Adaptive Band Extraction Based on Low Rank Approximated Nonnegative Tucker Decomposition for Anti-Friction Bearing Faults Diagnosis Using Measured Vibration Data

Haobin Wen [1], **Long Zhang** [2] **and Jyoti K. Sinha** [1,*]

[1] Dynamics Laboratory, School of Mechanical, Aerospace and Civil Engineering, The University of Manchester, Manchester M13 9PL, UK

[2] School of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, UK

[*] Correspondence: jyoti.sinha@manchester.ac.uk

Abstract: Condition monitoring and fault diagnosis are topics of growing interest for improving the reliability of modern industrial systems. As critical structural components, anti-friction bearings often operate under harsh conditions and are contributing factors of system failures. Efforts have been cast on bearing diagnostics under the sensor fusion and machine learning framework, whilst challenges remain open on the identification of incipient faults. In this paper, exploiting multi-way representations and decompositions of measured vibration data, a novel band separation method based on the factorization of spectrogram tensors using the low rank approximated nonnegative Tucker decomposition (LRANTD) is proposed and applied to identify detailed fault signatures from the spectral, temporal, and spatial dimensions, flexible for extracting multi-sensor features and multi-dimensional correlations. With the proposed method, informative frequency bands of the latent vibrational components can be automatically extracted, in accordance with the inherent temporal patterns that can be conveniently fed for spectral analysis and fault discrimination. Furthermore, an improved cross-spectrum can be calculated from multi-channel vibrations via LRANTD with enhanced fault features. Based on the real-world vibration data of the accelerated bearing life tests, detailed experimental studies and thorough comparisons to the conventional benchmarks have verified the effectiveness of the reported diagnostic methodology. The proposed method significantly improves the presence of the bearing frequency peaks distinctly over the background noises in the spectrum and hence improves the bearing defect detection process.

Keywords: anti-friction bearing; fault diagnosis; multi-sensor fusion; band extraction; tensor; Tucker decomposition; low-rank approximation; vibration analysis; low rank approximated nonnegative Tucker decomposition (LRANTD)

1. Introduction

In recent years, condition-based maintenance, involving minimally invasive condition measurements, massive data analytics, and decision making for complex engineering processes, has been considered a critical part of modern intelligent manufacturing for system optimization purposes [1], such as sustainable production and quality improvement. To prevent fatal system failures and allow scheduled maintenance, advanced prognostics and diagnostics are essential for the timely detection of anomalies.

Anti-friction bearings (AFBs) are critical structural components in various mechanical systems such as hydraulic pumps, induction motors, wind turbines, etc., but they are also sensitive parts that account for 45–55% of motor failures [2]. Without maintenance, defects in the AFBs could develop into critical disasters with considerable economic losses and physical injuries. In this context, condition monitoring and fault diagnosis for AFBs can be highly cost-efficient and of great practical significance.

Current approaches of bearing diagnosis can be divided into signal processing methods and machine learning (ML) methods. On the one hand, signal processing methods rely on a series of analytic and synthetic manipulations for processing signals to reveal specific fault characteristics from the time domain, frequency domain, and time-frequency (TF) domain. In this context, typical mainstream techniques include noise cancelation (e.g., time-synchronous average and minimum entropy deconvolution [3]), spectral analysis (e.g., envelope spectrum, spectral kurtosis [4], and cyclostationarity analysis [5,6]), time-frequency analysis (e.g., short-time Fourier transform [7]), sparsity diagnostics (e.g., sparse representations [8] and compressive sensing [9]), and matrix decomposition models for source separation (e.g., singular value decomposition (SVD) [10]). Among them, envelope analysis is one of the most successful approaches and plays an important role in bearing signal enhancement. On the other hand, machine learning methods attempt to identify the latent fault features from a data-driven perspective and learn the diagnostic model from the condition monitoring data in a reverse manner for state inference, such as support vector machines [11,12] and artificial neural networks [13,14]. Despite the successful applications in many data analysis challenges, ML methods are likely confronted with some last mile issues in industrial diagnostics, such as poor generalization to new machine sets and the data imbalance biased towards normal conditions. For both method categories, due to the complicated components and interferences inside the measured signals, challenges remain for robust solutions to bearing fault diagnosis.

With the advancement of sensing technology, more sensors are deployed for monitoring asset conditions and the resulting data exhibits in higher dimensions in the sensor networks. Based on the counteraction assumption of the random noise, sensor fusion and dimension reduction schemes are developed to enhance the inherent fault signatures from multiple measurements [15]. For example, Wodecki et al. [16] proposed a novel adaptive band selection method based on non-negative matrix factorization (NMF) of TF spectrograms and then further applied NMF to multi-sensor fusion for compound fault separation [17]. The diagnostic scope of NMF from the aspect of sparse regularization was studied for bi-spectrum decomposition by Liang et al. [18,19]. He et al. [20] introduced manifold learning for representing fault features in the TF domain by nonlinear mapping.

However, owing to the dimensional limitation of matrix representations, vectorization must be performed to further process higher-order (i.e., order larger than three, or multi-way) data, inevitably neglecting the inherent localized features such as multi-dimensional correlations. Hence, based on multilinear algebra, researchers resort to tensors for representing multi-way arrays and develop tensor decomposition (TD) for multidimensional data mining. As an extension to matrix component analysis, TD fills the gap from two-way analytics to multi-way component analytics. Examples of well-known TD models include the canonical polyadic decomposition or parallel factor analysis (CANDECOMP/PARAFAC decomposition or CPD in short) [21], the Tucker decomposition (aka the higher-order singular value decomposition (HOSVD) with orthogonal constraints) [22], nonnegative CPD (NCPD), nonnegative Tucker decomposition (NTD), etc. In [23], Cichocki and Mandic et al. gave a thorough review of TD and its applications for multi-way component analysis. A detailed guide of tensor methods for multi-sensor signal processing was summarized by Miron et al. [24]. Whereas in condition-based maintenance, limited applications of TD were found within current literature. Zhao et al. represented multiple intrinsic mode functions (IMF) into a Hankel tensor and applied tensor rank decomposition for bearing fault extraction [25]. The improved support tensor machine was proposed to classify the high-order condition monitoring data for rotating machinery diagnosis with imbalanced data [26]. Multi-dimensional noise reduction was introduced by Hu et al. to signal enhancement based on HOSVD for rotating machine diagnosis [27]. Sparse NCPD was applied to the impact feature extraction from the three-way TF representation of a single vibration signal [28] and to the fault identification using multi-channel measurements [29]. These examples have preliminarily demonstrated the

efficacy of the high-order tensor representations and the corresponding data mining techniques for fault diagnosis purposes. Following this line of research, some issues that limit the diagnostic performance of current TD models are notable and deserve necessary improvements:

(1) For CPD, the rank-one decomposition property requires the same number of lower ranks for each tensor mode (the dimensions of a tensor are often referred to as modes), which might be incompatible to express the inherent correlations where each mode has various number of latent components;

(2) Although different low-rank dimensions are allowed, the Tucker decomposition is often criticized for its poor uniqueness and the curse of dimensionality (i.e., the core tensor is dense whose entry number scales exponentially with the tensor order) [30];

(3) Diverse ways of representing vibration data as tensors can be found while a physically interpretable way is needed. To this end, the multi-dimensional correlation across different modes (such as the spectral, temporal, and spatial modes) might need more intuitive illustrations for diagnostic purposes.

To address the above issues, a novel diagnostic methodology has been formulated on the basis of time-frequency representations and the low rank approximated nonnegative Tucker decomposition [30,31], or LRANTD, a two-step model combining low-rank approximation (LRA) and NTD, which adaptively extracts the informative bands of the latent temporal patterns from the vibration signals. In the proposed method, a three-way TF tensor is generated from the single-channel or multi-channel vibration measurements by the introduced segmentation operation; it can be efficiently factorized by LRANTD into three factor matrices (embedding the latent features) and a core tensor (preserving the multi-dimensional correlations). The interpretation from the decomposed factors to the physical dimensions (including the spectral, temporal, and spatial dimensions) is elaborated in detail and the multi-dimension correlations preserved by the core tensor are illustrated graphically via the Hinton diagrams. Furthermore, the LRANTD model can be solved efficiently based on the iterative algorithms that integrate hierarchical alternating least squares (HALS) and multiplicative update (MU), which allow the flexible choice of the lower rank numbers for each tensor mode as compared to the CPD model and reduces the computational cost against the standard NTD.

For performance verification, the proposed method was applied to fault diagnosis using the bearing vibration data measured from multiple accelerated run-to-failure experiments [32]. Three typical health states including the incipient fault state, the maximal kurtosis state, and the severe fault state (later marked by Case A, B, and C) were analysed, whose results indicated the effectiveness of the proposed method for informative band extraction and fault signature enhancement. For the analysis using the single-channel signal, the diagnostic performance of LRANTD was verified through the comparative studies against a few successful benchmarks including smoothed envelope analysis, fast Kurtogram (FK), and empirical mode decomposition (EMD). For multi-channel sensor fusion, the cross spectrum computed based on LRANTD also outweighed the conventional approach based on the cross-power spectral density (CPSD).

This paper first provides the details of experimental vibration data and the observations made on their trend of the measured vibration RMS and Kurtosis values in Section 2. The principle of the LRANTD model is presented in Section 3, where the diagnostic methodology is also proposed for the defect detection in the anti-frication bearings. In Section 4, the LRANTD model is applied to identify fault signatures from the single-channel acceleration signal. Comparative studies are subsequently provided in Section 5 to show the effectiveness of the proposed method. Then, the cross-spectral analysis in Section 6 demonstrates the usage of the proposed LRANTD method for multi-sensors fusion to further reduce the background noises in the spectrum and hence improve the bearing fault detection process. Finally, the concluding remarks are given in Section 7. Basic notations are listed in Table 1.

Table 1. Basic notations and properties.

$\mathbf{0}$	The vector/matrix with all its entries being zero.
$U \geq 0$, u_r	Nonnegative matrix U and its r-th column vector.
$U_{(n)} \in \mathbb{R}^{I_n \times R_n}$, $u_r^{(n)} \in \mathbb{R}^{I_n \times 1}$	Mode-n matrix and its r-th column vector, $r = \{1, 2, \ldots, R_n\}$.
$\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$, $Y_{(n)} \in \mathbb{R}^{I_n \times \prod_{i \neq n} I_i}$	N-way arrays or tensor and its mode-n matricizaton.
$\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \cdots \times R_N}$, $G_{(n)} \in \mathbb{R}^{R_n \times \prod_{i \neq n} R_i}$	Core tensor and its mode-n matricizaton.
$(R_1, R_2, \cdots, R_N)$	Low-rank dimensions for NTD, $R_n < I_n$
$\mathbb{I}$	The index set of positive integers, $\mathbb{I} = \{1, 2, \ldots, N\}$.
$\mathcal{Y}' = \mathcal{Y} \times_n U_{(n)}$	Mode-n product, $\mathcal{Y}' \in \mathbb{R}^{I_1 \times \cdots I_{n-1} R_n I_{n+1} \cdots \times I_N}$
$K = P \otimes Q$	Kronecker product of $P \in \mathbb{R}^{I_1 \times I_2}$ and $Q \in \mathbb{R}^{R_1 \times R_2}$ yields $K = [p_{i_1 i_2} Q] \in \mathbb{R}^{I_1 R_1 \times I_2 R_2}$
$\circledast$, $\dfrac{U}{U'}$	Element-wise multiplication, division
$\mathcal{G} \times_{n \in \mathbb{I}} U_{(n)}$	is equal to $\mathcal{G} \times_1 U_{(1)} \cdots \times_N U_{(N)}$. If $\mathcal{Y} = \mathcal{G} \times_{n \in \mathbb{I}} U_{(n)}$, then $Y_{(n)} = U_{(n)} G_{(n)} (U_{(N)} \otimes \cdots U_{(n+1)} \otimes U_{(n-1)} \cdots \otimes U_{(1)})^{\mathrm{T}}$.
$y_i \in \mathbb{R}^L$, $i = 1, \ldots, M$	The i-th channel acceleration with L points in length, within M sensors measurement.
$y_{i,j} \in \mathbb{R}^{L_s}$, $y_{i,j}(j)$	The j-th segment resampled from the i-th channel signal, the j-th entry of $y_{i,j}$.
$Y_{i,j} \in \mathbb{R}^{I_1 \times I_2}$, $Y(i_1, i_2)$	Spectrogram matrix with I_1 frequency bins and I_2 time points, (i_1, i_2)th-entry of Y
S, L_s, I_3	The number of data segments resampled from one signal, the length of each segmentation, the total number of segments

2. Experimental Rig and Measured Vibration

The bearing vibration data collected from the accelerated life tests available in the XJTU-SY bearing database [32] are beneficial for the diagnostic performance validation under deteriorating health conditions and thus used in this study. A simple schematic of the experimental bearing test rig is shown in Figure 1. It mainly consists of an alternating current (AC) induction motor, an intermediate shaft supported through two ball (LDK UER204) bearings, and a hydraulic loading system to apply the load on the bearing in the horizontal direction. The vibration database [32] are composed of a total of 15 experimental groups based on the three different sets of operating conditions of the running speeds and load levels. Further details on the rig, conducted experiments, and the vibration data can be found in the study by Wang et al. [32].
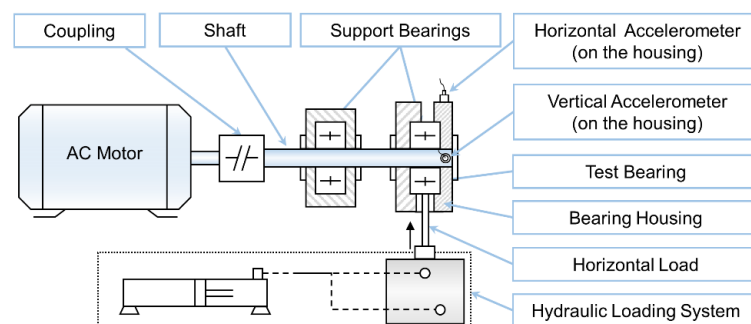
Figure 1. Schematic of the experimental rig with accelerometers locations.

In this work, the measured vibration acceleration data for the 'Bearing 1_1' [32] are considered and analysed to demonstrate the usefulness of the proposed method. The vibration acceleration data for the 'Bearing 1_1' were collected by Wang et al. [32] with the shaft speed of 35 Hz and a hydraulic load of 12 kN being applied to the bearing. The 'Bearing 1_1' data set contains 123 data records, each of which is acquired at an interval of 1 Minute and encompasses 1.28 s of vibration acceleration measured on the bearing in the vertical and horizontal directions. These 1.28 s data were collected at the sampling frequency, f_s of 25,600 Hz. The 123 data records means that the run-to-failure experiment lasted for two hours and 3 min until the stopping criteria was met (i.e., the maximum acceleration amplitude exceeding 10 times that in the normal state). The bearing specifications and the calculated faults characteristic frequencies are listed in Table 2.

Table 2. Test bearing specifications (f_r:shaft speed, f_c: cage frequency, f_b: ball spin frequency (BSF), f_o: ball pass frequency outer (BPFO), f_i: bass pass frequency inner (BPFI).).

Parameter	Value	Characteristics	Value
Mean diameter	34.55 mm	f_r	35 Hz
Ball diameter	7.92 mm	f_c	13.49 Hz
Number of balls	8	f_b	72.33 Hz
Contact angle	0°	f_o	107.91 Hz
Load rating	12 kN	f_i	172.09 Hz

To better observe the degradation process of the test bearing, the trends of overall bearing vibration in both directions are presented in Figure 2 based on two different indicators, i.e., Root Mean Square (RMS) and Kurtosis, for each acceleration record. From the normal state to severe-fault state, the RMS indicator follows an increasing trend and surges since the 78th minute, whereas Kurtosis is more fluctuant and rises much earlier from around the 66th minute, then peaking at the 78th. Based on this observation, as an impulsiveness indicator, Kurtosis is justified for recognizing the degradation of AFBs. By comparisons, both indicators have higher amplitudes in the horizontal direction than in the vertical direction, which is due to the hydraulic loading on the horizontal axis.

Based on the vibration trend, three typical timestamps indicating incipient-fault state, the maximum Kurtosis state, and the severe-fault state were selected for further analysis, i.e., the 66th, 78th, and 80th records, as denoted by the capital letters 'A', 'B', and 'C' in Figure 2.
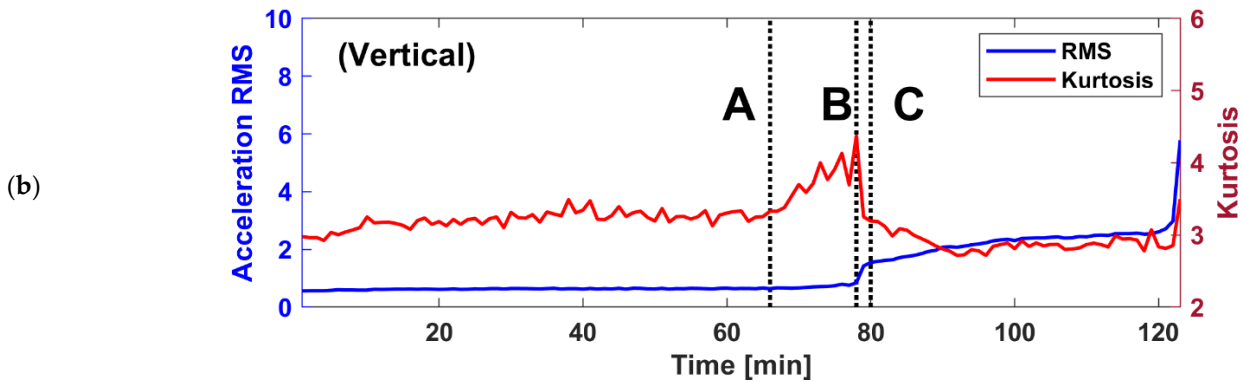
(**a**)

(b)
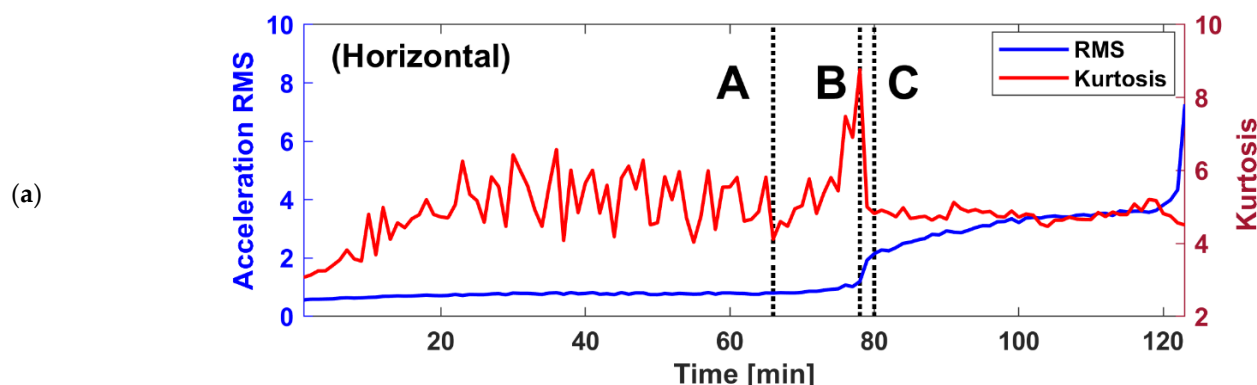
Figure 2. Trends of the overall bearing vibration by the acceleration RMS values and Kurtosis: (**a**) the horizontal-channel vibration and (**b**) the vertical-channel vibration.

3. Proposed Method

The concept of the LRANTD method is presented first and then the diagnostic approach using this concept is proposed.

3.1. Nonnegative Tucker Decomposition Based on Low-Rank Approximation

This section is divided into a number of steps to achieve the optimised LRANTD model to aid the understanding.

3.1.1. Standard Nonnegative Tucker Decomposition (NTD)

The tensor decomposition theory was first developed for factor analysis and data mining of multi-way psychometric data [22]. As an extension of matrix component analysis, the Tucker decomposition, or higher-order SVD of an N-way real-value tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$, can be defined as the mode-n products of the core tensor and the factor matrices [22],

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_{(1)} \times_2 \boldsymbol{U}_{(2)} \cdots \times_N \boldsymbol{U}_{(N)} \tag{1}$$

where $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \cdots \times R_N}$ is the core tensor and $\boldsymbol{U}_{(n)} \in \mathbb{R}^{I_n \times R_n}, n = 1, 2, \cdots, N$, represent N mode-n matrices, respectively. The multilinear rank of the Tucker decomposition can be denoted by the N-tuple $(R_1, R_2, \cdots, R_N)$. Note that $R_n \leq I_n$, a lot of multi-dimensional data, e.g., RBG images and multi-sensor signals, could be well represented by low-rank representations. To that end, the HOSVD method is the workhorse for computing such representations for unconstrained Tucker decomposition. Similar to the column-wise orthogonality of the singular matrices in SVD, the mode-n matrices of the HOSVD are also semi-orthogonal, i.e., $\boldsymbol{U}_{(n)}^T \boldsymbol{U}_{(n)} = \boldsymbol{I}$. Due to the rotational indeterminacy, the unconstrained Tucker decomposition can be restricted by the lack of uniqueness and the curse of dimensionality [30].

For the better uniqueness, Nonnegative Tucker Decomposition (NTD) is put forward by relaxing the orthogonality requirement in Tucker decomposition to approximate the original tensor via a set of nonnegative factors, i.e., $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{U}_{(n)}$ are subject to elementwise nonnegativity constraint [33],

$$\widehat{\boldsymbol{\mathcal{Y}}} = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_{(1)} \times_2 \boldsymbol{U}_{(2)} \cdots \times_N \boldsymbol{U}_{(N)}, \qquad \text{s.t.} \, \boldsymbol{\mathcal{G}}, \boldsymbol{U}_{(n)} \geq \boldsymbol{0} \tag{2}$$

or equivalently in a matricization form,

$$\widehat{\boldsymbol{Y}}_{(n)} = \boldsymbol{U}_{(n)} \big(\boldsymbol{U}_{(N)} \otimes \boldsymbol{U}_{(N-1)} \cdots \otimes \boldsymbol{U}_{(n+1)} \otimes \boldsymbol{U}_{(n-1)} \otimes \cdots \boldsymbol{U}_{(1)} \big) \boldsymbol{G}_{(n)}^{\mathrm{T}} = \boldsymbol{U}_{(n)} \boldsymbol{V}_{(n)}^{\mathrm{T}} \tag{3}$$

where $\widehat{\boldsymbol{Y}}_{(n)} \in \mathbb{R}^{I_n \times \prod_{i \neq n} I_i}$ and $\boldsymbol{G}_{(n)} \in \mathbb{R}^{R_n \times \prod_{i \neq n} R_i}$ are the mode-n matricizations of tensor $\widehat{\boldsymbol{\mathcal{Y}}}$ and $\boldsymbol{\mathcal{G}}$, and $\boldsymbol{V}_{(n)}$ is an auxiliary notation defined by,

$$V_{(n)} = (U_{(N)} \otimes \cdots U_{(n+1)} \otimes U_{(n-1)} \cdots \otimes U_{(1)})G_{(n)}^{\mathrm{T}} \tag{4}$$

The above nonnegative approximation is achieved by minimizing the approximation error based on the squared Frobenius norm, as in Formula (5) that further unfolds into matrix form, which can be solved effectively by a few alternating algorithms in the coordinate descent family.

$$\min_{\mathcal{G}, U_{(n)} \geq \mathbf{0}} \frac{1}{2} \|\mathcal{Y} - \widehat{\mathcal{Y}}\|_F^2 = \frac{1}{2} \|Y_{(n)} - U_{(n)} V_{(n)}^{\mathrm{T}}\|_F^2, n \in \{1, \dots, N\} \tag{5}$$

Two main attributes are endowed by the imposed nonnegativity: (1) only additive operations are allowed without subtractions, thus enhancing the interpretability in feature engineering applications; (2) since many entries of the core tensor becomes zeros, the resulting factors are often sparse, exhibiting parts-based features and alleviating the curse of dimensionality.

3.1.2. The Low Rank Approximated NTD (LRANTD) Model

Many efforts have been cast to reduce the computational cost in NTD for the low-rank approximation (LRA) of large-scale data, especially the computation of the gradients for the tensor-based cost function. Given its flexibility to other well-established LRA techniques and noise robustness, the low rank approximated nonnegative Tucker decomposition, or LRANTD, developed by Zhou, Cichocki et al. [30,31] is introduced as the key tensor factorization model in this work, which consists of two main steps:

(1) LRA. Perform unconstrained Tucker decomposition as in (1) using HOSVD to obtain an approximation of the original tensor, $\mathcal{Y}$, which gives

$$\mathcal{Y} \approx \widetilde{\mathcal{Y}} = \widetilde{\mathcal{G}} \times_1 \widetilde{U}_{(1)} \times_2 \widetilde{U}_{(2)} \cdots \times_N \widetilde{U}_{(N)} \tag{6}$$

and the mode-n unfolding of the LRA tensor, $\widetilde{\mathcal{Y}}$,

$$\widetilde{Y}_{(n)} = \widetilde{U}_{(n)} \widetilde{G}_{(n)} (\widetilde{U}_{(N)} \otimes \cdots \widetilde{U}_{(n+1)} \otimes \widetilde{U}_{(n-1)} \cdots \otimes \widetilde{U}_{(1)})^{\mathrm{T}} = \widetilde{U}_{(n)} \widetilde{V}_{(n)}^{\mathrm{T}} \tag{7}$$

(2) NTD. Minimize the new cost functions with the nonnegativity constraints on the factors as in (8). Note that given the tensor order N, a total of N unfolding forms from $\widetilde{\mathcal{Y}}$ are defined in (7) and the sequence of N minimization problems should be calculated for each factor matrix, $U_{(n)}$.

$$\min_{\mathcal{G}, U_{(n)} \geq \mathbf{0}} D_{NTD} = \frac{1}{2} \|\widetilde{\mathcal{Y}} - \widehat{\mathcal{Y}}\|_F^2 = \frac{1}{2} \|\widetilde{U}_{(n)} \widetilde{V}_{(n)}^{\mathrm{T}} - U_{(n)} V_{(n)}^{\mathrm{T}}\|_F^2, n \in \{1, \dots, N\}. \tag{8}$$

Without frequent visits to the original big tensor, the LRA step not only significantly reduces the computation complexity for the subsequent optimizations, but also suppresses the interference in the raw data tensor [30]. Rather than directly implementing NTD of a tensor, LRANTD combines HOSVD with NTD to provide a trade-off between efficiency and approximation error.

3.1.3. Optimization Algorithms

Both (5) and (8) are known as nonnegative least squares problems, which can be solved by a few well-known first-order methods from the area of NMF, such as multiplicative update (MU) [34], accelerated proximal gradient (APG) [35], the hierarchical alternating least squares (HALS) algorithm [36], active set methods [37], etc. By a crafty choice of learning step, these methods avoid the searching process for an appropriate step size in gradient descent and can be further extended to various tensor decomposition models.

For ease of comprehension, the following optimization algorithm is provided for three-way tensor decomposition, i.e., $N = 3$. The lower rank dimensions of the decomposition are given as (R_1, R_2, R_3) ideally in accordance with the number of latent components. And the desired variables $U_{(n)}, n \in \{1, 2, 3\}$ and $\mathcal{G}$ are initialized randomly within

the interval of 0 and 1. See Figure 3 for an intuitive illustration of the three-way tensor decomposition and consult [30] for further details of extensions to higher-order tensor ($N > 3$).
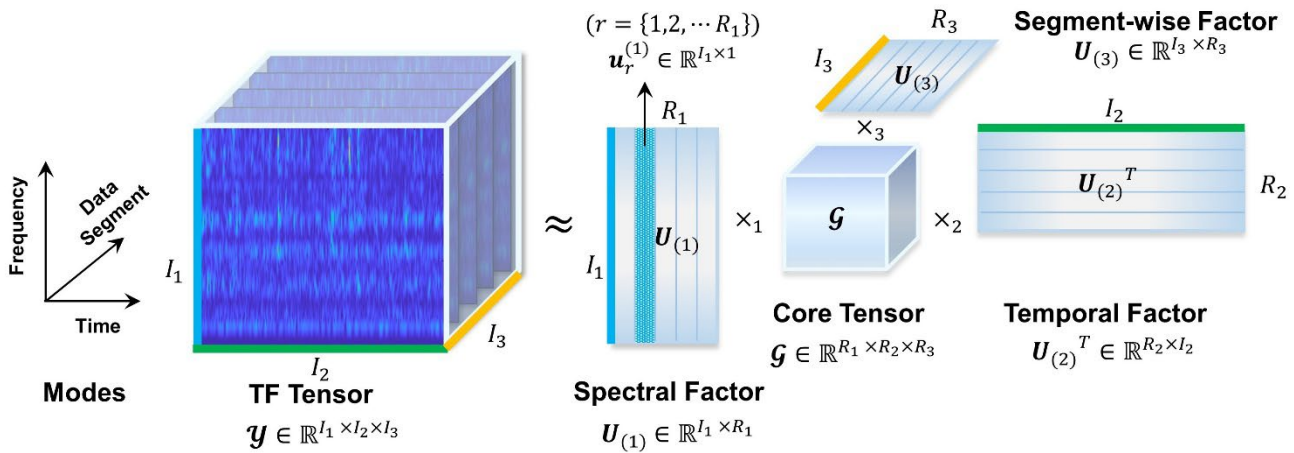

Figure 3. An illustration of the tensor decomposition model for TF tensor factorization.

- The iterative update rule for mode-n (factor) matrices $\boldsymbol{U}_{(n)}$

The LRANTD of the three-way tensor is computed based on the matricization operation undertaken for each tensor mode. For example, $\boldsymbol{Y}_{(1)} \in \mathbb{R}^{I_1 \times I_2 I_3}$ is the mode-1 matricization form of the tensor, $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. Taking the LRA step in (7), the approximation term $\widetilde{\boldsymbol{Y}}_{(1)}$ can be obtained and the mode-1 factor matrix $\boldsymbol{U}_{(1)} \in \mathbb{R}^{I_1 \times R_1}$ can be calculated via the NTD step in (8).

To generalize, for the mode-n matrix, $\boldsymbol{U}_{(n)} \in \mathbb{R}^{I_n \times R_n}, n \in \{1, 2, 3\}$, the low-rank approximation $\widetilde{\boldsymbol{Y}}_{(n)} \in \mathbb{R}^{I_n \times \prod_{i \neq n} I_i}$ is first attained in the LRA step, and the NTD step applies the HALS algorithm to update only one column of $\boldsymbol{U}_{(n)}$ at a time subsequently. By defining the residue term as

$$\widetilde{\boldsymbol{Y}}_{(n)}^r = \widetilde{\boldsymbol{Y}}_{(n)} - \sum_{i \neq r} \boldsymbol{u}_i^{(n)} \boldsymbol{v}_i^{(n)\mathrm{T}}, r \in \{1, \cdots, R_n\}, n \in \{1, 2, 3\}, \tag{9}$$

Equation (8) expands to a series of minimization problems that sequentially optimize $\boldsymbol{u}_r^{(n)}$ while fixing the remaining entries and variables, as given by

$$\min_{\boldsymbol{u}_r^{(n)} \geq \boldsymbol{0}} D_{NTD} = \frac{1}{2} \left\| \widetilde{\boldsymbol{Y}}_{(n)}^r - \boldsymbol{u}_r^{(n)} \boldsymbol{v}_r^{(n)\mathrm{T}} \right\|_F^2, r \in \{1, \cdots, R_n\}, \tag{10}$$

where $\boldsymbol{u}_r^{(n)} \in \mathbb{R}^{I_n \times 1}$ and $\boldsymbol{v}_r^{(n)} \in \mathbb{R}^{(\prod_{i \neq n} I_i) \times 1}$ are the r-th column vectors of $\boldsymbol{U}_{(n)}$ and $\boldsymbol{V}_{(n)}$ with $r \in \{1, 2, \cdots, R_n\}$. Note that $\boldsymbol{v}_r^{(n)}$ always serves as an auxiliary term and is decoupled from $\boldsymbol{U}_{(n)}$ as in (4).

Then, the gradient of D_{NTD} with respect to $\boldsymbol{u}_r^{(n)}$ can computed as

$$\frac{\partial D_{NTD}}{\partial \boldsymbol{u}_r^{(n)}} = \boldsymbol{u}_r^{(n)} \boldsymbol{v}_r^{(n)\mathrm{T}} \boldsymbol{v}_r^{(n)} - \widetilde{\boldsymbol{Y}}_{(n)}^r \boldsymbol{v}_r^{(n)} \tag{11}$$

Assigning zero to the gradient and imposing the nonnegativity constraints further gives

$$\boldsymbol{u}_r^{(n)} = \max\{0, \widetilde{\boldsymbol{Y}}_{(n)}^r \boldsymbol{v}_r^{(n)} (\boldsymbol{v}_r^{(n)\mathrm{T}} \boldsymbol{v}_r^{(n)})^{-1}\} \tag{12}$$

Substituting the residue in (7) further gives the following update rule,

$$\boldsymbol{u}_r^{(n)} \leftarrow \boldsymbol{u}_r^{(n)} + \frac{1}{\boldsymbol{v}_r^{(n)\mathrm{T}} \boldsymbol{v}_r^{(n)}} \max\{0, \widetilde{\boldsymbol{Y}}_{(n)} \boldsymbol{v}_r^{(n)} - \boldsymbol{U}_{(n)} \boldsymbol{V}_{(n)}^{\mathrm{T}} \boldsymbol{v}_r^{(n)}\} \tag{13}$$

Thus, by updating a total of R_n columns, the mode-n matrix $\boldsymbol{U}_{(n)}$ can be obtained. Consequently, the three factor matrices, $\boldsymbol{U}_{(1)}$, $\boldsymbol{U}_{(2)}$, and $\boldsymbol{U}_{(3)}$ can be successively computed, as in Figure 3.

- The iterative update rule for core tensor $\boldsymbol{\mathcal{G}}$

Next, the MU algorithm can be applied in the NTD step to update the core tensor, $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$. From (6) and (8), the gradient of the cost function with respect to $\boldsymbol{\mathcal{G}}$ can be computed as in

$$\frac{\partial D_{NTD}}{\partial \boldsymbol{\mathcal{G}}} = \widehat{\boldsymbol{\mathcal{Y}}} \times_{n \in \mathbb{I}} \boldsymbol{U}_{(n)}^{\mathrm{T}} - \widetilde{\boldsymbol{\mathcal{Y}}} \times_{n \in \mathbb{I}} \boldsymbol{U}_{(n)}^{\mathrm{T}} \tag{14}$$

where $\mathbb{I}$ reduces to $\mathbb{I} = \{1, 2, 3\}$ as $N = 3$. By taking the step size in the gradient descent as $\eta = \frac{\boldsymbol{\mathcal{G}}}{\widehat{\boldsymbol{\mathcal{Y}}} \times_{n \in \mathbb{I}} \boldsymbol{U}_{(n)}^{\mathrm{T}}}$, the additive update rule reduces to pure multiplicative operations, providing an efficient computation algorithm for the core tensor,

$$\boldsymbol{\mathcal{G}} \leftarrow \boldsymbol{\mathcal{G}} \circledast \frac{max\{0, \ \widetilde{\boldsymbol{\mathcal{Y}}} \times_{n \in \mathbb{I}} \boldsymbol{U}_{(n)}^{T}\}}{max \ \{0, \ \widehat{\boldsymbol{\mathcal{Y}}} \times_{n \in \mathbb{I}} \boldsymbol{U}_{(n)}^{T}\}} = \boldsymbol{\mathcal{G}} \circledast \frac{max\{0, \ \widetilde{\boldsymbol{\mathcal{G}}} \times_{n \in \mathbb{I}} \widetilde{\boldsymbol{U}}_{(n)} \times_{n \in \mathbb{I}} \boldsymbol{U}_{(n)}^{T}\}}{max \ \{0, \ \boldsymbol{\mathcal{G}} \times_{n \in \mathbb{I}} \boldsymbol{U}_{(n)} \times_{n \in \mathbb{I}} \boldsymbol{U}_{(n)}^{T}\}} \tag{15}$$

where $\circledast$ denotes the element-wise multiplication.

3.2. Diagnostic Scope of the LRANTD Model

In this section, the diagnostic application of the LRANTD model is illustrated by analysing multi-way vibrational data. The fundamentals of the tensor-based diagnostics are explained with practical considerations. Then, the methodology flowchart for fault diagnosis of AFBs is presented.

3.2.1. Data Pre-Processing: Moving Segmentation

Given the input vibration data, either from single-channel or multi-channel sensor measurements, the pre-processing operation termed moving segmentation is introduced to the raw signals to generate multi-way arrays that allow the subsequent multilinear data mining. Given a signal from i-th channel measurement, $\boldsymbol{y}_i \in \mathbb{R}^L, i \in \{1, \cdots, M\}$, where M denotes the total number of sensor channels and L represents the signal length in data points, a total of S segments can be resampled from $\boldsymbol{y}_i$ by a moving-window fashion. By segmentations, the j-th segment, $\boldsymbol{y}_{i,j} \in \mathbb{R}^{L_s}$, of the source signal, $\boldsymbol{y}_i$, is defined by

$$\boldsymbol{y}_{i,j} = [y_i(j), y_i(j+1), \cdots, y_i(L_s + j - 1)], \qquad j = 1, 2, \cdots, S \tag{16}$$

where L_s $(L_s \leq L - S + 1)$ controls the length of each segment and is suggested to be an integral multiple of the shaft-rotation period. The step size of the segmentation is set at 1 such that the adjacent segments are highly overlapped and self-similar. The above settings are important to enhance the inherent components in the signals as the random interference could be suppressed by the subsequent decomposition in ideal conditions.

3.2.2. Tensorization: Generating the Time-Frequency Tensor

The vibration signals from a localized fault in the AFBs are known to be non-stationary, i.e., the statistical behaviours of the signals are inconstant, and are indeed pseudo-cyclostationary as pointed out by Antoni and Randall [38]. As effective non-stationarity analysis tools, time-frequency (TF) representations provide a thorough depiction of signal variations by balancing the time-domain and frequency-domain features. On each one-way signal segment, $y_{i,j}$, the short-time Fourier transform (STFT) is first performed, as given by

$$\text{STFT}(i_1, i_2) = \sum_{\tau=0}^{L_w-1} y_{i,j}(i_2 + \tau) w(\tau) e^{-j2\pi\tau i_1/I_1} \tag{17}$$

where $i_1 = 0, 1, \cdots, I_1 - 1$ denotes the frequency bin with I_1 bins in total, $i_2 = 0, 1, \cdots, I_2 - 1$ is the time bin for I_2 bins in total, and $w(\tau)$ is the window function of length L_w. In practice, STFT can be efficiently computed by implementing fast Fourier transform for each windowed frame. Then, by taking the absolute value of the STFT result, the corresponding two-way TF spectrogram, $Y_{i,j}$, is attained with each entry defined as

$$Y_{i,j}(i_1, i_2) = |\text{STFT}(i_1, i_2)| \tag{18}$$

Consequently, by arranging the adjacent TF spectrograms, one can construct a three-way TF tensor from single or multi-sensor measurements by absorbing a total of $I_3 = MS$ spectrograms into the last mode. For illustration, Figure 3 provides an example of the TF tensor, $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, spanned by three physical modes, i.e., the frequency mode, the time mode, and the data-segment mode.

Extensive research utilized multiple sensors for vibration measurements under the assumption that comprehensive information, including the fault features, could be obtained, while it might be more often a case where a single measurement is analysed without detailed clarification. The motivations of tensorization the raw vibrational signals are two folds:

- Tensors are natural representations for multi-way data beyond the dimensional restriction of matrix analysis, which makes possible the data mining for higher-order connections among multi-channel vibration measurements;
- The time-frequency signatures of localized faults in AFBs are likely to be low-rank, as inherited by the repetitive occurrences of the impact-like structures in the TF spectrogram. It is therefore straight-forward to represent the signature-of-interest through low-rank learning models.

3.2.3. Tensor Decomposition by LRANTD

After selecting the low-rank dimensions as (R_1, R_2, R_3) for decomposition, one can further apply the presented LRANTD algorithm to the TF tensor for multilinear data mining and dimensionality reduction, which returns four factors with interpretable physical meaning (recall Figure 3 for illustrations):

(1) Based on the proposed method, the latent frequency bands of the TF spectrograms can be automatically extracted without subjective selection. The mode-1 matrix $\boldsymbol{U}_{(1)} \in \mathbb{R}^{I_1 \times R_1}$ is called the spectral factor that preserves the R_1 dominant frequency bands or spectral subfactors (SS) in its columns, $\boldsymbol{u}_r^{(1)} \in \mathbb{R}^{I_1}$. Each SS is normalized within the interval $[0, 1]$ to indicate the informative bands, which in turn rescales the corresponding temporal pattern to have meaningful amplitude;

(2) The mode-2 matrix $\boldsymbol{U}_{(2)} \in \mathbb{R}^{I_2 \times R_2}$ serves as the temporal factor that adaptively identifies the R_2 principal encoding patterns or the temporal subfactors (TS) embedded in the corresponding frequency bands. Essentially, these temporal components are the envelope profiles that reveal the variation in TF energy within the informative bands as extracted in the corresponding SS;

(3) Similarly, the mode-3 matrix $\boldsymbol{U}_{(3)} \in \mathbb{R}^{I_3 \times R_3}$ clusters I_3 data segments into R_3 segment-wise subfactors (SWS). Each SWS gives a hint about its relevance to the segments and sensor channels (or spatial locations). For multi-sensor measurements, the multiple SWS represent the dominating spatial components;

(4) Therefore, the entries of the core tensor $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ reflects the correlations across the spectral, the temporal, and the spatial dimensions, which can be visualized intuitively. In other words, the core tensor provides direct information about the spectral locations of the linked temporal patterns.

Note that the multilinear rank is always smaller than the input dimension $(R_n \ll I_n)$, the presented LRANTD also provides an efficient multi-sensor fusion scheme that greatly reduces the computational storage, from $O(\prod_n I_n)$ to $O(\sum_n I_n R_n + \prod_n R_n)$, which is of practical value in modern manufacturing systems with sensor networks and cloud-based computing for condition monitoring.

3.2.4. Evaluation Based on Spectrum Inspection

Finally, based on the separated TS, Kurtosis values are used to assess their impulsiveness and the frequency spectra are calculated to identify the characteristic frequencies of bearing faults. Overall, the proposed the diagnostic methodology based on LRANTD can be summarized into the flowchart in Figure 4.
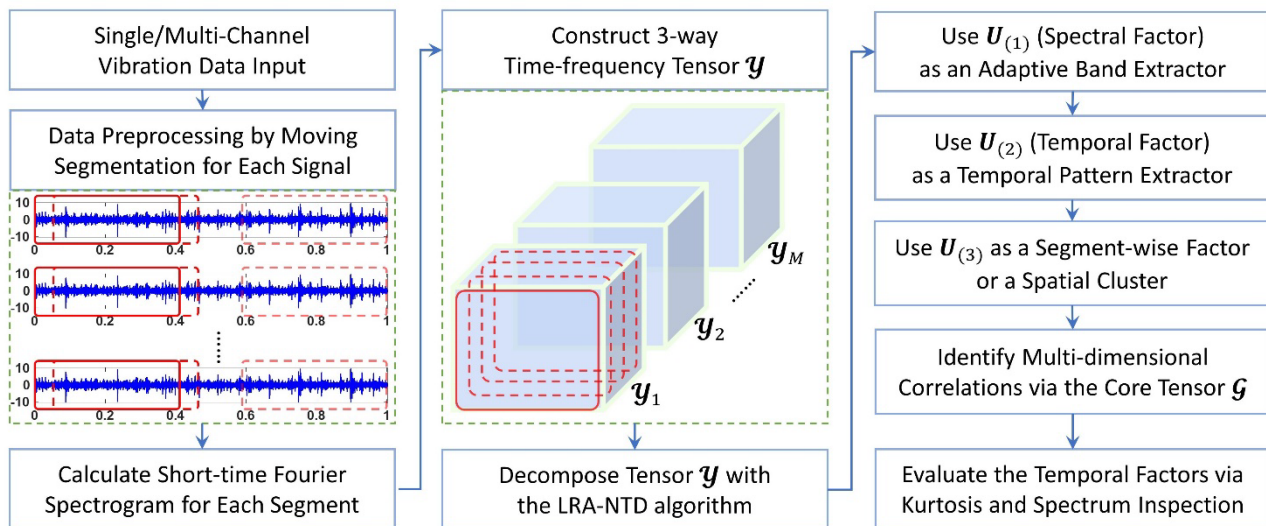

Figure 4. Flowchart of the diagnostic methodology.

4. Data Analysis and Case Studies

In this section, the proposed method was first applied to single-channel vibration (i.e., $M = 1$) for AFB diagnosis under three typical health conditions. The diagnostic results based on LRANTD were elaborated. Further comparative studies against a few benchmark models were provided in Section 5.

4.1. Case A

It can be seen from the vibration trend in Figure 2 that the Kurtosis at timestamp A just begins to climb during the incipient stage of bearing degradation. Provided that the hydraulic load was applied from the horizontal axis, the horizontal acceleration was selected for further analysis. In Figure 5a, a few irregular impacts appear in the acceleration signals, where judgment can hardly be made about the health state of the test bearing. In the amplitude spectrum in Figure 5b, the spectral peak points at around 1.2 kHz with

multiple bands rich in energy. Then, the signal was high-pass filtered with the cut-off frequency at 1 kHz. After performing STFT with the 63-point Hanning window (using consistent settings hereafter), the TF spectrogram of the filtered signal is still complex as shown in Figure 6a, where some impacts signatures are slightly clearer in the bands centred at 1.2 kHz, 5.5 kHz, 6.5 kHz, and 11 kHz. Hence, the early fault signal is hardly recognizable due to considerable background contaminations.
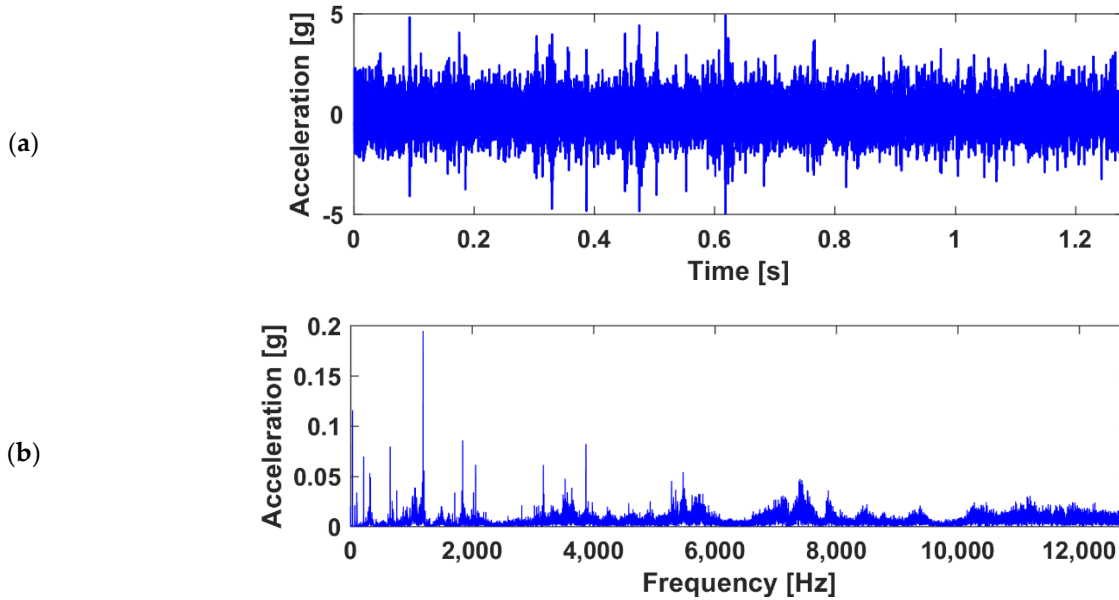
(**a**)

(**b**)

Figure 5. (**a**) The horizontal acceleration signal and (**b**) its amplitude spectrum in Case A.

(**a**)

(**b**)
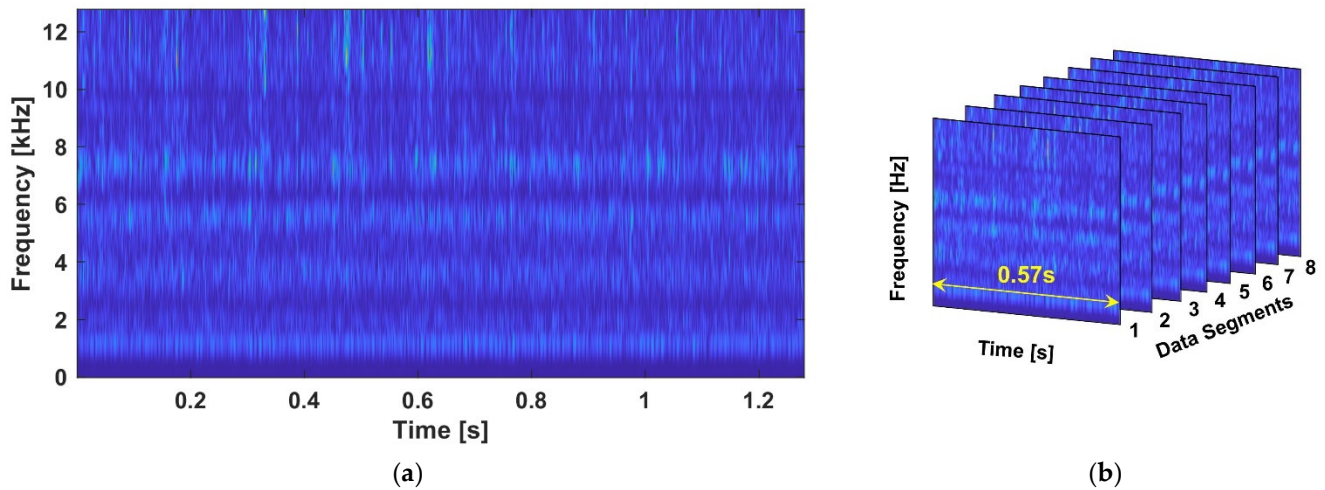
Figure 6. Time–frequency representations of the analysed data: (**a**) STFT spectrogram. (**b**) Time–frequency tensor reorganized from the spectrogram segments, $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{256 \times 14640 \times 8}$.

Setting the segmentation length to 20 times of rotation periods, i.e., $20 f_s / f_r \approx 14640$ (0.57 s), the frequency bins of STFT as 256, and the number of segments as $S = 8$ ($I_3 = 1S = 8$), the TF tensor in Case A, $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{256 \times 14640 \times 8}$, could be constructed according to the procedure in Section 3, as shown in Figure 6b.

As the number of latent features cannot be priorly known and the source determination remains a major challenge in current research, the low-rank dimensions (R_1, R_2, R_3) were selected based on trial and error as $(4, 4, 2)$. Then, LRANTD was performed on $\boldsymbol{\mathcal{Y}}$ for multi-way data mining based on the algorithms combining the HALS and MU update

rules, which returned a factor set including the spectral factor $\boldsymbol{U}_{(1)} \in \mathbb{R}^{256 \times 4}$, the temporal factor $\boldsymbol{U}_{(2)} \in \mathbb{R}^{14640 \times 4}$, the segment-wise factor $\boldsymbol{U}_{(3)} \in \mathbb{R}^{8 \times 2}$, and the core tensor $\mathcal{G} \in \mathbb{R}^{4 \times 4 \times 2}$, as in Figure 7.
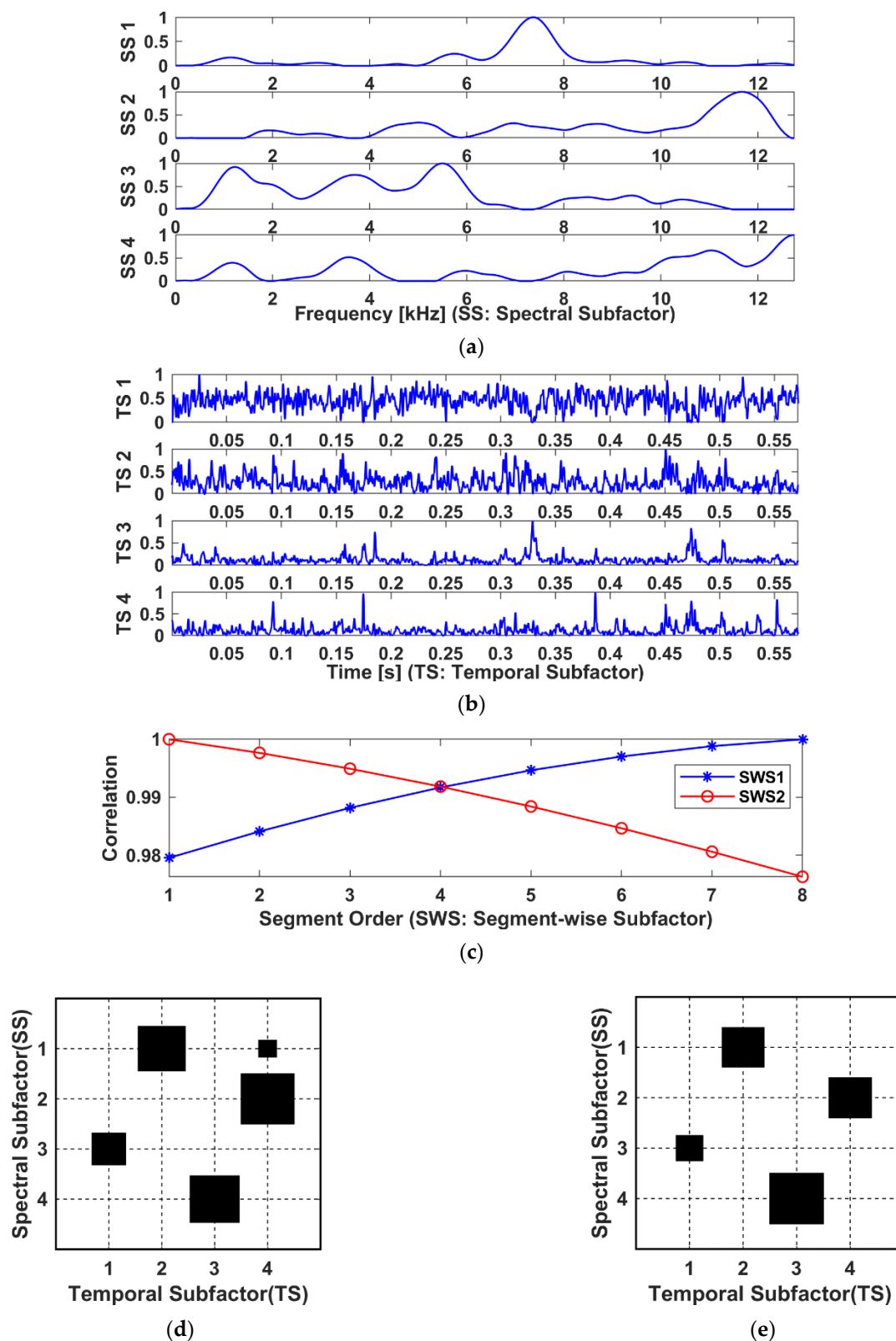

Figure 7. Factors of LRANTD: (**a**) the Spectral Factor, (**b**) the Temporal Factor, consisting of 4 temporal subfactors with kurtosis of 2.98, 4.1, 15.7, and 13.3, respectively, (**c**) the Segment-wise Factor

with 2 subfactors, and (**d**) the Core Tensor represented via its frontal slice w.r.t. SWS1, and (**e**) that w.r.t. SWS2.

(1) Note that each SS in $\boldsymbol{U}_{(1)}$ had been normalized to intuitively reveal the spectral weight of the correlated TS, $\boldsymbol{U}_{(1)}$ thereby adaptively extracted multiple dominant bands from the TF tensor, as shown in Figure 7a;

(2) For the temporal factor $\boldsymbol{U}_{(2)}$, it was found that the TS 3 and TS 4 exhibit apparent impulsive features and have relatively higher kurtosis, as shown in Figure 7b:

(3) Each SWS in the segment-wise factor $\boldsymbol{U}_{(3)}$ was found to have a correlation of over 97% to all data segments. This is because all data segments are sampled from the same channel measurement and hence resemble one another;

(4) Unlike the NMF factors, the mode matrices of LRANTD are implicitly linked with one another due to the flexible choice of lower rank dimensions. Such interconnections are embedded in entries of the core tensor, $\boldsymbol{\mathcal{G}}$. For visualization, the Hinton diagram of each frontal slice of the core tensor is shown in Figure 7d,e. For example, from Figure 7d, TS 4 was found to have a more distinctive correlation with SS 2 against SS 1, indicating TS 4 is more likely to contain high-frequency components as inferred from Figure 7a.

Then, the amplitude spectrum of each TS was computed and that with respect to the 4th TS is shown in Figure 8. Both BPFO and its 2nd harmonics are clear and dominant, as marked by the green dash-dot line. Tracing the spectral property of this temporal pattern (TS 4) according to the Hinton diagram in Figure 7, it was found that the dominant frequency band was extracted by the spectral factor in SS 2, indicating that the band from 11 to 12 kHz is likely to comprise the outer-race fault components.
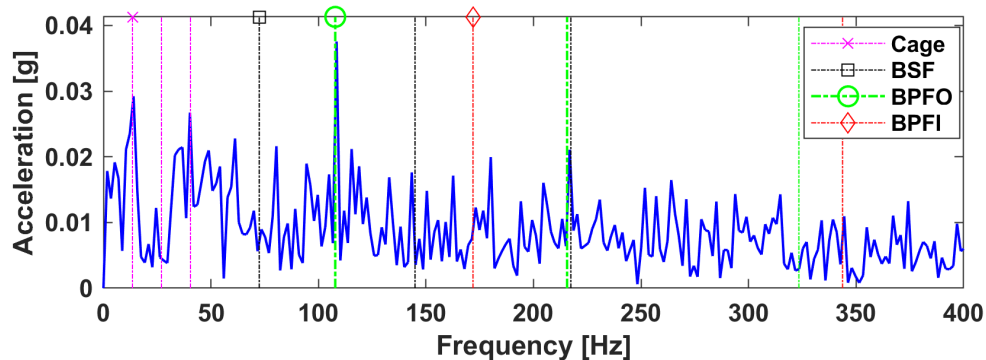

Figure 8. Case A result via LRANTD: The amplitude spectrum of TS 4 with clear outer-race fault component.

Therefore, the above result has shown the efficacy of the proposed diagnostic methodology based on LRANTD for informative bands extraction, which is further verified by comparative studies in Section 5.

4.2. Case B

In Case B, the bearing vibration data with the maximum kurtosis (at the 78th minute) were analysed. From the horizontal acceleration signal in Figure 9a, it can be recognized that more impacts appear in repetitive but uneven patterns as compared to the Case A signal (Figure 5a). In the corresponding frequency spectrum, the dominant bands remain similar to the early stage as in Figure 5b, whereas the bands centred at 7.5 kHz and 12 kHz have gained more intense energy. Correspondingly, the TF spectrogram based on STFT is given in Figure 9c. The impact-like signatures with high amplitudes lie above 10 kHz in the TF domain. While the 1k Hz components, dominating the frequency spectrum, show more regular patterns but much lower energy in the spectrogram. Given the maximum kurtosis of the analysed signal, no pivotal diagnostic information can be learned from the

above observations directly. By TF analysis, one can better recognize the spectral locations of the inherent temporal behaviours, while the complex TF features still limit its diagnostic performance, for which more advanced analysis should be introduced.
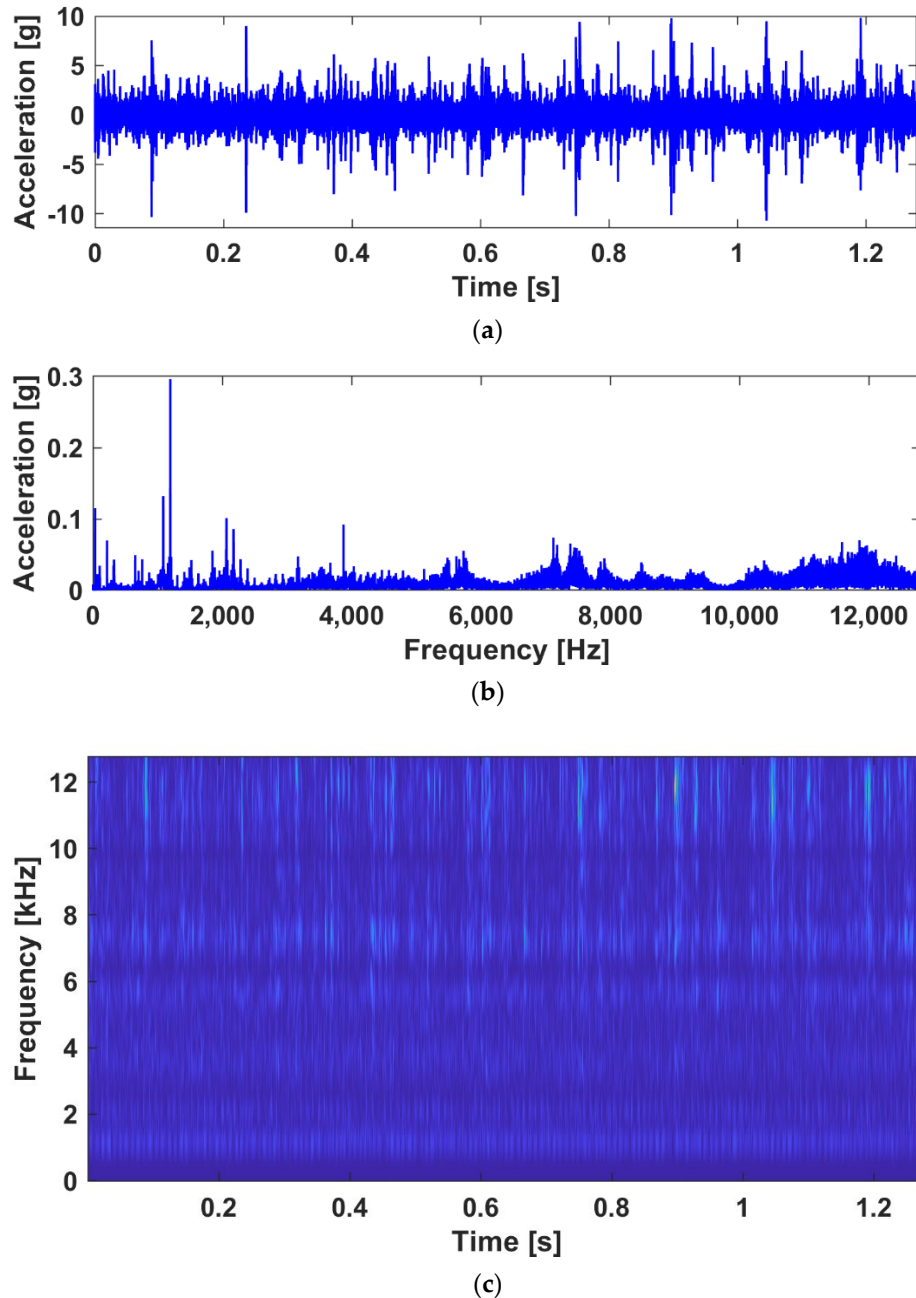

(a)

(b)

(c)

Figure 9. (**a**) The horizontal acceleration signal, (**b**) the amplitude spectrum, and (**c**) the corresponding STFT spectrogram in Case B.

Presumably, the fault signatures in Case B (maximum kurtosis state) are more prominent than that in Case A (at the early stage). Therefore, the low-rank dimensions were reduced to attain more concentrated fault features, i.e., (R_1, R_2, R_3) was set to $(3, 3, 2)$. Keeping other settings same to the Case A, LRANTD was performed for factorizing the TF tensor, $\mathcal{Y} \in \mathbb{R}^{256 \times 14640 \times 8}$, obtained in the similar fashion according to Section 3.

The resulting factors of the LRANTD are shown in Figure 10. Three SS with different dominant frequency bands were extracted in the spectral factor, i.e., $[11, 12.5]$, $[1,6]$, and

[7,8] kHz. The Kurtosis values of the three TS in Figure 10b are 19.5, 3.54, and 8.36, respectively. Furthermore, the segment-wise factor in Figure 10c shows an opposite trend between the two SWS, indicating apparent differences between the separated components.
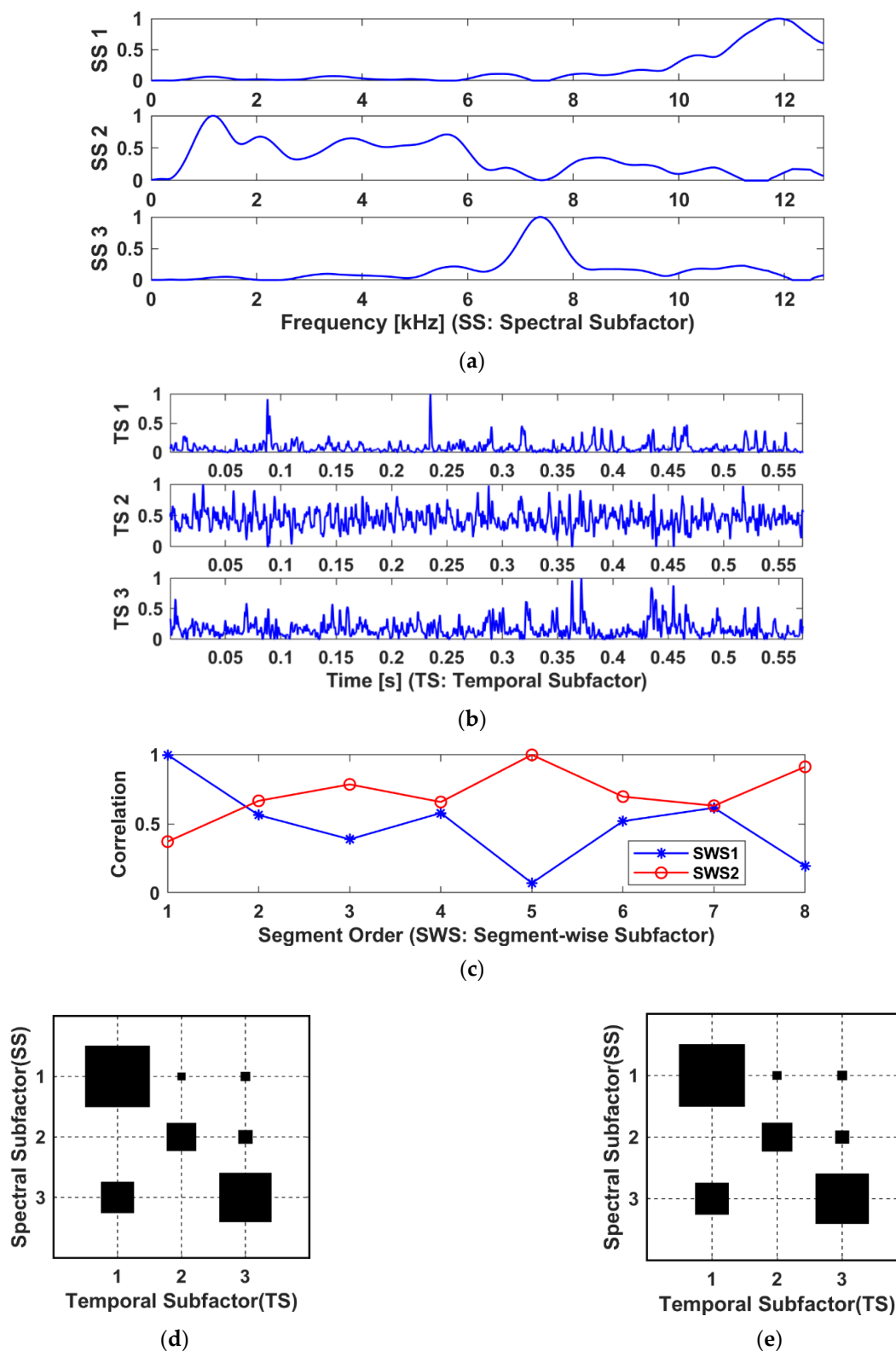

(**a**)

(**b**)

(**c**)

(**d**)

(**e**)

Figure 10. Factors of LRA-NTD: (**a**) Spectral Factor, (**b**) Temporal Factor, (**c**) Segment-wise Factor, and (**d**) the Core Tensor represented via the frontal slice 1 and (**e**) the frontal slice 2.

Next, the amplitude spectrum of each TS was computed and those of the first two temporal factors were shown in Figure 11. As the test bearing had further degraded in Case B, both the amplitude spectra of the TS 1 and TS 2 have captured the characteristics of the outer-race fault, whereas the spectrum of TS 2 shows more prominent signs with higher peaks at the BPFO and its 2nd multiple. Recalling the core tensor in Figure 10d,e, the TS 1 are dominated by SS 1 and TS 2 by SS 2, indicating that the outer-race fault component is more likely to locate itself in the bands of $[1,6]$ and $[11, 12.5]$ kHz.

(a)

(b)
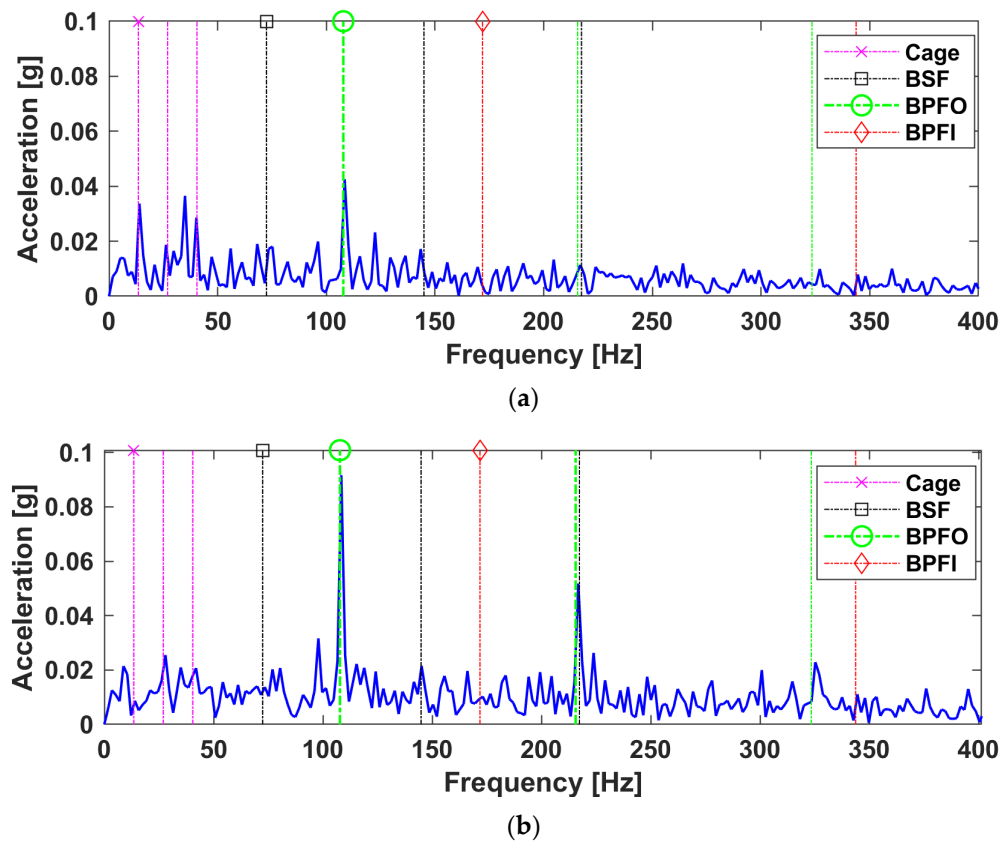
Figure 11. Case B results via LRANTD: (**a**) The amplitude spectrum of TS1 and (**b**) that of TS2.

4.3. Case C

In Case C, the source separation capability of the diagnostic algorithm based on LRANTD was demonstrated using a severe fault case (at the 80th minute) as the outer-race fault developed. For consistency, the horizontal channel acceleration signal was analysed. As shown in Figure 12, the time-domain acceleration signal exhibits impulsive behaviour of high-amplitude (over 10 g) in a cyclic (or regular) manner, while the frequency spectrum shows that the most dominant band lies in $[1, 2]$ kHz. Moreover, a closer look at the STFT spectrogram in Figure 12c provides that the dominant band centred at around 1.2 kHz contains a series of impact signatures, while other bands remain highly complex and may scarcely help diagnosis.
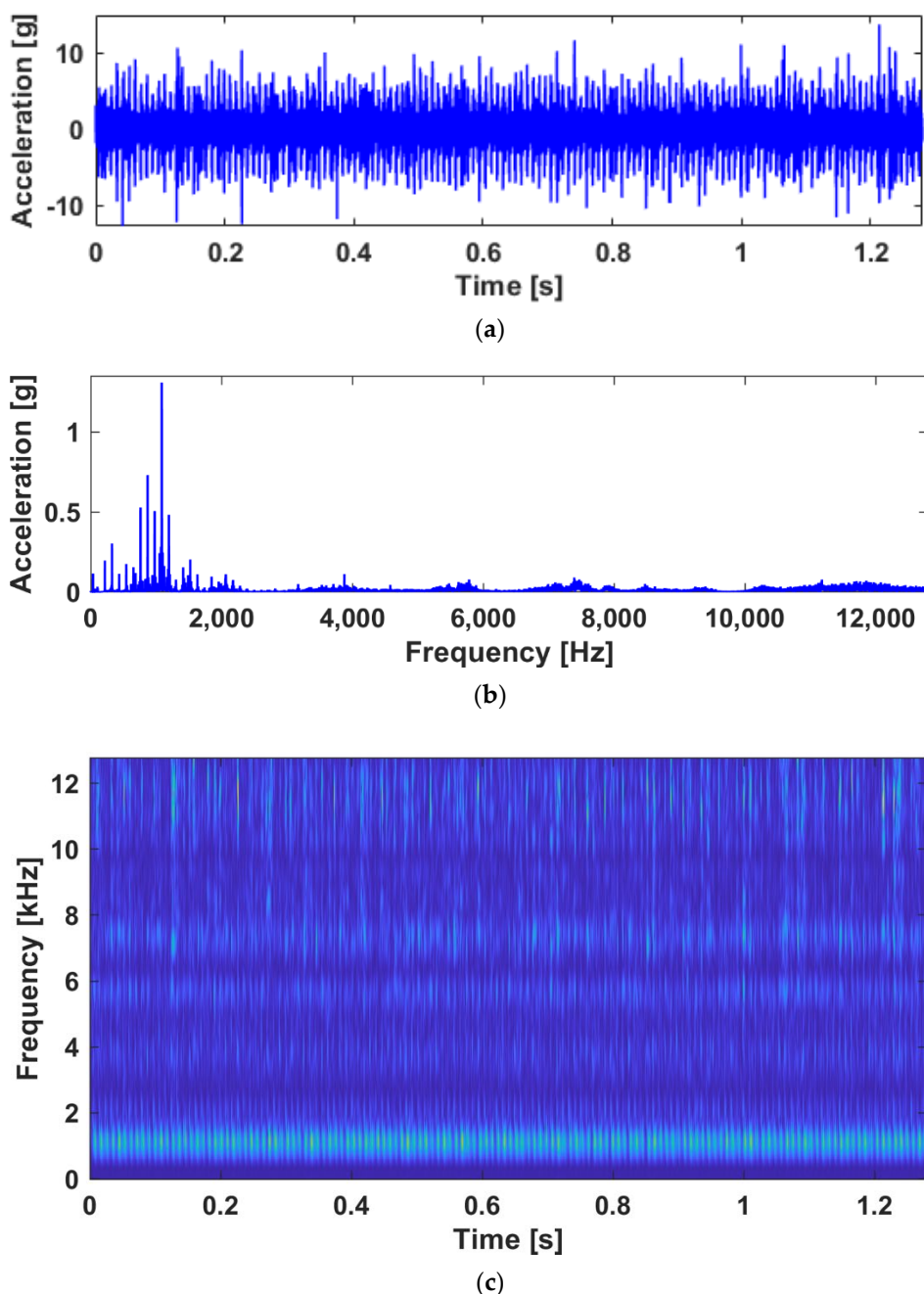
(a)

(b)

(c)

Figure 12. (**a**) The horizontal acceleration signal, (**b**) the amplitude spectrum, and (**c**) the corresponding STFT spectrogram in Case C.

Next, the same settings as in the flowchart in Figure 4 were used to obtain the three-way TF tensor $\mathcal{Y} \in \mathbb{R}^{256 \times 14640 \times 8}$ and to perform LRANTD for tensor factorization. The resulting four factors are shown in Figure 13. The spectral factor in Figure 13a has extracted three SS whose dominant frequency bands are centred at 1.2 kHz, 11.5 kHz, and 7.5 kHz, respectively. Three TS in Figure 13b show obvious impulsive patterns whose kurtosis values from TS 1 to TS 3 are 2.56, 8.06, and 4.30. Note that the core tensor in Figure 13d,e have large values in the diagonal entries, indicating the one-to-one correspondence between each SS and TS (e.g., SS 1 is most relevant to TS 1, so is SS 2 to TS 2, etc.).
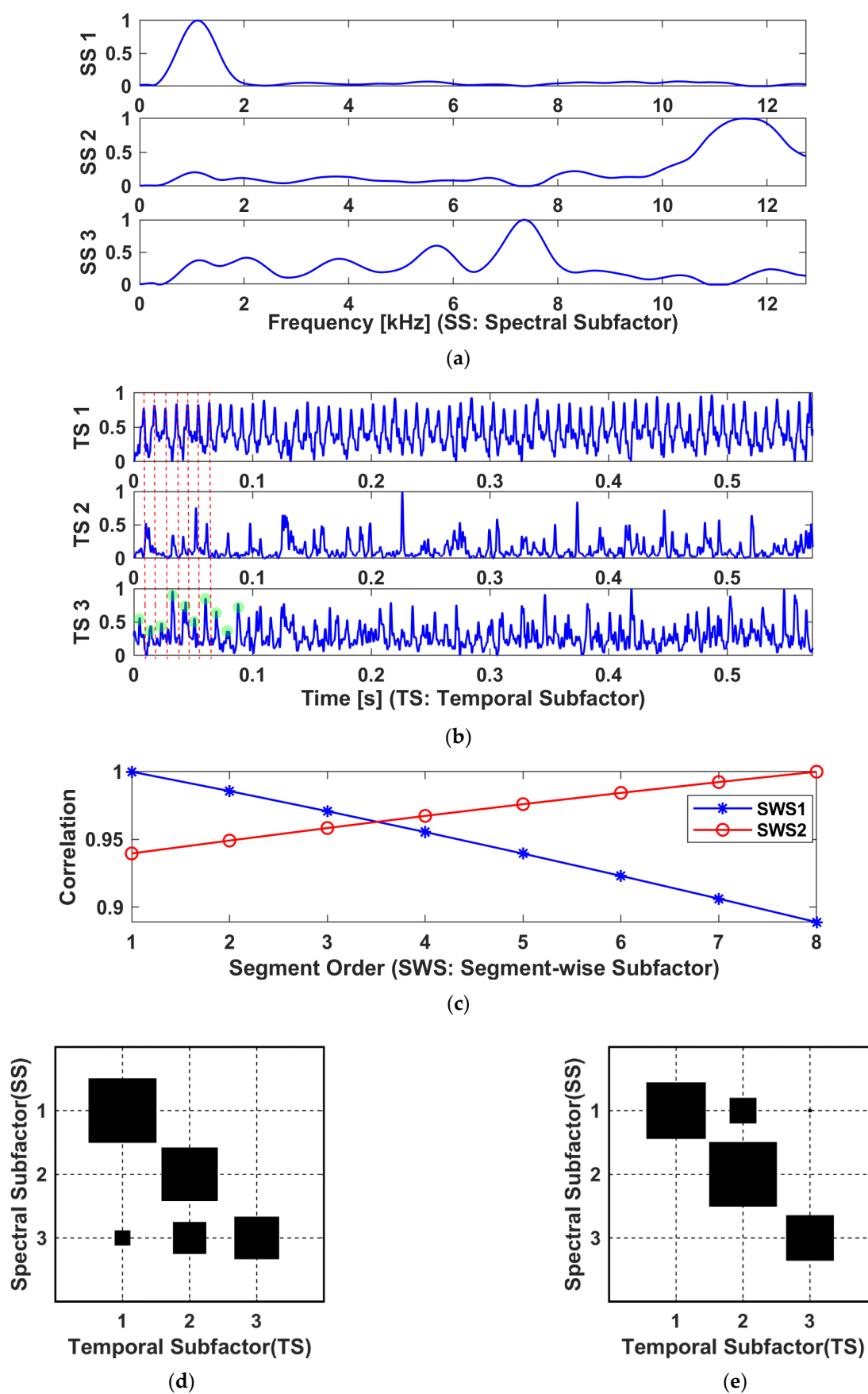
Figure 13. Factors of LRANTD: (**a**) Spectral Factor, (**b**) Temporal Factor, (**c**) Segment-wise Factor, and (**d**) the Core Tensor represented via the frontal slice 1 and (**e**) the frontal slice 2.

Further, the amplitude spectrum of each TS was calculated and those of the 1st and 3rd TS (or TS1 and TS3) were given in Figure 14. Both spectra contain apparent frequency component at BPFO and its high-order harmonics. Interestingly, it is notable that the impact sequences in TS1 and TS3 are indeed quite different, as they possess different phase properties. In Figure 13b, this difference is highlighted by marking the local peaks of TS1 by red dash lines and those of TS 3 by green dots. The observations suggest that although TS1 and TS3 are similarly modulated by BPFO, they are intrinsically two different sources of impact components with different spectral excitations. Such spectral differences are revealed by the extracted spectral factor in Figure 13a, which shows that the fault impacts of TS1 was excited at about 1.2 kHz while the fault impacts of TS 3 are dominated by a broad band that peaks at 7.5 kHz.

On the above basis, the source separation capability of LRANTD is demonstrated. As the test bearing had undergone a severe outer-race fault, it is justifiable to infer that these two different impact patterns were resulted by the rolling elements entering and exiting the defect zone.
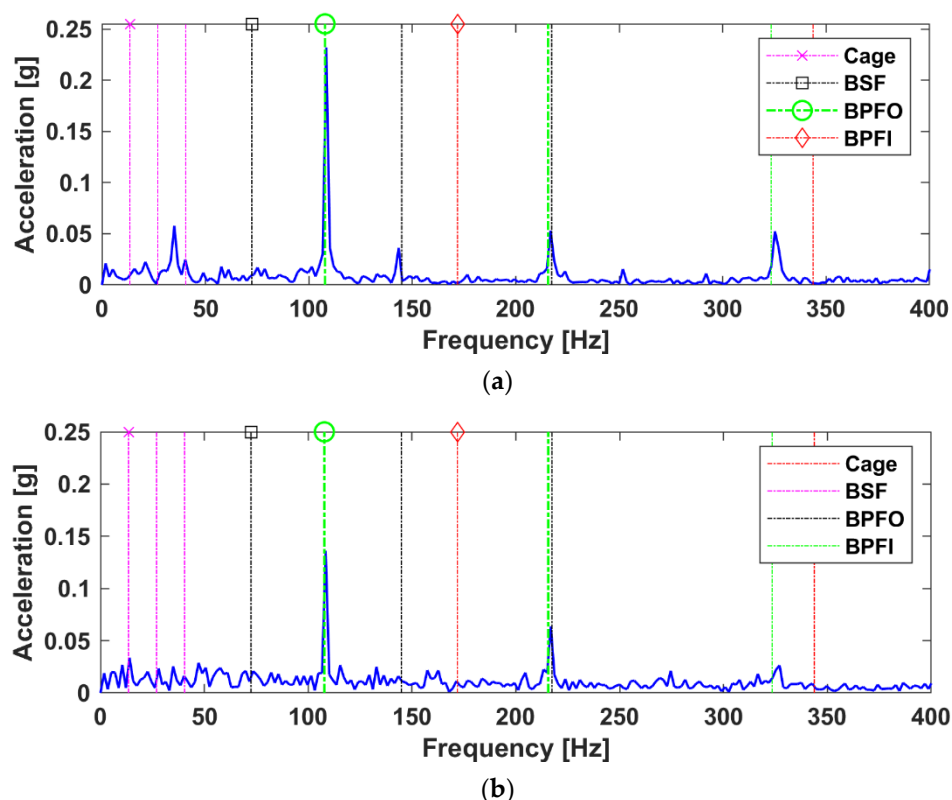

(**a**)

(**b**)

Figure 14. Case C results via LRANTD. (**a**) The amplitude spectrum of TS1 and (**b**) that of TS3.

5. Comparisons with Benchmark Methods

In this section, to verify the performance of the proposed diagnostic methodology, the result of the LRANTD in the early fault case (Case A) was compared against a few practical benchmarks including envelope analysis, FK, and EMD. For illustration, the spectral result based on the LRANTD in Case A is recalled and shown in Figure 15.
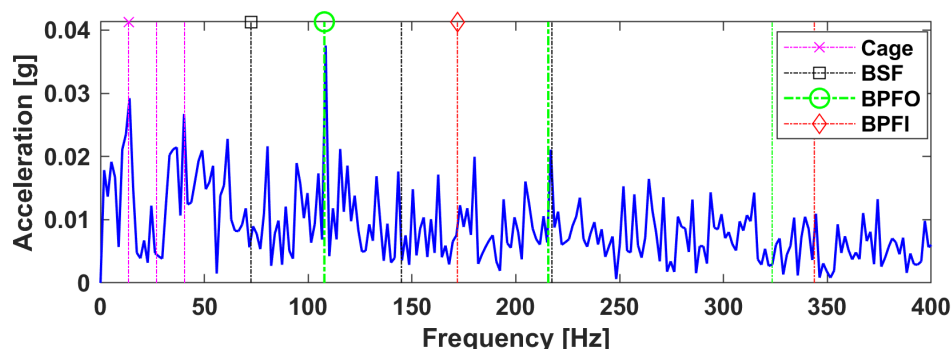
Figure 15. The amplitude spectrum attained by the LRANTD method in Case A (Proposed method).

5.1. Envelope Analysis

Figure 16 shows the smoothed envelope spectrum of the filtered (high pass above 1 kHz) signal of the horizontal acceleration in Case A. For noise reduction, the averaging operation was performed based on 90% overlap between the adjacent Fourier frames. Apparently, the envelope spectrum peaks at the shaft rotation frequency, as marked by the red dot, whereas the BPFO component is less prominent due to the masking of ambient interferences.
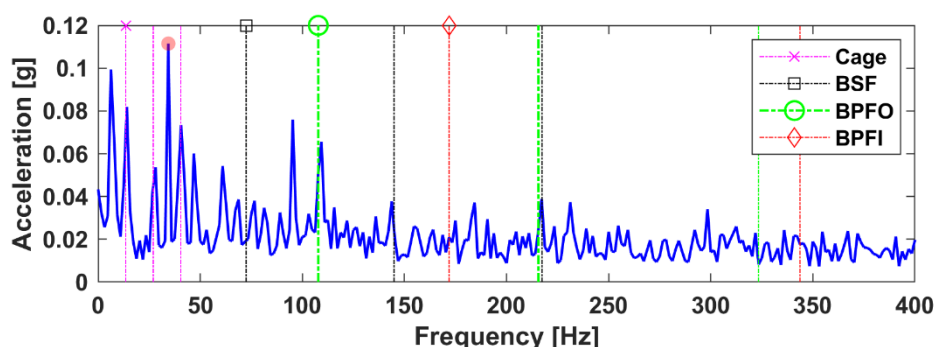

Figure 16. Smoothed envelope spectrum (with 90% overlap and resolution at 1.5625 Hz).

Compared to the envelope analysis, the spectral result of LRANTD in Figure 15 effectively removes the cyclic interference from the shaft rotation and presents clear fault characteristics for diagnosis.

5.2. Fast Kurtogram

As an effective band selector, fast Kurtogram is another pragmatic benchmark for rotating machine diagnosis [4]. The Kurtogram was computed based on the decimated filter banks and shown in Figure 17a. The result indicates that the impulsive components are more likely to dominate over 10 kHz, which is consistent with the spectral extraction result of LRANTD, as the SS 2 in Figure 7a.

Based on the returned optimal filter band with the highest kurtosis value, i.e., $[11200, 12800]$ Hz, the envelope spectrum of the filtered signal is further given in Figure 17b. Here, the square operation of the envelope has been omitted for comparisons. In this result, the BPFO component is also dominated by the rotation frequency (dotted in red) and is of much lower energy than the envelope spectrum in Figure 16 due to a small band width.
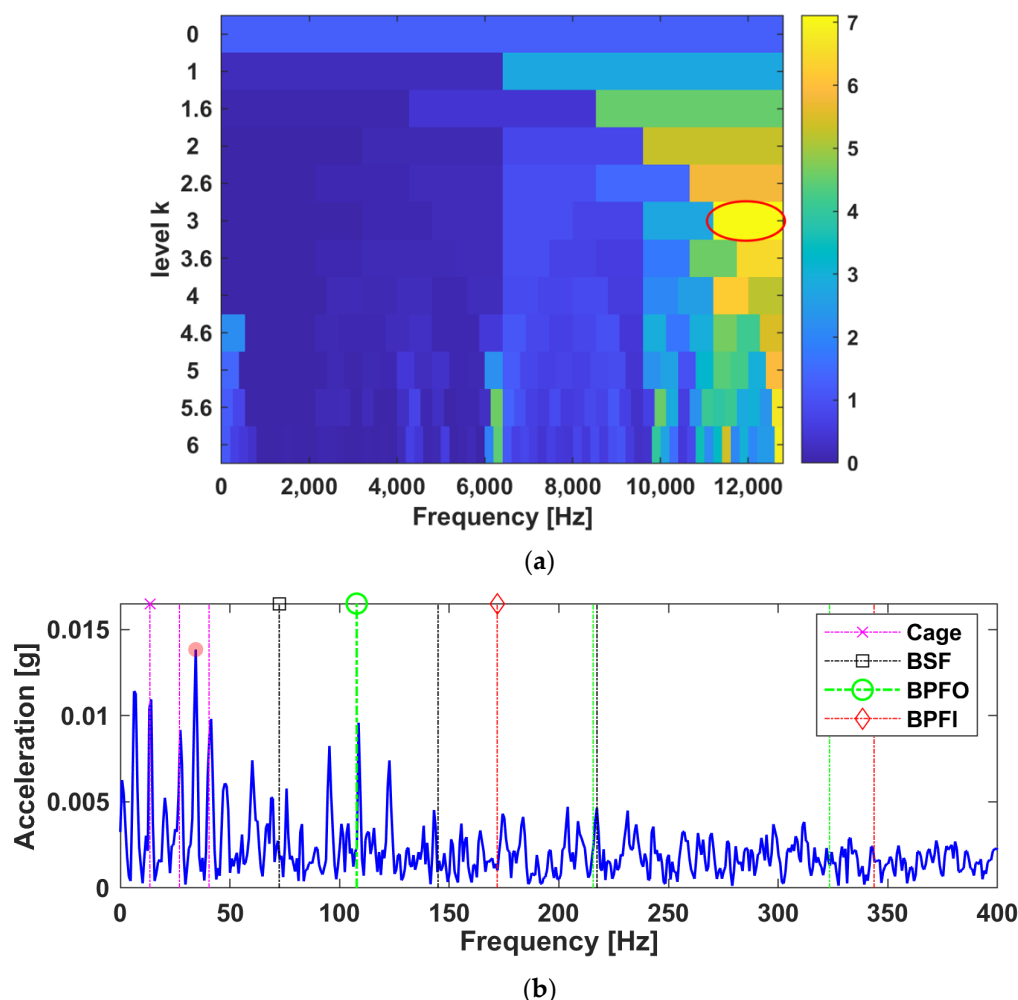
(**a**)

(**b**)

Figure 17. Results of Fast Kurtogram: (**a**) Kurtogram with the maximum kurtosis, 7.1, found at level 3, with band width 1600Hz, and the central frequency at 12000Hz. (**b**) The envelope spectrum of the filtered signal using the aforementioned optimal filter band.

In contrast, the LRANTD method benefits from the source separation capability of tensor decomposition that facilitates adaptive band extraction and hence provides more prominent diagnostic signatures.

5.3. Empirical Mode Decomposition

Lastly, the results were compared for the Case A signal using the empirical mode decomposition and the LRANTD method. Figure 18 shows the envelope spectrum of the 1st intrinsic mode function (IMF) (for comparison, the best result among 10 IMFs was selected), where the similar masking behaviours to the BPFO can be found from the shaft speed (dotted in red) and the other low-frequency components. This can be attributed to the incipient degradation condition where the BPFO component is low in energy and heavily contaminated.
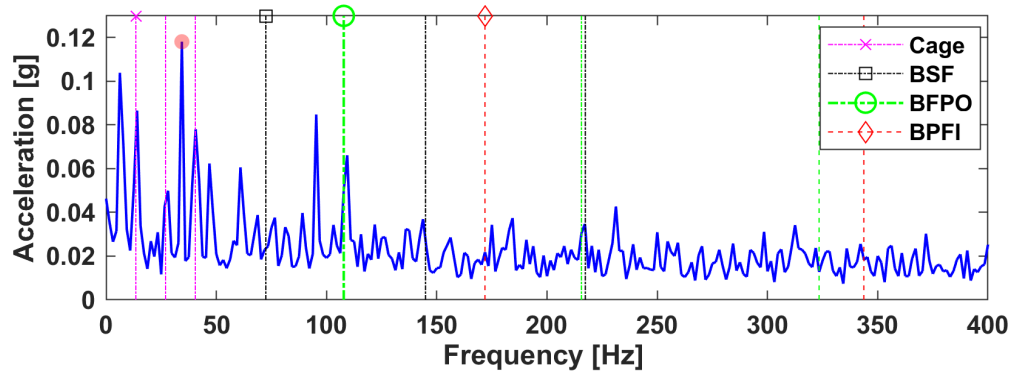
Figure 18. The envelope spectrum of the 1st IMF.

On the other hand, the proposed method effectively enhanced the inherent signatures within the original measurement by fusing multiple signal segments in an overlapping fashion and further mining the multi-way arrays by the LRANTD model.

6. Cross-Spectrum Methods Based on Sensor Fusion

Next, the proposed methodology was applied to multi-sensor fusion based on the two-channel vibration data. Since both horizontal and vertical acceleration signals were measured, we set $M = 2, S = 4$, and the total number of data segments $I_3 = MS = 8$, to create a three-way TF tensor, $\mathcal{Y} \in \mathbb{R}^{256 \times 14640 \times 8}$, in a consistent manner as in Section 3.

In the following, the three typical cases (from Case A to C) were revisited but analysed with two-channel fusion via LRANTD to present a novel cross-spectrum. The cross spectral results from the LRANTD model are shown in Section 6.1 and the comparative results based on the conventional cross power spectral density are given in Section 6.2.
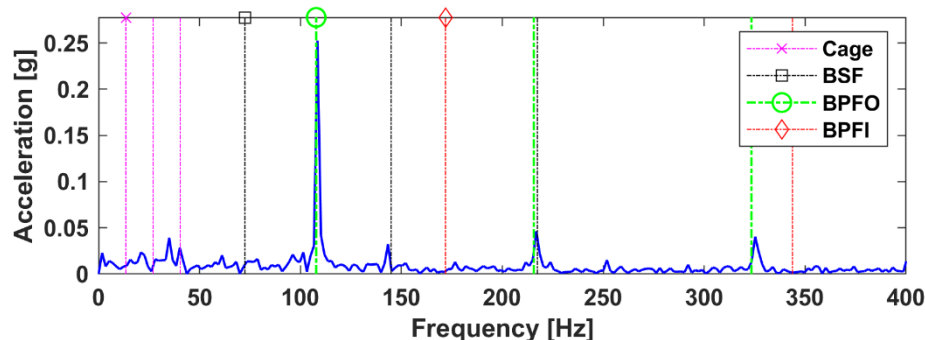
6.1. The Cross-Spectrum Based on LRANTD

By performing the proposed method on the data in Cases A to C, two-directional acceleration signals were first fused and then factorized using the LRANTD model. Consequently, the cross spectra based on the fast Fourier transform of the temporal factors were obtained in succession, as shown in Figure 19. From Case A to Case C, it can be observed that the BPFO component gradually increases in spectral amplitude, and the harmonics of BPFO also become more prominent as the out-race fault developed.

(**a**)
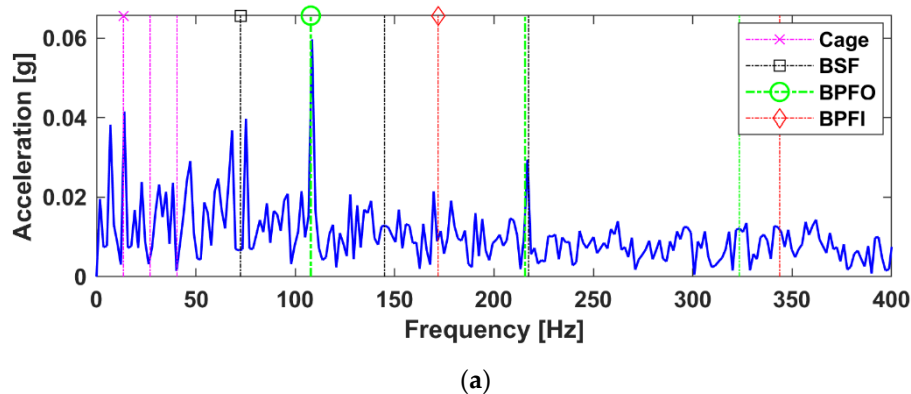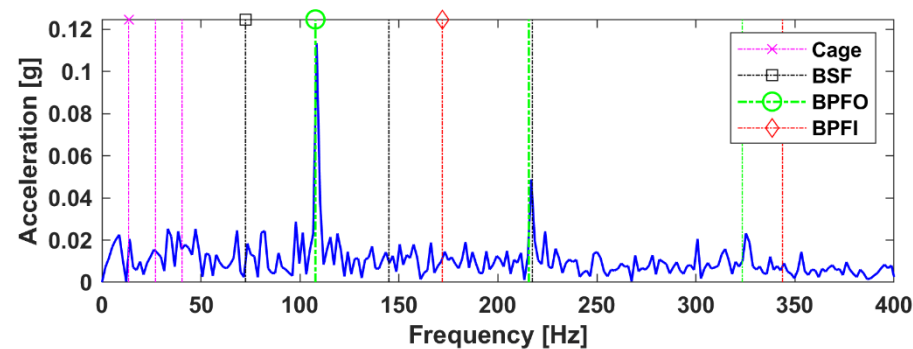
(**b**)

(**c**)

Figure 19. (**a**–**c**): The cross-spectra based on LRANTD fusion for Cases A to C. From Case A to Case C, the low-rank dimensions are, respectively, (3,4,2), (4,4,2), and (3,3,2).

6.2. The Conventional Method

For comparisons, the cross power spectral density (CPSD) of the envelope signals from two orthogonal acceleration measurements was calculated using the Welch's averaged periodogram method with 90% overlap. The resulting spectra of Case A, B, and C are, respectively, shown in Figure 20 and adjusted to the same scale as those in Figure 19. For Case A and Case B, the BPFO component and its harmonics in the CPSD are interfered by other cyclic components, in particular the shaft rotation frequency, as marked by the red dots in Figure 20a,b.

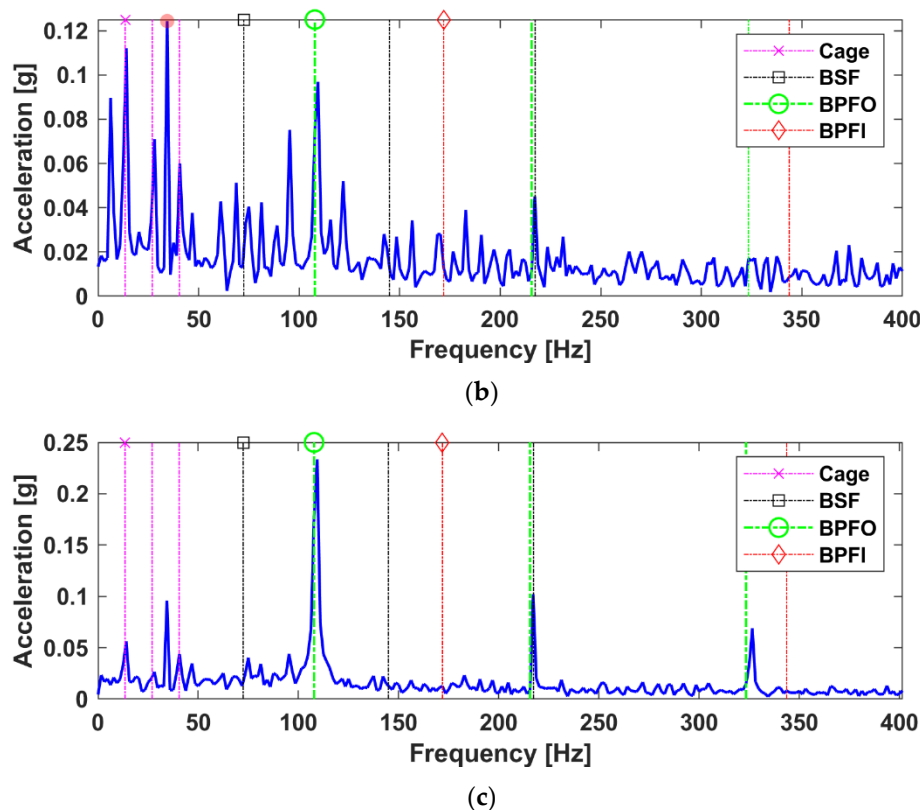(**a**)

(b)

(c)
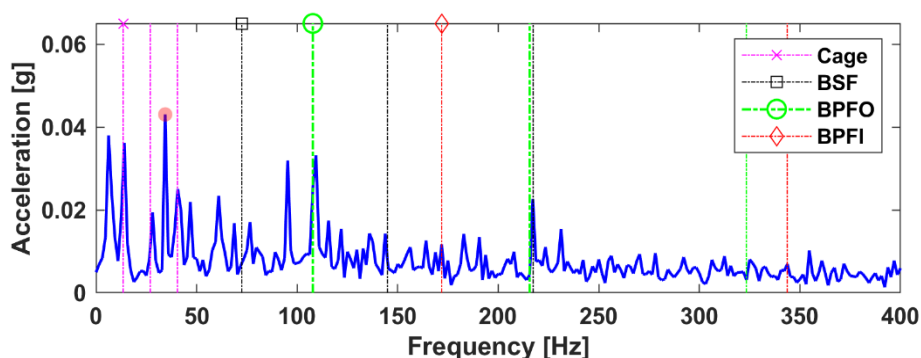
Figure 20. (a–c): Results based on the cross-power spectral density of the two-channel envelope signals, from Case A to Case C.

In contrast, the cross spectra obtained by the LRANTD-based sensor-fusion in Figure 19 provide more prominent fault features in terms of spectral amplitude and concentration, which eventually confirms the effectiveness of the proposed diagnostic method.

7. Summary of Observations

Tables 3 and 4 summarise the observations made from this study using one sensor and 2-sensors data fusion. The results of the proposed LRANTD method are also compared. The background noise is calculated based on the mean of spectral amplitude within the range 0-400 Hz. The ratio in the bracket is calculated using the ratio of the spectral amplitude at a frequency to the mean background noise. This study clearly indicates the predominance of the outer-race defect in the bearing. The "Remark" in Tables 3 and 4 is given by comparing the BPFO amplitude ratio (BPFO amplitude by background noise) with respect to the BPFO Baseline ratio of the envelope analysis as the reference (underlined in the table) within each case study. The following advantages are clearly observed for the proposed LRANTD method that are expected to improve the bearing fault diagnosis process significantly:

(i)   Reducing the presence of the shaft speed related frequency;

(ii) The signal-to-noise ratio is much better for the proposed LRANTD method, hence improving the appearance of the bearing defect frequencies in the spectrum;

(iii) Significant improvements in the BPFO amplitude, generally in the order of 90% for the sensor fusion cases (B and C) compared to the envelope cross-spectrum.

Table 3. Characteristic spectral amplitude in g (with its ratio to the mean background noise) for Case A.

Component	Case A (Single-Channel)				Case A (Sensor-Fusion)	
	Envelope Spectrum	Fast Kurtogram	EMD	LRANTD	(Envelope) CPSD	LRANTD
Shaft Speed f_r	0.1115 (5.21)	0.0138 (6.27)	0.1181 (5.15)	0.0211 (2.63)	0.043 (5.58)	0.0214 (2.05)
Cage Freq. f_c	0.0819 (3.82)	0.0109 (4.95)	0.0866 (3.78)	0.0292 (3.65)	0.0362 (4.70)	0.0416 (4.00)
BSF f_b	0.0202 (0.94)	0.0019 (0.86)	0.0244 (1.06))	0.0090 (1.12)	0.0103 (1.33)	0.0069 (0.66)
BPFO f_o	0.0656 (3.06)	0.0096 (4.36)	0.0661 (2.88)	0.0375 **(4.68)**	0.0332 (4.31)	0.0597 **(5.74) ***
BPFO 2x	0.0388 (1.81)	0.0047 (2.13)	0.0348 (1.51)	0.0211 **(2.63)**	0.0226 (2.93)	0.0294 **(2.82) ***
BPFO 3x	0.0230 (1.07)	0.0021 (0.95)	0.018 (0.78)	0.0060 (0.75)	0.0080 (1.03)	0.0118 **(1.13) ***
BPFI f_i	0.0209 (0.97)	0.0043 (1.95)	0.0218 (0.95)	0.0123 (1.53)	0.0117 (1.51)	0.0115 (1.10)
Background Noise	0.0214	0.0022	0.0229	0.008	0.0077	0.0104
Remark	BPFO Baseline	+1.30 (+42.48%)	−0.18 (−5.88%)	+1.62 **(+52.94%)**	BPFO Baseline	+1.43 **(+33.18%)**

* The best amplitude ratio for the specific spectral component in each case

Table 4. Characteristic spectral amplitude in g (with its ratio to the mean background noise) for Case B and Case C.

Component	Case B (Sensor-Fusion)		Case C (Sensor-Fusion)	
	CPSD	LRANTD	CPSD	LRANTD
Shaft Speed f_r	0.1244 (7.36)	0.0217 (2.08)	0.0958 (6.02)	0.0394 (4.37)
Cage Freq. f_c	0.112 (6.62)	0.0205 (1.97)	0.0562 (3.53)	0.0157 (1.74)
BSF f_b	0.0405 (2.39)	0.0184 (1.76)	0.0404 (2.54)	0.0137 (1.52)
BPFO f_o	0.0969 (5.73)	0.1134 **(10.90) ***	0.2335 (14.68)	0.2525 **(28.05) ***
BPFO 2x	0.0451 (2.66)	0.0488 **(4.69) ***	0.1021 **(6.42) ***	0.0464 (5.15)
BPFO 3x	0.0170 (1.01)	0.0232 **(2.23) ***	0.0690 (4.33)	0.0405 **(4.50) ***
BPFI f_i	0.0128 (0.75)	0.0146 (1.40)	0.0126 (0.79)	0.0054 (0.60)
Background Noise	0.0169	0.0104	0.0159	0.009
Remark	BPFO Baseline	+5.17 (+90.23%)	BPFO Baseline	+13.37 (+91.07%)

* The best amplitude ratio for the specific spectral component in each case

8. Conclusions

In the presented work, a novel diagnostic methodology based on the LRANTD model was proposed to tensorize the single-channel or multi-channel vibration into multi-way time-frequency tensors and to identify the latent fault signatures with different spectral excitations, temporal patterns, and sensor locations. Detailed observations and diagnostic analysis were provided based on the XJTU-SY bearing run-to-failure data, including the illustrations on three typical cases during the bearing lifespan to demonstrate its superiority in adaptive frequency band extraction and fault feature separation.

The LRANTD model was first introduced to bearing fault diagnostics to overcome the drawbacks of unconstrained tucker decomposition while enhancing the performance of NTD by low-rank approximation. The higher-order correlations among the spectral, temporal, and segment-wise dimensions were intuitively visualized using the Hinton diagrams. Moreover, the cross-spectrum derived from LRANTD also shows improved capability for sensor-fusion. The experimental and comparative studies have confirmed the diagnostic performance of the proposed tensor-based method. The proposed LRANTD method for the bearing fault diagnosis is significantly increasing the signal-to-noise ratio and reducing the presence of the shaft speed frequency and its harmonics. This improves the appearance of the bearing defect frequencies in the spectrum and hence eases the bearing defect detection process. The multi-sensors fusion further helps in improving the signal-to-noise ratio and the amplification in the peak amplitude of the bearing defect frequency by nearly 90% as compared to the envelope analysis. The proposed method is

showing encouraging potential for the reliable bearing fault diagnosis, and useful for the industrial application. Further studies are underway with more case studies to observe the robustness of the proposed method for the industrial application. Since the computational time is not significant compared to the conventional approaches, the method can further be explored for the online monitoring.

Author Contributions: Conceptualization, H.W., L.Z. and J.K.S.; methodology, software, writing—original draft preparation, H.W.; resources, supervision, project administration, L.Z. and J.K.S.; writing—review and editing, L.Z. and J.K.S. All authors have read and agreed to the published version of the manuscript.

References

1. Wang, J.; Xu, C.; Zhang, J.; Zhong, R. Big Data Analytics for Intelligent Manufacturing Systems: A Review. *J. Manuf. Syst.* **2022**, *62*, 738–752. https://doi.org/10.1016/j.jmsy.2021.03.005.
2. Loutas, T.H.; Roulias, D.; Georgoulas, G. Remaining Useful Life Estimation in Rolling Bearings Utilizing Data-Driven Probabilistic E-Support Vectors Regression. *IEEE Trans. Reliab.* **2013**, *62*, 821–832. https://doi.org/10.1109/TR.2013.2285318.
3. McDonald, G.L.; Zhao, Q. Multipoint Optimal Minimum Entropy Deconvolution and Convolution Fix: Application to Vibration Fault Detection. *Mech. Syst. Signal Process.* **2017**, *82*, 461–477. https://doi.org/10.1016/j.ymssp.2016.05.036.
4. Antoni, J. Fast Computation of the Kurtogram for the Detection of Transient Faults. *Mech. Syst. Signal Process.* **2007**, *21*, 108–124. https://doi.org/10.1016/j.ymssp.2005.12.002.
5. Antoni, J.; Xin, G.; Hamzaoui, N. Fast Computation of the Spectral Correlation. *Mech. Syst. Signal Process.* **2017**, *92*, 248–277. https://doi.org/10.1016/j.ymssp.2017.01.011.
6. Randall, R.B.; Antoni, J.; Gryllias, K. Alternatives to Kurtosis as an Indicator of Rolling Element Bearing Faults. In Proceedings of the ISMA 2016—International Conference on Noise and Vibration Engineering and USD2016—International Conference on Uncertainty in Structural Dynamics, Leuven, Belgium, 19–21 September 2016.
7. Ding, Y.; He, W.; Chen, B.; Zi, Y.; Selesnick, I.W. Detection of Faults in Rotating Machinery Using Periodic Time-Frequency Sparsity. *J. Sound Vib.* **2016**, *382*, 357–378. https://doi.org/10.1016/j.jsv.2016.07.004.
8. Feng, Z.; Zhou, Y.; Zuo, M.J.; Chu, F.; Chen, X. Atomic Decomposition and Sparse Representation for Complex Signal Analysis in Machinery Fault Diagnosis: A Review with Examples. *Measurement* **2017**, *103*, 106–132. https://doi.org/10.1016/j.measurement.2017.02.031.
9. Chen, Z. Rolling Bearing Fault Diagnosis with Compressed Signals Based on Hybrid Compressive Sensing. *J. Vibroeng.* **2022**, *24*, 18–29. https://doi.org/10.21595/jve.2021.22061.
10. Yang, B.; Liu, R.; Chen, X. Fault Diagnosis for a Wind Turbine Generator Bearing via Sparse Representation and Shift-Invariant K-SVD. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1321–1331. https://doi.org/10.1109/TII.2017.2662215.
11. Arcos Jiménez, A.; Zhang, L.; Gómez Muñoz, C.Q.; García Márquez, F.P. Maintenance Management Based on Machine Learning and Nonlinear Features in Wind Turbines. *Renew. Energy* **2020**, *146*, 316–328. https://doi.org/10.1016/j.renene.2019.06.135.
12. Senanayaka, J.S.L.; Van Khang, H.; Robbersmyr, K.G. Toward Self-Supervised Feature Learning for Online Diagnosis of Multiple Faults in Electric Powertrains. *IEEE Trans. Ind. Inform.* **2021**, *17*, 3772–3781. https://doi.org/10.1109/TII.2020.3014422.
13. Sepulveda, N.E.; Sinha, J. Parameter Optimisation in the Vibration-Based Machine Learning Model for Accurate and Reliable Faults Diagnosis in Rotating Machines. *Machines* **2020**, *8*, 66. https://doi.org/10.3390/machines8040066.
14. Espinoza-Sepulveda, N.; Sinha, J. Mathematical Validation of Experimentally Optimised Parameters Used in a Vibration-Based Machine-Learning Model for Fault Diagnosis in Rotating Machines. *Machines* **2021**, *9*, 155. https://doi.org/10.3390/MACHINES9080155.
15. Luwei, K.C.; Sinha, J.K.; Yunusa-Kaltungo, A.; Elbhbah, K. Data Fusion of Acceleration and Velocity Features (DFAVF) Approach for Fault Diagnosis in Rotating Machines. *MATEC Web Conf.* **2018**, *211*, 21005.
16. Wodecki, J.; Kruczek, P.; Bartkowiak, A.; Zimroz, R.; Wyłomańska, A. Novel Method of Informative Frequency Band Selection for Vibration Signal Using Nonnegative Matrix Factorization of Spectrogram Matrix. *Mech. Syst. Signal Process.* **2019**, *130*, 585–596. https://doi.org/10.1016/j.ymssp.2019.05.020.

17. Wodecki, J.; Michalak, A.; Zimroz, R.; Wyłomańska, A. Separation of Multiple Local-Damage-Related Components from Vibration Data Using Nonnegative Matrix Factorization and Multichannel Data Fusion. *Mech. Syst. Signal Process.* **2020**, *145*, 106954. https://doi.org/10.1016/j.ymssp.2020.106954.

18. Liang, L.; Ding, X.; Wen, H.; Liu, F. Impulsive Components Separation Using Minimum-Determinant KL-Divergence NMF of Bi-Variable Map for Bearing Diagnosis. *Mech. Syst. Signal Process.* **2022**, *175*, 109129. https://doi.org/10.1016/j.ymssp.2022.109129.

19. Liang, L.; Shan, L.; Liu, F.; Niu, B.; Xu, G. Sparse Envelope Spectra for Feature Extraction of Bearing Faults Based on NMF. *Appl. Sci.* **2019**, *9*, 755. https://doi.org/10.3390/app9040755.

20. He, Q.; Ding, X. Time-Frequency Manifold for Machinery Fault Diagnosis. In *Smart Sensors, Measurement and Instrumentation*; Structural Health Monitoring; Springer: Cham, Switzerland, 2017; Volume 26, pp. 131–154.

21. Harshman, R. a Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multimodal Factor Analysis. *UCLA Work. Pap. Phon.* **1970**, *16*, 1–84.

22. Tucker, L.R. Some Mathematical Notes on Three-Mode Factor Analysis. *Psychometrika* **1966**, *31*, 279–311. https://doi.org/10.1007/BF02289464.

23. Cichocki, A.; Mandic, D.; De Lathauwer, L.; Zhou, G.; Zhao, Q.; Caiafa, C.; PHAN, H.A. Tensor Decompositions for Signal Processing Applications: From Two-Way to Multiway Component Analysis. *IEEE Signal Process. Mag.* **2015**, *32*, 145–163. https://doi.org/10.1109/MSP.2013.2297439.

24. Miron, S.; Zniyed, Y.; Boyer, R.; de Almeida, A.L.F.; Favier, G.; Brie, D.; Comon, P. Tensor Methods for Multisensor Signal Processing. *IET Signal Process.* **2020**, *14*, 693–709. https://doi.org/10.1049/iet-spr.2020.0373.

25. Zhao, H.; Zhang, W.; Wang, G. Fault Diagnosis Method for Wind Turbine Rolling Bearings Based on Hankel Tensor Decomposition. *IET Renew. Power Gener.* **2019**, *13*, 220–226. https://doi.org/10.1049/iet-rpg.2018.5284.

26. He, Z.; Shao, H.; Cheng, J.; Zhao, X.; Yang, Y. Support Tensor Machine with Dynamic Penalty Factors and Its Application to the Fault Diagnosis of Rotating Machinery with Unbalanced Data. *Mech. Syst. Signal Process.* **2020**, *141*, 106441. https://doi.org/10.1016/j.ymssp.2019.106441.

27. Hu, C.; Wang, Y. Multidimensional Denoising of Rotating Machine Based on Tensor Factorization. *Mech. Syst. Signal Process.* **2019**, *122*, 273–289. https://doi.org/10.1016/j.ymssp.2018.12.012.

28. Liang, L.; Wen, H.; Liu, F.; Li, G.; Li, M. Feature Extraction of Impulse Faults for Vibration Signals Based on Sparse Non-Negative Tensor Factorization. *Appl. Sci.* **2019**, *9*, 3642. https://doi.org/10.3390/app9183642.

29. Wen, H.; Liang, L.; Niu, B.; Shan, L.; Li, M.; Li, G. The Separation of Vibration Components Based on Sparse Nonnegative Tensor Factorization. In *Smart Innovation, Systems and Technologies: Advances in Asset Management and Condition Monitoring*; Springer: Berlin/Heidelberg, 2020; pp. 1297–1306, ISBN 9783030577445.

30. Zhou, G.; Cichocki, A.; Zhao, Q.; Xie, S. Efficient Nonnegative Tucker Decompositions: Algorithms and Uniqueness. *IEEE Trans. Image Process.* **2015**, *24*, 4990–5003. https://doi.org/10.1109/TIP.2015.2478396.

31. Zhou, G.; Cichocki, A.; Xie, S. Fast Nonnegative Matrix/Tensor Factorization Based on Low-Rank Approximation. *IEEE Trans. Signal Process.* **2012**, *60*, 2928–2940. https://doi.org/10.1109/TSP.2012.2190410.

32. Wang, B.; Lei, Y.; Li, N.; Li, N. A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings. *IEEE Trans. Reliab.* **2018**, *69*, 401–412. https://doi.org/10.1109/TR.2018.2882682.

33. Kim, Y.D.; Choi, S. Nonnegative Tucker Decomposition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2007, Minneapolis, MN, USA, 17–22 June 2007.

34. Lee, D.D.; Seung, H.S. Algorithms for Non-Negative Matrix Factorization. *Adv. Neural Inf. Process. Syst.* **2001**, *13*, 556–562.

35. Phan, A.H.; Cichocki, A. Extended HALS Algorithm for Nonnegative Tucker Decomposition and Its Applications for Multiway Analysis and Classification. *Neurocomputing* **2011**, *74*, 1956–1979. https://doi.org/10.1016/j.neucom.2010.06.031.

36. Cichocki, A.; Zdunek, R.; Amari, S.I. Hierarchical ALS Algorithms for Nonnegative Matrix and 3D Tensor Factorization. In Proceedings of the 7th International Conference, ICA 2007, London, UK, 9–12 September 2007; Volume 4666, pp. 169–176. https://doi.org/10.1007/978-3-540-74494-8.

37. Kim, H.; Park, H. Nonnegative Matrix Factorization Based on Alternating Nonnegativity Constrained Least Squares and Active Set Method. *SIAM J. Matrix Anal. Appl.* **2008**, *30*, 713–730. https://doi.org/10.1137/07069239X.

38. Antoni, J.; Randall, R.B. Differential Diagnosis of Gear and Bearing Faults. *J. Vib. Acoust. Trans. ASME* **2002**, *124*, 165–171. https://doi.org/10.1115/1.1456906.