



HAL
open science

Sparse representations in vibration-based rolling element bearing diagnostics

Ge Xin

► **To cite this version:**

Ge Xin. Sparse representations in vibration-based rolling element bearing diagnostics. Vibrations [physics.class-ph]. Université de Lyon, 2017. English. NNT : 2017LYSEI051 . tel-02063268

HAL Id: tel-02063268

<https://theses.hal.science/tel-02063268>

Submitted on 11 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N°d'ordre NNT : 2017LYSEI051

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
L'INSA LYON

Ecole Doctorale N° 162
MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE

Spécialité/ discipline de doctorat : Acoustique

Soutenue publiquement le 22/06/2017, par :
Ge XIN

**Sparse Representations in Vibration-
based Rolling Element Bearing
Diagnostics**

Devant le jury composé de :

CHIEMENTIN Xavier, Maître de Conférences – HDR, URCA
CORBIER Christophe, Maître de Conférences – HDR, Université de Lyon
DRON Jean-Paul, Professeur des Universités, URCA
KAFTANDJIAN Valérie, Professeur des Universités, INSA-Lyon

Rapporteur
Rapporteur
Examineur
Examinatrice

ANTONI Jérôme, Professeur des Universités, INSA-Lyon
HAMZAOUI Nacer, Professeur des Universités, INSA-Lyon

Directeur de thèse
Co-directeur de thèse

Sparse Representations in Vibration-based Rolling Element Bearing Diagnostics



Ge XIN

Laboratoire Vibrations Acoustique
INSA-Lyon

This dissertation is submitted for the degree of
Doctor of Philosophy

Université de Lyon

July 2017

I would like to dedicate this thesis to my loving parents, my dear wife and my cute daughter
Yucheng.



献给我唯一的光源 — 瑀橙
For Yucheng, the only light I need.

Acknowledgements

This thesis has been made possible thanks to the generous help of many people, and I wish to offer my most heartfelt thanks to them.

First and foremost, I would like to thank my supervisors, Prof. Jérôme Antoni and Prof. Nacer Hamzaoui. Thanks for giving me this excellent opportunity to work on this exciting project. It has been an exceptional privilege to work with such kind, insightful, energetic, and knowledgeable advisers. Jérôme, no words can express my appreciation for your outstanding support, relentless encouragement and unwavering guidance throughout my academic career. Along all these, I understand such famous expression: “If I have seen further, it is by standing on the shoulders of giants” and for all these my deepest gratitude goes to you. Nacer, you are the most enthusiastic person I have ever met. Your patience, enthusiasm and kindness always inspire me to think positive and overcome any difficulty. Once again, I sincerely thank you.

I specially express my thanks to Dr. Xavier Chimentin and Dr. Christophe Corbier for reviewing this manuscript and Prof. Jean-Paul Dron and Prof. Valérie Kaftandjian for accepting as the jury of my thesis.

I would like to thank all my colleagues in the laboratory for their help during my stay in France. My thanks go especially to Dr. Liang Yu, who provided me many valuable scientific suggestions. Prof. Erliang Zhang and Dr. Bin Dong who gave me a lot of advice and shared their experiences. I miss the time spent with Edouard Cardenas Cabada, Souhayb Kass, Philéas Maigrot, Edgar Sierra Alonso, Marco Buzzoni and Charles Vanwynsberghe, who gave me warm encouragement and generous assistance. I would like to extend my thanks to many friends in Lyon, Emmanuel Roux, Inés Mérida, Corinne Lotto, Fulbert Mbailassem and Ting Su. We passed so much wonderful time together.

I would also show my gratitude to China Scholarship Council (CSC) for the financement and Laboratoire Vibrations Acoustique (LVA) for the support during my Ph.D. study.

Finally, my sincere thanks go to my family for their unconditional love and support in everything I do. Thanks my parents and my parents-in-law for taking care of my daughter and helping us to go through a very difficult time in our life. To my beloved wife, Miaomiao, thanks for your love and being there in my life. To my daughter, Yucheng, thanks for always making me feel like the happiest dad in the world.

Abstract

Although vibration-based rolling element bearing diagnostics is a very well-developed field, the research on sparse representations of vibration signals is yet new and challenging for machine diagnosis. As a desired property – representation of a signal in highly organized structure – sparsity enables us to reveal the natural signature of singular events embedded in a signal so as to reduce the demand on the user's expertise, even though it involves advanced theory of stochastic processes. In this thesis, several novel methods have been developed, by means of different stochastic models, associated with their effective algorithms so as to serve the industry in rolling element bearing diagnostics.

First, the sparsity-based model (sparse code, in natural image processing) is investigated based on the current literature. By summarizing its successful reasons from three points, the historical background of sparse representations has been inquired in the field of natural scenes. Along such three aspects, its mathematical model with corresponding algorithms has been categorized and presented as a fundamental premise; the main publications are therefore surveyed in the literature on machinery fault diagnosis; finally, by discussing the pros and cons of sparse representations, an interpretation of sparse structure in the Bayesian viewpoint is proposed which then gives rise to two novel models for machinery fault diagnosis.

Second, a new stochastic model is introduced to address this issue: it introduces a hidden variable to indicate the occurrence of the impacts and estimates the spectral content of the corresponding transients together with the spectrum of background noise. This gives rise to an automatic detection algorithm – with no need of manual prefiltering as is the case with the envelope spectrum – from which fault frequencies can be revealed. The same algorithm also makes possible to filter out the fault signal in a very efficient way as compared to other approaches based on the stationary assumption. The performance is investigated on synthetic signals with a high noise-to-signal ratio and also in the case of a mixture of two independent transients. The effectiveness and robustness of the method are also verified on vibration signals measured on a test-bench (gears and bearings). Results are found superior or at least equivalent to those of conventional envelope analysis and fast kurtogram.

Third, a novel scheme for extracting cyclostationary (CS) signals is proposed. It introduces a periodic-variance based stochastic model to recover the CS component in the

masking of interfering signals. This proceeds from the property that, for a CS signal, the STFT evidences periodic flows of energy in and across its frequency bins. By considering the periodic variance as hidden variables, a time-varying filter is designed so as to achieve the full-band reconstruction of CS signals characterized by some pre-set characteristic frequency, which can be obtained by the prior knowledge or the estimator of Spectral Correlation. The performance of the proposed scheme has been demonstrated on synthetic and experimental cases. Of particular interest is the robustness on experimental data sets and superior extraction capability over the conventional Wiener filter. It not only deals with the bearing fault at an incipient stage, but it even works for the installation problem and the case of two sources, i.e. bearing and gear faults together. Eventually, these experimental examples evidence its versatile usage on diagnostic analysis of compound signals.

Fourth, a benchmark analysis by using the fast computation of the spectral correlation is provided. This study benefits from a big data set – which is publicly available and widely used – supplied by the Case Western Reserve University (CWRU) Bearing Data Center. Particularly, it contains plenty of cases ranges from very easily diagnosable to not diagnosable; while some signals are showing the typical harmonic structure of bearing fault signatures, others are quite blurry or even display other fault symptoms. This is one crucial point that makes this work challenging and charming – moving forward the benchmark study of the CWRU data set – by uncovering its own unique characteristics.

Résumé

Bien que le diagnostic des roulements par analyse vibratoire soit un domaine très développé, la recherche sur les représentations parcimonieuses des signaux de vibration est encore nouvelle et difficile pour le diagnostic des machines tournantes. En tant que propriété souhaitée – la représentation d'un signal dans une structure hautement organisée – la parcimonie nous permet de révéler la signature naturelle d'événements singuliers intégrés dans un signal afin de réduire la demande sur l'expertise de l'utilisateur, même si elle implique une théorie avancée des processus stochastiques. Dans cette thèse, de méthodes nouvelles ont été développées, au moyen de différents modèles stochastiques, associées à des algorithmes efficaces afin de servir l'industrie dans le diagnostic des roulements.

Tout d'abord, les modèles parcimonieux présentés dans la littérature sont revus. En résumant leurs avantages en trois points, le contexte historique des représentations parcimonieuses a été examiné, notamment dans le domaine des scènes naturelles. En suivant ces trois aspects, les modèles mathématiques ainsi que les algorithmes associés ont été classés et présentés comme une prémisse fondamentale. Les principales publications concernant le diagnostic des machines tournantes ont également été considérées. Enfin, en discutant des avantages et des inconvénients des représentations parcimonieuses, une interprétation des structures creuses d'un point de vue Bayésien est proposée, ce qui donne lieu à deux nouveaux modèles de diagnostic des machines tournantes.

Dans un second temps, un nouveau modèle stochastique est proposé : il introduit une variable cachée relative à l'apparition d'impacts et estime le contenu spectral des transitoires correspondants ainsi que le spectre du bruit de fond. Cela donne lieu à un algorithme de détection automatique - sans besoin de pré-filtrage manuel comme c'est le cas avec le spectre d'enveloppe - à partir duquel les fréquences de défaut peuvent être révélées. Le même algorithme permet également de filtrer le signal de défaut de manière très efficace par rapport à d'autres approches basées sur l'hypothèse stationnaire. La performance de l'algorithme est étudiée sur des signaux synthétiques avec un rapport signal à bruit faible et également dans le cas d'un mélange de deux transitoires indépendants. L'efficacité et la robustesse de la méthode sont également vérifiées sur les signaux de vibration mesurés sur un banc

d'essai (engrenages et paliers). Les résultats sont meilleurs ou au moins équivalents à ceux de l'analyse d'enveloppes classique et du kurtogramme rapide.

Dans un troisième temps, un nouveau schéma pour l'extraction de signaux cyclostationnaires (CS) est proposé. Il introduit un modèle stochastique basé sur la variance périodique pour récupérer la composante CS masquée par des signaux parasites. Cela découle de la propriété statuant que, pour un signal CS, la transformée de Fourier à court terme met en évidence des flux d'énergie temporellement périodiques. En considérant la variance périodique en tant que variable cachée, un filtre temporel est conçu de manière à obtenir la reconstruction intégrale des signaux CS caractérisés par une fréquence cyclique préétablie, qui peut être connue a priori ou estimée à partir de la corrélation spectrale. La performance du schéma proposé a été évaluée sur des cas synthétiques et expérimentaux. Un intérêt particulier de la méthode est sa robustesse lorsqu'elle est appliquée sur des données expérimentales ainsi qu'une capacité d'extraction supérieure par rapport au filtre de Wiener conventionnel. Cette approche se révèle efficace non seulement sur des défauts de roulement à des stades précoces, mais également sur des problèmes de montage dans le cas où plusieurs sources sont présentes (par exemple les engrenages et les roulements ensemble). Finalement, ces exemples expérimentaux témoignent de l'utilisation polyvalente de la méthode à des fins de diagnostic de signaux composés.

Pour finir, une analyse comparée utilisant le calcul rapide de la corrélation spectrale est réalisée sur une base de données publiquement disponible et largement utilisée. Cette base bénéficie d'un grand ensemble de données fournies par le centre de données de l'Université Case Western Reserve (CWRU). En particulier, elle contient de nombreux cas d'étude, allant de diagnostics simples aux plus complexes. Tandis que certains signaux présentent une structure harmonique typique des signatures de défauts de roulement, d'autres sont assez ambiguës ou présentent d'autres symptômes inattendus de défaut. C'est un point crucial qui fixe un défi non-trivial à résoudre.

Table of contents

List of figures	xv
List of tables	xxv
1 Introduction	1
1.1 Context	1
1.2 Problem statement	3
1.3 State of the art	5
1.4 Focus and Contribution	6
1.5 Outline	8
2 Literature review on sparse representations of vibration signal	9
2.1 Derivation of sparse code	9
2.2 Sparsity-based models	11
2.2.1 Synthesis model	13
2.2.2 Analysis model	14
2.2.3 Comparison and evolution	14
2.3 Algorithms for sparse representation	15
2.3.1 Sparse coding	16
2.3.2 Learning-based dictionary	17
2.4 Applications on rolling element bearing diagnostics	18
2.5 Discussion and conclusion	21
3 Automatic Spectrum Matching of Repetitive Transients Based on Hidden Markov Model	25
3.1 Spectral mixture model	25
3.1.1 Model and assumption	26
3.1.2 Hidden Markov model	27
3.1.3 Structured covariance matrix	32

3.2	Reconstruction scheme for repetitive transients	33
3.3	Parameter selection	35
3.3.1	Initial settings	35
3.3.2	Cases 1 & 2: demonstration of parameter selection	38
3.3.3	Case 3: Comparison of different spectral correlation assumptions	42
3.3.4	Case 4: diagnostics and separation of a mixture of independent transients	47
3.4	Validation on vibration signals	51
3.4.1	Case 5: diagnosis of a ball fault	51
3.4.2	Cases 6 & 7: diagnostics of bearing and gears	57
3.4.3	Case 8: diagnostics of bearing in nonstationary regime	62
3.4.4	Case 9: diagnosis in the presence of multiple components	67
3.5	Conclusion	72
4	Extraction of cyclostationary signals	73
4.1	Periodic-variance based model	73
4.1.1	Model and assumptions	73
4.1.2	Periodic-variance based model	74
4.2	Extraction scheme for cyclostationary signals	76
4.3	Parameter selection	79
4.3.1	Initial settings	79
4.3.2	Cases 1 & 2: demonstration of parameter selection	81
4.4	Validation on experimental examples	89
4.4.1	Example 1: extraction of bearing signals with good gears	89
4.4.2	Example 2: separation of bearing signals with broken gears	103
4.5	Conclusion	113
5	Benchmark survey on using the fast spectral correlation	115
5.1	Background on applied diagnostic methods	115
5.2	Case Western Reserve University data	118
5.2.1	Experimental set-up	121
5.2.2	Examples and discussion	121
5.3	Result tables and conclusion	130
6	Conclusions and perspectives	133
6.1	Conclusions	133
6.2	Perspectives	134

Table of contents	xiii
References	137
Appendix A Commonly used symbols and statistical quantities	145
Appendix B Proof of the mixture distribution with two Gaussian distributions giving rise to a super-Gaussian distribution	149
Appendix C Proof of the inverse gamma distribution giving rise to a conjugate prior for the variance	151
Appendix D Tables of results	153
Publications	159

List of figures

1.1	Typical fault signals in rolling element bearings.	2
1.2	Illustration of the influence of random slips and fluctuations: (a) and (d) stand for synthetic signals in time domain; (b) and (e) for raw spectra calculated by the Fourier transform; (c) and (f) for squared envelope spectra.	4
2.1	(Fig.1 in Ref. [35]) Sparse coding. (a) An image is represented by a small number of “active” code elements, a_i , out of a large set. Which elements are active varies from one image to the next. (b) Since a given element in a sparse code will most of the time be inactive, the probability distribution of its activity will be highly peaked around zero with heavy tails. This is in contrast to a code where the probability distribution of activity is spread more evenly among a range of values (such as a Gaussian).	10
2.2	(Table 1.1 in [37]) Some robust penalty functions. For all cases, $C(\gamma) = \sum_i \rho(\gamma_i)$	13
2.3	Graphical illustration of the <i>synthesis</i> model, which represents a sparse vibration signal as a linear combination of three activated atoms in <i>Dictionary</i> . As can be seen, noisy measurements in blue, reconstructed signal in red, reference signal in green, activated atoms in black. It is notice that not all the atoms are displayed in the block of <i>Dictionary</i> , alternatively, only ten identifiable waveforms of trained atoms are displayed.	14
2.4	General vibration fault diagnosis procedure.	18

2.5	Illustration of Spectral Kurtosis analysis: the repetitive series of transients passes a filter-bank analysis implemented by the STFT; as can be seen in $K_y(f_b)$, there exist two different distributions which indicate the fault component (centered at frequency f_1) and the noisy component (centered at frequency f_0), respectively. It is highlighted that along the time instant the occurrence of transients can be characterized by means of a latent variable (i.e. the spectral mixture model in Chapter 3). In particular, for cyclostationary signals, its energy flow displays a periodical variation which indicates cyclostationarity in second-order statistics (i.e. the periodic-variance based model in Chapter 4).	23
3.1	Explicit steps of the EM algorithm to infer the parameters in the HMM. . .	30
3.2	Graphical model in the case $K = 2$	32
3.3	Illustration of how to select the window length N_w and shift R with respect to transient durations T_l and cycle Δt	36
3.4	Spectrogram of the signal simulated in Case 1 with resonance frequency $f_0 = 0.15$ Hz, $r = 0.95$ and fault frequency $\alpha_0 = 1.25 \times 10^{-3}$ Hz ($T = 1/\alpha_0 = 800$ s, $N_w = 2^7$ and $R = 20$).	38
3.5	(a) Synthetic signal of Case 1 with white noise (noise-to-signal-ratio= 6 dB). (b) Synthetic repetitive transients. (c) Initialized latent variable $\hat{\zeta}(i)^{[0]}$ (black line) with threshold (red dotted line) set to $0.01 \max(\hat{\zeta}(i)^{[0]})$ with $\hat{\mu}^{[0]} = 0.112$. (d) Estimated latent variable $\hat{\zeta}(i)$	39
3.6	(a) Initialized spectra (diagonals of covariance matrices $\hat{\mathbf{C}}_n^{[0]}$ and $\hat{\mathbf{C}}_x^{[0]}$ (red line and blue asteriks) and (b) estimated ones from EM (diagonals of covariance matrices $\hat{\mathbf{C}}_n^{[k+1]}$ and $\hat{\mathbf{C}}_x^{[k+1]}$ (red lines) together with the theoretical frequency response $H(z)$ (black dashed line and blue asteriks).	40
3.7	(a) LLR and (b) latent variable $\hat{\zeta}(i)$ in the time domain.	40
3.8	Spectrogram of the signal simulated in Case 2 with resonance frequency $f_0 = 6 \times 10^{-3}$ Hz, $r = 0.9$ and fault frequency $\alpha_0 = 1.25 \times 10^{-3}$ Hz.	41
3.9	Spectra of the LLR (a) and of the latent variable (b) presented in Fig. 3.7 together with the SES of actual fault signal (red dotted line). The limit $\alpha_{max} = 7.8 \times 10^{-3}$ Hz is indicated by a vertical black dotted line (Normalization to unit maximum value).	41
3.10	Spectrogram of the signal simulated in Case 3 with resonance frequency $f_0 = 0.2$ Hz, $r = 0.9$ and fault frequency $\alpha_0 = 1.25 \times 10^{-3}$ Hz.	42

3.11 Spectra of the LLR together with the theoretical SES (red dotted line) (a) full matrix, (b) diagonal matrix, (c) tridiagonal matrix, (d) low rank matrix, (e) Toeplitz matrix (normalization to unit maximum value).	43
3.12 Spectra of the latent variable $\hat{\zeta}(i)$ together with the theoretical SES (red dotted line) (a) full matrix, (b) diagonal matrix, (c) tridiagonal matrix, (d) low rank matrix, (e) Toeplitz matrix (normalization to unit maximum value).	43
3.13 Full matrix: (a) diagonals (the spectrum of the fault signal is indicated by a red solid line and that of the noise by blue asterisks); (b) absolute value; (c) real part; (d) imaginary part of the estimated covariance matrix.	44
3.14 Diagonal matrix: (a) diagonals (the spectrum of the fault signal is indicated by a red solid line and that of the noise by blue asterisks); (b) absolute value; (c) real part; (d) imaginary part of the estimated covariance matrix.	45
3.15 Tridiagonal matrix: (a) diagonals (the spectrum of the fault signal is indicated by a red solid line and that of the noise by blue asterisks); (b) absolute value; (c) real part; (d) imaginary part of the estimated covariance matrix.	45
3.16 Low rank matrix: (a) diagonals (the spectrum of the fault signal is indicated by a red solid line and that of the noise by blue asterisks); (b) absolute value; (c) real part; (d) imaginary part of the estimated covariance matrix.	46
3.17 Toeplitz matrix: (a) diagonals (the spectrum of the fault signal is indicated by a red solid line and that of the noise by blue asterisks); (b) absolute value; (c) real part; (d) imaginary part of the estimated covariance matrix.	46
3.18 (a) Component $\mathbf{X}^1(i)$ with period $T_1 = 530$ samples ($\sigma_\tau^1 = 0.02T_1$ and $\sigma_A^1 = 0.1$). (b) Component $\mathbf{X}^2(i)$ with period $T_2 = 470$ samples ($\sigma_\tau^2 = 0.02T_2$ and $\sigma_A^2 = 0.1$). (c) Noisy measurement (noise-to-signal ratio = 0 dB).	47
3.19 Spectrogram of the signal simulated in Case 4 with two components ($T_1 = 530$ samples and $T_2 = 470$ samples respectively).	48
3.20 Estimated spectra of component $\mathbf{X}^1(i)$ (red solid line), component $\mathbf{X}^2(i)$ (green solid line) and noise (blue dashed line). The two theoretical frequency responses of $H(z)$ are indicated by black dashed lines.	48
3.21 (a) Estimated latent variable $\zeta^1(i)$ (red solid line) together with the actual first component (black solid line). (b) Estimated latent variable $\zeta^2(i)$ (green solid line) together with the actual second component (black solid line).	49
3.22 Spectra of the estimated latent variable (a) $\zeta^1(i)$ and (b) $\zeta^2(i)$ (normalized to unit maximum value).	50

3.23	Reconstructed repetitive transients corresponding to (a) the first component $\mathbf{X}^1(i)$ with period $T_1 = 530$ samples and (b) the second component $\mathbf{X}^2(i)$ with period $T_2 = 470$ samples. (c) Summation of two components $\mathbf{X}^1(i)$ and $\mathbf{X}^2(i)$	50
3.24	Spectrogram (logarithmic scale) of signal of Case 5 (frequency resolution $\Delta f = 375$ Hz).	52
3.25	SES of the raw signal (normalized to unit maximum value).	53
3.26	Diagonals of covariance matrices (frequency resolution $\Delta f = 375$ Hz); the spectrum of the fault signal is indicated by a red solid line and that of the noise by a blue dashed line.	53
3.27	Spectrum of the <i>LLR</i> with markers at the theoretical fault frequency and harmonics; the limit $\alpha_{max} = 375$ Hz is indicated by a vertical black dotted line (normalization to unit maximum value).	54
3.28	Kurtogram of signal of Case 5 computed over $K = 7$ levels with a 1/3-binary tree and an 8 coefficient prototype filter. Several local maxima are presented. One relevant maximum is found at dyad $\{f_{10}; (\Delta f)_4\} = \{14250; 1500\}$ Hz.	55
3.29	SES in frequency band [13500; 15000] Hz returned by the kurtogram (normalization to unit maximum value).	55
3.30	(a) Vibration signal of Case 5 and (b) its band-pass filtered version in the frequency band [13500; 15000] Hz. (c) Full-band reconstructed fault signal from the proposed HMM-based time-varying filter.	56
3.31	(a) Enlarged view of vibration signal of Case 5 (with indication of the period corresponding to the peak f_g in Fig. 3.25) and (b) enlarged view of its band-pass filtered version in the frequency band [13500; 15000] Hz. (c) Enlarged view of the full-band reconstructed fault signal from the proposed HMM-based time-varying filter.	57
3.32	Test rig setup.	58
3.33	Spectrum of the <i>LLR</i> with markers at the suspected fault frequency BPFO and its harmonics (normalization to unit maximum value).	59
3.34	Fast kurtogram of signal of Case 6 computed over $K = 7$ levels with a 1/3-binary tree and an 8 coefficient prototype filter. One relevant maximum is found at dyad $\{f_{13}; (\Delta f)_4\} = \{20k; 1600\}$ Hz.	60
3.35	SES of complex envelope in dyad $\{f_{13}; (\Delta f)_4\} = \{20k; 1600\}$ Hz corresponding to the frequency band [18400; 21600] Hz with markers at the suspected fault frequency BPFO and its harmonics (normalization to unit maximum value).	60

3.36	(a) Vibration signal of Case 6 in time interval [1.9 2.3] s and (b) its band-pass filtered version in frequency band [18400; 21600] Hz. (c) Reconstructed transients from the HMM-based time-varying filter.	61
3.37	Spectrum of the <i>LLR</i> with markers at the suspected fault frequencies BPFO, its harmonics and $f_{(rot,2)}$ (normalization to unit maximum value).	62
3.38	(a) Estimated instantaneous speed of signal in Case 8 and (b) its corresponding acceleration signal which undergoes speed-dependent magnitude modulation. (c) Division of the estimated instantaneous speed in 5 regimes.	63
3.39	(a) Raw signal in regime No. 1 and (b) the corresponding reconstructed transients from the HMM-based time-varying filter.	64
3.40	(a) Raw signal in regime No. 2 and (b) the corresponding reconstructed transients from the HMM-based time-varying filter.	64
3.41	(a) Raw signal in regime No. 3 and (b) the corresponding reconstructed transients from the HMM-based time-varying filter.	65
3.42	(a) Raw signal in regime No. 4 and (b) the corresponding reconstructed transients from the HMM-based time-varying filter.	65
3.43	(a) Raw signal in regime No. 5 and (b) the corresponding reconstructed transients from the HMM-based time-varying filter.	66
3.44	Order spectrum of the <i>LLR</i> with markers at the suspected fault frequency BPOO and its harmonics (normalization to unit maximum value).	66
3.45	Measured signal (a) from 0 to 2.5 s and (b) from 2.5 to 5 s which evidence of instantaneous interfering component in time interval [3.4 4] s.	68
3.46	Spectrogram (logarithmic scale) of signal in time intervals (a) [1.9 2.3] s and (b) [3.45 3.85] s of Case 9 with evidence of two states: a global distribution with spectral content in band [1.8 3.8] kHz and a local distribution with spectral content in band [1.8 3.8] kHz and [9 11] kHz.	69
3.47	Diagonals of the covariance matrices of the three components, $\mathbf{X}^1(i)$ (red solid line), $\mathbf{X}^2(i)$ (green solid line) and noise (blue dashed line).	69
3.48	Spectra of the estimated latent variable (a) $\zeta^1(i)$ and (b) $\zeta^2(i)$, respectively (normalization to unit maximum value).	70
3.49	(a) Measured signal from 3.45 to 3.85 s. Reconstructed components (b) $\mathbf{X}^1(i)$ and (c) $\mathbf{X}^2(i)$ from the HMM-based time-varying filter.	70
3.50	(a) Measured signal from 1.9 to 2.3 s. Reconstructed components (b) $\mathbf{X}^1(i)$ and (c) $\mathbf{X}^2(i)$ from the HMM-based time-varying filter.	71

4.1	Spectrogram of the signal simulated in Case 1 with resonance frequency $f_0 = 0.1$ Hz, $r = 0.95$ and fault frequency $\alpha_0 = 5 \times 10^{-3}$ Hz ($T = 1/\alpha_0 = 000$ s, $N_w = 2^5$ and $R = 4$).	82
4.2	(a) Synthetic signal of Case 1 with white noise (noise-to-signal-ratio= 6 dB. (b) Synthetic transient signal. (c) Recovered time signal $\hat{x}[n]$	83
4.3	(a) Initialized time-dependent variance of the CS signal, $\hat{\sigma}_x^2(i; f_b)^{[0]}$, and (b) the estimated $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm. (c) Spectrogram of the estimated CS signal, $\hat{X}(i, f_b)$, with (d) its periodic time-varying filter $1/(1 + H(i; f_b))$ as defined in Eqs. 4.15-4.16.	83
4.4	(a) Initialized spectra (variance of noise and CS signals $\hat{\sigma}_n^2(f_b)^{[0]}$ and $\hat{\sigma}_x^2(f_b)^{[0]}$, in blue dotted line and red line, respectively) and (b) estimated spectra from the EM algorithm, ($\hat{\sigma}_n^2(f_b)^{[k+1]}$ and $\hat{\sigma}_x^2(f_b)^{[k+1]}$, in blue dotted line and red line, respectively) together with the theoretical squared magnitude frequency response $ H(z) ^2$ (black dashed line).	84
4.5	Cyclostationary index: function of the frequency f_b by means of taking the standard deviation of $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ over the time instant i	85
4.6	(a) Enlarged view of the recovered time signal $\hat{x}[n]$ of Fig. 4.2 (b) and (c). (b) Initialized time-dependent variance of the CS signal, $\hat{\sigma}_x^2(i; f_b)^{[0]}$, and (c) the estimated $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.	85
4.7	(a) Synthetic signal of Case 2 with white noise (noise-to-signal-ratio= 6 dB. (b) Synthetic repetitive transient signal. (c) Recovered time signal $\hat{x}[n]$	86
4.8	(a) Initialized time-dependent variance of the CS signal, $\hat{\sigma}_x^2(i; f_b)^{[0]}$, and (b) the estimated $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm. (c) Spectrogram of the estimated CS signal, $\hat{X}(i, f_b)$, with (d) its periodic time-varying filter $1/(1 + H(i; f_b))$ as defined in Eqs. 4.15-4.16.	87
4.9	(a) Enlarged view of the recovered time signal $\hat{x}[n]$ of Fig. 4.7 (b) and (c). (b) Initialized time-dependent variance of the CS signal, $\hat{\sigma}_x^2(i; f_b)^{[0]}$, and (c) the estimated $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.	88
4.10	Test rig setup of Example 1.	89
4.11	(a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of signal of Case 3 ($\Delta f = 800$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in full band $[0; 25.6]$ kHz.	91
4.12	(a) Vibration signal of Case 3 divided into (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = f_{rot,1} = 16$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).	91

- 4.13 (a) Spectrogram (magnitude of the STFT) of the raw signal, $|Y(i, f_b)|$. (b) Spectrogram of the estimated noise signal, $|Y(i, f_b) - \widehat{X}(i, f_b)|$, and (c) that of the estimated CS signal $|\widehat{X}(i, f_b)|$ with their spectra ($\widehat{\sigma}_n^2(f_b)^{[k+1]}$ and $\widehat{\sigma}_x^2(f_b)^{[k+1]}$) in black dash-dot lines. (d) The estimated time-dependent variance $\widehat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm. 93
- 4.14 Enlarged view of (a) the vibration signal of Case 3, (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = f_{rot,1} = 16$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$). 93
- 4.15 Squared Envelope Spectrum of (a) the vibration signal of Case 3: $S_y^{SES}(\alpha)$, (b) the recovered time signal: $S_{\hat{x}}^{SES}(\alpha)$ and (c) the noise (residual) signal: $S_{\hat{n}}^{SES}(\alpha)$ 94
- 4.16 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of the signal of Case 4 ($\Delta f = 800$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in full band $[0; 25.6]$ kHz. 95
- 4.17 (a) Vibration signal of Case 4 divided into (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = BPF0 = 48$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$). 96
- 4.18 (a) Spectrogram (magnitude of the STFT) of the raw signal, $|Y(i, f_b)|$. (b) Spectrogram of the estimated noise signal, $|Y(i, f_b) - \widehat{X}(i, f_b)|$, and (c) that of the estimated CS signal, $|\widehat{X}(i, f_b)|$, with their spectra ($\widehat{\sigma}_n^2(f_b)^{[k+1]}$ and $\widehat{\sigma}_x^2(f_b)^{[k+1]}$) in black dash-dot lines. (d) The estimated time-dependent variance $\widehat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm. 97
- 4.19 Enlarged view of (a) the vibration signal of Case 4, (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = BPF0 = 48$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$). 97
- 4.20 Squared Envelope Spectrum of (a) the vibration signal of Case 4: $S_y^{SES}(\alpha)$, (b) the recovered time signal: $S_{\hat{x}}^{SES}(\alpha)$ and (c) the noise (residual) signal: $S_{\hat{n}}^{SES}(\alpha)$ 98
- 4.21 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of the signal of Case 5 ($\Delta f = 800$ Hz, $\Delta\alpha = 0.1$ Hz). Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ (b) in the band $[7.6; 25.6]$ kHz and (c) in the band $[3.6; 7.6]$ kHz. 99
- 4.22 (a) Vibration signal of Case 5 divided into (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = BPF0 = 52$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$). 100

4.23	(a) Spectrogram (magnitude of the STFT) of the raw signal, $ Y(i, f_b) $. (b) Spectrogram of the estimated noise signal, $ Y(i, f_b) - \widehat{X}(i, f_b) $, and (c) that of the estimated CS signal, $ \widehat{X}(i, f_b) $, with their spectra $(\hat{\sigma}_n^2(f_b))^{[k+1]}$ and $\hat{\sigma}_x^2(f_b)^{[k+1]}$ in black dash-dot lines. (d) The estimated time-dependent variance $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.	101
4.24	Enlarged view of (a) the vibration signal of Case 5, (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = BPF0 = 52$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).	101
4.25	Squared Envelope Spectrum of (a) the vibration signal of Case 5: $S_y^{SES}(\alpha)$, (b) the recovered time signal: $S_{\hat{x}}^{SES}(\alpha)$ and (c) the noise (residual) signal: $S_{\hat{n}}^{SES}(\alpha)$	102
4.26	(a) Vibration signal of Case 6 divided into (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = BPF0 = 48.9$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).	104
4.27	(a) Spectrogram (magnitude of the STFT) of the raw signal, $ Y(i, f_b) $. (b) Spectrogram of the estimated noise signal, $ Y(i, f_b) - \widehat{X}(i, f_b) $, and (c) that of the estimated CS signal, $ \widehat{X}(i, f_b) $, with their spectra $(\hat{\sigma}_n^2(f_b))^{[k+1]}$ and $\hat{\sigma}_x^2(f_b)^{[k+1]}$ in black dash-dot lines. (d) The estimated time-dependent variance $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.	105
4.28	Spectrogram (logarithmic scale) of the signal of Case 7 (frequency resolution $\Delta f = 750$ Hz) with evidence of spectral content in band $[9.75; 23.25]$ kHz.	106
4.29	(a) Vibration signal of Case 7 divided into (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = 2 \times BSF = 51.29$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).	107
4.30	(a) Spectrogram (magnitude of the STFT) of the raw signal, $ Y(i, f_b) $. (b) Spectrogram of the estimated noise signal, $ Y(i, f_b) - \widehat{X}(i, f_b) $, and (c) that of the estimated CS signal, $ \widehat{X}(i, f_b) $, with their spectra $(\hat{\sigma}_n^2(f_b))^{[k+1]}$ and $\hat{\sigma}_x^2(f_b)^{[k+1]}$ in black dash-dot lines. (d) The estimated time-dependent variance $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.	107
4.31	Enlarged view of (a) the vibration signal of Case 7, (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = 2 \times BSF = 51.29$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).	108
4.32	Squared Envelope Spectrum of (a) the vibration signal of Case 7: $S_y^{SES}(\alpha)$, (b) the recovered time signal: $S_{\hat{x}}^{SES}(\alpha)$ and (c) the noise (residual) signal: $S_{\hat{n}}^{SES}(\alpha)$	108

4.33	(a) Vibration signal of Case 8 divided into (b) the recovered time signal of source 1: $\hat{x}^1[n]$ characterized by $1/T_1 = BPF1 = 42.77$ Hz and (c) that of source 2: $\hat{x}^2[n]$ characterized by $1/T_2 = f_g = f_{rot} \times \text{No. of teeth} = 192.2$ Hz.	110
4.34	(a) Spectrogram (magnitude of the STFT) of the raw signal, $ Y(i, f_b) $. (b) Spectrogram of the estimated CS signal from source 1: $ \hat{X}^1(i, f_b) $ and (c) that from source 2: $ \hat{X}^2(i, f_b) $ with their related spectra (variance of the CS signal $\hat{\sigma}_x^2(f_b)^{[k+1]}$) in black dash-dot lines. (d) The estimated spectra from the EM algorithm (variance of the CS signal $\hat{\sigma}_x^2(f_b)^{[k+1]}$).	111
4.35	Cyclostationary indexes of (a) source 1 and (b) source 2: function of the frequency f_b by means of taking the standard deviation of $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ over the time instant i	111
4.36	Enlarged view of (a) the vibration signal of Case 8, (b) the recovered time signal of source 1: $\hat{x}^1[n]$ characterized by $1/T_1 = BPF1 = 42.77$ Hz and (c) that of source 2: $\hat{x}^2[n]$ characterized by $1/T_2 = f_g = f_{rot} \times \text{No. of teeth} = 192.2$ Hz.	112
4.37	Squared Envelope Spectrum of (a) the vibration signal of Case 8: $S_y^{SES}(\alpha)$, the recovered time signal of (b) source 1: $S_{\hat{x}^1}^{SES}(\alpha)$ and (c) that of source 2: $S_{\hat{x}^2}^{SES}(\alpha)$	112
5.1	Connections between the spectral quantities handled in the paper [78].	117
5.2	CWRU bearing test rig [1].	119
5.3	(a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 125DE ($\Delta f = 1500$ Hz, $\Delta \alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in the band [750; 5250] Hz.	123
5.4	(a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 203DE ($\Delta f = 46.88$ Hz, $\Delta \alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in the band [500; 1500] Hz.	124
5.5	(a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 275DE ($\Delta f = 11.72$ Hz, $\Delta \alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in full band $[0; F_s/2]$ Hz.	125
5.6	(a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 277DE ($\Delta f = 46.88$ Hz, $\Delta \alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in the band [4.3; 5.5] kHz.	127
5.7	Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in selected band [2.3; 3.7] kHz with $\Delta \alpha = 0.1$ Hz.	127

5.8	(a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 282DE ($\Delta f = 5.86$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in the band [4.1; 4.3] kHz.	128
5.9	(a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 290DE ($\Delta f = 46.88$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in the band [3.9; 4.4] kHz.	129
5.10	Method comparison: diagnosis outcomes for the “benchmark data set” (Method 1-3 are outlined in Section 5.1; Y = successful, P = partially successful, N = not successful, more details can be found in Table 5.1.).	130
5.11	Pie charts of diagnosis outcomes for the “benchmark data set”: (a)-(c) respectively corresponded to benchmark method 1-3 and (d) to tested method Fast-SC (Method 1-3 are outlined in Section 5.1; Y = successful, P = partially successful, N = not successful, more details can be found in Table 5.1.). . .	131

List of tables

1.1	Bearing characteristic frequencies (d = bearing roller diameter; D = pitch circle diameter; n = number of rolling elements; θ = contact angle.)	3
3.1	Parameter settings in Case 5.	51
3.2	Parameter settings in Case 6 and Case 7.	58
3.3	Parameter settings in Case 9.	67
4.1	Parameter settings in Case 3, Case 4 and Case 5.	90
4.2	Parameter settings used in Case 6 and Case 7.	103
4.3	Parameter settings used in Case 8.	109
5.1	Criteria for categorizing the diagnosis outcomes.	120
5.2	Bearing details and fault frequencies.	121
5.3	Parameter settings used in the experiment of record 125DE, 203DE, 275DE.	122
5.4	Parameter settings used in the experiment of record 277DE, 282DE, 290DE.	126
A.1	List of symbols	145
A.2	Commonly used statistical quantities	148
D.1	12 k drive-end bearing fault analysis results; benchmark data sets: P and N categorise only; measurement position: DE only.	153
D.2	48 k drive-end bearing fault analysis results; benchmark data sets: P and N categorise only; measurement position: DE only.	154
D.3	12 k fan-end bearing fault analysis results; benchmark data sets: P and N categorise only; measurement position: DE only.	156

Chapter 1

Introduction

From a general perspective, information is most often conveyed by singular events embedded in a signal. In this thesis, the “signal” refers to a possible physical quantity of interest, such as vibration-based measurement – acceleration; and the “information” refers to human’s prior knowledge desire to explore on the object – from bearings or rolling element bearings. As an intermediate bridging the previous two, the hidden “singular events” probably will be captured through complex stochastic models and advanced signal processing techniques. Particularly, this thesis aims to address such issue from the view of “sparse structure”.

1.1 Context

Rolling element bearings are not only common components used in various types of mechanical systems, but they are also crucial due to the foremost cause of machinery breakdown. Such rotating or reciprocating machine includes engines of aircraft or automobile, compressors, turbines and pumps, which typically have a variety of mechanical parts such as the shaft, bearing, gearbox, blade, coupling and belt. For these reasons, vibration signals in practice usually consist of very many frequencies occurring simultaneously produced by some periodic mechanisms and no matter which sort of operating conditions – e.g. healthy, misalignment, damage on bearing or gear, etc. – it undergoes.

For the failure of bearings, typical defects are caused by crack, breakage, spall or uneven wear (pitting, scuffing, abrasion, erosion), often located on the mating surface of the inner race, the outer race or the rolling elements. As the rolling element strikes a defect, ensuing vibration transients occur at a specific rate called “bearing characteristic frequencies”. Since the bearing fault signatures have been well-investigated [2], its characteristic frequencies can be estimated with some kinematic and geometric parameters as presented in Table 1.1. At the incipient stage, each transient resembles a damped impulse response with high frequency

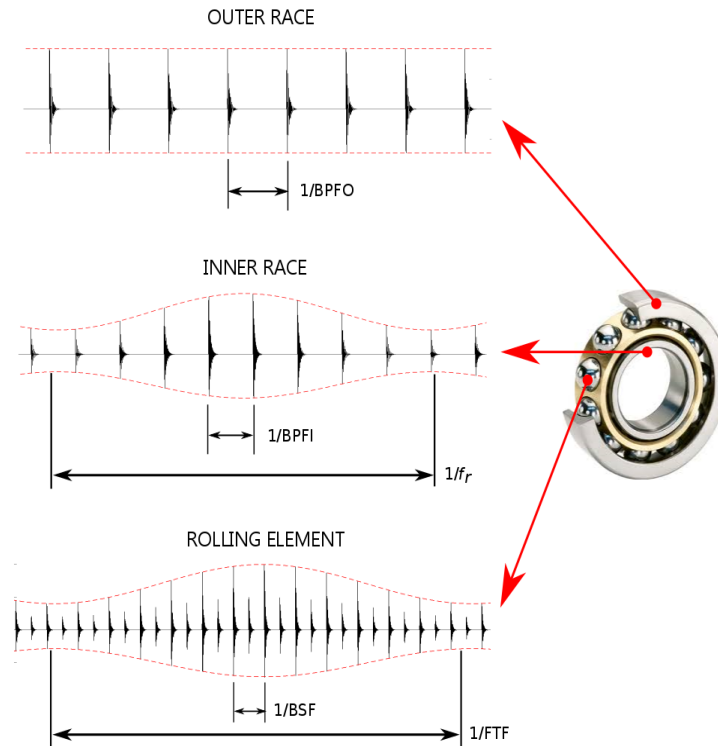


Fig. 1.1 Typical fault signals in rolling element bearings.

content corresponding to the excited structural resonances. Due to the load distribution, the series of impulse responses are simultaneously amplitude modulated by the passing period into and out of the load zone.

Figure 1.1 shows three typical fault signals with their corresponding “bearing characteristic frequencies”, which can directly reveal the occurrence and location of the bearing fault. In the case of stationary outer race defect, as the damage occurs on the surface of stator and in the presence of radial load, there exists uniform amplitude modulation. In the case of inner race defect, as the damage movement coincides to the shaft speed and the load distribution, it is amplitude modulated by the inner race rotation. In the case of rolling element defect, as the damaged rollers are guided by the cage, it experiences a periodic amplitude modulation at the cage speed; additionally, due to the double shock (inner and outer race) in each cycle, the even multiples of BSF are therefore dominant which will be further demonstrated by real data later.

Table 1.1 Bearing characteristic frequencies (d = bearing roller diameter; D = pitch circle diameter; n = number of rolling elements; θ = contact angle.)

Ballpass frequency, outer race (BPFO)	$\frac{n}{2}\Omega(1 - \frac{d}{D}\cos\theta)$
Ballpass frequency, inner race (BPFI)	$\frac{n}{2}\Omega(1 + \frac{d}{D}\cos\theta)$
Ball (roller) spin frequency (BSF)	$\frac{\Omega d}{D}(1 - (\frac{d}{D}\cos\theta)^2)$
Shaft rotation speed (f_{rot})	Ω
Fundamental train frequency (FTF)	$\frac{\Omega}{2}(1 - \frac{d}{D}\cos\theta)$

It is noteworthy that there exist two practical tips (invariant relations): $BPFO = FTF \times n$; $BPFI = n \times (f_{rot} - FTF)$, resulting in the sum of BPFO and BPFI always equal to $n \times f_{rot}$, regardless of the level of slip.

1.2 Problem statement

Based on the description above, a well-known and simple harmonic model for bearing signal (i.e. as can be seen in Fig. 1.2 (a)) is obtained by [3]:

$$y(t) = \sum_{i=-\infty}^{+\infty} h(t - iT)q(iT) + n(t) \quad (1.1)$$

where $h(t)$ represents the impulse response of system, $q(t) = q(t + P)$ indicates the periodic modulation of period P due to load distribution; T means the inter-arrival time between two consecutive impacts on the fault, index i denotes the occurrence of the i^{th} impact and $n(t)$ accounts for an additive background noise that embodies all other vibration sources.

Hence the power spectral density of the fault signal in Eq. (1.1) results in discrete spectrum series as seen in Fig. 1.2 (b) whose fault signature is distinctly carried in high frequency region of resonances excited by the internal impacts. Next, a conventional and popular method – squared envelope spectrum (SES) – is further executed to reveal the modulation frequency in order to obtain typical harmonic structure of fault signature in Fig. 1.2 (c). Unfortunately, it is merely an ideal model of the real world, and obviously limited for engineering applications. First, the transients produced by impacts are not strictly periodic due to random slips, possibly caused by speed fluctuations and variations of the axial to radial load ratio; second, the magnitudes of impacts are not uniform because of the random fluctuations, such as tiny changes in load and contact angle with time; third, the fault-induced

transients are often immersed in strong background noise which possibly derives from other rotating components. More precisely, two dominant challenges are highlighted:

- The masking effect from external sources (i.e. $n(t)$ in Eq. 1.1) due to the complex operating environment (i.e. bearings are most often associated with gears). Although the repetitive transients are usually carried in high frequency bands due to the high stiffness of bearings, they are fairly weak compared with surrounding noise or even completely submerged in other interfering signals (e.g. the deterministic components of other rotating parts such as gears);
- Uncertainties in the internal excitation mechanism (i.e. transients in Eq. 1.1). Even if the operating speed is controlled, there exists some kind of slight variations (i.e. random slips and fluctuations corresponding to intervals and magnitude of transients, respectively) caused by clearance and lubrication in actual operation, as illustrated in Fig. 1.2. It is proven that a small slip would completely destroy the harmonic structure of fault frequencies beyond the cut-off frequency; particularly at high speeds and light loads, slip may be as high as 50% [4–6].

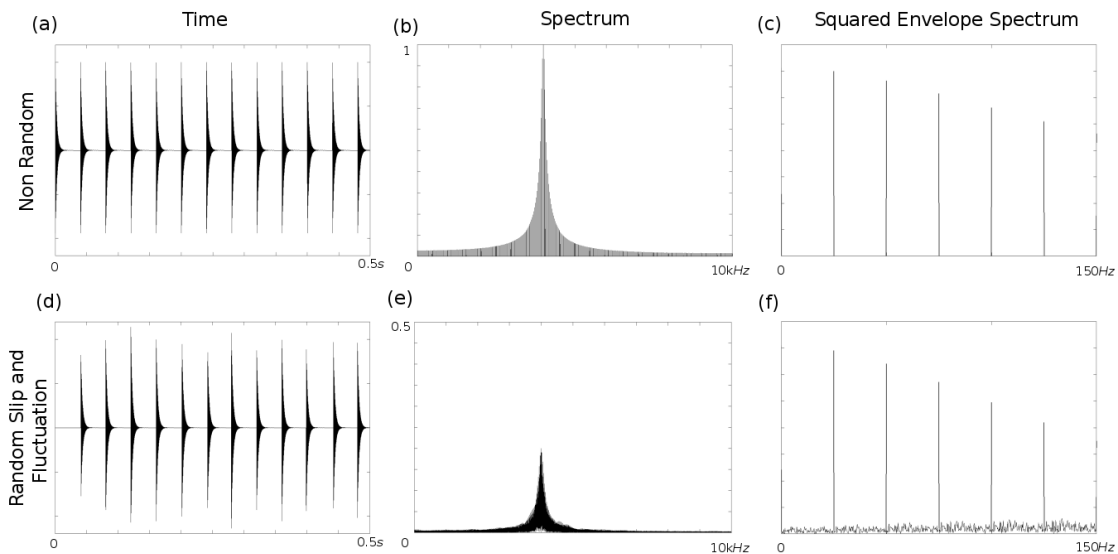


Fig. 1.2 Illustration of the influence of random slips and fluctuations: (a) and (d) stand for synthetic signals in time domain; (b) and (e) for raw spectra calculated by the Fourier transform; (c) and (f) for squared envelope spectra.

To see these clearly, a more realistic model for bearing signal is presented below:

$$y(t) = \sum_{i=-\infty}^{+\infty} h(t - iT - \tau_i)q(iT)A_i + n(t) \quad (1.2)$$

where τ_i and A_i accounts for the uncertainties on the arrival time and on the magnitude of the i^{th} impact, respectively. Without loss of generality, $\{\tau_i\}_{-\infty}^{\infty}$ and $\{A_i\}_{-\infty}^{\infty}$ are modelled as two random sequences $\tau_i \sim \mathcal{N}(\mu_\tau = 0, \sigma_\tau = 0.02T)$ and $A_i \sim \mathcal{N}(\mu_A = 0, \sigma_A = 0.1)$ in Fig.1.2 (d).

For all these reasons, the purpose of this thesis is to develop novel stochastic models whereby the topic involves the feature extraction, fault detection and identification, severity assessment and full-band time signal recovery in the field of machine condition monitoring and fault diagnostics.

1.3 State of the art

According to aforementioned mechanism – on operating bearings, feature extraction of fault-induced transients is an essential and crucial task for the following, i.e. fault detection and identification, particularly for precise identification of damage such as in inner race, outer race or rolling element. This is especially true in the incipient stage, so that various diagnostic algorithms have been proposed for decades [2, 7–16].

Traditionally, they can be divided into three categories: time-domain, frequency-domain and time-frequency domain. In particular, the characteristic features of time waveforms are rather visible and intuitive, because they are directly obtained from the acquired measurement, such as the Time Synchronous Average (TSA) [7], Autoregressive Model (AM) [8], Matching Pursuit (MP) [9] and so on [10]. In the time domain, fault detection mainly involves scalar quantities, i.e. variance, skewness, kurtosis, or some complex combination or variation of them – commonly used statistical quantities are shown in Table A.2. In addition, vibration signal can be decomposed into the spectral content so as to reveal its sinusoidal composition, whereby one can see the bearing fault signature from the vector quantities such as power spectral density, envelope analysis or cepstrum prewhitening [2, 4, 11]. More recently, various time-frequency analysis methods have been developed and applied to machinery fault diagnosis, such as the Short-Time Fourier Transform (STFT) [12, 13], Wigner-Ville distribution [17, 18], Wavelet Transform (WT) [19–21] and high-order spectral analysis [22]. All of them can benefit from the time-frequency framework that reveals the spectral features while identifying their variant occurrences in time, especially for representations of non-stationary signals. Although they have proven to succeed in many specified situations,

in the field of bearing fault diagnosis, it still requires ongoing developments for addressing a more general issue: spectrum feature extraction, automatic fault detection, full-band time signal recovery to reduce the demand on the user's expertise.

So far the discussion has primarily focused on the physical quantity or phenomenon that derives from tailored transforms (i.e. Fourier basis, Gabor transform, various types of wavelets or frames) in the *analytic* route. Such frame usually indicates significant physical meaning so as to lead to fast and effective algorithms, but lack adaptability and flexibility, i.e. they can not properly match the structure of the analyzed data in general. Nowadays, numerous signal analysis methods for machinery fault diagnosis have been investigated and used in the *learning-based* route [9, 10, 14–16], which infers the data-driven transforms obtained from signal realizations via machine-learning techniques [23] or advanced theory of stochastic processes [6]. Contrary to the *analytic* way, the *learning-based* route driven by observed data undergoes a high computational cost, but enjoys a better degree of freedom for signal representations, such as the shift-invariant sparse coding (SISC) model [24], group-sparsity model [25] and so on [26, 27].

This thesis thereby investigates both *analytic* and *learning-based* routes on sparse representations of vibration signal. Chapter 2 intends to share and merge the information on observations and prior knowledge in the general sense. With the idea of decomposing signals into some “sparse structure”, novel stochastic models are explored and applied in Chapter 3 and Chapter 4 for rolling element bearing diagnostics.

1.4 Focus and Contribution

This thesis focuses on sparse representations of vibration signal which aims to explore intelligent methods in rolling element bearing diagnostics; and it mainly makes four contributions as following:

- This thesis presents a literature survey on sparse representations of vibration-based signals for machinery fault diagnosis. First, an appealing study on primary visual cortex is investigated, of particular interest is the success of applying *sparse coding* model; and its contributions are further summarized by three aspects relative to vibration signals. The sparsity-based models are therefore studied with two crucial topics: on one hand, the sparse coding process that involves how to measure the sparsity of signal of interest; on the other hand, the dictionary designing process which can be divided into *analytic* and *learning-based* route. By following the review of sparse representations in machinery fault diagnosis, the conclusion on using “sparsity” is drawn with three crucial points.

- A new stochastic model is introduced to address this issue: it introduces a hidden variable to indicate the occurrence of the impacts and estimates the spectral content of the corresponding transients together with the spectrum of background noise. This gives rise to an automatic detection algorithm – with no need of manual prefiltering as is the case with the envelope spectrum – from which fault frequencies can be revealed. The same algorithm also makes possible to filter out the fault signal in a very efficient way as compared to other approaches based on the stationary assumption. The performance is investigated on synthetic signals with a high noise-to-signal ratio and also in the case of a mixture of two independent transients. The effectiveness and robustness of the method are also verified on vibration signals measured on a test-bench (gears and bearings). Results are found superior or at least equivalent to those of conventional envelope analysis and fast kurtogram.
- A new method is proposed to extract cyclostationary (CS) signals masked by interfering signals. First, it introduces a stochastic model that regularizes the second-order statistical descriptors as hidden variables so as to recover the CS component characterized by pre-set cyclic frequency. Meanwhile, it provides a CS indicator to assess the level of CS components along carrier frequency. The validity of the proposed scheme has been demonstrated on synthetic and experimental cases. Of particular interest is the robustness on experimental dataset and superior extraction capability over the conventional Wiener filter. It not only deals with the bearing fault at an incipient stage, but it even works for the installation problem and in the case of two sources, i.e. bearing and gear faults. Eventually, these experimental examples evidence its versatile usage on diagnostic analysis of compound signals.
- A benchmark study on proposed fast estimator of the spectral correlation, the *Fast Spectral Correlation* (Fast-SC), has been proposed and applied to a big data set. It contains plenty of experimental situations that ranges from very easily diagnosable to not diagnosable. Of particular interest is the cases that display unexpected mechanical symptoms, where the conventional methods (e.g. envelope analysis, cepstrum prewhitening and benchmark method in Ref. [11]) are obviously deficient. Finally the proposed Fast-SC is systematically tested in all the cases and its robustness and practicality have therefore been validated by the given result tables. The improved results therefore demonstrate more details on the CWRU data set so as to make the Fast-SC a more widely spread tool in condition monitoring. Another contribution of this work is to move forward the benchmark study of public and commonly used data

set – Case Western Reserve University (CWRU) Bearing Data Center – by uncovering its own unique characteristics.

1.5 Outline

The relevant works of this thesis are divided into six chapters:

Chapter 1 provides an overview of the issue with the research background, problem statement and the state-of-the-art in machinery fault diagnosis. It also highlights some interesting points to develop in the following contents.

Chapter 2 concerns the literature review on sparse representations of vibration signals. With the idea of decomposing signals with a “sparse structure”, a historical background on sparse coding is first provided. Next, a literature survey leads to the state-of-the-art on sparse representations particularly for vibration signals. Eventually, a few crucial points are drawn for its applications on machinery diagnostics – an interpretation on the Bayesian viewpoint is given so as to illuminate the following explorations.

Chapter 3 exploits a novel stochastic model that works under the assumption of non-stationary regime. It allows automatic detection of bearing faults without need of manual prefiltering. In addition, it also makes possible a full-band reconstruction of repetitive transients in the time domain. By discussing the initialization and parameter selection, its effectiveness and robustness are verified on vibration signals measured on a test-bench (gears and bearings).

Chapter 4 proposes an extraction scheme for CS signals particularly in the presence of competing sources. It first introduces a periodic-variance based stochastic model to extract the feature, recover the time signal, and detect the fault. The performance of the proposed scheme is demonstrated on synthetic and experimental cases.

Chapter 5 gives the benchmark survey on using the fast computation of the spectral correlation. It aims to make the spectral correlation a more widely spread tool in condition monitoring, thereby it has selected a big data set and tested all the cases. The performance of proposed Fast-SC is therefore assessed by the partial or non-diagnosable cases in Ref. [11]. In addition, the validating experiment is also demonstrated on non trivial vibration signals (very weak bearing signatures).

Chapter 6 gives the conclusion of the thesis and discusses the perspectives of future work.

Chapter 2

Literature review on sparse representations of vibration signal

Representing signals compactly with overcomplete dictionary has proven to be advantageous for many applications such as pattern recognition, image processing, machine learning, signal processing, and computer vision, etc [28–33]. Sparsity is a desired property that gives rise to an efficient representation for uncovering the intrinsic structure in signals, because it possesses a higher degree of statistical independence among its outputs. For machinery fault diagnosis, sparse representations of vibration signal is still a new and challenging topic [23, 26]. Hence, in this chapter, a literature review on this topic has been investigated based on current literature and furthermore a preliminary conclusion concerning its application is drawn for machinery fault diagnosis.

2.1 Derivation of sparse code

Sparse representation was originally inspired by the research on primary visual cortex (area V1) which uses a sparse code to efficiently represent natural scenes (Olshausen and Field, 1996) [34–36]. Here, their works are briefly summarized and categorized into three points:

- First, it was found that the spatial receptive fields of simple cells in the human vision system can be characterized as being *localized*, *oriented*, and *bandpass*; and these response properties of visual neurons can be accounted for producing a sparse distribution of output activity in response to natural images.
- Second, a statistical model, named *sparse code*, was built to capture the “sparse structure” in terms of a collection of statistically independent events; and furthermore

a coding strategy was investigated by maximizing sparseness which proves to be sufficient to account for all three of the above properties.

- Third, of particular case of interest is when the code is overcomplete, an unsupervised learning algorithm was studied for training the basis functions.

Interestingly, these points happen to answer the three relevant questions to sparse representations of vibration signals: a) WHY does the sparse code for natural images succeed in the field of cortical simple cells; b) WHAT is the sparse model and c) HOW to solve it. With these questions, the rest of this chapter will explore the current literature in order to find a proper way of sparsely representing vibration-based signals for machinery fault diagnosis.

Along such lines, it is reminded that the goal of efficient coding is to find a set of subcodes that forms a complete code (that is, spans the image space) and results in the coefficient values being as statistically independent as possible over an ensemble of natural images. The reasons for desiring statistical independence have been summarized in [34], but can be re-emphasized by taking into account higher-order statistical structure in the data. The appropriate form for this structure is that it is “sparse”, meaning that only a few subcodes can be “active” out of a very large set, as illustrated in Fig. 2.1 (Fig. 1 in Ref. [35]). It was shown in [34] that when such a code is sought for natural images, the basis functions that emerge are qualitatively similar in form to simple cell receptive fields and also the basis functions of certain wavelet transforms.

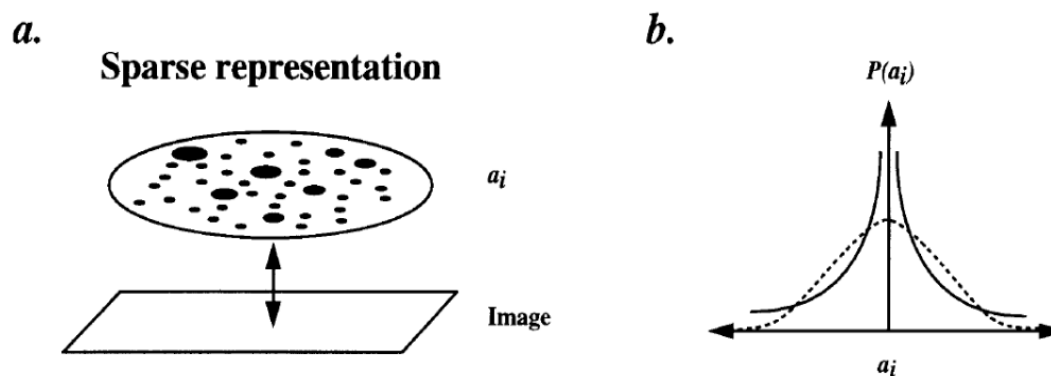


Fig. 2.1 (Fig.1 in Ref. [35]) Sparse coding. (a) An image is represented by a small number of “active” code elements, a_i , out of a large set. Which elements are active varies from one image to the next. (b) Since a given element in a sparse code will most of the time be inactive, the probability distribution of its activity will be highly peaked around zero with heavy tails. This is in contrast to a code where the probability distribution of activity is spread more evenly among a range of values (such as a Gaussian).

In summary, this strategy, referred to as “sparse coding”, could possibly confer several advantages. First, it allows for increased storage capacity in associative memories; second, it makes the structure in natural signals explicit; third, it represents complex data in a way that is easier to read out at subsequent levels of processing; and fourth, it saves energy [36].

2.2 Sparsity-based models

Signal models are a cornerstone of contemporary signal processing, because they are a fundamental tool for facilitating this distinctiveness of the interesting signals. A good signal model formulates a mathematical description of the family of interesting signals, $\Omega \in \mathbb{R}^N$, which allows to distinguish them from the rest of the signal space. There are various mathematical forms of signal models, one of the simplest and most common forms is as a penalty function

$$R(x) : \mathbb{R}^N \rightarrow \mathbb{R}^+ \quad (2.1)$$

which assigns smaller penalties to signals more likely to belong to Ω . It is preferable to identify signal models with the definition in Eq. 2.1, which allows more general constructions. In statistical theory, such a penalty function gives rise to some a-priori probability distribution assumed on the signal space,

$$P(x) = \frac{1}{Z} \cdot e^{-R(x)} \quad (2.2)$$

where the penalty function R determines the shape of the distribution and Z is a normalizing constant. For instance, choosing $R(x) = \log(1 + x^2)$ corresponds to specifying a Cauchy distribution for the prior, which has the desired sparse shape as well as these will be later discussed in Table 2.2 (Table 1.1 in [37]).

Let return back to the realistic model for the vibration signal in Eq. 1.2, reforming it as

$$y(t) = x(t) + n(t) \quad (2.3)$$

where $x(t)$ indicates the fault-induced transients which is of particular interest for fault detection. Separating it from measurement $y(t)$ – contaminated by strong background noise – is an impossible task without further assumptions on $x(t)$. Hence it introduces another important use of signal models – denoising problem regularization – by which the missing information can be filled in, such as the prior knowledge of “sparse structure” on the signal of interest $x(t)$. Specifically, by penalizing undesired signals, the penalty function $R(x)$ gives

rise to the following cost function:

$$\hat{x} = \operatorname{argmin}_{\hat{x}} \frac{1}{2} \|y - \hat{x}\|_2^2 + \lambda R(\hat{x}) \quad (2.4)$$

where $\lambda > 0$ is a regularization parameter balancing the fidelity and regularity terms. The first term measures how well this model describes the signal, and it is selected to be the mean square of the error between the measurement and the reconstructed signal. The second term assesses the sparseness of the reconstructed signal by assigning a penalty function depending on how activity is distributed (i.e. two common routes: *synthesis* model penalizes the coefficients of the reconstructed signal; *analysis* model penalizes inner products of the dictionary atoms and the reconstructed signal). In terms of Bayesian statistics, considering Eq. 2.2 as a prior probability distribution, the optimization process in Eq. 2.4 can be interpreted as a maximum-a-posterior (MAP) estimator of $x(t)$ [38, 39]. As can be seen, $R(x)$ in the above expresses all our knowledge about the set Ω , and its accuracy directly determines the success of the model.

To motivate the definition of the penalty function $R(x)$, let consider the desired property – sparse representation – which has proven to be of wide existence in natural images. All the notion focuses on the assumption of a higher degree of statistical independence of the desired signal, which gives to filter out the other statistical sources, such as the background noise $n(t)$. In other words, the sparsity-based model imposes a “sparse structure” through a sparsifying transform $x \rightarrow \gamma(x)$, where $\gamma(x) \in \mathbb{R}^L$ may have a different length than x (specifically, $L \gg N$ in image processing). For signals in Ω , the representation $\gamma(x)$ is expected to be sparse, in the sense that its sorted coefficients decay rapidly; for signals not in Ω , the representation vectors should become denser. Concerning the sparsifying transform $\gamma(x)$, there are mainly two ways to understand the sparsity: *synthesis* and *analysis* model, which will be further introduced in the following content.

Given the transform $\gamma(x)$, the sparsity of γ describes the estimated likelihood of $x(t)$ belonging to Ω . Now let consider the measure of *sparseness* $C(\gamma)$, which penalizes denser representations:

$$R(x) = C(\gamma(x)). \quad (2.5)$$

The purpose of regularizing the sparse signal is to approach the heavy-tailed probability distribution as displayed in Fig. 2.1 (b). Therefore it penalizes more the non-vanishing small coefficients, while tolerating a limited number of large ones. Such examples of robust functions include the Huber, Cauchy, and Tukey functions, as well as the family of l^p cost function with $0 \leq p \leq 1$, as shown in Table 2.2 (Table 1.1 in [37]). It is notice that the l^1 norm is an exception, which equally penalizes all magnitudes of coefficients. As such, it




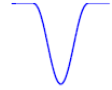
ℓ^p	$\rho(x) = x ^p$	
Huber	$\rho(x) = \begin{cases} x^2/2 & x \leq c \\ c(x - c/2) & x \geq c \end{cases}$	
Cauchy	$\rho(x) = \log(1 + (x/c)^2)$	
Tukey	$\rho(x) = \begin{cases} 1 - (1 - (x/c)^2)^3 & x \leq c \\ 1 & x \geq c \end{cases}$	

Fig. 2.2 (Table 1.1 in [37]) Some robust penalty functions. For all cases, $C(\gamma) = \sum_i \rho(\gamma_i)$.

establishes the boundary between robust and non-robust sparsity measures, which has been widely applied in image processing [40, 33].

In terms of the sparsifying transform $\gamma(x)$, the literature splits into two routes as follows: *synthesis* and *analysis* model which probably give rise to different physical or geometric meanings quite depending on applications.

2.2.1 Synthesis model

Over the past two decades, various models have been proposed and investigated to improve the solution of sparse representations [29–31, 40]. The *synthesis* model is one of the most common descriptions of signals, in which the signal x is represented as a linear combination of the atoms of dictionary D :

$$x = D\gamma_s, \quad (2.6)$$

where the coefficients γ_s are assumed to be *sparse*.

A graphical illustration of the *synthesis* model is presented in Fig.2.3, which provides a preliminary result for sparse representation of vibration signals¹. There are many studies and appealing applications based on this model due to its more intuitive and versatile structure [29–31, 40], it is a mature and stable field with clear theoretical foundations.

¹In the case of strong background noise, none of unsupervised learning algorithms can avoid extracting features of interfering signals. In other words, the trained dictionary probably contains some atoms from background noise $n(t)$.

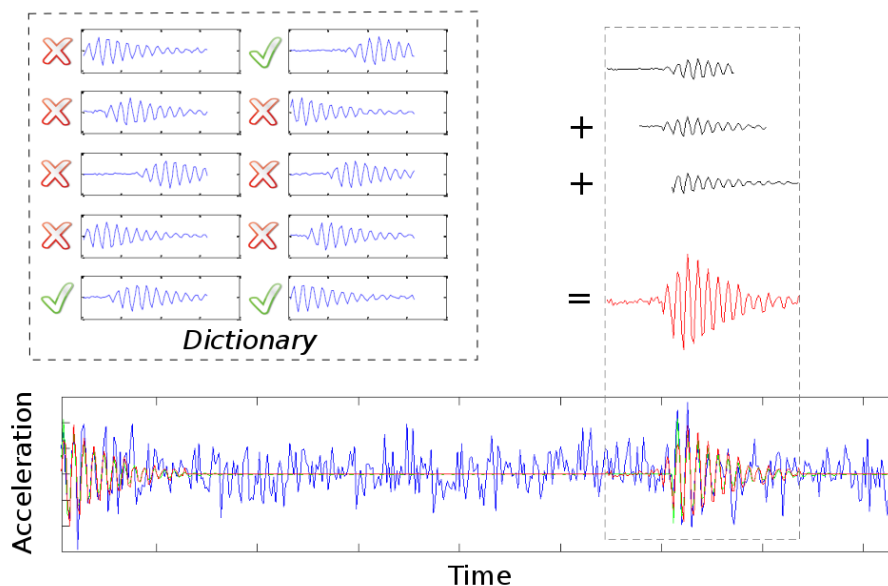


Fig. 2.3 Graphical illustration of the *synthesis* model, which represents a sparse vibration signal as a linear combination of three activated atoms in *Dictionary*. As can be seen, noisy measurements in blue, reconstructed signal in red, reference signal in green, activated atoms in black. It is notice that not all the atoms are displayed in the block of *Dictionary*, alternatively, only ten identifiable waveforms of trained atoms are displayed.

2.2.2 Analysis model

Interestingly, the *synthesis* model has a “twin” that takes an “*analysis*” point of view. This alternative describes the signal x via its inner products with the dictionary atoms Ω ,

$$\gamma_a = \Omega x, \quad (2.7)$$

where the analyzed vector γ_a is expected to be *sparse*. There are also many successful applications based on this model [28, 38, 41–43]. Particularly, the superiority of the *analysis* model in signal denoising was demonstrated in [38].

2.2.3 Comparison and evolution

There are some papers on the comparison between the two models [38], which indicate that for a square and invertible dictionary, the *synthesis* and the *analysis* models are equivalent with $D = \Omega^{-1}$. When the dictionary forms a basis, it is said to be *complete* in linear algebra theory. In this case every signal has a unique representation, such that in *synthesis* model $x = D\gamma_s$, where $\gamma_s(x) = D^{-1}x$ can be equivalently viewed as the inner products of x and the atoms of $\Omega = D^{-1}$, known as the *bi-orthogonal* dictionary [37]. A conclusion was drawn that

although when reducing to be *complete* and *under-complete* formulations the two models become equivalent, in their more interesting *overcomplete* formulation they are different [38].

For instance, in image processing, of great interest is the case when the dictionary is *overcomplete* ($L > N$) [35, 38, 41–43]. As it was explained in [35], the reason why it is more pertinent is that an *overcomplete* dictionary will allow greater flexibility in matching the generative model to the input structure. This is especially important for images, because there is little reason to believe *a priori* that images are composed of N discrete independent causes (where N is the dimensionality of the input). Hence, it allows for smooth interpolation along this continuum. Concerning the denoising problem, using a sparse and redundant representation as driving forces has drawn a lot of research attention in the past decade [37, 38, 40], such as the *sparseland* model in [40]. In the following contents, unless otherwise mentioned, the *overcomplete* dictionary will be exclusively investigated.

It is worthy to mention that there are also some other appealing models, such as double-sparse model, which employs a pair of analysis and synthesis dictionaries for thresholding-based image recovery [44]; structured sparse model, which is based on Gaussian mixture models, estimated via a computationally efficient MAP-EM algorithm [45]; shift-invariant sparse coding model, which allows a sparse representation using basis atoms in all possible time shifts [46, 47]; group-sparsity model, which addresses signal denoising when large-amplitude coefficients form clusters (groups) [48, 49]; convolutional sparse coding model, which provides a multi-layer model by imposing a special structure – a union of banded and circulant matrices – on the involved dictionary [50].

2.3 Algorithms for sparse representation

With prespecified models, the success of applications is dependent on the designed transform which gives rise to a sparse distribution. In particular, it primarily depends on the designed set of basis atoms, which is the so-called dictionary in machine learning. The dictionary is arranged as a matrix, with the atoms constituting its columns or rows (*synthesis* or *analysis* model, respectively), i.e. $D = [a_1 \ a_2 \ \dots \ a_L] \in \mathbb{R}^{N \times L}$ or $\Omega = [a_1 \ a_2 \ \dots \ a_L]^T \in \mathbb{R}^{L \times N}$. Both are composed of L atoms, wherein each atom of the dictionary is noted as $a_l \in \mathbb{R}^{N \times 1}$.

There are two routes for designing a sparse transform, either a prespecified set of functions or trained set of functions driven by a given set of signal examples. In this thesis, let take *analytic* dictionary and *learning-based* dictionary for notation. Choosing an *analytic* dictionary is appealing because in many cases it leads to simple and fast algorithms for the estimation of sparse representations. In addition, a prespecified transform also specifies a

clear physical meaning or typical waveforms that closely associates with signal analysis. This is indeed the case for overcomplete wavelets [51], curvelets [52], contourlets [53], short-time Fourier transforms [54], and so on. However, on the other hand, the *learning-based* route also gives rise to some interesting properties such as an explicit expression, flexibility and adaptability to the data itself. There are also some algorithms for this type, such as Principle Component Analysis (PCA) [55], Generalized PCA [56], the K-SVD [33], and others. No matter which route is chosen, there exists one important criterion to validate its performance – by which giving rise to a “sparse structure”. To reduce the demand on the user’s expertise, the *learning-based* route is one hot and challenging choice, which has attracted a lot of researches and led to the state-of-the-art [23, 33, 38, 40, 43–45]. In this work, it is preferable to exploit the *learning-based* dictionary as the focus of the following contents.

Concerning the optimization problem in Eq. 2.4, it is not jointly convex, but convex with respect to one variable while holding the other one fixed. Thus, it can be iteratively solved by two convex optimization problems: regularizing the coefficients γ with the dictionary fixed, and solving the atoms of the dictionary with the given coefficients γ . Especially, both of them are key topics in sparse representations which will be further investigated in the following contents.

2.3.1 Sparse coding

Given the *overcomplete* dictionary D , if it is a full-rank matrix, there will be infinite solutions with respect to $x = D\gamma_s$. Therefore the sparsity measure $C(\gamma)$ must be set to pursue the sparsest possible representation which gives rise to the objective function

$$\gamma_s = \underset{\gamma_s}{\operatorname{argmin}} C(\gamma_s) \quad \text{Subject To} \quad x = D\gamma_s. \quad (2.8)$$

This process corresponds to the so-called *sparse coding* step and is related to pursuit algorithms [32, 57, 58]. Of specific interest is the l^0 case, which leads to a strict sparse distribution by counting the number of non-zeros in the representation. Nevertheless, the l^0 minimization proves to be an NP-hard problem and sparsest representations is intractable in practice [59]. Many approximate solutions are therefore considered instead, such as the orthogonal matching pursuit (OMP) algorithms [57], basis pursuit (BP) [32], focal underdetermined system solver (FOCUSS) [58] and so on². A common and efficient way to penalize denser representations is the l^1 norm, which provides a powerful combination of robustness and convexity. It is also a stable approximation of the l^0 case, which leads to

²There are also some other ways to approximate the sparseness, e.g. to promote sparsity more strongly; some works employ non-convex regularization term, more details can be found in [49].

$$\gamma_s = \underset{\gamma_s}{\operatorname{argmin}} \|\gamma_s\|_1 \quad \text{Subject To} \quad \mathbf{x} = D\gamma_s. \quad (2.9)$$

As mentioned before, this approximate solution establishes the boundary between robust and non-robust sparsity measures, which leads to wide applications in image processing [33, 40]. The sparse coding algorithms provide not only a good approximation to the “sparse structure”, but a crucial step to the interactive procedure of dictionary learning.

2.3.2 Learning-based dictionary

More generally, the sparse coding gives rise to the following optimization problem,

$$\underset{\gamma, D}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \sum_{l=1}^L a_l \gamma_l\|_2^2 + \lambda \sum_{l=1}^L \|\gamma_l\|_1 \quad \text{s.t.} \quad \|a_l\|_2^2 \leq c, \quad 1 \leq l \leq L. \quad (2.10)$$

The normalization constraint c^3 prevents trivial solutions where γ_l becomes very small and a_l becomes very large. To solve the objective function in Eq. 2.10⁴, the first step is to compute the maximum-a-posteriori (MAP) estimates of the hidden coefficients $\gamma = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_L]^T \in \mathbb{R}^{L \times 1}$ and then estimate the learning-based dictionary $D = [a_1 \ a_2 \ \dots \ a_L] \in \mathbb{R}^{N \times L}$ by maximizing the complete likelihood function. This problem is convex in either γ_l or a_l (though not jointly convex in both). Hence it can be alternately solved by two convex optimization problems: 1) sparse coding to solve γ , which is a L_1 regularized least squares problem as shown in Eq. 2.9 and 2) redundant dictionary design to update D , which reduces the objective function 2.10 to a L_2 constrained optimization problem as demonstrated below:

$$\underset{D}{\operatorname{argmin}} \|\mathbf{y} - \sum_{l=1}^L a_l \gamma_l\|_2^2 \quad \text{s.t.} \quad \|a_l\|_2^2 \leq c, \quad 1 \leq l \leq L. \quad (2.11)$$

An intuitive way of understanding the algorithm is that it is seeking a set of γ_l for which the a_l can tolerate “sparsification” with minimum reconstruction error.

Concerning the *learning-based* dictionary, it will be referred, as expeditiously as possible, to the conventional algorithms: K-SVD [33] and analysis K-SVD [43], based on *synthesis* and *analysis* model respectively. The K-SVD is an iterative method for the *synthesis* model that alternates between sparse coding of the examples based on the current dictionary and an

³Norm constraints are important since there always exists a linear transformation of γ_l and a_l which keeps $\sum_{l=1}^L a_l \gamma_l$ invariant while allowing γ_l to approach zero. Another consideration is the comparison among the sparse coefficients, which accounts for the energy of each components.

⁴Rigorously speaking, it is just an intermediate of the complete process, which corresponds to maximizing the average log-likelihood of target signals $P(y | D)$ (more details to follow).

update process for the dictionary atoms so as to better fit the data. Particularly, it is flexible and can work with any pursuit method (e.g., basis pursuit, FOCUSS, or matching pursuit) [33]. The analysis K-SVD is parallel and similar to the previous one, it provides a training algorithm for the *analysis* dictionary – multiplies the signal, leading to a sparse outcome [43]. It is noteworthy that there are also some efficient algorithms to address the sparse coding, such as the feature-sign search algorithm [60] and CoefROMP algorithm [61].

2.4 Applications on rolling element bearing diagnostics

In recent years, the idea of sparsely representing vibration signals with pre-trained redundant dictionary has grabbed a lot of attention in machinery fault diagnosis [9, 10, 14–16, 23–27, 62, 63]. The procedure of fault diagnosis starts with data acquisition, followed by feature extraction, fault detection and identification as shown in Fig.2.4. As can be seen, feature extraction has the responsibility of extracting characteristic waveforms from the acquired measurement, which motivates the following classifier to recognize the type of fault.

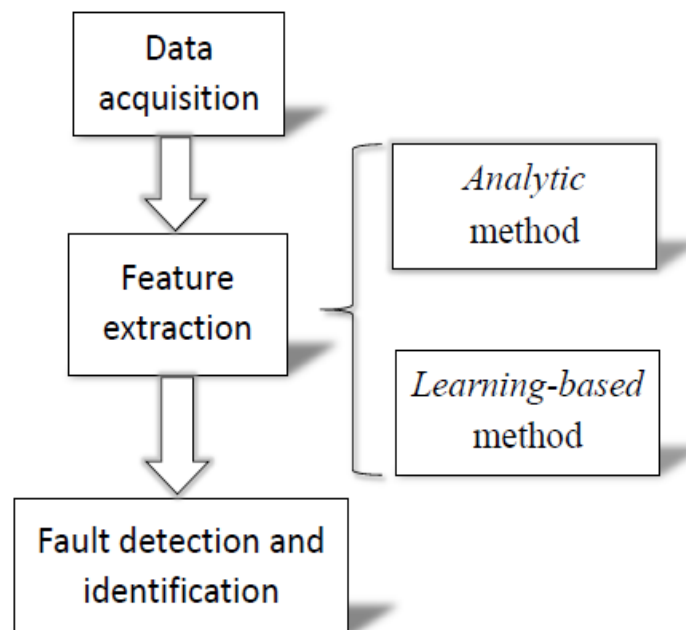


Fig. 2.4 General vibration fault diagnosis procedure.

Concerning the route of feature extraction, the current literature mainly consists of *analytic* and *learning-based* methods as mentioned in Section 1.3. In terms of *analytic*

method, the wavelet-based methods have been widely studied for decades and have made great progress in fault diagnosis. In particular, reviews of the wavelet-based methods were presented and various applications have been proposed in Ref. [19–21]. These proposed models match the vibration signals of bearing faults to some degree, but a one-to-one correspondence between the parameters in the models and the parameters and operational status of the analyzed bearing is nearly impossible. Especially in engineering applications, such methods are limited in their abilities to adapt to different types of data. In other words, special cares must be taken in choosing the basis functions, and the success of such dictionaries in applications depends on how suitable they are to describe target signals. To reduce the demand on the user's expertise, the *learning-based* route is one promising and challenging choice. Hence the sparsity-based feature extraction techniques – often associating with novel models – are investigated which can enable the following step to achieve the machinery fault diagnosis.

Along these lines, a number of studies have attempted to train unsupervised learning dictionary on vibration signals which derived from real data instead of predefined transform – in the hope of extracting the feature of fault-induced signals [9, 14, 15, 23]. The sparse representation was first illustrated in [14], which exploited the sparsity of vibration signals in the time-frequency domain in order to detect and extract fault features. For extracting the impulse features of damaged gear vibration, four sparse representation algorithms have been compared to analyze the vibration signals of both healthy and faulty gearboxes [15]. According to the characteristics of bearing faults, an adaptive matching pursuit method using an adaptive impulse dictionary is presented [9]. Based on the sparse model of compressed sensing, a sparse extraction of impulses by an adaptive dictionary scheme has been proposed to detect gearbox fault [23]. More recently, an on-line dictionary learning and sparse approximation denoising algorithm (ODL-SAD) has been proposed that intends to replace the traditional wavelet transform-based methods in which the performance is greatly impacted by the mother wavelet [16].

Another interest in research focuses on developing novel models – by imposing constraints or assumptions on vibration signals to achieve a good result of fault detection [10, 24–27, 62, 63]. The driving force behind different models is *sparsity* – the rapid decay of the representation coefficients over the dictionary – which results in a more efficient structure motivated by the higher-level of statistical independence of target signals as follows.

Of special interest is the shift-invariant sparse coding (SISC) model [24], which was originally applied to model the underlying principle of the mammalian auditory system in coding natural sounds. In particular, it allows a sparse representation using basis atoms in all possible time so that it matches well with the fact – when the fault-induced contact occurs,

then the characteristic pattern would be excited at any time-location which is expected to be the impulse response of the system. With the help of shift-invariant sparse coding, the same defective impulse at different time locations can be represented by just one basis function which is used for vibration analysis. In particular, a redundant dictionary is successfully built by merging all sub-dictionary learning from one class of bearings data [10]. Following the SISC model, a sparse representation based latent components decomposition method is proposed [27]. The main differences are that the fault feature is selected by the latent component without requiring any prior knowledge and a more efficient algorithm is applied – the feature-sign search algorithm [60]. After the property of shift-invariance has proven to be suitable to extract periodic impulses, the shift-invariant dictionary learning (SIDL) method is introduced as an adaptive feature extraction technique, which can extract a double-impulse in real data [26]. Then a signal can be decomposed into a collection of latent components, each of which is reconstructed by one basis atom and its corresponding time-shifts. The energy of each latent component is computed to form a sparse feature set, and finally the hidden Markov model (HMM) is used to identify the bearing fault type.

Another interesting work is based on the mentioned group-sparsity model [48], which aims to capture the signals where the variables (signal/coefficients) are not only sparse but also exhibit a clustering or grouping property. Then this model was initially improved by utilizing a non-convex regularization term to denoise group-sparse signals in speech enhancement [49]. By modeling vibration signals as only one repetitive group-sparse component, its potential for extracting periodic transient pulses is demonstrated in [25]. The periodicity-induced overlapping group shrinkage (POGS) approach is developed for rotating machinery fault diagnosis based on a periodic group-sparse signal representation. More recently, it has been generalized to extract multiple components which have different periods of the repetitive transient pulses corresponding to different fault frequencies [62]. In a similar work [63], the oscillatory fault features are extracted as periodically structured groups of coefficients in the time–frequency domain. The proposed algorithm allows non-convex regularization to promote the resulting sparsity, and it is guaranteed to converge to a local minimum even though the objective function is not convex.

In summary, the success of sparse representation in applications depends on how suitable it is to sparsely describe the vibration signals of interest, such as the localized faults of the gear teeth and bearing components in rotating machinery. Recently, the sparsity-based feature extraction techniques have been developed for machinery fault diagnosis and have demonstrated its potential in the *learning-based* route. According to prior knowledge about the signal behavior, different structures (models) have been applied in sparse representations

of vibration signals. However, in the perspective of applying it in the real world, it still requires further discussion as follows.

2.5 Discussion and conclusion

Concerning the sparse representation of vibration signals, it imposes a prior knowledge or inherent property aiming to activate the high-level structure of some periodic mechanisms. In addition, the desired “sparse structure” in vibration signals has proved to exist as illustrated in section 2.4. For instance, a series of transients can be effectively represented by impulsive activated coefficients based on a proper dictionary, thus yielding a sparse representation, as illustrated in Fig.2.3. These mechanical features are not limited to the oscillating, reciprocating, or other periodic waveforms, which sparsify uniformly smooth signals. Particularly, for such vibration signals, the Fourier and DCT transforms have proved good performances, as well as the wavelet transforms which sparsifies piecewise-smooth 1-D signals with a finite number of discontinuities [51].

Of particular interest is the *learning-based* dictionary instead of the *analytic* dictionary. Unlike the decomposition based on a predefined dictionary, each atom of the *learning-based* one is tailored to the data which leads to a high adaptability. Since this framework does not impose too many additional constraints or assumptions, it allows more flexibility to adapt the representation to target signals. For instance, the conventional sparsity-based model assumes that all the segments for training are independently distributed and the prior distribution of the coefficient γ_s is assumed to be a super-Gaussian (i.i.d.) distribution that favors sparsity. However, there are no further considerations on the structure (assumption) of dictionary or coefficients, e.g. some intrinsic coherence among dictionary elements or the periodic or repetitive patterns along time series, etc. With such a large degree of freedom, the sparse model seems more like a black-box, which may cause artifacts and loss of information, such as classifying some unwanted signals as sparse, or misclassify some signals in Ω as dense.

In rolling element bearing diagnostics, in the author’s viewpoint, there exists few crucial points that affect the success of its application:

- Can the signal of interest be characterized by the “sparse structure”? If it were a positive answer, then which model is more suitable (i.e. additive assumption or constraint in the specified structure) and how to measure the sparseness of vibration signals (i.e. penalty function or probability distribution)?
- If the signal of interest was sparse, then what is the sparsifying transform best adapted to vibrations (i.e. sparsity-based feature extraction: *analytic* or *learning-based* route)

and how to obtain it (i.e. for a predefined transform, it needs to specify the sparse transforms; for a learning-based transform, does it require more constraints to promote the sparsity)?

- Concerning the current literature, most of the researches focus on a redundant dictionary to extract features derived from the fault-induced transients; is it the only way to enforce a sparse representation of vibration signals? Or alternatively, is there some more efficient way to represent vibration signals, such as *complete* formulation.

In summary, sparse representation is a general principle for finding statistically independent components in vibration signals rather than a special and formulaic model for matching more physical characteristics. As a versatile model with a high degree of freedom, it embodies an interesting potential to deal with multi-component signals; however, the effectiveness of the technique is subjected to the ratio between impulses and other interference in training examples, especially when training examples are generated under complex mechanisms.

In the Bayesian viewpoint, the optimization problem in Eq. 2.8 results from a maximum-a-posteriori (MAP) estimation. The unknown coefficients of each sample are estimated as hidden variables by maximizing the posterior $P(\gamma_s | x, D) \propto P(x | D, \gamma_s)P(\gamma_s)$, i.e. two well-known pursuit algorithms: basis pursuit (BP) and focal underdetermined system solver (FOCUSS) [32, 58]. In order to calculate the probability of target signals arising from the sparsity-based model, $P(y | D)$, it requires to specify the prior probability distribution over the coefficients, $P(\gamma_s)$, as well as the probability of observations arising from a certain state of the coefficients in the model, $P(y | D, \gamma_s)$. Once these two distributions are specified, then the probability of the observations arising from the model is given by:

$$P(y | D) = \int P(y | D, \gamma_s)P(\gamma_s)d\gamma_s, \quad (2.12)$$

which assesses how closely the model matches the observations in the real world. Alternatively, it is related to the maximum likelihood estimation (MLE) of the dictionary D . In particular, the probability of an observation arising from a particular choice of coefficients, $P(y | D, \gamma_s)$, expresses the level of noise or uncertainty during the acquisition process. Without loss of generality, let assume $n(t)$ obeys a zero-mean Gaussian distribution, $\mathcal{N}(0, \sigma_n^2)$, then it comes:

$$P(y | D, \gamma_s) = \frac{1}{\sigma_n \sqrt{2\pi}} e^{-\frac{\|y - D\gamma_s\|_2^2}{2\sigma_n^2}}. \quad (2.13)$$

Hence, the goal of dictionary learning will be to find a set D that maximizes the average log-likelihood of target signals under a sparse prior distribution as in formula 2.10.

Along these lines, let remind WHY sparseness in vibration signals? One can see evidence for sparse structure in vibration signals by filtering $y(t)$ with filterbanks and selecting the best frequency band which maximizes the impulsive signal $x(t)$ with respect to background noise $n(t)$. The resulting output distribution of band-pass filtered signal typically shows high kurtosis (spectral kurtosis) – an effective measure of the “impulsiveness” hidden in the signal – which is indicative of sparse structure. In other words, the selected band specifies the impulsive component (super-Gaussian distribution) – composition of a mixture $x(t)$ and $n(t)$ – with high signal-to-noise ratio; inversely, the residual component (Gaussian distribution) only contains $n(t)$ as illustrated in Fig. 2.5.

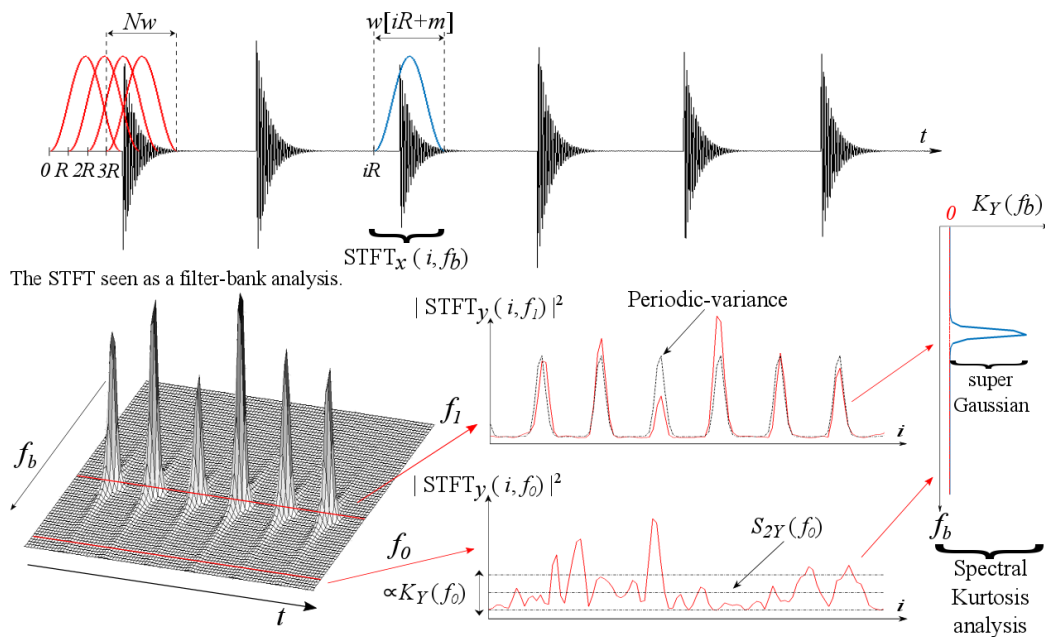


Fig. 2.5 Illustration of Spectral Kurtosis analysis: the repetitive series of transients passes a filter-bank analysis implemented by the STFT; as can be seen in $K_Y(f_b)$, there exist two different distributions which indicate the fault component (centered at frequency f_1) and the noisy component (centered at frequency f_0), respectively. It is highlighted that along the time instant the occurrence of transients can be characterized by means of a latent variable (i.e. the spectral mixture model in Chapter 3). In particular, for cyclostationary signals, its energy flow displays a periodical variation which indicates cyclostationarity in second-order statistics (i.e. the periodic-variance based model in Chapter 4).

However, the spectral kurtosis analysis provides a band-pass filter of repetitive transients, rather than a full-band one that contains all spectral contents. In addition, although the fault signatures are usually carried in high frequency bands because of the high stiffness of bearings, they are fairly weak compared with surrounding noise. In fact, the shown “sparse structure” (super-Gaussian distribution) is consisting of a mixture of the spectra of $x(t)$ and

$n(t)$. Hence the “impulsiveness” is actually hidden in vibration signals with respect to a linear mixture of two different distributions⁵; then seeking the fault-induced signal can be viewed as a sparse representation motivated by the *analysis* model.

This thesis addresses the issue by proposing novel stochastic models in the following Chapter 3 and 4; and by means of a new fast estimator of the spectral correlation in Chapter 5. It is noteworthy that all of them are rooted on the Short-Time-Fourier-Transform (STFT). One important reason, which arises in practical applications, is that the STFT of cyclostationary signals evidences periodic flows of energy *in* and *across* the frequency bins which promotes the way of sparsely representing vibration signals.

⁵It is proved that a linear mixture of two Gaussian random variables gives rise to a super-Gaussian random variable, more details are provided in Appendix B.

Chapter 3

Automatic Spectrum Matching of Repetitive Transients Based on Hidden Markov Model

Hidden Markov Models (HMMs) have been proposed in 1970s as statistical models for time series and so far they have been applied in a wide range of fields including speech recognition, computer vision, pattern recognition and many other areas. Numerous works based on HMM have been reported in fault diagnosis during the last decade [64–71]. However, they are mainly concerned with the use of HMMs as classifiers for fault or condition recognition [64, 65] [66, 67], which is different from the objective of modelling the vibration signal itself.

This chapter aims to introduce a HMM as a flexible stochastic model of some types of non-stationary signals. In particular, it will allow the automatic identification of a mixture of transients and of stationary background noise. In addition, it will make possible to filter out the fault signal by means of a time-varying filter and to estimate its full-band spectral content. The proposed model only assumes that each state obeys a stationary condition and relaxes the assumption of cyclostationarity used in other approaches [6, 72, 73]. It is also shown that it can deal with non-stationary regimes.

3.1 Spectral mixture model

This section begins with a presentation of the stochastic model and its corresponding assumptions. Then the HMM is introduced and its parameters are estimated by the Expectation-

Maximization (EM) algorithm. The reconstruction of the fault signal is addressed in section 3.2.

3.1.1 Model and assumption

The objective is to detect the informative signal $x(t)$ – which contains the fault – when it is contaminated by additive background stationary noise $n(t)$ which comprises all other vibration sources. Let denote the noisy measurement $y(t)$ as

$$y(t) = x(t) + n(t). \quad (3.1)$$

In the following, sampled signals with sampling frequency F_s and n temporal index will be considered as $x(n/F_s) \rightarrow x[n]$.

The STFT is used hereafter to decompose the vibration signal in the time-frequency plane. The STFT of signal $x(t)$ over a time interval of length N_w is defined as

$$STFT_x(i, f_b) = \sum_{m=0}^{N_w-1} w[m] \cdot x[iR + m] \cdot e^{-j2\pi f_b \frac{m}{F_s}} \quad (3.2)$$

where $\{w[m]\}$ denotes a positive and smooth N_w -long data-window which truncates a segment of the L -long signal $x(t)$ at time datum i ($i = 1, \dots, N$, $N = \text{floor}[(L - N_w)/R + 1]$) with window shift R ($1 < R < N_w$) and where $f_b = b \cdot \Delta f$ denotes the frequency (from 0 to $F_s/2$) with frequency resolution $\Delta f = F_s/N_w$ and bin index $b = 1, \dots, N_f$ with $N_f = N_w/2 + 1$.

In the proposed stochastic model, the phase information will play a significant role. Hence it is required to phase all the segments to zero at the beginning of the signal, at time instant $t = 0$. This phase correction can also be interpreted as the Gabor transform

$$X(i, f_b) = \sum_{n=0}^{L-1} w[n - iR] \cdot x[n] \cdot e^{-j2\pi f_b \frac{n}{F_s}} = STFT_x(i, f_b) \cdot e^{-j2\pi f_b \frac{iR}{F_s}}. \quad (3.3)$$

It is noteworthy that $X(i, f_b)$ is related to the “instantaneous complex envelope” of the signal of interest $x[n]$ described in both time and frequency. More precisely, its squared magnitude reflects the energy flow which is mapped by time index i and frequency f_b centered in a narrow frequency band Δf [74].

It is reminded at this juncture that a fault signal in its early stage is well modeled as a series of impacts that repetitively excite resonances of the bearing and of its receiving structure, in relatively high frequencies, and leading to damped impulse responses. It is assumed that transients can be well captured in segments of the STFT. Their rate of repetition is expected to embody the fault information. On the contrary, other vibration components

will be modeled as stationary noise spread all over the time-frequency plane. The following subsection introduces a two-state HMM to account for the different probability distributions of the transients and of the noise.

3.1.2 Hidden Markov model

A HMM is a probabilistic model of the joint probability distribution of random variables which depend on some hidden states. Hereafter, the measurement $Y(i, f_b)$ is merged along the spectral contents with respect to the i^{th} time instant and denoted by a vector, $\mathbf{Y}(i) = [Y(i, f_1) \dots Y(i, f_{N_f})]^T$. Next, vector $\mathbf{Y}(i)$ is further represented by a linear combination of K components – whose events are assumed mutually exclusive – denoted by $\mathbf{X}^k(i) = [X^k(i, f_1) \dots X^k(i, f_{N_f})]^T$ and contaminated by a noisy component $\mathbf{N}(i)$. All components in the model are allowed to have different probability distributions and are controlled by a vector of latent variables, $\boldsymbol{\zeta}(i) = [\zeta^1(i) \dots \zeta^K(i)]^T$, each of which acting as a switch taking only values 0 or 1. Thus, the proposed model reads

$$\mathbf{Y}(i) = \mathbf{X}(i)\boldsymbol{\zeta}(i) + \mathbf{N}(i) \quad (3.4)$$

where $\mathbf{X}(i) = [\mathbf{X}^1(i) \dots \mathbf{X}^K(i)]$ is a matrix consisting of K column vectors $\mathbf{X}^k(i)$ and $\mathbf{N}(i) = [N(i, f_1) \dots N(i, f_{N_f})]^T$.

a) One component model

In order to start simple, let us first consider from now on the model with only one component of interest, that is

$$\mathbf{Y}(i) = \zeta(i)\mathbf{X}(i) + \mathbf{N}(i). \quad (3.5)$$

Since the latent variable $\zeta(i)$ can take only two values, it is assumed to follow a Bernouilli distribution, $\zeta(i) \sim \text{Bernouilli}(\pi)$:

$$\begin{cases} p(\zeta(i) = 0 \mid \pi) = 1 - \pi \\ p(\zeta(i) = 1 \mid \pi) = \pi \end{cases}. \quad (3.6)$$

Here $\zeta(i) = 0$ means the presence of noise only, i.e. “State 0: $\mathbf{Y}(i) = \mathbf{N}(i)$ ”, whereas $\zeta(i) = 1$ indicates the presence of noise *and* the signal of interest, i.e. “State 1: $\mathbf{Y}(i) = \mathbf{X}(i) + \mathbf{N}(i)$ ”. Let us introduce the likelihood function:

$$p(\mathbf{Y}(i) \mid \zeta(i), \mathbf{C}_n, \mathbf{C}_x) \sim \mathcal{CN}(\mathbf{Y}(i); \mathbf{0}, \mathbf{C}_n + \zeta(i)\mathbf{C}_x) \quad (3.7)$$

where $\mathcal{CN}(\mathbf{Y}; \boldsymbol{\mu}, \mathbf{C})$ denotes the circularly-symmetric complex normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{C} applied to variable \mathbf{Y} . Without loss of generality, it is assumed that $\boldsymbol{\mu} = \mathbf{0}$ (as obtained after first centering the signal).

Meanwhile, all the unknown parameters of the proposed model are denoted as $\boldsymbol{\theta} = \{\mathbf{C}_n, \mathbf{C}_x, \zeta(i), \pi\}$. It is highlighted that the latent variables $\zeta(i)$ are hidden in the sense that they are not observed directly. This is what will make the estimation of parameters difficult. Assuming independent segments in the STFT, the complete log-likelihood function is evaluated from Eq. 3.7 as

$$\log L_C(\boldsymbol{\theta}) = \sum_{i=1}^N \log(p(\mathbf{Y}(i) | \boldsymbol{\theta})). \quad (3.8)$$

Developing further, one has

$$\sum_{i=1}^N \log(p(\mathbf{Y}(i) | \zeta(i) = 0, \mathbf{C}_n)p(\zeta(i) = 0 | \pi) + p(\mathbf{Y}(i) | \zeta(i) = 1, \mathbf{C}_n, \mathbf{C}_x)p(\zeta(i) = 1 | \pi)) \quad (3.9)$$

$$= \sum_{i=1}^N \log((1 - \pi) \times \mathcal{CN}(\mathbf{0}, \mathbf{C}_n) + \pi \times \mathcal{CN}(\mathbf{0}, \mathbf{C}_{x+n})) \quad (3.10)$$

where it has been assumed that all states are a priori equally probable.

The parameters $\boldsymbol{\theta}$ are estimated by maximizing the above likelihood function. In theory, this completely solves the problem since the estimated latent variable $\zeta(i)$ will then return the times of occurrence of the impacts on the faults and therefore the bearing characteristic frequency. Since it is difficult to find a closed-form solution, the EM algorithm [75] is used as an iterative method to find the maximum likelihood estimates. The EM algorithm makes use of the following quantities.

First, the posteriori probability distribution of the latent variable is formed as

$$p(\zeta(i) | \mathbf{Y}(i), \mathbf{C}_n, \mathbf{C}_x^1, \pi) = \frac{p(\mathbf{Y}(i) | \zeta(i), \mathbf{C}_n, \mathbf{C}_x^1)p(\zeta(i) | \pi)}{p(\mathbf{Y}(i) | \mathbf{C}_n, \mathbf{C}_x^1, \pi)} \quad (3.11)$$

where \mathbf{C}_x^1 denotes the covariance matrix of the 1st component. According to Eq. 3.11, the expectation of the latent variable $\zeta(i)$ given the measurement is then computed as

$$\mathbf{E}\{\zeta(i) | \mathbf{Y}(i), \mathbf{C}_n, \mathbf{C}_x^1, \pi\} = \frac{\pi \mathcal{CN}(\mathbf{0}, \mathbf{C}_{x+n})}{(1 - \pi) \mathcal{CN}(\mathbf{0}, \mathbf{C}_n) + \pi \mathcal{CN}(\mathbf{0}, \mathbf{C}_{x+n})} \quad (3.12)$$

where \mathbf{C}_{x+n} denotes $\mathbf{C}_n + \mathbf{C}_x^1$. After simple arrangement, this is expressed as

$$\mathbf{E}\{\zeta(i) \mid \mathbf{Y}(i), \mathbf{C}_n, \mathbf{C}_x^1, \pi\} = \frac{1}{1 + \frac{1-\pi}{\pi} \times e^{LLR}} \quad (3.13)$$

where LLR denotes the natural logarithm of the likelihood ratio between the noise and the component of interest as calculated by

$$LLR = \log \frac{|\mathbf{C}_{x+n}|}{|\mathbf{C}_n|} e^{-\mathbf{Y}(i)^H (\mathbf{C}_n^{-1} - \mathbf{C}_{x+n}^{-1}) \mathbf{Y}(i)}. \quad (3.14)$$

As a useful intermediate, the LLR may be interpreted as the likelihood that a transient is present at any time datum i . Therefore, it will be demonstrated as a powerful indicator for fault detection in section 3.3. Its Fourier transform will turn out a valid alternative to the SES to evidence bearing characteristic frequencies.

The EM algorithm is summarized in Fig. 3.1. It is noted that there are two factors to set: k_{max} indicates the maximum number of iterations and STOP-CRIT means the expected relative tolerance between $\hat{\mathbf{C}}_n^{[k+1]}$ and $\hat{\mathbf{C}}_n^{[k]}$. The estimation of the covariance of the signal of interest is finally obtained as $\hat{\mathbf{C}}_x^1 = (\hat{\mathbf{C}}_{x+n} - \hat{\mathbf{C}}_n)_+$ where operator $(\dots)_+$ keeps only the positive eigenvalues of a matrix.

With these estimated parameters, an automatic fault detection scheme and a time-varying filter for filtering out the signal of interest are proposed in the following sections.

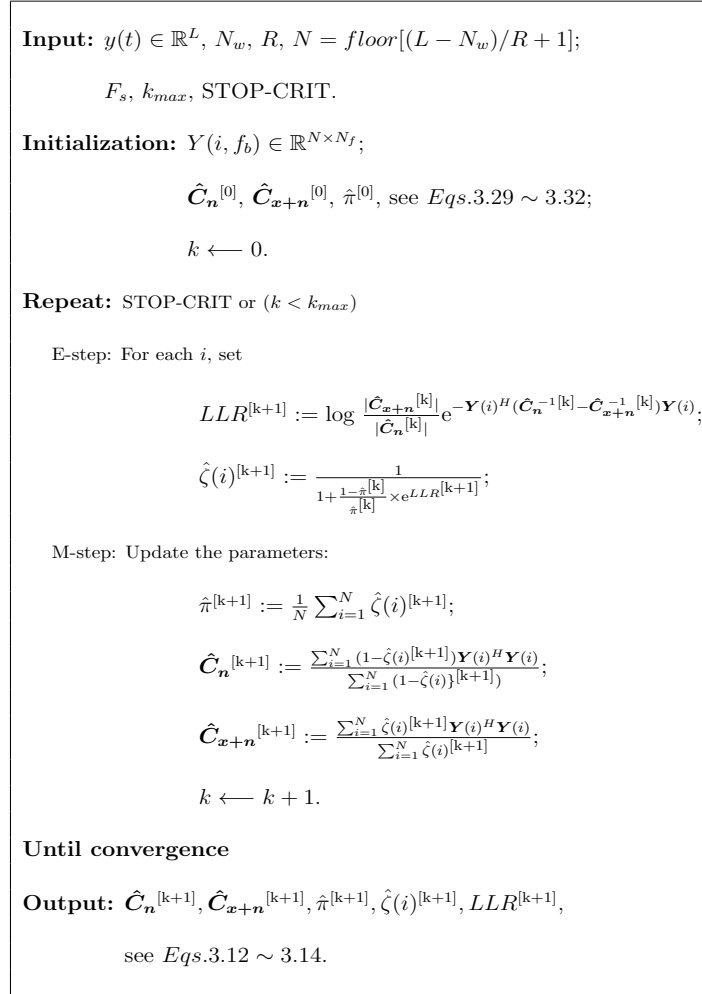


Fig. 3.1 Explicit steps of the EM algorithm to infer the parameters in the HMM.

b) *K* component model

Let us now consider the general case with $K > 1$ signals of interest, $\mathbf{X}^k(i)$, $k = 1, \dots, K$. In this case, the probability that two (or more) components occur together will be assumed so small that such events will be disregarded (this may be seen as an extreme sparse representation where only one state is allowed at a time). Therefore, under the mutually exclusive assumption, the occurrence of the k^{th} state is defined as

$$A_k = \{\zeta^k(i) = 1; \zeta^l(i) = 0 \mid 1 \leq l \leq K; l \neq k\} \quad (3.15)$$

where $\zeta^k(i) = 1$ indicates the presence of the k^{th} signal $\mathbf{X}^k(i)$. The pure noise case – i.e. $A_{K+1} = \{\boldsymbol{\zeta}(i) = \mathbf{0}\}$ – is denoted as the $(K+1)^{\text{th}}$ state. Thus, there is a total of $K+1$ possible states in the model. Therefore, the marginal probability distribution reads

$$p(\mathbf{Y}(i) \mid \mathbf{C}_n, \mathbf{C}_x^k, \pi^k, k = 1, \dots, K) = \sum_{k=1}^{K+1} p(\mathbf{Y}(i) \mid A_k, \mathbf{C}_n, \mathbf{C}_x^k) p(A_k \mid \pi_k) \quad (3.16)$$

where the k^{th} latent variable is assumed to follow a Bernoulli distribution, $\zeta^k(i) \sim \text{Bernoulli}(\pi^k)$:

$$\begin{cases} p(\zeta^k(i) = 0) = 1 - \pi^k \\ p(\zeta^k(i) = 1) = \pi^k \end{cases} \quad (3.17)$$

The posteriori probability distribution then reads

$$p(\boldsymbol{\zeta}(i) \mid \mathbf{Y}(i), \mathbf{C}_n, \mathbf{C}_x^k, \pi^k, k = 1, \dots, K) = \frac{p(\mathbf{Y}(i) \mid \boldsymbol{\zeta}(i), \mathbf{C}_n, \mathbf{C}_x^k) p(\boldsymbol{\zeta}(i) \mid \boldsymbol{\pi}^k)}{p(\mathbf{Y}(i) \mid \mathbf{C}_n, \mathbf{C}_x^k, \pi^k)} \quad (3.18)$$

Assuming mutually exclusive states, the expectation of the k^{th} latent variable is thus

$$\mathbf{E}\{\zeta^k(i) \mid \mathbf{Y}(i), \mathbf{C}_n, \mathbf{C}_x^k, \pi^k, k = 1, \dots, K\} = \frac{\pi^k \mathcal{E}\mathcal{N}(\mathbf{0}, \mathbf{C}_n + \mathbf{C}_x^k)}{\mathcal{E}\mathcal{N}(\mathbf{0}, \mathbf{C}_n) \prod_{k=1}^K (1 - \pi^k) + \sum_{k=1}^K \pi^k \mathcal{E}\mathcal{N}(\mathbf{0}, \mathbf{C}_n + \mathbf{C}_x^k)} \quad (3.19)$$

Figure 3.2 illustrates the situation with $K = 2$, which involves three states.

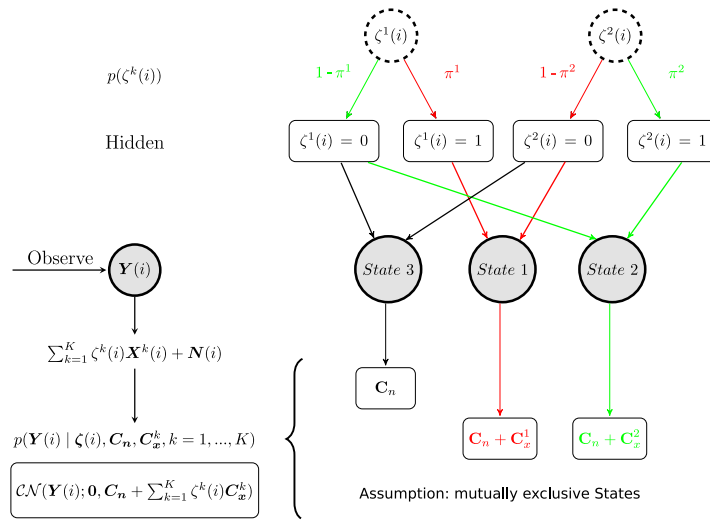


Fig. 3.2 Graphical model in the case $K = 2$.

3.1.3 Structured covariance matrix

The proposed model assigns to the signal of interest a specific probability distribution with a given covariance matrix. For the sake of generality, five different structures of covariance matrices are investigated hereafter and tested on simulated signal. More details and further comparison among them will be demonstrated in subsection 3.3.3. It is noteworthy that our goal is to simplify the model and achieve as low computational cost as possible.

First, since the frequency components of a stationary process are uncorrelated, the covariance matrix of stationary background noise is diagonal, i.e. $\mathbf{C}_n = \text{diag}(C_n(f_1) \dots C_n(f_{N_f}))$.

On the other hand, several assumptions are possible for the transient part of the signal. Five possible structures are considered: full matrix, diagonal matrix, tridiagonal matrix, low rank matrix, Toeplitz matrix.

The first model is a full covariance matrix. This is obviously the most general assumption, yet not the easiest one to estimate due to the high number of free parameters. The other assumptions listed hereafter aim to simplify the model by imposing a given structure.

The second model is returned by a diagonal covariance matrix, i.e. $\mathbf{C}_x^1 = \text{diag}(\mathbf{C}_x^1)$, which only considers the auto-spectrum of the transients. This means that there is no correlation between different frequency components. Thus, this simple structure is expected to be the easiest to estimate.

The third model considers possible correlation between spectral components spaced apart by one frequency bin. This corresponds to a tridiagonal matrix, i.e. $\mathbf{C}_x^1 = \text{tridiag}(\mathbf{C}_x^1)$. It enlarges a little bit the concerned range of relevant spectral correlation as compared with the previous diagonal structure.

The fourth model is a low rank structure $\mathbf{C}_x^1 = \text{lowr}(\mathbf{C}_x^1, r)$. It allows full correlation between frequency components while still imposing a constraint on the complexity of the model that fits the data.

The fifth model is inspired by the covariance of stochastic processes modeled as the response of a causal linear time invariant filter to stationary white noise. This corresponds to a Toeplitz matrix with a constant diagonal structure in the time domain that leads to only $2n - 1$ degrees of freedom instead of n^2 . In the frequency domain, this reads $\mathbf{C}_x^1 = 2\mathcal{F}\{\text{toep}(2\mathcal{F}^{-1}\{\mathbf{C}_x^1\})\}$, where $2\mathcal{F}$ and $2\mathcal{F}^{-1}$ stand for the 2D Discrete Fourier Transform and its inverse and $\text{toep}(A)$ for the operator that imposes a Toeplitz structure, i.e. $A_{i,j} = A_{i+1,j+1} = a_{i-j}$.

In all cases, the structured covariance matrices are estimated by systematically imposing the required constraints (i.e. operators *diag*, *tridiag*, *lowr* or *toep*) in the iterations of the EM algorithm in Fig. 3.1.

The five aforementioned structures will be discussed and compared in subsection 3.3.3.

3.2 Reconstruction scheme for repetitive transients

The HMM provides a versatile tool for characterizing the non-stationarity of a signal from its second-order statistics only, regardless of the operating speed. With the estimated parameters $\hat{\boldsymbol{\theta}}^k$ of the HMM, the fault signal can be reconstructed in full band based on the latent variable vector $\boldsymbol{\zeta}(i)$ and the covariance matrix \mathbf{C}_x^k of transients.

First, let introduce the posterior probability distribution of the k^{th} signal of interest $\mathbf{X}^k(i)$ as

$$p(\mathbf{X}^k(i) | \mathbf{Y}(i), \hat{\boldsymbol{\theta}}^k) \propto p(\mathbf{Y}(i) | \mathbf{X}^k(i), \hat{\boldsymbol{\theta}}^k) p(\mathbf{X}^k(i) | \hat{\boldsymbol{\theta}}^k) \quad (3.20)$$

where $\boldsymbol{\mu}_x^k = [\mu_x^k(f_1) \dots \mu_x^k(N_f)]^T$ denotes the N_f -dimensional mean vector and \mathbf{C}_x^k the covariance matrix.

Next, for simplicity, let us assume that the covariance matrix $\hat{\mathbf{C}}_x^k$ is diagonal. Therefore, the problem decouples for each frequency bin. The posterior probability density at a given

frequency then reads

$$p(X^k(i, f_b) | Y(i, f_b), \hat{\boldsymbol{\theta}}^k) = \frac{e^{-\frac{|Y(i, f_b) - \zeta^k(i)X^k(i, f_b)|^2}{\hat{C}_n^k(f_b)}} e^{-\frac{|X^k(i, f_b)|^2}{\hat{C}_x^k(f_b)}}}{\pi^2 \hat{C}_n^k(f_b) \hat{C}_x^k(f_b)} \quad (3.21)$$

where $\hat{C}_n^k(f_b)$ and $\hat{C}_x^k(f_b)$ stand for the diagonal element of $\hat{\mathbf{C}}_n^k$ and $\hat{\mathbf{C}}_x^k$ related to frequency f_b , respectively. After some manipulations, Eq. 3.21 can be expressed as

$$p(X^k(i, f_b) | Y(i, f_b), \hat{\boldsymbol{\theta}}^k) = \frac{e^{-\frac{|X^k(i, f_b) - \mu_{x|y}^k(f_b)|^2}{C_{x|y}^k(f_b)}}}{\pi C_{x|y}^k(f_b)} = \mathcal{CN}(X^k(i, f_b); \mu_{x|y}^k(f_b), C_{x|y}^k(f_b)) \quad (3.22)$$

with

$$\begin{cases} C_{x|y}^k(f_b) = \left(\frac{\zeta^k(i)^2}{\hat{C}_n^k(f_b)} + \frac{1}{\hat{C}_x^k(f_b)} \right)^{-1} \\ \mu_{x|y}^k(f_b) = \frac{C_{x|y}^k(f_b)}{\hat{C}_n^k(f_b)} \zeta^k(i) Y(i, f_b) \end{cases} \quad (3.23)$$

Therefore the expectation of the k^{th} signal of interest $X^k(i, f_b)$ is estimated as

$$\mathbf{E}\{X^k(i, f_b) | Y(i, f_b), \hat{\boldsymbol{\theta}}^k\} = \mu_{x|y}^k(f_b) = \frac{\zeta^k(i)}{\zeta^k(i)^2 + \frac{\hat{C}_n^k(f_b)}{\hat{C}_x^k(f_b)}} Y(i, f_b). \quad (3.24)$$

Finally, the time signal $\hat{x}^k[n]$ is obtained from Eq. 3.24 by using the inverse STFT.

Two remarks are noteworthy. First, it is seen that Eq. 3.24 corresponds to a time-varying filter from which superior performance is expected than from a conventional time-invariant filter. Second, the standard Wiener filter appears as a particular case under the assumption of stationarity, that is

$$\mathbf{E}\{X^k(i, f_b) | Y(i, f_b)\} = \frac{1}{1 + \frac{C_n^k(f_b)}{C_x^k(f_b)}} Y(i, f_b), \quad (3.25)$$

where the latent variable $\zeta(i) = 1$ for all time instants. In other words, Eq. 3.25 then corresponds to the case where ‘‘State 1: $\mathbf{Y}(i) = \mathbf{X}(i) + \mathbf{N}(i)$ ’’ only occurs.

The performance of the proposed reconstruction filter will be demonstrated and further compared with the fast kurtogram in section 3.4.

3.3 Parameter selection

This section discusses the initialization and setting of the HMM parameters introduced in the previous section. Thereby an effective data-driven initialization for the EM algorithm is proposed. Meanwhile several synthetic signals are generated and analyzed for demonstration. The performance of different covariance structures is also investigated in order to retain the best model in term of compromise between simplicity and efficiency. To demonstrate the potential of the proposed HMM, an extended case is eventually tested with two simultaneous components of interest.

Before proceeding further, let remind and summarize the proposed fault detection scheme. The detailed steps are presented as follows:

1. Transform the raw signal into the time-frequency domain using the phase-corrected STFT (see Eq. 4.3) which reflects the “instantaneous complex envelope” of the signal $x(t)$ in a narrow frequency band Δf centered on f_b and sampled at time instant i .
2. Represent the noisy measurement $\mathbf{Y}(i)$ as a linear combination of K components of interest contaminated by a noisy component $\mathbf{N}(i)$ with different covariance matrices (see Eq. 3.4 and 3.5) by introducing a predefined latent variable vector $\boldsymbol{\zeta}(i)$ as a switch to indicate the presence of transients or not.
3. Initialize the parameters of the EM algorithm (see Eqs. 3.29-3.32) and estimate the required latent variables $\zeta(i)$, the *LLR* and others parameters (\mathbf{C}_n , \mathbf{C}_x^1 , $\boldsymbol{\pi}$) as seen in Eqs. 3.12-3.14.
4. The estimated $\boldsymbol{\zeta}(i)$ indicates the presence of underlying fault occurrences and the Fourier transform of the *LLR* is calculated to find the bearing characteristic frequencies.

3.3.1 Initial settings

The basic idea of the proposed scheme is to model the signal of interest by means of a latent random variable which reflects the presence or not of a transient due to an impact on the fault. This is achieved in the domain of the STFT. Therefore, the first required parameters to tune are the window length N_w and the window shift R .

Selecting the window length N_w

The selected value of N_w directly controls the frequency resolution,

$$\Delta f = F_s/N_w, \quad (3.26)$$

which characterizes the carrier frequency. It is required to cover at least the duration T_I of a transient, which implies the condition

$$\Delta f < 1/T_I. \tag{3.27}$$

As the STFT is subjected to the uncertainty principle, $\Delta t \Delta f \geq 1$, the highest switching frequency of the latent variable, $\alpha = 1/\Delta t$, is bounded upward by Δf [74]. Therefore the available range of the latent variable is limited by

$$\alpha \leq \Delta f. \tag{3.28}$$

Therefore, N_w should be taken short to allow a high switching rate in Eq. 3.28, but long enough to satisfy Eq. 3.27, i.e. $F_s \cdot T_I < N_w \leq F_s \cdot \Delta t$ as illustrated in Fig. 3.3.

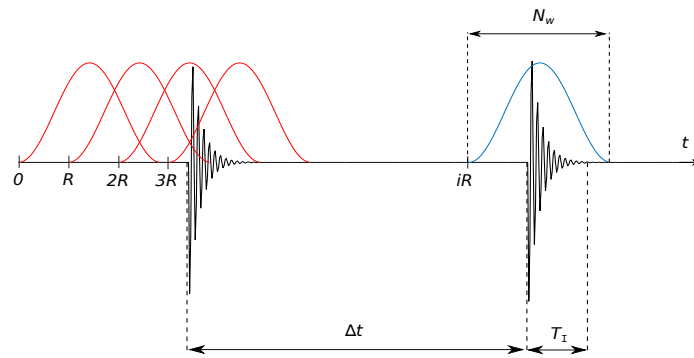


Fig. 3.3 Illustration of how to select the window length N_w and shift R with respect to transient durations T_I and cycle Δt .

Other reasons for taking N_w small is to reduce the computation time required by the STFT and also to ensure sufficient segments for accurate parameter estimation.

It is noticed here that the rule for setting N_w also works in the special case where the interval Δt between adjacent transients is close to the impulse duration T_I . In other words, it is robust enough to balance the trade-off between a fine resolution and a high switching frequency of the latent variable. These facts will be experimentally verified in subsection 3.3.2.

Selecting the window shift R

There are two considerations for the window shift R :

- firstly, for reconstructing the fault signal, the phase-corrected STFT is required to be invertible, therefore it is recommended to take 75% overlap with a Hanning window,
- secondly, if the first item is not necessary, R should be taken sufficiently small to keep more information while not increasing too much computational cost and dependence between segments, such as 50% or 75% overlap with a Hanning window.

Initializing parameters for the EM algorithm

The EM algorithm generally requires a good initialization step. A simple solution is given hereafter to obtain initial values for the covariance matrices of the two states in the HMM.

The initial diagonal of the covariance matrix of noise, $\hat{\mathbf{C}}_n^{[0]}$, is obtained by taking the median value of the natural logarithm of the squared magnitude of measurement $\mathbf{Y}(i)$ with respect to time instant i ,

$$\hat{\mathbf{C}}_n^{[0]} = \exp(\text{median}\{\log|\mathbf{Y}(i)|^2\}). \quad (3.29)$$

This approximation is based on the fact that State 0 is characterized by a high probability π so that the median in the above equation is almost unaffected by the occurrence of the fault. Besides, the initialization of the covariance matrix in State 1 makes use of extreme values that exceeds a threshold. Specifically, a sample set $I^{(1)}$ for State 1 is obtained from collecting all indices of the STFT such that the quantity

$$\text{Indicator}(i) = \frac{1}{N_f} \sum_{f_b=1}^{N_f} \log |Y(i, f_b)|^2 - \frac{1}{N_f} \sum_{f_b=1}^{N_f} \log \hat{\mathbf{C}}_n(f_b)^{[0]} \quad (3.30)$$

is found greater than a given threshold. Therefore the initial diagonal of the covariance matrix in State 1 is calculated as

$$\hat{\mathbf{C}}_{x+n}^{[0]} = \frac{1}{N_1} \sum_{i \in I^{(1)}} |\mathbf{Y}(i)|^2 \quad (3.31)$$

where N_1 is the cardinal of set $I^{(1)}$, from which $\hat{\mathbf{C}}_x^{[0]}$ is obtained as $(\hat{\mathbf{C}}_{x+n}^{[0]} - \hat{\mathbf{C}}_x^{[0]})_+$. The corresponding probability is initialized to

$$\hat{\pi}^{[0]} = \frac{N - N_1}{N}. \quad (3.32)$$

It has been observed in numerous experiments that the proposed initializations are often quite close to the maximum likelihood estimates.

3.3.2 Cases 1 & 2: demonstration of parameter selection

To demonstrate the performance of the proposed algorithm and initialization, a synthetic signal is generated with a resonance frequency $f_0 = 0.15$ Hz, which is further modulated by a relatively high fault frequency $\alpha_0 = 1.25 \times 10^{-3}$ Hz ($T = 1/\alpha_0 = 800$ s, the sampling frequency is normalized to $F_s = 1$ Hz). More precisely, the produced signal is described as:

$$y(t) = \sum_{j=-\infty}^{+\infty} h(t - jT - \tau_j)A_j + n(t) \quad (3.33)$$

$$H(z) = \frac{b_1 + b_2 \cdot z^{-1}}{a_1 + a_2 \cdot z^{-1} + a_3 \cdot z^{-2}} \quad (3.34)$$

where $\tau_j \sim \mathcal{N}(\mu_\tau = 0, \sigma_\tau = 0.02T)$ and $A_j \sim \mathcal{N}(\mu_A = 0, \sigma_A = 0.1)$ account for the uncertainties on the arrival time and on the magnitude of the j^{th} transient, respectively. The white noise $n(t)$ is set to a noise-to-signal-ratio of 6 dB and the signal length is $L = 10^5$ samples. A second-order system is defined by Eq. 3.34, whose numerator and denominator coefficients are $\mathbf{b} = [-1, 1]$ and $\mathbf{a} = [1, -2\cos(2\pi f_0)r, r^2]$ with $r = 0.95$, respectively. Fig. 3.4 shows the spectrogram (magnitude of the STFT) of the raw signal whose record in time is displayed in Fig. 3.5 (a).

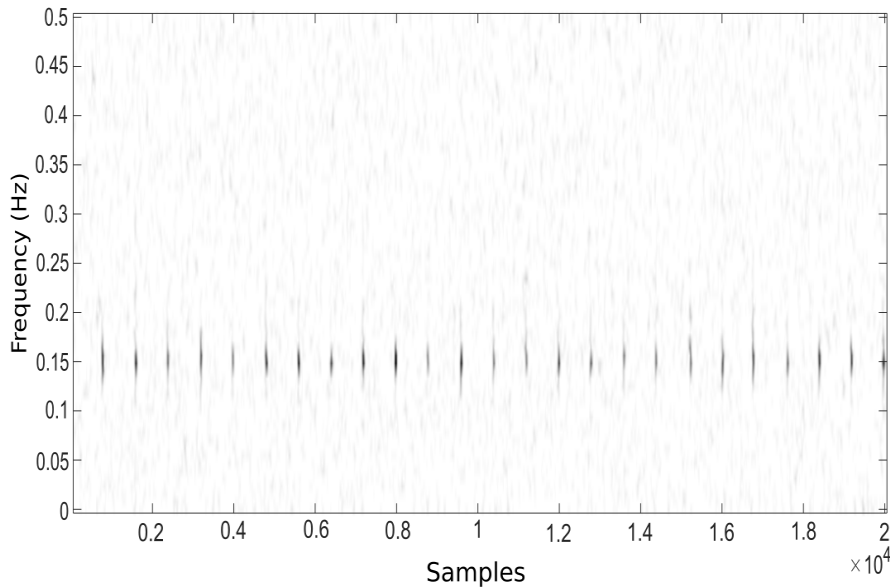


Fig. 3.4 Spectrogram of the signal simulated in Case 1 with resonance frequency $f_0 = 0.15$ Hz, $r = 0.95$ and fault frequency $\alpha_0 = 1.25 \times 10^{-3}$ Hz ($T = 1/\alpha_0 = 800$ s, $N_w = 2^7$ and $R = 20$).

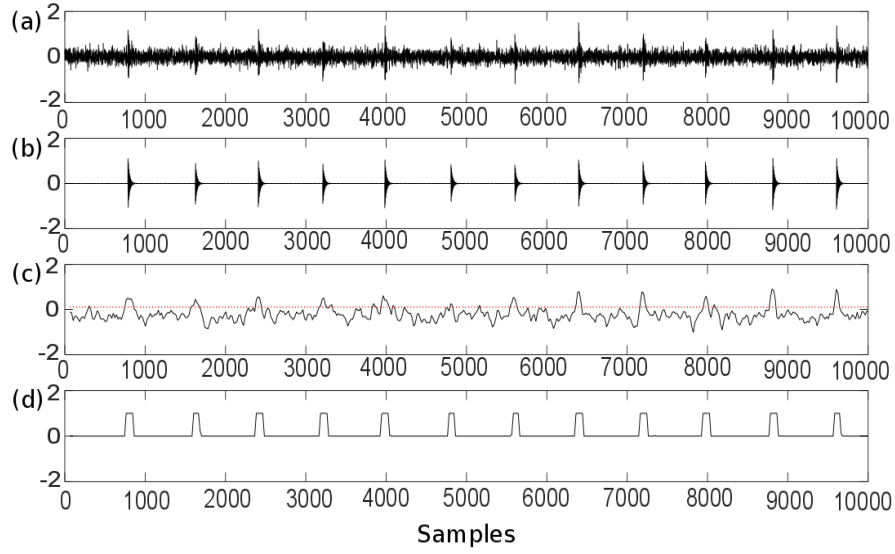


Fig. 3.5 (a) Synthetic signal of Case 1 with white noise (noise-to-signal-ratio= 6 dB. (b) Synthetic repetitive transients. (c) Initialized latent variable $\hat{\zeta}(i)^{[0]}$ (black line) with threshold (red dotted line) set to $0.01 \max(\hat{\zeta}(i)^{[0]})$ with $\hat{\pi}^{[0]} = 0.112$. (d) Estimated latent variable $\hat{\zeta}(i)$.

For simplicity, a diagonal covariance matrix is assumed. Following Eqs. 3.29-3.32, one can initialize the parameters $\hat{\mathbf{C}}_n^{[0]}$, $\hat{\mathbf{C}}_x^{[0]}$, $\hat{\zeta}(i)^{[0]}$ and $\hat{\pi}^{[0]}$ as shown in Fig. 3.6 (a) and Fig. 3.5 (c). It is seen that the proposed initialization is simple and effective, even though the estimated spectrum of the signal of interest still contains a significant contribution from noise especially below 0.08 Hz. After convergence of the EM algorithm, the estimation of the signal and noise spectra are close to the real values as can be seen in Fig. 3.6 (b). The very good estimation of the latent variable is verified in Fig. 3.5 (d).

In order to demonstrate the performance when dealing with a coarse frequency resolution, a second synthetic signal is generated with a resonance frequency $f_0 = 6 \times 10^{-3}$ Hz and a fault frequency $\alpha_0 = 1.25 \times 10^{-3}$ Hz. All the other parameters are as in Case 1. Figure 3.8 shows the spectrogram of the raw signal.

Since the modulation frequency is now close to the resonance one, this case encounters a trade-off between a fine spectral content and a large fault frequency range. A coarse resolution is chosen as $N_w = 2^7$ ($\Delta f = 7.8 \times 10^{-3}$ Hz), although it cannot resolve the resonance frequency $f_0 = 6 \times 10^{-3}$ Hz, embodies at least 6 integer multiples of $\alpha_0 = 1.25 \times 10^{-3}$ Hz. The corresponding limit on the detection of the fault frequency is $\alpha_{max} \leq \Delta f = 7.8 \times 10^{-3}$ Hz.

Figure 3.7 (a) displays the *LLR*, which accurately localizes the fault occurrences: the function sharply goes to negative infinity when it identifies an impulse. Besides, the estimated

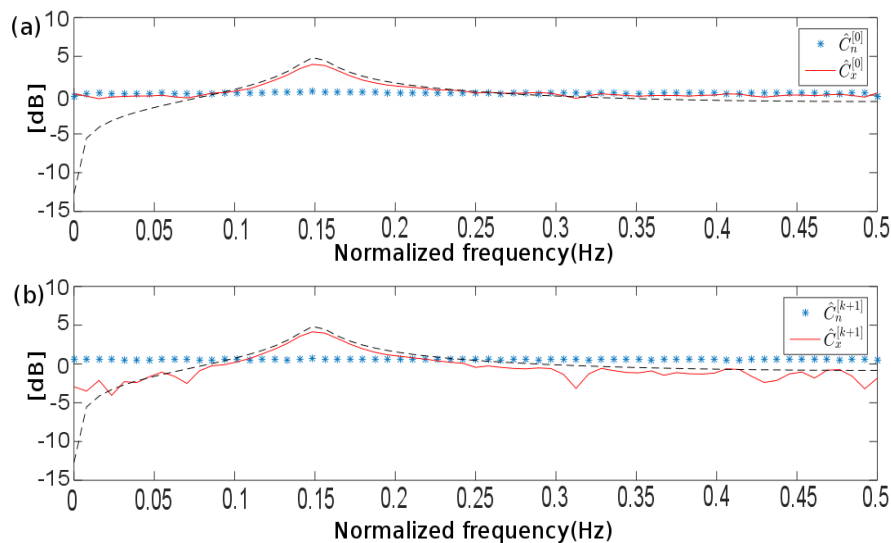


Fig. 3.6 (a) Initialized spectra (diagonals of covariance matrices $\hat{C}_n^{[0]}$ and $\hat{C}_x^{[0]}$ (red line and blue asteriks) and (b) estimated ones from EM (diagonals of covariance matrices $\hat{C}_n^{[k+1]}$ and $\hat{C}_x^{[k+1]}$ (red lines) together with the theoretical frequency response $H(z)$ (black dashed line and blue asteriks).

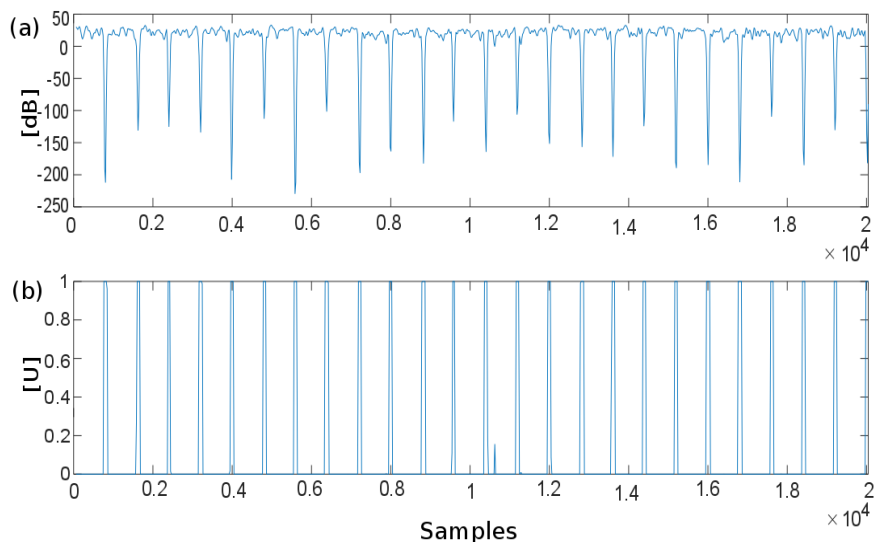


Fig. 3.7 (a) *LLR* and (b) latent variable $\hat{\zeta}(i)$ in the time domain.

latent variable $\hat{\zeta}(i)$ locates exactly all the STFT segments that contain a fault occurrence, as shown in Fig. 3.7 (b). To further identify the fault type, the spectra of the *LLR* and of the

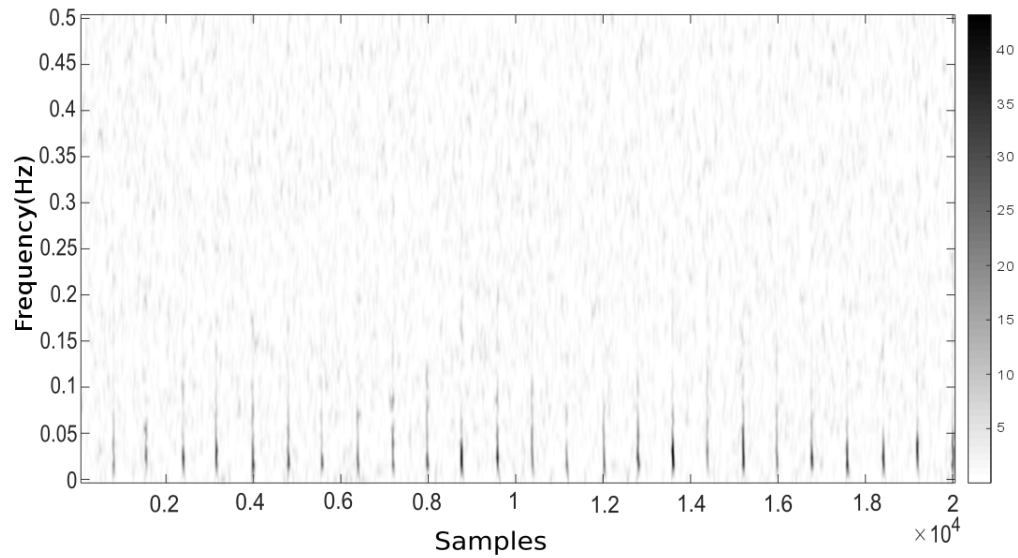


Fig. 3.8 Spectrogram of the signal simulated in Case 2 with resonance frequency $f_0 = 6 \times 10^{-3}$ Hz, $r = 0.9$ and fault frequency $\alpha_0 = 1.25 \times 10^{-3}$ Hz.

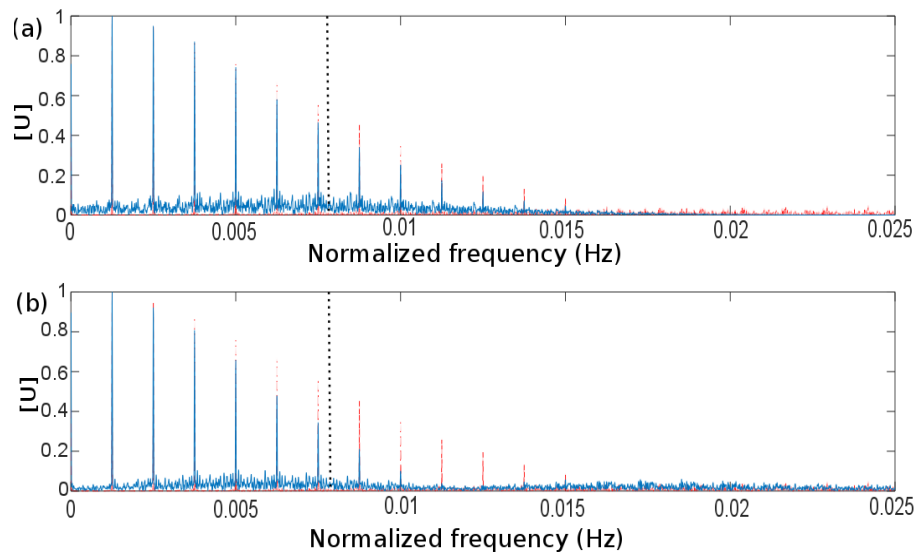


Fig. 3.9 Spectra of the *LLR* (a) and of the latent variable (b) presented in Fig. 3.7 together with the SES of actual fault signal (red dotted line). The limit $\alpha_{max} = 7.8 \times 10^{-3}$ Hz is indicated by a vertical black dotted line (Normalization to unit maximum value).

latent variable together with the SES of the actual fault signal are displayed in Fig. 3.9 (note that all spectra have been normalized to unit maximum), with the limit α_{max} indicated by a vertical black dotted line. It is noteworthy that the spectrum of the *LLR* (Fig. 3.9 (a)) shows the largest number of harmonics due to the fact that the *LLR* has sharper peaks at the position of the impulses. This superiority will be further verified in the following section.

Despite the coarse frequency resolution used in this case, it has been demonstrated that the proposed scheme can still detect the expected fault frequency with very good accuracy.

3.3.3 Case 3: Comparison of different spectral correlation assumptions

This subsection now investigates the different possible structures of the covariance matrix previously listed in subsection 3.1.3, i.e. a full matrix, a diagonal matrix, a tridiagonal matrix, a low rank matrix, a Toeplitz matrix.

A synthetic signal with resonance frequency $f_0 = 0.2$ Hz is employed as shown in Fig. 3.10. All parameters are set as in Case 1.

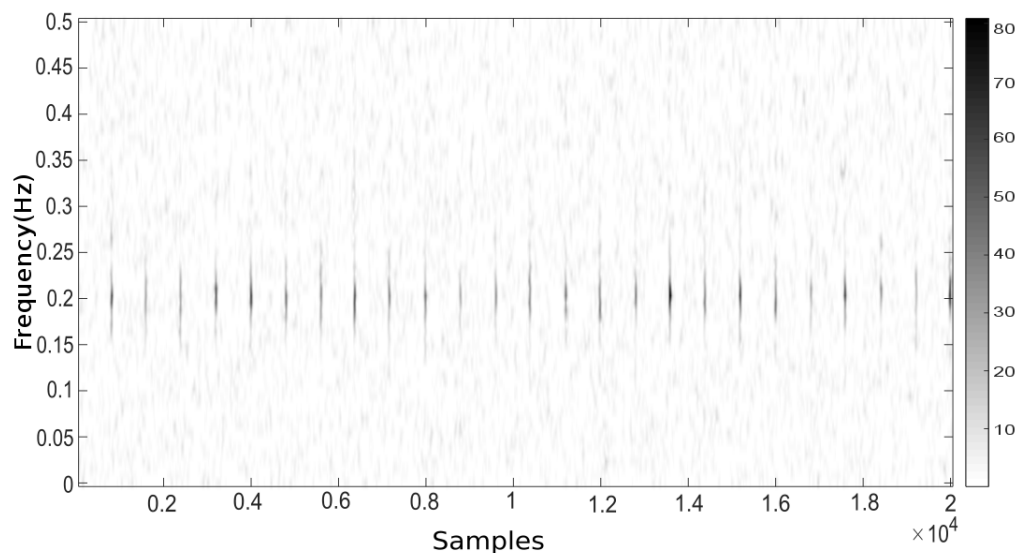


Fig. 3.10 Spectrogram of the signal simulated in Case 3 with resonance frequency $f_0 = 0.2$ Hz, $r = 0.9$ and fault frequency $\alpha_0 = 1.25 \times 10^{-3}$ Hz.

The spectra of the *LLR* and of the latent variable $\hat{\zeta}(i)$ for the five different structures of covariance matrices are displayed in Figs. 3.11-3.12, respectively, together with the

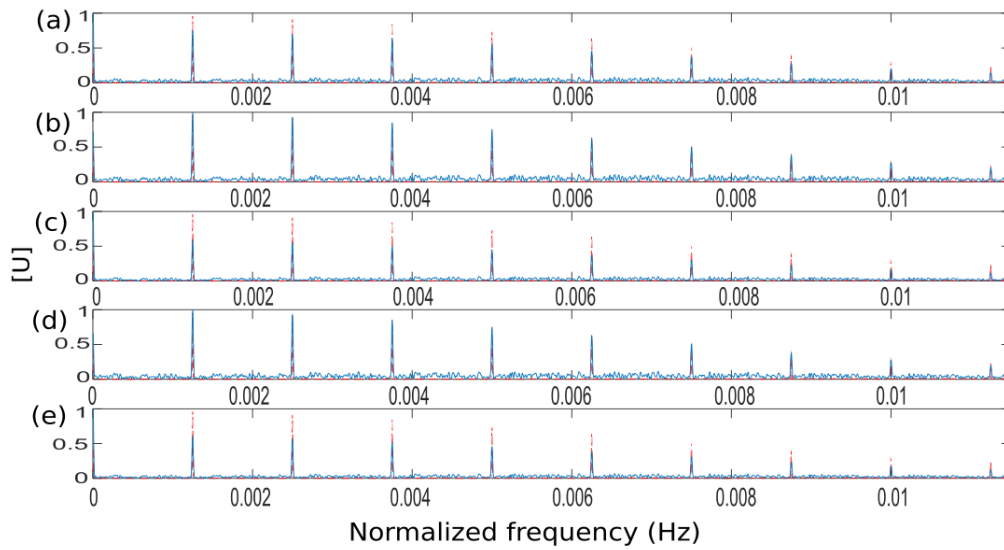


Fig. 3.11 Spectra of the LLR together with the theoretical SES (red dotted line) (a) full matrix, (b) diagonal matrix, (c) tridiagonal matrix, (d) low rank matrix, (e) Toeplitz matrix (normalization to unit maximum value).

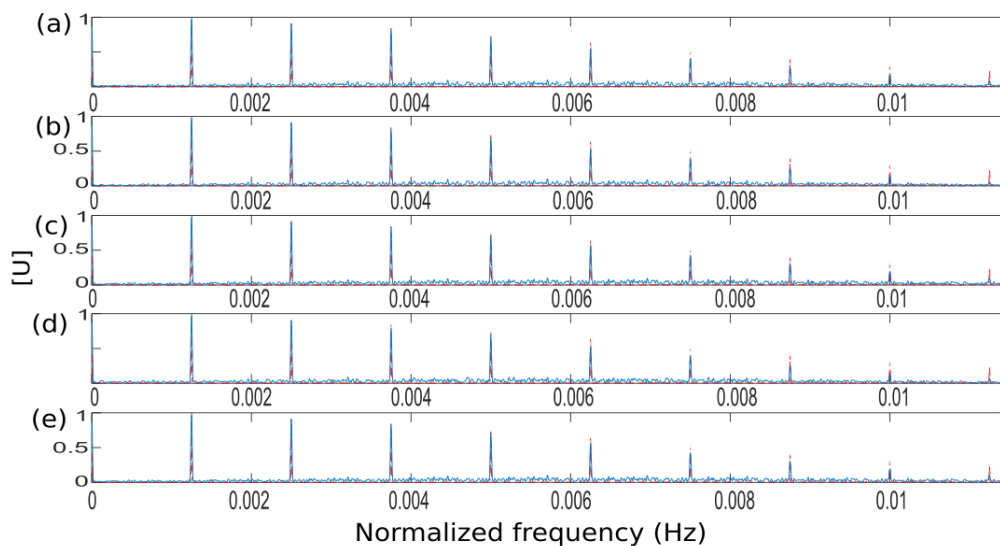


Fig. 3.12 Spectra of the latent variable $\hat{\zeta}(i)$ together with the theoretical SES (red dotted line) (a) full matrix, (b) diagonal matrix, (c) tridiagonal matrix, (d) low rank matrix, (e) Toeplitz matrix (normalization to unit maximum value).

theoretical SES. Meanwhile, the five different structures of covariance matrices are estimated and demonstrated separately in Fig. 3.13-3.17. As far as the latent variable is concerned, it is seen that there is no significant difference between the five structures. As for the *LLR*, it seems that the diagonal and low rank structures return slightly superior results to the others.

As a conclusion, it seems that the simplest structure based on a diagonal covariance matrix should be retained. This has a further advantage since it corresponds to accounting only for the spectrum of the transients when constructing the covariance matrix.

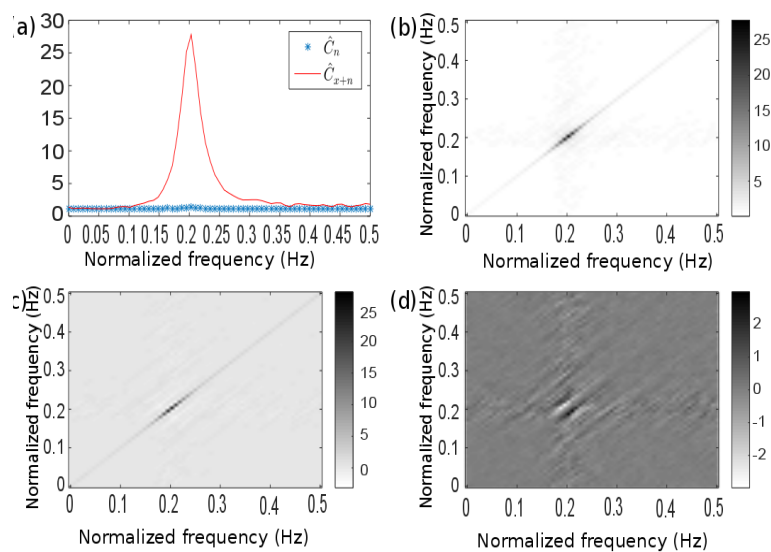


Fig. 3.13 Full matrix: (a) diagonals (the spectrum of the fault signal is indicated by a red solid line and that of the noise by blue asterisks); (b) absolute value; (c) real part; (d) imaginary part of the estimated covariance matrix.

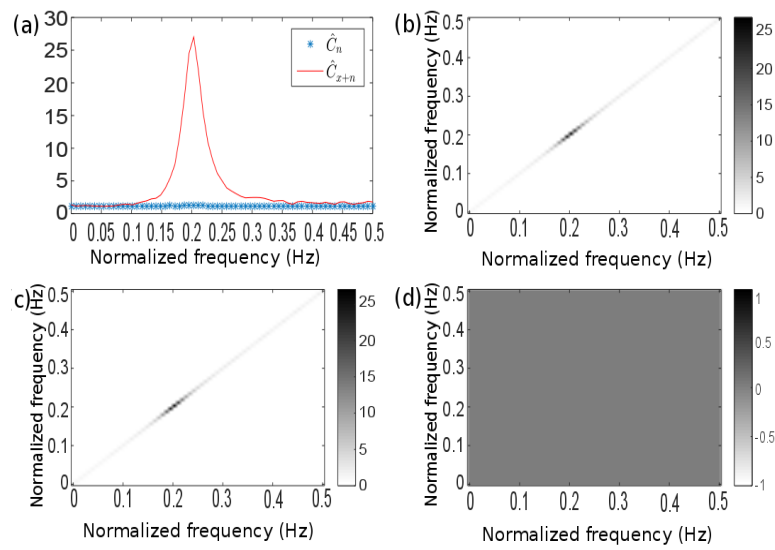


Fig. 3.14 Diagonal matrix: (a) diagonals (the spectrum of the fault signal is indicated by a red solid line and that of the noise by blue asterisks); (b) absolute value; (c) real part; (d) imaginary part of the estimated covariance matrix.

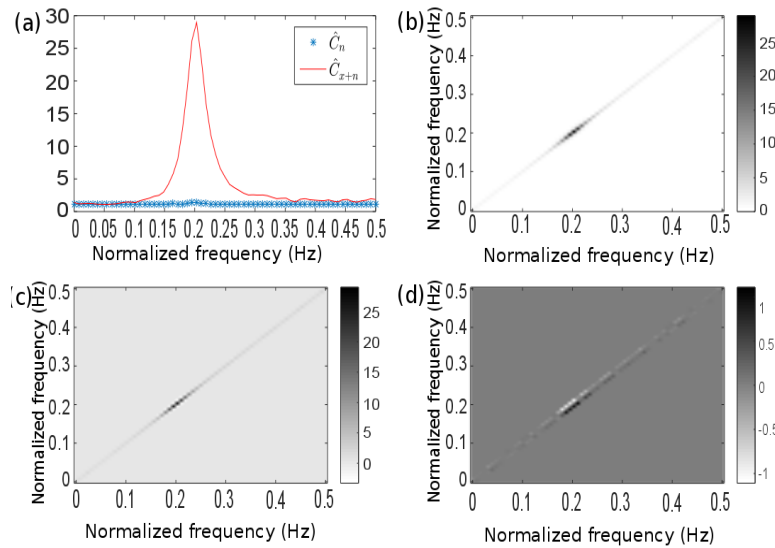


Fig. 3.15 Tridiagonal matrix: (a) diagonals (the spectrum of the fault signal is indicated by a red solid line and that of the noise by blue asterisks); (b) absolute value; (c) real part; (d) imaginary part of the estimated covariance matrix.

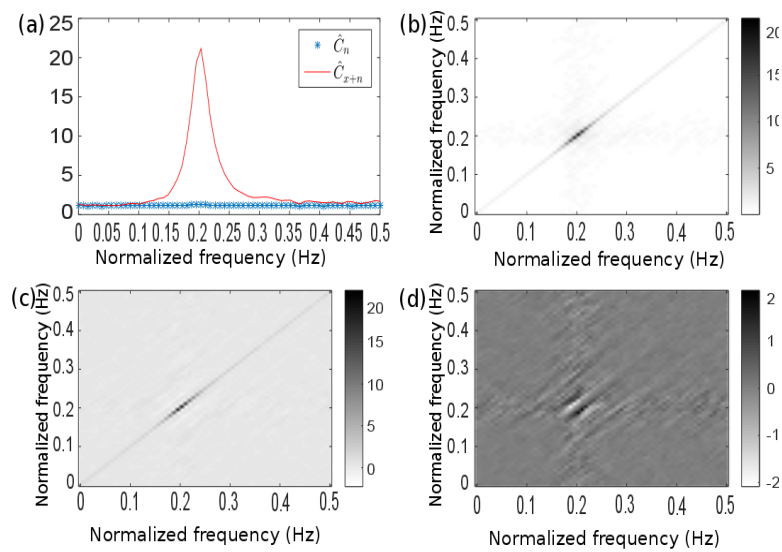


Fig. 3.16 Low rank matrix: (a) diagonals (the spectrum of the fault signal is indicated by a red solid line and that of the noise by blue asterisks); (b) absolute value; (c) real part; (d) imaginary part of the estimated covariance matrix.

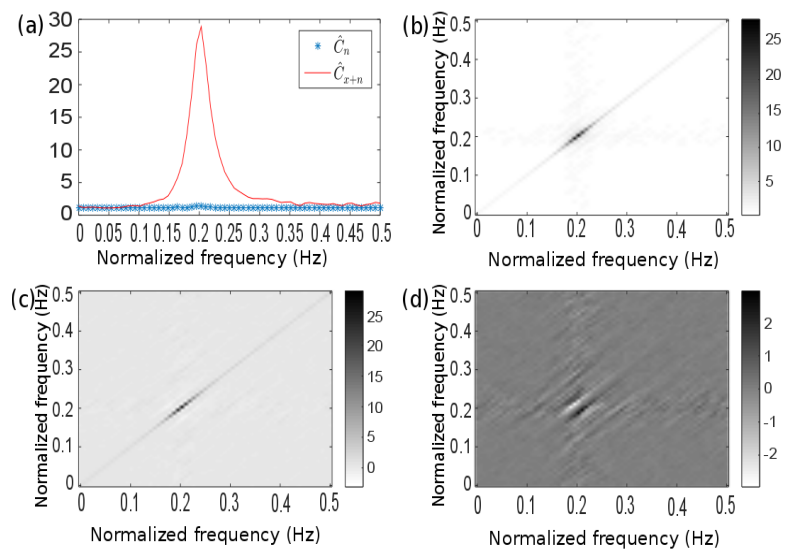


Fig. 3.17 Toeplitz matrix: (a) diagonals (the spectrum of the fault signal is indicated by a red solid line and that of the noise by blue asterisks); (b) absolute value; (c) real part; (d) imaginary part of the estimated covariance matrix.

3.3.4 Case 4: diagnostics and separation of a mixture of independent transients

This subsection intends to demonstrate the potential of proposed HMM to deal with $K = 2$ simultaneous components in background noise. A synthetic signal with two different components is generated as shown in Fig. 3.18. The first component $\mathbf{X}^1(i)$ has a resonance frequency $f_0^1 = 0.25$ Hz with $r = 0.9$ and its cyclic frequency is $\alpha_0^1 = 1.9 \times 10^{-3}$ Hz ($T_1 = 1/\alpha_0^1 = 530$ samples), whereas the second one $\mathbf{X}^2(i)$ has a resonance frequency $f_0^2 = 0.35$ Hz with $r = 0.7$ and cyclic frequency $\alpha_0^2 = 2.1 \times 10^{-3}$ Hz ($T_2 = 1/\alpha_0^2 = 470$ samples) – see Fig. 3.18 (a) and (b), respectively. All other parameters are set as in Case 1, with $\sigma_\tau^{1,2} = 0.02T_{1,2}$ and $\sigma_A^{1,2} = 0.1$. The noise-to-signal ratio is 0 dB.

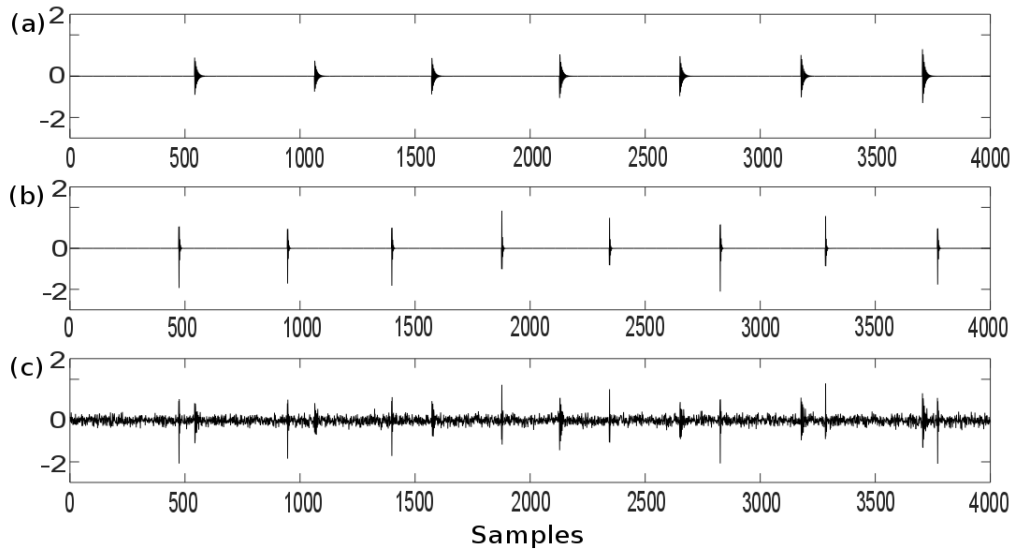


Fig. 3.18 (a) Component $\mathbf{X}^1(i)$ with period $T_1 = 530$ samples ($\sigma_\tau^1 = 0.02T_1$ and $\sigma_A^1 = 0.1$). (b) Component $\mathbf{X}^2(i)$ with period $T_2 = 470$ samples ($\sigma_\tau^2 = 0.02T_2$ and $\sigma_A^2 = 0.1$). (c) Noisy measurement (noise-to-signal ratio = 0 dB).

The spectrogram of the measurement is computed with $N_w = 2^6$ and $R = 23$ (overlapping ratio $R/N_w = 0.359$) is shown in Fig. 3.19.

The estimated diagonals of the three covariance matrices are displayed in Fig. 3.20. It is seen that the spectra of all components are correctly identified. The corresponding latent variables $\zeta^1(i)$ and $\zeta^2(i)$ are displayed in Fig. 3.21. Compared with the reference signals, it is obvious that all the times of occurrence of each component have been correctly located.

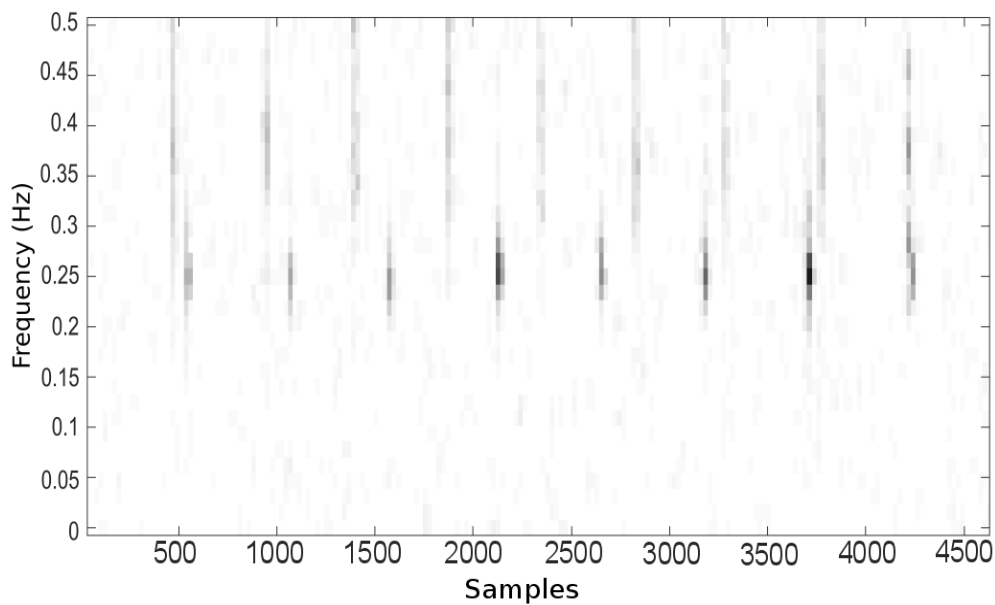


Fig. 3.19 Spectrogram of the signal simulated in Case 4 with two components ($T_1 = 530$ samples and $T_2 = 470$ samples respectively).

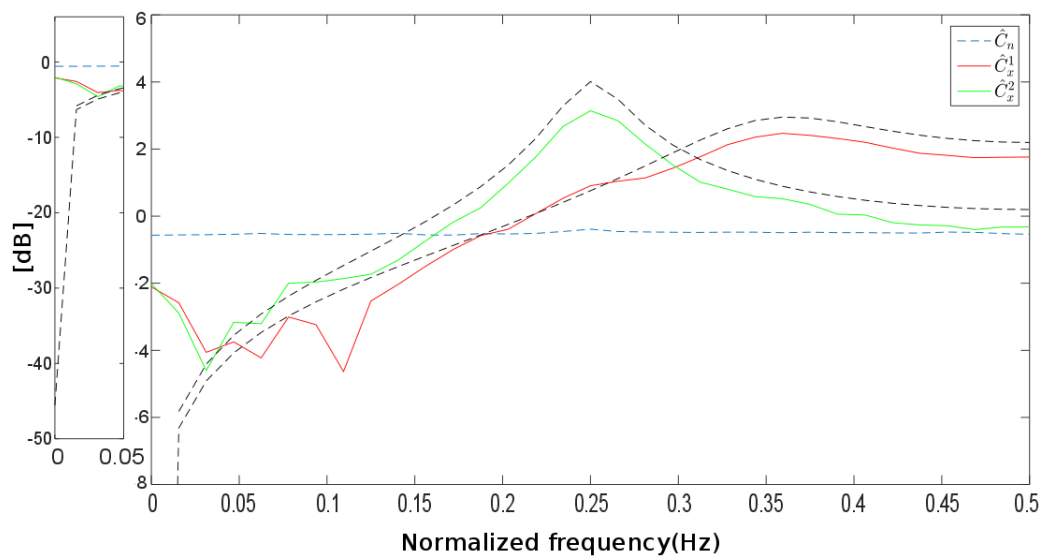


Fig. 3.20 Estimated spectra of component $\mathbf{X}^1(i)$ (red solid line), component $\mathbf{X}^2(i)$ (green solid line) and noise (blue dashed line). The two theoretical frequency responses of $H(z)$ are indicated by black dashed lines.

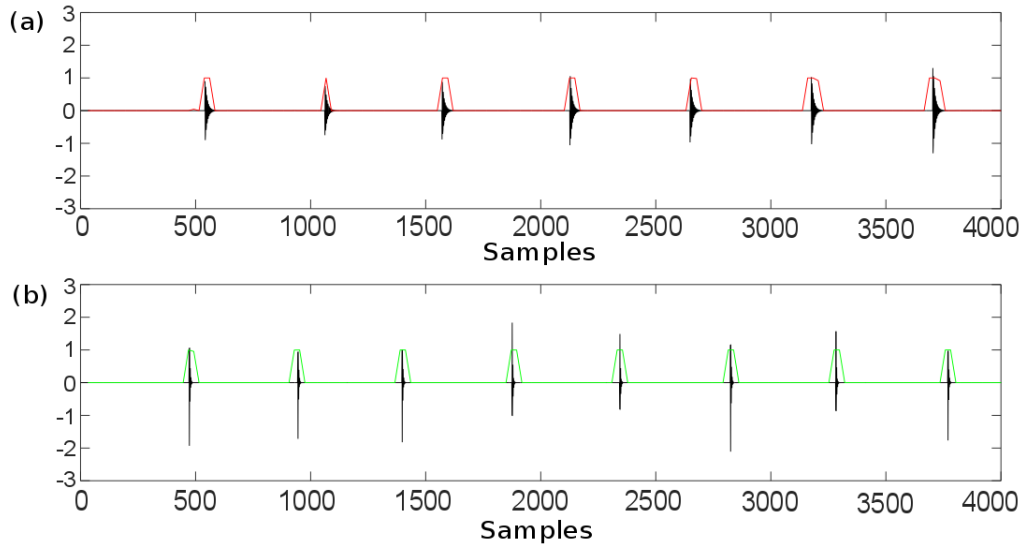


Fig. 3.21 (a) Estimated latent variable $\zeta^1(i)$ (red solid line) together with the actual first component (black solid line). (b) Estimated latent variable $\zeta^2(i)$ (green solid line) together with the actual second component (black solid line).

The spectra of the latent variables in Fig. 3.22 further reveal the fault characteristic frequencies at $\alpha_0^1 = 1.9 \times 10^{-3}$ Hz and $\alpha_0^2 = 2.1 \times 10^{-3}$ Hz. Finally, Fig. 3.23 displays the reconstructed repetitive transients $\mathbf{X}^1(i)$ and $\mathbf{X}^2(i)$ as well as their summation. Very good reconstruction is obtained, which demonstrates the performance of the proposed algorithm.

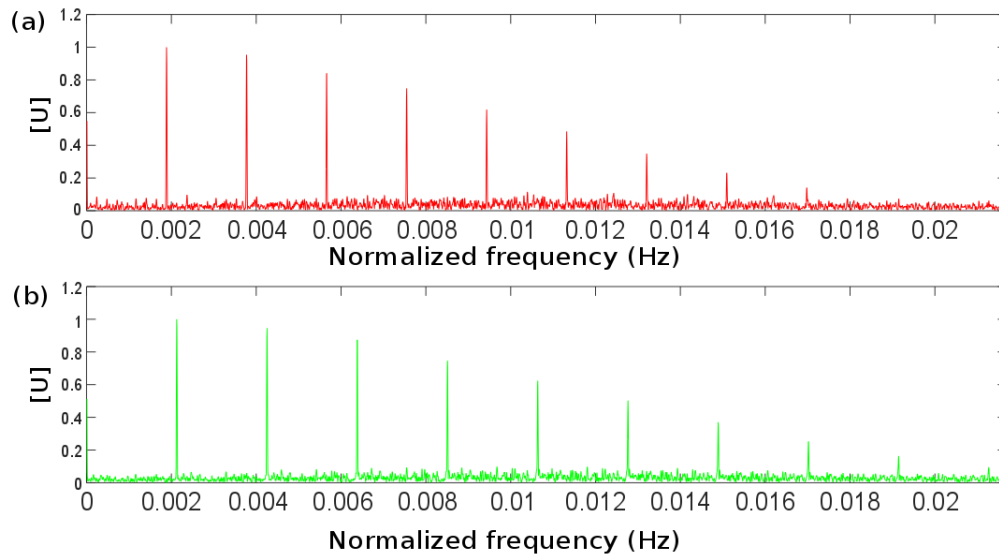


Fig. 3.22 Spectra of the estimated latent variable (a) $\zeta^1(i)$ and (b) $\zeta^2(i)$ (normalized to unit maximum value).

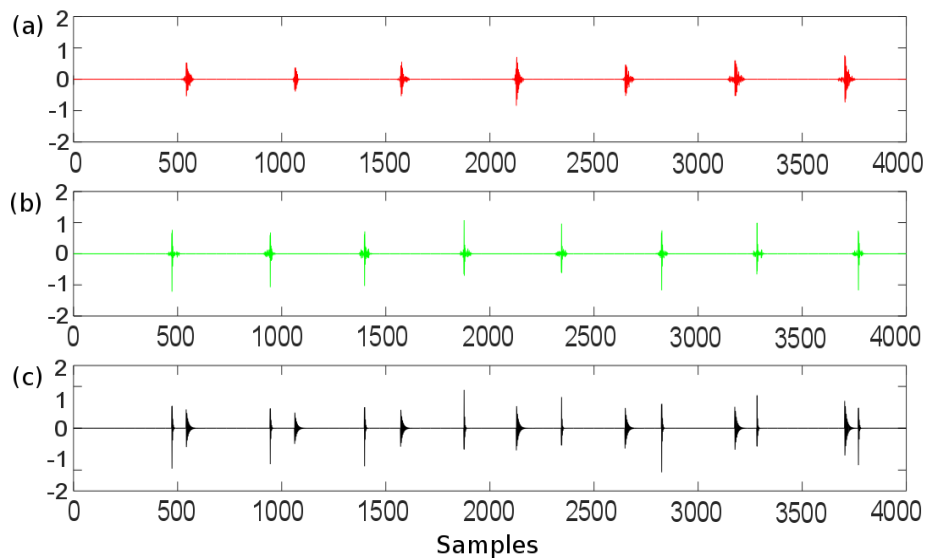


Fig. 3.23 Reconstructed repetitive transients corresponding to (a) the first component $\mathbf{X}^1(i)$ with period $T_1 = 530$ samples and (b) the second component $\mathbf{X}^2(i)$ with period $T_2 = 470$ samples. (c) Summation of two components $\mathbf{X}^1(i)$ and $\mathbf{X}^2(i)$.

3.4 Validation on vibration signals

This section illustrates the application of the proposed scheme on actual vibration signals for the detection of incipient bearing faults. Intended to be an automatic detection, its superior performance is first demonstrated as compared to classical envelope analysis. Furthermore, compared with the fast kurtogram, similar results are achieved by the spectrum of the *LLR* regarding the identification of the bearing fault frequency. Besides, the reconstruction of repetitive transients is achieved by a time-varying filter controlled by the latent variable $\zeta(i)$. Finally the feasibility of extending the proposed method to non-stationary regime is demonstrated on data captured during a run-up. It is important to notice that all the validation is based on the proposed simple model (see Eq. 3.5). The multiple-component model (see Eq. 3.4 and subsection 3.3.4) is also verified by real data.

3.4.1 Case 5: diagnosis of a ball fault

To demonstrate the effectiveness of the proposed scheme on real data, three typical types of fault (i.e. inner race, outer race and ball fault) are investigated in a dataset from the Vibrations and Acoustics Laboratory of the University of New South Wales (Sydney) [76]. The test-rig is a one-stage gearbox with primary and secondary shafts supported by ball bearings. Since it is often more difficult to identify a ball defect, particularly at incipient stage, this case is tested here. The parameter settings are listed in Table 3.1.

Table 3.1 Parameter settings in Case 5.

Sampling frequency F_s (Hz)	48000
Duration (s)	1.365
N_w	2^7
R	20
Rotation frequency – f_{rot} (Hz)	10
Ball spin frequency - BSF (Hz)	26.11
Fundamental train frequency - FTF (Hz)	4.08

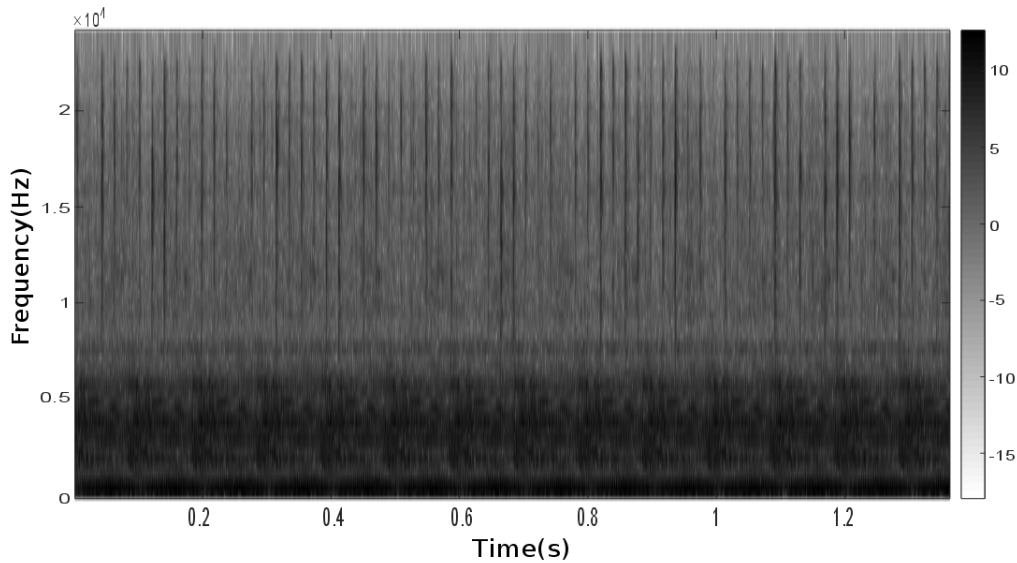


Fig. 3.24 Spectrogram (logarithmic scale) of signal of Case 5 (frequency resolution $\Delta f = 375$ Hz).

The spectrogram of the raw signal is displayed in Fig. 3.24. It is seen that there exists a non-stationary component in the high frequency band above 10 kHz, whereas the low frequency range is dominated by high energy components related to the gearbox vibrations.

Comparison with classical envelope analysis

Next, the SES of the raw signal is displayed in Fig. 3.25. It is noted that there exists a relatively high value at $f_g = 319.7$ Hz surrounded by shaft speed sidebands ($f_{rot} = 10$ Hz) which originates from the gearbox. Clearly, the information of the bearing fault is completely masked by high-energy components from the gearbox in the SES of the full-band signal. At this point, it is important to resort to methods that optimally and automatically select the frequency bands in the signal where the signal-to-noise ratio is maximum.

Figure 3.26 displays the diagonals of the covariance matrices $\hat{\mathbf{C}}_n$ and $\hat{\mathbf{C}}_x^1$. It is noteworthy that there is a crossing of the two spectra around 8 kHz which is consistent with the two frequency bands identified in Fig. 3.24. This reflects the fact that the fault signal and surrounding noise belong to different probability distributions. More precisely, the high energy vibrations of the gearbox dominates below the crossing frequency, whereas the repetitive transients dominate above the crossing frequency. The spectrum of the *LLR* is displayed in Fig. 3.27. It is noted that for the ball fault case there are harmonics of BSF

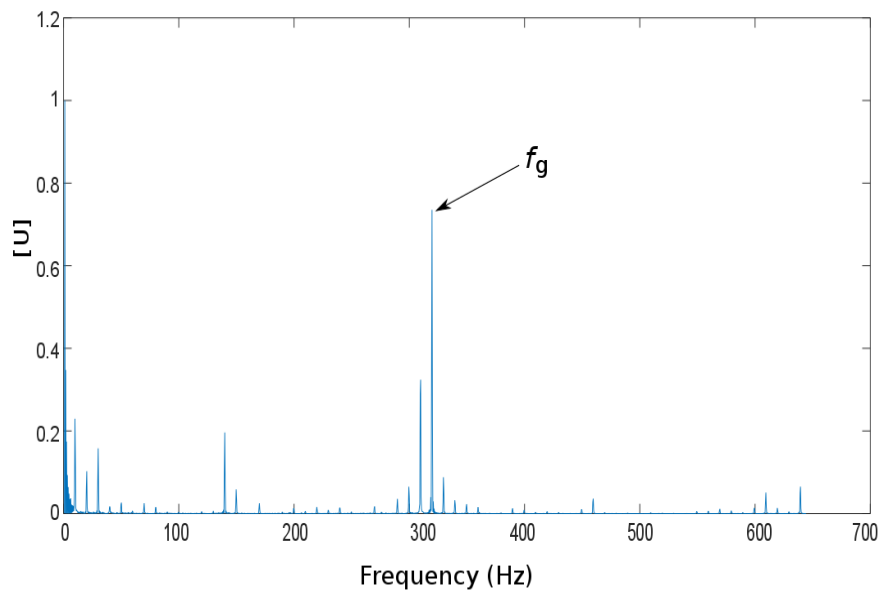


Fig. 3.25 SES of the raw signal (normalized to unit maximum value).

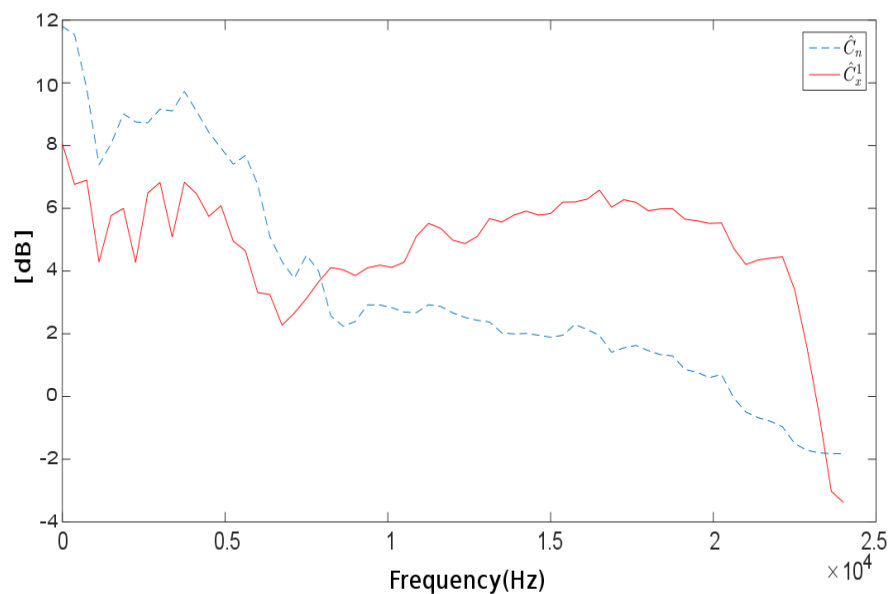


Fig. 3.26 Diagonals of covariance matrices (frequency resolution $\Delta f = 375$ Hz); the spectrum of the fault signal is indicated by a red solid line and that of the noise by a blue dashed line.

(with dominant even harmonics of BSF) surrounded by modulation sidebands at the cage speed (FTF). The superior performance as compared to full-band SES is also verified, which

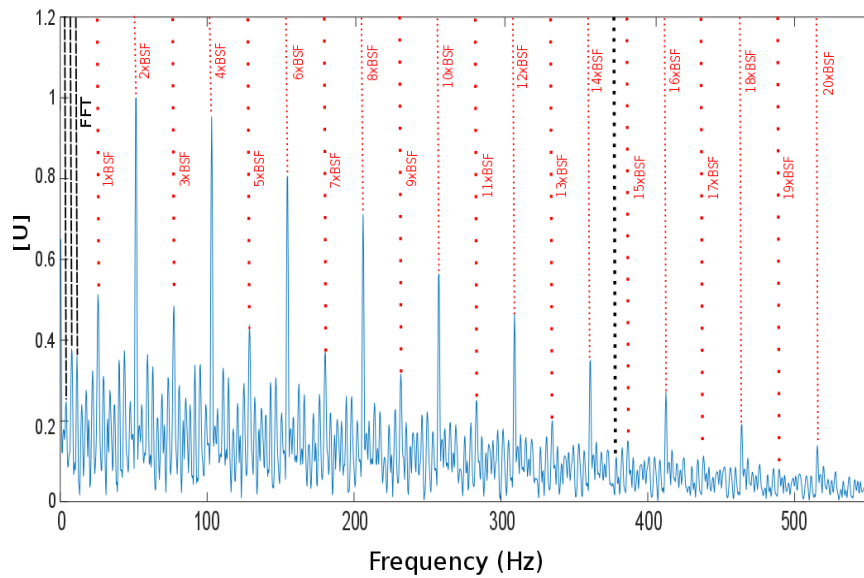


Fig. 3.27 Spectrum of the *LLR* with markers at the theoretical fault frequency and harmonics; the limit $\alpha_{max} = 375$ Hz is indicated by a vertical black dotted line (normalization to unit maximum value).

confirms the capability of the method to detect the fault without manual pre-processing such as band-pass filtering.

Comparison with the fast kurtogram

As stated in Ref. [76], the fast kurtogram has proved a powerful fourth-order spectral analysis tool for detecting and characterizing impulses in a signal. The fast kurtogram is applied to the signal of Case 5 with $K = 7$ decomposition levels in a 1/3-binary tree. As seen in Fig. 3.28, there exists several local maxima in the kurtogram. They are coherent with the spectrogram of Fig. 3.24 which evidences a clear non-stationary activity above 10 kHz. All the dyads with very high kurtosis values have been checked to have similar complex envelopes. Therefore one relevant maximum is taken at 71.83 whose corresponding dyad is $\{f_{10}; (\Delta f)_4\} = \{14250; 1500\}$ Hz compared to the frequency band [13500; 15000] Hz. The SES in that band is displayed in Fig. 3.29. It clearly reveals the even BSF surrounded by modulation sidebands at cage speed (FTF). Comparing Fig. 3.29 with Fig. 3.27, there exists no major difference between the two results, except for the better enhancement of the odd harmonics of BSF in Fig. 3.27. Therefore, the conclusion in that case is that the proposed method achieved a detection as good as the fast kurtogram followed by standard envelope analysis. One advantage, however, is that it returns the full spectral content of the fault signal

and possibly allows its complete reconstruction in time (and not only in a pass-band). This is further investigated in the next subsection.

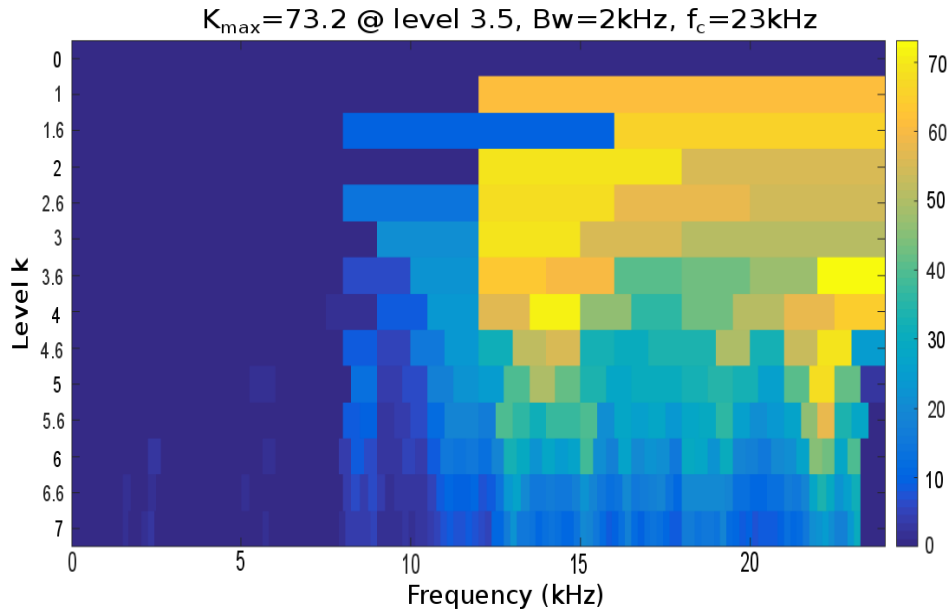


Fig. 3.28 Kurtogram of signal of Case 5 computed over $K = 7$ levels with a 1/3-binary tree and an 8 coefficient prototype filter. Several local maxima are presented. One relevant maximum is found at dyad $\{f_{10}; (\Delta f)_4\} = \{14250; 1500\}$ Hz.

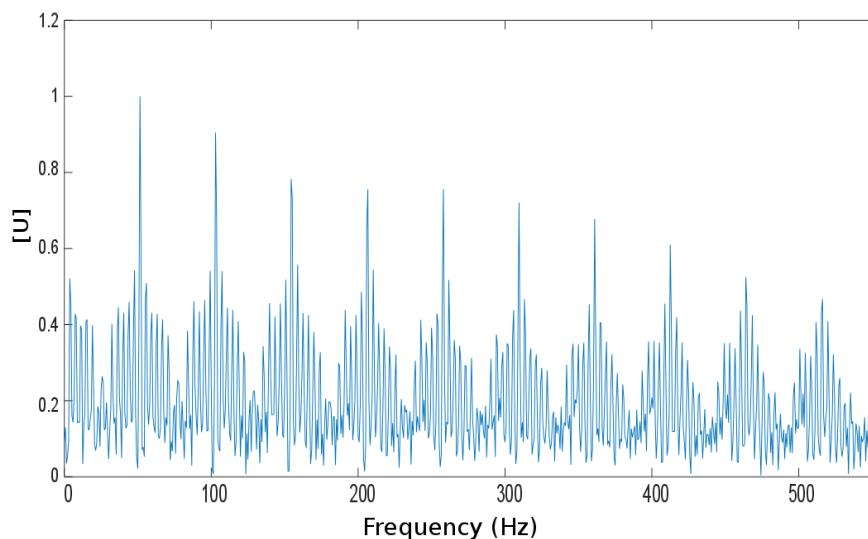


Fig. 3.29 SES in frequency band $[13500; 15000]$ Hz returned by the kurtogram (normalization to unit maximum value).

Full-band reconstruction of the fault signal

The capability of the proposed HMM method to reconstruct the full-band fault signal is now demonstrated and compared with the band-pass results obtained from the kurtogram.

Figure 3.30 (a) displays 1.36 s of the analyzed vibration signal of Case 5 and Fig. 3.31 (a) and an enlarged view in the vicinity of a transient. The band-pass filtered signal in band [13500; 15000] Hz obtained from the kurtogram dyad $\{f_{10}; (\Delta f)_4\} = \{14250; 1500\}$ Hz is displayed in Fig. 3.30 (b) and its enlarged view in Fig. 3.31 (b); it clearly evidences the presence of transients with maximum signal-to-noise ratio. The reconstructed signal from the proposed HMM-based time-varying filter is displayed in Fig. 3.30 (c) and its enlarged view in Fig. 3.31 (c). As compared to the filtered signal based on the kurtogram, the reconstructed signal achieves an exact location of the transients with their full-band spectral content and is therefore closer to the actual fault signal. This may be used advantageously to better characterize the fault signature, infer the fault dimension and spectral content, and possibly update trend models for prognostics.

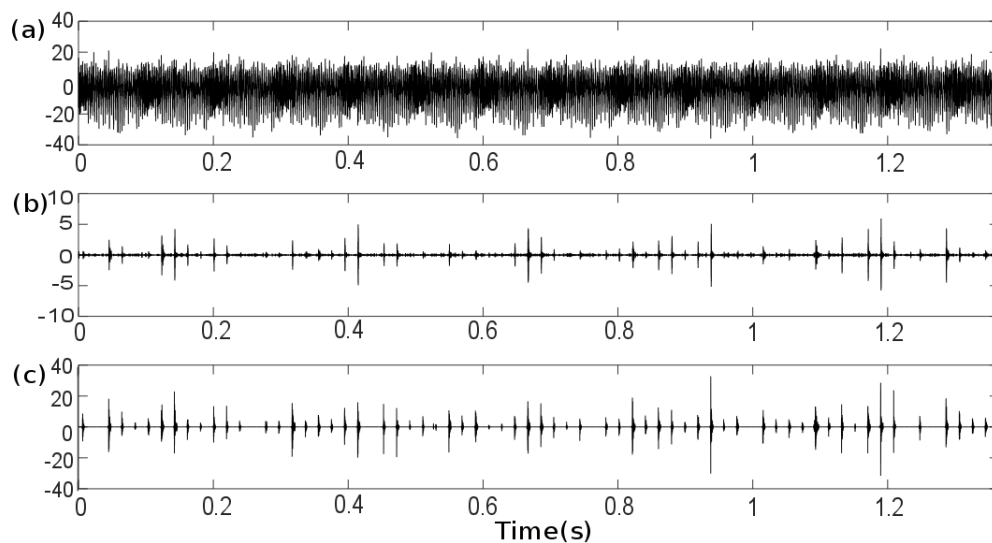


Fig. 3.30 (a) Vibration signal of Case 5 and (b) its band-pass filtered version in the frequency band [13500; 15000] Hz. (c) Full-band reconstructed fault signal from the proposed HMM-based time-varying filter.

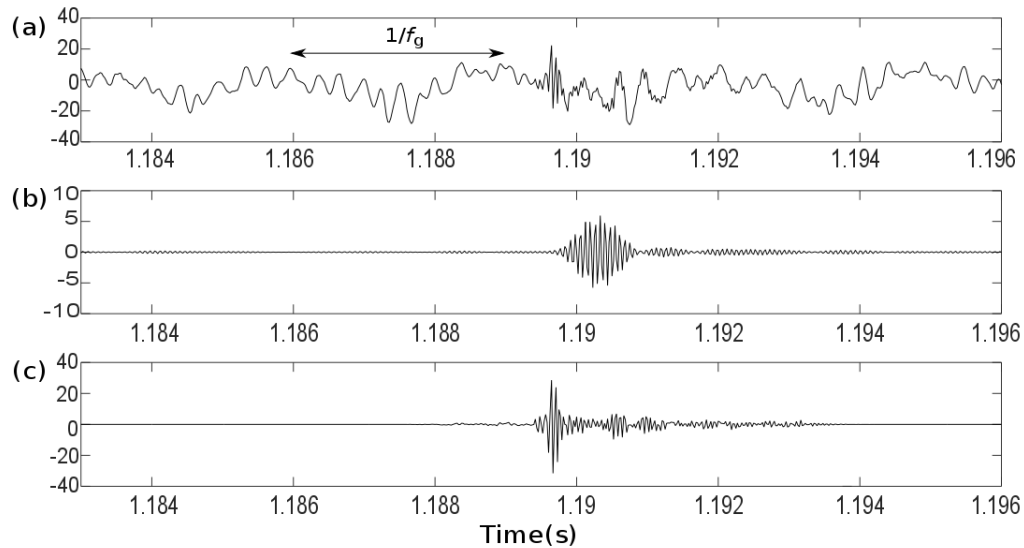


Fig. 3.31 (a) Enlarged view of vibration signal of Case 5 (with indication of the period corresponding to the peak f_g in Fig. 3.25) and (b) enlarged view of its band-pass filtered version in the frequency band [13500; 15000] Hz. (c) Enlarged view of the full-band reconstructed fault signal from the proposed HMM-based time-varying filter.

3.4.2 Cases 6 & 7: diagnostics of bearing and gears

Signal recorded on a test rig are now considered. The test rig shown in Fig. 5.2 has been designed by company DYNAE. It mainly consists of an electric asynchronous motor, a rotary encoder, 4 accelerometer sensors, a speed variator, a driving gear with 45 teeth, four bearings (3 healthy and 1 outer race fault) and two pinions (healthy and broken).

The test rig has been used to produce two types of fault: Case 6 relates to an outer race fault of the bearing and Case 7 to the combination of an outer race fault and a broken pinion connected to the driving gear. The parameter settings are listed in Table 3.2.

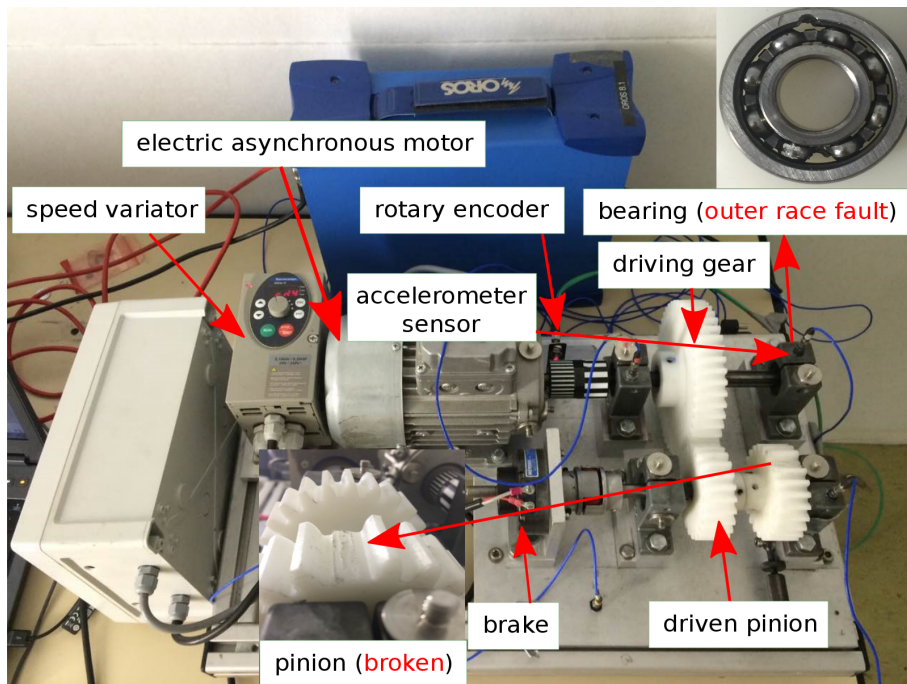


Fig. 3.32 Test rig setup.

Table 3.2 Parameter settings in Case 6 and Case 7.

	Case 6	Case 7
Sampling frequency F_s (Hz)	51200	
Duration (s)	10	
N_w	2^7	
R	45	
Main shaft rotation frequency – $f_{rot,1}$ (Hz)	15.3-16.4	22.2-24.5
Secondary shaft rotation frequency – $f_{rot,2}$ (Hz)	$1.875 \times f_{rot,1}$ (28.7-30.8)	$1.875 \times f_{rot,1}$ (41.6-45.9)

a) *Analysis of Case 6*

Figure 3.33 shows the spectrum of the *LLR* which is to be compared with the results of the Fast kurtogram displayed in Figs. 3.34-3.35. The maximum of the Fast kurtogram takes value 46 in dyad $\{f_{13}; (\Delta f)_4\} = \{20k; 1600\}$ Hz corresponding to frequency band [18400; 21600] Hz. The SES in this band is displayed in Fig. 3.35.

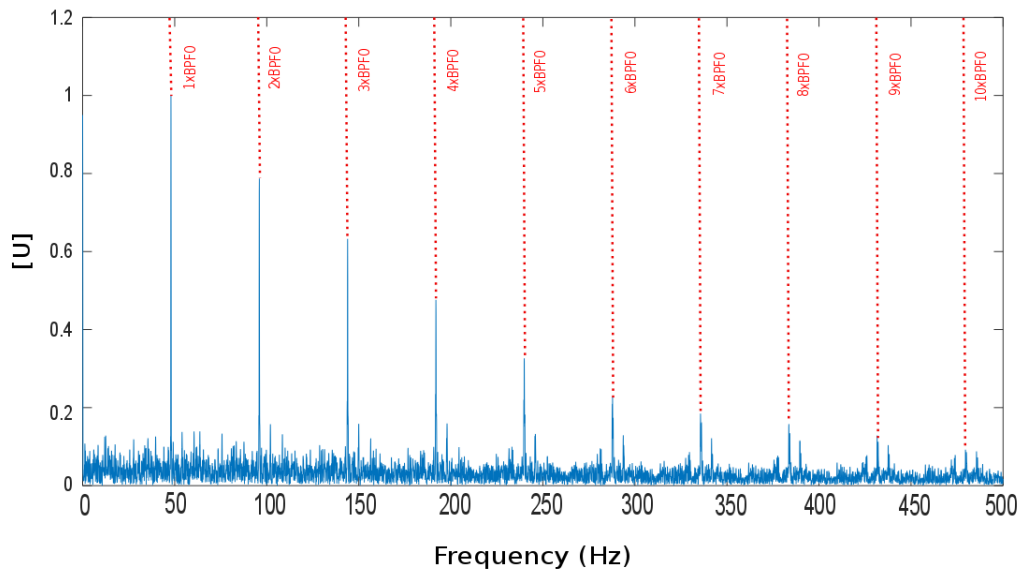


Fig. 3.33 Spectrum of the *LLR* with markers at the suspected fault frequency BPFO and its harmonics (normalization to unit maximum value).

Further comparison are now carried on in the time domain. Figure 3.36 a) displays the original signal which clearly shows the presence of transients out of background noise. Figure 3.36 b) displays the signal filtered in band [18400; 21600] Hz selected from the kurtogram, which maximizes the signal-to-noise ratio. Finally, Fig. 3.36 c) displays the signal reconstructed from the HMM-based time-varying filter with full-band content. The latter seems slightly richer than the former.

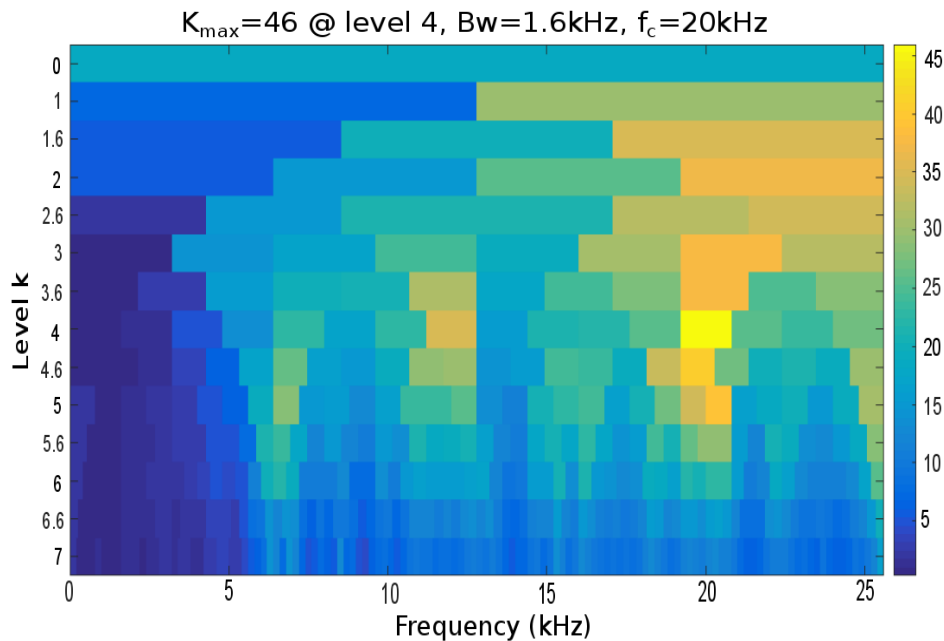


Fig. 3.34 Fast kurtogram of signal of Case 6 computed over $K = 7$ levels with a 1/3-binary tree and an 8 coefficient prototype filter. One relevant maximum is found at dyad $\{f_{13}; (\Delta f)_4\} = \{20k; 1600\}$ Hz.

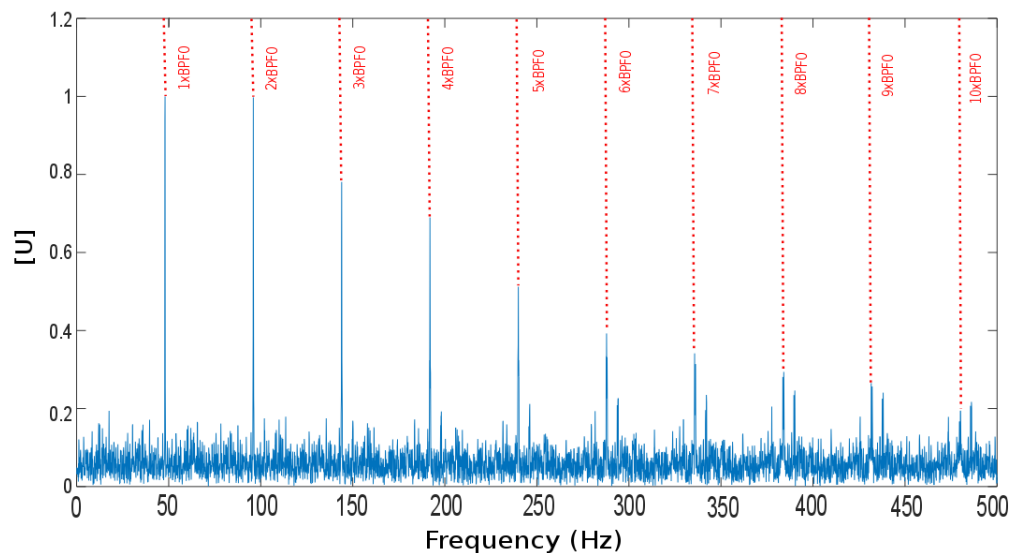


Fig. 3.35 SES of complex envelope in dyad $\{f_{13}; (\Delta f)_4\} = \{20k; 1600\}$ Hz corresponding to the frequency band $[18400; 21600]$ Hz with markers at the suspected fault frequency BPFO and its harmonics (normalization to unit maximum value).

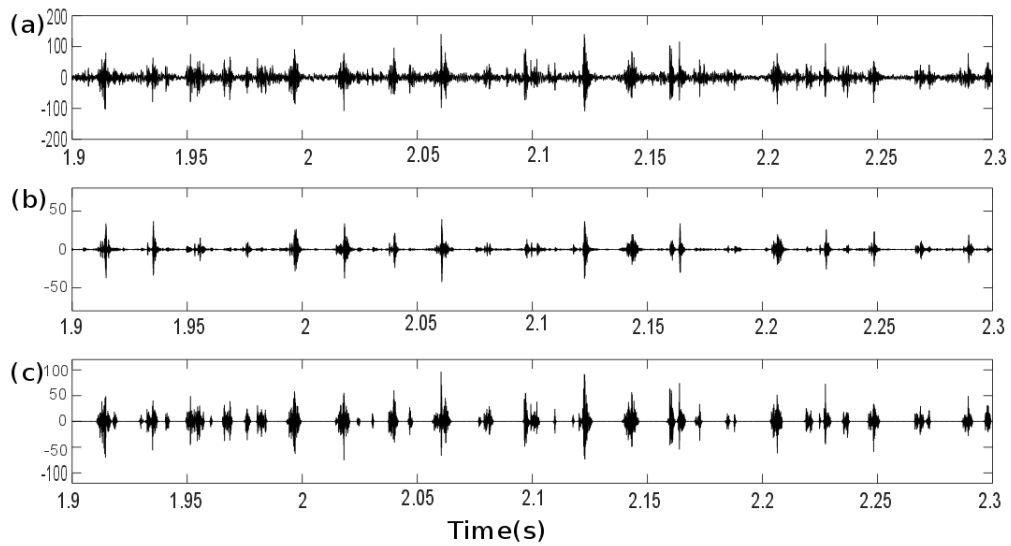


Fig. 3.36 (a) Vibration signal of Case 6 in time interval [1.9 2.3] s and (b) its band-pass filtered version in frequency band [18400; 21600] Hz. (c) Reconstructed transients from the HMM-based time-varying filter.

b) Analysis of Case 7

Using the same test rig and the same parameter settings, a compound source of vibration that contains gear and bearing fault together is diagnosed in Case 7. Figure 3.37 shows the spectrum of the *LLR* which reveals the harmonics of BPF0 and $f_{(rot,2)}$, thus demonstrating the presence of the outer race defect and of the secondary gear defect.

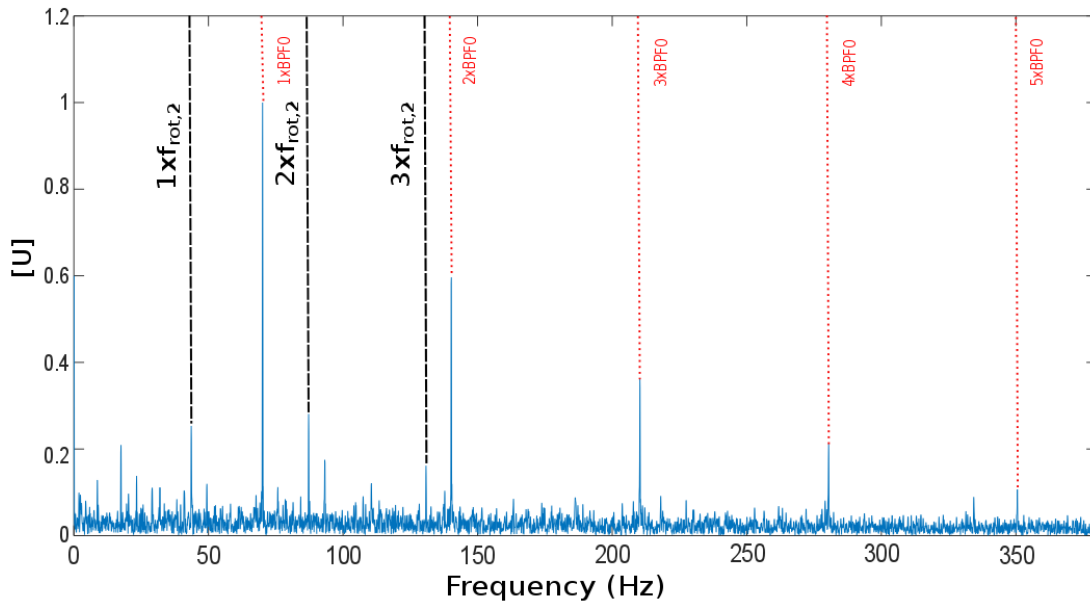


Fig. 3.37 Spectrum of the *LLR* with markers at the suspected fault frequencies BPF0, its harmonics and $f_{(rot,2)}$ (normalization to unit maximum value).

3.4.3 Case 8: diagnostics of bearing in nonstationary regime

This subsection intends to illustrate the potential of the proposed model in variable operating conditions. A runup of 15 s from 2 Hz to 25 Hz has been manually produced with the test rig of Fig. 5.2. The instantaneous speed is displayed in Fig. 3.38 (a) and the corresponding acceleration signal in Fig. 3.38 (b). The latter undergoes speed-dependent variations in magnitude and phase. Since there is no constraint on the distribution of the time instants of the impacts, the proposed model can be freely extended to deal with machine signals recorded in time varying regimes. However, there may exist speed-dependent variations of the probability distribution of the states which should be taken into account. Hence the whole signal is divided into consecutive speed segments of 12 rotations as suggested in Ref.

[77] – see Fig. 3.38 (c). Then, the parameters of the HMM are estimated separately on each segment. The probability of State 1 in the five regimes is estimated to $\hat{\pi} = 0.28, 0.30, 0.30, 0.33$ and 0.37 .

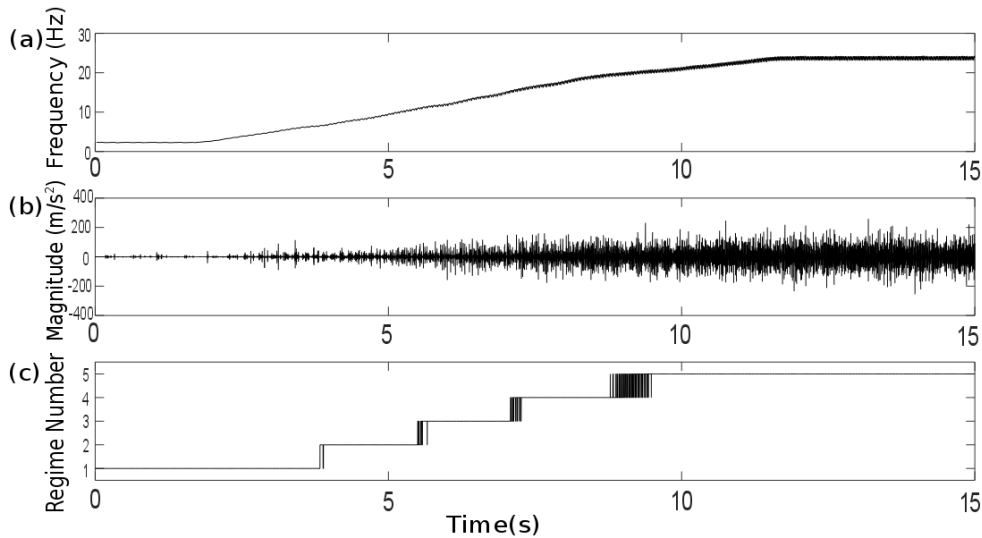


Fig. 3.38 (a) Estimated instantaneous speed of signal in Case 8 and (b) its corresponding acceleration signal which undergoes speed-dependent magnitude modulation. (c) Division of the estimated instantaneous speed in 5 regimes.

Figs. 3.39-3.43 display the raw signal on each segment together with the corresponding reconstructed transients. It is seen that the transients are all well identified, although their behavior slightly changes with rotational speed: the 4th segment corresponding to a speed around 20Hz exhibits the cleanest signature of the fault.

The detection of the fault characteristic frequency requires some specific processing under a nonstationary regime. Since the defect impacts occur periodically with respect to the angular position, its frequency is to be computed in the order domain. Therefore, the phase-corrected STFT in Eq. 4.3 has been resampled (using cubic splines interpolation) from the time to the angular domain, while maintaining a constant spectral bandwidth. The order spectrum of the *LLR* has then been computed on the resampled data. As seen in Fig. 3.44, it clearly reveals the presence of the Ball Pass Order on the Outer race (BPOO).

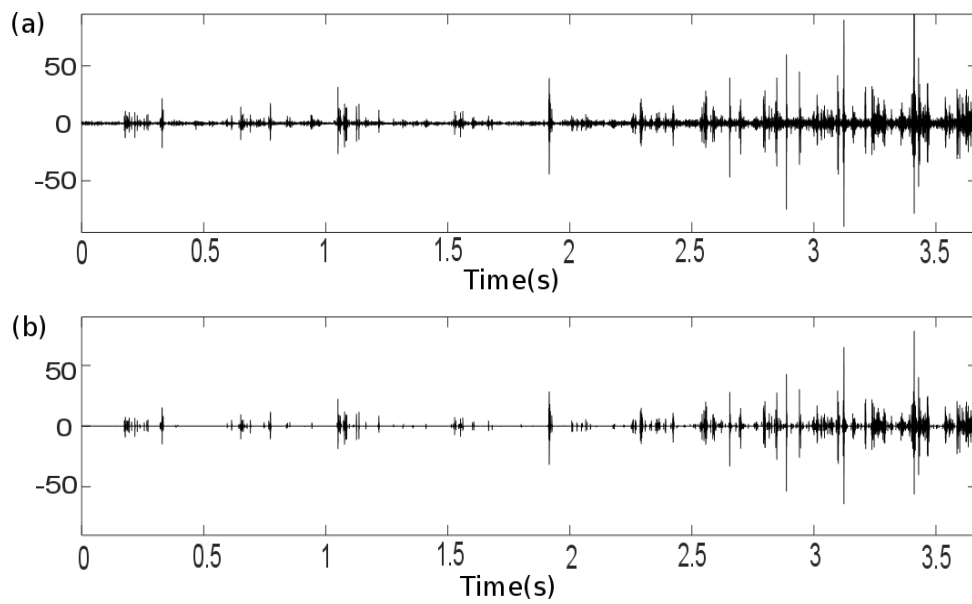


Fig. 3.39 (a) Raw signal in regime No. 1 and (b) the corresponding reconstructed transients from the HMM-based time-varying filter.

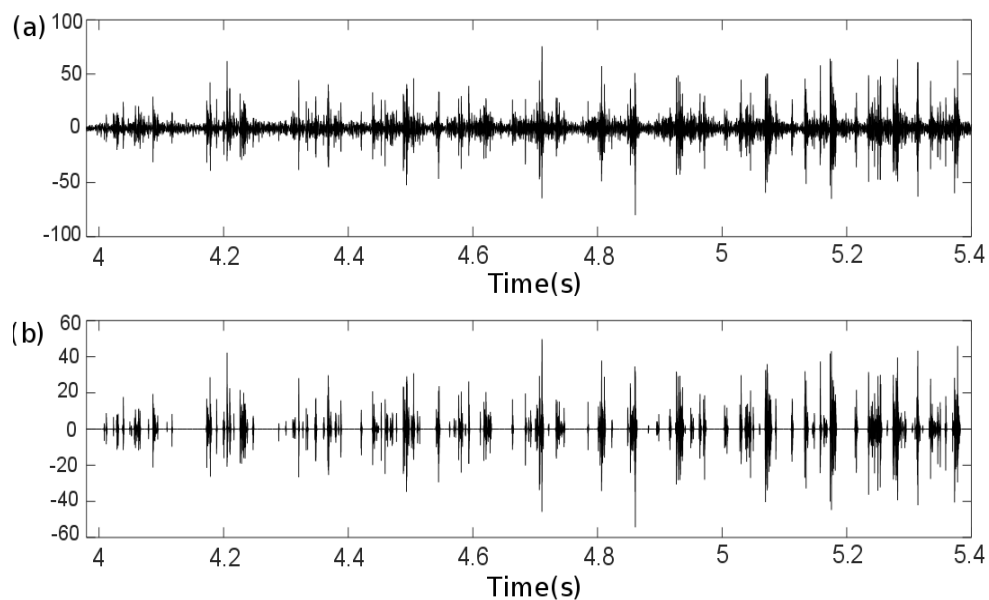


Fig. 3.40 (a) Raw signal in regime No. 2 and (b) the corresponding reconstructed transients from the HMM-based time-varying filter.

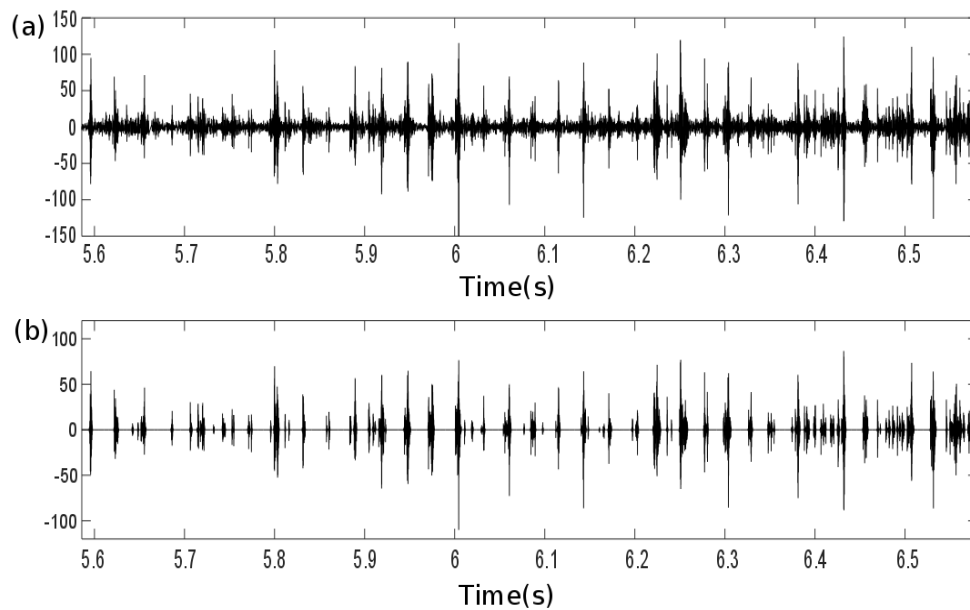


Fig. 3.41 (a) Raw signal in regime No. 3 and (b) the corresponding reconstructed transients from the HMM-based time-varying filter.

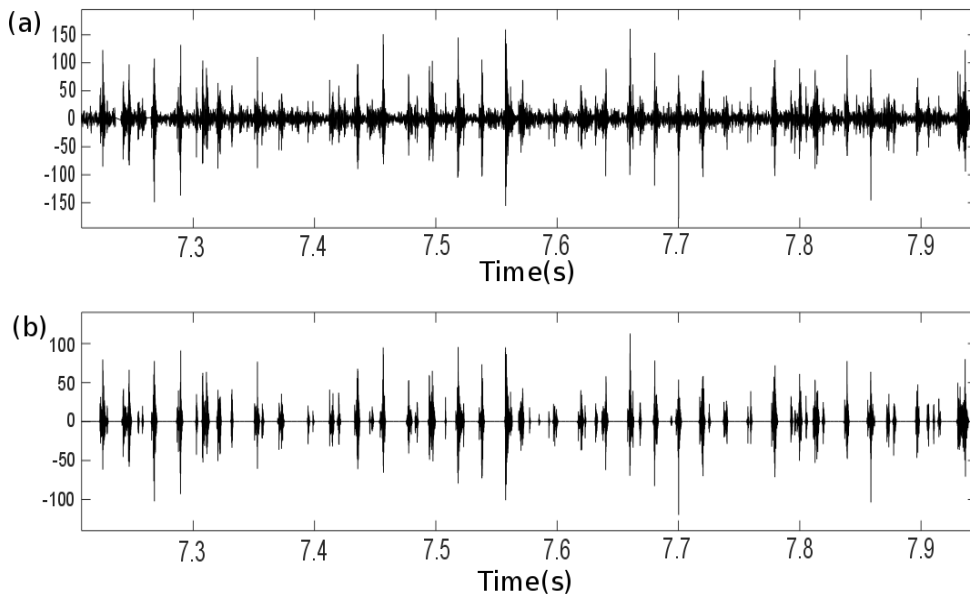


Fig. 3.42 (a) Raw signal in regime No. 4 and (b) the corresponding reconstructed transients from the HMM-based time-varying filter.

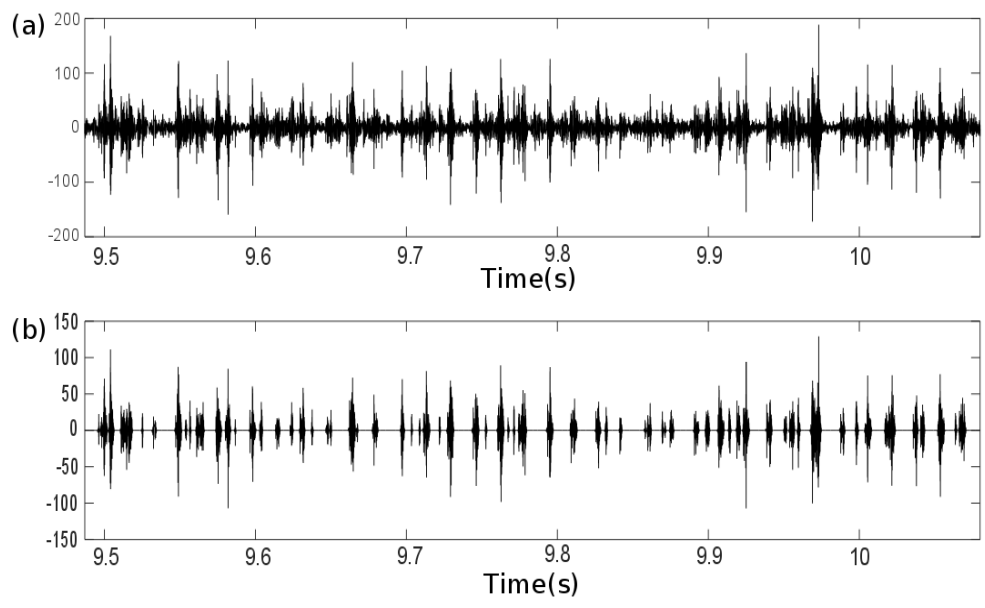


Fig. 3.43 (a) Raw signal in regime No. 5 and (b) the corresponding reconstructed transients from the HMM-based time-varying filter.

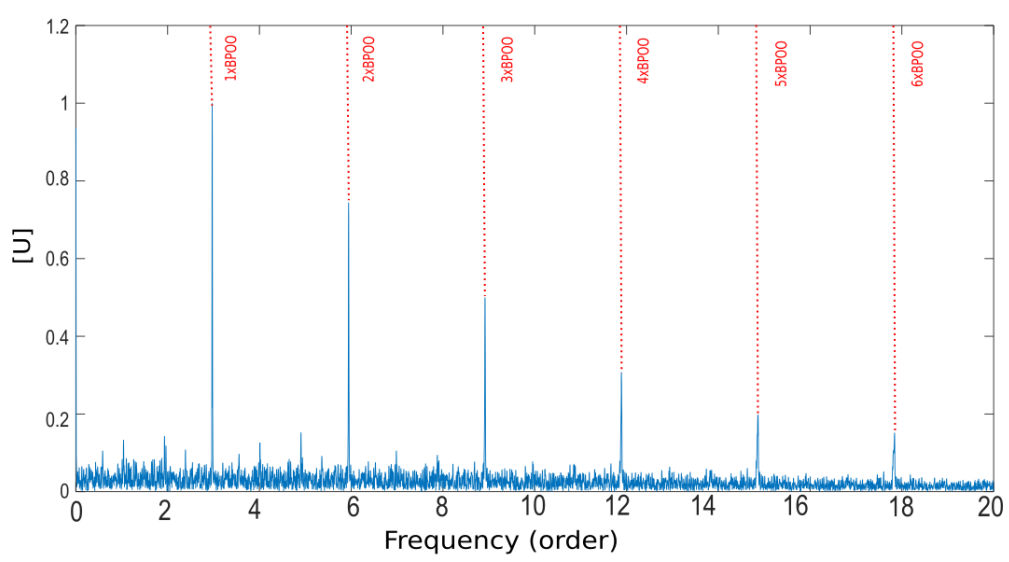


Fig. 3.44 Order spectrum of the *LLR* with markers at the suspected fault frequency BPOO and its harmonics (normalization to unit maximum value).

3.4.4 Case 9: diagnosis in the presence of multiple components

In order to demonstrate the performance of the HMM in the case of multiple-components, the dataset supported by the Department of Mechanical Engineering of Curtin University (Bentley) and made available online (<http://data-acoustics.com/measurements/bearing-faults/bearing-1/>) has been used. It corresponds to radial vibration measurements taken on the bearing housing of the SpectraQuest Machinery Fault Simulator test rig with a known outer race bearing fault. This case is interesting since there exist two different probabilities of states, as shown in Fig. 3.45. Their corresponding spectrograms are displayed in Fig. 3.46. Table 3.3 presents the parameter settings used in Case 9.

Table 3.3 Parameter settings in Case 9.

Sampling frequency F_s (Hz)	51200
Duration (s)	10
N_w	2^7
R	45
Rotation frequency – f_{rot} (Hz)	29
Ballpass frequency, outer race - BPFO (Hz)	103.6
Ballpass frequency, inner race - BPFI (Hz)	157.4
Fundamental train frequency – FTF (Hz)	11.5
Ball (roller) spin frequency – BSF (Hz)	67.3

It is seen that two families of transients occur with different frequency contents. The proposed multiple-component model introduced in Section 3.1.2 has thus been used with $K = 2$. The estimated probabilities are $\hat{\pi}_3 = 0.634$ (noise only), $\hat{\pi}_1 = 0.029$ and $\hat{\pi}_2 = 0.337$ for States 0, 1 and 2. Figure 3.47 displays the estimated diagonals of the corresponding covariance matrices. The noise spectrum is found fairly flat, whereas the first component has a high energy around [1.8 3.8] kHz and [9 11] kHz and the second component has its energy concentrated around [1.8 3.8] kHz. The spectrum of the second latent variable $\hat{\zeta}^2(i)$ reveals the BPFO of a bearing fault, while that of the first latent variable $\hat{\zeta}^1(i)$ shows some smeared

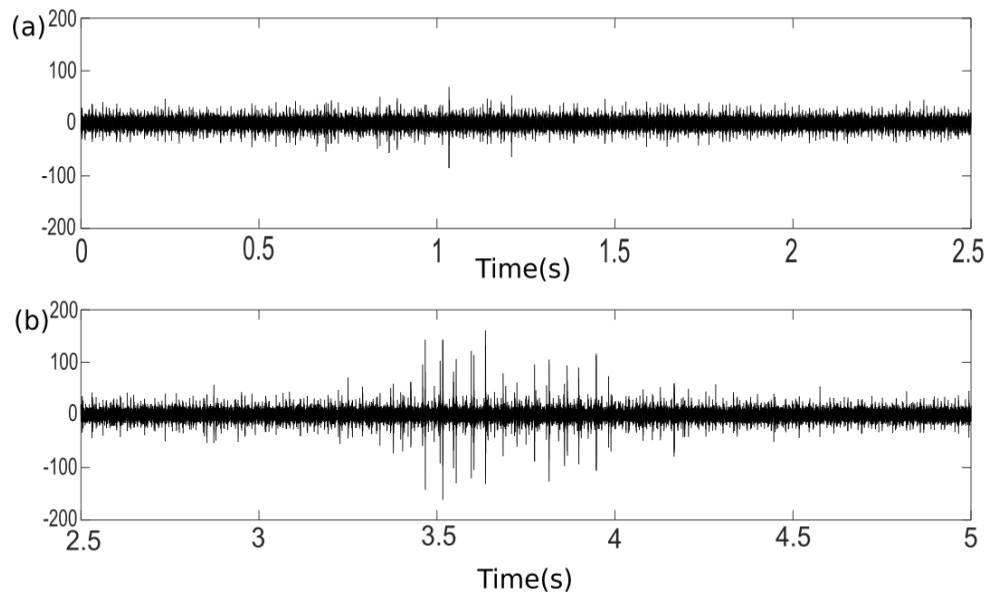


Fig. 3.45 Measured signal (a) from 0 to 2.5 s and (b) from 2.5 to 5 s which evidence of instantaneous interfering component in time interval $[3.4\ 4]$ s.

component in the low frequency, around $f_1 = 22.86$ and $f_2 = 124.4$ Hz (see Fig. 3.48) with their counterparts T_1 and T_2 showed in Fig. 3.49 (b).

The reconstructed components are displayed in Fig. 3.49 and 3.50 in intervals $[3.45\ 3.85]$ s and $[1.9\ 2.3]$ s. The first interval contains the interfering component already noticeable in Figs. 3.45 (b) and 3.46 (b), whereas the second one contains only the uniformly distributed fault signature as shown in Figs. 3.45 (a) and 3.46 (a). From Fig. 3.50 (b), it can be seen that the first component $\mathbf{X}^1(i)$ is an instantaneous interference which may be caused by structural looseness or clearance. Since frequency f_1 happens to be twice the cage speed, it might indicate that the bearing is misaligned.

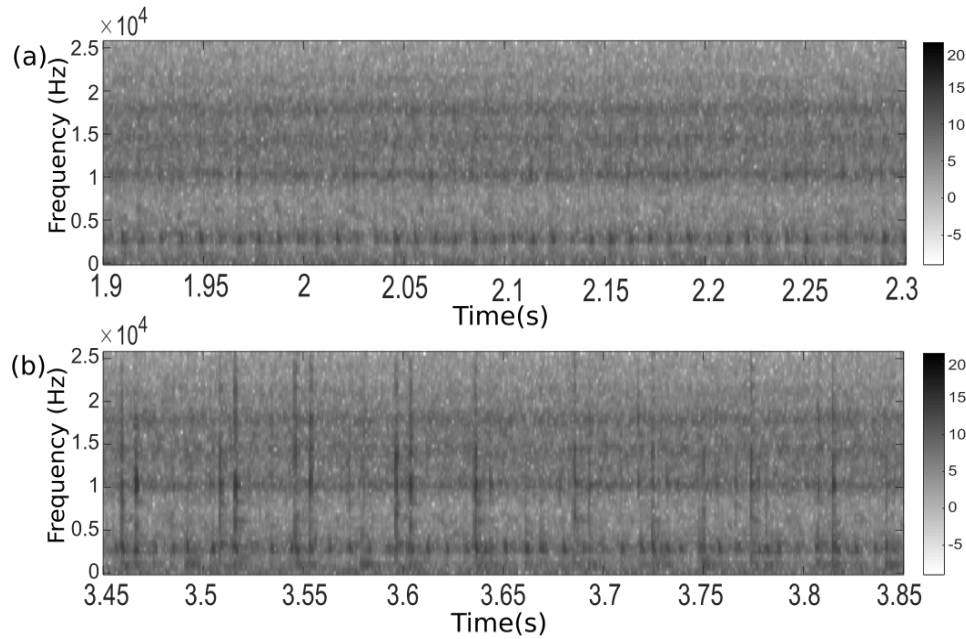


Fig. 3.46 Spectrogram (logarithmic scale) of signal in time intervals (a) [1.9 2.3] s and (b) [3.45 3.85] s of Case 9 with evidence of two states: a global distribution with spectral content in band [1.8 3.8] kHz and a local distribution with spectral content in band [1.8 3.8] kHz and [9 11] kHz.

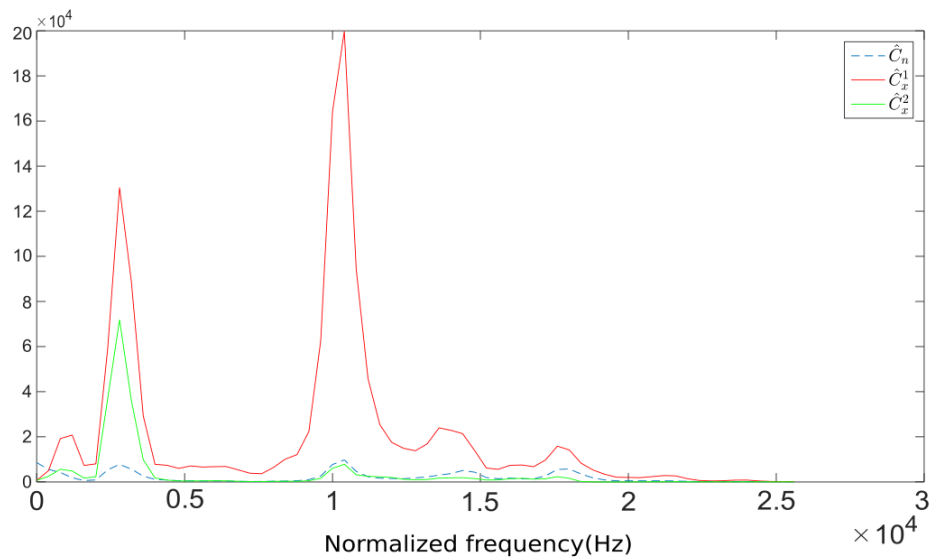


Fig. 3.47 Diagonals of the covariance matrices of the three components, $\mathbf{X}^1(i)$ (red solid line), $\mathbf{X}^2(i)$ (green solid line) and noise (blue dashed line).

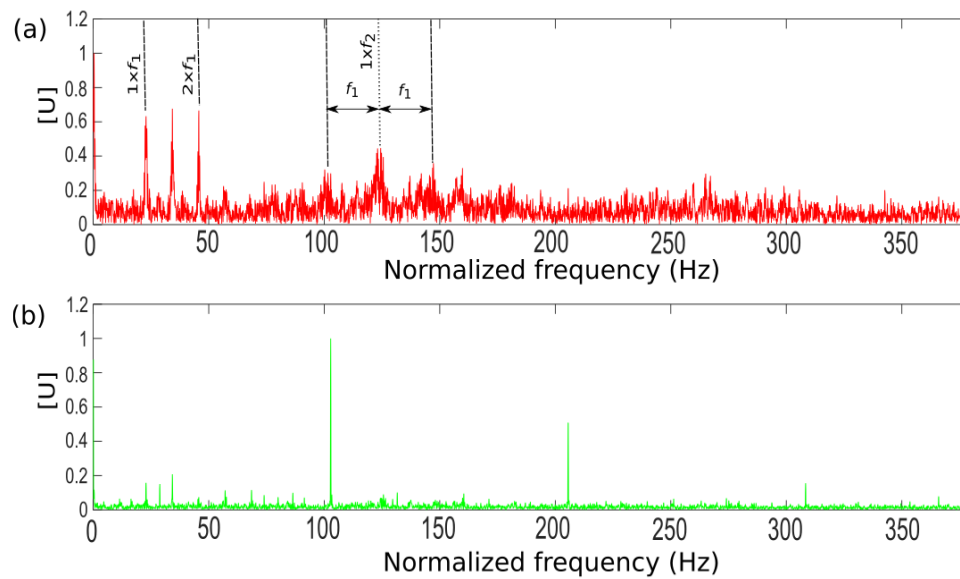


Fig. 3.48 Spectra of the estimated latent variable (a) $\zeta^1(i)$ and (b) $\zeta^2(i)$, respectively (normalization to unit maximum value).

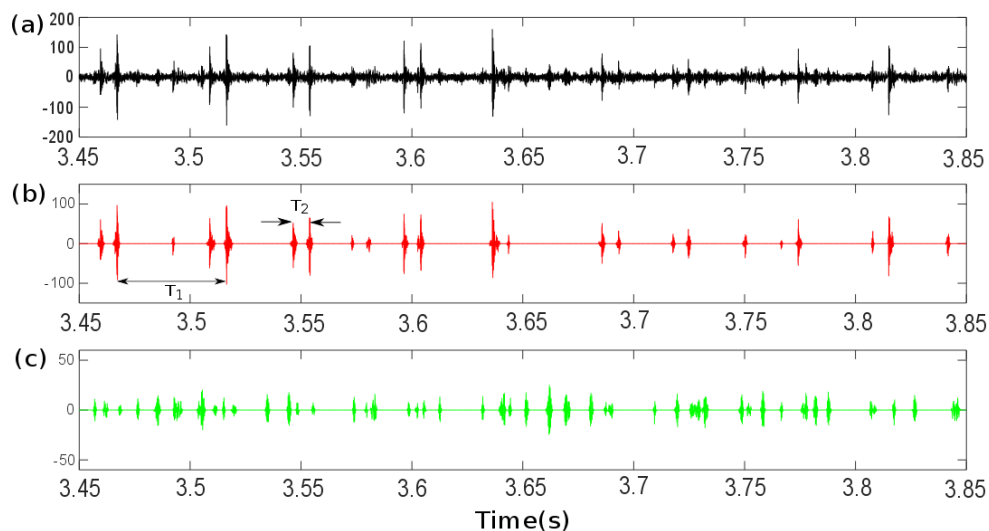


Fig. 3.49 (a) Measured signal from 3.45 to 3.85 s. Reconstructed components (b) $\mathbf{X}^1(i)$ and (c) $\mathbf{X}^2(i)$ from the HMM-based time-varying filter.

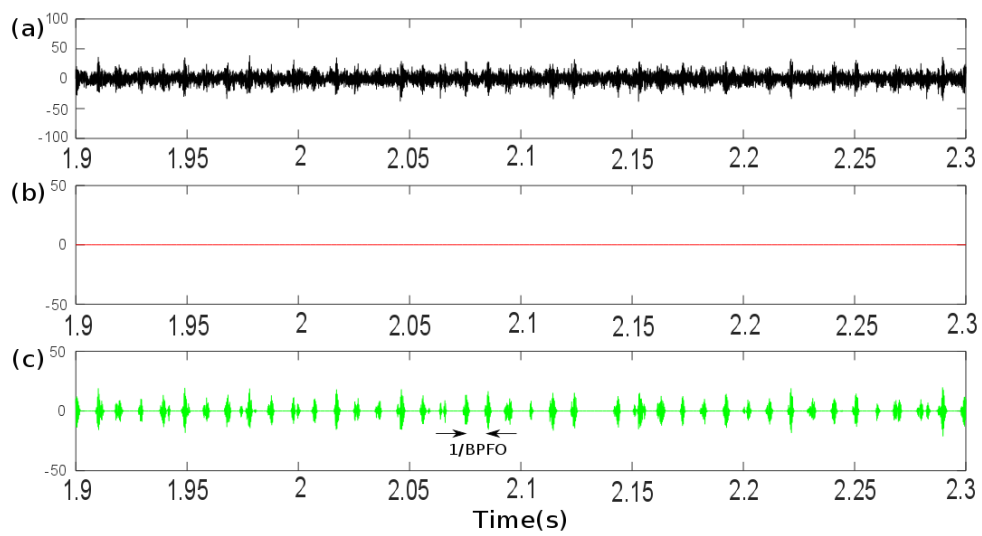


Fig. 3.50 (a) Measured signal from 1.9 to 2.3 s. Reconstructed components (b) $\mathbf{X}^1(i)$ and (c) $\mathbf{X}^2(i)$ from the HMM-based time-varying filter.

3.5 Conclusion

This chapter has introduced a new stochastic model for representing weak repetitive transients hidden in stationary noise, such as encountered with incipient faults in rolling element bearings. By using a HMM, the proposed method can detect transients in the time-frequency plane as well as estimating accurately their instants of occurrence and their spectral content. One advantage is to provide an automatic method without requirement for signal pre-processing. Another advantage is to allow the full-band reconstruction of the fault signal in the time domain. The performance of the proposed method has been demonstrated on several vibration signals from test rigs (gear and bearing). Superior or equivalent results to conventional envelope analysis and fast kurtogram have been observed in terms of fault detection. Results are clearly superior in terms of reconstruction of the fault signal when compared to band-pass filtering as achieved with the fast kurtogram. The proposed model also deals with the case where there exists multiple components and to signals which have been acquired under non-stationary regimes of the machine.

Chapter 4

Extraction of cyclostationary signals

In rotating machines, vibration-based signals primarily consist of periodic and random components (e.g. periodic signals, stationary signals, periodically-modulated signals, repetitive transients, etc). Due to the cyclic collision on the surface of defects, a series of impulses is periodically produced (i.e. each impact independently introduces a transient returned in both the amplitude and the phase). Hence the diagnostic information – particularly for bearing’s fault frequencies – are often carried in the random part and further characterized as cyclostationary processes. In particular, one can find that second-order statistics (e.g. the auto-correlation function) varies periodically as a function of time. Derived from such a statistical behavior, a novel stochastic model is proposed whose goal is to detect and recover the desired signals (i.e. cyclostationary (CS) signals on the second-order) in the presence of competing sources (background noise and interfering signals).

4.1 Periodic-variance based model

This section begins with a description of the proposed model with its corresponding assumptions. Since the model highly depends on unobserved latent variables, an expectation-maximization (EM) algorithm is introduced for finding maximum-a-posteriori (MAP) estimates of parameters in the proposed statistical model; meanwhile an extraction scheme is also inserted inside the iterative steps of the EM algorithm so as to improve the convergence as presented in section 4.2.

4.1.1 Model and assumptions

The idea of modelling the vibration signal $y(t)$ – which contains the fault component $x(t)$ immersed in background noise $n(t)$ – as a CS signal is motivated by the physical display

of some hidden repetitiveness of its energy flow, which has been illustrated in Fig. 2.5. Hereafter $x(t)$ is assumed to be a second-order CS signal and $n(t)$ a stationary signal; thereby the measurement $y(t)$ is constituted by the linear combination

$$y(t) = x(t) + n(t). \quad (4.1)$$

In particular, the repetitive transients are characterized by resonances of the bearing and of their receiving structure, which are often in relatively high frequencies. Therefore it is assumed that such transients can be well captured in segments of the STFT of $x(t)$:

$$STFT_x(i, f_b) = \sum_{m=0}^{N_w-1} w[m] \cdot x[iR + m] \cdot e^{-j2\pi f_b \frac{m}{F_s}} \quad (4.2)$$

where $\{w[m]\}$ denotes a positive and smooth N_w -long data-window which truncates a segment of the L -long signal $x(t)$ at time datum i ($i = 1, \dots, N$, $N = \text{floor}[(L - N_w)/R + 1]$) with window shift R ($1 < R < N_w$) and where $f_b = b \cdot \Delta f$ denotes the frequency (from 0 to $F_s/2$) with frequency resolution $\Delta f = F_s/N_w$ and bin index $b = 1, \dots, N_f$ with $N_f = N_w/2 + 1$.

Hereafter the vibration signal is first decomposed in the time-frequency plane. Then it is required to phase all the segments to zero at the beginning of the signal, at time instant $t = 0$. This phase correction can also be interpreted as the Gabor transform

$$Y(i, f_b) = \sum_{n=0}^{L-1} w[n - iR] \cdot y[n] \cdot e^{-j2\pi f_b \frac{n}{F_s}} = STFT_y(i, f_b) \cdot e^{-j2\pi f_b \frac{iR}{F_s}}. \quad (4.3)$$

Let us consider f_b corresponding to the carrier frequency under the assumption of stationarity over a time interval of length N_w/F_s . The interpretation of $Y(i, f_b)$ is the ‘‘complex envelope’’ of signal $y[n]$ in a narrow frequency band of bandwidth Δf centered on f_b and sampled at time instant iR/F_s . It is reminded that its second-order statistics are expected to embody the cyclic frequency which corresponds to bearing fault signatures. The following subsection introduces a periodic-variance based model to extract the CS signal from the background noise with respect to different probabilistic properties.

4.1.2 Periodic-variance based model

As opposed to stationary signals, CS signals contain extra information due to their hidden periodicities. In the time domain, such extra information is carried by the periodic variations of statistical descriptors such as the instantaneous auto-correlation function and its related quantities (instantaneous power, the envelope function, etc.) [73, 78]. This means that the

Wigner-Ville *spectrum*,

$$\mathcal{W}\mathcal{V}_{XX}[i, f_b] = \mathbb{E} \left\{ \mathcal{F}_{\tau \rightarrow f_b} \{X[i + \tau/2]X[i - \tau/2]^*\} \right\}, \quad (4.4)$$

(where i and τ stand for the instantaneous time sampling and the time-lag, \mathbb{E} for the ensemble average operator, \mathcal{F} for “Fourier transform” and $*$ for the complex conjugate) is a periodic function of time instant i with some period T ; therefore it can also be expressed by a Fourier series

$$\mathcal{W}\mathcal{V}_{2X}[i, f_b] = \mathcal{W}\mathcal{V}_{2X}[i + T, f_b] = \sum_{\alpha_j \in \mathcal{A}} \mathcal{S}_{2X}(f_b; \alpha_j) e^{j2\pi\alpha_j i R / F_s} \quad (4.5)$$

over the spectrum $\mathcal{A} = \{\alpha_j\}$ of cyclic frequencies α_j associated with the non-zero Fourier coefficients, $\mathcal{S}_{2X}(f_b; \alpha_j)$, which denotes the cyclic power spectrum (units = power/Hz) [73].

The interpretation of $\mathcal{W}\mathcal{V}_{2X}[i, f_b]$ is an instantaneous power spectrum at time instant i which reflects a joint time-frequency energy distribution. Different from the current statistical descriptors, the proposed model recognizes the non-stationarity of signal $y(t)$ as a time-dependent distribution over time index i . From now on, the “complex envelope” of measurement $y(t)$, $Y(i, f_b)$, is assumed to be a function of the time datum i only and to be indexed by frequency f_b which follows a time-dependent complex normal distribution

$$Y(i, f_b) \sim \mathcal{CN}(0, \sigma_y^2(i); f_b) \quad (4.6)$$

where $\mathcal{CN}(\mu, \sigma^2(i); f_b)$ denotes the circular-symmetric complex normal distribution with mean μ and instantaneous variance $\sigma^2(i)$ applied to frequency f_b . Without loss of generality, it is assumed that $\mu = 0$ (as obtained after first centering the signal). It is hereby reminded that in this chapter the proposed model will consider the frequency f_b as a parameter rather than a variable¹.

Now let us consider the unknown time-dependent variance $\sigma^2(i)$ as hidden variables, which is further regularized by the inverse gamma distribution

$$\sigma^2(i) \sim \text{Inv} - \text{Gamma}(\alpha, \beta) \quad (4.7)$$

with shape parameter α and scale parameter β . For instance, $\sigma_x^2(i) \sim \text{Inv} - \text{Gamma}(\alpha_x, \beta_x)$ and $\sigma_n^2(i) \sim \text{Inv} - \text{Gamma}(\alpha_n, \beta_n)$ stand for the time-dependent variance of $x(t)$ and $n(t)$, respectively.

¹This means that the frequency f_b is separable and nondistinctive from other frequencies. For notational simplicity, it will be dropped off in cases wherein it will not cause any confusion.

In the case of CS signals, $\sigma_x^2(i; f_b)$ reflects the hidden repetitiveness of the energy flow in the frequency band f_b . Then the time-dependent variance of $Y(i, f_b)$ reads

$$\sigma_y^2(i; f_b) = \sigma_x^2(i; f_b) + \sigma_n^2(f_b) \quad (4.8)$$

where $i = i + j \times T$ ($T \times R/F_s$ is the periodic time of impacts and j is the integer) and $\sigma_n^2(f_b)$ denotes the variance of $N(i, f_b)$ which is assumed time-invariant for simplicity.

Meanwhile, all the unknown parameters of the proposed model are denoted as $\boldsymbol{\theta} = \{\sigma_x^2(i), \sigma_n^2, T, \alpha_x, \beta_x, \alpha_n, \beta_n\}$. It is highlighted that the periodic-variance based model embodies hidden probability distributions varying periodically with time index i , i.e. $\sigma_x^2(i)$ are hidden in the sense that they are not observed directly. This is what will make the estimation of parameters difficult. Hence the hidden period T is considered as the prior knowledge² that corresponds to “bearing characteristic frequencies” as presented in Table 1.1. In addition, the unknown shape and scale parameters $\{\alpha_x, \beta_x, \alpha_n, \beta_n\}$ will be estimated and kept constant in the procedure of data-driven initialization. The following section provides the inference of the parameters as well as the extraction of the CS signal.

4.2 Extraction scheme for cyclostationary signals

Now, let introduce the posterior probability distribution of the CS signal $X(i, f_b)$ and of the unknown parameters $\boldsymbol{\theta}$ given the measurement $Y(i, f_b)$ as

$$p(X(i, f_b), \boldsymbol{\theta} | Y(i, f_b)). \quad (4.9)$$

Since it is a possible task to extract $X(i, f_b)$ and estimate $\boldsymbol{\theta}$ together, it will be useful to consider the following conditional

$$p(X(i, f_b), \boldsymbol{\theta} | Y(i, f_b)) \propto \begin{cases} p(X(i, f_b) | Y(i, f_b), \boldsymbol{\theta}); & \text{knowing : } \boldsymbol{\theta} \\ p(\boldsymbol{\theta} | Y(i, f_b), X(i, f_b)); & \text{knowing : } X(i, f_b) \end{cases}. \quad (4.10)$$

Since it is difficult to find a closed-form solution, the EM algorithm [75] is used as an iterative method to find the maximum posterior estimates. In particular, the EM iteration alternates between performing an *expectation* (E) step and a *maximization* (M) step as shown in Eq. 4.10. The EM algorithm makes use of the following quantities.

²To quickly fix the hidden period, we strongly recommend the Fast Spectral Correlation (Fast-SC) which provides a fast computation of the Spectral Correlation [79]. Especially in practical applications, it illuminates the performance of extraction scheme as demonstrated in section 4.4.

(a) *Expectation step : extraction of the CS signal*

First, let introduce the posterior probability distribution of the signal of interest $X(i, f_b)$ with the current $\hat{\boldsymbol{\theta}}$ as

$$p(X(i, f_b) | Y(i, f_b), \hat{\boldsymbol{\theta}}) \propto p(Y(i, f_b) | X(i, f_b), \hat{\boldsymbol{\theta}})p(X(i, f_b) | \hat{\boldsymbol{\theta}}) \quad (4.11)$$

where $\boldsymbol{\mu}_x(i) = [\mu_x(i, f_1) \dots \mu_x(i, f_{N_f})]^T$ denotes the N_f -dimensional mean vector at time index i , and $\boldsymbol{\mu}_x = [\boldsymbol{\mu}_x(1) \dots \boldsymbol{\mu}_x(N)]^T$ denotes the $N \times N_f$ time-dependent mean matrix; $\boldsymbol{\sigma}_x(i) = [\sigma_x(i, f_1) \dots \sigma_x(i, f_{N_f})]^T$ denotes the N_f -dimensional variance vector at time index i , and $\boldsymbol{\sigma}_x = [\boldsymbol{\sigma}_x(1) \dots \boldsymbol{\sigma}_x(N)]^T$ denotes the $N \times N_f$ time-dependent variance matrix.

Then the posterior probability density at a given frequency f_b reads

$$p(X(i, f_b) | Y(i, f_b), \hat{\boldsymbol{\theta}}) = \frac{\exp\left\{-\frac{|Y(i, f_b) - X(i, f_b)|^2}{\hat{\sigma}_n^2(f_b)}\right\} \exp\left\{-\frac{|X(i, f_b)|^2}{\hat{\sigma}_x^2(i, f_b)}\right\}}{\pi^2 \hat{\sigma}_n^2(f_b) \hat{\sigma}_x^2(i, f_b)}. \quad (4.12)$$

After some manipulations, Eq. 4.12 can be expressed as

$$p(X(i, f_b) | Y(i, f_b), \hat{\boldsymbol{\theta}}) = \frac{\exp\left\{-\frac{|X(i, f_b) - \mu_x(i, f_b)|^2}{\sigma_x^2(i, f_b)}\right\}}{\pi \sigma_x^2(i, f_b)} = \mathcal{C} \mathcal{N}(X(i, f_b); \mu_x(i, f_b), \sigma_x(i, f_b)) \quad (4.13)$$

with

$$\begin{cases} \sigma_x^2(i, f_b) = \left(\frac{1}{\hat{\sigma}_n^2(f_b)} + \frac{1}{\hat{\sigma}_x^2(i, f_b)}\right)^{-1} \\ \mu_x(i, f_b) = \frac{\hat{\sigma}_x^2(i, f_b)}{\hat{\sigma}_n^2(f_b)} Y(i, f_b) \end{cases}. \quad (4.14)$$

Therefore the expectation of the signal of interest $X(i, f_b)$ reads

$$\mathbb{E}\{X(i, f_b) | Y(i, f_b), \hat{\boldsymbol{\theta}}\} = \mu_x(i, f_b) = \frac{1}{1 + H(i, f_b)} Y(i, f_b) \quad (4.15)$$

where $H(i, f_b)$ denotes the time-dependent variance ratio between the noise and the CS signal in the frequency band f_b calculated by

$$H(i, f_b) = \frac{\hat{\sigma}_n^2(f_b)}{\hat{\sigma}_x^2(i, f_b)}. \quad (4.16)$$

Finally, the time signal $\hat{x}[n]$ is obtained from Eq. 4.15 by using the inverse STFT and this is also the end of the E step.

Three remarks are noteworthy. First, it is seen that Eq. 4.15 corresponds to a periodic time-varying filter from which superior performance is expected than from a conventional time-invariant filter. Second, the standard Wiener filter appears as a particular case under the

assumption of stationarity, that is

$$\mathbb{E}\{X(i, f_b) | Y(i, f_b)\} = \frac{1}{1 + \frac{\hat{\sigma}_n^2(f_b)}{\hat{\sigma}_x^2(f_b)}} Y(i, f_b), \quad (4.17)$$

where the time index i of the time-dependent variance $\hat{\sigma}_x^2(i; f_b)$ stays constant for all time instants. In other words, Eq. 4.17 then corresponds to the case where $\sigma_y^2 = \sigma_x^2 + \sigma_n^2$ instead of Eq. 4.8. Last but not least, Eq. 4.15 can be seen as an analogue to Eq. 3.24 (in the case where the covariance $\hat{\mathbf{C}}_x^k$ is diagonal), however, the latent variable $\zeta^k(i)$ is replaced by a periodic-variance based variable $\hat{\sigma}_x^2(i; f_b)$.

(b) *Maximization step : estimation of unknown parameters*

Next, let introduce the posterior probability distribution of the unknown parameters $\boldsymbol{\theta}$ with the current $\hat{X}(i, f_b)$ as

$$p(\boldsymbol{\theta} | Y(i, f_b), \hat{X}(i, f_b)) \propto p(Y(i, f_b) | \hat{X}(i, f_b), \boldsymbol{\theta}) p(\boldsymbol{\theta} | \hat{X}(i, f_b)) \quad (4.18)$$

where $\boldsymbol{\theta}$ consists of such unknown parameters $\{\sigma_x^2(i), \sigma_n^2\}$. As mentioned above, the estimation of the shape and scale parameters $\{\alpha_x, \beta_x, \alpha_n, \beta_n\}$ will be introduced in the procedure of data-driven initialization for the EM algorithm in section 4.3.

In the proposed stochastic model, let us further assume that $\sigma_x^2(i)$ and σ_n^2 are assigned informative priors³ in the form of analytically tractable conjugate distributions⁴

$$p(\sigma_x^2(i; f_b) | Y(i, f_b), \hat{X}(i, f_b)) \sim \text{Inv-Gamma}(\alpha_x + \frac{N}{2}, \beta_x + \frac{N}{2} \sum_{\alpha_j \in \mathcal{A}} \mathcal{S}_{2X}(f_b; \alpha_j) e^{j2\pi\alpha_j i R / F_s}) \quad (4.19)$$

where $\mathcal{S}_{2X}(f_b; \alpha_j)$ denotes the cyclic power spectrum with known spectrum $\mathcal{A} = \{\alpha_j\}$ of cyclic frequencies α_j , as calculated by

$$\mathcal{S}_{2X}(f_b; \alpha_j) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\hat{X}(i, f_b))^2 e^{-j2\pi\alpha_j i R / F_s}. \quad (4.20)$$

It is highlighted that for the CS stochastic process the non-zero Fourier coefficients only appear on a countable set of frequencies, which are relative to the hidden periodicities of

³This expresses specific, definite information about a variable that is determined largely by pre-existing evidence rather than any original assumption; the terms ‘‘prior’’ and ‘‘posterior’’ are generally relative to a specific datum or observation. A reasonable approach is to make the prior parametrization with expected value and variance value as schemed in Eq. 4.21.

⁴It is proved that the inverse gamma distribution is a conjugate prior of the complex normal distribution where $p(\sigma^2 | y[n], \mu) \sim \text{Inv-Gamma}(\alpha + \frac{n}{2}, \beta + \sum_{i=1}^n \frac{|y_i - \mu|^2}{2})$, more details are shown in Appendix C.

impacts corresponding to the characteristic frequencies in Table 1.1. This is important for the CS signal spectral contents to be estimated and recovered by the proposed model provided that the prior assumption of known hidden period T applies to all frequency bins f_b .

Now let us remind the property of *Inv – Gamma*(α, β) parametrization with

$$\begin{cases} \mathbb{E}\{\sigma^2\} = \frac{\beta}{\alpha-1} & \text{for } \alpha > 1 \\ \text{Var}\{\sigma^2\} = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} & \text{for } \alpha > 2 \end{cases}. \quad (4.21)$$

Therefore the expectation of the hidden time-dependent variance $\sigma_x^2(i; f_b)$ reads

$$\mathbb{E}\{\sigma_x^2(i; f_b)\} = \frac{\beta_x + \frac{N}{2} \sum_{\alpha_j \in \mathcal{A}} \mathcal{S}_{2X}(f_b; \alpha_j) e^{j2\pi\alpha_j i R / F_s}}{\alpha_x + \frac{N}{2} - 1}. \quad (4.22)$$

Likewise, the expectation of the time-invariant variance $\sigma_n^2(f_b)$ is

$$\mathbb{E}\{\sigma_n^2(f_b)\} = \frac{\beta_n + \sum_{i=1}^N \frac{|Y(i, f_b) - \hat{X}(i, f_b)|^2}{2}}{\alpha_n + \frac{N}{2} - 1} \quad (4.23)$$

which appears as a particular case under the stationary assumption of the noise.

By now, the M step is completely introduced and it will alternate with the E step in the iterative process, thus driving the convergence of the proposed EM algorithm. The performance of the proposed reconstruction scheme of CS signals will be demonstrated on the synthetic and experimental signals of sections 4.3 and 4.4, respectively.

4.3 Parameter selection

This section discusses the initialization and setting of the periodic-variance based model. Thereby an effective data-driven initialization for the EM algorithm is proposed. Meanwhile several synthetic signals of interest (modulated white noise, modulated narrow-band noise, the transient signal) are generated and analyzed for demonstration. To demonstrate the potential of the proposed scheme, an extended case (repetitive transient signal) is eventually tested with two random variables on the arrival time and the magnitude of transients.

4.3.1 Initial settings

The basic idea of the proposed scheme is to model the contaminated signal of interest (i.e. CS signals on the second-order) by means of latent stochastic variables, which reflects the

periodic variations on its energy distribution applied to the full-band spectra; in addition, by maximizing the posterior probability distribution of the CS signal, it enables the hidden variations to recover the desired signal with its full-band spectral content. This is achieved in the domain of the STFT. Therefore, the first required parameters to tune are the window length N_w and the window shift R . Since they have been discussed in subsection 3.3.1, let us continue with the procedure of initialization of the EM algorithm.

Initializing parameters of the EM algorithm

The EM algorithm generally requires a good initialization step. A simple solution is given hereafter to obtain initial values one-by-one along the frequency indexes f_b . In the periodic-variance based model, let start with the first spectral content of the unknown variance $\{\sigma_x^2(i; f_1), \sigma_n^2(f_1)\}$.

a) First frequency initialization

The initial variance of noise, $\mathbb{E}\{\sigma_n^2(f_1)\}^{[0]}$, is obtained by taking the mean value of the squared magnitude of measurement $Y(i, f_1)$ with respect to time instant i ,

$$\mathbb{E}\{\sigma_n^2(f_1)\}^{[0]} = \frac{1}{N} \sum_{i=1}^N |Y(i, f_1)|^2. \quad (4.24)$$

Then the initial time-dependent variance of the CS signal, $\mathbb{E}\{\sigma_x^2(i; f_1)\}^{[0]}$, is obtained by the subtraction

$$\mathbb{E}\{\sigma_x^2(i; f_1)\}^{[0]} = \left(\sum_{\alpha_j \in \mathcal{A}} \left(\frac{1}{N} \sum_{i=1}^N |Y(i, f_1)|^2 e^{-j2\pi\alpha_j i R / F_s} \right) e^{j2\pi\alpha_j i R / F_s} - \mathbb{E}\{\sigma_n^2(f_1)\}^{[0]} \right)_+ \quad (4.25)$$

where operator $(\dots)_+$ keeps only the positive values of a vector. This approximation is based on the assumption that the measurement $Y(i, f_1)$ embodies the CS signal $X(i, f_1)$ in the presence of stationary noise $N(f_1)$.

Next, according to Eq. 4.21, one can estimate the shape and scale parameters of noise from

$$\begin{cases} \hat{\alpha}_n(f_1) = \frac{(\mathbb{E}\{\sigma_n^2(f_1)\}^{[0]})^2}{\text{Var}\{\sigma_n^2(f_1)\}^{[0]}} + 2 \\ \hat{\beta}_n(f_1) = (\hat{\alpha}_n(f_1) - 1) \times \mathbb{E}\{\sigma_n^2(f_1)\}^{[0]} \end{cases} \quad (4.26)$$

where the initial variance of $\sigma_n^2(f_1)$ can be taken as large as possible to achieve a wide dispersion e.g. $\text{Var}\{\sigma_n^2(f_1)\}^{[0]} = 10 \times \mathbb{E}\{\sigma_n^2(f_1)\}^{[0]}$. This estimation also works for the shape and scale parameters of the CS signal at frequency f_b , wherein the parameters are taken as $\hat{\alpha}_x(f_1) = 0$ and $\hat{\beta}_x(f_1) = 0$ to avoid the risk of overfitting.

b) *Spectral components initialization* : iterative procedure

The initialization of spectral components is implemented by means of iterative way, i.e. $f_{b-1} \rightarrow f_b$ (bin index $b = 2, \dots, N_f$), so as to start with the initial variance of noise

$$\mathbb{E}\{\sigma_n^2(f_b)\}^{[0]} = \frac{1}{N} \sum_{i=1}^N |Y(i, f_{b-1}) - \widehat{X}(i, f_{b-1})|^2. \quad (4.27)$$

Then the initial time-dependent variance of the CS signal reads

$$\mathbb{E}\{\sigma_x^2(i; f_b)\}^{[0]} = \sum_{\alpha_j \in \mathcal{A}} \left(\frac{1}{N} \sum_{i=1}^N |Y(i, f_b)|^2 e^{-j2\pi\alpha_j i R/F_s} \right) e^{j2\pi\alpha_j i R/F_s} - \mathbb{E}\{\sigma_n^2(f_b)\}^{[0]}. \quad (4.28)$$

Next, the shape and scale parameters of noise are initialized as

$$\begin{cases} \hat{\alpha}_n(f_b) = \frac{(\mathbb{E}\{\sigma_n^2(f_b)\}^{[0]})^2}{\text{Var}\{\sigma_n^2(f_b)\}^{[0]}} + 2 \\ \hat{\beta}_n(f_b) = (\hat{\alpha}_n(f_b) - 1) \times \mathbb{E}\{\sigma_n^2(f_b)\}^{[0]} \end{cases} \quad (4.29)$$

where the initial variance of $\sigma_n^2(f_b)$ can be taken as $\text{Var}\{\sigma_n^2(f_b)\}^{[0]} = 10 \times \mathbb{E}\{\sigma_n^2(f_b)\}^{[0]}$. Similarly, for the CS signal,

$$\begin{cases} \hat{\alpha}_x(f_b) = \frac{(\mathbb{E}\{\sigma_x^2(f_b)\}^{[0]})^2}{\text{Var}\{\sigma_x^2(f_b)\}^{[0]}} + 2 \\ \hat{\beta}_x(f_b) = (\hat{\alpha}_x(f_b) - 1) \times \mathbb{E}\{\sigma_x^2(f_b)\}^{[0]} \end{cases} \quad (4.30)$$

where the initial variance of $\sigma_x^2(f_b)$ can be taken as $\text{Var}\{\sigma_x^2(f_b)\}^{[0]} = 10 \times \mathbb{E}\{\sigma_x^2(f_b)\}^{[0]}$ and the initial expectation of $\sigma_x^2(f_b)$ as $\mathbb{E}\{\sigma_x^2(f_b)\}^{[0]} = \frac{1}{N} \sum_{i=1}^N \mathbb{E}\{\sigma_x^2(i; f_b)\}^{[0]}$.

It has been observed in numerous experiments that the proposed initializations are often quite close to the maximum a posteriori estimates.

4.3.2 Cases 1 & 2: demonstration of parameter selection

To demonstrate the performance of the proposed algorithm, a synthetic signal is generated with a resonance frequency $f_0 = 0.1$ Hz which is further modulated by a relatively high fault frequency $\alpha_0 = 5 \times 10^{-3}$ Hz ($T = 1/\alpha_0 = 200$ s, the sampling frequency is normalized to $F_s = 1$ Hz). More precisely, the produced signal is described as:

$$y(t) = \sum_{j=-\infty}^{+\infty} h(t - jT) + n(t) \quad (4.31)$$

$$H(z) = \frac{b_1}{a_1 + a_2 \cdot z^{-1} + a_3 \cdot z^{-2}} \quad (4.32)$$

where $n(t)$ denotes white noise that achieves a noise-to-signal-ratio of 6 dB and the signal length is $L = 10^4$ samples. A second-order system is defined by Eq. 4.32, whose numerator and denominator coefficients are $\mathbf{b} = [1]$ and $\mathbf{a} = [1, -2\cos(2\pi f_0)r, r^2]$ with $r = 0.95$, respectively. Fig. 4.1 shows the spectrogram (magnitude of the STFT) of the raw signal, with its periodic energy flow, with period T (dashed line in red). The time record is displayed in Fig. 4.2 (a).

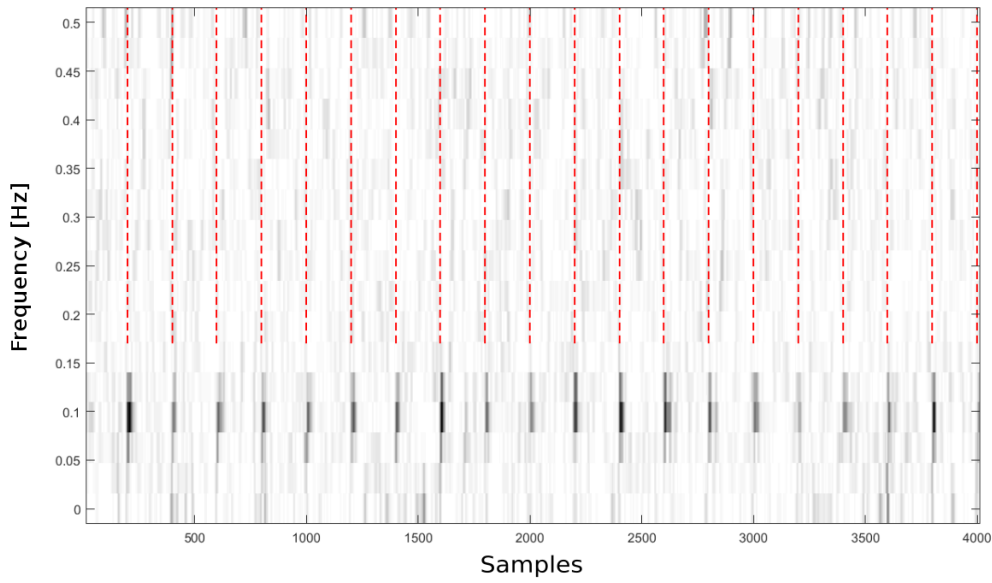


Fig. 4.1 Spectrogram of the signal simulated in Case 1 with resonance frequency $f_0 = 0.1$ Hz, $r = 0.95$ and fault frequency $\alpha_0 = 5 \times 10^{-3}$ Hz ($T = 1/\alpha_0 = 000$ s, $N_w = 2^5$ and $R = 4$).

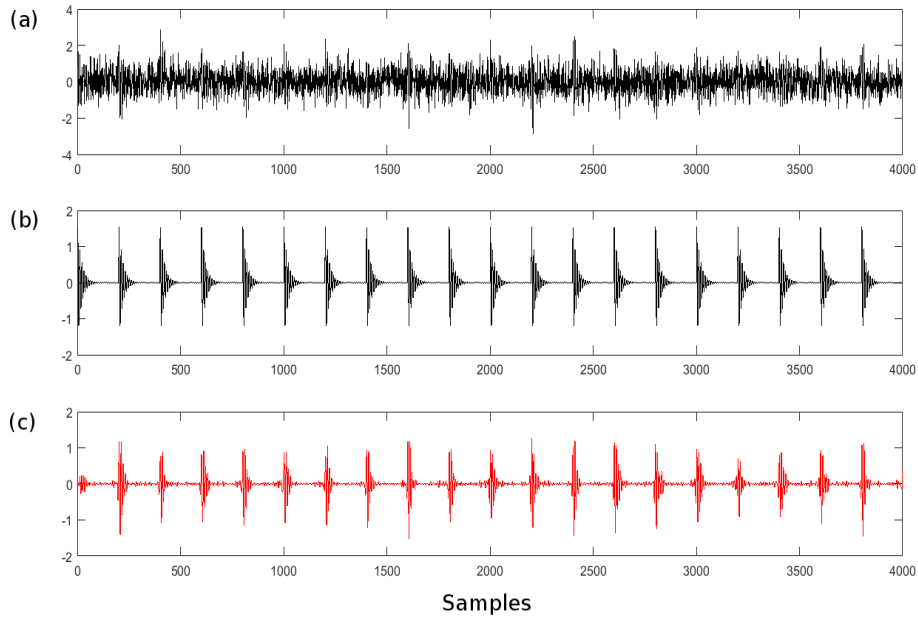


Fig. 4.2 (a) Synthetic signal of Case 1 with white noise (noise-to-signal-ratio= 6 dB). (b) Synthetic transient signal. (c) Recovered time signal $\hat{x}[n]$.

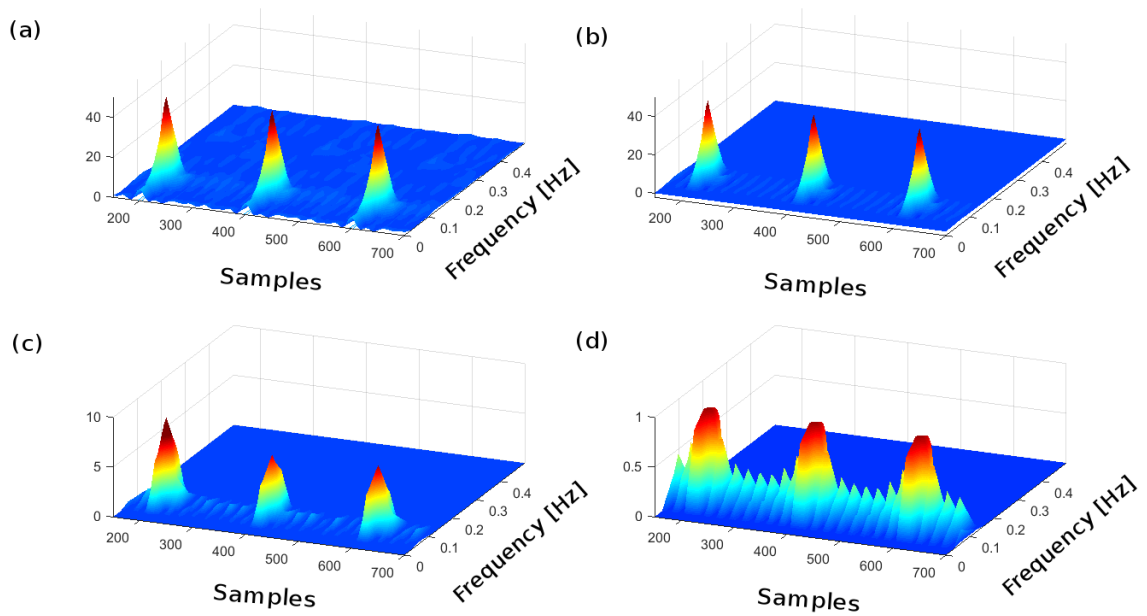


Fig. 4.3 (a) Initialized time-dependent variance of the CS signal, $\hat{\sigma}_x^2(i; f_b)^{[0]}$, and (b) the estimated $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm. (c) Spectrogram of the estimated CS signal, $\hat{X}(i, f_b)$, with (d) its periodic time-varying filter $1/(1 + H(i; f_b))$ as defined in Eqs. 4.15-4.16.

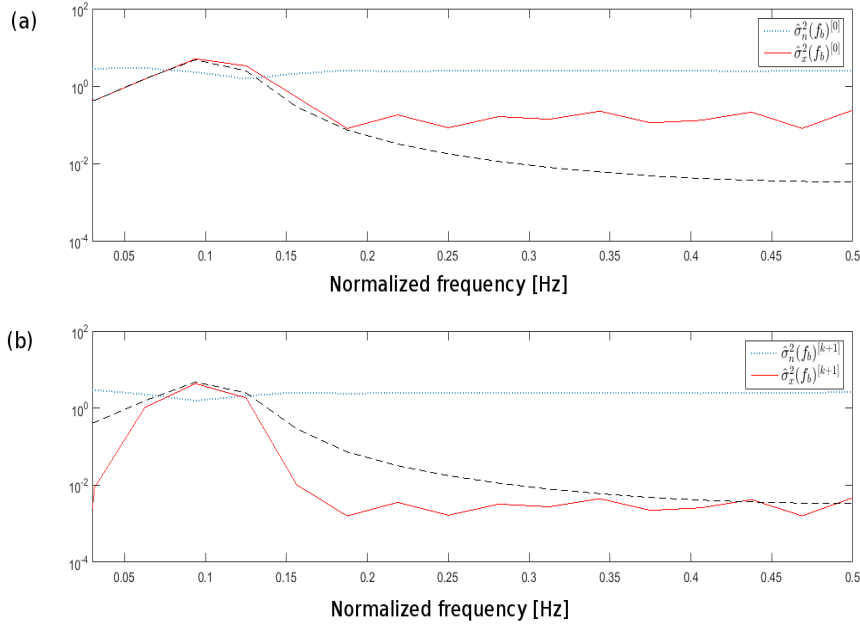


Fig. 4.4 (a) Initialized spectra (variance of noise and CS signals $\hat{\sigma}_n^2(f_b)^{[0]}$ and $\hat{\sigma}_x^2(f_b)^{[0]}$, in blue dotted line and red line, respectively) and (b) estimated spectra from the EM algorithm, ($\hat{\sigma}_n^2(f_b)^{[k+1]}$ and $\hat{\sigma}_x^2(f_b)^{[k+1]}$, in blue dotted line and red line, respectively) together with the theoretical squared magnitude frequency response $|H(z)|^2$ (black dashed line).

Following Eqs. 4.24-4.30, one can initialize the parameters $\hat{\sigma}_x^2(i; f_b)^{[0]}$, $\hat{\sigma}_x^2(f_b)^{[0]}$ and $\hat{\sigma}_n^2(f_b)^{[0]}$ as shown in Fig. 4.3 (a) and Fig. 4.4 (a). It is seen that the proposed initialization is simple and effective, even though the estimated spectrum of the signal of interest still contains a significant contribution from noise especially beyond 0.18 Hz as seen in Fig. 4.4 (a). After convergence of the EM algorithm, the estimation of the signal and noise spectra, $\hat{\sigma}_x(f_b)^{[k+1]}$ and $\hat{\sigma}_n(f_b)^{[k+1]}$ are close to the real values as can be seen in Fig. 4.4 (b). Particularly, the very good estimation of the hidden variance $\hat{\sigma}_x(i; f_b)^{[k+1]}$ is displayed in Fig. 4.3 (b) and further verified by the cyclostationary index in Fig. 4.5.

Finally, the recovered time signal $\hat{x}[n]$ is displayed in Fig. 4.2 (c) and obtained by using the inverse STFT of the estimated CS signal, $\hat{X}(i, f_b)$, which passes a time-varying filter over the full-band spectrum as shown in Fig. 4.3 (c) and (d), respectively.

To highlight and summarize the performance of proposed extraction scheme, Fig. 4.6 (a) shows a better view of the recovered time signal $\hat{x}[n]$ associated with synthetic transient signal. It illustrates that although the proposed scheme addresses the extraction of the CS signal with the full-band spectrum, its output can independently and adaptively recognize the weight along its frequency f_b with regards to the CS index. This property also can be verified by the initialized parameter $\hat{\sigma}_x^2(i; f_b)^{[0]}$, even though it still embodies some marginal spectral

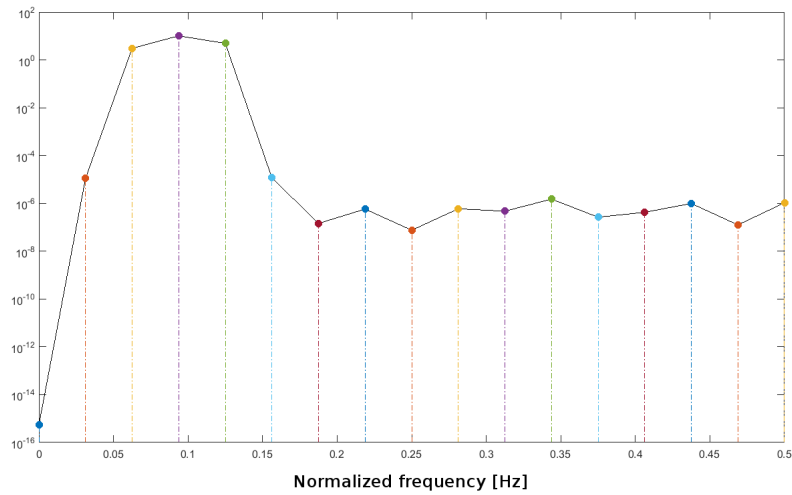


Fig. 4.5 Cyclostationary index: function of the frequency f_b by means of taking the standard deviation of $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ over the time instant i .

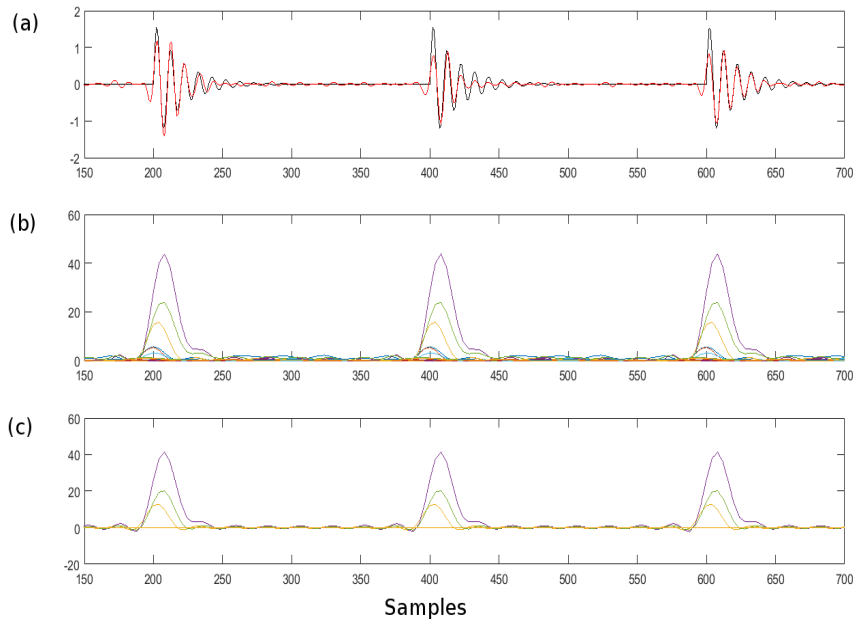


Fig. 4.6 (a) Enlarged view of the recovered time signal $\hat{x}[n]$ of Fig. 4.2 (b) and (c). (b) Initialized time-dependent variance of the CS signal, $\hat{\sigma}_x^2(i; f_b)^{[0]}$, and (c) the estimated $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.

contents in Fig. 4.6 (b). After the iterative steps of the EM algorithm, there exist only the components from the CS signal as can be seen in Fig. 4.6 (c) associated with Fig. 4.5.

It is well known that in rotating machines various rotations of mechanical components are likely to produce periodic modulations of vibration signals. Besides the transient signal, the

periodically-modulated signals (modulated white noise and modulated narrow-band noise) are also tested so that as good performances are achieved as in Case 1.

In order to demonstrate the potential of the proposed scheme, Case 2 (repetitive transient signal) is generated by means of considering two random variables described as:

$$y(t) = \sum_{j=-\infty}^{+\infty} h(t - jT - \tau_j)A_j + n(t) \quad (4.33)$$

$$H(z) = \frac{b_1}{a_1 + a_2 \cdot z^{-1} + a_3 \cdot z^{-2}} \quad (4.34)$$

where $\tau_j \sim \mathcal{N}(\mu_\tau = 0, \sigma_\tau = 0.05T)$ and $A_j \sim \mathcal{N}(\mu_A = 0, \sigma_A = 0.1)$ account for the uncertainties on the arrival time and on the magnitude of the j^{th} transient, respectively. The white noise $n(t)$ is set to a noise-to-signal-ratio of 6 dB and the signal length is $L = 10^4$ samples. A second-order system is defined by Eq. 4.34, whose numerator and denominator coefficients are $\mathbf{b} = [1]$ and $\mathbf{a} = [1, -2\cos(2\pi f_0)r, r^2]$ with $r = 0.95$, respectively.

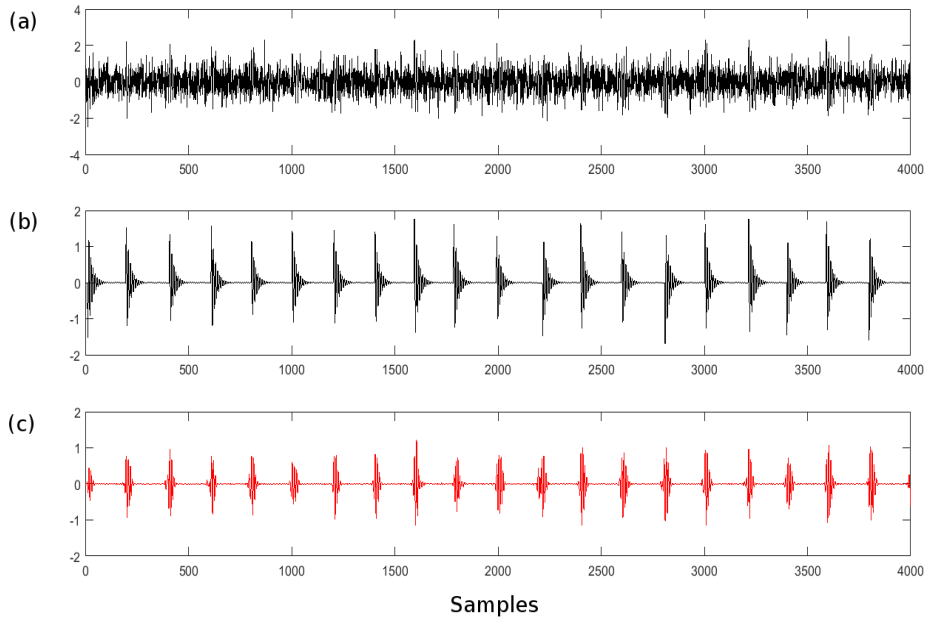


Fig. 4.7 (a) Synthetic signal of Case 2 with white noise (noise-to-signal-ratio= 6 dB). (b) Synthetic repetitive transient signal. (c) Recovered time signal $\hat{x}[n]$.

Fig. 4.7 (a) and (b) show the record of the raw signal and synthetic repetitive transient signal in time. It is highlighted that although these transients are not exactly periodic because the rolling elements experience some random slips, its corresponding parameters $\hat{\sigma}_x^2(i; f_b)^{[0]}$ and $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ both provide good performances as displayed in Fig. 4.8 (a) and (b). In

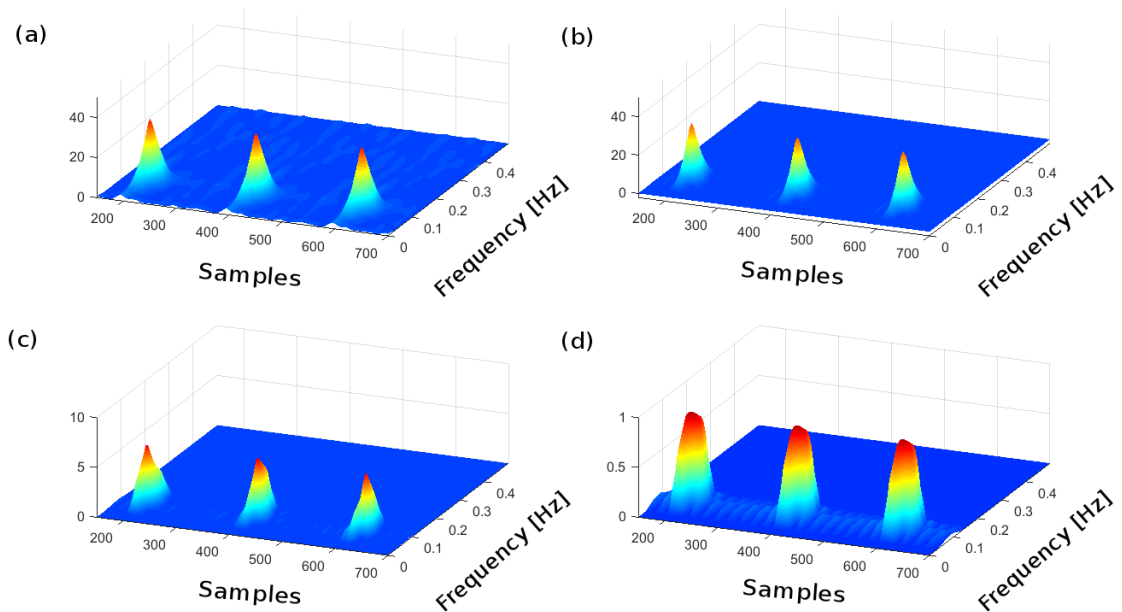


Fig. 4.8 (a) Initialized time-dependent variance of the CS signal, $\hat{\sigma}_x^2(i; f_b)^{[0]}$, and (b) the estimated $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm. (c) Spectrogram of the estimated CS signal, $\hat{X}(i, f_b)$, with (d) its periodic time-varying filter $1/(1 + H(i; f_b))$ as defined in Eqs. 4.15-4.16.

addition, the spectrogram of the estimated CS signal $\hat{X}(i, f_b)$ still reveals distinct spectral contents with its periodic time-varying filter as seen in Fig. 4.8 (c) and (d), which are further verified by the recovered time signal $\hat{x}[n]$ in Fig. 4.7 (c). It is seen from the enlarged view of $\hat{x}[n]$ in Fig. 4.9 (a) that the transients are close to the reference, even though the frequency of these modulations are not integrally related to the mean transient repetition rate⁵. The corresponding parameters $\hat{\sigma}_x^2(i; f_b)^{[0]}$ and $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ are estimated as well as those of Case 1 but considerably wider as marked in Fig. 4.9 (b) and (c).

⁵This is also called a quasi-cyclostationary signal to be more precise.

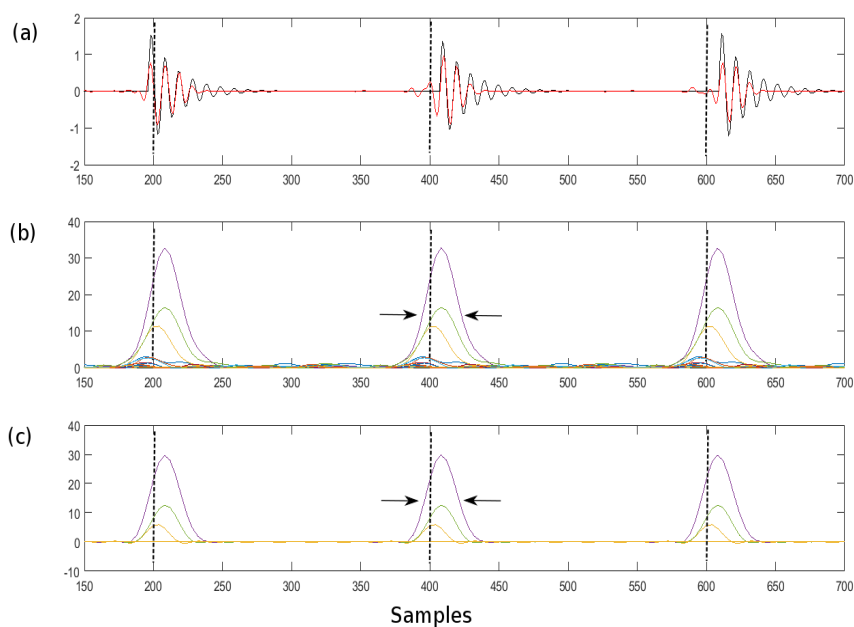


Fig. 4.9 (a) Enlarged view of the recovered time signal $\hat{x}[n]$ of Fig. 4.7 (b) and (c). (b) Initialized time-dependent variance of the CS signal, $\hat{\sigma}_x^2(i; f_b)^{[0]}$, and (c) the estimated $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.

4.4 Validation on experimental examples

This section proposes two experimental examples for a validation of the proposed technique and a visual illustration of the benefits introduced by the periodic-variance based model. The first example is recorded on a test rig designed by company DYNAE, while the second one derives from a dataset offered by the Vibrations and Acoustics Laboratory of the University of New South Wales (Sydney) [76] and addresses the particular case of two sources with bearing and gear signals. It is important to notice that all validations are based on the prior knowledge of the characteristic frequencies as obtained from the Spectral Correlation (e.g. the Cyclic Modulation Spectrum or the Fast Spectral Correlation [79]).

4.4.1 Example 1: extraction of bearing signals with good gears

Signals recorded on a test rig are now considered. The test rig shown in Fig. 4.10 has been designed by company DYNAE. It mainly consists of an electric asynchronous motor, a rotary encoder, 4 accelerometer sensors, a speed variator, a driving gear with 45 teeth, four bearings (3 healthy and 1 outer race fault) and two pinions (healthy and broken). In addition, the first experiment involves almost new gears; therefore it is only expected to recover the CS2 components related to the bearings.

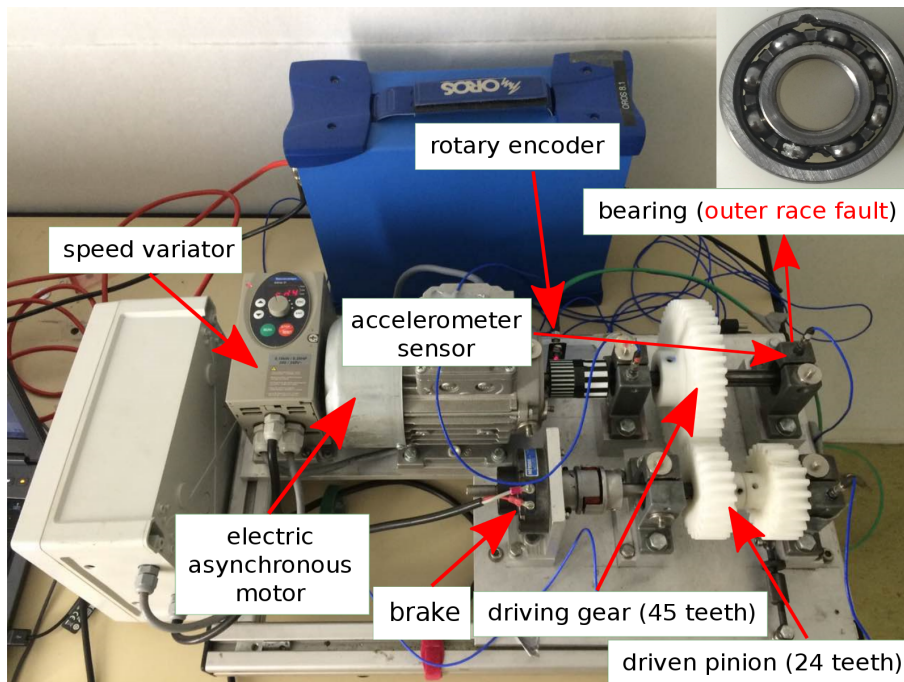


Fig. 4.10 Test rig setup of Example 1.

The test rig has been used to produce three types of machine condition: Case 3 relates to very weak bearing signatures (or to very early stage of a fault); Case 4 corresponds to an outer race fault of the bearing and Case 5 to the effect of “load”. The parameter settings used in the experiment are listed in Table 4.1.

Table 4.1 Parameter settings in Case 3, Case 4 and Case 5.

	Case 3	Case 4	Case 5
Sampling frequency F_s (Hz)	51200		
Duration (s)	10		
N_w	2^6		
R	8		
Main shaft rotation frequency — $f_{rot,1}$ (Hz)	15.5-16.5	15.4-16.3	16.8-18.6
Secondary shaft rotation frequency — $f_{rot,2}$ (Hz)	$1.875 \times f_{rot,1}$ (29.1-30.9)	$1.875 \times f_{rot,1}$ (28.9-30.6)	

Case 3: healthy bearings

To demonstrate the effectiveness of the proposed scheme on real data, the first test is executed on almost new bearings, in normal condition, which perhaps corresponds to some very incipient signatures. The hidden period $T = 1/f_{rot,1} = 1/16$ s is found from the corresponding Spectral Coherence and Enhanced Envelope Spectrum (EES)⁶ in Fig. 4.11. It indicates the harmonic structure of the main shaft speed with sidebands at an unknown frequency of 2 Hz. In addition, they spread the full-band in $[0; 25.6]$ kHz.

Inspection of the vibration signal in Case 3 shows that it undergoes a strong amplitude modulation located at 2 Hz as indicated in Fig. 4.12 (a). This effect is more distinct in

⁶In this work, the cyclic spectral analysis is executed by the Fast-SC which benefits from increased computational efficiency while still maintaining a very fine carrier frequency resolution [79]. In addition, the Spectral Coherence and the Enhanced Envelope Spectrum are displayed so as to directly assess the “depth” of a modulation with frequency α and carrier f and better enhance non-zero cyclic components than the Squared Envelope Spectrum.

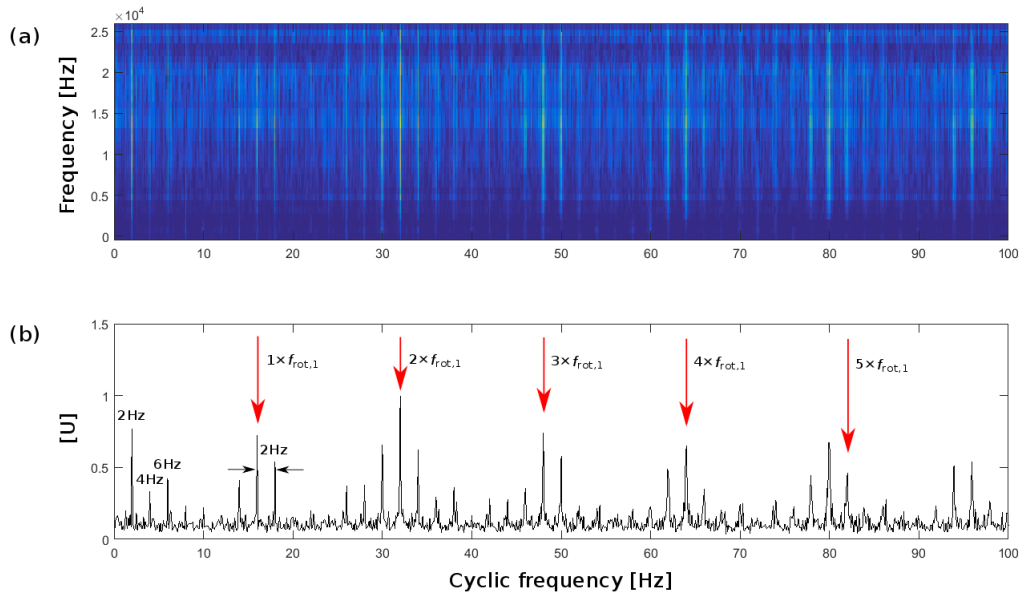


Fig. 4.11 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of signal of Case 3 ($\Delta f = 800$ Hz, $\Delta \alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in full band $[0; 25.6]$ kHz.

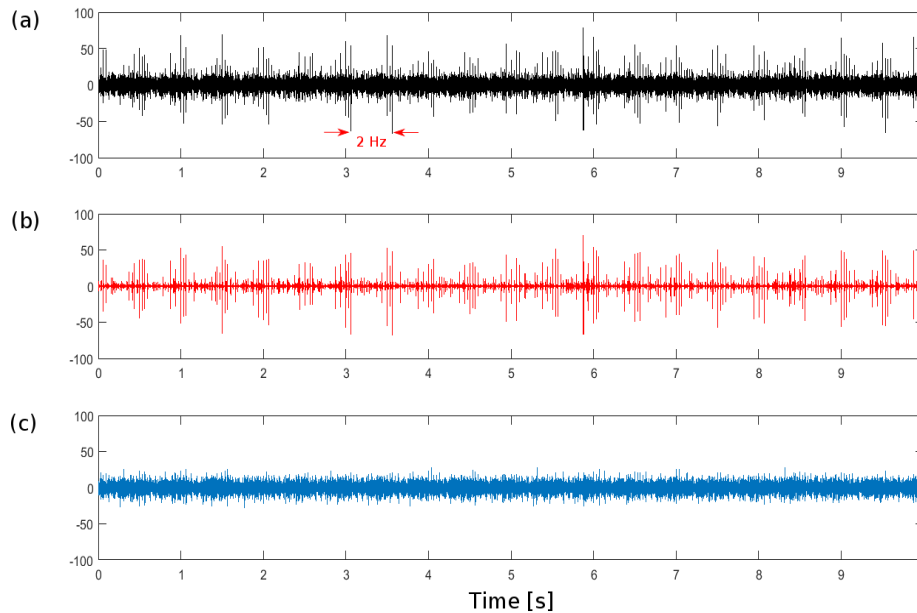


Fig. 4.12 (a) Vibration signal of Case 3 divided into (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = f_{rot,1} = 16$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).

the recovered time signal $\hat{x}[n]$, which is characterized by $1/T = f_{rot,1} = 16$ Hz, than that in the noise (residual) signal displayed in Fig. 4.12 (b) and Fig. 4.12 (c). Furthermore, one can read more details from their spectrograms in Fig. 4.13, i.e. $|Y(i, f_b)|$, $|\hat{X}(i, f_b)|$, $|Y(i, f_b) - \hat{X}(i, f_b)|$, which reflect the energy distributions in the time-frequency plane. The spectrum (energy distribution) of the CS signal is obtained by integrating its variance along the time instants as displayed in black dash-dot line in Fig. 4.13 (c). Compared with the spectra of the noise signal (black dash-dot line in Fig. 4.13 (b)), one can see that the CS signal dominates the frequency domain in $[3; 11]$ kHz and the noise signal in $[0; 1]$ kHz. Fig. 4.13 (d), the periodic time-dependent variance, also provides the evidence of energy distributions characterized by $1/T = f_{rot,1} = 16$ Hz.

Of particular interest, from the closer inspection of Fig. 4.12, is the repetitive transients (relatively sharp damping) in Fig. 4.14 (b) that are characterized by twice of the main shaft speed. Even though the CS component has been removed, the noise (residual) signal is still characterized by the main shaft speed with the carrier frequency of 1.11 kHz in Fig. 4.14 (c).

Finally, the squared envelope spectra (SES) are shown in Fig. 4.15. They evidence two points: 1) the CS signal is successfully extracted by means of the proposed model (see Fig. 4.11 (b) and Fig. 4.15 (b)); 2) the noise (residual) signal is mainly characterized by the unknown frequency of 2 Hz.

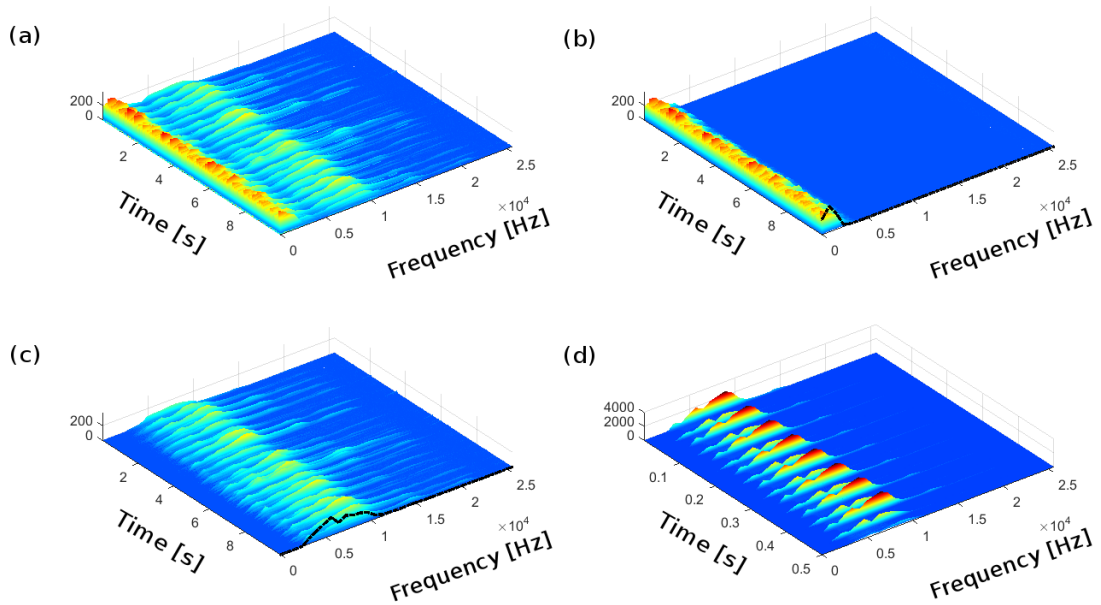


Fig. 4.13 (a) Spectrogram (magnitude of the STFT) of the raw signal, $|Y(i, f_b)|$. (b) Spectrogram of the estimated noise signal, $|Y(i, f_b) - \widehat{X}(i, f_b)|$, and (c) that of the estimated CS signal $|\widehat{X}(i, f_b)|$ with their spectra ($\hat{\sigma}_n^2(f_b)^{[k+1]}$ and $\hat{\sigma}_x^2(f_b)^{[k+1]}$) in black dash-dot lines. (d) The estimated time-dependent variance $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.

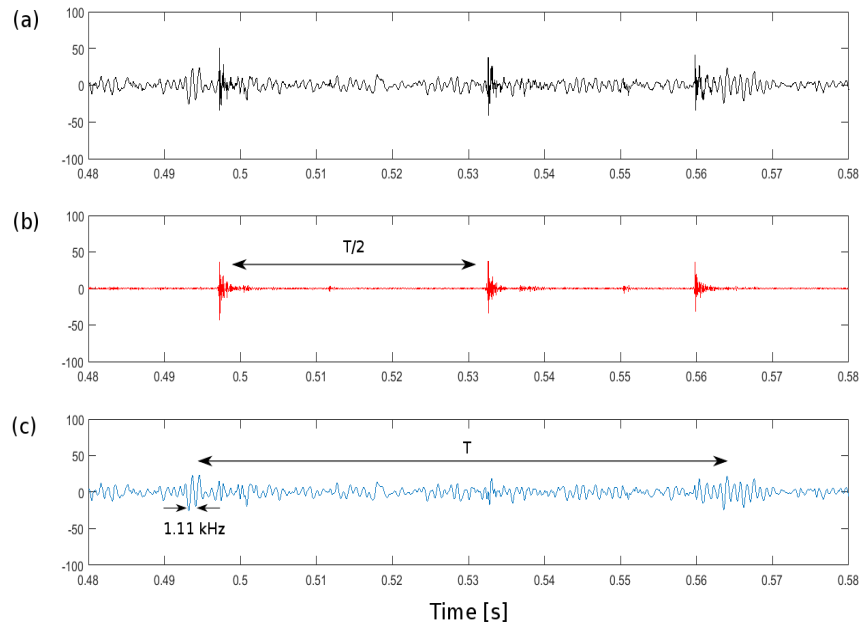


Fig. 4.14 Enlarged view of (a) the vibration signal of Case 3, (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = f_{rot,1} = 16$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).

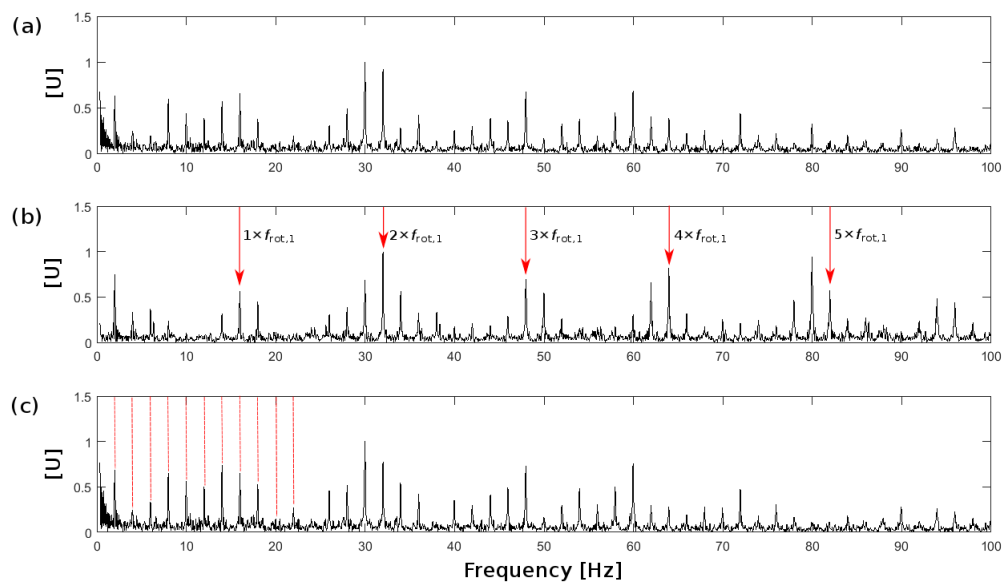


Fig. 4.15 Squared Envelope Spectrum of (a) the vibration signal of Case 3: $S_y^{SES}(\alpha)$, (b) the recovered time signal: $S_{\hat{x}}^{SES}(\alpha)$ and (c) the noise (residual) signal: $S_{\hat{n}}^{SES}(\alpha)$.

Case 4: an outer race fault

In the following case, a defect (indentation) is introduced on the outer race of one of the bearing as installed in Fig. 4.10. First, the Spectral Correlation of Case 4 is calculated and shows a clear presence of BPFO in Fig. 4.16. In addition, it spreads almost the full-band and dominates the high frequency in [7.6; 25.6] kHz. Therefore, the hidden period $T = 1/BPFO = 1/48$ s is selected.

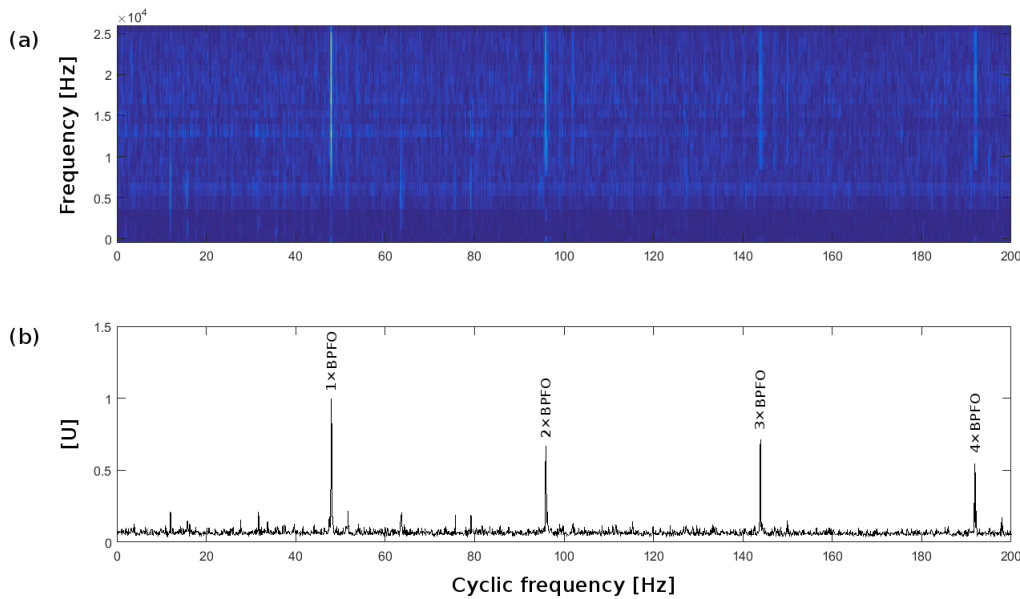


Fig. 4.16 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of the signal of Case 4 ($\Delta f = 800$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in full band [0; 25.6] kHz.

From careful inspection of the time record, one can see the weak amplitude modulation of 2 Hz due to the heavy shocks in Fig. 4.17. In order to display more details, their spectrograms are shown in Fig. 4.18, which demonstrates the energy distributions in the time-frequency domain. It is noted that the CS component (characterized by $1/T = BPFO = 48$ Hz) dominates the high frequency above 7.6 kHz, even if there exists a strong noise (residual) component in the low frequency. More evidence is provided by the further analysis in Fig. 4.19 and Fig. 4.20. Hereby, one conclusion is first drawn that the CS signal (characterized by $1/T = BPFO = 48$ Hz) is recovered with its full-band spectral content (see Fig. 4.19 (b) and 4.20 (b)).

Based on these analyses, one can draw a preliminary conclusion that the tested rig undergoes an amplitude modulation (unknown frequency at 2 Hz) and contains a CS component characterized by $1/T = f_{rot,1} = 16$ Hz. In addition, such a CS signal has a wide spectral

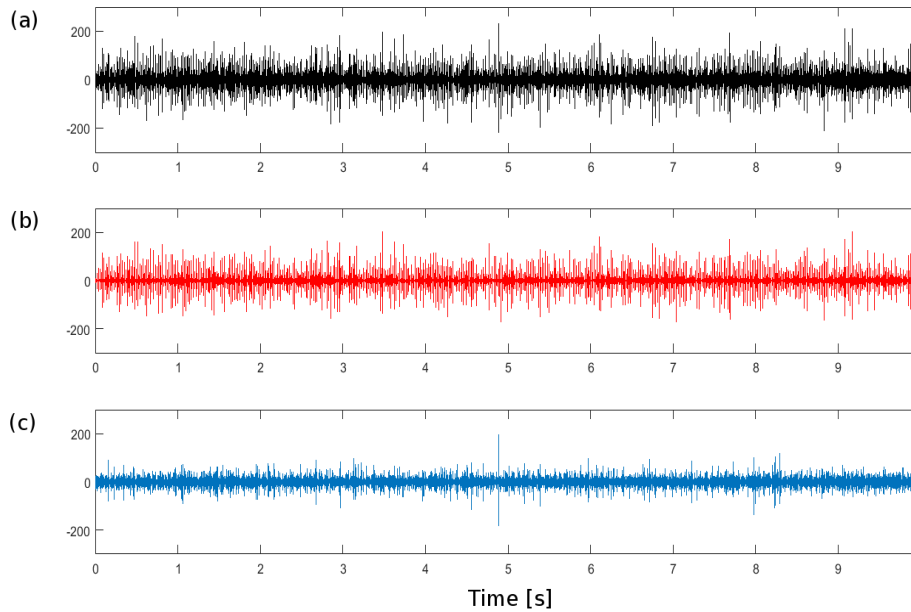


Fig. 4.17 (a) Vibration signal of Case 4 divided into (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = BPF0 = 48$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).

content (relatively fast damping in time) which is characterized by the main shaft speed with enhancement at even harmonics. This effect probably corresponds to the misalignment of the bearing⁷, mechanical looseness, etc. Another suspected source, particularly in complex machines, is the torque loads applied by the brake as indicated in Fig. 4.10. Since the two parallel shafts are connected by the driving gear and the driven pinion, the torque load is converted to the radial contact (gear and pinion). And the last but not least possibility is that the resulting harmonics of rotation speed may be generated by tooth faults at a very early stage (wear, pitting, etc.). To find the answer, the following test is designed.

⁷Its vibration symptoms is similar to angular misalignment, i.e. it attempts to realign coupling or to balance the shaft by which the problem will be alleviated. Such self-regulation will cause a twisting motion, with approximate 180 degree phase, shifts from side to side or top to bottom.

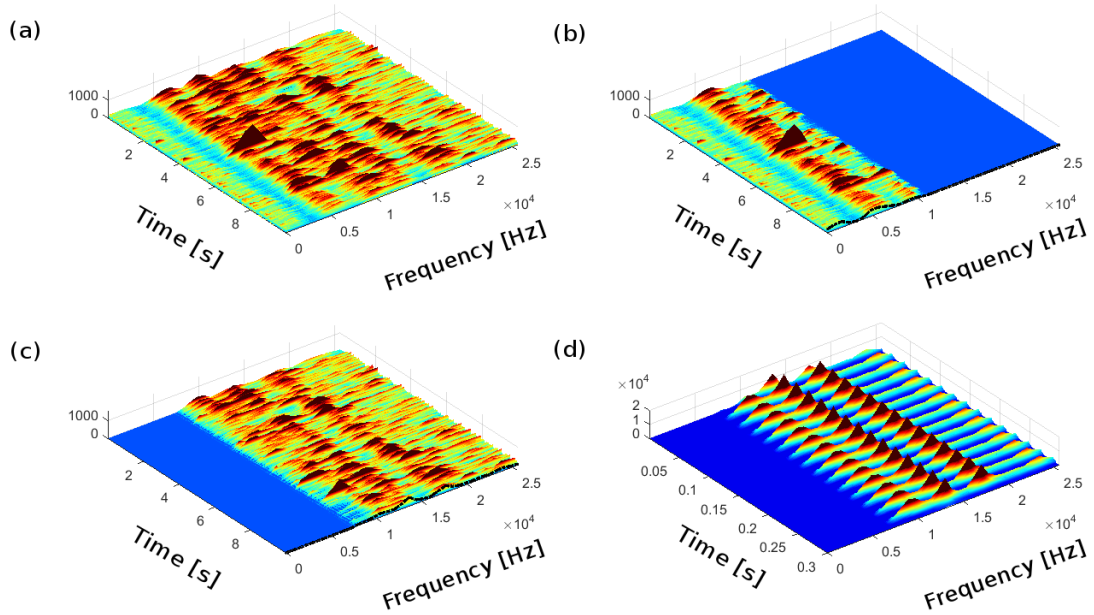


Fig. 4.18 (a) Spectrogram (magnitude of the STFT) of the raw signal, $|Y(i, f_b)|$. (b) Spectrogram of the estimated noise signal, $|Y(i, f_b) - \widehat{X}(i, f_b)|$, and (c) that of the estimated CS signal, $|\widehat{X}(i, f_b)|$, with their spectra ($\hat{\sigma}_n^2(f_b)^{[k+1]}$ and $\hat{\sigma}_x^2(f_b)^{[k+1]}$) in black dash-dot lines. (d) The estimated time-dependent variance $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.

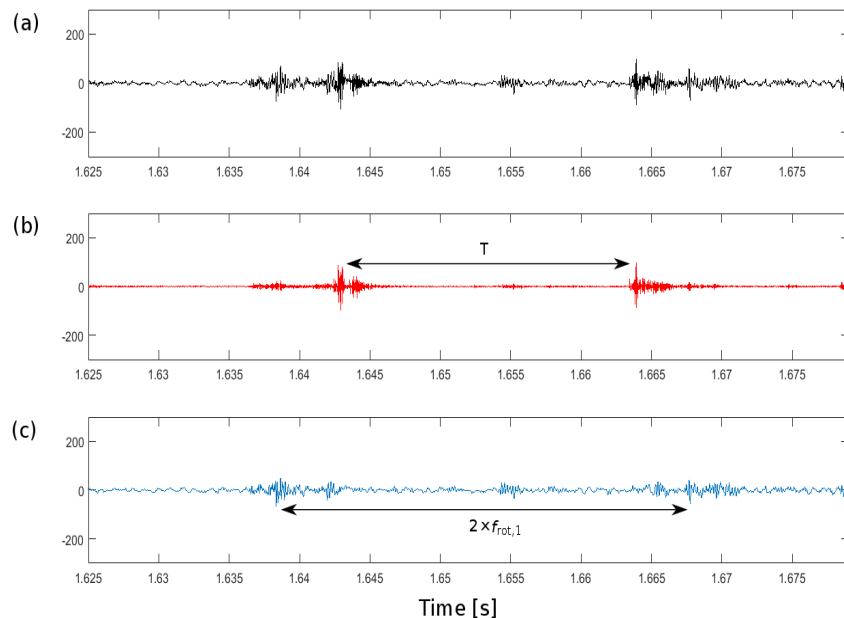


Fig. 4.19 Enlarged view of (a) the vibration signal of Case 4, (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = BPFO = 48$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).

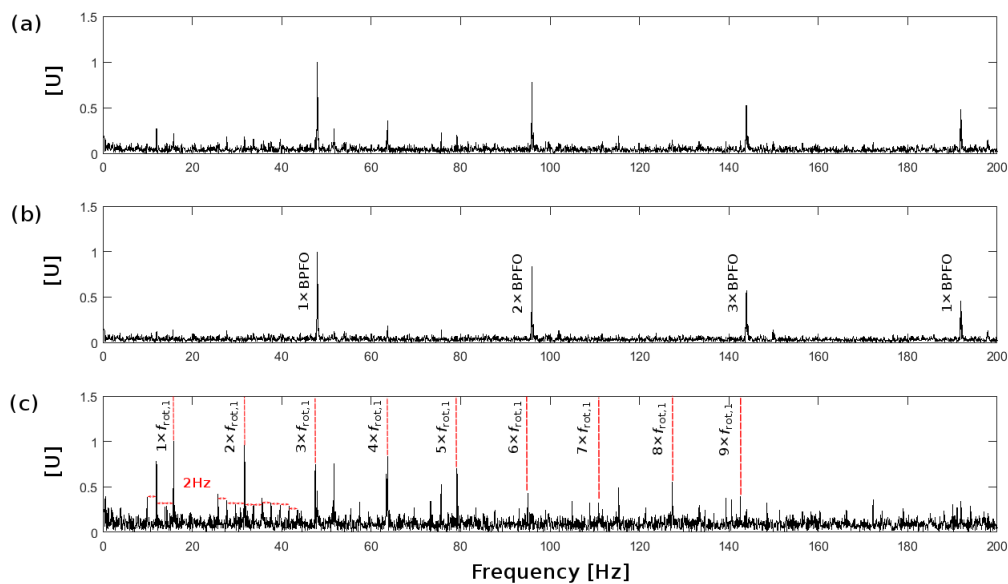


Fig. 4.20 Squared Envelope Spectrum of (a) the vibration signal of Case 4: $S_y^{SES}(\alpha)$, (b) the recovered time signal: $S_x^{SES}(\alpha)$ and (c) the noise (residual) signal: $S_n^{SES}(\alpha)$.

Case 5: an outer race fault under no load condition

In this case, the test is executed in the condition of no load to demonstrate the robustness of the proposed scheme. To achieve zero-load, the driven pinion is decoupled with the driving gear. It is noteworthy that without the connection from the secondary shaft there is neither the dynamic load (torque force) nor the effect from the acting force. Different from previous cases, there is only (in theory) static gravitational load from the weight of the shaft and related components. Hence it rotates relatively faster as seen in Table 4.1 and the no force condition brings a relatively light damping (narrow spectral content) while undergoing the damage of bearing. Compared with Case 4, it presents very distinct harmonics of the main shaft speed (relatively strong), whereas the harmonics of the BPFO is relatively weak in Fig. 4.21. Through the Enhanced Envelope Spectrum in Fig. 4.21 (b) and Fig. 4.21 (c), one can fix them in the band [7.6; 25.6] kHz and [3.6; 7.6] kHz, respectively.

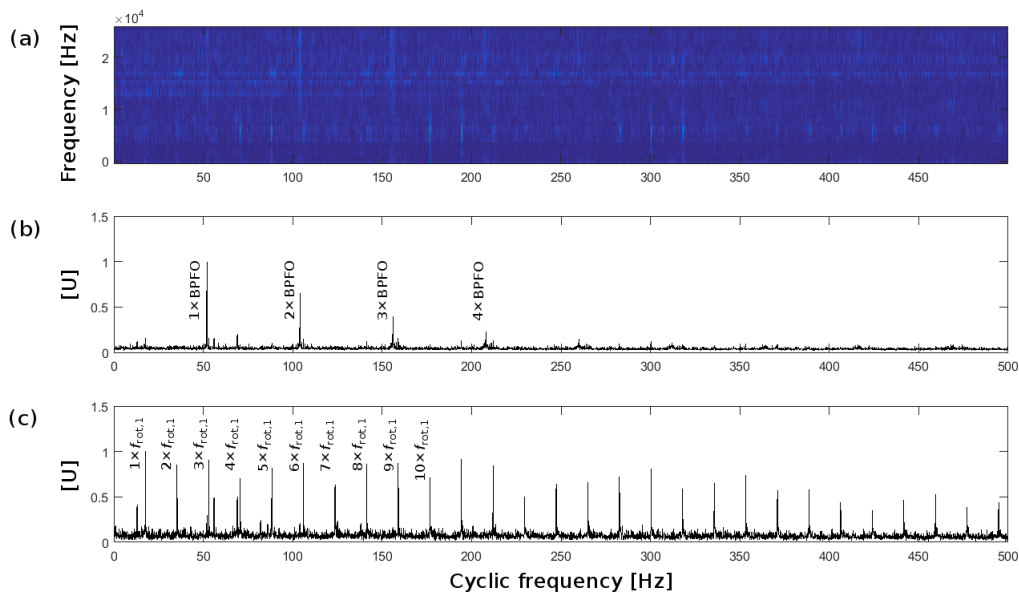


Fig. 4.21 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of the signal of Case 5 ($\Delta f = 800$ Hz, $\Delta\alpha = 0.1$ Hz). Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ (b) in the band [7.6; 25.6] kHz and (c) in the band [3.6; 7.6] kHz.

In the record of Case 5, one can see heavy shocks with wide spectral contents by inspecting Fig. 4.22 (a). Then the raw signal is further separated into the CS component (characterized by $1/T = BPFO = 52$ Hz) and the noise (residual) component as displayed in Fig. 4.22 (b) and Fig. 4.22 (c). It is seen that they are characterized by different damping ratio, light for $\hat{x}[n]$ and heavy for $\hat{n}[n]$, which can also be verified by their spectrograms in Fig. 4.23. By inspecting the recovered time signal, Fig. 4.24 (b) shows more details on the

transients with a random period⁸. For the noise (residual) part, the transients (characterized by $f_{rot,1} = 16$ Hz) are indicated in 4.24 (c).

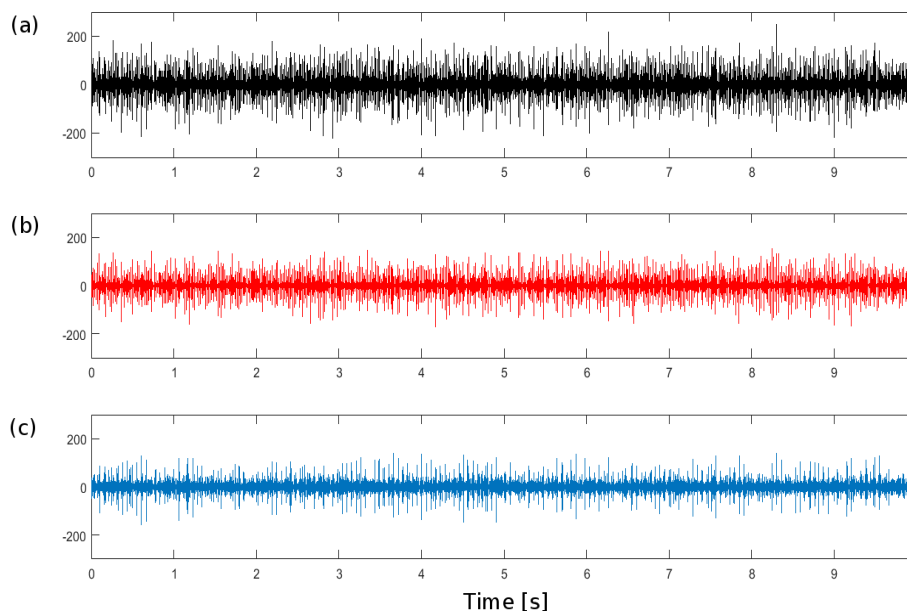


Fig. 4.22 (a) Vibration signal of Case 5 divided into (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = BPF0 = 52$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).

Finally, their squared envelope spectrum is displayed in Fig. 4.25. One can see the CS component (characterized by $1/T = BPF0 = 52$ Hz) shows the same harmonics as in Fig. 4.21 (b), but it is reconstructed in full-band content as seen in Fig. 4.23 (d). It also indicates that the amplitude modulation (suspected frequency at 2 Hz) disappears completely (see Fig. 4.22 (b)) which comes from the connection with the driven pinion. For the CS component (characterized by $1/T = f_{rot,1} = 16$ Hz), it derives from the installation problem of the driving gear, i.e. misalignment of bearing.

⁸This means that the repetition of these shocks is not integrally related to the hidden period T , i.e. it should be modelled as variable [6]. Especially in Case 5, there is no acting force to reduce variations of speed while rotating relatively faster than Case 4. This is the reason why this case presents relatively weak harmonics of the BPF0.

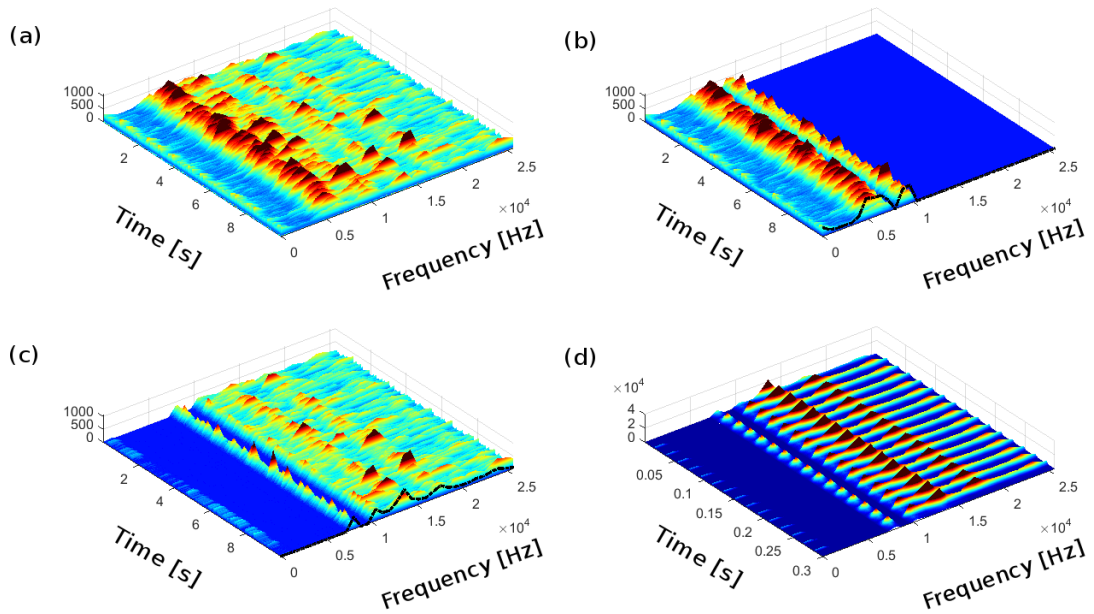


Fig. 4.23 (a) Spectrogram (magnitude of the STFT) of the raw signal, $|Y(i, f_b)|$. (b) Spectrogram of the estimated noise signal, $|Y(i, f_b) - \hat{X}(i, f_b)|$, and (c) that of the estimated CS signal, $|\hat{X}(i, f_b)|$, with their spectra ($\hat{\sigma}_n^2(f_b)^{[k+1]}$ and $\hat{\sigma}_x^2(f_b)^{[k+1]}$) in black dash-dot lines. (d) The estimated time-dependent variance $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.

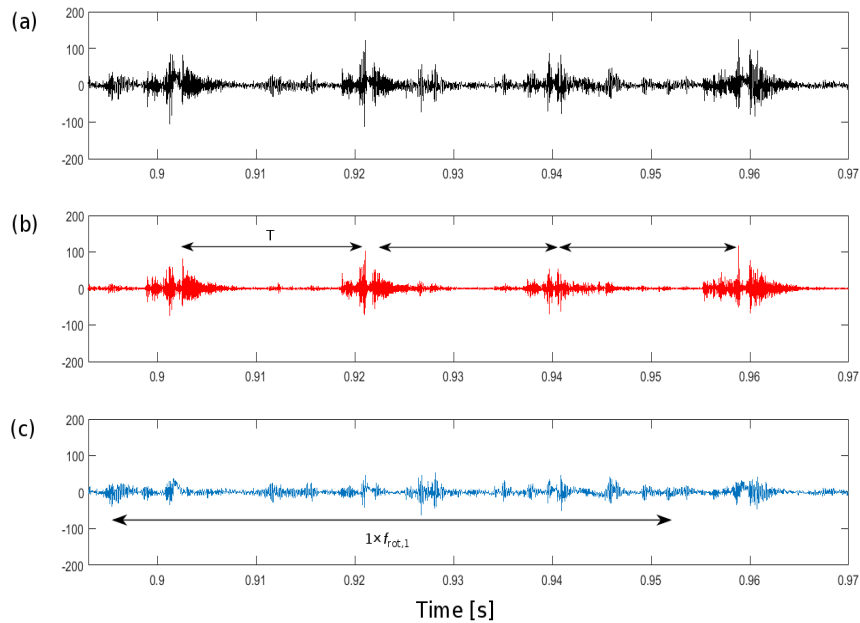


Fig. 4.24 Enlarged view of (a) the vibration signal of Case 5, (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = BPFO = 52$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).

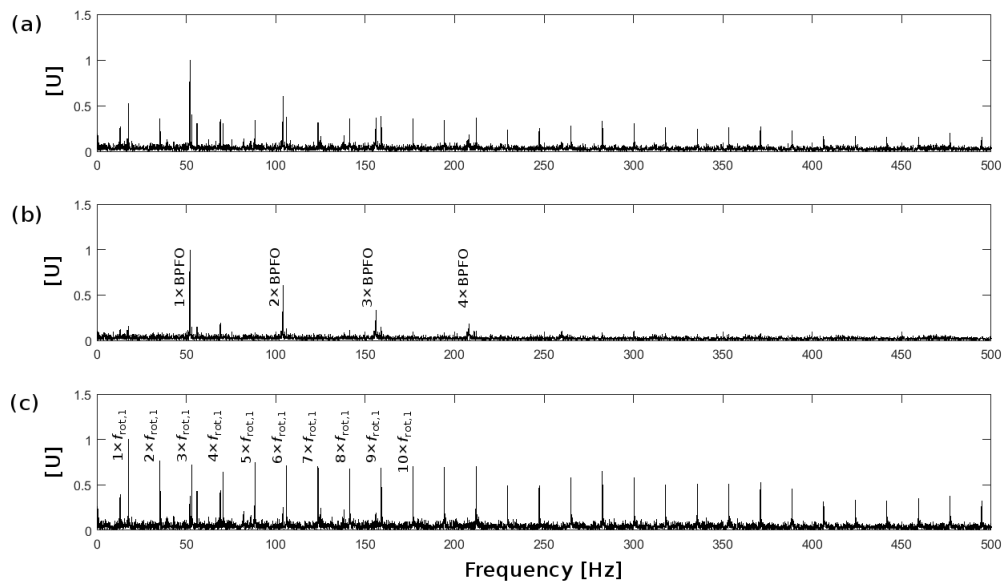


Fig. 4.25 Squared Envelope Spectrum of (a) the vibration signal of Case 5: $S_y^{SES}(\alpha)$, (b) the recovered time signal: $S_x^{SES}(\alpha)$ and (c) the noise (residual) signal: $S_n^{SES}(\alpha)$.

4.4.2 Example 2: separation of bearing signals with broken gears

The test-rig is a one-stage gearbox with primary and secondary shafts supported by ball bearings. In addition, this experiment corresponds to broken gears; it is therefore expected to recover the CS2 components related to the bearings in the presence of strong interfering components. The parameter settings used in the experiment are listed in Table 4.2 for Case 6 and Case 7.

Table 4.2 Parameter settings used in Case 6 and Case 7.

	Case 6	Case 7
Sampling frequency F_s (kHz)	48	
Duration (s)	1.365	
N_w	2^6	
R	8	
Rotation frequency — f_{rot} (Hz)	10	
Ballpass frequency, outer race – BPFO (Hz)	48.9	
Ballpass frequency, inner race – BPFI (Hz)	71.1	
Fundamental train frequency — FTF (Hz)	4.1	
Ball (roller) spin frequency — BSF (Hz)	26.1	

Case 6: healthy bearings

To demonstrate the robustness of the proposed scheme on real data, the first test is executed on almost new bearings. Therefore no CS2 components related to the damage are expected in Case 6. Since it is indeed difficult to identify the damage at incipient stage, particularly in the case (gears and bearings), this test is designed to detect and extract very weak bearing signatures at the possible characteristic frequencies listed in Table 4.2.

The time record is shown in Fig. 4.26 (a), which is divided into the CS component (characterized by one possible fault frequency $1/T = BPFO = 48.9$ Hz) and the noise

(residual) component in Fig. 4.26 (b) and Fig. 4.26 (c), respectively. It is noted that the recovered time signal is always null for all the characteristic frequencies, as can be seen in Fig. 4.27.

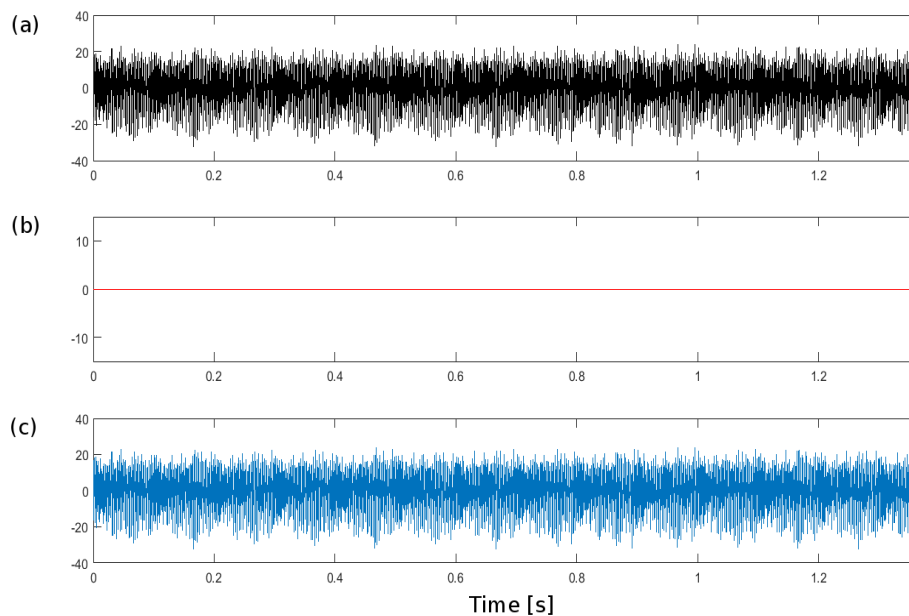


Fig. 4.26 (a) Vibration signal of Case 6 divided into (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = BPF0 = 48.9$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).

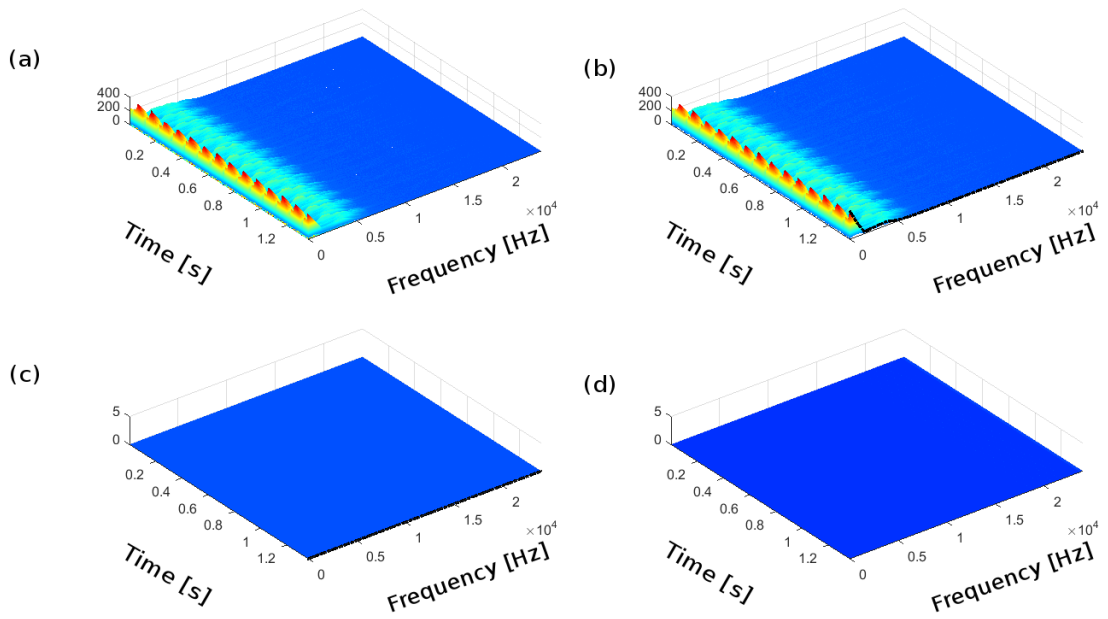


Fig. 4.27 (a) Spectrogram (magnitude of the STFT) of the raw signal, $|Y(i, f_b)|$. (b) Spectrogram of the estimated noise signal, $|Y(i, f_b) - \hat{X}(i, f_b)|$, and (c) that of the estimated CS signal, $|\hat{X}(i, f_b)|$, with their spectra $\hat{\sigma}_n^2(f_b)^{[k+1]}$ and $\hat{\sigma}_x^2(f_b)^{[k+1]}$ in black dash-dot lines. (d) The estimated time-dependent variance $\hat{\sigma}_x^2(i, f_b)^{[k+1]}$ from the EM algorithm.

Case 7: an incipient ball fault

The second test is executed on the same test-rig with three typical types of fault (i.e. inner race, outer race and ball fault) to demonstrate the effectiveness of the proposed scheme. Case 7 analyzes an incipient damage of bearings in the presence of strong gear vibrations and extracts a CS signal related to a ball defect.

Fig. 4.28 shows the spectrogram (logarithmic scale) of the raw signal with evidence of periodic energy flow ($1/T = 2 \times BSF$, dashed line in red). Its time record is displayed in Fig. 4.29 (a) which is then divided into the recovered time signal $\hat{x}[t]$ characterized by $1/T = 2 \times BSF = 51.29$ Hz and the noise (residual) part in Fig. 4.29 (b) and Fig. 4.29 (c). It is seen that the CS component is heavily smeared in the low frequency, especially below 5 kHz, as demonstrated in Fig. 4.30. By zooming its time record, more details can be seen in Fig. 4.31. Of particular interest is the distinct amplitude modulation in the noise (residual) component which is carried by the low frequency (below 5 kHz). Then it is verified to be the gear mesh frequency ($f_g = f_{rot} \times \text{No. of teeth} = 10 \times 32 = 320$ Hz) which originates from the gearbox. Due to such effect in relatively heavy energy, the classic envelope analysis (SES) of the raw signal presents only the fault signatures of the gearbox, whereas the recovered signal purely reveals the characteristic frequency of the ball fault in Fig. 4.32.

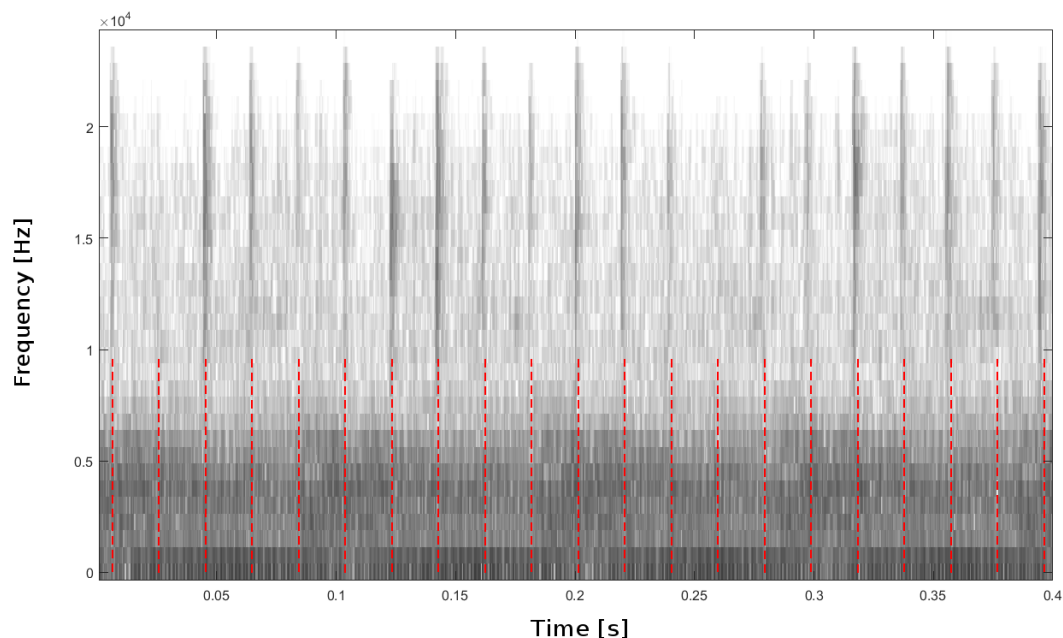


Fig. 4.28 Spectrogram (logarithmic scale) of the signal of Case 7 (frequency resolution $\Delta f = 750$ Hz) with evidence of spectral content in band $[9.75; 23.25]$ kHz.

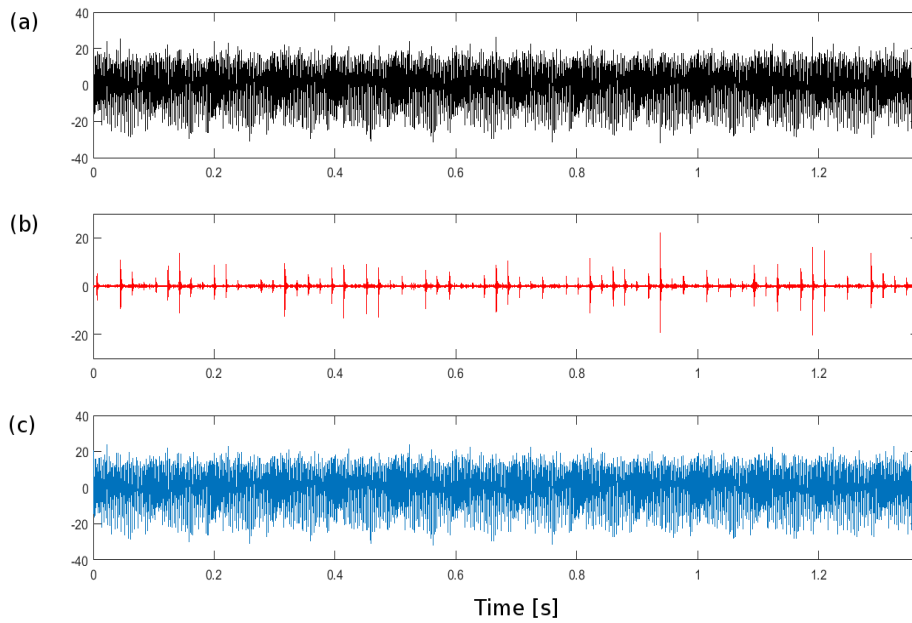


Fig. 4.29 (a) Vibration signal of Case 7 divided into (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = 2 \times BSF = 51.29$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).

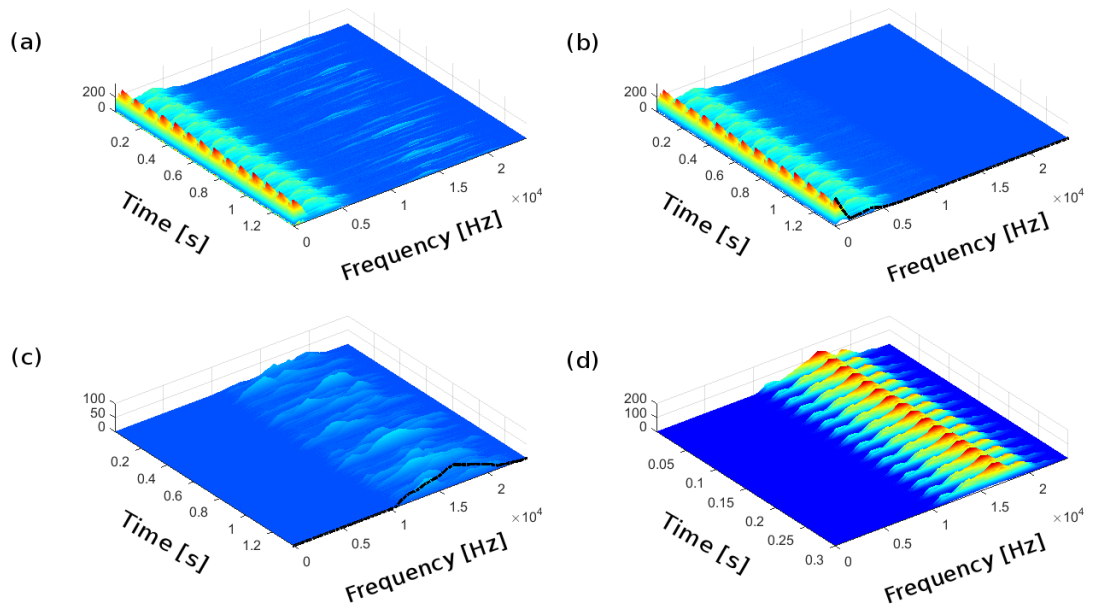


Fig. 4.30 (a) Spectrogram (magnitude of the STFT) of the raw signal, $|Y(i, f_b)|$. (b) Spectrogram of the estimated noise signal, $|Y(i, f_b) - \hat{X}(i, f_b)|$, and (c) that of the estimated CS signal, $|\hat{X}(i, f_b)|$, with their spectra $(\hat{\sigma}_n^2(f_b))^{[k+1]}$ and $(\hat{\sigma}_x^2(f_b))^{[k+1]}$ in black dash-dot lines. (d) The estimated time-dependent variance $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ from the EM algorithm.

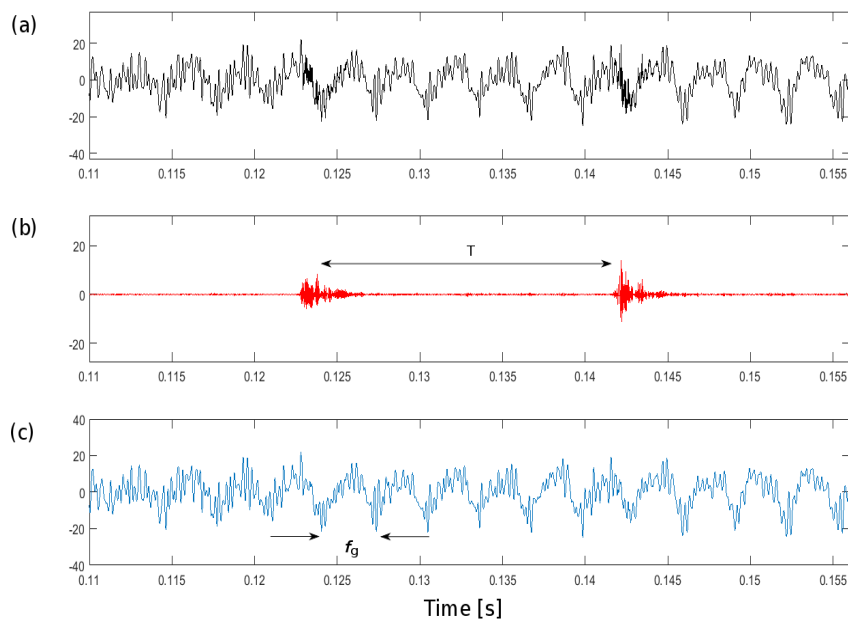


Fig. 4.31 Enlarged view of (a) the vibration signal of Case 7, (b) the recovered time signal $\hat{x}[n]$ characterized by $1/T = 2 \times BSF = 51.29$ Hz and (c) the noise (residual) signal ($= y[n] - \hat{x}[n]$).

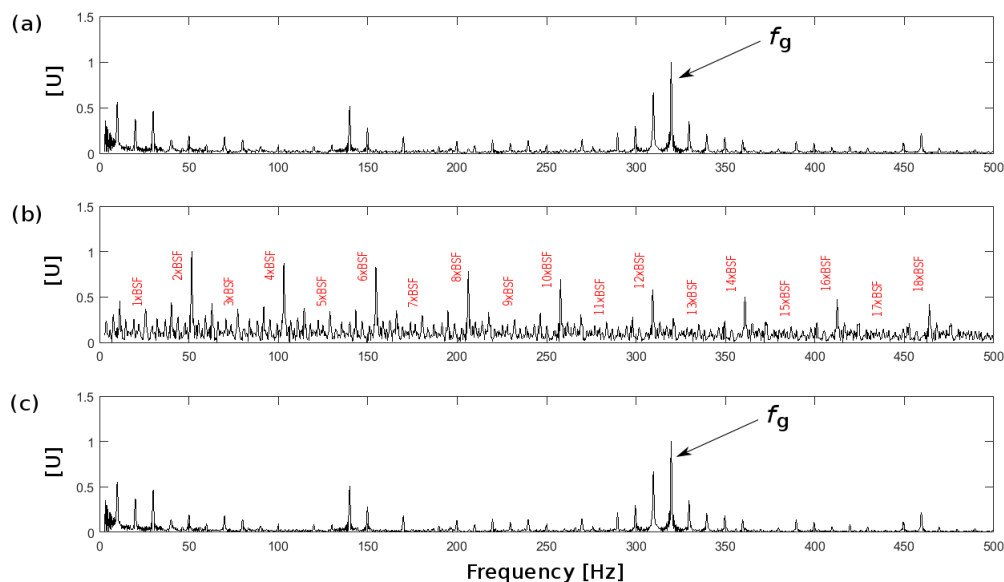


Fig. 4.32 Squared Envelope Spectrum of (a) the vibration signal of Case 7: $S_y^{SES}(\alpha)$, (b) the recovered time signal: $S_x^{SES}(\alpha)$ and (c) the noise (residual) signal: $S_n^{SES}(\alpha)$.

Cases 8: separation of bearing and gear signals

Although gearboxes⁹ are often common and critical equipment of machines (e.g. milling machines, wind turbines, automobiles, etc.), their analysis in condition monitoring is still complex and complicated, especially for industrial applications. This is because there can be many possible sources of vibration signals, e.g. bearings and gears meshing, and some of them are related to the fault signatures. In order to identify them at an early stage, it is essential to first separate them according to their relevant characteristic frequency. Hereafter, Case 8 provides an example that separately extracts two CS signals corresponding to bearings and gears, respectively. The parameter settings used in Case 8 are listed in Table 4.3.

Table 4.3 Parameter settings used in Case 8.

Sampling frequency F_s (kHz)	48
Duration (s)	2.082
N_w	2^7
R	16
Rotation frequency — f_{rot} (Hz)	6
Ballpass frequency, inner race – BPF1 (Hz)	42.77
Gear mesh frequency – f_g (Hz)	192

From the time record, one can first identify the bearing characteristic frequency $1/T_1 = BPF1 = 42.77$ Hz. The corresponding reconstruction of transients is displayed in Fig. 4.33 (b). Next, the gear signal characterized by $1/T_2 = f_g = f_{rot} \times \text{No. of teeth} = 192.2$ Hz is shown in Fig. 4.33 (c).

Furthermore, Fig. 4.34 shows the illustration of extracting two sources (characterized by different characteristic signatures) in the time-frequency plane. As it is highlighted above, the extraction of the CS signals is achieved by the full-band filter associated with the CS indexes of two sources shown in Fig. 4.35.

⁹It is an assembly of gears, shafts, bearings, keys, bearing covers, oil rings/oil seals and a casing to house all these machines elements.

Finally, an enlarged view of Fig. 4.33 and the envelope analysis (SES) are displayed in Fig. 4.36 and Fig. 4.37, respectively. Both of them prove a good performance in extracting the CS signals.

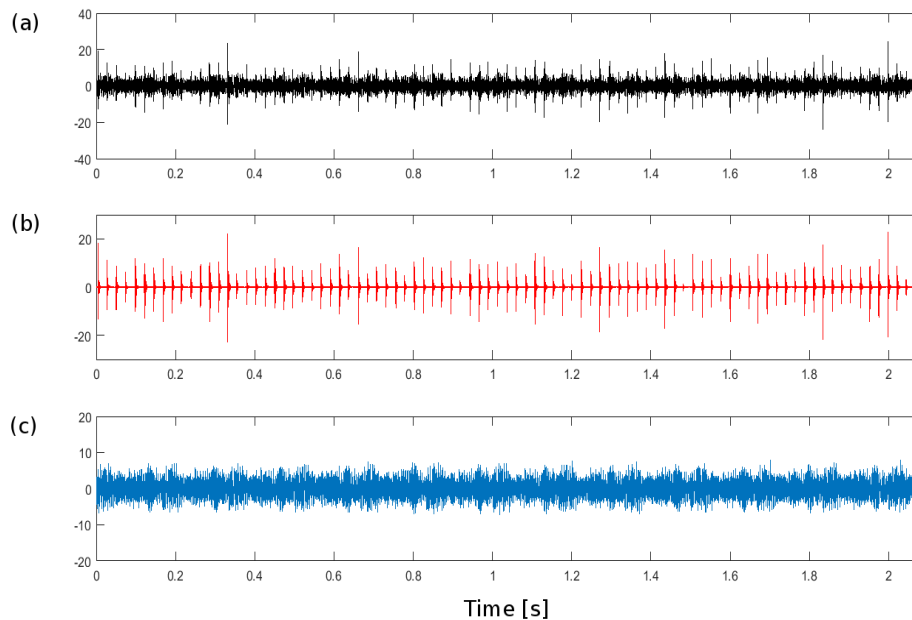


Fig. 4.33 (a) Vibration signal of Case 8 divided into (b) the recovered time signal of source 1: $\hat{x}^1[n]$ characterized by $1/T_1 = BPF1 = 42.77$ Hz and (c) that of source 2: $\hat{x}^2[n]$ characterized by $1/T_2 = f_g = f_{rot} \times \text{No. of teeth} = 192.2$ Hz.

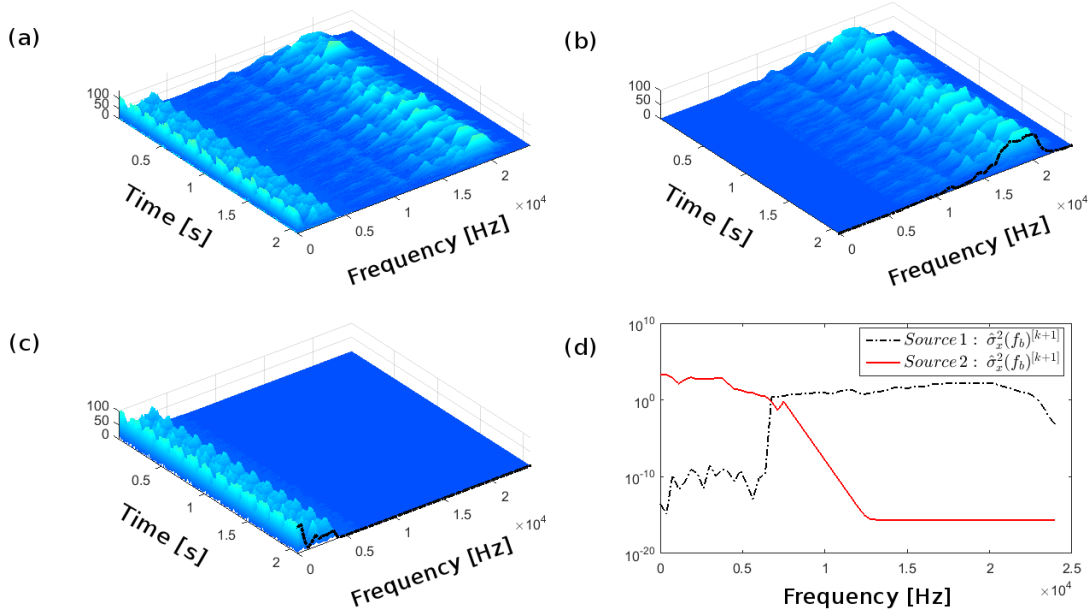


Fig. 4.34 (a) Spectrogram (magnitude of the STFT) of the raw signal, $|Y(i, f_b)|$. (b) Spectrogram of the estimated CS signal from source 1: $|\hat{X}^1(i, f_b)|$ and (c) that from source 2: $|\hat{X}^2(i, f_b)|$ with their related spectra (variance of the CS signal $\hat{\sigma}_x^2(f_b)^{[k+1]}$) in black dash-dot lines. (d) The estimated spectra from the EM algorithm (variance of the CS signal $\hat{\sigma}_x^2(f_b)^{[k+1]}$).

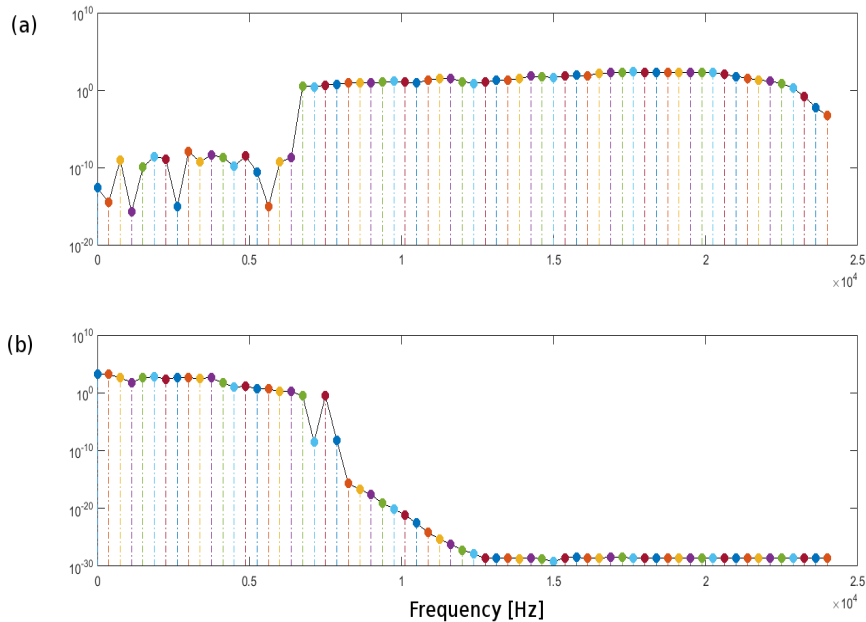


Fig. 4.35 Cyclostationary indexes of (a) source 1 and (b) source 2: function of the frequency f_b by means of taking the standard deviation of $\hat{\sigma}_x^2(i; f_b)^{[k+1]}$ over the time instant i .

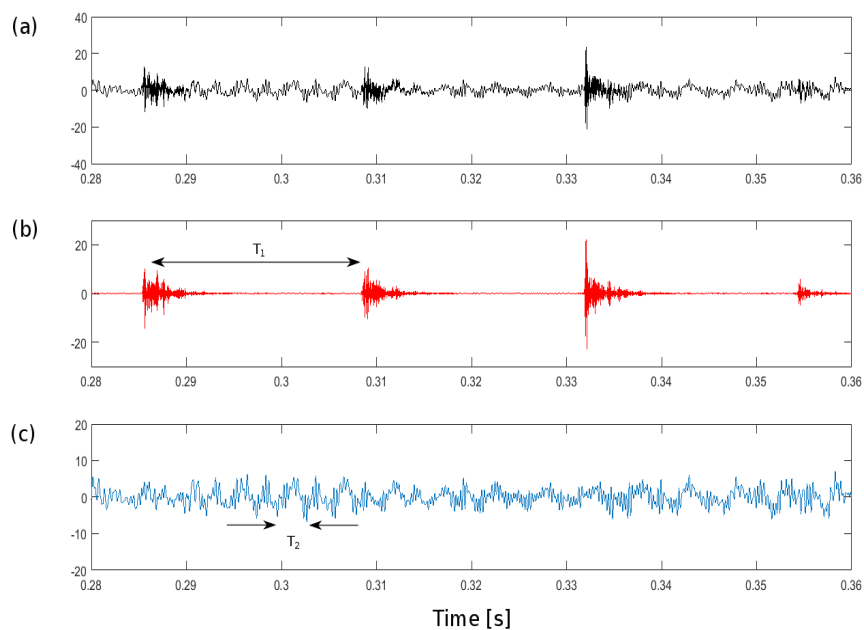


Fig. 4.36 Enlarged view of (a) the vibration signal of Case 8, (b) the recovered time signal of source 1: $\hat{x}^1[n]$ characterized by $1/T_1 = BPF1 = 42.77$ Hz and (c) that of source 2: $\hat{x}^2[n]$ characterized by $1/T_2 = f_g = f_{rot} \times \text{No. of teeth} = 192.2$ Hz.

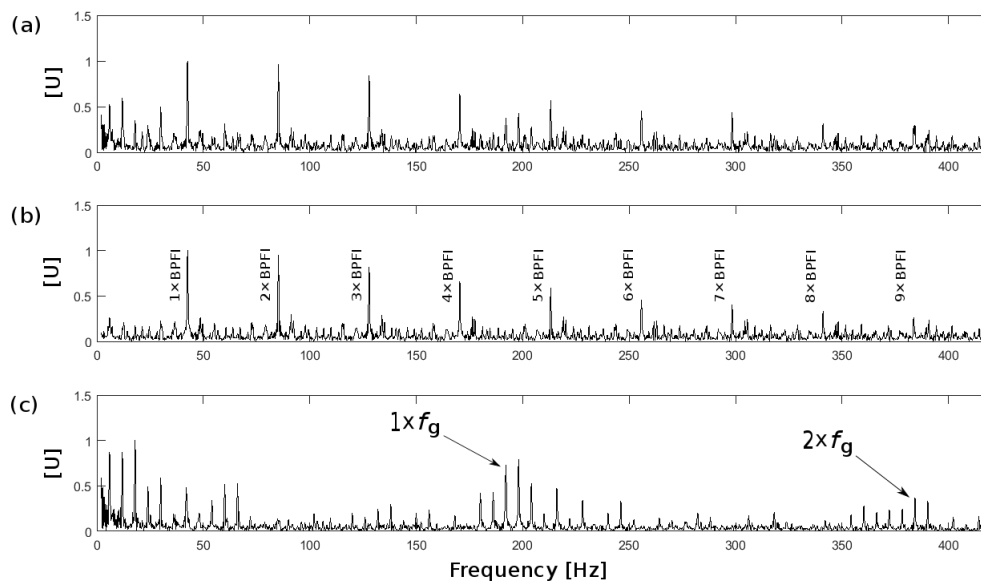


Fig. 4.37 Squared Envelope Spectrum of (a) the vibration signal of Case 8: $S_y^{SES}(\alpha)$, the recovered time signal of (b) source 1: $S_{\hat{x}^1}^{SES}(\alpha)$ and (c) that of source 2: $S_{\hat{x}^2}^{SES}(\alpha)$.

4.5 Conclusion

This chapter has proposed a novel periodic-variance based stochastic model to extract a cyclostationary signal in the masking of other interfering signals. It proceeds from the property that, for a CS signal, the STFT evidences periodic flows of energy in and across its frequency bins. By characterizing the periodic variance as hidden variables, a time-varying filter is designed so as to achieve the full-band reconstruction of the CS signals characterized by some pre-set characteristic frequencies, which can be obtained by prior knowledge or by the Spectral Correlation. Meanwhile, it provides a CS indicator to assess the level of CS components along the carrier frequency. The performance of the proposed scheme has been demonstrated on synthetic and experimental cases. Of particular interest is the robustness on experimental datasets and superior extraction capability over the conventional Wiener filter. It not only deals with the bearing fault at an incipient stage, but even works for the relevant problems to the bearings (such as the misalignment, the mechanical looseness, etc.) and the case of two sources, i.e. bearing and gear signal. Eventually, these experimental examples evidence its versatile usage on diagnostic analysis of compound signals.

Chapter 5

Benchmark survey on using the fast spectral correlation

Since the cyclostationary (CS) framework is actually rather large and includes many types of signals produced by periodic mechanisms, it has indeed become a very valuable and standard tool in practice. Although the Spectral Correlation (SC) is one of the most versatile spectral tools to analyze CS signals (i.e. signals comprising hidden periodicities or repetitive patterns), its use in condition monitoring has so far been hindered by its high computational cost. Most recently, this situation has been fixed by a new fast estimator of the SC, the *Fast Spectral Correlation* (Fast-SC), rooted on the short-time Fourier transform (STFT) [79]. In particular, it not only makes considerable the gain in computational cost, but it makes possible the analysis of long CS signals over a wide cyclic frequency range. To see this, the CWRU data sets have been entirely tested and the highlight is dealing with the more challenging cases. By means of systematic experiments and analyses, this chapter proposes a benchmark study on the Fast-SC and demonstrates the improvement on the Enhanced Envelope Spectrum (EES) through the comparison with three diagnostic methods. Finally, a whole picture to assess the performance of the Fast-SC is drawn in the form of tables while intending to make the SC a more widely spread tool in condition monitoring.

5.1 Background on applied diagnostic methods

According to the aforementioned description in Section 1.2, the fault detection can be intuitively addressed as a process of signal demodulation. As a very popular method, the envelope spectrum is a mean to demodulate high carrier frequencies and detect periodic modulations associated with the bearing characteristic frequencies [2, 4, 80, 81]. In its most

frequent version, the envelope is estimated by squaring the signal – or better, the analytical signal – and its spectrum is then computed [82]. This is called the squared envelope spectrum (SES). However, the SES of the raw signal is often not a good diagnostic indicator because several vibration components (categorized as background noise) are likely to mask the fault signature [5, 6, 83]. In addition, it can be greatly improved by pre-filtering the signal, hence the SES is applied as the fundamental diagnostic tool¹ in this chapter.

In order to reduce or remove the effect of masking sources, such as strong deterministic components, the benchmark study in Ref. [11] employs two practical and effective preprocessing steps to improve the performance of the SES.

- On the one hand, it sets all frequency components to the same magnitude, by using the Cepstrum Prewhitening method, so as to highlight only the pure random part whose phase information is initialized by pseudo-periodic impacts. In other words, both discrete frequencies and resonances will be removed together, no matter whether fault signature are present or not. Such an approach was first proposed in Ref. [84] and has been applied to variable speed applications in Ref. [85]. In particular, this method is easily implemented by setting the whole real cepstrum to zero, except for the zero frequency; the zeroed cepstrum is then transformed back to the time domain.
- On the other hand, it determines the best frequency band which maximizes the energy of the impulses with respect to the background noise. In other words, the band-pass filtered signal reduces spurious components while increasing the pure CS part. This issue has been well addressed and led to the use of indicators such as the kurtosis, which is viewed as an effective measure of the “impulsiveness” hidden in a signal [2, 13, 86]. In particular, the spectral kurtosis and its computation by the fast kurtogram provide a robust way to design almost optimal band-pass filters for incipient fault detection even in the presence of strong masking noise [86, 76]. Interestingly the kurtosis has been demonstrated to be equivalent to the sum of the peaks in the SES divided by the zero-frequency SES [87].

Another powerful tool for rolling element bearing diagnostic is based on the SC which proceeds from the CS property of rolling element bearing signals as demonstrated by Randall et al. [5] and further developed by Antoni [6, 72]. The SC ideally decomposes the signal in terms of modulation and carrier frequencies, which makes easy the detection of repetitive transients [6, 72, 88, 89]. Note that the SES has been demonstrated to be related to the SC – its integral over frequency – in Ref. [6]. Whereas the SC $S_x(\alpha, f)$ is a theoretical quantity,

¹For instance, in Ref. [11] three diagnostic techniques have been applied based on the SES, the differences between them being the preprocessing steps applied, as discussed in detail later.

the ACP $S_x^{ACP}(\alpha, f)$, the CMS $S_x^{CMS}(\alpha, f)$, and the Fast-SC $S_x^{Fast}(\alpha, f)$ are three different estimators of the SC. As it is demonstrated in Ref. [79], the connections between the spectral quantities are schemed in Fig. 5.1.

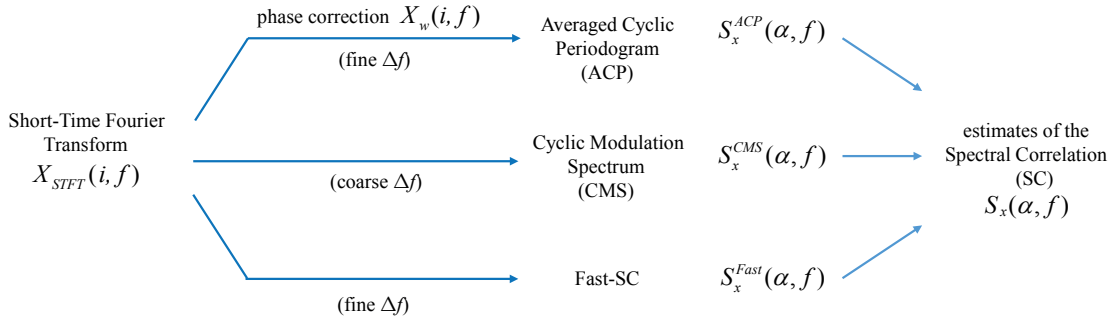


Fig. 5.1 Connections between the spectral quantities handled in the paper [78].

A popular estimator of the SC is obtained from the so-called “time-smoothed cyclic periodogram” [89][90] or, equivalently, the Averaged Cyclic Periodogram (ACP) [73] which is an extension of Welch’s method (also known in spectral analysis as the “Weighted-Overlapped-Segment-Averaging” method) to CS signals. Particular advantage of the ACP estimator is to achieve a high-resolution version of the envelope spectrum – arbitrary measure of cross-correlation between spectral components². However, to see the relatively fast periodic modulations, its usage may be hindered by the high computational cost that mainly consumed in the loop over cyclic frequencies $\alpha_k = k\Delta\alpha$. In practice such problem is quite common and troublesome, especially for naive users, while roughly selecting a large cyclic frequency range.

The Cyclic Modulation Spectrum (CMS) [91], being essentially a waterfall of envelope spectra at the output of a filterbank, stands as a much faster alternative, yet it suffers from the uncertainty principle and is thus limited to detect relatively slow periodic modulations. While the CMS is simply interpreted as the detection of periodic flows of energy *in* frequency bands, the Fast-SC extends it to the detection of periodic flows *across* different frequency bands [79]. In addition, the Fast-SC may be seen as a correction of the CMS such as to make it approach the ideal SC³. For practical considerations, its main advantage is that the computational effort is considerably alleviated as compared to the ACP since it essentially

²By synchronously recruiting the spectral components at frequencies f and $f - \alpha$, the available range of cyclic frequency – “modulation frequency” α that can shift apart away from the “carrier frequency” f – is therefore possible as high as the Nyquist frequency $F_s/2$ with a very fine resolution $\Delta\alpha = F_s/L$ [79].

³It is proved in [79] that the Fast-SC is an asymptotically convergent (unbiased and nil variance) estimator of the SC with similar statistical performance as the ACP.

relies on calculating the FFT of STFT products. Particularly, in terms of computational gains, their connections can be roughly summarized as [79]

$$\frac{\mathcal{C}_{Fast}}{\mathcal{C}_{CMS}} \sim P + 1, \quad \text{and} \quad \frac{\mathcal{C}_{ACP}}{\mathcal{C}_{Fast}} \sim \frac{L}{2R} \quad (5.1)$$

where P denotes the maximum value of p obtained by taking the nearest whole number rounded down to $N_w/(2R)$, L stands for the signal length and R for block shift in STFT when $L \gg R$.

This chapter aims to entirely assess the performance of the Fast-SC so that the CWRU data sets are tested and its results are then compared with benchmark results (the three diagnostic methods as mentioned in [11]). To see the “depth” of a modulation with frequency α and carrier f , the Spectral Coherence (SC)⁴ will serve as a basis to define the Squared Envelope Spectrum (SES) [5],

$$S_x^{SES}(\alpha) = \left| \int_{f_1}^{f_2} \gamma_x(\alpha, f) df \right|, \quad (5.2)$$

measured in a given frequency band $[f_1; f_2]$ and, a newly proposed spectral quantity, the “Enhanced Envelope Spectrum” (EES)

$$S_x^{EES}(\alpha) = \int_{f_1}^{f_2} |\gamma_x(\alpha, f)| df. \quad (5.3)$$

It is noteworthy that $S_x^{SES}(\alpha) \leq S_x^{EES}(\alpha)$ in general. The EES is expected to better enhance non-zero cyclic components than the SES because the latter integrates complex values, a process which may possibly a convergence towards zero in the case of fast rotating phases [79].

5.2 Case Western Reserve University data

For the last decade, the data set provided by the Case Western Reserve University (CWRU) Bearing Data Center [1] has been widely used as a standard reference in the bearing diagnostics field, e.g. Smith et al. [11] counted its use in 41 papers in Mechanical System and Signal Processing between 2004 and early 2015. To be a famous and standard data set, it indeed provides plenty of practical cases, with the author counting 161 test cases⁵ are

⁴It may also be interpreted as the SC of the whitened signal, which tends to equalize regions with very different energy levels and thus to magnify weak CS signals.

⁵The CWRU data set provides a number of bearing faults, i.e. seeded faults for Inner, Outer and Ball damage with various loading conditions (0-3 horsepower, approximate motor speeds of 1797-1720 rpm) and

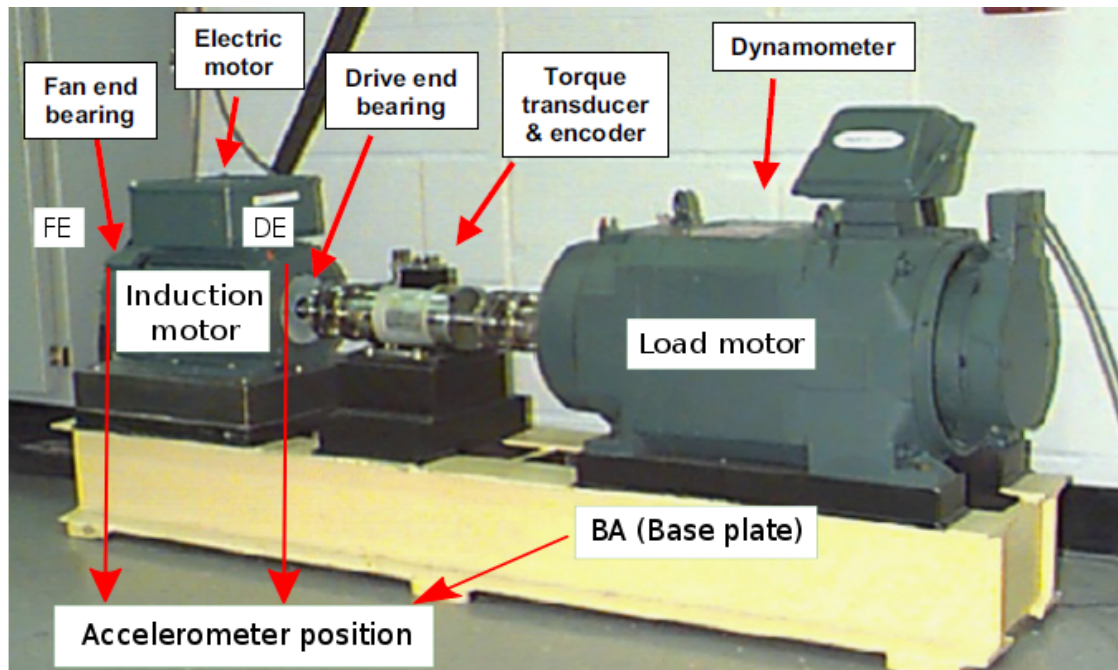


Fig. 5.2 CWRU bearing test rig [1].

publicly available online. Reference [11] provides a valuable description of the vibration signals found in the database together with the identification of the difficult cases which are worth consideration when trying to improve upon results obtained from state-of-the-art methods, such as the SES with possible prewhitening [85] and optimal prefiltering with the kurtogram [76]. The Fast-SC and the EES have been systematically computed for all cases investigated in Ref. [11]. The conclusion was that the Fast-SC never performed worse than the reference methods, but could improve the diagnosis outcomes⁶ in some difficult cases.

damage severity (faults ranging in diameter from 0.007 to 0.028 in. corresponding to 0.18-0.71 mm). For each test case, the measurement was possibly taken in the vertical direction on the housing of the drive-end bearing (DE), on the fan-end bearing housing (FE) and on the motor supporting base plate (BA) as indicated in Fig. 5.2. This means the entire data set contains more than 161 cells. Further details can be found at the CWRU Bearing Data Center website [1].

⁶The criteria for categorizing the diagnosis outcomes are outlined in Table 5.1 [11].

Table 5.1 Criteria for categorizing the diagnosis outcomes.

Diagnostic result	Category	Rules and details
Data clearly diagnosable	Y1	Evidence of classic characteristics for the given bearing fault in both the time and frequency domains
	Y2	Evidence of non-classic characteristics in either or both of the time and frequency domains
Data probably diagnosable	P1	Evidence of discrete components at the expected fault frequencies but they are not dominant in the spectrum
Data potentially diagnosable	P2	Evidence of smeared components that appear to coincide with the expected fault frequencies
Data not diagnosable	N1	Evidence of nothing for the specified bearing fault but with other identifiable problems (e.g., looseness)
	N2	Evidence of nothing for the specified bearing fault with the possible exception of shaft harmonics

5.2.1 Experimental set-up

The basic layout of the test rig is shown in Fig. 5.2. It consists of a 2 horsepower Reliance Electric motor driving a shaft on which a torque transducer and encoder are mounted. The torque is applied to the shaft via a dynamometer and electronic control system. Since there is no mechanism (e.g. gears) to convert such torque to the radial axis, the effect of ‘load’ here is almost inconsiderable for diagnosis outcomes as mentioned in Ref. [11]. It means that there is no radial load (in theory) borne by the bearings except the static gravitational load (6.00 o’clock position) from the weight of the shaft and any attached components. To see these, the tests for Outer damage are further grouped into three categories according to the fault position relative to the load zone: ‘centred’ (fault in the 6.00 o’clock position), ‘orthogonal’ (3.00 o’clock) and ‘opposite’ (12.00 o’clock).

For the tests, faults were seeded on the drive- and fan-end bearings (SKF deep-groove ball bearings: 6205-2RS JEM and 6203-2RS JEM, respectively) of the motor using electrodischarge machining (EDM). Table 5.2 shows the relevant bearing details and fault frequencies. The sample rates used were 12 kHz for some tests and 48 kHz for others. Information for all 161 test cases is divided into four categories – 48k baseline, 12k drive-end fault, 48k drive-end fault, and 12k fan-end fault – according to the sample rate and fault location.

Table 5.2 Bearing details and fault frequencies.

Position on rig	Model number	Fault frequencies (multiple of shaft speed)			
		BPFI	BPFO	FTF	BSF
Drive-end	SKF 6205-2RS JEM	5.415	3.585	0.3983	2.357
Fan-end	SKF 6203-2RS JEM	4.947	3.053	0.3816	1.994

Note: an equivalent NTN bearing, relative to the SKF 6205-2RS JEM, with the same fault frequencies was used for the 0.028 in. drive-end faults.

5.2.2 Examples and discussion

This section illustrates the use of the Fast-SC in combination with several examples while intending to move forward the benchmark study on the CWRU data set – uncovering its own

unique characteristics. Particularly, it has mainly investigated the “benchmark data set”⁷ in order to assess the newly proposed diagnostic algorithm, i.e. the Fast-SC [79]. And therein lie the reported 6 examples, marked in red (see Table D.2 and D.3) which illustrate quite well the general observation.

The parameter settings are given in Table 5.3 for record 125DE, 203DE, 275DE and Table 5.4 for record 277DE, 282DE, 290DE.

Table 5.3 Parameter settings used in the experiment of record 125DE, 203DE, 275DE.

Data Set	125DE	203DE	275DE
Sampling frequency F_s (kHz)	48		12
Duration (s)	10		
α_{max} (Hz)	300	500	600
N_w in Fast-SC	2^5	2^{10}	
R in Fast-SC	8	48	10
Rotation frequency f_{rot} (Hz)	28.68	29.15	29.63
Fundamental train frequency — FTF (Hz)	11.42		
Ball (roller) spin frequency — BSF (Hz)	67.61		
Ballpass frequency, outer race – BPFO (Hz)		104.5	
Ballpass frequency, inner race – BPF1 (Hz)			145.8

First, one example connected with the ball fault is presented which relates to record 125DE. It was categorised as “unsuccessful diagnoses” for all the 3 benchmark methods. As summarized in Ref. [11], the ball fault cases are certainly the most difficult to diagnose. Such phenomena are especially evident in some consecutive cases, e.g. 122DE, 123DE, 124DE, 125DE – with the smallest fault width (0.007 in.). For instance, all of their diagnosis

⁷It consists of the more challenging data sets, whose diagnosis outcomes are categorised by “partially successful” or even worse in Ref. [11], under the criteria in Table 5.1. Particularly, all the 68 cases were selected and divided into 3 tables in Appendix D.

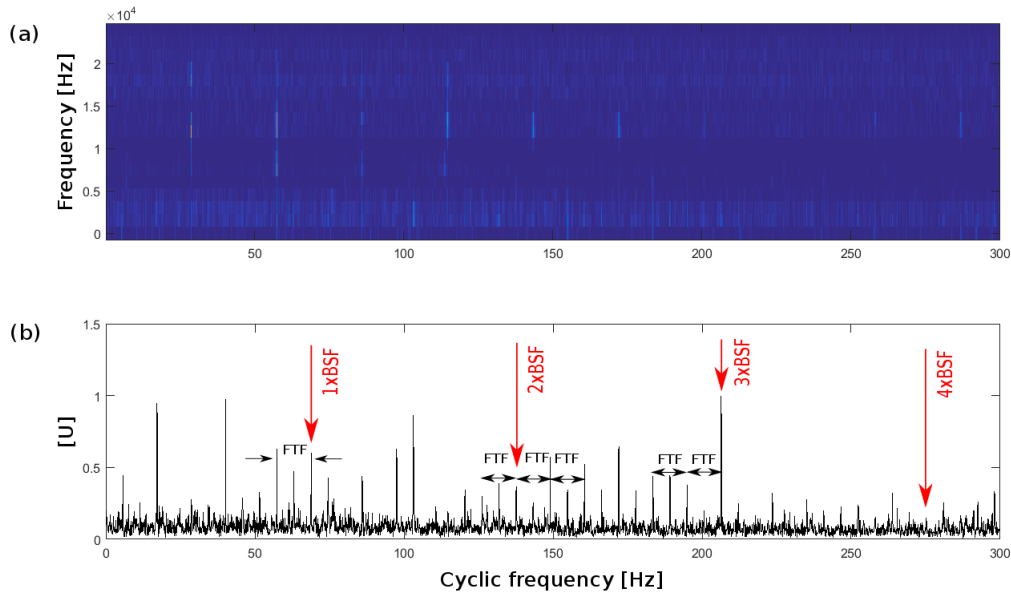


Fig. 5.3 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 125DE ($\Delta f = 1500$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in the band [750; 5250] Hz.

outcomes were “not diagnosable” with any of the techniques used in benchmark study (see Table D.2).

Figure 5.3 (a) displays the Spectral Coherence based on the Fast-SC in the full band [0; 24] kHz. Clearly, it reveals the harmonic structure of the shaft speed, particularly in the high “carrier frequency” of about [5; 20] kHz. Although this component dominates nearly the full band and strongly masks the fault signature, the smeared information has been found in the band [750; 5250] Hz. The Enhanced Envelope Spectrum is next computed by integrating the Spectral Coherence according to formula 5.3 in the latter band with maximal signal-to-noise ratio (SNR) as shown in Fig. 5.3 (b). The harmonics of BSF surrounded by modulation sidebands at the cage speed (FTF) are therefore verified for a ball damage. Since they are not dominant in the spectrum, the case is categorised by “partially successful”. Specially, for the consecutive 4 cases, they were measured under the same operating conditions except the loads. As explained above, the effect of ‘load’ is primarily on the shaft speed, which is reduced by almost 4% in the maximum load case. Hence the same phenomena – fault signature in [750; 5250] Hz – have also been found for the other cases with 0.007 in. fault width. The presented example – record 125DE – illustrates quite well the evident improvement in the consecutive cases, i.e. 122DE, 123DE, 124DE, 125DE in connection with ball faults.

The second example relates to record 203DE in connection with an outer-race fault, which was categorised as “P2” for all the 3 benchmark methods. As mentioned in Ref. [11], it shows impulse responses subject to impulsive amplitude modulation which gives rise to smearing effects in the envelope spectrum (see Fig. 26 therein). Figure 5.4 (a) shows the Spectral Coherence based on the Fast-SC with $\Delta f = 46.88$ Hz. It is noted that the aforementioned smeared components are dominant in the band of about [2; 5] kHz. The Enhanced Envelope Spectrum is then computed by integrating the Spectral Coherence in the band [500; 1500] Hz with maximal SNR in Fig. 5.4 (b).

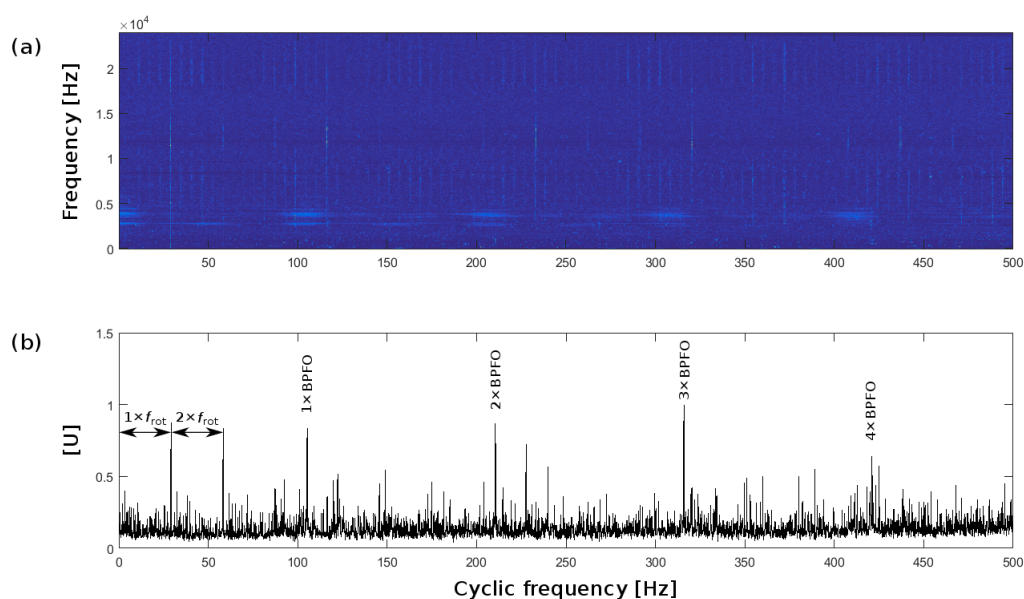


Fig. 5.4 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 203DE ($\Delta f = 46.88$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in the band [500; 1500] Hz.

Of particular interest is the 12k fan-end bearing faults (see Table D.3). For such fan-end bearings (SKF type 6203-2RS JEM deep groove in Table 5.2), it is noteworthy that three of the bearing characteristic frequencies are close to integer multiples of shaft speed, with values of 4.947, 3.053, 1.994 ($\times f_{rot}$) for BPFI, BPFO and BSF, respectively. As mentioned in Ref. [11], in some of the results, the bearing frequencies appear to lock onto these shaft harmonics, making it difficult to establish a definite diagnosis, though in others there is a small difference. Another interesting thing (cf. Table D.3) is that the accelerometric sensor is located on the drive-end bearing and is therefore far from the faulty bearing located on the fan-end, on the other side of a large – and possibly noisy – electrical motor. Above all, some of the data sets are dominated by non-classical features, especially for the cases with its own

unique characteristics. The following examples therefore demonstrate some typical cases contained therein.

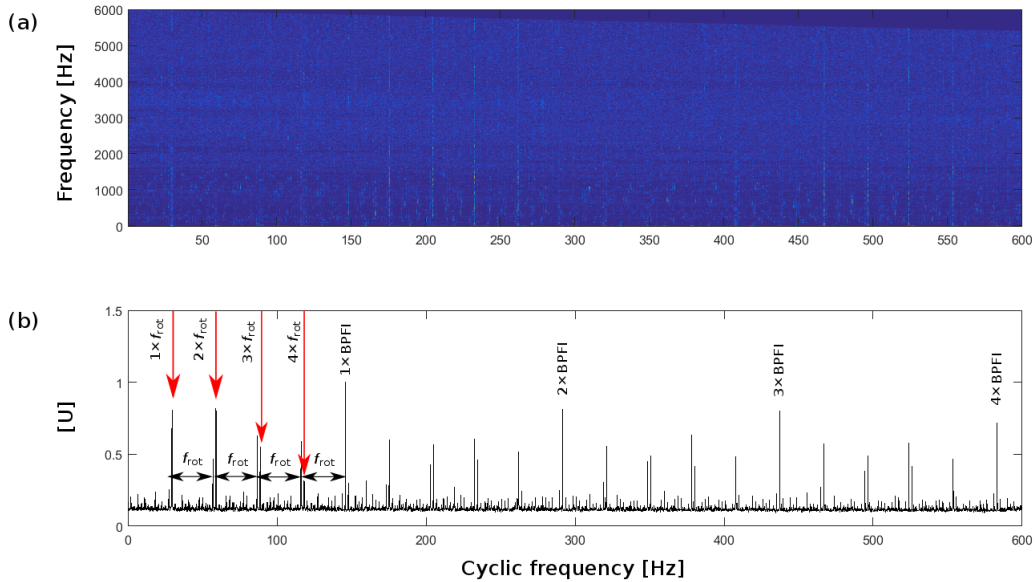


Fig. 5.5 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 275DE ($\Delta f = 11.72$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in full band $[0; F_s/2]$ Hz.

The third example derives from an inner-race fault, 275DE, whose measurement position is located on the drive-end bearing. As mentioned in Ref. [11], a small number of cases (275-277, drive-end measurements) are difficult to diagnose with any of the methods. For instance, they were categorised as P1 or N1 (see Table D.3).

Figure 5.5 (a) and (b) display the Fast-SC-based Spectral Coherence and Enhanced Envelope Spectrum in the full band $[0; 6]$ kHz. On closer inspection, the shaft harmonics are found and vanish rapidly after the fourth order, as marked by red arrows in Fig. 5.5 (b). It is also noted that the shaft-speed sidebands from BPFI ($= 4.947 \times f_{rot}$) can be distinguished therein (see black double-arrow). Concerning the small difference between BPFI and $5 \times f_{rot}$, Ref. [11] gives a physical interpretation on beat-type effect and provides a valuable description associated with the fan-end measurement (270FE, see Fig. 30 therein). In addition, it noted that such phenomenon – beating effect – was found to be clearest on the fan-end measurements and least clear on the base measurements. Hereafter, with the improvement of Fast-SC, it proves that the BPFI ($= 4.947 \times f_{rot}$) is not locking onto $5 \times f_{rot}$ in all the records (observed throughout the inner-race fault cases), and on closer inspection such two components are indeed distinguishable (see in Table D.3).

Table 5.4 Parameter settings used in the experiment of record 277DE, 282DE, 290DE.

Data Set	277DE	282DE	290DE
Sampling frequency F_s (kHz)	12		
Duration (s)	10		
α_{max} (Hz)	600	400	
N_w in Fast-SC	2^8	2^{11}	2^8
R in Fast-SC	10	15	
Rotation frequency f_{rot} (Hz)	28.88	29.95	29.92
Ballpass frequency, inner race (fan end) – BPF1 (Hz)	142.3		
Ballpass frequency, inner race (drive end) – BPF1 (Hz)	156.3		
Fundamental train frequency — FTF (Hz)		12	11.7
Ball (roller) spin frequency — BSF (Hz)		59.88	59.83

Another interesting example is one which illustrates quite atypical observation. It relates to record 277DE, a case with an inner-race fault, denoted as “partially successful” for all the 3 methods tested in Ref. [11] (see Table B4 therein). This is a difficult case because the accelerometric sensor is located on the drive-end bearing and is therefore far from the faulty bearing located on the fan-end, on the other side of a large – and possibly noisy – electrical motor.

Figure 5.6 (a) displays the Spectral Coherence based on the Fast-SC in the full band [0; 6] kHz. As demonstrated in Ref. [79], it seems to maximize the SNR in the band [4.3; 5.5] kHz. Next, the detection of the fault is further demonstrated by means of the EES computed in the band in Fig. 5.6 (b). The Fast-SC-based EES clearly evidences the dominant harmonics of the BPF1 ($= 4.947 \times f_{rot}$) with sidebands at the shaft rotation. Here, the very fine cyclic frequency resolution $\Delta\alpha = 0.1$ Hz in addition to the detection of higher order harmonics of the fault prevent us from such a confusion – beating effect.

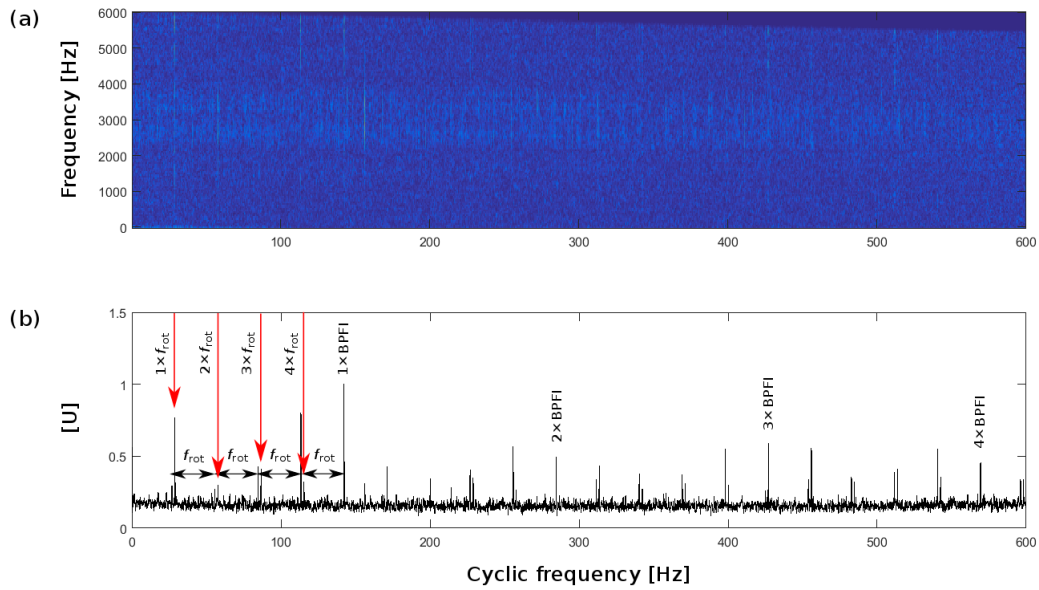


Fig. 5.6 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 277DE ($\Delta f = 46.88$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in the band [4.3; 5.5] kHz.

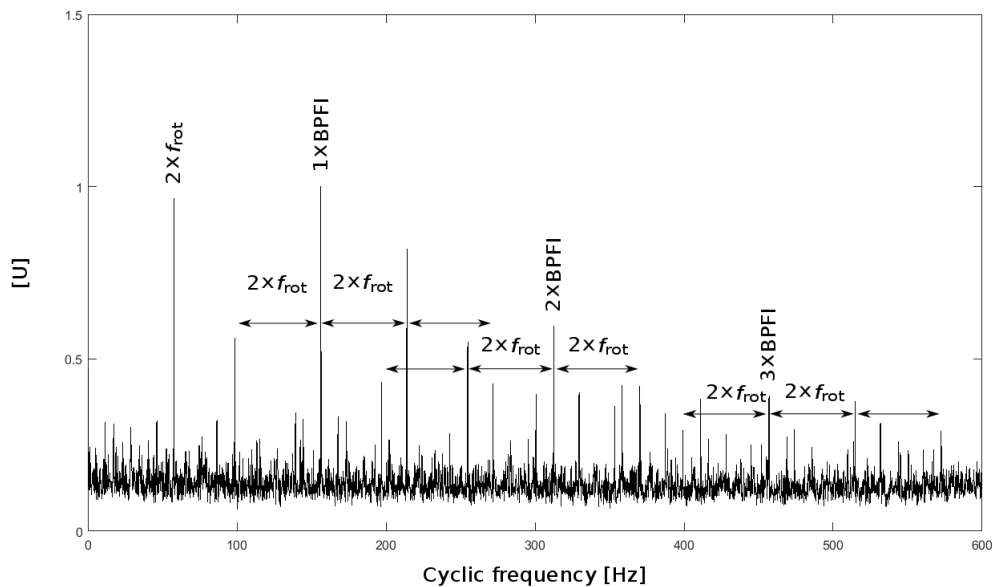


Fig. 5.7 Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in selected band [2.3; 3.7] kHz with $\Delta\alpha = 0.1$ Hz.

Incidentally, the Spectral Coherence in Fig. 5.6 (a) also shows a high interference at $\alpha = 156.3$ Hz which dominates in the band $[2.3; 3.7]$ kHz. The Fast-SC-based EES computed in this band is shown in Fig. 5.7. It displays the signature of an inner-race fault in the drive-end bearing with marked side-bands at twice the rotation speed. Although no such fault is reported in the literature for the configuration relating to record 277DE, it is believed that the accelerometric sensor (which is close to drive-end bearing) sees a misalignment of the drive-end bearing due to the numerous dismantling operations carried out in the experiment – a loose misaligned inner-race would then have a potential signature at the BPF with modulations at twice the rotation speed.

As aforementioned for the drive-end bearings, the ball faults are quite often connected with the most difficult cases for diagnostics. Here, this situation is further complicated by the fact that the BSF ($= 1.994 \times f_{rot}$) is really close to the second multiple of the shaft speed, which might be troublesome for fault identification. The following two examples intend to illustrate how the Fast-SC can improve the results and deepen user's understanding of the CWRU data sets.

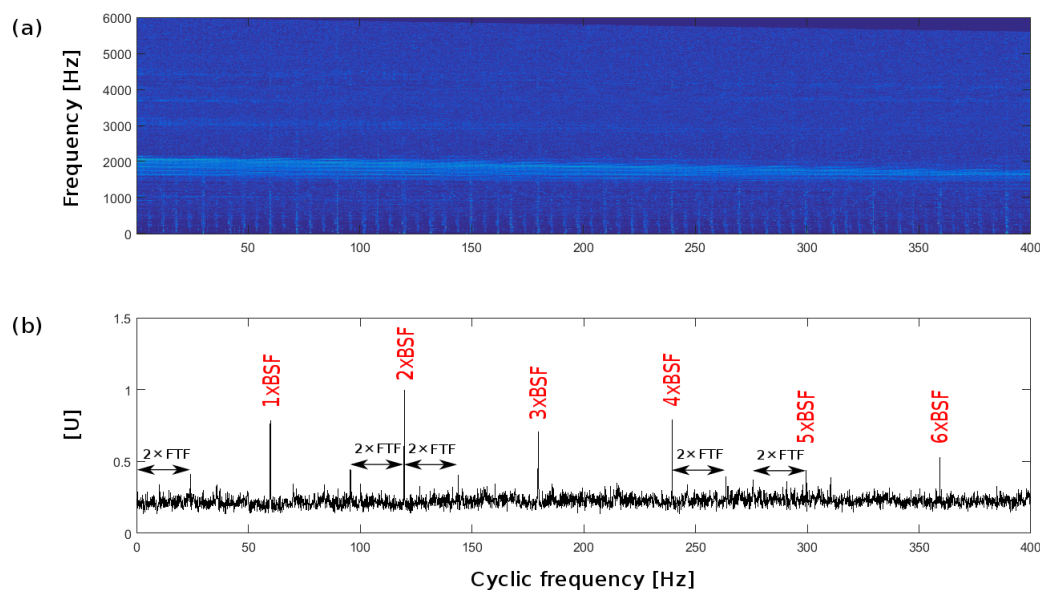


Fig. 5.8 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 282DE ($\Delta f = 5.86$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in the band $[4.1; 4.3]$ kHz.

The fifth example relates to record 282DE, a case with a ball fault, denoted as “partially successful” or even worse for the benchmark result in Ref. [11] (see Table D.3). Fig. 5.8 (a) displays the Spectral Coherence based on the Fast-SC in the full band $[0; 6]$ kHz

($\Delta f = 5.86$ Hz, $\Delta\alpha = 0.1$ Hz). It is evident that smeared components are observed over a narrow frequency range of about [1.5; 2] kHz. This phenomenon is possibly attributable to two separate reasons, one mechanical looseness and one electro-magnetic interference (EMI). The interference was first thought to have been caused by EMI, but closer inspection on its time record revealed that it contains seemingly random pulses⁸, again perhaps attributable to mechanical looseness.

Particularly, the Spectral Coherence in Fig. 5.8 (a) also shows a clear harmonic of 59.88 Hz which dominates in the band [4.1; 4.3] kHz. Next, the Fast-SC-based EES computed in such band is displayed in Fig. 5.8 (b). It reveals the harmonic structure of BSF (with dominant even multiples thereof) surrounded by twice the cage speed. As will be seen in the next example, such side-bands at twice the cage speed are also present and possibly caused by mechanical looseness.

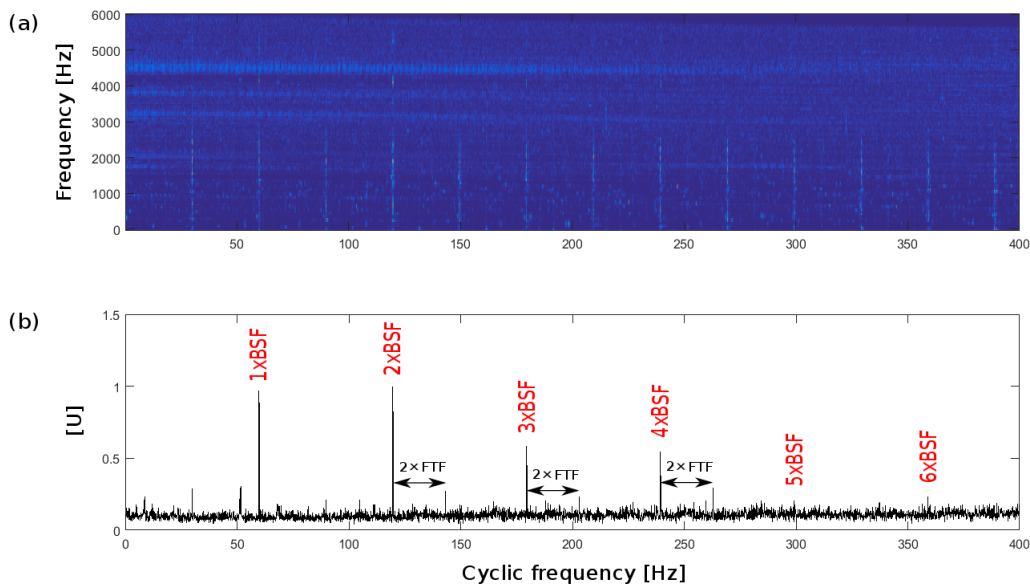


Fig. 5.9 (a) Spectral Coherence based on the Fast-SC, $S_y^{Fast}(\alpha, f)$, of record 290DE ($\Delta f = 46.88$ Hz, $\Delta\alpha = 0.1$ Hz). (b) Fast-SC-based Enhanced Envelope Spectrum $S_y^{EES}(\alpha)$ in the band [3.9; 4.4] kHz.

The sixth example is also a ball fault case with deeper fault width (0.021 in.) – 290DE. It was categorised as “N1” for all the 3 benchmark methods (see in Table D.3). Fig. 5.9 (a) displays the Spectral Coherence based on the Fast-SC in the full band [0; 6] kHz ($\Delta f = 46.88$ Hz, $\Delta\alpha = 0.1$ Hz). It shows clear harmonics of the shaft speed which dominates a low

⁸As discussed in Ref. [11], it is suspected to be a result of mechanical looseness, causing impulsive modulation of random amplitude at intervals of one revolution, but not necessarily phase-locked to the rotation.

frequency range from 0 Hz up to 3 kHz. Such harmonics are not unfamiliar, especially in connection with the discussion above regarding the effect of mechanical looseness.

In addition, closer inspection on the Spectral Coherence also reveals a distinct harmonic of 59.83 Hz carried in the band [3.9; 4.4] kHz. The Fast-SC-based EES is next computed in such band with maximal SNR as shown in Fig. 5.9 (b). It displays the typical signature of a ball fault – with twice the cage speed ($2 \times \text{FTF}$) as the previous case – possibly caused by mechanical looseness.

5.3 Result tables and conclusion

It was mentioned in Section 5.1 that the Fast-SC has been selected as an ideal tool for condition monitoring. This is because of its high capability to unwrap complicated signals onto a two-dimensional map that clearly reveals the presence of modulations and makes easy the identification of fault frequencies. In addition, the Fast-SC-based EES computed in selected bands provides an improved version of the SES and gives a versatile tool especially for complicated cases.

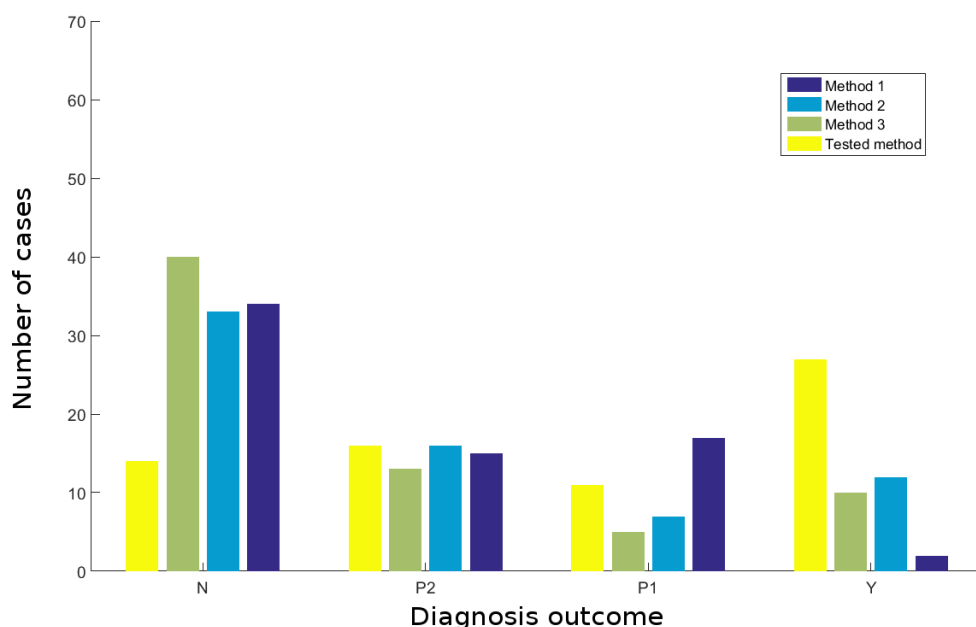


Fig. 5.10 Method comparison: diagnosis outcomes for the “benchmark data set” (Method 1-3 are outlined in Section 5.1; Y = successful, P = partially successful, N = not successful, more details can be found in Table 5.1.).

As for the tested CWRU data sets, many of the consecutive cases exhibit similar characteristics and in particular for some challenging ones, they have their own unique fault signatures. To examine the newly proposed Fast-SC, the “benchmark data set” is built and therein lies the 6 examples which result in a more comprehensive diagnosis of data sets in the P categories or in a successful diagnosis for data sets in the N categories.

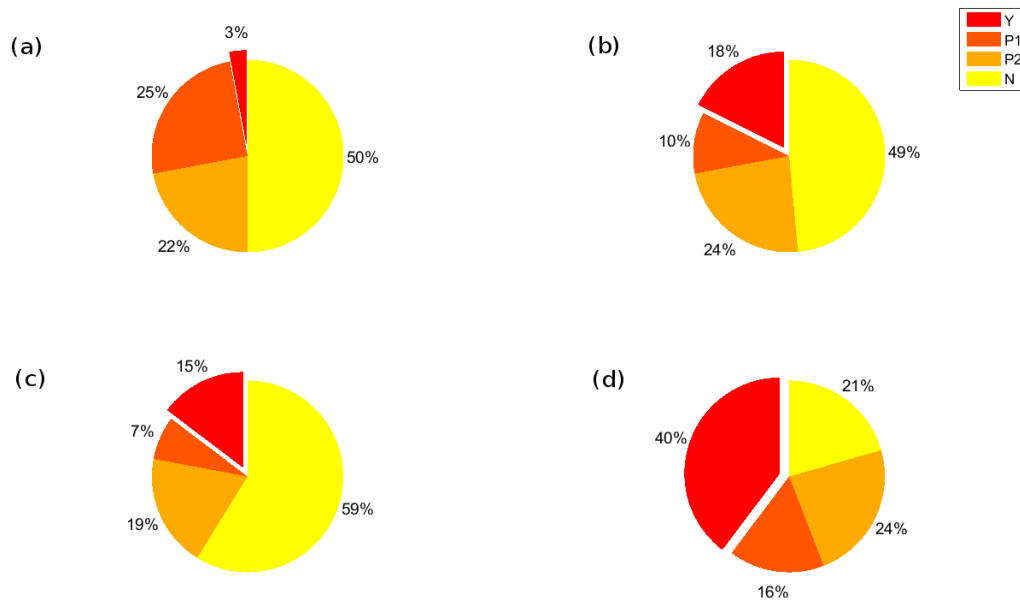


Fig. 5.11 Pie charts of diagnosis outcomes for the “benchmark data set”: (a)-(c) respectively corresponded to benchmark method 1-3 and (d) to tested method Fast-SC (Method 1-3 are outlined in Section 5.1; Y = successful, P = partially successful, N = not successful, more details can be found in Table 5.1.).

In addition, the tables in Appendix D (Table D.1-D.3) gives a whole picture of the diagnostic performance of the techniques of Section 5.1. Figure 5.10 draws the bar chart on diagnosis outcomes for the “benchmark data set” by counting the number of cases categorised as ‘Y’, ‘P1’, ‘P2’, ‘N’. In the figure, the ‘N1’ and ‘N2’ categories have been merged into the ‘N’ for not diagnosable; the ‘Y1’ and ‘Y2’ into the ‘Y’ for successfully diagnosable. It can be seen that the tested method achieves at least as twice the number of successful diagnosis as the benchmark methods; inversely, for the ‘N’ diagnosis, it is less than the half of any other methods. Furthermore, the pie charts give more details for quantitatively assessing their performances.

This chapter has given the benchmark survey on using the fast computation of the spectral correlation. It has benefited from a big data set – which is publicly available and widely used – supplied by the Case Western Reserve University (CWRU) Bearing Data Center.

Particularly, it contains plenty cases ranged from very easily diagnosable to not diagnosable, while some signals show the typical harmonic structure of bearing fault signature, others are quite blurry or even display other fault symptoms. This study aims to make the Faset-SC a more widely spread tool in condition monitoring, thereby it has selected a big data set and tested all the cases. The performance of proposed Fast-SC has therefore been assessed by the partial or non-diagnosable cases for the benchmark result. Another contribution of this work has been to move forward the benchmark study of public and commonly used data set – Case Western Reserve University (CWRU) Bearing Data Center – by uncovering its own unique characteristics.

Chapter 6

Conclusions and perspectives

6.1 Conclusions

In this PhD work, we investigated vibration-based rolling element bearing diagnostics via stochastic models that exploit sparse representations of vibration signals. Sparsity is a property that enables us to reveal the natural signature of singular events embedded in a signal so as to reduce the demand on the user's expertise. It is noteworthy that sparse representations is a general principle for finding statistically independent components in vibration signals, rather than a special and formulaic model for matching more physical characteristics. It is demanding of an appropriate transform that promotes the sparsity, rather than a redundant and complicated basis function. A literature survey on sparse representations of vibration signals has been proposed and few crucial points were drawn for its applications in machinery diagnostics.

This thesis mainly studied two stochastic models – one is based on the hidden Markov model and the other is periodic-variance based model – and a fast estimator of the spectral correlation which proceeds from the fact that the STFT evidences time-dependent flows of energy in and across its frequency bins. As a consequence, all of them benefit from the computational advantage of the Short-Time-Fourier-Transform. This is the reason why the Fast-SC can offer a substantial computational gain which makes it very practical when it comes to analyse long records over a wide cyclic frequency range. Furthermore, it helped to move forward the benchmark survey on the CWRU data set – by uncovering its own unique characteristics. Of particular interest is that the proposed periodic-variance based model can be intuitively viewed as a special case – under periodic assumptions instead of non-stationarity – of the hidden Markov model. Whereas they have an alike interpretation from the Bayesian viewpoint, yet they were illuminated by different properties.

The hidden Markov model introduces a hidden variable to indicate the occurrence of the impacts and estimates the spectral content of the corresponding transients together with the spectrum of background noise. This gives rise to an automatic detection algorithm – with no need of manual prefiltering as is the case with the envelope spectrum – from which fault frequencies can be revealed. The same algorithm also makes possible to filter out the fault signal in a very efficient way as compared to other approaches based on the stationary assumption. The performance is investigated on synthetic signals with a high noise-to-signal ratio and also in the case of a mixture of two independent transients. The effectiveness and robustness of the method are also verified on vibration signals measured on a test-bench (gears and bearings). Results are found superior or at least equivalent to those of conventional envelope analysis and fast kurtogram.

The periodic-variance based model aims to extract cyclostationary (CS) signals in the masking of interfering signals. First, it introduces a stochastic model that characterizes the second-order statistical descriptors as hidden variables so as to recover the CS component characterized by pre-set cyclic frequency. Meanwhile, it provides a CS indicator to assess the level of the CS components along carrier frequency. The validity of the proposed scheme has been demonstrated on synthetic and experimental cases. Of particular interest is the robustness on experimental datasets and superior extraction capability over the conventional Wiener filter. It not only deals with the bearing fault at an incipient stage, but even works for the relevant problems to the bearings (such as the misalignment, the mechanical looseness, etc.) and the case of two sources, i.e. bearing and gear signal. Eventually, these experimental examples evidence its versatile usage on diagnostic analysis of compound signals.

In summary, by discussing the pros and cons of sparse representations, this thesis thereby investigated both *analytic* and *learning-based* routes in the Bayesian viewpoint to promote the “hidden” sparsity of vibration signals. It intends to share and merge the information on observations and prior knowledge in the general sense so as to reduce the demand on the user’s expertise. With the idea of decomposing signals into some “sparse structure”, novel stochastic models have been explored and applied for rolling element bearing diagnostics.

6.2 Perspectives

Due to limited time, many further works have remained open and should be considered in the future.

- We have proposed an automatic detection algorithm – with no need of manual prefiltering as is the case with the envelope spectrum – from which fault frequencies can be revealed. In addition, we also discussed the initialization and setting of the

required parameters, i.e. the window length N_w and the window shift R . However, in terms of N_w , it still depends on manual tuning. To achieve an intelligent method, it needs to be solved. Another interesting point to develop is related to proposed model. It only assumes that each state obeys a stationary condition which relaxes the assumption of cyclostationarity for a wider application. Nevertheless in consideration of cyclostationary signals (i.e. signals comprising hidden periodicities or repetitive patterns), its performance can still be improved by recognizing hidden variables as periodic variation so as to promote the accuracy for capturing repetitive transients.

- In recent decades, more and more interesting studies have been reported on sparsely representing rotating machines signals. Some promising successes will keep its developments growing in this area. Exploitation of dictionary learning in vibration signals is valuable and promising, yet too much interest is payed into *overcomplete* for training the basis functions. As discussed in Chapter 2, the application of sparsity in machine diagnosis still has to consider a few crucial points. Particularly, for the *learning-based* route, with such a large degree of freedom, it seems more like a black-box, which possibly causes artifacts and loss of information, such as wrong classification of some unwanted signals as sparse, or misclassification of some signals in Ω as dense. Anyway, the exploratory study – *learning-based* dictionary algorithm – of sparse representations of vibration signals is still going on. The notion focuses on the assumption of a higher degree of statistical independence of the desired signal, which gives rises to filtering out the other statistical structures, such as background noise $n(t)$.

References

- [1] Kenneth A Loparo. Bearing data center, case western reserve university, 2003.
- [2] Robert B Randall and Jérôme Antoni. Rolling element bearing diagnostics—a tutorial. *Mechanical Systems and Signal Processing*, 25(2):485–520, 2011.
- [3] PD McFadden and JD Smith. Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of sound and vibration*, 96(1):69–82, 1984.
- [4] PD McFadden and JD Smith. Vibration monitoring of rolling element bearings by the high-frequency resonance technique—a review. *Tribology international*, 17(1):3–10, 1984.
- [5] Robert B Randall, Jérôme Antoni, and S Chobsaard. The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. *Mechanical systems and signal processing*, 15(5):945–962, 2001.
- [6] Jérôme Antoni. Cyclic spectral analysis of rolling-element bearing signals: facts and fictions. *Journal of Sound and vibration*, 304(3):497–529, 2007.
- [7] S Braun. The synchronous (time domain) average revisited. *Mechanical Systems and Signal Processing*, 25(4):1087–1102, 2011.
- [8] DC Baillie and J Mathew. A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, 10(1):1–17, 1996.
- [9] Lingli Cui, Jing Wang, and Seungchul Lee. Matching pursuit of an adaptive impulse dictionary for bearing fault diagnosis. *Journal of Sound and Vibration*, 333(10):2840–2862, 2014.
- [10] Haining Liu, Chengliang Liu, and Yixiang Huang. Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 25(2):558–574, 2011.
- [11] Wade A Smith and Robert B Randall. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical Systems and Signal Processing*, 64:100–131, 2015.
- [12] C Ottonello and S Pagnan. Modified frequency domain kurtosis for signal processing. *Electronics Letters*, 30(14):1117–1118, 1994.

- [13] Jérôme Antoni. The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing*, 20(2):282–307, 2006.
- [14] Hongyu Yang, Joseph Mathew, and Lin Ma. Fault diagnosis of rolling element bearings using basis pursuit. *Mechanical Systems and Signal Processing*, 19(2):341–356, 2005.
- [15] Zhipeng Feng and Fulei Chu. Application of atomic decomposition to gear damage detection. *Journal of Sound and Vibration*, 302(1):138–151, 2007.
- [16] Liang Guo, Hongli Gao, Jun Li, Haifeng Huang, and Xiaochen Zhang. Machinery vibration signal denoising based on learned dictionary and sparse representation. In *Journal of Physics: Conference Series*, volume 628, page 012124. IOP Publishing, 2015.
- [17] PC Russell, J Cosgrave, D Tomtsis, A Vourdas, L Stergioulas, and GR Jones. Extraction of information from acoustic vibration signals using gabor transform type devices. *Measurement science and technology*, 9(8):1282, 1998.
- [18] Qingfeng Meng and Liangsheng Qu. Rotating machinery fault diagnosis using wigner distribution. *Mechanical Systems and Signal Processing*, 5(3):155–166, 1991.
- [19] ZK Peng and FL Chu. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical systems and signal processing*, 18(2):199–221, 2004.
- [20] Zhipeng Feng, Ming Liang, and Fulei Chu. Recent advances in time–frequency analysis methods for machinery fault diagnosis: a review with application examples. *Mechanical Systems and Signal Processing*, 38(1):165–205, 2013.
- [21] Ruqiang Yan, Robert X Gao, and Xuefeng Chen. Wavelets for fault diagnosis of rotary machines: a review with applications. *Signal Processing*, 96:1–15, 2014.
- [22] VK Rai and AR Mohanty. Bearing fault diagnosis using fft of intrinsic mode functions in hilbert–huang transform. *Mechanical Systems and Signal Processing*, 21(6):2607–2615, 2007.
- [23] Xuefeng Chen, Zhaohui Du, Jimeng Li, Xiang Li, and Han Zhang. Compressed sensing based on dictionary learning for extracting impulse components. *Signal Processing*, 96:94–109, 2014.
- [24] Evan C Smith and Michael S Lewicki. Efficient auditory coding. *Nature*, 439(7079):978–982, 2006.
- [25] Wangpeng He, Yin Ding, Yanyang Zi, and Ivan W Selesnick. Sparsity-based algorithm for detecting faults in rotating machines. *Mechanical Systems and Signal Processing*, 72:46–64, 2016.
- [26] Haitao Zhou, Jin Chen, Guangming Dong, and Ran Wang. Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden markov model. *Mechanical Systems and Signal Processing*, 72:65–79, 2016.

- [27] Haifeng Tang, Jin Chen, and Guangming Dong. Sparse representation based latent components analysis for machinery weak fault detection. *Mechanical Systems and Signal Processing*, 46(2):373–388, 2014.
- [28] Javier Portilla. Image restoration through l0 analysis-based sparse optimization in tight frames. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 3909–3912. IEEE, 2009.
- [29] Michael Elad, J-L Starck, Philippe Querre, and David L Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Applied and Computational Harmonic Analysis*, 19(3):340–358, 2005.
- [30] Anna Llagostera Casanovas, Gianluca Monaci, Pierre Vandergheynst, and Rémi Gribonval. Blind audiovisual source separation based on sparse redundant representations. *IEEE Transactions on Multimedia*, 12(5):358–371, 2010.
- [31] Mark D Plumbley, Thomas Blumensath, Laurent Daudet, Rémi Gribonval, and Mike E Davies. Sparse representations in audio and music: from coding to source separation. *Proceedings of the IEEE*, 98(6):995–1005, 2010.
- [32] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.
- [33] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.
- [34] Bruno A Olshausen et al. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.
- [35] Bruno A Olshausen and David J Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37(23):3311–3325, 1997.
- [36] Bruno A Olshausen and David J Field. Sparse coding of sensory inputs. *Current opinion in neurobiology*, 14(4):481–487, 2004.
- [37] Ron Rubinstein. *Analysis versus synthesis sparse modeling method in image processing*. PhD thesis, Israel Institute of Technology, 2011.
- [38] Michael Elad, Peyman Milanfar, and Ron Rubinstein. Analysis versus synthesis in signal priors. *Inverse problems*, 23(3):947, 2007.
- [39] Rafael Molina, Jorge Núñez, Francisco José Cortijo, and Javier Mateos. Image restoration in astronomy: a bayesian perspective. *IEEE Signal Processing Magazine*, 18(2):11–29, 2001.
- [40] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.

- [41] Ivan W Selesnick and Mario AT Figueiredo. Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. In *SPIE Optical Engineering+ Applications*, pages 74460D–74460D. International Society for Optics and Photonics, 2009.
- [42] Sangnam Nam, Mike E Davies, Michael Elad, and Rémi Gribonval. The cosparsity model and algorithms. *Applied and Computational Harmonic Analysis*, 34(1):30–56, 2013.
- [43] Ron Rubinstein, Tomer Peleg, and Michael Elad. Analysis k-svd: a dictionary-learning algorithm for the analysis sparse model. *IEEE Transactions on Signal Processing*, 61(3):661–677, 2013.
- [44] Ron Rubinstein, Michael Zibulevsky, and Michael Elad. Double sparsity: Learning sparse dictionaries for sparse signal approximation. *IEEE Transactions on signal processing*, 58(3):1553–1564, 2010.
- [45] Guoshen Yu, Guillermo Sapiro, and Stéphane Mallat. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *IEEE Transactions on Image Processing*, 21(5):2481–2499, 2012.
- [46] Roger Grosse, Rajat Raina, Helen Kwong, and Andrew Y Ng. Shift-invariance sparse coding for audio classification. *arXiv preprint arXiv:1206.5241*, 2012.
- [47] Boris Mailhé, Sylvain Lesage, Rémi Gribonval, Frédéric Bimbot, and Pierre Vandergheynst. Shift-invariant dictionary learning for sparse representations: extending k-svd. In *Signal Processing Conference, 2008 16th European*, pages 1–5. IEEE, 2008.
- [48] Po-Yu Chen and Ivan W Selesnick. Translation-invariant shrinkage/thresholding of group sparse signals. *Signal Processing*, 94:476–489, 2014.
- [49] Po-Yu Chen and Ivan W Selesnick. Group-sparse signal denoising: Non-convex regularization, convex optimization. *IEEE Transactions on Signal Processing*, 62(13):3464–3478, 2014.
- [50] Papyan Vardan, Yaniv Romano, and Michael Elad. Convolutional neural networks analyzed via convolutional sparse coding. *arXiv preprint arXiv:1607.08194*, 2016.
- [51] Stéphane Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [52] Emmanuel J Candes and David L Donoho. Curvelets: A surprisingly effective non-adaptive representation for objects with edges. Technical report, DTIC Document, 2000.
- [53] Minh N Do and Martin Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on image processing*, 14(12):2091–2106, 2005.
- [54] Jont B Allen and Lawrence R Rabiner. A unified approach to short-time fourier analysis and synthesis. *Proceedings of the IEEE*, 65(11):1558–1564, 1977.

- [55] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.
- [56] Rene Vidal, Yi Ma, and Shankar Sastry. Generalized principal component analysis (gpca). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1945–1959, 2005.
- [57] Joel A Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information theory*, 50(10):2231–2242, 2004.
- [58] Irina F Gorodnitsky and Bhaskar D Rao. Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm. *IEEE Transactions on signal processing*, 45(3):600–616, 1997.
- [59] Geoff Davis, Stephane Mallat, and Marco Avellaneda. Adaptive greedy approximations. *Constructive approximation*, 13(1):57–98, 1997.
- [60] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. In *Advances in neural information processing systems*, pages 801–808, 2006.
- [61] Leslie N Smith and Michael Elad. Improving dictionary learning: Multiple dictionary updates and coefficient reuse. *IEEE Signal Processing Letters*, 20(1):79–82, 2013.
- [62] Wangpeng He, Yin Ding, Yanyang Zi, and Ivan W Selesnick. Repetitive transients extraction algorithm for detecting bearing faults. *Mechanical Systems and Signal Processing*, 84:227–244, 2017.
- [63] Yin Ding, Wangpeng He, Binqiang Chen, Yanyang Zi, and Ivan W Selesnick. Detection of faults in rotating machinery using periodic time-frequency sparsity. *Journal of Sound and Vibration*, 382:357–378, 2016.
- [64] Zhinong Li, Zhaotong Wu, Yongyong He, and Chu Fulei. Hidden markov model-based fault diagnostics method in speed-up and speed-down process for rotating machinery. *Mechanical Systems and Signal Processing*, 19(2):329–339, 2005.
- [65] Diego A Tobon-Mejia, Kamal Medjaher, Noureddine Zerhouni, and Gerard Tripot. A mixture of gaussians hidden markov model for failure diagnostic and prognostic. In *2010 IEEE International Conference on Automation Science and Engineering*, pages 338–343. IEEE, 2010.
- [66] Qiang Miao and Viliam Makis. Condition monitoring and classification of rotating machinery using wavelets and hidden markov models. *Mechanical systems and signal processing*, 21(2):840–855, 2007.
- [67] Hasan Ocak and Kenneth A Loparo. Hmm-based fault detection and diagnosis scheme for rolling element bearings. *Journal of Vibration and Acoustics*, 127(4):299–306, 2005.
- [68] Theo Heyns, Philippus Stephanus Heyns, and Johan Pieter De Villiers. Combining synchronous averaging with a gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox. *Mechanical Systems and Signal Processing*, 32:200–215, 2012.

- [69] Hasan Ocak and Kenneth A Loparo. A new bearing fault detection and diagnosis scheme based on hidden markov modeling of vibration signals. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 5, pages 3141–3144. IEEE, 2001.
- [70] Diego Alejandro Tobon-Mejia, Kamal Medjaher, Noureddine Zerhouni, and Gerard Tripot. A data-driven failure prognostics method based on mixture of gaussians hidden markov models. *IEEE Transactions on reliability*, 61(2):491–503, 2012.
- [71] Jianbo Yu. Health condition monitoring of machines based on hidden markov model and contribution analysis. *IEEE Transactions on Instrumentation and Measurement*, 61(8):2200–2211, 2012.
- [72] Jérôme Antoni, Frédéric Bonnardot, A Raad, and Mohamed El Badaoui. Cyclostationary modelling of rotating machine vibration signals. *Mechanical systems and signal processing*, 18(6):1285–1314, 2004.
- [73] Jérôme Antoni. Cyclic spectral analysis in practice. *Mechanical Systems and Signal Processing*, 21(2):597–630, 2007.
- [74] Jérôme Antoni. Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, 23(4):987–1036, 2009.
- [75] Jeff A Bilmes et al. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4(510):126, 1998.
- [76] Jérôme Antoni. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, 21(1):108–124, 2007.
- [77] Dany Abboud, Jérôme Antoni, Mario Eltabach, and Sophie Sieg-Zieba. Speed-spectral whitening for enhancing envelope analysis in speed varying conditions. In *VISHNO 2014*, 2014.
- [78] William A Gardner. Exploitation of spectral redundancy in cyclostationary signals. *IEEE Signal processing magazine*, 8(2):14–36, 1991.
- [79] Jérôme Antoni, Ge Xin, and Nacer Hamzaoui. Fast computation of the spectral correlation. *Mechanical Systems and Signal Processing*, 92:248–277, 2017.
- [80] Michael Feldman. Hilbert transform in vibration analysis. *Mechanical systems and signal processing*, 25(3):735–802, 2011.
- [81] Mark S Darlow, Robert H Badgley, and GW Hogg. Application of high-frequency resonance techniques for bearing diagnostics in helicopter gearboxes. Technical report, DTIC Document, 1974.
- [82] D Ho and RB Randall. Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mechanical systems and signal processing*, 14(5):763–788, 2000.

- [83] J Antoni and RB Randall. Differential diagnosis of gear and bearing faults. *Journal of Vibration and Acoustics*, 124(2):165–171, 2002.
- [84] N Sawalhi and RB Randall. Signal pre-whitening using cepstrum editing (liftering) to enhance fault detection in rolling element bearings. In *Proceedings of the 24 international congress on condition monitoring and diagnostic engineering management (comadem2011), may*, pages 330–336, 2011.
- [85] P Borghesani, P Pennacchi, RB Randall, N Sawalhi, and R Ricci. Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions. *Mechanical Systems and Signal Processing*, 36(2):370–384, 2013.
- [86] Jérôme Antoni and RB Randall. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, 20(2):308–331, 2006.
- [87] P Borghesani, P Pennacchi, and S Chatterton. The relationship between kurtosis-and envelope-based indexes for the diagnostic of rolling element bearings. *Mechanical Systems and Signal Processing*, 43(1):25–43, 2014.
- [88] B. Boashash, E.J. Powers, and A.M. Zoubir. *Higher-order statistical signal processing*. Longman Cheshire, 1995.
- [89] W Gardner. Measurement of spectral correlation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1111–1123, 1986.
- [90] Randy S Roberts, William A Brown, and Herschel H Loomis. Computationally efficient algorithms for cyclic spectral analysis. *IEEE Signal Processing Magazine*, 8(2):38–49, 1991.
- [91] Jerome Antoni and David Hanson. Detection of surface ships from interception of cyclostationary signature with the cyclic modulation coherence. *IEEE Journal of Oceanic Engineering*, 37(3):478–493, 2012.

Appendix A

Commonly used symbols and statistical quantities

Table A.1 List of symbols

CS	Cyclostationary
HMM	Hidden Markov Model
MLE	Maximum Likelihood Estimation
MAP	Maximum A Posteriori
EM	Expectation Maximization
LLR	Natural Logarithm of the Likelihood Ratio
\mathcal{N}	Normal Distribution
\mathcal{CN}	Circularly-Symmetric Complex Normal Distribution
SC	Spectral Correlation
CMS	Cyclic Modulation Spectrum
Fast-SC	Fast Spectral Correlation

DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
STFT	Short-Time Fourier Transform
SES	Squared Envelope Spectrum
EES	Enhanced Envelope Spectrum
$x(t_n)$	signal of interest
$w[n]$	data window (function of time index n)
$X_w(i, f)$	Gabor coefficient at time index i and frequency f
$X_{STFT}(i, f)$	STFT coefficient at time index i and frequency f
L	signal length
N_w	window length in STFT
N_0	central time index of window
R	window shift in STFT
K	total number of blocks used in spectral estimates
F_s	sampling frequency
t_n	n -th discrete time instant (in s)
τ	time-lag (in s)
T	cyclic period of a cyclostationary signal (in s)
α	cyclic (or modulation) frequency (in Hz)
α_{max}	maximum scrutinizable cyclic frequency (in Hz)
f	spectral (or carrier) frequency (in Hz)

f_k	k -th discrete frequency (in Hz)
$\Delta\alpha$	cyclic frequency resolution in α (in Hz)
Δf	frequency resolution in f (in Hz)
p	index of STFT frequency closest to a given cyclic frequency α
P	index of STFT frequency closest to α_{max}
$R_x(t_n, \tau)$	instantaneous autocorrelation function of signal x
$R_w(\alpha)$	discrete Fourier transform of $ w[n] ^2$
$S_x(\alpha, f)$	Spectral Correlation of signal x
$\gamma_x(\alpha, f)$	Spectral Coherence of signal x
$S_x^{CMS}(\alpha, f)$	Cyclic Modulation Spectrum of signal x
$S_x(\alpha, f; p)$	Scanning Spectral Correlation of signal x
$S_x^{Fast}(\alpha, f)$	Fast Spectral Correlation of signal x
$\gamma_x^{Fast}(\alpha, f)$	Fast Spectral Coherence of signal x
$S_x^{SES}(\alpha)$	Squared Envelope Spectrum of signal x
$S_x^{EES}(\alpha)$	Enhanced Envelope Spectrum of signal x
BPFO	Ballpass frequency, outer race
BPMI	Ballpass frequency, inner race
BSF	Ball (roller) spin frequency
f_{rot}	Shaft rotation speed
FTF	Fundamental train frequency

Table A.2 Commonly used statistical quantities

Moment (central)	Estimator	Measures
2^{nd} moment:		
$\sigma_x^2 = E [X - \mu_x ^2]$	$m_2 = \frac{\sum_{i=1}^N x_i - \bar{x} ^2}{N-1}$	Variance (spread or dispersion)
3^{rd} moment:		
$M_3 = E [X - \mu_x ^3]$	$m_3 = \frac{\sum_{i=1}^N x_i - \bar{x} ^3}{(N-1)\sigma_x^3}$	Skewness (degree of asymmetry)
4^{th} moment:		
$M_4 = E [X - \mu_x ^4]$	$m_4 = \frac{\sum_{i=1}^N x_i - \bar{x} ^4}{(N-1)\sigma_x^4}$	Kurtosis (degree of flattening)

where x_i denotes the i^{th} sample, from 1 to N; \bar{x} denotes the sample mean (which estimates the arithmetic mean, μ_x).

Appendix B

Proof of the mixture distribution with two Gaussian distributions giving rise to a super-Gaussian distribution

Let start by expressing the probability distribution of observation $y(t)$ as a linear mixture of two different Gaussian distributions: $x_1(t) \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $x_2(t) \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Hence, the mixture distribution reads

$$f(y) = \pi f_1(y) + (1 - \pi) f_2(y), \quad (\text{B.1})$$

where π and $(1 - \pi)$ denote the mixture weights (probabilities) corresponding to $x_1(t)$ and $x_2(t)$, respectively.

Without loss of generality, it is assumed that $\mu_1 = \mu_2 = 0$ (as obtained after first centering the signal). In addition, one can obtain the second and fourth central moment of $x_i(t)$ as $E\{X_i^2(t)\} = \sigma_i^2$ and $E\{X_i^4(t)\} = 3\sigma_i^4$ ($i = 1, 2$). Then the fourth central moment of $y(t)$ is estimated by the following:

$$E\{Y^4(t)\} = \int y^4 f(y) dy = \pi \int y^4 f_1(y) dy + (1 - \pi) \int y^4 f_2(y) dy = 3(\pi\sigma_1^4 + (1 - \pi)\sigma_2^4). \quad (\text{B.2})$$

The second central moment of $y(t)$ is

$$E\{Y^2(t)\} = \int y^2 f(y) dy = \pi \int y^2 f_1(y) dy + (1 - \pi) \int y^2 f_2(y) dy = \pi\sigma_1^2 + (1 - \pi)\sigma_2^2. \quad (\text{B.3})$$

Therefore the kurtosis of $Y(t)$ is calculated by

$$Kurt[Y(t)] = \frac{E\{Y^4(t)\}}{(E\{Y^2(t)\})^2} - 3 = 3 \frac{\pi\sigma_1^4 + (1-\pi)\sigma_2^4}{(\pi\sigma_1^2 + (1-\pi)\sigma_2^2)^2} - 3. \quad (\text{B.4})$$

Let further prove the following inequality:

$$\pi\sigma_1^4 + (1-\pi)\sigma_2^4 \geq (\pi\sigma_1^2 + (1-\pi)\sigma_2^2)^2. \quad (\text{B.5})$$

Next, let remind the Lemma of Minkowski's inequality:

Lemma 2 (Hölder) *If $p > 1$ and $q > 1$ are such that*

$$\frac{1}{p} + \frac{1}{q} = 1 \quad (\text{B.6})$$

then for all \mathbf{x} and $\mathbf{y} \in \mathbb{R}^n$ we have

$$\sum_{i=1}^n \|x_i y_i\| \leq \left(\sum_{i=1}^n \|x_i\|^p \right)^{1/p} \left(\sum_{i=1}^n \|y_i\|^q \right)^{1/q}. \quad (\text{B.7})$$

Now, let take the $n = p = q = 2$, $y_1 = \sqrt{\pi}$, $y_2 = \sqrt{1-\pi}$, $x_1 = \sigma_1^2 \sqrt{\pi}$ and $x_2 = \sigma_2^2 \sqrt{1-\pi}$, then one can obtain

$$(\pi\sigma_1^2 + (1-\pi)\sigma_2^2)^2 \leq (\pi\sigma_1^4 + (1-\pi)\sigma_2^4)(\pi + (1-\pi)). \quad (\text{B.8})$$

Since the kurtosis of $Y(t)$ in Eq. B.4 is greater or equal to 0, the mixture distribution with two Gaussian distributions gives rise to a super-Gaussian distribution.

Appendix C

Proof of the inverse gamma distribution giving rise to a conjugate prior for the variance

This is a demonstration of the inverse gamma distribution arising as a conjugate prior for the variance σ^2 of a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with known mean μ . It is noteworthy that this property is under the assumption of an informative prior, as mentioned in footnote 3 in subsection 4.2.

Let start by expressing the likelihood function of observation $y[n]$ as

$$p(y[n] | \mu, \sigma^2) = p(y_1, \dots, y_n | \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \mu)^2}{2\sigma^2}\right\} \quad (\text{C.1})$$

where σ^2 denotes the unknown variance of the normal distribution, and it is assumed to follow an inverse gamma distribution

$$\sigma^2 \sim \text{Inv} - \text{Gamma}(\alpha, \beta) \quad (\text{C.2})$$

with shape parameter α and scale parameter β as the prior. Next, let introduce the posterior probability distribution of the unknown variance σ^2 as

$$p(\sigma^2 | y[n], \mu) \propto p(y[n] | \mu, \sigma^2)p(\sigma^2 | \mu). \quad (\text{C.3})$$

152 Proof of the inverse gamma distribution giving rise to a conjugate prior for the variance

Then the posterior probability density reads

$$p(\sigma^2 | y[n], \mu) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{\sigma^2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{2}\right\} \times \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} \exp\left\{-\frac{\beta}{\sigma^2}\right\} \quad (\text{C.4})$$

where $\Gamma(\cdot)$ denotes the gamma function. Let us ignore constant terms in the prior $\frac{\beta^\alpha}{\Gamma(\alpha)}$ and in the likelihood $\frac{1}{(2\pi)^{n/2}}$. After some manipulations, Eq. C.4 can be expressed as

$$p(\sigma^2 | y[n], \mu) = \text{constant} \times (\sigma^2)^{-(\alpha + \frac{n}{2} + 1)} \exp\left\{-\frac{1}{\sigma^2} \left(\beta + \sum_{i=1}^n \frac{(y_i - \mu)^2}{2}\right)\right\}, \quad (\text{C.5})$$

alternatively, the posterior probability distribution of the unknown variance σ^2 is proportional to

$$p(\sigma^2 | y[n], \mu) = \frac{(\beta + \sum_{i=1}^n \frac{(y_i - \mu)^2}{2})^{\alpha + \frac{n}{2}}}{\Gamma(\alpha + \frac{n}{2})} (\sigma^2)^{-(\alpha + \frac{n}{2} + 1)} \exp\left\{-\frac{1}{\sigma^2} \left(\beta + \sum_{i=1}^n \frac{(y_i - \mu)^2}{2}\right)\right\} \quad (\text{C.6})$$

that arises as an inverse gamma distribution, *Inv - Gamma*(α_{pos} , β_{pos}), with parameters

$$\begin{cases} \alpha_{pos} = \alpha + \frac{n}{2} \\ \beta_{pos} = \beta + \sum_{i=1}^n \frac{(y_i - \mu)^2}{2} \end{cases} \quad (\text{C.7})$$

Appendix D

Tables of results

Table D.1 12 k drive-end bearing fault analysis results; benchmark data sets: P and N categorise only; measurement position: DE only.

Benchmark data set		Fast-SC-based EES	Benchmark Result (M1, M2, M3)
Inner race faults	171	Y2	P1 / Y2 / Y2
	3001	N1	N1 / N1 / N1
	3002	N1	N1 / N1 / N1
	3003	N1	N1 / N1 / N1
	3004	N1	N1 / N1 / N1
Ball faults	118	N1	N1 / N1 / N1
	119	N1	N1 / N2 / N1
	120	N1	N1 / N2 / N1
	121	N1	N1 / N1 / Y2
	185	P2	P2 / P2 / N1
	186	P2	P2 / P2 / P2

Benchmark data set		Fast-SC-based EES	Benchmark Result (M1, M2, M3)
Ball faults	187	P2	N1 / P2 / N1
	188	P2	P2 / P2 / P2
	222	Y2	P1 / Y2 / Y2
	223	Y2	Y2 / Y2 / P2
	224	N1	N1 / N1 / N1
	225	N1	N1 / N1 / N1
Outer race faults	197	Y2	N1 / N1 / Y2
	198	P1	P2 / N2 / N2
	199	Y2	P1 / N2 / N2
	200	P2	N1 / N1 / N1

DE = drive end acceleration; M1-3 = Method 1-3; diagnosis categories: Y = successful, P = partially successful, N = not successful, more details can be found in Table 5.1.

Table D.2 48 k drive-end bearing fault analysis results; benchmark data sets: P and N categorise only; measurement position: DE only.

Benchmark data set		Fast-SC-based EES	Benchmark Result (M1, M2, M3)
Inner race faults	174	N1	N1 / N1 / N1
	175	Y2	P1 / Y2 / Y2
	176	Y2	P1 / Y2 / P1
	177	Y2	P1 / Y2 / N1

Benchmark data set		Fast-SC-based EES	Benchmark Result (M1, M2, M3)
Ball faults	122	P1	N1 / N2 / N1
	123	P1	N1 / N2 / N1
	124	P1	N1 / N2 / N1
	125	P1	N1 / N2 / N1
	189	P2	N1 / P2 / P2
	190	P2	N1 / P2 / N1
	191	P2	N1 / P2 / N1
	192	P2	N1 / N1 / N1
	226	P1	N1 / P1 / N1
	227	P1	P2 / P1 / P2
	228	P2	N1 / N1 / N1
	229	P2	N1 / N1 / N1
Outer race faults	202	Y2	P1 / N1 / P1
	203	Y2	P2 / P2 / P2
	204	Y2	Y2 / N2 / Y2
	262	Y2	P1 / Y2 / Y2
	264	Y2	P1 / Y2 / P1

DE = drive end acceleration; M1-3 = Method 1-3; diagnosis categories: Y = successful, P = partially successful, N = not successful, more details can be found in Table 5.1.

Table D.3 12 k fan-end bearing fault analysis results; benchmark data sets: P and N categorise only; measurement position: DE only.

Benchmark data set		Fast-SC-based EES	Benchmark Result (M1, M2, M3)
Inner race faults	275	Y2	P1 / P1 / N1
	276	Y2	P1 / P1 / N1
	277	Y2	P1 / P1 / P1
	270	Y2	P1 / Y2 / Y2
	271	Y1	P1 / Y1 / Y2
	273	Y1	P1 / Y1 / Y1
Ball faults	282	Y2	P1 / P2 / N1
	283	Y2	P2 / P2 / N1
	284	Y2	N1 / Y2 / P2
	285	Y2	P1 / N1 / P1
	286	P1	P2 / P2 / P2
	287	P1	P2 / P2 / P2
	288	P1	P2 / P1 / P2
	289	P2	P2 / P2 / P2
	290	Y2	N1 / N1 / N1
	291	Y2	P2 / N1 / P2
	292	Y2	N1 / P2 / N1
	293	Y2	N1 / N1 / N1

Benchmark data set		Fast-SC-based EES	Benchmark Result (M1, M2, M3)
Outer race faults	298	P2	N1 / N2 / P2
	299	P1	P2 / P1 / N1
	300	P2	P2 / P2 / N2
	301	P2	N1 / P2 / N1
	302	N1	N1 / N2 / N1
	305	P2	P2 / N1 / N1
	306	N1	N1 / N1 / N1
	307	N1	N1 / N2 / N1

DE = drive end acceleration; M1-3 = Method 1-3; diagnosis categories: Y = successful, P = partially successful, N = not successful, more details can be found in Table 5.1.

Publications

Journal article

1. J. Antoni, **G. Xin**, N. Hamzaoui, "Fast computation of the spectral correlation", *Mechanical Systems and Signal Processing*, 92:248–277, 2017.
2. **G. Xin**, J. Antoni, N. Hamzaoui, "Automatic spectrum matching of repetitive transients based on hidden Markov model ", *Journal of Sound and Vibration*, in revision.
3. **G. Xin**, J. Antoni, N. Hamzaoui, "Extraction of second-order cyclostationary sources by instantaneous power spectrum matching and stochastic model – application to vibration analysis", in preparation.

Conference paper

1. **G. Xin**, J. Antoni, N. Hamzaoui, "Extracting transient signal from strong noise in rolling bearing diagnosis", *Journées des Jeunes Chercheurs en vibrations, Acoustique et Bruit (JJCAB)*, Lyon, France, 2014.
2. **G. Xin**, J. Antoni, N. Hamzaoui, "An Exploring Study of Hidden Markov Model in Rolling Element Bearing Diagnostis", *Surveillance 8*, Roanne, France, 2015.
3. **G. Xin**, J. Antoni, N. Hamzaoui, "Latent Variable Analysis Based on Hidden Markov Model in Rolling Element Bearing Diagnostics", *The 13th French Acoustics Congress (CFA 2016)*, Le Mans, France, 2016.
4. J. Antoni, D. Abboud, **G. Xin**, "Cyclostationarity in Condition Monitoring : 10 years after", *International Conference on Noise and Vibration Engineering (ISMA)*, Leuven, Belgium, 2016.



FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : XIN

DATE de SOUTENANCE : 22/06/2017

Prénoms : Ge

TITRE : Sparse Representations in Vibration-based Rolling Element Bearing Diagnostics

NATURE : Doctorat

Numéro d'ordre : 2017LYSEI051

Ecole doctorale : MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE

Spécialité : Acoustique

RESUME :

Although vibration-based rolling element bearing diagnostics is a very well-developed field, the research on sparse representations of vibration signals is yet new and challenging for machine diagnosis. As a desired property – representation of a signal in highly organized structure – sparsity enables us to reveal the natural signature of singular events embedded in a signal so as to reduce the demand on the user's expertise, even though it involves advanced theory of stochastic processes. In this thesis, several novel methods have been developed so as to serve the industry in rolling element bearing diagnostics.

First, the sparsity-based model is investigated based on the current literature. By discussing the pros and cons of sparse representations, an interpretation of sparse structure in the Bayesian viewpoint is proposed for machinery fault diagnosis.

Second, a new stochastic model is introduced to address this issue: it introduces a hidden variable to indicate the occurrence of the impacts and estimates the spectral content of the corresponding transients together with the spectrum of background noise. The performance is investigated on synthetic signals with a high noise to-signal ratio and also in the case of a mixture of two independent transients. The effectiveness and robustness of the method are also verified on vibration signals measured on a test-bench (gears and bearings). Results are found superior or at least equivalent to those of conventional envelope analysis and fast kurtogram.

Third, a novel scheme for extracting cyclostationary (CS) signals is proposed. It introduces a periodic-variance based stochastic model to recover the CS component in the masking of interfering signals. Of particular interest is the robustness on experimental data sets and superior extraction capability over the conventional Wiener filter. Eventually, these experimental examples evidence its versatile usage on diagnostic analysis of compound signals.

Fourth, a benchmark analysis by using the fast computation of the spectral correlation is provided. This study benefits from a big data set supplied by the Case Western Reserve University (CWRU) Bearing Data Center. Diagnostic results have moved forward the benchmark study of the CWRU data set -- by uncovering its own unique characteristics.

MOTS-CLÉS : Machinery fault diagnosis, Sparse representations, Hidden Markov model, Cyclostationary signals, Spectral correlation

Laboratoire (s) de recherche : Laboratoire Vibrations Acoustique (LVA) – INSA Lyon

Directeur de thèse : ANTONI Jérôme, Professeur des Universités
HAMZAOUI Nacer, Professeur des Universités

Président de jury : DRON Jean-Paul, Professeur des Universités

Composition du jury : CHIEMENTIN Xavier (Rapporteur), CORBIER Christophe (Rapporteur), DRON Jean-Paul (Examinateur), KAFTANDJIAN Valérie (Examinatrice), ANTONI Jérôme (Directeur de thèse), HAMZAOUI Nacer (Co-directeur de thèse).