CONDITION MONITORING OF MECHANICAL DRIVES:

FEATURE EXTRACTION AND FAULT DIAGNOSIS METHODS

Pavle Boškoski

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia, December 2011

Evaluation Board:
Prof. Dr. Stanislav Strmčnik, Chairman, Jožef Stefan Institute, Ljubljana, Slovenia
Prof. Dr. Len Gelman, Member, Cranfield University, Cranfield, Bedfordshire, United Kingdom
Prof. Dr. Jožef Vižintin, Member, Faculty of Mechanical Engineering, University of Ljubljana, Slovenia

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL
Ljubljana, Slovenia

Pavle Boškoski

Condition monitoring of mechanical drives:

Feature extraction and fault diagnosis methods

Doctoral Dissertation

Spremljanje stanja pogonskih sklopov:

Postopki generiranja značilk in diagnosticiranja poškodb

Doktorska disertacija

Supervisor: Prof. Dr. Đani Juričić

Co-supervisor: Prof. Dr. Mile Stankovski

Ljubljana, Slovenia, December 2011

Contents

List of abbreviations

ANN artificial neural networks.

BPFI bearing pass frequency inner race.

BPFO bearing pass frequency outer race.

BSF ball spin frequency.

c.r.v complex random variable.

CBM condition-based maintenance.

CM condition monitoring.

CWT continuous wavelet transform.

DSS decision support system.

DWT discrete wavelet transform.

EC electronically commutated.

EHL elasto-hydrodynamic lubrication.

ER evidential reasoning.

FDI fault detection and indetification.

FTF fundamental train frequency.

GMF gear mesh frequency.

i.i.d independent and identically distributed.

ICF information cost function.

IP instantaneous power.

MADM multiple attribute decision making.

PDF probability density function.

PHM prognostics and health management.

PWM pulse-width modulation.

r.v. random variable.

RMS root-mean square.

RUL remaining useful life.

RVM relevance vector machines.

SCD spectral correlation density.

SCOH spectral coherence function.

SK spectral kurtosis.

SNR signal-to-noise ratio.

STFT short-time Fourier transform.

SVM support vector machines.

TBM transferable belief model.

TSA time-synchronous averaging.

WPT wavelet packet transform.

WSS wide sense stationary.

WVD Wigner-Ville distribution.

Abstract

Mechanical drives are the most ubiquitous items of equipment in almost all industrial branches. Wear, excessive operational loads or errors in assembly can cause premature unexpected failures resulting in partial or total production downtime, damaged equipment or even loss of lives. Proper condition monitoring of such equipment is therefore of great practical importance.

The field of condition monitoring of mechanical drives is very well developed. Over the last decades, a large number of articles addressing every segment in the field were published: mathematical models describing various faults; signal processing methods for extraction of characteristic fault features and estimation methods for the remaining useful life of the monitored drive. Notwithstanding the significant results in every segment, several issues still remain open. Probably the most notable are robust fault detection and isolation and prognostics under variable and presumably unknown operating conditions. This dissertation addresses the problem of robust feature extraction under speed and load variations. Two novel techniques are suggested: one utilising the probability density of vibration envelope signal and the other attempting to identify specific vibrational patterns using the theory of temporal point processes. On top of that, this thesis presents a possible solution to the problems of fault isolation and evaluation based on evidential reasoning.

The analysis of the envelope distribution was done using wavelet packet coefficients of the vibration signals. Based on these coefficients, alterations in the distribution were quantified using Rényi entropy and α-Jensen divergence. Features based on these quantifiers turned to be sufficiently sensitive to the changes in the condition and, in the same time, they tend to be robust to the variations in the operating regime.

Temporal point processes were employed to model the repetitive impacts patterns that occur in cases of bearing surface damage. The analysis showed that these impacts can be modelled as temporal point process with Inverse Gaussian interevent distribution. The analysis of such a point process provides an explanation why Rényi entropy quantifiers are suitable choice for fault detection under variable operating conditions.

These two approaches greatly simplify the fault detection process since they require no information about the operating conditions. In the same time, the employed methods impose no limitations on the statistical properties of the analysed signals, hence making them applicable both for stationary as well as for non-stationary signals. Fault detection capabilities of the proposed methods are demonstrated on a two-stage gearbox operating under different rotational speeds and loads with various artificially built-in mechanical faults.

Finally, based on information contained in the calculated feature set, an evidential reasoning based solution for fault isolation and evaluation is proposed. This methodology mimics the reasoning process of the maintenance personnel thus making it intuitive and open for seamless integration of experts knowledge. The method fuses information contained in the feature set into a ranked list of possible faults as well as an single overall assessment of the utility of the monitored system. A prototype version of the system is validated on a test batch of 130 electronically commutated motors, demonstrating high diagnostic resolution and accuracy.

1 Introduction

Effective exploitation of today's complex industrial systems requires efficient maintenance techniques. These techniques provide means for maximising availability and reliability of industrial assets as well as preserving their optimal conditions. Despite the obvious benefits, these techniques are actively applied only in a limited number of highly sophisticated and expensive systems.

Surveys during the last 20 years reveal that direct maintenance costs in European industry amount to 4%–8% of the whole companies' income (Muller et al., 2007). In the power generation sector these figures go even higher, up to 11%. The most surprising is the fact that one-third to one-half of this expenditure is wasted through ineffective maintenance. In addition, the indirect costs caused by degraded product quality, reduced production efficiency, loss of customers etc. are at least of the same range of magnitude as direct costs.

These dire figures result mainly from inappropriate maintenance strategies that are still largely prevailing nowadays in the form of reactive and preventive/periodic maintenance. The first strategy, as its name says, takes action after a failure has occurred. Uncontrolled evolution of the fault leads towards unscheduled interruptions, and requires significant amount of spare parts to be constantly held on stock. The latter strategy employs scheduled maintenance actions that determine the system's condition as well as replacement of the critical components regardless of their condition. Although this strategy is the most common one, it suffers from two significant drawbacks:

- some of the scheduled maintenance actions including prescribed replacement of system components, is performed regardless of the component's condition and
- more importantly, at the end of the day, after the maintenance action, the systems keep running without permanent monitoring up to the next maintenance action.

Due to these drawbacks, modern manufacturing is now moving to novel and versatile maintenance strategies referred to as condition-based maintenance (CBM) and prognostics and health management (PHM). These strategies are based on synergy of many disciplines such as signal processing, computing, micro-electronics, communication and information sciences.

CBM strategy relies on data continuously acquired from the monitored machine, which serve to infer the machine's current condition (Vachtsevanos et al., 2006). Results of this process can be used to schedule maintenance actions prior to failure. PHM focuses on prediction of the remaining useful life (RUL) of the monitored machine, based on the available condition assessment. The synergy of these two segments results into the so-called CBM/PHM maintenance strategy that is expected to outperform all previous maintenance approaches by increasing the availability and reliability of the monitored systems and simultaneously lowering the operational and support costs. The scope of this thesis is focused on the CBM segment.

1.1 CBM system structure

The role of condition monitoring (CM) is to timely detect onset of failure, localise the root-cause and, possibly, trend its progression over time. This process utilises information gathered from physical sensors in combination with machine-related prior knowledge such as geometry of the mechanical parts, lubrication quality etc. Based on that CM procedures map the actual physical measurements into the health assessment indices of the system. CM usually consists of a sequence of tasks, as shown in Figure 1.1 (Isermann and Ballé, 1997):

Data acquisition of the sensor readings,

Feature extraction, which means calculation of special condition indicators from raw data; the indicators represent the fingerprint of the underlying fault,

Feature evaluation, which consists of associating a qualitative status to the features.

Fault isolation Determination of the kind, location and time of fault occurrence.

Fault identification Estimation of the size and time evolution of the fault.

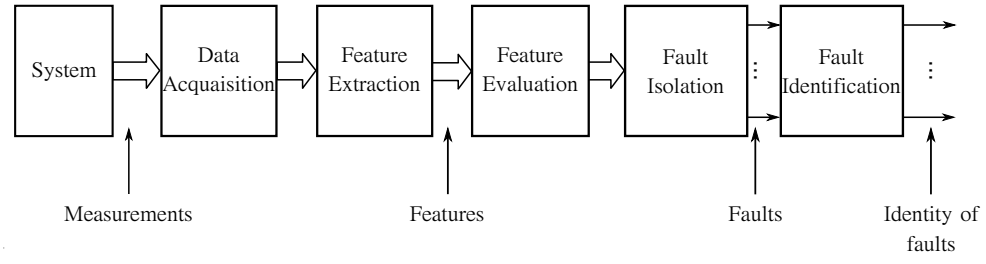


Figure 1.1: Standard fault diagnosis process (Rakar et al., 1999)

The effectiveness of the CM procedure above is assessed through several aspects (Gertler, 1998):

Sensitivity the ability to detect faults of sufficiently small size i.e. in its incipient phase;

Isolability a measure of the system's ability to distinguish between certain failure modes;

Accuracy is attribute indicating that the true fault makes part of the list of the suspected faults;

Reaction speed the ability to detect faults with a reasonably small delay; and

Robustness the ability to operate properly in a presence of noise, disturbances and modelling errors, with fewest possible false alarms.

1.2 Condition monitoring of mechanical drives

In the context of mechanical drives and rotational machinery only a restricted set of specific mechanical faults comprise the most probable causes of mechanical failure. According to early surveys (Albrecht et al., 1986; reliability working group MRWG, 1985a,b,c), bearing faults represent the most common cause for mechanical failure. Besides bearings, substantial percentage of mechanical failures are due to gear faults. The percentage of gear failures is especially high in gearboxes installed in wind turbines (Crabtree, 2010). Finally, the last group of possible causes for mechanical failures are damages that evolve due to improper initial assembly. A rich set of diagnostic methods tailored to deal with the faults mentioned above have been proposed in the literature. What follows is a brief overview of the most relevant ideas.

1.2.1 Feature extraction

The feature extraction task for condition monitoring of mechanical drives predominantly relies on analysis of vibration, sound and acoustic emission signals. The feature extraction techniques applied to these signals can be split into three main categories: time-domain analysis, frequency-domain analysis and time-frequency analysis (Jardine et al., 2006).

The features extracted by time-domain analysis reflect particular properties of the signal such as: root-mean square (RMS), peak-to-peak value, crest factor, kurtosis etc. More sophisticated time-domain methods include time-synchronous averaging (TSA) (Stander and Heyns, 2005; Zhan et al., 2006), ARMA modelling etc. These so-called time-domain features can serve solely as a general indicator about the signal's properties, and are unable to provide any deeper insight and more precise diagnostic information.

Commonly condition monitoring of mechanical drives is performed when the monitored drives are operating under constant and known load and speed. In such cases, faults tend to generate periodically reoccurring vibrational patterns. Consequently, the majority of the feature extraction approaches rely on spectral analysis of the acquired signals (Randall, 2002). The prevailing method is the spectral analysis of the signal's envelope (Ho and Randall, 2000). The main difficulty in this process is selection of the carrier frequency (Boashash, 1992). By addressing this issue, Antoni and Randall (2006) developed a method for detecting the most appropriate frequency band for the calculation of the envelope spectrum. Randall et al. (2001) showed that the same information contained in the signal's envelope can be extracted by analysing the signal's spectral correlations. Furthermore, Antoni et al. (2004) showed that these spectral correlations indicate that the generated vibration signals exhibit pseudo-cyclostationary characteristics. Hence, they defined an approach for bearing fault detection fully described in the framework of cyclostationary random processes. Despite the wide application, the spectral analysis methods suffer from one major deficiency, i.e. they are inapplicable for non-stationary signals.

The third category, time-frequency analysis, overcomes the problems related to non-stationary signals by investigating the signal's behaviour simultaneously in both time and frequency domain. Initially, the predominant tools were short-time Fourier transform (STFT) and Wigner-Ville distribution (WVD) with numerous applications. However, due to its superior characteristics, the wavelet transform has recently become the most commonly used method for time-frequency analysis. Staszewski (1998) was one of the first authors to apply wavelet transform for gear fault detection. Besides gears, the wavelet transform was successfully applied for bearing fault detection by numerous authors like (Feng and Schlindwein, 2009; Kar and Mohanty, 2006; Rubini and Meneghetti, 2001). Peng and Chu (2004) and Jardine et al. (2006) compiled a thorough review of the recent advances in wavelet analysis. In parallel with the wavelet transform, the appearance of Hilbert-Huang transform (HHT) developed by Huang et al. (1998) spun a brand new thread of methods based on the analysis the time-frequency characteristics of its intrinsic mode functions (IMF). This approach was applied for bearing fault detection (Peng et al., 2005; Yu et al., 2006) as well as for gear fault detection (Louitridis, 2004; Ricci and Pennacchi, 2011). Despite such a variety of methods, capable of analysing non-stationary signals, they are mainly employed for features extraction when the monitored drives operate under constant and known operating conditions. Thus the full potential of time-frequency methods is not properly exploited.

Feature extraction under variable operating conditions. Constant operating conditions are appealing since changes in the values of the extracted features can be directly

related to changes in the drive's condition. However, in reality, mechanical drives rarely operate under constant operating conditions and consequently the features extracted by these methods are ineffective for the process of condition monitoring, since both operating conditions and changes in the drive's state influence their value. Thus, the specification of features that are robust to variations in the operating conditions can be regarded as the most challenging task in the process of feature extraction.

In the small number of papers addressing the issue of feature extraction under variable operating conditions, the proposed solutions rely on precise information about operating speed and loads. Usually, these information are directly used in the process of calculating feature values. Order tracking is such a method, capable of coping with speed variations by observing the signals in angular domain (Stander and Heyns, 2005; Zhan et al., 2006). There are also cases when the analysis of higher order spectra is applied for the detection of various bearing faults under different load conditions (Parker et al., 2000). Bartelmus and Zimroz (2009) successfully performed fault detection in multi-stage gearboxes by taking into account information about variations in speed and load. On the other hand, there are approaches that exploit the non-stationary character of the vibration signals generated by drives operating under variable conditions. In that manner, Poulimenos and Fassois (2006) provided a comprehensive analysis of non-stationary vibration signals in time domain by specifying a set of different parametric methods for modelling non-stationary vibration signals acquired from a single vibration sensor. Padovese (2004) presented four time-frequency approaches for analysing transient signals. These approaches are based on autoregressive models applied for calculating time–frequency characteristics of the observed signal as well as an estimation of its power density. Baydar and Ball (2000) performed detection of gear deterioration under different loads using instantaneous power spectrum by employing Wigner-Ville distribution. They successfully performed fault detection of gear faults irrespective of the variable operating conditions. Similarly, Fan and Zuo (2006) combined Hilbert and wavelet packet transforms for detection of gear faults with the ability to handle non-stationary signals. Finally, there are methods that exploit some statistical properties of the acquired vibrations like instantaneous power spectrum (Chen et al., 2005; Wang et al., 2010).

Although the proposed solutions give satisfactory results, they heavily depend on accurate measurements of the current speed and load of the monitored drives. Such a requirement significantly limits the applicability of these solutions because accurate speed and load measurements are rarely available beyond the area of highly sophisticated machinery. Therefore, one of the main goals of this thesis is the development of features whose values will be sufficiently robust to variations in the operating conditions and in the same time sufficiently sensitive to changes in the drive's condition. The extraction process

of such features will rely neither on information about the current operating conditions nor on any physical data regarding the mechanical drive, such as gear characteristics, bearing dimensions etc.

1.2.2 Fault evaluation and isolation

Generally, fault evaluation and isolation tasks perform a comparison between the specific information contained in the extracted features with a anticipated set of possible faults. Two main trends can be identified in the literature. The first one relies on machine learning methods, in particular pattern recognition (Jardine et al., 2006). The second emerging trend introduces the so-called maintenance decision support systems which attempt to mimic the decision making process of the maintenance personnel (Vachtsevanos et al., 2006).

Generally, pattern recognition methods assign each input vector one finite number, and the process is known as classification. In cases when the number of choices is unknown, the classification problem becomes clustering problem (Russell and Norvig, 2009). The most exploited classification method are the artificial neural networks (ANN) (Worden et al., 2011). A significant number of the papers based on this approach were used for gear fault diagnosis (Rafiee et al., 2007; Samanta, 2004; Schlechtingen and Santos, 2011), bearing diagnosis (Castejón et al., 2010; Huang et al., 2007). Despite the successful implementations of ANN, they lack the ability of incorporating expert knowledge. This space is filled with the application of neuro-fuzzy based systems capable of encoding prior knowledge into the ANN structure. Similarly to the case of ANN, there are numerous applications that utilise fuzzy logic models (Sugumaran and Ramachandran, 2007). Recently support vector machines (SVM) and relevance vector machines (RVM) have been used to optimize the boundary between clusters (Caesarendra et al., 2010; Jack and Nandi, 2001; Widodo and Yang, 2007). An emerging method gaining popularity in the field of fault isolation are hidden Markov models (HMM). Initially, they were applied in the areas like speech and handwriting recognition. However, there are now successful examples of the application of HMM for fault isolation (Ge et al., 2004; Miao and Makis, 2007).

Although the listed machine learning methods are often used in practice, they suffer from several limitations. Apart from fuzzy systems, the mentioned algorithms require sufficiently high amount of training data. Additionally, all these methods regard the feature evaluation and isolation process as pure classification problem, ignoring the fact that the most needed information is the ranking of the overall condition as well as the most probable causes for its deterioration. Possibly the biggest problem is that the training process usually comprises only a limited set of faults, and the performance of these methods is unclear for the cases of unforeseen faults.

In terms of decision support system (DSS), the fault evaluation and isolation tasks are regarded as multi-attribute decision making problem, where attributes are the calculated features. Based on their values and a set of decision rules, the output of this approach is a ranked list of most probable faults. Examples of such systems include the transferable belief model (TBM) (Rakar et al., 1999), evidential reasoning (Boškoski et al., 2011; Wang and Elhag, 2008), Bayesian networks (Karlsson et al., 2008). There are also cases of fuzzy logic based DSS (Wang and Kanneg, 2009). Such implementation unifies the classification and ranking capabilities into a single operation. The resulting systems are able to imitate the decision making patterns of a human operator by providing a ranked list of possible failure alternatives as well as generating an overall assessment (utilisation rank) of the system as a whole. Therefore, this thesis presents one possible solution in the form of multi-attribute decision support system based on evidential reasoning.

1.3 Purpose and contribution of the dissertation

The field of condition monitoring of rotational machines has received significant attention in the past decades. The vast majority of contributions address the initial segment of the diagnostic process, i.e. feature extraction. Even in this segment the bulk of the results are limited to the cases when the monitored systems operate under constant and known operating conditions. Conversely, the main purpose of this dissertation is to present a complete approach that addresses both fault detection and identification, with special attention to the cases when the monitored system operates under variable and unknown operating conditions.

1.3.1 Problem statement

Prior to stating the problem, let us remind the concept of feature, which is essential in this context. Namely, by definition, feature is a variable which is calculated from available measured signals so that its value depends only on the condition of the underlying system and is independent on any other process variable. The posing question is: can we define features whose values are not sensitive to the variations of the operating condition but sensitive solely to the changes in the system's condition? Contingent the statement is true, the feature values will be comparable even in cases when they were acquired under different operating conditions. The main purpose of this thesis is to show that this goal is feasible by employing methods that assess some general statistical properties of the acquired signals without any prior knowledge about physical characteristics of the mechanical drive and, more importantly, without the requirement of constant and known operating conditions. Hopefully, the answer to this challenge is positive in a sense that

diagnostic resolution can be at least partially achieved in spite of missing knowledge of operating conditions.

1.3.2 Original scientific contributions

The main original scientific contributions of this dissertation are:

- **New model of bearing faults from the perspective of point processes** Bearing faults are characterised by repetitive impacts between the rolling elements and the damaged surface. The time intervals between two impacts are directly related with the type and location of the surface fault. In this dissertation these time intervals are interpreted as realisation of point processes with Inverse Gaussian distribution of the interevent times (Boškoski and Juričić, 2011).

- **Detection of improperly lubricated bearings** Nearly all methods for bearing fault detection, published so far, address surface mechanical faults. Only a handful of papers address the issue of lubrication faults. The detection of improperly lubricated bearings from vibration patterns is a difficult task especially when notoriously short operating records are available. The thesis proposes a novel approach to the detection of improperly lubricated bearings by monitoring the spectral components associated with the fundamental cage frequency and ball spin frequency (Boškoski et al., 2010b). As changes in these spectral components caused by improper lubrication are minute, the method relies on cyclostationary analysis and spectral kurtosis for the selection of a frequency band in which variations in vibration patterns are most distinct. The proposed approach was successfully tested on a production line of electronically commutated motors as an operating industrial prototype.

- **Development of a methodology for gear and bearing fault diagnosis under slow varying operating conditions** The majority of the approaches to fault diagnosis of rotational machines and drives assume constant and known operating conditions. As these assumptions are often violated in practice, this dissertation presents an approach to the diagnosis of gearboxes in presumably non-stationary and unknown operating conditions. The approach makes use of information indices derived from the coefficients of the wavelet packet transform of the measured vibration records (Boškoski et al., 2010a). These indices encompass characteristic statistical properties of the instantaneous power fluctuations of the generated vibrations signals that are largely unaffected by changes in the operating conditions.

- **A decision making approach to overall condition assessment implemented onto an industrial prototype** Features are just indicators that should be used in

a systematic way to assess the overall system's condition. The problem is addressed in the context of end-quality assessment of electronically commutated (EC) motors. The overall quality of an EC motor is equivalent to the overall condition of the device. Hence, the problem of quality assessment is regarded as multi-attribute decision making problem, with attributes defined as values of specially designed features (Boškoski et al., 2011). Unlike many other similar systems based on artificial intelligence methods, the proposed system can incorporate prior knowledge, needed for the definition of the corresponding quality thresholds, and it does not necessarily require any learning process.

The idea is implemented on an industrial prototype for quality assessment of EC motors at the end of the production line.

1.4 Dissertation structure

The organisation of the remaining of the text is as follows. In Chapter 2 a brief theoretical background of stochastic processes, is delivered first. Special attention in Chapter 3 is devoted to the specific subclass of stochastic processes called point processes. These processes are of particular importance for modelling random events occurring in cases of surface bearing faults. Chapter 4 introduces the wavelet transform and wavelet based indices as a general tool for processing non-stationary signals. This chapter defines the so-called information cost functions, that will serve as the main features in the process of fault detection.

Before going deeper in the fault diagnosis process, a general overview of the signal models for mechanical failures is given in Chapter 5. Chapter 6 focuses on fault diagnosis performed under constant and known operating conditions. The main results are presented along with the case study of fault diagnosis of electronically commutated motors. The applicability of point processes and statistical properties of the envelope of vibration signals for fault detection under variable and presumably unknown operating conditions are examined in Chapter 7. These concepts are applied on two different gearbox systems operating under variable conditions. Finally, the Chapter 8 presents an approach that fuses information contained in the calculated features into a single overall assessment of the machine condition. The presented system is based on evidential reasoning multi attribute decision theory, and was developed for end-quality control in manufacturing line of electronically commutated motors.

The final Chapter 9 contains a specific set of remarks that have to be further examined and researched. It also gives a glimpse of the way the results of this work can be generalised to other areas of fault diagnosis comprising of far more complicated examples than the

cases of rotational machines.

2 Overview of stochastic processes

A stochastic process, is a family of random variables, indexed by a parameter t which usually represents a time-sequence, however it can be generalised to any multi-dimensional sequence of interest (Papoulis, 1991). Throughout this thesis stochastic processes will be analysed in respect to time-sequence t.

For a specific t, $x(t)$ can be regarded as random variable (r.v.) with the distribution function

$$F(x,t) = \Pr\{x(t) \leq x\}.$$

However the complete characterisation of a stochastic process is contained in its n-order distribution

$$F(x_1,\ldots,x_n;t_1,\ldots,t_n) = \Pr\{x(t_1) \leq x_1,\ldots,x(t_n) \leq x_n\} \tag{2.1}$$

Depending on the properties of the nth-order distribution function (2.1) various classes of stochastic processes can be defined (see Figure 2.1). The subsequent sections present the most important classes of stochastic processes accompanied by the minimal required mathematical framework for their analysis.

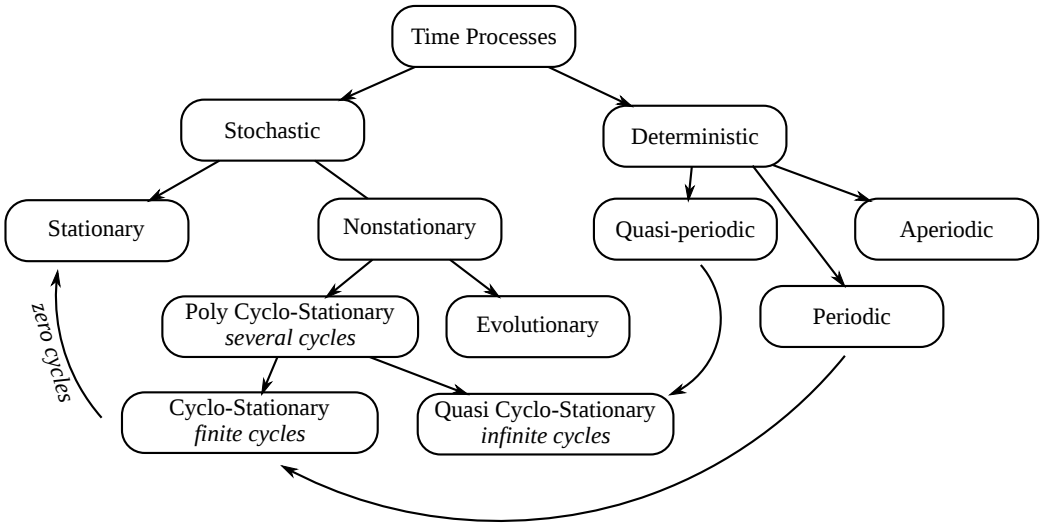


Figure 2.1: Ontology of signals

2.1 Stationary random processes

Stationary random processes form a class of random processes whose statistical characteristics remain constant over time. If a random process $x(t)$ belongs to this class, it implies (Papoulis, 1991)

$$F(x_1, \ldots, x_n; t_1, \ldots, t_n) \equiv F(x_1, \ldots, x_n; t_1 + c, \ldots, t_n + c). \tag{2.2}$$

The relation (2.2) states that the statistical characteristics of a stationary random process are time-shift invariant.

Wide-sense stationary random processes For practical purposes, the stationarity of random processes is checked only up to some finite order m. When the stationarity is checked and fulfilled for $m = 2$, then the corresponding process is referred to as wide sense stationary (WSS) process. For a WSS process, its mean μ_x is constant and its autocorrelation ρ_x depends only on the interval $\tau = t_2 - t_1$ and not on the location of t_1 and t_2 along the time axis:

$$\begin{aligned} E\{x(t)\} &= \mu_x(t) \\ E\left\{x(t+\tau)x^*(t)\right\} &= \rho_x(\tau), \end{aligned} \tag{2.3}$$

where $x^*(t)$ is complex conjugate of $x(t)$.

2.1.1 Spectral analysis

Spectral analysis of deterministic signals

The most common approach to signal analysis is to look at its spectral representation. From a physical viewpoint this representation describes the signal's energy(power) distribution over the observed frequency range. This representation is appealing since it decomposes the original signal into a sum of sine and cosine terms with appropriate amplitudes and frequencies. The decomposition is usually performed by applying Fourier transform to the original signal.

Fourier transform of a function $f(t)$ exists if the function $f(t) \in \mathbb{L}^2$, i.e.

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt < \infty. \tag{2.4}$$

Generally, the relation between a function $f(t)$, that satisfies the condition (2.4), and its spectral representation $\hat{f}(\omega)$ is given in the form of Fourier-Stieltjes transform:

$$f(t) = \int_{-\infty}^{+\infty} e^{j\omega t} d\hat{f}(\omega), \tag{2.5}$$

where $\hat{f}(\omega)$ is a complex valued function called Fourier-Stieltjes transform of $f(t)$. For cases when $\hat{f}(\omega)$ is differentiable, the Fourier-Stieltjes integration (2.5) reduces to Lebesgue integral, since $d\hat{f}(\omega) = \hat{f}(\omega)d\omega$. Such is the case when $f(t)$ is non-periodic. On the other hand, when $f(t)$ is periodic, $d\hat{f}(\omega)$ becomes discrete function with the following form

$$d\hat{f}(\omega) = \sum_{r=-\infty}^{\infty} A_r \delta(\omega - r\omega_0), \quad \omega_0 = \frac{2\pi}{T},$$

where T is the period of the signal.

Spectral analysis of stochastic processes

The main difficulty in spectral analysis of stochastic processes is that they can not be represented by an ordinary function, but rather by an infinite set of realisations of the process. Even a single realisation of a random process $x(t)$ can not be directly analysed neither by Fourier series nor by Fourier transform. The main problem with the former is that we can not assume that $x(t)$ has any periodic structure. Similarly, for the latter case, we can not assume that $x(t)$ will decay as time evolves, thus the process breaches the condition (2.4).

Consequently, the obvious choice is to limit the observation of $x(t)$ to a finite interval $[-T, T]$ i.e. to define a new version $x_T(t)$ as:

$$x_T(t) = \begin{cases} x(t), & -T \leq t \leq T \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

Thus, according to (2.5), $\hat{x}(\omega)$ can be defined as the Fourier-Stieltjes transform of $x_T(t)$. Since $x_T(t)$ is defined on a finite interval, instead of analysing signal's energy/frequency distribution, the analysis will focus on signal's power/frequency distribution. Consequently, if the limit

$$\lim_{T\to\infty} \frac{|\hat{x}_T(\omega)|^2}{2T}$$

exists it is interpreted as power density function. The values of $\hat{x}_T(\omega)$ will be valid only for a particular realisation of $x(t)$. In order to define a quantity that will characterise the whole stochastic process, then the obvious step is to calculate the average over all realisations. Hence, this leads to (Priestley, 1981)

$$\Phi(\omega) = \lim_{T\to\infty} \left[E\left\{ \frac{|\hat{x}_T(\omega)|^2}{2T} \right\} \right]. \tag{2.7}$$

The function $\Phi(\omega)$ is power spectral density of $x(t)$, and the values $\Phi(\omega)\, d\omega$ represent the average contribution to the total power in the frequency range between ω and $\omega + d\omega$.

Calculating $\Phi(\omega)$ through the formal definition (2.7) can be a formidable task. Hopefully, the same function $\Phi(\omega)$ can be calculated using the autocorrelation function $\rho_x(\tau)$ of the process $x(t)$. This relation is given with the following theorem.

Theorem 2.1 (Wiener-Khinchine Theorem (Priestley, 1981))**.** *A necessary and sufficient condition for $\rho(\tau)$ to be the autocorrelation function of some stochastically continuous stationary process $x(t)$ is that there exists a function $H(\omega)$, having the properties of a probability distribution function on $(-\infty, \infty)$ (i.e. $H(-\infty) = 0$, $H(\infty) = 1$ and $H(\omega)$ is non-decreasing), such that for all τ, $\rho(\tau)$ may be expressed as*

$$\rho(\tau) = \int_{-\infty}^{+\infty} e^{j\omega\tau} dH(\omega).$$

Function $H(\omega)$ is referred to as the integrated spectrum or spectral distribution function. It is related to the *power spectral density* (2.7) as follows

$$H(\omega) = \int_{-\infty}^{\omega} \Phi(\theta) d\theta. \tag{2.8}$$

The integrated spectrum (2.8) serves as the basis for spectral representation of a stationary stochastic process.

Theorem 2.2. *A zero mean stationary stochastic process $x(t)$ can be written in the form*

$$x(t) = \int_{-\infty}^{\infty} e^{j\omega t} dZ(\omega), \tag{2.9}$$

where $Z(\omega)$ has the following properties:

i) $E[dZ(\omega)] = 0$

ii) $E[|dZ(\omega)|^2] = dH(\omega)$

iii) $E[dZ(\omega)dZ^*(\omega')] = 0$ *for* $\omega \neq \omega'$

The proof of the theorem can be found in (Priestley, 1981, Chapter 4). Theorem 2.2 shows that any zero-mean stationary random process can be regarded as a sum of sine and cosine components with random amplitudes $dZ(\omega)$. Furthermore, it shows that the amplitudes at different ω are uncorrelated. Finally, the mean squared value of the complex valued random process $Z(\omega)$ at frequency ω is directly related to the increment of the integrated spectrum of the corresponding stochastic process $x(t)$.

2.1.2 Circular random variables

For the stationary random process $x(t)$ defined with (2.9), the corresponding complex valued spectral process $Z(\omega)$ has specific statistical properties that deserves further clarification. In order to describe these properties firstly we have to provide the notion of complex random variable.

For each ω, $Z(\omega)$ is a complex random variable (c.r.v). As such in can be decomposed as $Z(\omega) = X(\omega) + jY(\omega)$, where both X and Y represent real random variables. The distribution of the c.r.v is defined through the joint distribution of its real and imaginary parts i.e. $F_{\Re(Z),\Im(Z)}(x,y) = \Pr\{X \leq x, Y \leq y\}$. Consequently, the probability density function of c.r.v Z is defined as $f(z) \equiv f(x,y)$. Hence, a c.r.v $Z(\omega)$ is said to be normal (or Gaussian) of both X and Y are a pair of jointly Gaussian r.v. (Picinbono, 1994).

The characteristic function of the c.r.v Z is defined as (Eriksson et al., 2009)

$$\begin{aligned} \varphi_Z(z) \triangleq \varphi_{X,Y}(x,y) &= E_{X,Y}\left\{\exp\left(j(xX+yY)\right)\right\} \\ &= E\left\{\exp\left(\frac{j}{2}(z^*Z + zZ^*)\right)\right\} \end{aligned} \tag{2.10}$$

where $z = x + jy$.

As in the case of real r.v. the series expansion of the characteristic function (2.10) defines the moments of the c.r.v Z (Amblard et al., 1996)

$$\varphi_Z(z) = \sum_{n=1}^{\infty} \frac{1}{n!}\frac{j^n}{2^n} \sum_{m=0}^{n} \binom{n}{m} z^{n-m} u^{*m} E\left\{Z^{*n-m} Z^m\right\}.$$

The last term, $E\left\{Z^{*n-m}Z^m\right\}$, shows that there are $n+1$ different nth-order moments for a c.r.v Z.

Definition 2.1. *A complex random variable Z is circular if for any $\theta \in [0, 2\pi)$ the distribution of $Ze^{j\theta}$ is the same with the distribution of Z.*

Properties of circular complex random variables

i) A c.r.v Z is circular if and only it its probability density function depends only on the product ZZ^*, where Z^* represents the complex conjugate;

ii) A c.r.v Z is circular if and only it its characteristic function $\varphi_Z(w)$ is a function only of the product ww^*;

iii) For the case where the moments of circular c.r.v Z exist, the only non-null moments are those constructed with the same power in Z and Z^*.

Circularity of frequency components of stationary random process

The random amplitudes $dZ(\omega)$ defined by Theorem 2.2 are circular complex random variables. The circularity can be easily shown by testing whether the two c.r.v $dZ(omega)$ and $dZ(\omega)e^{-j\omega\tau}$ have the same probability distribution.

According to (2.9) $dZ(\omega)$ is associated with a random process $x(t)$. A time shifted version of the same process $x(t-\tau)$ would have the following spectral representation

$$x(t-\tau) = \int_{-\infty}^{\infty} e^{j\omega t} e^{-j\omega\tau} dZ(\omega).$$

Due to the stationarity of $x(t)$, any shift in the time has no influence on the statistical properties of the signal. Consequently, both spectral representations should have the same probability distribution. Hence, the c.r.v $dZ(\omega)$ associated with the spectral representation of a stationary random process are in essence circular complex variables.

2.2 Cyclostationary processes

A random process $x(t)$ whose statistical properties are invariant to origin shifts that can be represented as mT, where $m \in \mathbb{Z}$, are called *cyclostationary* processes with period T. The cyclostationary processes are in essence non-stationary, since their statistics vary in time. However, as these variations are specified there are tools that enable proper analysis of such signals.

The cyclostationary signals are common in telecommunications, telemetry, radar, and sonar applications, multiplexing, and coding operations (Gardner et al., 2006). In the area of mechanical systems such signals are produced by rotating parts (Randall et al., 2001).

2.2.1 Definition of a cyclostationary process

The random process $x(t)$ is said to be n^{th}-order cyclostationary with period T, if its n^{th}-order distribution satisfies

$$F(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n) = F(x_1, x_2, \cdots, x_n; t_1 + mT, t_2 + mT, \cdots, t_n + mT),$$

where $m \in \mathbb{Z}$. Similarly like in the case of stationary processes, in practice the analysis is limited to the so-called wide-sense cyclostationary processes, i.e. processes whose mean and autocorrelation, defined by (2.3), are periodic with period T:

$$\begin{aligned} \mu_x(t) &= \mu_x(t + mT) \\ \rho_x(t, \tau) &= \rho_x(t + mT, \tau). \end{aligned} \tag{2.11}$$

In addition to the periodicity of mean and autocorrelation function, the second-order cyclostationary processes can be viewed as random processes characterized by the correlation between a set of frequency components. The analysis of cyclostationary signals through this mechanism of frequency correlation enables the derivation of the spectral correlation density and spectral coherence, the two most significant signal processing concepts for second-order cyclostationary processes.

2.2.2 Spectral correlation density

As the autocorrelation function (2.11) of a cyclostationary process is periodic, with period T, it can be expanded using the Fourier series (Gardner et al., 2006)

$$\rho_x(t,\tau) = \sum_{n=-\infty}^{+\infty} \hat{\rho}_x(\tau;n) e^{j2\pi(n/T)t},$$

where $\hat{\rho}_x(\tau;n)$ are the Fourier coefficients calculated as

$$\hat{\rho}_x(\tau;n) \triangleq \frac{1}{T}\int_{-T/2}^{T/2} \rho_x(t,\tau)\, e^{-j2\pi(n/T)t} dt. \tag{2.12}$$

The obtained Fourier coefficients $\hat{\rho}_x(\tau;n)$ are referred to as cyclic autocorrelation function and the frequencies $\alpha_n = n/T$, $n \in \mathbb{Z}$ as cyclic frequencies. A more general case is the one where the autocorrelation function $\rho_x(t,\tau)$ is *almost periodic* in t for all τ. In such a case the possible values of α_n are incommensurate, and can take any integer multiple of all periods of interest. Therefore, for $\alpha \in \mathbb{R}$, the general cyclic autocorrelation function (2.12) becomes (Antoni, 2007a; Gardner et al., 2006):

$$\hat{\rho}_x(\tau;\alpha) \triangleq \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} \rho_x(t,\tau)\, e^{-j2\pi\alpha t} dt. \tag{2.13}$$

The spectral correlation density (SCD) between two spectral components located at $f - \bar{\beta}\alpha$ and $f + \beta\alpha$, where $\beta + \bar{\beta} = 1$, can be calculated using the spectral representation of a random process (2.9) as:

$$\begin{aligned} SC_x(f;\alpha) &= E\left\{dZ(f+\bar{\beta}\alpha)dZ^*(f-\beta\alpha)\right\} \\ &= E\left\{\int_{\mathbb{R}} x(t_1)e^{-j2\pi f t_1}e^{-j2\pi\bar{\beta}\alpha t_1}\, dt_1 \left[\int_{\mathbb{R}} x(t_2)e^{-j2\pi f t_2}e^{j2\pi\beta\alpha t_2}\, dt_2\right]^*\right\} \\ &= E\left\{\iint_{\mathbb{R}} x(t_1)x^*(t_2)e^{-j2\pi f(t_1-t_2)}e^{-j2\pi\alpha(\bar{\beta}t_1+\beta t_2)}\, dt_1 dt_2\right\}. \end{aligned} \tag{2.14}$$

By introducing the following substitutions:

$$\tau = t_1 - t_2$$
$$t = \bar{\beta} t_1 + \beta t_2$$

with the corresponding Jacobian determinant:

$$J = \begin{vmatrix} \frac{\partial \tau}{\partial t_1} & \frac{\partial \tau}{\partial t_2} \\ \frac{\partial t}{\partial t_1} & \frac{\partial t}{\partial t_2} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ \bar{\beta} & \beta \end{vmatrix} = 1,$$

the relation (2.14) transforms into

$$\begin{aligned} SC_x^{\alpha}(f) &= E\left\{ \iint\limits_{\mathbb{R}} x(t_1)x^*(t_2)e^{-j2\pi f(t_1-t_2)}e^{-j2\pi\alpha(\bar{\beta}t_1+\beta t_2)}\, dt_1 dt_2 \right\} \\ &= E\left\{ \iint\limits_{\mathbb{R}} x(t+\beta\tau)x^*(t-\bar{\beta}\tau)e^{-j2\pi f\tau}e^{-j2\pi\alpha t}\, dt d\tau \right\} \\ &= \iint\limits_{\mathbb{R}} E\left\{x(t+\beta\tau)x^*(t-\bar{\beta}\tau)\right\} e^{-j2\pi f\tau}e^{-j2\pi\alpha t}\, dt d\tau \\ &= \iint\limits_{\mathbb{R}} \rho_x(t,\tau)e^{-j2\pi f\tau}e^{-j2\pi\alpha t}\, dt d\tau \\ &= \iint\limits_{\mathbb{R}} \sum_{\alpha_i\in A} \hat{\rho}_x(\tau;\alpha_i)e^{j2\pi\alpha_i t}e^{-j2\pi f\tau}e^{-j2\pi\alpha t}\, dt d\tau \\ &= \sum_{\alpha_i\in A} \int\limits_{\mathbb{R}} \hat{\rho}_x(\tau;\alpha_i)e^{-j2\pi f\tau}\, d\tau \int\limits_{\mathbb{R}} e^{j2\pi\alpha_i t}e^{-j2\pi\alpha t}\, dt \\ &= \sum_{\alpha_i\in A} \mathcal{S}_x(f;\alpha_i)\delta(\alpha-\alpha_i). \end{aligned} \tag{2.15}$$

The last step of (2.15) used the relation

$$\mathcal{S}_x(f;\alpha) = \int_{\mathbb{R}} \hat{\rho}_x(\tau;\alpha)e^{-j2\pi f\tau}\, d\tau. \tag{2.16}$$

The relation (2.16) is referred to as cyclic spectrum at cyclic frequency α. It shows that cyclostationary processes exhibit correlation between spectral components at f and $f+\alpha$. Such processes clearly violate the condition for stationarity defined by (2.9), where for stationary process it is required that its corresponding complex process should have stationary increments. This condition in essence states that all spectral components should be uncorrelated.

At this point it should be noted that in the derivation (2.15) the autocorrelation $\rho_x(t,\tau) = E\left\{x(t+\beta\tau)x^*(t-\bar{\beta}\tau)\right\}$ is usually calculated with values $\beta = 1$ and $\bar{\beta} = 0$, in

accordance with the initial notation specified by (2.3). However there is a wide use of the symmetric form i.e. $\beta = \bar{\beta} = 1/2$. For such a case the autocorrelation becomes

$$\rho_x(t,\tau) = E\{x(t+\tau/2)x^*(t-\tau/2)\}. \tag{2.17}$$

The standard autocorrelation is invariant to time translation as well as the cyclic autocorrelation for given $\alpha = 0$. For the autocorrelation function defined with (2.17), the cyclic autocorrelation (2.13) becomes

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} E\{x(t-\tau/2)x(t+\tau/2)^*\} e^{-j2\pi\alpha t} dt = \hat{\rho}_x(\tau;\alpha) e^{-j\pi\alpha\tau}.$$

2.2.3 Spectral coherence

The linear dependence of two random wide-sense stationary processes $u(t)$ and $v(t)$ is determined by the coherence function. The coherence function is defined as the normalized cross-power spectral density (Penny, 2000)

$$\mathcal{C}_{uv}(f) \triangleq \frac{\Phi_{uv}(f)}{\sqrt{\Phi_{uu}(f)\Phi_{vv}(f)}}, \tag{2.18}$$

where $\Phi_{uu}(f)$ and $\Phi_{vv}(f)$ are the power spectral density of the signals $u(t)$ and $v(t)$ respectively, and $\Phi_{uv}(f)$ is their cross-spectral density defined as

$$\Phi_{uv}(f) = \mathcal{F}\left\{\rho_{uv}(t,\tau)\right\},$$

where $\rho_{uv}(t,\tau) = E\{u(t+\tau)v^*(t)\}$ is the cross-correlation between $u(t)$ and $v(t)$. Furthermore, as the coherence function is actually correlation coefficient, the following relation is valid

$$0 \leq |\mathcal{C}_{uv}(f)|^2 \leq 1$$

Consequently, $|\mathcal{C}_{uv}(f)|^2 = 1$ for some value f if and only if the frequency counterparts of $u(t)$ and $v(t)$ are linearly related for that particular frequency.

As cyclostationary processes exhibit frequency correlation (2.15), suitable forms for $u(t)$ and $v(t)$ can be defined as frequency shifted versions of $x(t)$

$$\begin{aligned} u(t) &= x(t)e^{-j\pi\alpha t} \\ v(t) &= x(t)e^{+j\pi\alpha t}. \end{aligned} \tag{2.19}$$

By applying the substitutions (2.19) for $u(t)$ and $v(t)$, the coherence function (2.18) will reveal whether a linear relation exists between the frequency components $\hat{x}(f+\alpha/2)$ and

$\hat{x}(f-\alpha/2)$ of the observed signal $x(t)$. As a result, the coherence function (2.18) can now be reformulated as

$$\begin{aligned} \mathcal{C}_x(f;\alpha) &= \frac{\mathcal{S}_x(f;\alpha)}{\sqrt{\Phi_{xx}(f+\alpha/2)\Phi_{xx}(f-\alpha/2)}} \\ &= \frac{\mathcal{S}_x(f;\alpha)}{\sqrt{\mathcal{S}_x(f+\alpha/2;\alpha=0)\mathcal{S}_x(f-\alpha/2;\alpha=0)}}, \end{aligned} \tag{2.20}$$

where $\Phi_{xx}(f)$ is the power spectral density of the cyclostationary signal $x(t)$, which can be also obtained from the spectral correlation density $\mathcal{S}_x(f;\alpha)$ defined by (2.16), for $\alpha = 0$. Thus, the relation (2.20) defines the spectral coherence function (SCOH).

2.3 Non-stationary random processes

Stationary random processes are intuitively appealing owing to the ergodicity property. This property guarantees that all statistical characteristics defined on ensemble averages can be determined just from a single realisation of the process by performing time averages. In reality, most of the observed experiments violate the ergodicity property. The vast group of random processes that have time varying statistical properties are called non-stationary.

Central to the non-stationary processes is that non-stationarity is non-property. It is defined solely on the bases that a non-stationary process breaches the necessary conditions for stationarity, thus it belongs to the vast super-set of non-stationary processes (see Figure 2.1).

Unlike the abundance of processing techniques applicable to stationary processes, the set of applicable tools for non-stationary ones is rather modest. As a first step in creating a way to analyse such processes, at least to some extent, is to define a suitable description. For stationary cases such a form is represented by Eq. (2.9), i.e. the spectral representation of stationary process. Priestley (1981) extended this form by allowing the amplitudes of particular spectral components in (2.9) to vary slowly in time, hence defining a subset of non-stationary processes called evolutionary processes. Evolutionary process can be expressed in the generalized form (Spanos and Failla, 2004):

$$x_e(t) = \int_{-\infty}^{\infty} A(\omega,t)e^{j\omega t}dZ(\omega), \tag{2.21}$$

where $A(\omega,t)$ is a slow varying time and frequency dependent modulating function such that its Fourier transform has an absolute maximum at the origin and $Z(\omega)$ is complex

random process with identical properties as in (2.9), i.e.

$$\begin{aligned} &E[dZ(\omega)] = 0 \\ &E[dZ(\omega)dZ^*(\omega')] = 0,\ \omega \neq \omega' \\ &E[|dZ(\omega)|^2] = dH(\omega),\ \forall\omega, \end{aligned}$$

where $dH(\omega)$ is the integrated power spectrum of a stationary process $x(t)$ that has spectral representation (2.9). In cases when $A(\omega, t)$ varies slowly with time it can be treated as amplitude modulation function. Another way of interpreting (2.21) is via frequency domain statement that non-stationary process $x_e(t)$ can be produced by passing a stationary process $x(t)$ through the time-variant and deterministic filter $A(\omega, t)$ (Michaelov et al., 1999). For the special case $A(\omega, t) \equiv 1$ the relation (2.21) defines a stationary process.

One can define the mean square of $x_e(t)$, which after some elementary derivation reads:

$$E[x_e^2(t)] = \int_{-\infty}^{\infty} |A(\omega, t)|^2 \, d\omega,$$

and $|A(\omega, t)|^2$ is referred to as evolutionary spectral density (Priestley, 1981). It defines the energy distribution of $x_e(t)$ in the vicinity of the time instant t and can be interpreted equivalently to the power spectral density for stationary processes.

The evolutionary processes, as a sub-class of non-stationary processes, are important in the area of fault detection of rotational machines. This sub-class contains: modulated stationary processes, filtered modulated white noise, and modulated filtered white noise. Specific details about the statistical properties of these special cases can be defined by using (2.21).

2.3.1 Uniformly modulated process

The simplest form of nons-tationary process, represented with (2.21), is obtained by modulating a stationary process $x(t)$ with stochastic envelope $C(t)$

$$x_e(t) = C(t)x(t), \tag{2.22}$$

where $x(t)$ is stationary process with spectral representation (2.9) and $C(t)$ is any non-negative real valued function for which the Fourier transform exists. Hence (2.22) becomes

$$x_e(t) = \int_{-\infty}^{\infty} C(t)e^{j\omega t} \, dZ(\omega).$$

For such a process the integrated spectrum (2.8) becomes

$$\begin{aligned} dH_e(\omega) &= E\left\{|C(t)dZ(\omega)|^2\right\} \\ &= |C(t)|^2 E\left\{|dZ(\omega)|^2\right\} \\ &= |C(t)|^2 dH(\omega) \end{aligned} \tag{2.23}$$

The relation (2.23) shows that all spectral components of the process (2.22) evolve in time exactly the same way because $A(\omega, t)$ depends only on t. Due to "uniform" frequency influence, processes with such a characteristic are called uniformly modulated processes.

2.3.2 Linear filtering of stochastic processes

Generally, linear filtering of a stationary process $x(t)$ relies on the convolution integral

$$y(t) = \int_{-\infty}^{\infty} g(u)x(t-u)\,du,$$

where $g(t)$ is impulse response of the filter. The associated transfer function is defined as:

$$\Gamma(\omega) = \int_{-\infty}^{\infty} g(u)e^{-ju\omega}\,du.$$

For the stationary case, the spectral density of $y(t)$ is completely defined by the spectral density of $x(t)$ and $\Gamma(\omega)$. The value at a particular frequency ω_k is not influenced by any other frequency from $x(t)$ apart from $dH_x(\omega_k)$.

For evolutionary process $x_e(t)$ that satisfies (2.21) the filtering process results in

$$y(t) = \int_{-\infty}^{\infty} \Gamma(\omega)A(\omega, t)e^{j\omega t}dZ(\omega).$$

Generally, the filter's impulse response $g(t)$ decays towards zero as $t \to \infty$. Contingent $A(\omega, t)$ varies sufficiently slowly, one can consider that $A(\omega, t)$ is almost constant for the duration of $g(t)$. Consequently, for the integrated power spectrum (2.8) $dH_y(\omega)$ of $y(t)$ becomes related to the integrated spectrum $dH_x(\omega)$ of the stationary process $x(t)$ as

$$dH_y(\omega) \approx |\Gamma(\omega)|^2 dH_X(\omega).$$

3 Point Processes

Point processes represent a segment of the theory of random processes that are most commonly used for characterising random collections of point occurrences (Cox and Isham, 1980). In the simplest form, these points usually represent the time moments of their occurrences. This class of point processes is also known as temporal point processes. Some examples that belong to this class are: emissions from radioactive sources, electrical surges in neural fibres, customer queues etc.

Each random occurrence of a point (event) can be characterised with additional information called mark. The underlying point processes are called marked temporal point processes. The values of such marks can be also considered as random variables, and are usually treated as statistically independent of the underlying point process. These marks normally carry vital physical information about the observed process. One obvious example is the time of occurrence and the magnitude of an earthquake.

This chapter reviews basic tools for analysis of marked temporal point processes. Focus will be on the definition of conditional intensity function and probability distribution of observing N events within a specific time interval. Special attention will be given to the definition and statistical characteristics of a point process in which the time occurrences are governed by Inverse Gaussian distribution.

3.1 Basics of point processes

The theory of point processes evolved from research conducted in four different areas (Daley and Vere-Jones, 2003): (i) renewal theory and lifetime expectancy, (ii) counting problems, (iii) particle physics problems and (iv) communication problems. Characterisation of point processes in each of these areas has been done in different ways.

The main research focus in the area of renewal and lifetime expectancy was the estimation of the probability that a failure will not occur between two consecutive service intervals. This process is characterised by the so-called survival function

$$s(t) = \Pr\{\text{lifetime} > t\}. \tag{3.1}$$

and hazard function defined as

$$q(t)\,dt = \Pr\{\text{lifetime terminates between } t \text{ and } t+dt | \text{it does not terminate before } t\} \tag{3.2}$$

The most common way to define a point process is through the specification of the statistical characteristics of the number of observed events. This approach is best described using the random measure framework (Daley and Vere-Jones, 2003; Schoenberg, 2001).

Let $N(A)$ be the number of events of the point process for a set $A \subset \mathbb{R}^+$. The probability distribution $\Pr\{N(A) = k\}$, for $k \in \mathbb{N}$ completely describes the underlying point process. Thus $N(A)$ is increasing and right-continuous, accepting only non-negative integer values. In cases when the subset A represents a time interval $(0, t]$ the notation can be changed to $N(t)$. Furthermore, the measure $N(A)$ should be finite for bounded set A, thus limiting the possibility of "too many" points occurring "too close" (Daley and Vere-Jones, 2003).

Besides the statistics of the number of events, a point process can be also defined by the distribution of the time t_k needed for observation of k events i.e.

$$t_k = \inf\{t > 0 : N(t) \geq k\}.$$

Finally, the same goal can be reached by specifying the statistical properties of the sequence of interarrival intervals τ_i between two consecutive events, defined as

$$\tau_i = t_i - t_{i-1}.$$

It can be noticed that there are several equivalent ways to define the statistical properties of a point process. Throughout the chapter the analysis was performed using the distribution of the interevent times and the number of events.

3.1.1 Stationarity of point processes

The concept of stationarity for point processes is the same as for the case of general stochastic process.

Definition 3.1. *For a the time interval $A_k \subset \mathbb{R}$ we define the measure $N(A_k)$ specifying the number of events in A_k. A point process is stationary if for all bounded intervals A_k, $k = 1, 2, \ldots$ the joint distribution of*

$$\{N(A_1 + \tau), \ldots, N(A_k + \tau)\}$$

does not depend on $\tau(-\infty < \tau < \infty)$.

In terms of the interevent times τ_k, the following definition is valid.

Definition 3.2. *A point process is stationary if for every $r = 1, 2, \cdots$ and $i_r \in \mathbb{Z}$ the joint distribution of the interevent times $\{\tau_{i_1+k}, \cdots, \tau_{i_r+k}\}$ is independent of $k \in \mathbb{Z}$.*

3.1.2 Intensity

The intensity λ of a point process describes the "frequency" of occurrence of events around time point t. It is formally defined by the Khinchin's Existence Theorem (Daley and Vere-Jones, 2003):

Theorem 3.1. *For a simple stationary point process, the intensity is*

$$\lambda = \lim_{h \to 0} \frac{\Pr\{N(0,h] > 0\}}{h}.$$

3.2 Types of point processes

Out of various types of point processes two specific ones are interesting in the context of this thesis. The first one is the Poisson process and the second one is a subgroup of point processes called renewal processes.

3.2.1 Poisson process

Poisson process is defined as a simple point process that satisfies the following three rules:

i) for an empty interval $A = \emptyset$, $N(A) = 0$,

ii) the number of events in any disjoint intervals are independent, and

iii) the number of points in any interval $A = [\tau, \tau + t]$ follows the Poisson distribution

$$\Pr\{N(A) = k\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Furthermore, the expected number of events for a time interval with length t is

$$E[N(t)] = \lambda t. \tag{3.3}$$

Relation (3.3) clarifies the meaning of the name *intensity* for the parameter λ. In many cases the intensity λ is time dependent and is written as $\lambda(t)$. Such a Poisson process is called *inhomogeneous Poisson process.* The Poisson process has some significant properties that makes it particularly interesting (Dekking et al., 2005).

Theorem 3.2. *The interevent times $\tau_n = t_n - t_{n-1}$ of Poisson process with $\lambda > 0$ are independent and identically distributed exponential random variables with mean $1/\lambda$.*

Theorem 3.3. *Given that a Poisson process has $N(A)$ events in interval A then the locations of the points are independently distributed with a uniform distribution on the interval A.*

These two theorems show that the occurrence of a single event has no influence on the time of occurrence of the next event. Although in some cases such an approach is suitable, for our purpose we need a point process capable to model dependencies between consecutive events.

3.2.2 Renewal processes, Markov processes and conditional intensity

A renewal process is a point process for which the interevent times are independent and identically distributed (i.i.d) variables with arbitrary non-negative distribution function $F(t)$. It is a generalisation of the Poisson process by allowing a wider family of distributions for the interevent times instead of just exponential one. For such processes the distribution of the τ_{n+1} interevent time interval depends on $\tau_n, \tau_{n-1}, \ldots$ i.e. the history of the point process up to specific time moment t. This dependence is usually marked as $\mathcal{D}_t$. The simplest form are the so-called Wold processes, for which the distribution of τ_{n+1} depends only on the τ_n i.e. the previous one. Hence, for such a process the interevent times are governed by first-order Markov chain.

An approach to modelling of Wold processes is by defining the conditional interevent density $f(t|\mathcal{D}_t)$ for the time moment t in respect to the history of the process $\mathcal{D}_t$ up to the particular moment t (Daley and Vere-Jones, 2003) i.e.

$$f(t|\mathcal{D}_t)\,dt = \Pr\{\text{event occurs between } t \text{ and } t + dt|\mathcal{D}_t\}.$$

The conditional interevent density $f(t|\mathcal{D}_t)$ can be derived through the so-called conditional intensity (hazard) function. The hazard function defined by (3.2), can be rewritten as

$$q(t|\mathcal{D}_t) = \frac{f(t|\mathcal{D}_t)}{S(t|\mathcal{D}_t)} = \frac{f(t|\mathcal{D}_t)}{1 - F(t|\mathcal{D}_t)},$$

where $S(t|\mathcal{D}_t)$ the survival function (3.1) and $F(t|\mathcal{D}_t)$ is the cumulative distribution function associated with $f(t|\mathcal{D}_t)$.

Definition 3.3. *The conditional intensity function for a regular point process defined on* $\mathbb{R}^+ = [0, \infty)$ *is*

$$\lambda(t|\mathcal{D}_t) = \begin{cases} q(t_1), & \text{for } 0 < t \leq t_1 \\ q(t|\mathcal{D}_t), & \text{for } t_{n-1} < t \leq t_n, \quad n \geq 2, \end{cases} \tag{3.4}$$

where t_1 *is the time moment of the first occurrence, and* t_n *is the time of the* n*th occurrence.*

The term conditional intensity comes from the analogy with the relation (3.3) i.e.

$$\lambda(t|\mathcal{D}_t)\,dt \approx E\{N(dt)|\mathcal{D}_t\}.$$

The conditional intensity function completely describes the underlying point process. In general, as shown by (3.4), this function depends on both the current time t as well as the complete point process history $\mathcal{D}_t$ up to the moment t.

By modelling different conditional intensity functions one can specify the most suitable point process for the observed events. In the simplest form, if $\lambda(t|\mathcal{D}_t)$ in (3.4) becomes independent of $\mathcal{D}_t$, then the interevent times will become independent from the past. Hence the resulting process becomes non-stationary (inhomogeneous) Poisson point process. It is called inhomogeneous because the intensity $\lambda(t)$ can vary in time. However, making $\lambda(t|\mathcal{D}_t)$ both independent from the past and constant in time will result in a stationary Poisson point process. In the subsequent sections, several not-trivial, conditional intensity functions will be used to model some randomly re-occurring patterns.

3.3 Modelling of re-occurring patterns

Re-occurring patterns are common phenomenon in many systems. Usually, these patterns are analysed through statistical characteristics of the interevent times. For point processes, the conditional intensity function $\lambda(t|\mathcal{D}_t)$ offers a suitable mechanism of defining the statistical characteristics of such re-occurring patterns. This section presents three different conditional intensity functions that can be used for modelling specific re-occurring patterns.

3.3.1 Self-exciting and self-correcting point processes

The name of this group of point processes originates from the way the interevent time intervals are correlated.

Definition 3.4. *Let $N(A)$ and $N(B)$ be the number of events in two non-overlapping intervals A and B. If the covariance $cov\{N(A), N(B)\} > 0$ then the process is said to be self-exciting, otherwise the process is self-correcting.*

A typical example of a self-exciting process is the so-called Hawkes process, defined by Ogata (1981) with its conditional intensity function

$$\lambda(t) = c\mu(t) + \alpha \sum_{t_i<t} e^{-(t-t_i)},$$

where $\mu(t)$ is the background intensity and t_i is the time of the occurrence of the ith event. Each time a new event occurs, the conditional intensity grows by α, and then decreases exponentially as time passes. Such a conditional intensity function causes the process to exhibit clustered bursts of events. This model is used in modelling earthquake

aftershocks since the events tend to cluster immediately after the first occurrence. A simple realisation is shown in Figure 3.1.

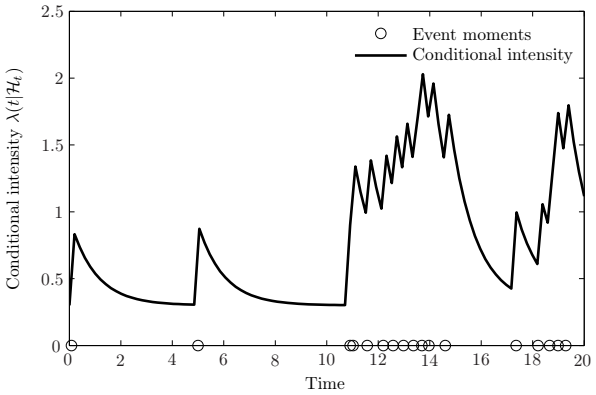


Figure 3.1: Simulation of a self-exciting process

Self correcting process, on the other hand, describes completely opposite behaviour. Its conditional intensity function has the form (Anagnos and Kiremidjian, 1984)

$$\lambda(t|\mathcal{D}_t) = \exp\left(\mu t - \sum_{t_i<t} \alpha\right),$$

where μ and α are positive constants and t_i is the time of the occurrence of the ith event. In such a case the conditional intensity increases as the time from the last event passes. A realisation of this process is shown in Figure 3.2.

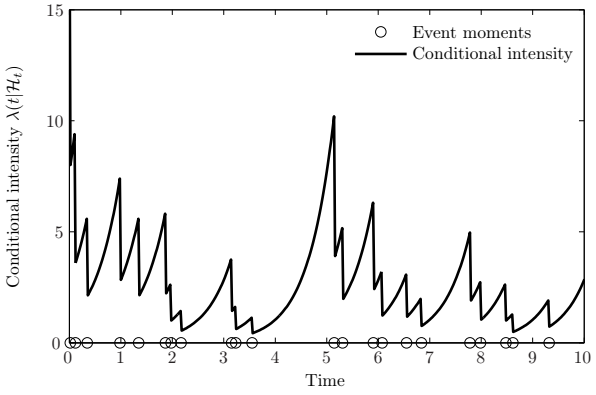


Figure 3.2: Simulation of a self-correcting process

These processes are readily applied for modelling earthquakes and their after shocks. However, the conditional intensity functions lack the capability to impose better structure on the corresponding point process, such as almost periodical re-occurrence. Since the main focus of this dissertation is the analysis of signals generated by rotational machines, the next logical step would be to define more appropriate shape of the conditional intensity

function that will enable modelling of randomly re-occurring events that may exhibit almost periodic structure.

3.3.2 Inverse Gaussian distribution of interevent times

The frequency of the re-occurring events in rotational machines is closely related with rotational speed. So, the proposed model should be able to accommodate for three different modes of operation:

i) constant,

ii) slowly varying and

iii) significantly variable rotational speed.

The simplest case is when events occur once per revolution. In such a case the most intuitive approach is to model the evolution of the angle of rotation, since an event will occur each time the angle of rotation reaches 2π. The occurrence of a new event will represent the start of a new revolution. This assumption is valid, since the events are due to mechanical surface faults and such faults have fixed position on the rotating element. Thus, the only varying parameter influencing the time of their occurrence is rotational speed.

By adopting this approach, the most suitable way of modelling the angle of rotation $\theta(t)$ is through the following relation:

$$\theta(t) = \theta_0(t) + \varepsilon(t), \tag{3.5}$$

where $\theta_0(t)$ is some deterministic process usually directly related to the rotational speed ν, hence satisfying the differential equation $\dot{\theta}_0(t) = \nu$ and $\varepsilon(t)$ is a stochastic process that accommodates all speed fluctuations as well as any additional unmodelled disturbances. Motivated by the central-limit theorem one may model the process (3.5) simply by adding a Gaussian noise scaled with some constant σ to the deterministic part $\theta_0(t)$. Hence, the rotational speed $\dot{\theta}(t)$ becomes

$$\dot{\theta}(t) = \nu + \sigma\dot{W}(t),$$

where $W(t)$ is standard Brownian motion, whose increments are normally distributed with zero mean and some constant variance (Matthews et al., 2002). Consequently, (3.5) becomes

$$\theta(t) = \nu t + \sigma W(t). \tag{3.6}$$

The parameter $\nu > 0$, also known as positive drift, represents the deterministic part of the rotational speed, and σ specifies the speed fluctuations (Whitmore and Seshadri, 1987).

The time occurrence of once-per-revolution events can be defined as:

$$t_k = \min\{t : t \geq 0; \theta(t) \geq 2k\pi\}.$$

A realisation of (3.5) with marked time intervals t_k is shown in Figure 3.3. Schrödinger has shown that the distribution of the time intervals t_k needed for the Wiener process (3.6) to reach a fixed threshold a follows the Inverse Gaussian distribution (Folks and Chhikara, 1978):

$$f(t) = \frac{a}{\sigma\sqrt{2\pi t^3}} \exp\left\{-\frac{(\nu t - a)^2}{2\sigma^2 t}\right\}, \tag{3.7}$$

usually denoted by $t \sim IG(a/\nu, a^2/\sigma^2)$. Tweedie (1957) introduced the term Inverse Gaussian based upon the relationship between the cumulant generating functions of the Gaussian and Inverse Gaussian distributions. Since this distribution is central in our analysis we will present a short overview of the most important statistical properties of a point process with IG distribution of interevent times.

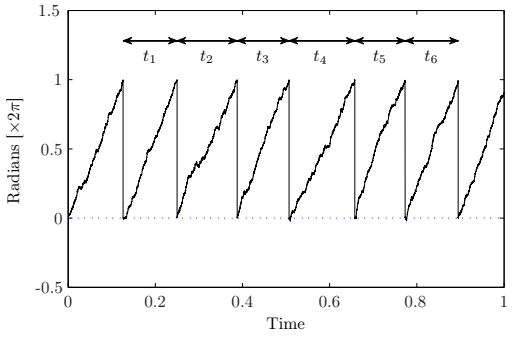


Figure 3.3: Realization of a renewal process with Inverse Gaussian interevent time distribution

3.4 Statistical characteristics of renewal processes with Inverse Gaussian distribution of interevent times

Although the distribution of the interevent time completely defines a point process, this section presents an additional set of statistical properties of point processes governed by IG distribution.

Probability of observing k consecutive events The probability of observing k consecutive events within the time interval $[t_0, t)$ is usually the most common way of describing a point process. For simplicity, the time moment t_0 is set to $t_0 = 0$.

Let the probability density of a single event occurring up to time t_1 be $f_1(t_1)$. The goal is to calculate the probability density $f_k(t_k)$ for k consecutive events occurring up to the time moment t_k. This probability density reads (van Vreeswijk, 2010):

$$f_k(t_k) = \int_0^{t_k} f_{k-1}(t_{k-1}) f_1(t_k - t_{k-1}) dt_{k-1}. \tag{3.8}$$

The Eq. (3.8) is a convolution of two probability densities defined on the non-negative real line, since both $t_k > 0$ and $t_k > t_{k-1}$. This convolution can be elegantly calculated using the Laplace transform of (3.8) as:

$$f_{L,k}(s) = f_{L,k-1}(s) f_{L,1}(s) = f_{L,1}^k(s), \tag{3.9}$$

where $f_{L,k-1}(s) = \mathcal{L}\{f_{k-1}(t)\}$, $f_{L,1}(s) = \mathcal{L}\{f_1(t)\}$ and $\mathcal{L}\{\cdot\}$ stands for the Laplace transform.

In case of Inverse Gaussian interevent times, $f_{L,1}(t)$ is the Laplace transform of (3.7) i.e.:

$$f_{L,1}(s) = \exp\left\{\frac{\nu a}{\sigma^2}\left[1 - \sqrt{1 + 2\frac{\sigma^2}{\nu^2}s}\right]\right\} \tag{3.10}$$

Calculating then the inverse $\mathcal{L}^{-1}\{f_{L,1}^k(s)\}$ of (3.10) results into (Tweedie, 1957):

$$f_k(t) = \frac{ka}{\sigma\sqrt{2\pi t^3}} \exp\left\{-\frac{(\nu t - ka)^2}{2\sigma^2 t}\right\}. \tag{3.11}$$

The obtained result (3.11) has quite intuitive explanation. Namely, in (3.7) the threshold for the Wiener process was set at a. Therefore the time t needed to observe k consecutive crossings has the same distribution as if one elevated the threshold up to ka.

Variations in the drift and variance

The model (3.5) for determining the interevent times is suitable for the cases when re-occurring patterns are truly periodic as well as when the periodicity assumption is breached. For periodic cases, the value of σ in (3.6) will become zero, hence the Inverse Gaussian distribution (3.7) becomes Dirac impulse i.e.

$$\lim_{\sigma \to 0} f(t|\nu, \sigma) = \delta(\nu t - a).$$

Consequently, the corresponding point process will represent truly periodic events.

Small variations in the re-occurrence period can be accommodated by allowing small values for σ in (3.7). The analysis of the autocorrelation of a point process governed by IG interevent distribution reveals the most important statistical characteristics of this case.

The autocorrelation $\rho_x(\tau)$ of a renewal process has the following form (van Vreeswijk, 2010):

$$\rho_x(\tau) = \nu\delta(\tau) + \tilde{P}_x(\tau),$$

where $\nu\delta(\tau)$ represents the probability of an event occurring at time t and *the same* event occurring in time $t+\tau$, whereas $\tilde{P}_x(\tau)$ represents joint probability density of an event occurring at time t and *another* one occurring at time $t+\tau$. Furthermore, the joint probability of the kth event occurring at time t while the $k+n$th event occurring at time $t+\tau$ is $f_n(\tau)$. Having this in hand the probability $\tilde{P}_x(\tau)$ becomes a sum of probabilities that an event occurs at time t whereas the nth consecutive event occurs at time $t+\tau$. Hence $\tilde{P}_x(\tau)$ becomes:

$$\tilde{P}_x(\tau) = \nu \sum_{n=1}^{\infty} f_n(\tau). \tag{3.12}$$

Using (3.8) for $f_n(\tau)$ (3.12) becomes

$$\tilde{P}_x(\tau) = \nu f(\tau) + \int_0^{\tau} f(\tau')\tilde{P}_x(\tau - \tau')\, d\tau'. \tag{3.13}$$

Calculating the Laplace transform of (3.13) transforms it into

$$\tilde{P}_{Lx}(s) = \frac{\nu f_L(s)}{1 - f_L(s)}.$$

Using the Final value theorem of the Laplace transform it can be readily shown that

$$\lim_{\tau\to\infty} \rho_x(\tau) = \frac{2\sigma^2}{a\nu} < \infty.$$

As already analysed by Antoni and Randall (2002), such a process can be treated as pseudo cyclostationary in cases when σ is sufficiently small.

The modelling of completely arbitrary speed variations can be done by allowing both $\nu_{shaft} = \nu(t)$ and $\sigma_{shaft} = \sigma(t)$ in (3.7) to become time dependent. The resulting process is called doubly stochastic process which in essence is non-stationary process.

Despite the non-stationary characteristics, for cases where $\nu(t)$ varies slowly, one can employ the so–called modified variability measure C_{V2}. This measure is fairly insensitive to variations in the firing rate of the point process and is defined as (Ponce-Alvarez et al., 2010):

$$C_{V2} = \frac{2|\tau_{i+i} - \tau_i|}{\tau_{i+1} - \tau_i},$$

where τ_i represents the interevent time between the events $i-1$ and i.

4 Wavelet transform of non-stationary stochastic processes

One of the most exploited techniques for signal processing is Fourier Transform. It decomposes the original signal into a sum of sinusoidal waves. This transform is suitable for regular signals (roughly stated smooth periodic signals). However for signals describing transient phenomena, located in a narrow time interval, the application of Fourier Transform turns to be inadequate.

The search for a signal processing technique, capable of analysing signals that may contain both smooth and periodic components as well as short term transients, resulted into techniques such as: STFT, Wigner-Ville distribution, wavelet transform etc. Here we focus on the properties of the time-frequency signal processing method called wavelet transform.

4.1 Brief history of wavelet transform

Haar (1910) initially introduced the wavelet transform by defining the simplest so-called Haar wavelet. Grossmann and Morlet (1984) made an additional improvement by circumventing the limitation of the short-time Fourier transform for fixed length windows. This achievement enabled significant increase of time-frequency resolution compared with that of STFT. Furthermore, they showed that the original signal can be reconstructed from the calculated wavelet coefficients. This discovery paved the way for new techniques of signal processing that operate solely on the wavelet coefficients.

4.2 Introduction to wavelet transform

Wavelet transform is based on a set of specifically designed functions called wavelets. The continuous wavelet transform (CWT) of a square integrable function $f(t) \in \mathbf{L}^2(\mathbb{R})$ is defined as (Mallat, 2008):

$$Wf(s,u) = \langle f(t), \psi_{u,s}(t) \rangle = \int_{-\infty}^{\infty} f(t)\psi_{u,s}(t)\, dt,$$

where $\psi_{u,s}(t)$ is a dilated and translated version of the mother wavelet $\psi(t)$:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \tag{4.1}$$

If we limit the values for the parameters s and u to a set of discrete values $s_j = 2^{-j}$ and $u_{j,k} = 2^{-j}k$, where $j, k \in \mathbb{Z}$, the mother wavelet (4.1) becomes:

$$\psi_{j,k}(t) = 2^{j/2}\psi\left(2^j t - k\right), \ j, k \in \mathbb{Z}. \tag{4.2}$$

The wavelet transform restricted with (4.2) is referred to as discrete wavelet transform (DWT) (Figliola and Serrano, 1997). The DWT $c_{j,k}$ for the same finite energy function $f(t)$ is defined as

$$c_{j,k} = Wf(s_j, u_{j,k}) = \langle f(t), \psi_{j,k}(t)\rangle. \tag{4.3}$$

Such a definition of DWT is a first step towards the multi-resolution analysis of signals.

4.3 Multi-resolution analysis

Definition 4.1 (Burrus et al. (1994))**.** *If a function $f(t) \in \mathbf{L}^2(\mathbb{R})$ can be expressed as $f(t) = \sum_k a_k\phi_k(t)$, then the components $\phi_k(t)$ are called the expansion set. For cases when this representation is unique the expansion is called the basis.*

The multi-resolution analysis is based on the definition of a scaling function $\phi(t) \in \mathbf{L}^2(\mathbb{R})$ that defines a subset $\mathcal{V}_0 \subseteq \mathbf{L}^2(\mathbb{R})$ so that

$$\mathcal{V}_0 = \text{Span}\{\phi_k(t)\}, \quad k \in \mathbb{Z},$$

where $\phi_k(t) = \phi(t-k)$. By introducing the scaling parameter j the initial space $\mathcal{V}_0$ can be expanded as

$$\phi_{j,k}(t) = 2^{j/2}\phi\left(2^j t - k\right),$$

and

$$\mathcal{V}_j = \text{Span}\{\phi_{j,k}(t)\}.$$

Given the resolution level 2^{-j}, it can be considered that the time axis is divided into intervals, each 2^j wide. The value at each interval represents the average of $f(t)$ in that neighbourhood. Increasing the resolution will decrease the width of these intervals and the resulting transformation will better resemble the original signal $f(t)$. Formally speaking, the approximation of $f(t)$ at resolution 2^{-j} is the orthogonal projection of $f(t)$ on the space $\mathcal{V}_j$ (Mallat, 2008).

The spaces $\mathcal{V}_j$ at different resolution levels have the following properties (Mallat, 2008):

P1) At higher resolutions the approximation contain all the information from the lower resolution approximations plus a new previously unseen details i.e. $\mathcal{V}_{j+1} \subset \mathcal{V}_j$.

P2) perfect reconstruction is obtained at infinite resolution ($j = -\infty$)

P3) Any time shifts in the original signal have no effect on the approximation at the particular resolution i.e.

$$f(t) \in \mathcal{V}_j \Leftrightarrow f(t - k2^j) \in \mathcal{V}_j, \ \forall (j,k) \in \mathbb{Z}$$

P4) Scaling of the original function alters the resolution level

$$f(t) \in \mathcal{V}_j \Leftrightarrow f(2^{-1}t) \in \mathcal{V}_{j+1}$$

P5) There always exists a function $b(t)$ such that $b(t-k)$, $k \in \mathbb{Z}$ is an orthonormal basis for $\mathcal{V}_0$.

The properties P4 and P5 state that one can define the scaling function for different resolution levels starting from the initial scaling function using the relation

$$\varphi(t) = \sqrt{2} \sum_n h(n) \varphi(2t - n), \quad n \in \mathbb{Z}, \tag{4.4}$$

where the coefficients $h(n)$ are sequence of real or complex numbers that define the so-called scaling function coefficients.

Detail subspaces

The transfer from a finer resolution level towards a coarser one results into loss of information i.e. a part that is visible in the higher level is not visible in the lower resolution level. This additional information can be considered to be included into another subspace such as (Gargour et al., 2009):

$$\mathcal{V}_{j-1} = \mathcal{V}_j \oplus \mathcal{W}_j,$$

where the subspace $\mathcal{W}_j$ is orthogonal to the subspace $\mathcal{V}_j$. The subspace $\mathcal{W}_j$ is called detail subspace. Furthermore, it can be shown that the detail space $\mathcal{W}_j$ is orthogonal to any other detail space i.e.

$$\mathcal{W}_j \perp \mathcal{W}_k \quad \forall j \neq k.$$

Since the subspace $\mathcal{W}_j \subset \mathcal{V}_j$ the basis spanning these detail subspaces can be calculated using the scaling function $\varphi(t)$ (Mallat, 2008)

$$\psi(t) = \sqrt{2} \sum_n g(n) \varphi(2t - n), \quad n \in \mathbb{Z}, \tag{4.5}$$

where the coefficients $g(n)$ are sequence of real or complex numbers similarly like $h(n)$ in (4.4).

The discrete wavelet analysis (4.3) can be implemented by defining the coefficients in (4.4) and (4.5) as a set of discrete conjugate mirror filters: $h(n)$ which is a low-pass filter and $g(n)$ which is a high-pass filter. Both filters are related to scaling function $\phi(t)$ and to wavelet function $\psi(t)$ correspondingly through the following relations

$$h(n) = \frac{1}{\sqrt{2}}\langle\phi(t), \phi(2t-n)\rangle$$
$$g(n) = \frac{1}{\sqrt{2}}\langle\psi(t), \phi(2t-n)\rangle.$$

In order to preserve orthogonality of the transformation, both filters are related by $g(n) = (-1)^n h(1-n)$.

4.4 Wavelet packet transform

In the case of DWT the resolution of the approximation space $\mathcal{V}_j$ is gradually decreased and the signal is analysed through it projections on the detailed spaces $\mathcal{W}_j$. An upgrade to this approach would be to recursively divide both detailed $\mathcal{W}_j$ and approximation $\mathcal{V}_j$ spaces which will result into the so-called binary wavelet packet transform (WPT), usually depicted with a wavelet packet tree, as shown in Figure 4.1. Nodes of the tree are marked as (d, n), where $d \in \{1, \cdots, D_0\}$ represents the depth of the tree, and $n = 0, \cdots, 2^d - 1$ represents the number of the node at depth d. Each of the node defines a space with orthonormal basis $\psi_d^n(t - 2^d k)$, $k \in \mathbb{Z}$, and contains N_d wavelet packet coefficients $P_j^n(t)$, $t = 0, \cdots, N_d - 1$ (Percival and Walden, 2000). These basis can be calculated using the splitting relations Mallat (2008)

$$\psi_{j+1}^{2p}(t) = \sum_{n=-\infty}^{+\infty} h(n)\psi_j^p(t - 2^j n)$$
$$\psi_{j+1}^{2p+1}(t) = \sum_{n=-\infty}^{+\infty} g(n)\psi_j^p(t - 2^j n).$$

For the first two steps, i.e. for $j = 0$ and $p = 0, 1$ the orthonormal basis correspond to the scaling function $\phi(t)$ and wavelet function $\psi(t)$ respectively. Projecting the square integrable signal $f(t)$ onto a space spanned by a single wavelet packet node will result into the following wavelet coefficients

$$P_j^n(k) = \langle f(t), \psi_d^n(2^j t - k)^*\rangle = \frac{1}{\sqrt{2^j}}\int_{-\infty}^{\infty} f(t)\psi_d^n(2^j t - k)^*\, dt, \tag{4.6}$$

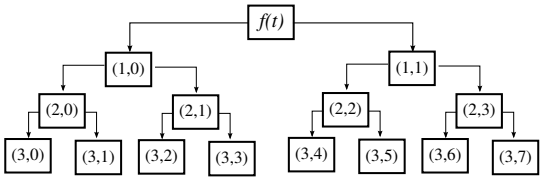


Figure 4.1: Example of full WPT tree with depth $D_0 = 3$

where $P_j^n(k)$ is the vector of wavelet coefficients at node n of level j for translation time moments k.

Wavelet packet coefficients contain information from the specific frequency segment of the analysed signal. The frequency segment covered by each wavelet packet function is determined by the scaling parameter j and the oscillation parameter n.

4.5 Wavelet packet analysis of non-stationary random signals

Of special interest is the wavelet analysis of the non-stationary signal $x_e(t)$ (2.21):

$$x_e(t) = \int_{-\infty}^{\infty} A(\omega, t) e^{j\omega t} dZ(\omega).$$

Similarly like for the case of Fourier transform, $x_e(t)$ violates the conditions (2.4) i.e. $x_e(t) \notin \mathbf{L}^2(\mathbb{R})$. Therefore, in the same manner as in the case of Fourier transform, the wavelet transform is applied only to limited version of $x_e(t)$ within the observed time interval as in (2.6)

$$x_{eT}(t) = \begin{cases} x_e(t), & -T \leq t \leq T \\ 0, & \text{otherwise.} \end{cases}$$

For simplicity of notation, however, in the subsequent analysis the index T is ommited.

Projecting the signal $x_e(t)$ onto the space spanned by a single wavelet packet node n of level j will result into the wavelet coefficients $P_j^n(k)$ (4.6) (Spanos and Failla, 2004):

$$\begin{aligned} P_j^n(k) &= 2^{-j/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\omega, t) e^{j\omega t} \psi_d^n (t - 2^d k)^* \, dt \, dZ(\omega) \\ &\approx \frac{1}{2^{j/2}} \int_{-\infty}^{\infty} A(\omega, 2^{-j}k) e^{j\omega k} \Psi^*(2^{-j}\omega) \, dZ(\omega), \end{aligned} \tag{4.7}$$

where $\Psi^*(\omega)$ is the Fourier transform of mother wavelet $\psi(t)$ for the corresponding node. The second step in (4.7) is a consequence of the time localization of mother wavelet in the vicinity of $2^{-j}k$. The wavelet coefficients (4.7) have the similar form as the non-stationary

oscillatory process (2.21), with time-frequency envelope $2^{-j/2}A(\omega, 2^{-j}k)\Psi^*(2^{-j}\omega)$. Hence, the variance $Var\{P_j^n(k)\}$ can be interpreted as a measure of instantaneous power at the time moment k (Priestley, 1981):

$$Var\{P_j^n(k)\} = 2^{-j} \int_{-\infty}^{\infty} |A(\omega, 2^{-j}k)|^2 |\Psi^*(2^{-j}\omega)|^2 \Phi_x(\omega)\, d\omega, \tag{4.8}$$

where $\Phi_x(\omega)$ is the power spectral density of a zero-mean stationary process $x(t)$ defined by (2.9), that has spectral representation $dZ(\omega)$. Since the wavelet coefficients for that particular node span a specific frequency band, the decomposition will give the power fluctuations contained within a specific set of frequencies.

Accuracy of the wavelet packet representation depends on the selection of mother wavelet. Spanos and Failla (2004) showed that good accuracy can be achieved when the time-frequency variations of the envelope $A(\omega, t)$ are slower than the duration of the wavelet for the particular node. Additionally, Unser and Tafti (2010) concluded that the crucial parameter, for sparse wavelet representation of signals containing repetitive impulse responses, is the number of vanishing moments of the mother wavelet rather then the selection of the "optimal" mother wavelet that will closely match the underlying signal.

4.6 Concluding remarks

The complete theory behind wavelet transform and their various applicability is enormous. This brief overview contained only the minimum required for the subsequent application of wavelet analysis for fault detection in mechanical drives. For a comprehensive background on the wavelets and wavelet transform the reader is referred to Burrus et al. (1994); Daubechies (1992); Mallat (2008).

The orthogonality of the wavelet transform ensures that the wavelet coefficients completely encompass the statistical characteristics of the analysed signal. Based on this property and the wavelet packet analysis of non-stationary random signals, Chapter 7 will present the so-called information cost functions that will serve as basic building blocks for the feature set robust to the variations in the operating conditions.

5 Signal models for mechanical faults

The effectiveness of the fault diagnosis process relies on the capability of extracting information about the state and behaviour of the monitored system from observed signals. The first step in this process is to determine the relationship between the patterns occurring in the signals and actual condition of the system. For that purpose this chapter will present models of the most frequent mechanical faults that occurr in rotational machinery and mechanical drives. The selection was based on a report of the reliability working group MRWG (1985a,b,c), showing that bearing and gear faults made more than 80% of the mechanical failures in rotational machines. Therefore the forthcoming sections address models of signals under various gear and bearing faults.

5.1 Gear vibration model

Gears represent the most intensive source of vibrations in a mechanical drive, hence their condition can significantly influence the behaviour of other components. The main source of vibrations in healthy gears are variations in mesh stiffness. The way these variations occur depends on the gear's geometrical characteristics. In this research the focus is on the simplest form of gears, the so-called spur gears. As soon as two teeth enter into the meshing contact the applied torque on the driving tooth causes both driven and driving teeth to bend. The amount of deflection is defined by the stiffness of the teeth, the elastohydrodynamic characteristics of the lubrication as well as the applied torque (Howard et al., 2001). As gears rotate, the number of teeth in contact varies. In cases when more than two teeth are in the meshing contact, the overall contact stiffness increases because the same torque is distributed over several (most commonly two) teeth. Consequently, the deflection is smaller. Conversely, when only two teeth are in meshing contact, the overall meshing stiffness decreases. Thus, for the cases of constant speed and load, variations in meshing stiffness generate the characteristic gear vibrational patterns dominated by the so-called gear mesh frequency (GMF) (Kuang and Li, 2003). This effect is schematically shown in Figure 5.1

Besides GMF components, gear vibrations are accompanied by modulations that originate from assembly errors and fluctuations in gears speed and load. Thus, vibrations

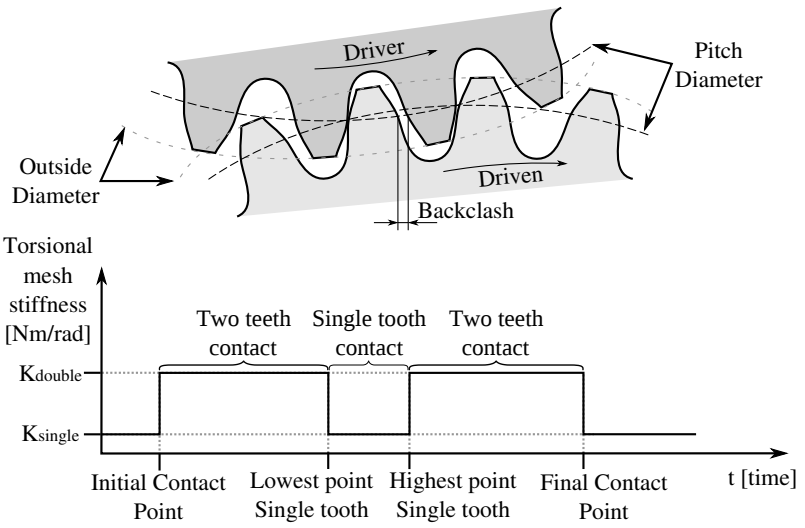


Figure 5.1: Variations in mesh stiffness

$y(t)$, produced by meshing gears that rotate with constant speed, can be represented as follows (Wang, 2001):

$$y(t) = \sum_{m=1}^{M} A_m\big(1 + a_m(t)\big)\cos\big(2\pi m \times f_{gmf}t + \beta_m + b_m(t)\big), \tag{5.1}$$

where f_{gmf} is gear mesh frequency, M is the highest harmonic of interest and A_m is amplitude of the m^{th} harmonic of the gear mesh frequency. The components $b_m(t)$ represent phase modulations, and β_m is the initial phase. The amplitude modulations $a_m(t)$ are due to incipient localized tooth damages, i.e. surface faults that occur only on a limited number of gear teeth. These modulations contain all the essential diagnostic information (Özguven and Houser, 1988; Randall, 1982).

Fault influence

Gear faults evolve through various severity stages (GfT, 2002):

Scuffing This stage is characterised by removal of flat particles from surfaces, for example due to material or manufacturing faults or overuse.

Initial pitting At this stage, locally formed dimples (pits) occur, mainly due to initial over-stressing,

Pitting Shell-shaped depressions form by removal of material particles due to surface disorder, and

Spalling This stage is characterised by breaking within or at the surface coating or surface layer.

Regardless of the severity stage, the presence of faults affects the physical characteristics of the damaged teeth. Localised surface faults, present only on small number of teeth, affect the stiffness of the meshing when the damaged teeth are in contact. Since these damaged teeth enter the mesh once per revolution, their influence is felt as a periodic modulation of the mesh stiffness. As a result of this effect, the vibration signal is enriched by an additional amplitude modulation around the GMF, i.e. change in $a_m(t)$ from (5.1). Thus, the spectrum of this altered vibration signal contains sidebands around the dominant f_{gmf} spectral components. The sideband spacing is directly related to the rotational speed of the damaged gear (Randall, 1982). It is worth noting that the presence of phase modulations, due to shaft torsional stiffness, can cause asymmetry in the sidebands amplitudes.

As the faults gradually spread accross the majority of gear's teeth, the influence of the amplitude modulation decreases. However, surface damages alter the meshing stiffness causing overall increase in the amplitudes of the dominant GMF components. The present amplitude modulations, on the other hand, will not vanish completely since the teeth are not equally damaged.

These observations about vibrational patterns due to healthy and damaged gears lead to conclusion that the presence of the fault affects the amplitude modulation $a_m(t)$ in (5.1). Therefore, the main goal of an effectual feature extraction procedure will be careful monitoring of the changes happening on $a_m(t)$. This, task is fairly simple in cases of known and constant operating conditions. Chapter 7 will present an approach for feature extraction using $a_m(t)$ for gears running under variable and unknown operating conditions.

5.2 Bearing vibration model

Bearings have more complicated structure than gears. Generally, bearings consist of inner ring, outer ring and rolling elements (balls) positioned between the two rings. Usually, the inner ring is the rotating one, while the outer ring is stationary. However, there are cases where the opposite is true. Bearings are specified by three characteristic dimensions (T.A.Harris, 2000): the radius of the outer ring D, the radius of the rolling element d and the contact angle α, as shown in Figure 5.2. The first two parameters determine the outer and inner shaft radii and the contact angle defines the axial deflection between the inner and outer ring. The value of the contact angle is related to the maximal allowed axial load of the bearing.

In comparison to gears, fault–free bearings produce negligible vibrations. One of the

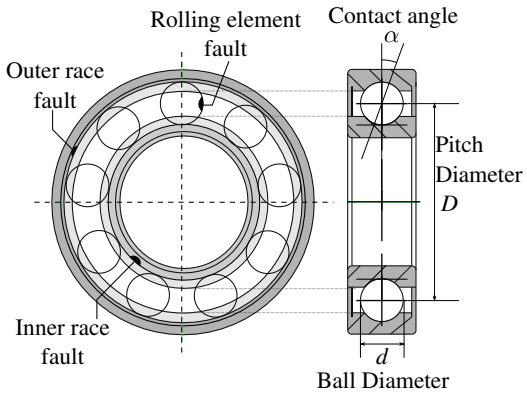


Figure 5.2: Bearing dimensions used for the calculation of the bearing characteristic frequencies

main causes for bearing faults is material fatigue. Over time, areas with surface damages occur in a form of small pits. These pits can appear on any bearing element. When rotating, impacts occur each time a rolling element passes across the damaged surface. These impacts excite the entire mechanical structure impulse responses $s(t)$, hence the generated vibrations can be modelled as (Randall et al., 2001) :

$$y(t) = \sum_{i=-\infty}^{+\infty} A_i s(t - \tau_i) + n(t), \tag{5.2}$$

where A_i is the amplitude of force that excites the entire structure and τ_i is the time of its occurrence. The final component $n(t)$ defines an additive random component that contains all non–modelled vibrations as well as environmental disturbances. The amplitudes A_i of the impulses should be considered as random process since the amplitude of each impact differs (Antoni and Randall, 2003).

Under constant operating conditions generated impacts may be considered as truly periodic, so the impact times τ_i in (5.2) become iT, where T is the period of occurrence. The period T is directly related to the type and location of the surface fault through the relations listed in Table 5.1 (Tandon and Choudhury, 1999).

bearing pass frequency inner race (BPFI)	$f_{BPFI} = \frac{Zf_{rot}}{2}\left(1 + \frac{d}{D}cos\alpha\right)$
bearing pass frequency outer race (BPFO)	$f_{BPFO} = \frac{Zf_{rot}}{2}\left(1 - \frac{d}{D}cos\alpha\right)$
fundamental train frequency (FTF)	$f_{FTF} = \frac{f_{rot}}{2}\left(1 - \frac{d}{D}cos\alpha\right)$
ball spin frequency (BSF)	$f_{BSF} = \frac{Df_{rot}}{2d}\left(1 - \left(\frac{d}{D}cos\alpha\right)^2\right)$

Table 5.1: Characteristic bearing fault frequencies

Conceptual model of lubrication influence on vibration

Besides mechanical faults, bearing lubrication has significant influence on the generated vibrations too. Despite this fact, the problem of detecting improper bearing lubrication via vibration analysis has not received wider attention. For that purpose this section presents a conceptual model of vibrations generated by lubrication starved bearings.

Bearing lubricant has four different functions (Howard, 1994):

i) provides a separation layer between adjacent surfaces thus preventing wear,

ii) acts as a coolant,

iii) prevents contamination of the bearing raceways with foreign debris and

iv) prevents corrosion of the bearing's elements.

Degraded or insufficient lubrication in bearings can deteriorate these functionalities. One can expect that changes in the lubricant can be presumably detected as alterations in the resulting vibrations.

The influence of elasto-hydrodynamic lubrication (EHL) on bearing vibrations can be observed by focusing on the influences of the lubrication on the microscopic material level (Aini et al., 2002; Dowson, 1995), or by applying rough estimates of the underlying influences on the generated vibrations (Igarashi, 1964). Such a conceptual model offers sufficiently good explanation of the origins of the observed vibration patterns for the cases of improper bearing lubrication.

The mechanism and intensity of impact that lubrication has on vibrations can be explained by combining the results from the model of Wijnant et al. (1999) with the results from the simplified bearing vibration model developed by Sawalhi and Randall (2008). Wijnant et al. (1999) modelled the contact among inner ring, outer ring, the rolling element and the elastohydrodynamic lubrication as a spring-damper mechanism. The damping contribution is mostly expressed in the unloaded zone of a lubricated bearing (Lynagh et al., 2000). Provided that there are no changes in the bearing's surface structure, the parameters of the spring-damper mechanism depend only on the lubricant characteristics. Additionally, lubrication layer also fills the clearance in the bearings or even in some cases adds a pre-tension to the bearing elements i.e. a negative clearance. Sawalhi and Randall (2008) used this fact in order to incorporate the EHL effect in the model of bearing vibrations.

By employing the spring–damper model for the individual contacts, Sawalhi and Randall (2008) developed a model for bearing vibrations for fault–free bearing as well as for different kinds of bearing faults. The model takes into account vibrations produced

by each rolling element passing through a load zone. Under constant rotational speed of the inner ring and stationary outer ring, the rolling elements pass through the load zone periodically with a predefined cage frequency. Consequently, as soon as a rolling element enters the load zone, elastic deformation occurs by slight movement of the inner ring towards the outer ring. In the presence of lubricant, this movement is very limited because the lubricant fills the clearance inside the bearing. Additionally, this movement is damped by the lubricant's spring–damper characteristics. Thus, in fault–free bearings these movements do not cause any significant increase in the bearing's vibrations.

In the case of lubrication starved bearing, the absence of lubricant film in the bearing clearance introduces an additional space when a ball passes through the load zone. Furthermore, the lack of lubricant removes the spring–damper mechanism from the model. As a consequence of these changes, vibrations produced by balls passing through the load zone should be stronger compared to the ones produced by a properly lubricated bearing. Thus, the increase of amplitude of vibrations produced by the rotating bearing balls can be used as indication of either deteriorated lubrication quality or a complete lack of lubricant.

Each time a rolling element enters the load zone it is subjected to a deformation under the pressure emanating from the inner ring. Usually this deformation is damped by the elastic characteristic of the rolling element itself as well as the spring–damper characteristic of the lubricant. In cases of improper lubrication the contact between the passing rolling element and the bearing's inner and outer ring excites the system's impulse response $s(t)$.

6 Bearing fault detection under quasi-stationary operating conditions

Among components of mechanical drives, bearings are mostly exposed to failures. Therefore, incipient bearing fault detection is of great practical importance. Commonly, the detection of such faults is performed under assumption of strictly constant operating conditions and the generated vibrations are regarded as stationary signals. Since the assumption of constant speed does not always hold, the actual vibration signals breach the condition of stationarity. However, when the speed fluctuations are minor, e.g. in the range of several percent, the generated vibration signals can be treated as (quasi)cyclostationary.

Notwithstanding the small variations in the rotational speed, the envelope analysis can be still applied for study of the generated vibrations. The fact that the underlying signals are cyclostationary offers an additional way to determine the most informative frequency band in which the demodulation process should take place. This chapter presents two methods that can be employed for the selection of the demodulation band: spectral coherence which relies on the cyclostationary nature of the signal and spectral kurtosis which relies on the signals' time-frequency statistical characteristics.

6.1 Frequency band selection based on cyclostationary analysis

Under mild variations in the rotational speed the bearing signal model (5.2) can be rewritten as:

$$y(t) = \sum_{i=-\infty}^{+\infty} A_i s(t - iT - \tau_i) + n(t), \tag{6.1}$$

where T is the period between two successive impacts in case of bearing fault in accordance with relations shown in Table 5.1, τ_i is a zero mean Gaussian random variable satisfying $E\{\tau_i \tau_j\} = \delta_{ij}\sigma_\tau^2$ with probability density function $f_\tau(\tau)$ and $n(t)$ is a random process that describes all the unmodelled effects and environmental disturbances. Furthermore, due to the periodicity of rotation, the statistical properties of the random amplitudes A_i can

be regarded periodic with period Q as:

$$\begin{aligned} E\{A_i\} &= E\{A_{i+Q}\} \\ E\{A_i^2\} &= \sigma_A^2 \\ Cov\{A_i, A_j\} &= Cov\{A_{i+Q}, A_{j+Q}\}. \end{aligned}$$

The first step in this analysis is the calculation of the spectral correlation $\mathcal{S}_x(f;\alpha)$ (2.15). This can be accomplished using the fact that the autocorrelation function of a cyclostationary signal is periodic i.e. for the signal (6.1):

$$\rho_y(t,\tau) = \rho_y(t+mT,\tau), \quad m \in \mathbb{Z}.$$

Due to the periodicity, the autocorrelation function $\rho_y(t,\tau)$ can be expanded into Fourier series as in (2.13). Having these coefficients, the cyclic power spectrum $\mathcal{S}_y(f;\alpha)$ can be calculated according to (2.15). Following this approach the cyclic power spectrum $\mathcal{S}_y(f;\alpha)$ for the signal (6.1) becomes (Randall et al., 2001, Eq. (18)):

$$\begin{aligned} \mathcal{S}_y(f;\alpha) = \frac{1}{T}\hat{S}\left(f+\frac{\alpha}{2}\right)\hat{S}\left(f-\frac{\alpha}{2}\right)^* \left((1-\sigma_A^2)\varphi_\tau(\alpha) - \varphi_\tau\left(f+\frac{\alpha}{2}\right)\varphi_\tau\left(f-\frac{\alpha}{2}\right)^*\right) \\ \times \sum_{i\in\mathbb{Z},q\in\mathbb{Z}} \mathcal{A}_q \delta\left(f-\frac{i}{T}-\frac{q}{Q}\right) + \delta(\alpha)\Phi_n(f), \end{aligned} \tag{6.2}$$

where $\hat{S}(f)$ is the Fourier transform of of the impulse response $s(t)$, $\varphi_\tau(f)$ is the Fourier transform of $f_\tau(\tau)$, $\Phi_n(f)$ is the power spectral density of $n(t)$ and $\mathcal{A}_q$ are the Fourier series coefficients of the periodic mean $E\{A_i\}$. It is assumed that the mean period of impacts T and the period of the load variations Q are incommensurable.

Taking into consideration that the power density $\Phi_n(f)$ of the noise $n(t)$ and the transfer function $\hat{S}(f)$ are sufficiently smooth for small frequency shifts, then

$$\begin{aligned} \Phi_n(f \pm \alpha/2) &= \Phi_n(f) \\ \hat{S}(f \pm \alpha/2) &= \hat{S}(f). \end{aligned}$$

Furthermore, the bandwidth of the $\varphi_\tau(f)$ is directly related to the standard deviation of the σ_τ of the random fluctuations. In particular, for bigger fluctuations the bandwidth of $\varphi_\tau(f)$ decreases. By assuming mild fluctuations $\sim 1\%$ and Gaussian distribution for τ_i, the bandwidth of $\varphi_\tau(f)$ is such that $\varphi(f+\alpha/2) \approx 0$. With these assumptions, inserting (6.2) into (2.20), the final form of the squared spectral coherence becomes (Antoni, 2007b, Eq. (15))

$$|\rho_x^\alpha(f)|^2 = \left|\frac{SNR(f)}{1+SNR(f)}\right|^2 |\Phi(\alpha)|^2 \sum_{i\in\mathbb{Z},q\in\mathbb{Z}} \frac{\mathcal{A}_q}{\mathcal{A}_0}\delta\left(f-\frac{i}{T}-\frac{q}{Q}\right), \tag{6.3}$$

where

$$SNR(f) = \frac{(1+\sigma_A^2)\mathcal{A}_0}{T}\frac{|S(f)|^2}{S_n(f)}.$$

The relation (6.3) shows that the amplitude of $|\rho_x^\alpha(f)|^2$ in f domain can serve as indication in which frequency band the signal-to-noise ratio (SNR) is the highest, hence determining the most suitable frequency band for envelope analysis. Additionally, the presence of cyclic components in α domain serves as the diagnostic information about the present fault.

Simulated example

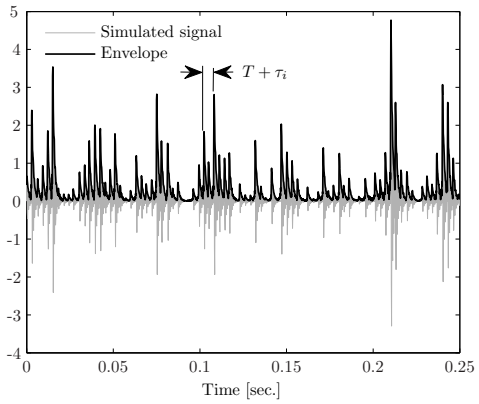
The properties of the spectral correlation density can be easily demonstrated by simulating a bearing fault using (6.1) as:

$$\begin{aligned} y(t) &= \sum_{i=-\infty}^{+\infty} A_i s(t - iT - \tau_i) + n(t) \\ s(t) &= e^{-100t}\sin(2\pi 1500 t) \end{aligned} \tag{6.4}$$

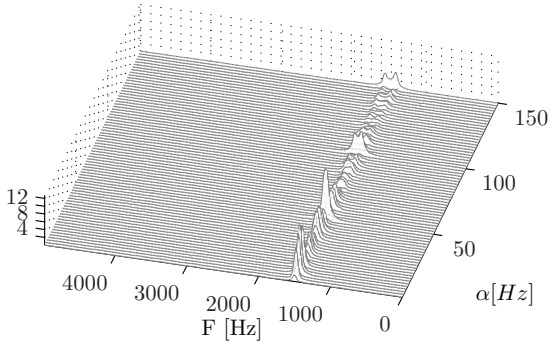
where $A_i \sim \mathcal{N}(0, \sigma_{A_i}^2 = 0.5)$, $\tau_i \sim \mathcal{N}(0, \sigma_{\tau_i}^2 = 0.05T)$, $T = 50$ Hz and $n(t)$ defines an additive random component so the overall SNR=1. One realisation of the signal (6.4) is shown in Figure 6.1a.

SCD (6.2) of the simulated vibration signal is shown in Figure 6.1b. As expected, the maximal values of SCD are reached at frequency $f \approx 1.5$ kHz. Although the SCD plot should show periodicity along the α axis, such effect is not visible. The maximal value is reached at $\alpha = 50$ Hz, however similar peaks are not clearly visible for other integer multiples of α. The anticipated cyclic periodicity is obscured by the comparably high amplitudes of SCD for other cyclic frequencies α. Therefore, due to the scaling problem it is not always possible to use SCD plots to detect the proper set of cyclic frequencies α for which the observed signal exhibits cyclostationary behaviour.

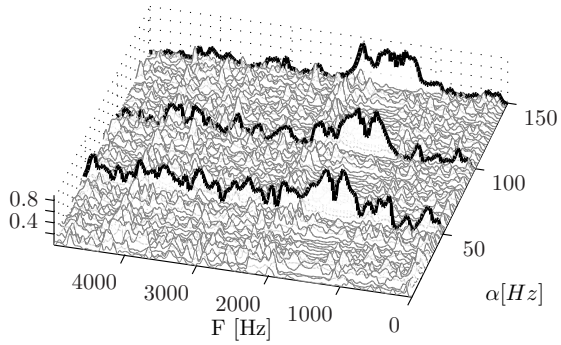
Contrary to SCD, the plot of SCOH (6.3) shows distinctive cyclic components α, as shown in Figure 6.1c. These frequencies, marked with thicker lines, are $\alpha \in \{50, 100, 250\ Hz\}$. Additionally, the amplitude of the spectral coherence for a specific cyclic frequency α is related to the SNR, as shown by (6.3). The frequency band in which the spectral coherence reaches maximal value is located correspond to $\alpha = 50$ Hz and frequency band around 1.5 kHz. These values are directly connected with the parameters T and the impulse response $s(t)$. Thus, the frequency bands, in which the values of SCOH become high, indicate the possible frequency bands in which the envelope analysis should be conducted.



(a) Simulated signal



(b) Spectral correlation density (6.2) with magnitude signal–units^2/Hz



(c) Spectral coherence (2.20) of signal with marked dominant cyclic frequencies α

Figure 6.1: Estimated SCD and SCOH for the simulated version of the signal (6.4)

6.2 Frequency band selection based on spectral kurtosis

spectral kurtosis (SK) was firstly introduced by Dwyer (1983), as a method for detection of randomly occurring events. The method exploits the possibility of using kurtosis as a measure of distance between an arbitrary random process and a Gaussian one. The value of SK of a random signal is calculated by applying these statistical concepts in the frequency domain i.e. by calculating kurtosis of each frequency component contained in the observed random signal.

As already stated in Section 2.1.2, complex spectral components $dZ(\omega)$, for the corresponding stationary random process $x(t)$, can be considered as circular complex r.v. Furthermore it was shown that for such variables the only non-null moments are those constructed with the same power in Z and Z^*, where Z^* is the complex conjugate. The same limitation is valid for the cummulants of circular complex r.v. i.e. for a $Z(\omega)$ there are $n+1$ different n-order cummulants:

$$\kappa_{Z,n}(\omega) = \kappa_n\{\underbrace{Z(\omega),\cdots,Z(\omega)}_{n-p},\underbrace{Z^*(\omega),\cdots,Z^*(\omega)}_{p}\}. \tag{6.5}$$

Consequently, the kurtosis value for each frequency ω is defined as (Vrabie et al., 2003)

$$\gamma_2(\omega) = \frac{\kappa_{Z,4}(\omega)}{\left[\kappa_{Z,2}(\omega)\right]^2},$$

where $\kappa_{Z,4}(\omega)$ and $\kappa_{Z,2}(\omega)$ are the fourth-order and the second-order cumulant respectively of a complex random variable $Z(\omega)$. Due to circularity the only non-null cumulants for the spectral components $Y(f)$ of the observed signal $y(t)$ defined by (6.5) will be:

$$\kappa_{Z,4}(\omega) = \kappa_4\{Z(\omega), Z(\omega), Z(\omega)^*, Z(\omega)^*\}$$
$$\kappa_{Z,2}(\omega) = \kappa_2\{Z(\omega), Z(\omega)^*\}.$$

Finally, by applying the relationship between cumulants and moments of a complex random variables (Mendel, 1991), and taking into consideration the circular nature of $Z(\omega)$ the final form of SK relation becomes (Antoni and Randall, 2006):

$$\gamma_2(\omega) = \frac{E\{|Z(\omega)|^4\}}{E\{|Z(\omega)|^2\}^2} - 2$$

Since the value of kurtosis indicates the presence of sporadic high-energy bursts in random fluctuations, the spectral components whose amplitude changes abruptly in time will obtain higher kurtosis. By selecting the frequency component of the signal that has the highest kurtosis we actually select the frequency band where the impulses within the signal are most expressed.

Estimation of spectral kurtosis of the simulated signal

The estimation of SK for a particular frequency f is done by calculating STFT of the observed signal. It should be noted that the selected window length and the percent of overlap used for the STFT calculation significantly influence the value of the spectral kurtosis. Therefore, for the purpose of selecting the optimal window length the spectral kurtosis is calculated for several different window lengths. The result of these calculations is the so–called *kurtogram* diagram Antoni (2007c). The window parameters for which the spectral kurtosis has its maximum value determine the band–pass filter parameters.

The time-frequency characteristic of signal (6.1) is shown in Figure 6.2. It is noticeable that the highest peaks are around 1.5 kHz, which was the chosen simulated eigenfrequency of the impulse response $s(t)$. The amplitudes of the spectral components around 1.5 kHz vary in time considerably, compared to the ones above and below this frequency band. At this moment we should note that by calculating the average over time we will obtain the standard power-spectral density.

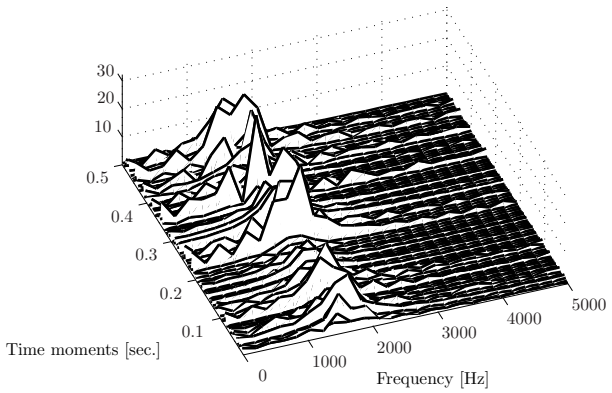


Figure 6.2: Short-time Fourier analysis of the signal $y(t)$, see Eq. (6.1)

The kurtogram for the signal (6.1) is shown in Figure 6.3. The maximal value of SK is obtained in the frequency band with central frequency $f_c = 1500$ Hz and bandwidth $B_w = 625$ Hz, and the maximum of SK in that frequency band is 1.4.

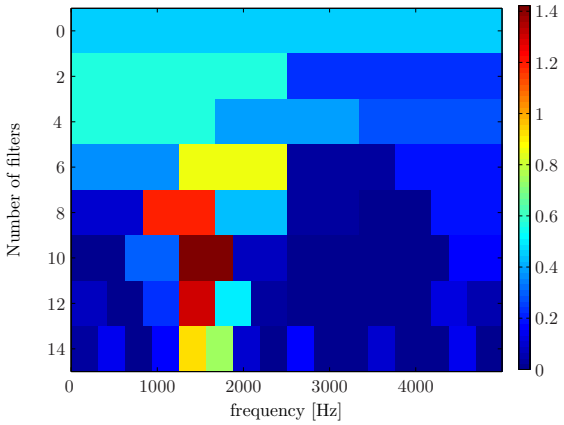


Figure 6.3: Kurtogram of the simulated signal $y(t)$ defined by Eq. (6.1)

The obtained frequency band coincides with the one calculated from the cyclic coherence plot (see Figure 6.1 (c)). Generally, the use of only one of the proposed methods is sufficient for the task of bearing fault detection. However, the subsequent results will show that the detection of incipient bearing faults as well as detection of improperly lubricated bearings, can be enhanced by analysing the generated vibrations using both approaches.

6.3 The experimental results

To study the problem of bearing faults and improper lubrication detection, we conducted experiments on a set of 130 EC motors. The mechanical damages were introduced using electric discharge machining (EDM). The experimental set contained damaged specimens

with different fault severity. Furthermore, the specimens covered all possible bearing surface faults as specified in Chapter 5. The lubrication starved bearings were prepared by cleaning with *tetrachloroethylene*. This procedure effectively removed all grease in the bearing. During the cleaning process, a special care was taken to prevent entry of any foreign debris within the bearing raceways. The reason was that the cleaning process required removal of bearing's shields.

Suitably prepared bearings were mounted on electronically commutated motors, which were subsequently tested using the test–rig shown in Figure 6.4. The rig consisted of a fixed pedestal on top of which a metal disk was positioned. The rubber dampers suspended the test EC motor. The experiment commenced by positioning the EC motor vertically on the rubber dampers, in such a way that the drive–end bearing was on the bottom side. Afterwards, two accelerometers were positioned on the motor housing nearest to the both bearings. Such a test–rig minimized the environmental influence, thus providing considerably constant experimental conditions.

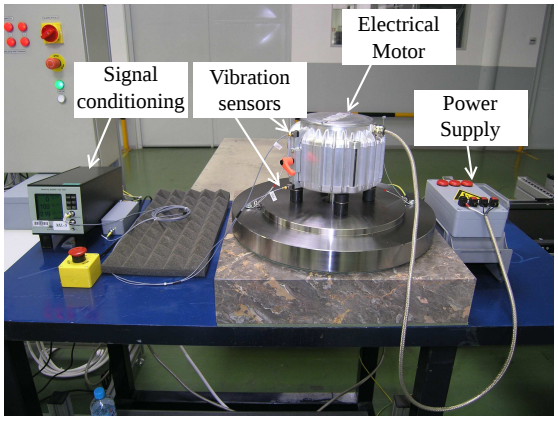


Figure 6.4: The test rig

The data acquisition process commenced as soon as the nominal rotational speed was reached. Firstly, both vibration signals were low–pass filtered with cut–off frequency at 22 kHz. Afterwards, both signals were sampled at 60 kHz. During the whole data acquisition process the nominal rotational speed of $f_{rot} = 38$ Hz was maintained. Each acquisition process lasted 8 seconds. After finishing the acquisition the motor was decelerated down to the stop position. All experiments were conducted using FAG 6205 bearings. The principle fault components are given in Table 6.1.

The analysis of vibration signals acquired from the experimental runs was conducted using a two step approach. Firstly the proper band–pass filter parameters were determined by combining the results from cyclostationary analysis and spectral kurtosis method. After that, features were extracted from the envelope spectra of the filtered vibration signals. Features extracted in such a manner provided an unambiguous distinction between fault–

Bearing fault	f/f_{rot}
bearing pass frequency inner race (BPFI)	5.415
bearing pass frequency outer race (BPFO)	3.585
fundamental train frequency (FTF)	0.398
ball spin frequency (BSF)	2.375

Table 6.1: Calculated characteristic bearing fault frequencies for FAG 6205

free and lubrication starved bearing.

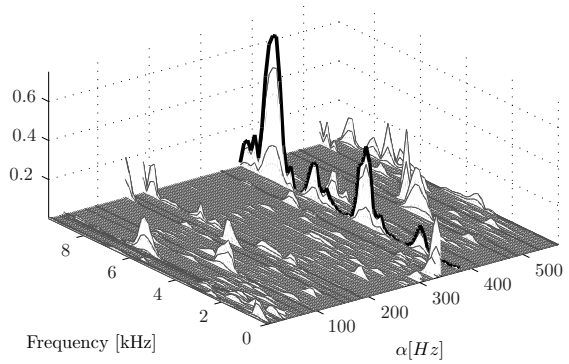
6.3.1 Cyclostationary analysis of the vibration signals

The cyclostationary analysis was based on spectral coherence of the vibration signal. The spectral coherence plots of vibration signals taken from a motor with fault–free bearings and those with damaged ones are shown in Figure 6.5. The plots are calculated using cyclic resolution $\Delta\alpha = 1$ Hz.

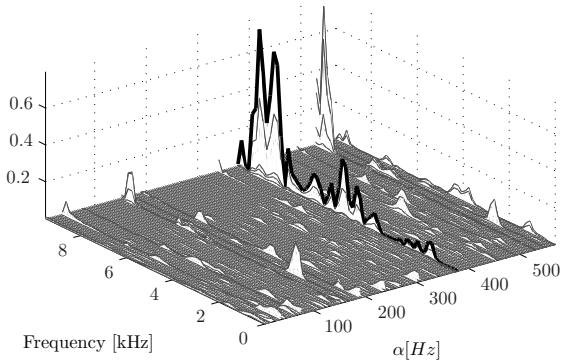
The SCOH plot of vibration signal with fault-free bearings has distinctive cyclic components at $\alpha = 378$ Hz, as marked in Figure 6.5a. The origin of this cyclic frequency can be attributed to the motor's power supply. The EC motor in question has 5 pole stator windings powered by pulse-width modulation (PWM). The frequency of the PWM is $f_{pwm} = 193.5$ Hz, as a consequence of which the most dominant cyclic frequency $\alpha = 2 \times f_{pwm}$.

The SCOH plot of vibrations of improperly lubricated bearing is similar to the one of fault-free runs. The most dominant cyclic component is at $\alpha = 378$ Hz, as marked in Figure 6.5b. Since lubrication starved bearings contain no mechanical surface damage, the main source of vibrations are uneven rotation due to the PWM supply. However, by comparing the amplitude of spectral coherence in f domain for fixed $\alpha = 387$ Hz it becomes noticeable that spectral coherence plots differ in the frequency interval [8.2-9 kHz]. In that interval the spectral coherence of lubrication starved bearing shows significant increase compared to the one calculated for the properly lubricated bearing.

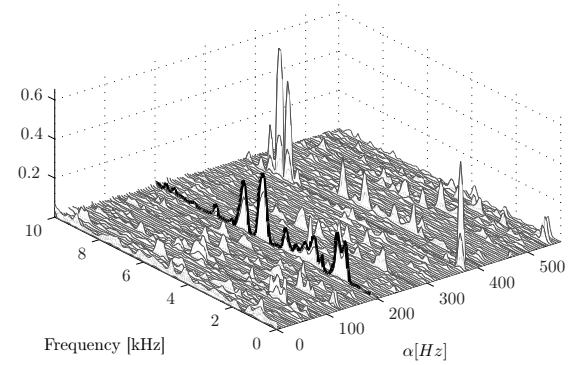
Unlike the case of improperly lubricated bearings, the cases of inner and outer ring damages exhibit distinctive changes in their spectral coherence plots. The case of inner ring damage, shown in Figure 6.5c, has distinctive spectral component located at $\alpha \approx 204$ Hz. Considering the characteristic bearing fault frequencies from Table 6.1 and the fact that rotational frequency was kept constant $f_{rot} = 38$ Hz, the corresponding cyclic frequency matches the characteristic fault frequency for inner race damage. Similar conclusion is valid for the case of outer race fault, shown in Figure 6.5d. In this case,



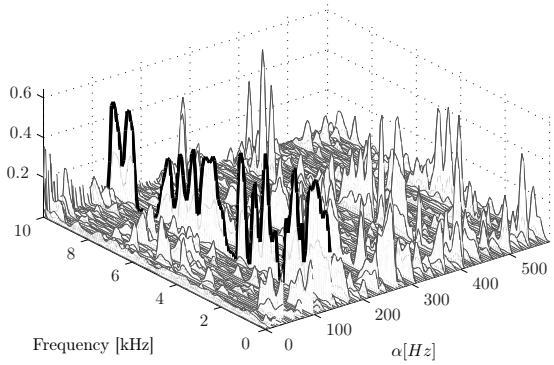
(a) Fault-free bearing



(b) Lubrication starved bearing



(c) Bearing with surface damage on the inner ring



(d) Bearing with surface damage on the outer ring

Figure 6.5: Spectral coherence plots

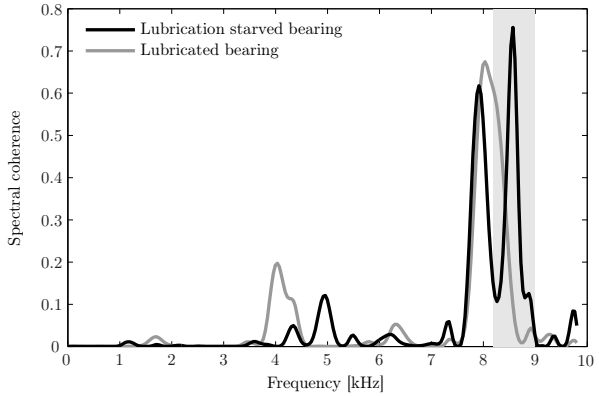
the dominant cyclic frequency is $\alpha \approx 136$ Hz and it corresponds to the characteristic frequency of damaged outer ring.

Similarly like for the lubrication starved bearing, the amplitude of SCOH for the dominant cyclic frequencies can be used to determine the most informative frequency band for demodulation. The plot for SCOH at $\alpha = 204$ Hz for inner race fault is shown in Figure 6.6b. The maximum of SCOH is located in the frequency band around 5 kHz, as marked on the figure. The SCOH plot for the case of outer ring damage, shown in Figure 6.5d, has several peaks with similar amplitude. Therefore, all these frequency bands can be considered for the demodulation and consequent envelope analysis.

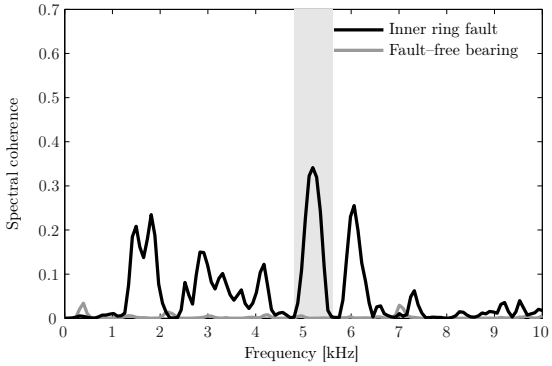
6.3.2 Analysis of spectral kurtosis of the vibration signals

The band–pass filter parameters obtained by cyclostationary analysis were confirmed by applying SK method to the vibration signals for the same cases. The SK was estimated by calculating a wavelet based kurtogram diagram developed by Sawalhi et al. (2007).

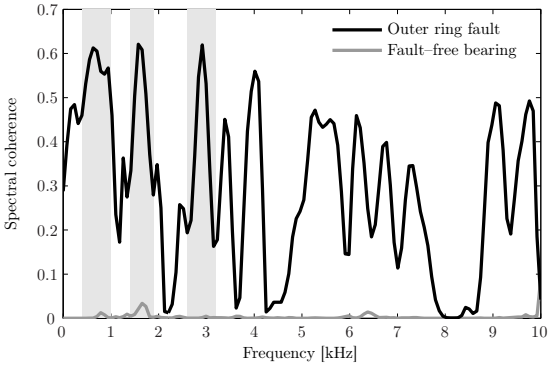
The kurtogram diagrams for all four cases are shown in Figure 6.7. The frequency



(a) Lack of lubrication fault at $\alpha = 387$ Hz



(b) Inner ring damage at $\alpha = 204$ Hz



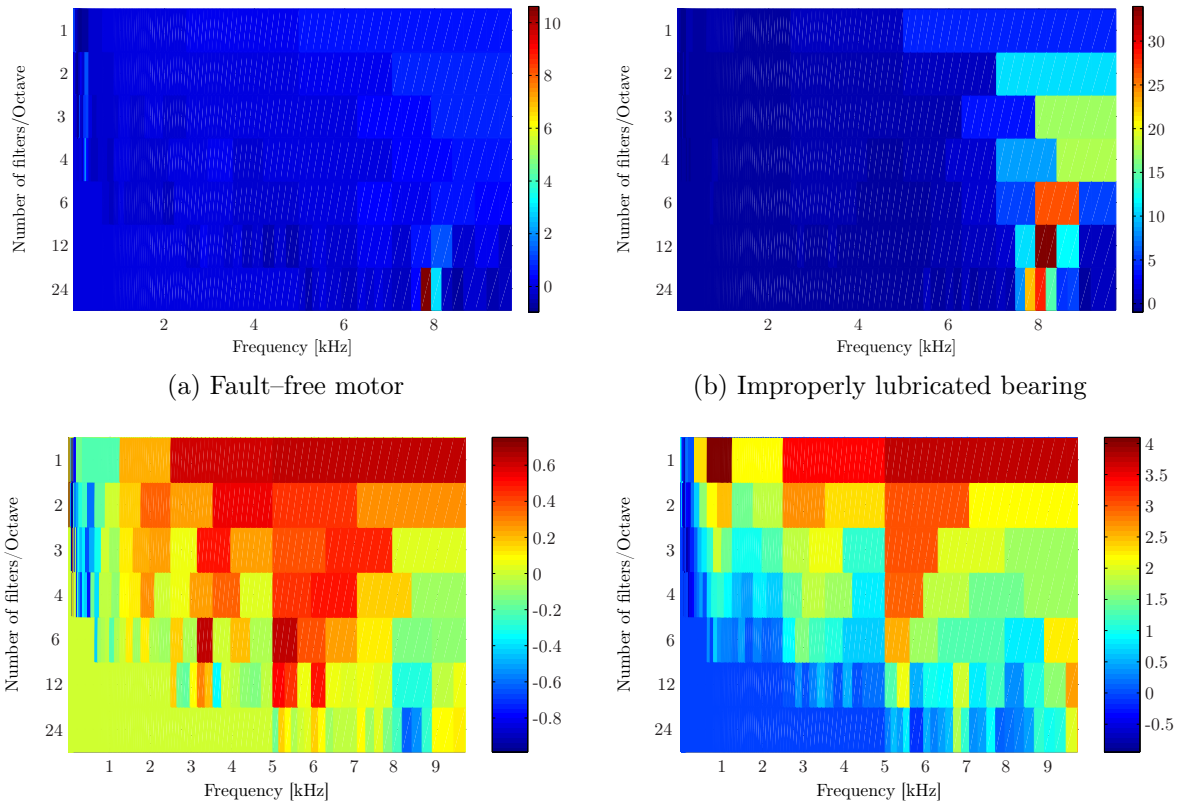
(c) Outer ring damage at $\alpha = 136$ Hz

Figure 6.6: Spectral coherence plot for the dominant cyclic frequencies

band in which the spectral kurtosis reaches its maximum determines the parameters for the demodulation and envelope analysis. In case of lubrication starved bearings the maximal spectral kurtosis is reached at 8.5 kHz, as shown in Figure 6.7b. This result is almost the same as the one reached by the analysis of the spectral coherence.

Unlike improperly lubricated case, the kurtogram diagrams of runs with inner and outer race faults have several distinctive peaks, marked with dark red color in Figure 6.7 (c) and (d). For the case of inner race fault, the two frequency bands with the highest values of SK are at level 6 and centred around 3 and 5 kHz. These two intervals correspond to the two extremes in the spectral coherence plot from Figure 6.5c. Similarly, the outer race fault kurtogram reaches maximal values at level 1 around 1 kHz and in the strip 3–5 kHz, which coincides with the local maxima marked in Figure 6.5d.

Although the usual approach is to select the frequency band which maximises the value of SK, the kurtogram diagrams shown in Figures 6.7 (c)-(d), show SK values are comparably high in multiple frequency bands. So by comparing these values with the values calculated from the spectral coherence plots one can determine which of these possible candidates is the most suitable.



(a) Fault–free motor

(b) Improperly lubricated bearing

(c) Bearing with inner race damage

(d) Bearing with outer race damage

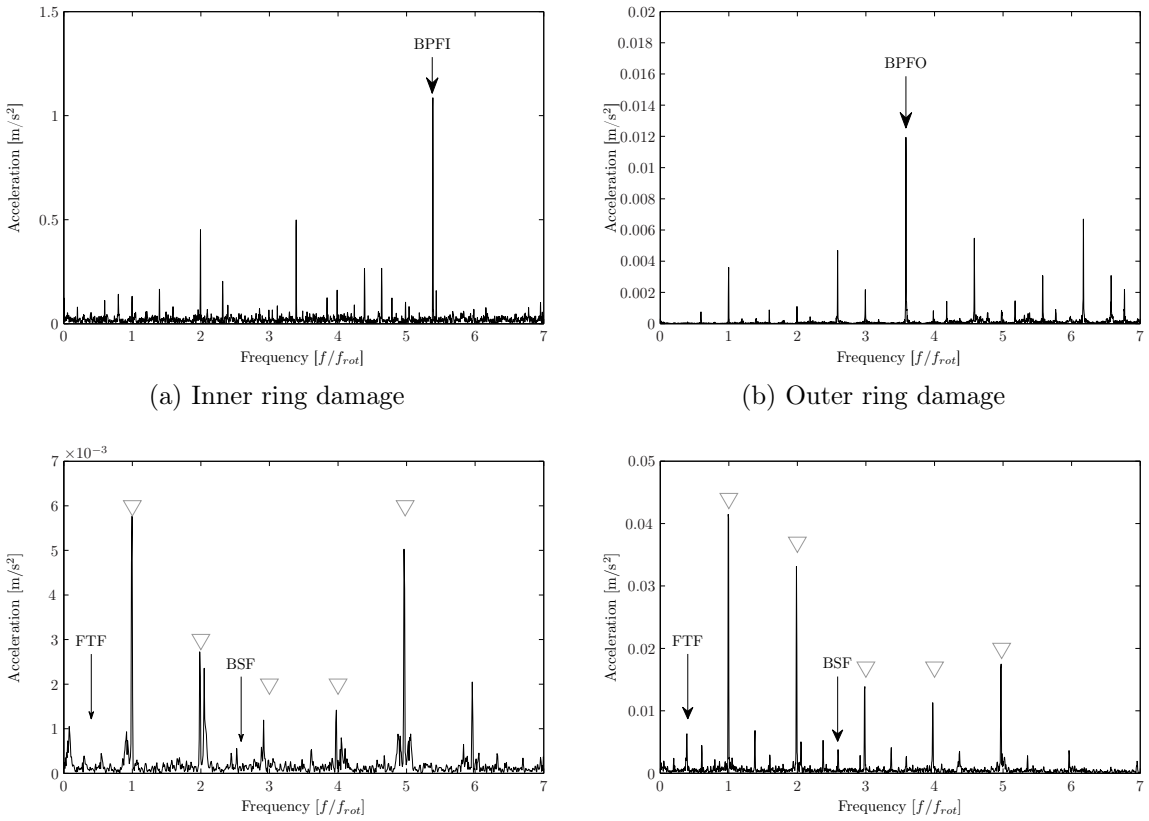
Figure 6.7: Kurtogram diagrams of vibration signal

6.3.3 Feature extraction for bearing faults

Localisation of a particular bearing fault is based on its vibration excitation patterns. For the most common bearing faults these patterns can be identified as specific modulation frequencies of the vibration signals. These modulation frequencies can be calculated based on the bearing's geometrical characteristics and the bearing rotational speed (cf. Table 5.1). Consequently, the presence of any of these frequencies in the vibration envelope is a clear sign for a particular bearing fault.

The envelope spectra for inner and outer race faults are shown in Figure 6.8 (a) and (b). The spectral components at the characteristic bearing frequencies for inner and outer race faults (BPFI and BPFO) dominate the corresponding spectra. Such a result is a clear evidence for fault presence. So the amplitudes of these two spectral components, calculated from the corresponding envelope spectra, can be regarded as two features in the condition monitoring system.

Conversely, the set of appropriate spectral components for estimating the lubrication quality of the bearing has not been clearly specified. Therefore, the selection of features



(a) Inner ring damage

(b) Outer ring damage

(c) Fault–free bearing

(d) Improperly lubricated bearing

Figure 6.8: Envelope spectra for different bearing faults. The spectral component of the rotational speed and its higher harmonics are marked with ▽.

relevant for detecting lubrication starved bearings has been done by comparing the envelope spectra of the filtered vibration signals from the properly and improperly lubricated bearing. Both spectra, calculated from the filtered in the determined frequency band, are shown in Figure 6.8 (c) and (d). Although both spectra are similar, the spectrum of the lubrication starved bearing shows significant increase in the amplitudes of the spectral components at FTF and BSF which are absent from the spectrum of properly lubricated bearing. Since the lubrication layer acts as a damper and, at the same time, fills the gaps of the bearing clearance, its absence allows bearing's cage to oscillate more freely between the bearing's inner and outer ring.

Based solely on the two spectral components at FTF and BSF we have conducted a test on a batch consisting of 84 motors, from which 63 motors were fitted with properly lubricated bearings and 21 motors with lubrication starved bearings. The amplitude of the FTF spectral component was recorded as the relevant feature. The histogram presenting the distribution of these amplitudes is shown in Figure 6.9. It is clearly visible that the motors with properly lubricated bearings have distinctively lower amplitudes of the

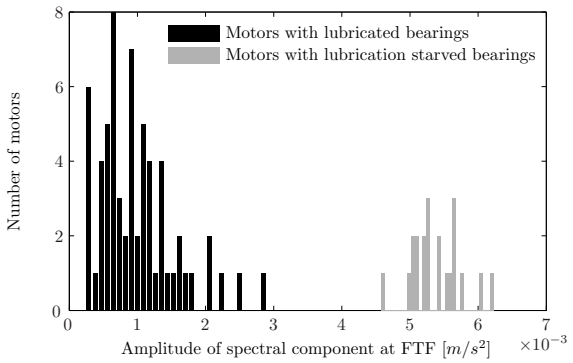


Figure 6.9: Histogram of measured motors

FTF components than the corresponding cases with lubrication starved bearings. Such distinction allows seamless detection of motors with lubrication starved bearings.

7 Fault diagnosis under variable operating conditions

The traditional vibration-based diagnostic approaches for rotational machines are largely designed for stationary and known operating conditions. Consequently, a change in the pattern of fault signatures can be directly associated with the changes in condition of the monitored machine. In case of fluctuations in the operating conditions, these associations become ambiguous, because changes in the operational regime usually influence the produced vibrational patterns. Therefore, reliable fault detection and isolation, irrespective to changing and unmeasurable operating conditions, is of significant practical merit.

The non-stationary nature of the signals, generated by machines operating under variable conditions, significantly limits the number of applicable signal processing tools. Generally, such signals are analysed using time-frequency methods such as Wigner-Vile distribution or wavelet transform, since these tools impose no constraints regarding the statistical characteristics of the analysed signal. Despite this fact there are not many examples that utilise these methods for the purpose of fault detection. Some of the most notable achievements is the work of Poulimenos and Fassois (2006) with their comprehensive analysis of non-stationary vibration signals in the time domain, Padovese (2004) for the application of autoregressive models for transient signal analysis and Baydar and Ball (2000), who performed detection of gear deterioration under different loads using instantaneous power spectrum by employing WVD. These contributions support the idea that a suitable solution for the problem of feature extraction under variable operating conditions may be found.

In this chapter we attempt to deliver some novel results on the way the operating conditions influence the instantaneous power distribution of the generated vibrations and the way the changes in the distribution can serve as features in the process of fault detection. The instantaneous power is analysed from two different viewpoints:

i) directly from the envelope distribution and

ii) by utilising the point process framework.

In both cases, wavelet packet analysis is employed for describing the statistical properties of the instantaneous power of the generated vibrations.

7.1 Mechanical faults signatures

Vibrations in rotational drives originate from complex interactions between rotating parts and excitation of the structural eigenmodes. Under constant operating conditions these interactions demonstrate almost periodic repetitive pattern. Therefore, the most natural approach to fault detection is to look for the changes in the set of spectral components that are related to such repetitive events.

However, the presence of unknown and possibly variable operating conditions makes fault detection and isolation very difficult. Moreover, this situation is very realistic in many applications. In such a case the generated vibration signals are essentially non-stationary and the traditional methods designed for stationary processes are not applicable any longer.

One way to alleviate these difficulties is to observe the signal's instantaneous power (IP). The instantaneous power is directly related to the signal's envelope (Antoni, 2009; Ochiai, 2011), which is state-of-the art approach to fault detection of rotational machines under constant operating conditions. Therefore, by monitoring the instantaneous power of the generated non-stationary vibration signals we have access to the same information that is contained in signal's envelope.

In what follows we investigate the distribution of IP of the vibration signals generated by gearboxes operating under variable loads and speeds. Since a signal generated by running gears and damaged bearings can be regarded as the sum of sine signals with random phase and amplitude, we will limit our analysis only to non-stationary signals that belong to this class. In particular, the aim is to analyse the distribution of the IP in the presence of faults and under changes in operating conditions.

As the first step we will review the most commonly adopted models for vibrations generated by running gears and bearings. Based on these two models we will analyse how the presence of faults and variations in the operating conditions influence the statistical characteristics of IP.

The influence of faults and variable operating conditions on the distribution of signal's envelope

Variations in the operating conditions and faults directly affect signals in two profound ways (see models (5.1) and (5.2)):

1. changed amplitude modulations and

2. modified frequency signature due to the new spectral components that result from excitation of additional eigenmodes $s(t)$ in (5.2).

Under constant and known operating conditions these changes are easily detectable by performing spectral analysis of the underlying vibration signals. In the case of unknown and presumably variable operating conditions the generated vibration signals become non-stationary and can be expressed in the generalized form (2.21)

$$y(t) = \int_{-\infty}^{\infty} A(\omega, t) e^{j\omega t} dZ(\omega). \tag{7.1}$$

The spectral representation (7.1) defines an evolutionary non-stationary process for deterministic $A(\omega, t)$. However, as already noticed by Antoni and Randall (2006), in many cases $A(\omega, t)$ has stochastic nature which is either a result of random temporal variations in the process or is due to random observation intervals in respect to the process dynamics.

Our goal is to investigate the way the presence of mechanical faults and variable operating conditions influence the distribution of $A(\omega, t)$. We will assume that during a short measurement section the operating conditions will be nearly constant. Our main focus will be on tracing the changes in the distribution of $A(\omega, t)$ over measurement sections when either machine state or the operating conditions differ substantially.

Influence of fault on vibration signal

By observing the models (5.1) and (5.2) we can notice that the overall vibrations can be described as a sum of sine waves with random amplitudes and phases in the presence of additive noise. Any change in their statistical characteristics will influence the envelope of the observed signal. Therefore, we will derive the probability density function of the envelope of a signal modelled as a sum of sine waves.

The problem of determining the envelope's probability density function (PDF) has been addressed by many authors e.g. Abdi et al. (2000); Esposito and Wilson (1973); Helstrom (1999); Simon (1985). As noted by Abdi et al. (2000), the value of random sinusoidal signal at any time moment is characterized by its amplitude and phase. Their values at each time moment can be described as a vector sum of n random amplitudes and phases, whose time evolution will follow the original signal. The number n determines the number of random components present in the observed signal. Hence, a sum of sine components can be written in the following form Abdi et al. (2000):

$$z(t) = \sum_{k=1}^{n} r_k(t) e^{j\theta_k t}, \tag{7.2}$$

where r_k are arbitrarily dependent positive random variables, θ_k is uniformly distributed i.e. $\theta_k \sim \mathcal{U}(0, 2\pi)$ and r_k are independent of θ_k. The envelope s_n of (7.2) is $s_n = |\sum_{k=1}^{n} r_k(t) e^{j\theta_k t}|^2$. Hence the PDF of the envelope s_n for a sum of n components becomes

Abdi et al. (2000)

$$f(s_n) = 2\beta s_n \exp\left(-\beta s_n^2\right) \sum_{m=0}^{M_n} c_m \mathcal{L}_m\left(\beta s_n^2\right) \quad 0 < \beta < \infty, \tag{7.3}$$

where $M_n \to \infty$ represents the number of terms used for numeric calculations, β is positive constant, $\mathcal{L}_m(\cdot)$ is the Laguerre polynomial of order m and

$$c_m = E_{s_n}\left[\mathcal{L}_m\left(\beta s_n\right)\right], \tag{7.4}$$

where the expectation $E[\cdot]$ is calculated over random amplitudes s_n.

Numerical simulation The density function (7.3) depends only on the values of amplitudes s_n via (7.4) and the number n of components in the summation (7.2). We have performed numerical simulations for four different numbers of components n, keeping the highest order of Laguerre polynomial fixed at $M_n = 9$. The resulting density functions are shown in Figure 7.1. From these plots it is evident that the number of components has significant influence on the shape of the resulting envelope's PDF.

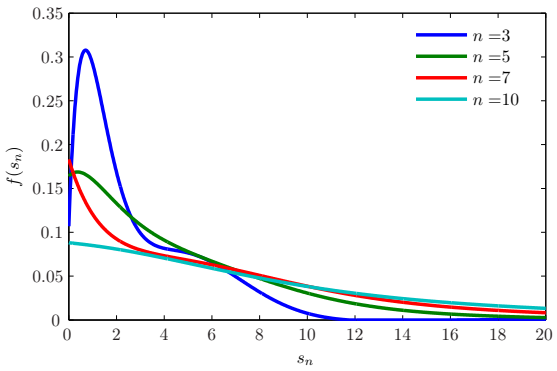


Figure 7.1: Probability density function of a sum of sinusoidal in dependance of the number of components calculated via (7.4) and Laguerre polynomials of order 9

A vibration signal has usually a broad spectrum. Splitting the original signal into a set of narrowband intervals allows envelope analysis of the sum of sine components belonging to the corresponding frequency band. In the presence of fault, new components will occur in several of these narrowband signals that will contribute to a significant change in the shape of envelope's PDF.

Influence of the variable operating condition

Do variations in operating conditions cause significant change in the envelope PDF? As already presented, the most influential factor that contributes to the distribution of envelope is the number of sine components in the observed signal. Changes in the operating

conditions can be modelled by the term $b(\omega, t)$ added to the complex envelope $A(\omega, t)$ from (7.1)

$$A_1(\omega, t) = A(\omega, t) + b(\omega, t).$$

Since we assumed that the operating conditions can be considered as almost constant for the short observation time, we can consider $b(\omega, t)$ to be independent of t. The presence of $b(\omega)$ will cause amplification or attenuation of some of signal's components regardless of whether the cause is load and/or speed variation.

Baydar and Ball (2000) showed that the impact of gear faults is much more visible than the influence of the load variations. In the case of fault-free bearings, load changes should have almost no effect since fault-free bearings produce negligible vibrations. Despite this influence, the difference in the envelope distribution of fault-free and faulty bearings dominates the influence caused by load variations. Regardless of the influence of load variations, the presence of faults introduces more significant change in the envelope PDF, and this change can be exploited for fault diagnosis.

Herewith, we can conclude that by analysing the probability distribution of the instantaneous power (signal's envelope) of narrowband filtered parts of the produced vibration signals one can infer about the machine condition without any knowledge of the operating conditions. Therefore, we employed wavelet packet analysis and tracked the probability distribution of the wavelet coefficients as features in the process of performing sufficiently accurate fault detection.

7.2 Information cost functions

According to the analysis conducted in Chapter 4.5, wavelet packet analysis is capable of analysing evolutionary non-stationary signals (2.21). The wavelet packet coefficients at each wavelet packet node can serve as indicators about the statistical characteristics of the signal's instantaneous power using (4.8). Thus one can estimate the probability density of the signal's envelope contained in that particular node. In order to quantify alterations in the PDF caused by fault presence, we have selected a set of entropy functions referred to as information cost functions (ICFs).

The first step is calculation of the signal's energy $E_{d,n}$ contained within one node (d, n) as (Blanco et al., 1998):

$$E_{d,n} = \sum_{k=0}^{N_d-1} \|P_d^n(k)\|^2 \tag{7.5}$$

For the set of terminal nodes T, the relative energy for each node $(d,n) \in T$ is:

$$p_{d,n}(k) = \frac{\|P_d^n(k)\|^2}{E_{d,n}},\ k = 0, \cdots, N_d - 1. \tag{7.6}$$

The total signal's energy can be obtained by summing the energy contained within the set of terminal nodes T:

$$E_{tot} = \sum_{\substack{k=0 \\ d,n \in T}}^{N_d-1} \|P_d^n(k)\|^2 \tag{7.7}$$

The ratios (7.5) and (7.7) specify the energy distribution over the set of nodes $(d,n) \in T$:

$$p_{d,n}(f) = \frac{E_{d,n}}{E_{tot}} \tag{7.8}$$

Since each terminal node is associated to a specific frequency interval, the relation (7.8) can be interpreted as the signal's energy distribution over the observed frequency range.

Based on the two distributions (7.6) and (7.8) ICFs borrowed from (Rosso et al., 2001; Zunino et al., 2007) read:

$$C_t(d,n) = -\sum_{t=0}^{N_d-1} p_{d,n}(t) \log_2 \big(p_{d,n}(t)\big),\ \text{for each } (d,n) \in T \tag{7.9}$$

and

$$C_f = -\sum_{d,n \in T} p_{d,n}(f) \log_2 \big(p_{d,n}(f)\big). \tag{7.10}$$

7.2.1 Rényi entropy

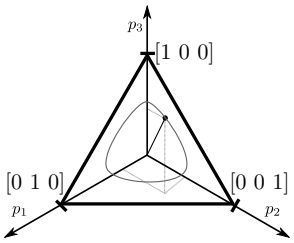
The Rényi entropy is a generalisation of the Shannon entropy used for the definition of the ICFs(7.9) and (7.10). The Rényi entropy for a discrete probability distribution p defined on a finite set $\mathcal{D}$ is given by Rényi (1960)

$$H_\alpha(P) = \frac{1}{1-\alpha} \ln \sum_{x \in \mathcal{D}} p^\alpha(x), \tag{7.11}$$

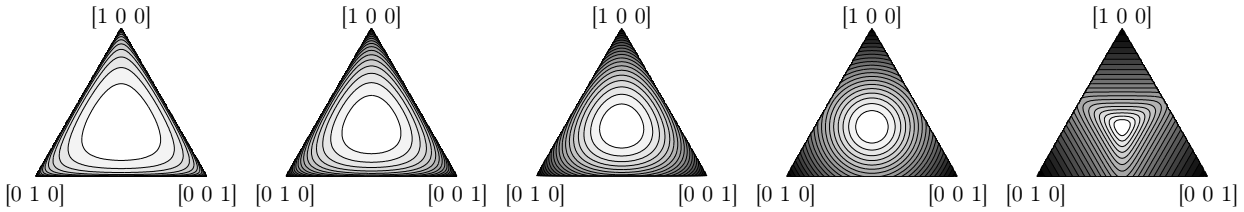
where the exponent $\alpha > 0$ and $\alpha \neq 1$. For $\alpha \to 1$ it reduces to Shannon entropy. The variations in the α parameter alters the sensitivity of the Rény entropy to particular changes in the shape of the analysed probability density function (Hero et al., 2002).

The α exponent in the Rény entropy specifies the relative importance of small values versus large values of the probability mass. This effect can be visualised using the isoentropy plots of Xu and Erdogmuns (2010) for all possible probability distributions over N bins. For the case where $N = 3$, we have $p_1 + p_2 + p_3 = 1$ which defines a plane in the first octant (see Figure 7.2 (a)). For the cases where $\alpha < 1$, as shown in Figure 7.2 (b)–(c),

the isoentropy lines are dense near the edges of the triangle, showing high sensitivity to the low probability components. Conversely as the parameter α increases the isoentropy lines spread throughout the triangle's surface, as shown in Figures 7.2 (d)–(f). The area around the centre of the triangle becomes smaller, hence the entropy sensitivity shifts towards highly probable events.



(a) Space covered by PDFs with three bins



(b) $\alpha = 0.2$ (c) $\alpha = 0.5$ (d) $\alpha = 1$ Shannon entropy (e) $\alpha = 2$ (f) $\alpha = 10$

Figure 7.2: Rény isoentropy plots for probability distributions with three components as in Xu and Erdogmuns (2010)

Finally, using the Rényi entropy (7.11) instead of Shannon entropy the original ICFs (7.9) and (7.10) can be rewritten as (Rosso et al., 2006)

$$\begin{aligned} C_t^{\alpha}(d,n) &= \frac{1}{1-\alpha} \ln \left(\sum_{t=0}^{N_d-1} p_{d,n}^{\alpha}(t) \right), \text{ for each } (d,n) \in T \\ C_f^{\alpha} &= \frac{1}{1-\alpha} \ln \left(\sum_{d,n \in T} p_{d,n}^{\alpha}(f) \right). \end{aligned} \tag{7.12}$$

7.2.2 Jensen-Rényi divergence

The most common approach to comparing probability densities is by means of divergence as dissimilarity measure (Cichocki and Amari, 2010). Two distinctive classes of such divergences exist (Basseville and Cardoso, 1995): f-divergences, which includes Kullback-Leibler divergence, and divergences built upon the concavity of entropy functionals which

are defined as

$$J_H^{\beta}(p,q) \triangleq H\left((1-\beta)p+\beta q\right)-(1-\beta)H(p)-\beta H(q),$$

where $0<\beta<1$ and $H(p)$ is entropy functional. This relation can be also generalised for a population of probability densities as (Basseville, 2010; Nielsen and Boltz, 2011):

$$J_H^{w}(p_1,\cdots,p_n)=H\left(\sum_{i=1}^{n}w_ip_i\right)-\sum_{i=1}^{n}w_iH(p_i), \tag{7.13}$$

where $w_i \geq 0$ and $\sum_{i=1}^{n} w_i = 1$. Jensen-Rényi divergence is defined using (7.13) by applying $H(p) = -H_\alpha(p)$, where $H_\alpha(p)$ is the Rényi entropy (7.11) (Ben Hamza et al., 2001). The application of Rény entropy introduces the α parameter that allows control of the sensitivity of this divergence.

The Jensen-Rény divergence (7.13) becomes monotonically related to a true distance metric when $\alpha \in (0, 2]$ and $\beta = 1/2$ (Briët and Harremoës, 2009). The value of α parameter in (7.13) influences the sensitivity towards specific changes in analysed PDFs (Hero et al., 2002):

- for $\alpha \to 0$ (7.13) becomes highly sensitive to changes in the tails of the distribution; and
- for $\alpha \to 1$ (7.13) acts similarly like Kullback-Leibler divergence and is sensitive to changes where the bulk of the probability mass is located.

The Jensen-Rényi divergence for the probability distribution of the wavelet coefficients (7.6) represents the third and final feature in this analysis. Together with the Rényi entropy based ICFs (7.12), they have the capability of focusing their sensitivity to a specific shape alterations of the analysed PDFs using the α parameter. The subsequent analysis will present the most applicable values for this parameter for the purpose of detecting gear and bearing faults.

7.2.3 Simulated vibrational signal

Two simulated signals were analysed with the aim to demonstrate:

- the capability of ICFs to respond to the changes in the shape of the envelope distribution, and
- to show robustness of the method to the changes in load and speed.

We used wavelet packet tree with depth 4, resulting into 16 terminal nodes. The mother wavelet was Daubechies 10 and sampling frequency is chosen $f_s = 5$ kHz.

Fault-free signal For the baseline we selected an amplitude modulated signal contaminated with Gaussian noise.

$$\begin{aligned} x(t) &= (1 + a(t)) \cos(2\pi f_c t) + n(t) \\ a(t) &= A \sin(2\pi f_a t). \end{aligned} \tag{7.14}$$

The carrier sinusoidal signal had random frequency $f_c \sim \mathcal{N}(300, 3)$ Hz. The modulating signal was taken as realisation of a random process with amplitude $A_t \sim \mathcal{N}(10, 0.2)$, frequency $f_a \sim \mathcal{N}(45, 3)$ and $n(t) \sim \mathcal{N}(0, \sigma^2)$. The value of σ was chosen so that the overall signal-to-noise ratio was SNR=1. The prominent spectral components of the simulated signal (7.14), the carrier frequency f_c and the left sidebands, reside in the frequency band spanned by the 2nd wavelet node. The histogram and the estimated PDF of the instantaneous power from the 2nd node is shown in Figure 7.3a. The shape of the PDF is similar to the one calculated for a signal containing three random sine waves as shown in Figure 7.1.

Simulated signal in the presence of fault and variable operating conditions The simulations of fault presence were done by adding new components to the base-line signal (7.14) that characterise gear and bearing faults. Gear faults were simulated, in accordance to (5.1), by introducing new modulation component with frequency $f_{mod} \sim \mathcal{N}(40, 3)$ as follows

$$\begin{aligned} x(t) &= (1 + a(t) + a_{new}(t)) \cos(2\pi f_c t) + n(t) \\ a_{new}(t) &= 0.2 \sin(2\pi f_{mod} t), \end{aligned} \tag{7.15}$$

The presence of bearing surface faults, was simulated according to (5.2), i.e. by adding randomly occurring impulse responses $s(t)$ defined as:

$$\begin{aligned} f(t) &= \sum_{i=-\infty}^{+\infty} A_i s(t - \tau_i) \\ s(t) &= e^{-10^{-3} t} \sin(2\pi 1500 t) \end{aligned} \tag{7.16}$$

where $\tau_i \sim IG(1/15, 2\pi/0.2^2)$[1] following (Antoni and Randall, 2003; Boškoski and Juričić, 2011) and $A_i \sim N(1, 0.2)$.

The PDF of the instantaneous power of the signal (7.15) with simulated gear faults is shown in Figure 7.3b. By comparing it with the baseline signal, shown in Figure 7.3a, one can notice a significant change in the tail of the PDF. The occurrence of an additional sine wave component causes similar alterations in the shape of the PDF as once already shown in Figure 7.1. The influence of load changes, on the other hand, has little influence on

[1] IG stands for Inverse Gaussian distribution

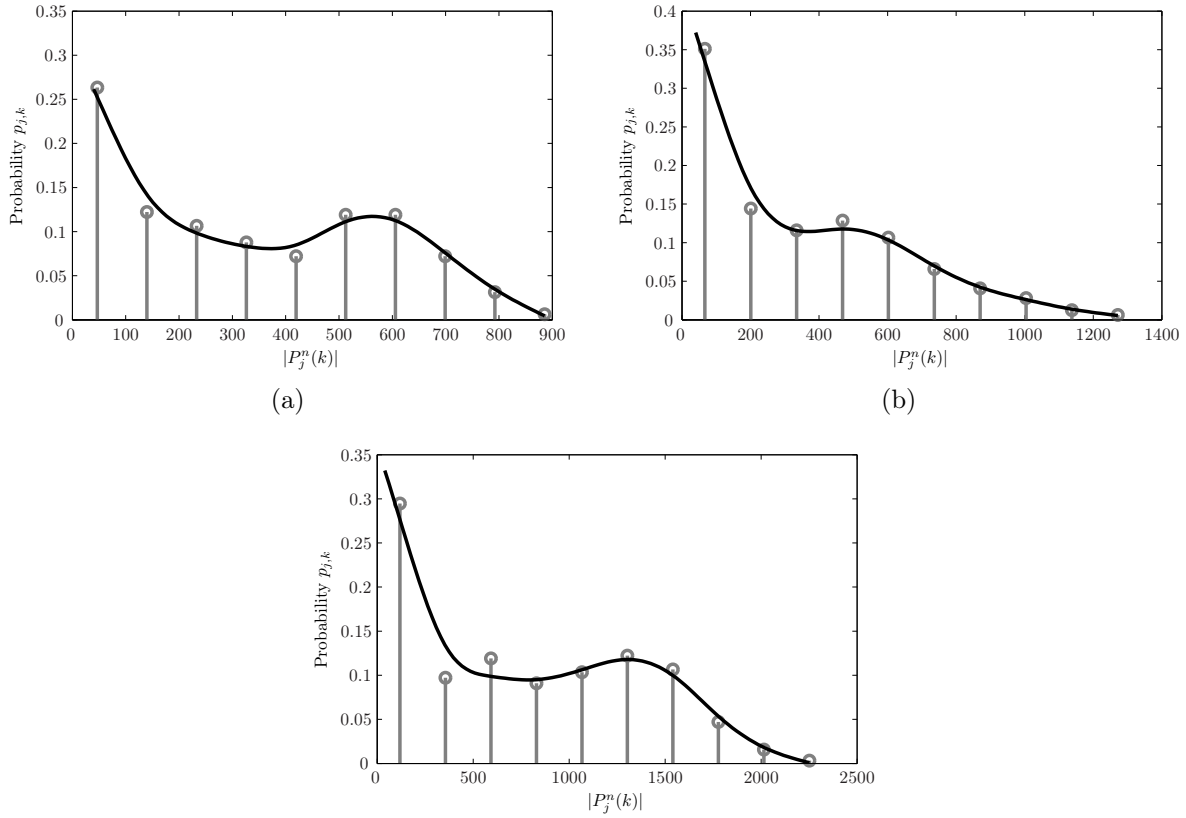
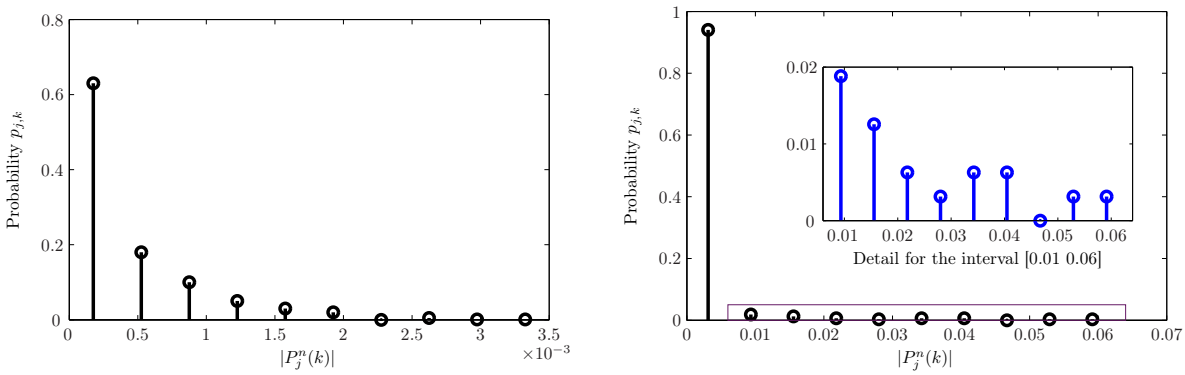


Figure 7.3: Probability densities of the wavelet packet coefficients for the simulated signals: a) "fault-free" signal (7.14); b) presence of an additional modulation frequency (7.15); c) signal (7.14) with different load $\tilde{A} = A + b$. Note the similarity with the analytical simulations in Figure 7.1

the shape of the envelope distribution. As shown in Figure 7.3c, the PDF of the envelope under higher load is quite similar to the baseline signal, shown in Figure 7.3a. The only difference is the interval covered by the wavelet coefficients $|P_j^n(k)|$. Since the values of ICFs (7.12) are estimated using the histograms of the corresponding PDFs, their values remain almost the same regardless of the applied load.

The occurrence of bearing faults (7.16), on the other hand, introduces a new random component in the frequency band spanned by the 11^{th} node. This node spans the frequency interval that contains the simulated impulse response $s(t)$ from (7.16). The detection of their presence is fairly simple since the random excitations of the impulse responses $s(t)$ cause the overall signal power to be concentrated in a rather small number of wavelet coefficients with significantly high value, as shown in Figure 7.4. In particular, for the case when the impulses (7.16) are present more than 30% of the energy in the 11^{th} node is contained within 6% of wavelet coefficients. This is visible from the interval

spanned by the tail of the distribution as shown in Figure 7.4 (b). On the other hand, the distribution of the wavelet coefficients, from the same node for the case without impulses (7.16), covers significantly narrower interval as shown in Figure 7.4 (a). This shows that the signals energy is almost evenly spread in time, and each wavelet coefficient represents almost the same percentage of the overall signal energy.



(a) (b)

Figure 7.4: Probability densities of the wavelet packet coefficients for the 11^{th} node: a) the "fault-free" signal; b) impulses present in the signal (7.16)

Simulations above illustrate some of the problems regarding the detection of changes in real signals. Fault, either gear or bearing, alters the distribution of the wavelet coefficients. These changes affect the tails rather than the "stomach" of the PDFs. Consequently, feature set should be able to distinguish these small variations. As a result of this influence, the value exponent α of the two Rényi entropy ICF (7.12) was chosen to be in the interval $0 < \alpha < 1$.

7.3 Case studies

The effectiveness of ICFs for fault detection was examined under assumption that no information about rotational speed and load is given. For that purpose we used a two-stage gearbox operating under different speeds and loads with various combinations of embedded gear and bearing faults.

In order to keep the presentation consistent, we will present the results that contain the values of C_t^α calculated only from a subset of terminal nodes. Only nodes that exhibit the largest changes in the corresponding ICF are selected. Normally, the values of ICFs from every terminal node should be taken into account.

	Gear				Bearing[1]			Shaft
	1	2	3	4	1	2	3	
#1	Fault Free (FF)							
#2	Chipped	FF	Eccentric	FF	Fault Free (FF)			FF
#3	FF	FF	Eccentric	FF	Fault Free (FF)			FF
#4	FF	FF	Eccentric	Broken	Inner	Ball	FF	FF
#5	Chipped	FF	Eccentric	Broken	Inner	Ball	Outer	FF
#6	FF	FF	FF	Broken	Inner	Ball	Outer	Imbalance
#7	FF	FF	FF	FF	Inner	FF	FF	Keyway Sheared
#8	FF	FF	FF	FF	FF	Ball	Outer	Imbalance

[1] Faults were introduced only on Bearings 1–3 (see Figure 7.5). The other three bearings were kept fault-free during all the experimental runs.

Table 7.1: Faults in each experimental run. Same set was used for both load levels

Experimental results

The experimental data are taken from PHM benchmark (PHM, 2009), which consists of a laboratory two–stage gearbox (see Figure 7.5). The test runs include 7 different fault combinations and one fault–free reference run. Each set of faults was tested under 5 different rotational speeds of the input shaft: 30, 35, 40, 45 and 50 Hz. Additionally, two different load levels were applied. The detailed list of the introduced faults is provided in Table 7.1. The signals were sampled with sampling frequency $f_s = 66.6$ kHz and the sampling horizon was 4 seconds long.

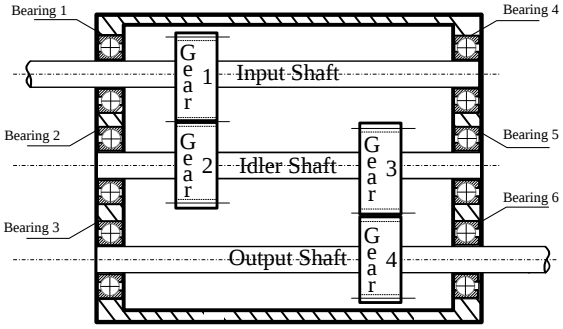
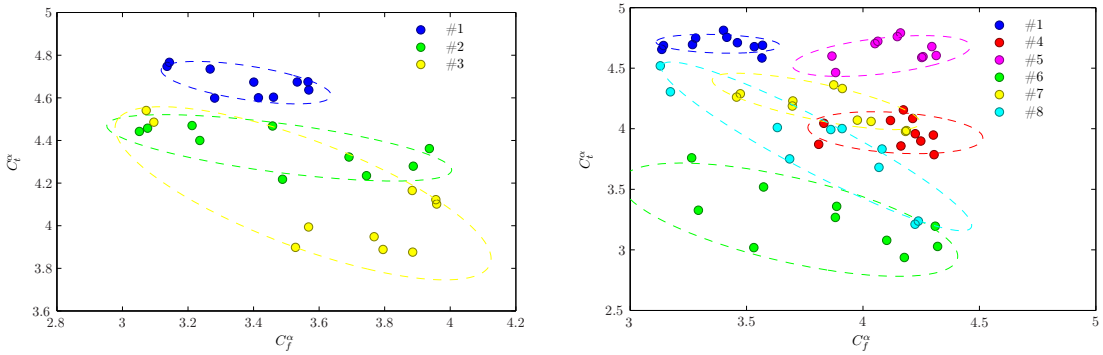


Figure 7.5: Schematic description of the experimental two–stage gearbox (PHM, 2009)

Both load cases were analysed using WPT with *db10* mother wavelet. Each experimental run was analysed using 4-level WPT tree, which produced 16 terminal nodes.



(a) Experimental runs with gear faults only

(b) Experimental runs with mixed gear and bearing faults

Figure 7.6: Values of the ICFs for the experimental runs conducted under *high load*.

For the purpose of clearer presentation, we will first present the results from experiments performed under high load followed by the experiments performed under low load.

High Load Values of both ICFs (7.12) for experiments performed under high load are shown in Figure 7.6. The values of ICFs from experiments with the same faults tend to group into clusters regardless of the changes in the rotational speed. The cluster from the reference fault-free run (#1) is located in the upper left corner of the plots. With respect to the fault-free cluster, clusters from other faults have two distinct characteristics: higher value for C_f^α (abscissa) and lower value for C_t^α (ordinate).

The increase in the value of C_f^α is an indicator that vibration spectrum is spread over a broader range of frequencies compared with the fault-free run. Changes in C_t^α indicate changes in the distribution of signal's envelope caused by alterations in the modulation components.

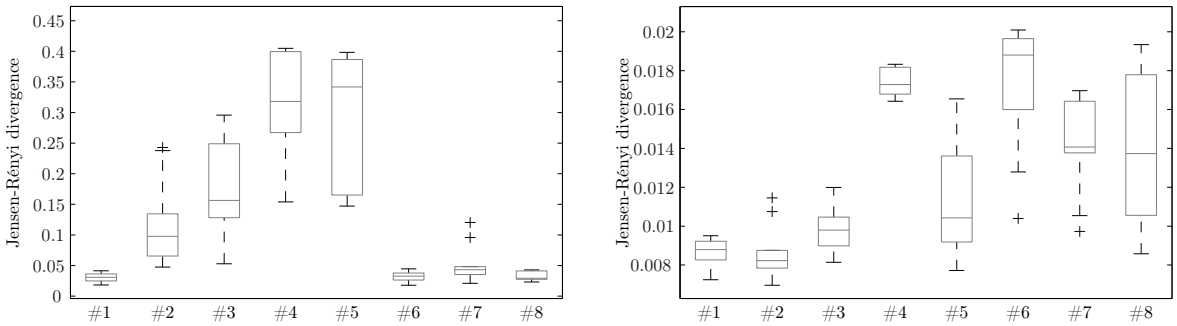
The results of the experimental runs #2 and #3 are shown in Figure 7.6a. As these runs contain only gear faults, it is noticeable that these runs tend to have lower values for C_t^α (ordinate) than the corresponding values of the fault-free runs. On the other hand, as gear faults do not excite additional eigenfrequencies, the produced vibrations retain the same energy distribution over the observed frequency range. Consequently, the values of C_f^α remain roughly in the same interval as in the fault-free run #1.

Unlike the experiments containing only gear faults, the results of experiments with mixed gear and bearing faults exhibit significant changes both in C_t^α and C_f^α (see Figure 7.6b). The presence of bearing faults introduces new high-frequency components due to the excited eigenfrequencies according to (5.2), hence increasing the bandwidth of the produced vibrations. Similarly to the cases of pure gear faults, clusters related to gear faults exhibit lower values for C_t^α than the fault-free one.

The application of Jensen-Rényi divergence (7.13) to the same set of data provides additional information regarding the fault detection process. From all 16 nodes in the WPT tree Figure 7.7 shows the Jensen-Rényi divergence for the 1st and the 13th node since these nodes exhibit the highest sensitivity to gear and bearing faults for these experiments.

The acquired 8 seconds long vibration signals were split into 0.25 seconds long non-overlapping segments, and the distributions of the wavelet packet coefficients $p_{d,n}(t)$ were calculated for each segment. The Jensen-Rényi divergence was calculated among these distributions using the fault-free case as a reference point.

The Jensen-Rényi divergence among different distributions for the fault-free cases #1 is negligible, regardless under which operating conditions the distributions of $p_{d,n}(t)$ were calculated. The analysis of these results is fairly straightforward.



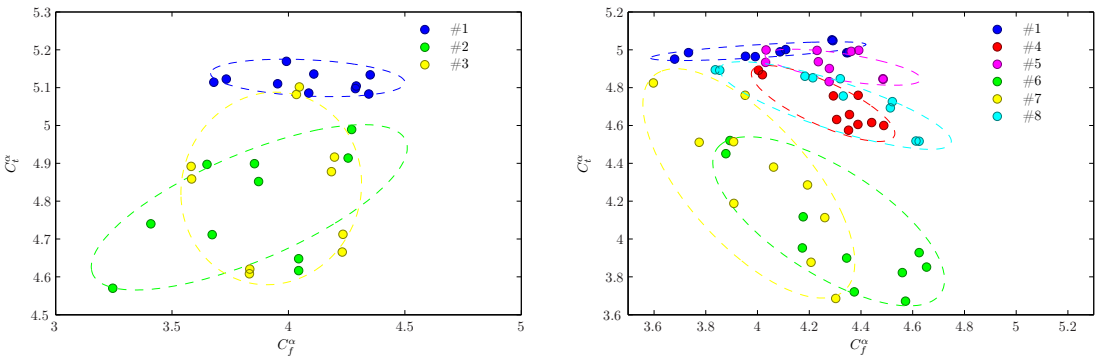
(a) the 1st node (b) the 13th node

Figure 7.7: Jensen-Rényi divergence between the datasets taken from the fault-free machine and datasets obtained under various faults explained in Table 7.1. All experiments were performed under high load.

The distances calculated for the 1st WPT node, shown in Figure 7.7 (a) indicate that cases #2–#5 exhibit the biggest distance from the fault-free case #1. As the 1st node contains the low-frequency part of the signal, these changes can be attributed to gear faults. The cases with bearing faults #6–#8 have similar values as the fault-free case, as their influence is more expressed in the high-frequency band. This effect is visible in the 13th node, shown in Figure 7.7 (b). For this node the cases containing only gear faults #2–#3 show no changes, whereas all the other cases #4–#8 exhibit significant departure from the fault-free signal. It should be noticed that runs #4 and #5 show increases in both WPT nodes, since they contain both gear and bearing faults.

Low Load The distribution of the ICFs for low load is shown in Figure 7.8. Obviously, the fault isolation capability of the proposed pair of entropy indices drops down. Similarly to the case with high load, clusters associated with the fault-free runs tend to locate in

the upper left corner of the plots, due to high values for C_t^α and low values for C_f^α.



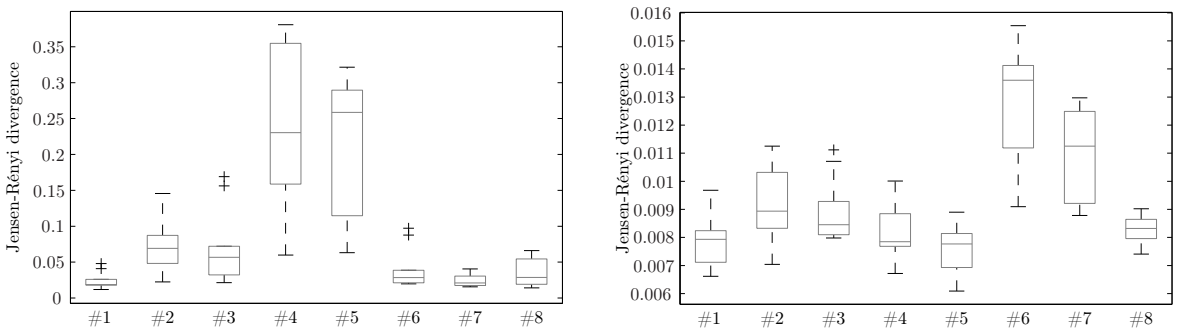
(a) Experimental runs with gear faults only

(b) Experimental runs with mixed gear and bearing faults

Figure 7.8: Values of the ICFs for the experimental runs conducted under *low load*.

Under low load, fault clusters are compressed in a narrower C_t^α interval compared to the experiments performed under high load. This effect is most visible in the runs containing gear faults. Unlike the changes in the values of C_t^α, variations in the load have little effect on the values of C_f^α. By comparing clusters from the corresponding experimental runs performed under high and low load one can notice that the corresponding clusters share similar spread of C_f^α values. This behaviour is well in accordance with the conceptual model of bearing faults (5.2).

The values of Jensen-Rényi divergence calculated for wavelet coefficients $p_{d,n}(t)$ in respect to the fault-free run for the low load cases are shown in Figure 7.9. Similarly to the high-load case the gear faults are mostly visible in the low-frequency node, shown in Figure 7.9 (a). The load change has no influence on the fault-free run #1. The distance of the runs #2 and #3 from #1 is somewhat smaller than the corresponding one from the high-load case. On the other hand, the runs #4 and #5 remain sufficiently distinguishable. The same observation is valid for the runs containing bearing faults, shown in Figure 7.9 (b). The runs #6 and #7 clearly show departure from the fault-free run. However the load change smears the detection for the run #8 based on Jensen-Rényi divergence. Regardless of this effect the values of C_f^α and C_t^α suffice for the detection of this fault.



(a) the 1st node (b) the 13th node

Figure 7.9: Jensen-Rényi divergence between the datasets taken from the fault-free machine and datasets obtained under various faults explained in Table 7.1. All experiments were performed under low load.

7.4 Point processes and their application to the detection of bearing faults

According to the bearing fault model (5.2), the presence of bearing faults can be detected by analysing vibration envelope calculated in the frequency band that contains the excited eigenmode with impulse response $s(t)$. The same signal can be analysed through the distribution of time intervals between two successive impacts $\Delta T_i = \tau_i - \tau_{i-1}$. Additional information is contained in the distribution $p(N; T_w)$ of the number of impacts N that occur within a fixed window of observation T_w. Both parameters are equivalent in defining a temporal point process whose statistical characteristics are closely related to the underlying bearing fault.

Impacts between rolling elements and the corresponding damaged site on the surface occur with frequencies characteristic for a particular fault. Hence, the necessary information for fault detection is contained in the statistical properties of the time intervals ΔT_i i.e. interevent times between impacts. The time of occurrence of the next impact τ_i depends only on the time from the last impact τ_{i-1} and the rotational speed ω_{shaft}. As firstly noticed by Antoni and Randall (2003), the difference among the interevent times can be modelled as a Brownian motion. Based on this observation, one can specify the actual distribution of the interevent times and employ this information in the process of fault detection.

The relation with the Brownian motion becomes apparent when observing the angle that a rolling element has to traverse in order to reach the damaged cite. As bearing cage keeps the rolling element equidistant, the traversed angle between two consecutive impacts is constant. Hence impact will occur as soon as a rolling element traverses a

specific arc distance. The angle can be modelled as (3.6)

$$d\theta(t) = \nu dt + \sigma dW(t), \tag{7.17}$$

where $W(t)$ is standard Brownian motion. So whenever the angle $\theta(t)$ reaches the corresponding angular value a an impact occurs. According to (3.7), the distribution of the corresponding interevent times ΔT_i between two consecutive impacts is governed by the Inverse Gaussian distribution $IG(a/\nu, a^2/\sigma^2)$.

Having in hand the statistical properties of the governing renewal process we can now analyse the way the model performs under different specific operating conditions. The goal of this analysis is to show that the model works for constant as well as for variable operating speed and in cases of single as well as of multiple bearing faults.

7.4.1 Single bearing fault

Regardless of the location of a bearing damage the time of two consecutive impacts is governed by (7.17). A single bearing fault $k \in \mathcal{K} = \{1$ (Inner ring), 2 (Outer Ring), 3 (Bearing Cage), 4 (Ball spin)$\}$ is characterised by a specific arc distance a_k that has to be covered by the rolling element and the positive drift ν_k. However, the parameters ν_k and σ_k are directly related to the speed of the rotating shaft and can be rewritten as

$$\begin{aligned} \nu_k &= C_k \nu_{shaft} \\ \sigma_k &= C_k \sigma_{shaft}, \end{aligned} \tag{7.18}$$

where the multiplication constant C_k is determined by the geometrical characteristics of the bearing which determine the ratio between the angular speed of the rotating ring and a specific bearing element. Consequently, the distribution of the interevent times for the point process associated with the damage on the kth component according to (3.7) with the limitations (7.18) becomes:

$$t_k \sim IG\left(\frac{a}{C_k \nu_{shaft}}, \frac{a^2}{C_k^2 \sigma_{shaft}^2}\right) \tag{7.19}$$

7.4.2 Multiple faults on different bearing components

In cases of multiple bearing faults we look at the overall produced vibrations as a sum of several point processes each associated with a specific bearing fault and governed by its own IG probability distribution defined by (7.19). In the general case the sum of IG r.v. does not necessarily results in a r.v. governed by IG distribution. Tweedie (1957) showed through the analysis of the characteristic function that the sum of random variables

$$x = \sum_i x_i \quad \text{where } x_i \sim IG(a/\nu_i, a^2/\sigma_i^2),$$

will have IG distribution only when

$$\nu_i = w_i \nu_0 \quad \text{and} \quad \sigma_i^2 = w_i^2 \sigma_0^2, \tag{7.20}$$

where ν_0 and σ_0^2 are positive constants defining the process (7.17). Under conditions (7.20), the ratio between the mean and the variance is constant i.e.

$$\frac{Var\{x_i\}}{E\{x_i\}} = \frac{\sigma_0^2}{\nu_0^2}.$$

The sum of r.v.'s governed by the distribution (7.19) fulfils the necessary condition (7.20) and the ratio between the mean and the variance

$$\frac{Var\{t_k\}}{E\{t_k\}} = \frac{a\sigma_{shaft}^2}{C_k \nu_{shaft}^3} \frac{C_k \nu_{shaft}}{a} = \frac{\sigma_{shaft}^2}{\nu_{shaft}^2}$$

remains constant, i.e. independent of C_k. Thus the sum of such renewal processes results into a new renewal process with IG interevent distribution:

$$\begin{aligned} S &= \sum_k t_k \\ &\sim IG\left(\frac{a}{\nu_{shaft}} \sum_{k \in \mathcal{K}} C_k, \frac{a^2}{\sigma_{shaft}^2} \left(\sum_{k \in \mathcal{K}} C_k\right)^2\right) \end{aligned} \tag{7.21}$$

Expression (7.21) shows that the case of multiple faults can be also treated as a point process with IG interevent distribution. Furthermore, each combination of faults will be governed by IG with different parameters since the sums of C_{k_f} in (7.21) will differ according to the set of present faults.

7.4.3 Distribution of the number of impacts

Within a fixed observed window of length T_w one can use the distribution of the number of impacts N as information that is closely related with the underlying fault. In order to detect impacts we will use wavelet transform of the vibration signal. As already shown in Section 7.2.3, the presence of impacts in the vibration signal causes the bulk of signal's energy to be concentrated in a small number of wavelet coefficients. Thus the entropy of the wavelet coefficient will decrease. By characterizing the probability $p(N; T_w)$ of N impacts within a time window with length T_w, one can correlate the changes in the entropy of the wavelet coefficients with a particular bearing fault.

The probability $p(N; T_w)$ can be determined by the survival probability $s_N(t)$ (3.1) as in (van Vreeswijk, 2010). The survival probability $s_N(t)$ gives the probability of observing N impacts after a time moment t:

$$s_N(t) = \int_t^{+\infty} f_N(t')dt', \tag{7.22}$$

where $f_N(t)$ is the PDF of observing N events in the time interval $[0, t)$, as defined by (3.8). Based on (7.22), the probability of observing N impacts within a time window of length T becomes

$$p(N, T) = s_{N+1}(T) - s_N(T). \tag{7.23}$$

The Laplace transform of (7.22) combined with (3.9) becomes

$$S_{L,N}(s) = \frac{1 - f_{L,N}(s)}{s} = \frac{1 - f_{L,1}^N(s)}{s}, \tag{7.24}$$

where $f_{L,k}(s) = \mathcal{L}\{f_k(t)\}$, similarly like in (3.9). Calculating the Laplace transform of (7.23) and inserting (7.24) results into:

$$p(N, s) = \frac{1 - f_{L,1}(s)}{s} f_{L,1}^N(s).$$

In order to simplify the analysis we will concentrate only on mean and variance of the distribution $p(N, T_w)$. These values can be approximated by taking into account only a limited number of Taylor expansion terms. Hence for the mean $E\{N|T_w\}$ and variance $Var\{N|T_w\}$ when $p(t) \sim IG(a/\nu, a^2/\sigma^2)$ read

$$\begin{aligned} E\{N|T_w\} &= \nu T_w + \frac{\sigma^2\nu - 1}{2} \\ Var\{N|T_w\} &= \sigma^2\nu^2 T_w \end{aligned} \tag{7.25}$$

In case of bearing vibrations, as already shown by (7.21), each bearing fault differ by the factor C_k multiplying the shaft rotational speed and its fluctuation. As a result of this dependence, each bearing fault is governed by different interevent distribution $p(t)$, thus the number of expected impulses within a fixed time window of size T_w will differ among different fault combinations.

Apart from the changes caused by different faults, the distribution $p(N, T_w)$ will change with rotational speed. However, according to (7.21) the variations in the rotational speed will influence every bearing fault in the same manner, i.e. by adding and additional constant to each coefficient C_k in (7.21). Consequently, notwithstanding the variations in the speed, the entropy values for particular bearing fault will always be distinguishable, since the underlying IG distributions will remain different among various bearing faults.

7.5 Summary of the results for bearing faults

We will summarise the experimental results related to the runs that contained bearing faults from the experiment with two-stage gearbox, presented in Section 7.3. The values of C_t for the corresponding experimental runs are shown in Figure 7.10. The results are

grouped in "bins" of experiments conducted with same speed. Each bin contains two measurements, one performed under high load and the other under low load.

The most evident feature is the constant value of C_t^α for the fault-free experimental runs, regardless of the operating conditions. Conversely, the fault #7 shows the highest value, followed by the fault #8 and #6 having lower value. By examining the fault details from Table 7.1, one can notice that fault #7 contains only a single damaged element, fault #8 two damaged elements and fault #6 contains three damaged elements. The occurrence of multiple faults can be treated as the sum of several r.v.'s governed by IG distribution. Thus, according to (7.21) the resulting random process will have higher firing rate. A higher firing rate in essence contributes to increased number of expected impact occurrences N within a time window T_w, according to (7.25). This alters the shape of the wavelet energy distribution and due to which its overall entropy decreases. So different bearing faults tend to have sufficiently different values for C_t^α, which explains the effect of clustering of the results as shown in Figures 7.6 and 7.8.

Although different faults tend to cluster, the clusters for bearing faults tend to be inclined (having negative gradient) for C_t^α values. The explanation for this effect is similar to the explanation of the influence of multiple faults. With the increase of the rotational speed one will observe an increase in the firing rate of the IG process. As already stated this increase leads towards lower wavelet energy entropy. However, as such an increase influences all the bearing faults in the same manner, the resulting effect is an increased difference in the entropies of particular faults. This effect can be explained if we analyse the joint effect of the rotational speed and a combination of bearing faults on the statistical properties of the resulting IG renewal process. According to (7.19) each bearing fault generates a renewal process $\sim IG(a/C_k\nu_{shaft}, a/C_k^2\sigma_{shaft}^2)$. A change in the speed can be easily modelled as a new multiplying factor to ν_{shaft} i.e. through altering the values of C_k. Although the C_t values of a particular fault will decrease with the increase of the rotational speed, in the same time the difference between the entropies of two particular faults will increase. Consequently, the faults will become more distinguishable as the rotational speed increases.

Comments on results

The experimental results support the hypothesis that bearing faults can be detected through the concepts of random point processes. One of the main advantage of the approach is that it requires no information about the operating conditions. Additionally, the analysis of bearing faults provides a possible explanation of the results presented in Chapter 7.3. Those results showed that, based solely on the entropy of wavelet packet coefficients, one can perform accurate fault detection of gears and bearings regardless of

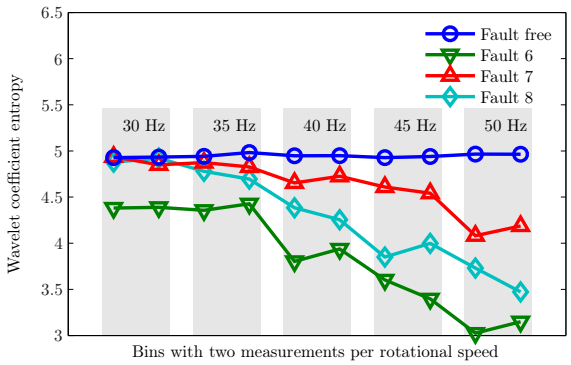


Figure 7.10: C_t^{α} values for selected bearing faults

the operating conditions. The relations (7.21) and (7.25) explain how different bearing faults change the probability distribution of the wavelet coefficients hence modifying its entropy.

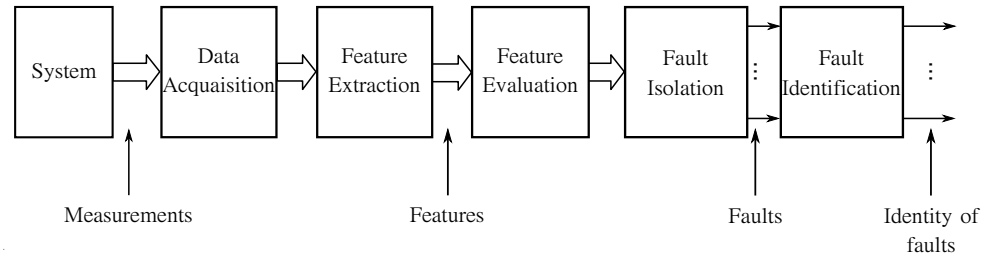
8 An extended fault isolation scheme based on evidential reasoning

The majority of the research in the field of fault diagnosis is focused on problems regarding the feature extraction task. The advances in the sensing and signal processing techniques alleviate some of the problems, hence broadening the applications of fault detection and indetification (FDI) systems. In the same time, these improvements result into significant increase of the information contained in the extracted feature set, thus causing an information overflow. Therefore, a variety of techniques has been applied for fusing these information in order to isolate the present fault and assess the potential impact on the overall system condition. This chapter presents one implementation of fault detection and identification system based on evidential reasoning approach. The system was tested as an end-quality assessment system on a production line of EC motors.

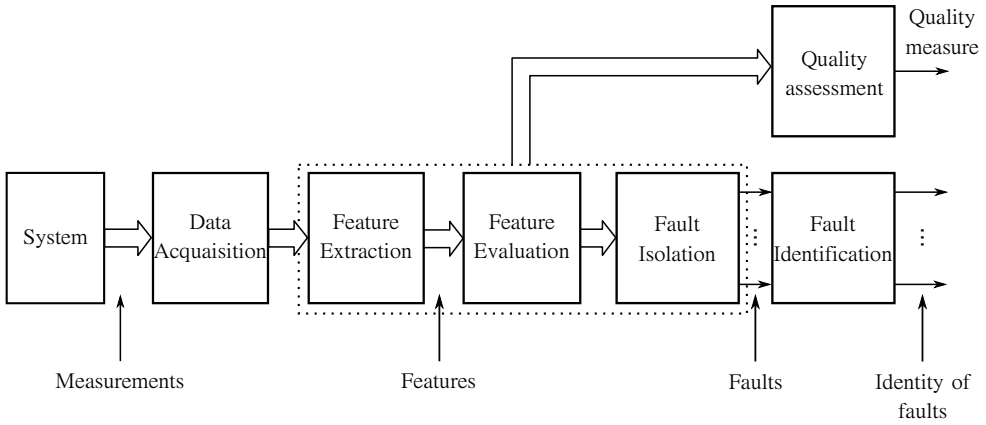
8.1 An extended FDI system (EFDI) architecture

The majority of FDI systems employ the standard structure that consists of four modules (Rakar et al., 1999): feature extraction, feature evaluation, fault isolation and fault identification (see Figure 8.1a). For the purpose of end-quality assessment systems the classical concept of FDI system is usually extended by adding a new functionality, referred to as the overall assessment. This chapter presents such an expansion module that performs overall quality assessment, as shown in Figure 8.1b.

The proposed end-quality assessment system employs evidential reasoning (ER) based on Dempster-Shafer theory (Yang and Singh, 1994). The application of ER approach is justified due to its efficient evidence combination rule and the application of utility function. The evidence combination rule allows the problem of overall quality assessment to be divided into smaller tasks of assessing the quality of the specific system's components. Then, the final decision is reached by aggregating the quality estimates of each component. As the end of the process, the utility value is calculated based on this final decision. This value can be regarded as an assessment of the overall quality of the system. A prototype version of the system was validated on a test batch of 130 electronically commutated motors, demonstrating high diagnostic resolution and accuracy.



(a) Standard fault diagnosis process (Rakar et al., 1999)



(b) Expanded fault diagnosis process by adding the Quality estimation module

Figure 8.1: Classical fault detection and isolation and EFDI system

8.2 Quality assessment by means of evidential reasoning

In any decision making context, some particular choices or actions may be regarded as more preferable then others. For the selection of the most appropriate choice or action one has to take into account all possible contributing attributes (Belton and Steward, 2002). Cases where the available *alternatives* are characterized by multiple attributes are referred to as the multiple attribute decision making (MADM) problems. The problem of quality assessment of EC motors can be regarded as MADM problem, since the overall motor quality is estimated from the values acquired by a specific set of attributes in the feature set. One common approach to MADM problems is through application of ER methods based on Dempster-Shafer (D-S) theory (Yang and Singh, 1994).

Method overview

The ER method distinguishes two types of attributes: basic and general attributes. Basic attributes are the ones that can be directly assessed, usually by means of direct measure-

ment. In the case of EC motor, basic attributes are represented by the extracted feature set (see Figure 8.3). General attributes, on the other hand, represent more abstract concepts and are usually assessed by aggregating the set of basic attributes. Relations between the basic and general attributes are usually represented by a hierarchical structure similar to the one used for EC motor, shown in Figure 8.4.

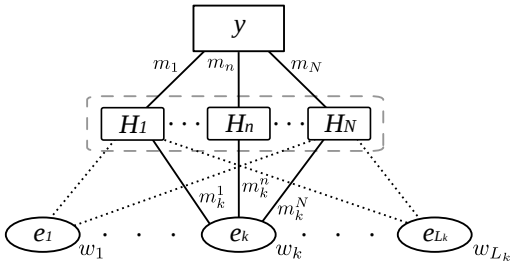


Figure 8.2: Tree depicting the connections between the particular motor components

The aggregation process can be illustrated on a simple two-level hierarchical structure with one general attribute y and a set E of basic attributes, shown in Figure 8.2. The set E is defined as follows:

$$E = \{e_1, e_2, \cdots, e_k, \cdots, e_{L_k}\},$$

where L_k is the cardinal number of the basic attribute set. Based on the information contained in E, the general attribute is assessed on a set of distinct evaluation grades H, which usually contains values like "*excellent*" or "*poor*", i.e.

$$H = \{H_1, H_2, \cdots, H_n, \cdots, H_N\}, \tag{8.1}$$

where H_1 and H_N represent the worst and the best evaluation grade respectively, and consequently H_{n+1} is preferred to H_n. This assessment of the general attribute y is based on the degree of belief $\beta_{n,k}$ that the general attribute y may be assessed to the evaluation grade H_n based solely on the information contained by the single basic attribute e_k. The degree of belief $\beta_{n,k}$ has the following two properties:

$$\beta_{n,k} \geq 0 \quad \text{and} \quad \sum_{n=1}^{N} \beta_{n,k} \leq 1. \tag{8.2}$$

In addition to the degree of belief, weights factors w_k are introduced in order to incorporate information about the relative importance of each basic attribute e_k. These weight factors have the following properties (Yoon and Hwang, 1995):

$$0 \leq w_k \leq 1, \quad \sum_{i=1}^{n} w_i = 1.$$

Consequently, using the degree of belief $\beta_{n,k}$ and the weight factor w_k, we can define the basic probability mass, m_k^n, that shows to what extent the attribute e_k supports the hypothesis that the general attribute y can be assessed to the evaluation grade H_n. The basic probability mass m_k^n is defined as (Yang and Sen, 1994):

$$m_k^n = w_k \beta_{n,k} \tag{8.3}$$

Since the sum in (8.2) might not add up to one, we can consider the remaining unassigned basic probability mass as the system's uncertainty. Such a remaining basic probability mass of the set of evaluation grades (8.1) is marked by m_k^H and it is defined as

$$m_k^H = 1 - \sum_{n=1}^{N} m_k^n = 1 - \sum_{n=1}^{N} w_k \beta_{n,k}, \tag{8.4}$$

where H in the superscript of m_k^H marks that this belief is assigned to the whole set of evaluation grades as opposed to the m_k^n which is assigned to a specific evaluation grade H_n.

As we have defined the basic probability mass for each attribute e_k the basic probability mass for the general attribute y can be obtained by applying the D-S evidence combination rule using the set of evaluation grades (8.1). The recursive implementation is applied (Yang and Xu, 2002):

$$\begin{aligned}
&m_{n,I(i+1)} = K_{I(i+1)} \left[m_{n,I(i)} m_{n,i+1} + m_{H,I(i)} m_{n,i+1} + m_{n,I(i)} m_{H,i+1} \right] \\
&m_{H,I(i)} = \tilde{m}_{H,I(i)} + \bar{m}_{H,I(i)} \\
&\tilde{m}_{H,I(i+1)} = K_{I(i+1)} \left[\tilde{m}_{H,I(i)} \tilde{m}_{H,i+1} + \bar{m}_{H,I(i)} \tilde{m}_{H,i+1} + \tilde{m}_{H,I(i)} \bar{m}_{H,i+1} \right] \\
&\bar{m}_{H,I(i+1)} = K_{I(i+1)} \left[\bar{m}_{H,I(i)} \bar{m}_{H,i+1} \right] \\
&K_{I(i+1)} = \left[1 - \sum_{t=1}^{N} \sum_{\substack{j=1 \\ j \neq t}}^{N} m_{t,I(i)} m_{j,i+1} \right]^{-1} \\
&i = 1, \cdots, L_k - 1, \quad m_{n,I(1)} = m_k^n, \quad m_{H,I(1)} = m_k^H,
\end{aligned} \tag{8.5}$$

where $I(i)$ defines the subset of the basic attributes containing only the first i elements, i.e. $E_{I(i)} = \{e_1, e_2, \cdots, e_i\}$, where $i \leq L_k$, and $m_{n,I(i)}$ represents the basic probability mass supporting the hypothesis that the general attribute y can be assigned to the evaluation grade H_n by observing only the evidence contained in basic attribute subset $E_{I(i)}$. Similarly to (8.4), the remaining unassigned probability mass is marked as $m_{H,I(i)}$. Finally, after iterating through all L_k basic attributes, the degrees of belief of the general

attribute y are calculated as:

$$\begin{aligned}\beta_n &= \frac{m_{n,I(L_k)}}{1-\bar{m}_{H,I(L_k)}},\ n=1,2,\cdots,N\\ \beta_H &= \frac{\tilde{m}_{H,I(L_k)}}{1-\bar{m}_{H,I(L_k)}}.\end{aligned} \tag{8.6}$$

The value β_n represents the degree of belief to which the general attribute y is associated to the evaluation grade H_n, and β_H represents the unassigned belief. Using results obtained by (8.6), we can define a set containing pairs of evaluation grades and the corresponding assigned beliefs, that completely describes the assessment of the general attribute y. Such a set is defined as:

$$S(y) = \{(H_n, \beta_n), n = 1, \cdots, N\}. \tag{8.7}$$

Although the assessment of the general attribute $S(y)$ offers a detailed insight into the assessment for each evaluation grade H_n, it does not offer a mechanism for alternative comparison. Such a comparison is an essential step in the process of alternative ranking. One way around this problem is the assignment of utility value $u(H_n)$ to each evaluation grade H_n that reflects its preference i.e.

$$u(H_{n+1}) > u(H_n) \quad \text{if } H_{n+1} \text{ is preferred to } H_n.$$

Based on these values and the values in the assessment of the general attribute, given by (8.7), the *expected utility value* is calculated as (Yang and Xu, 2002):

$$u\left[S(y)\right] = \sum_{n=1}^{N} u(H_n)\beta_n, \tag{8.8}$$

However, the expected utility function defined by (8.8) ignores the information contained in the unassigned belief β_H, which may contribute to the belief of any of the evaluation grades H. There are two extreme cases: assign β_H the maximum utility value or assign β_H to the minimum utility value. As a result of these assignments we obtain the following two variations of the expected utility value (Yang, 2001):

$$\begin{aligned}u_{max}[S(y)] &= \sum_{n=1}^{N-1} \beta_n u(H_n) + \Big(\beta_N + \beta_H\Big)u(H_N)\\ u_{min}[S(y)] &= \Big(\beta_1 + \beta_H\Big)u(H_1) + \sum_{n=2}^{N} \beta_n u(H_n)\end{aligned} \tag{8.9}$$

Consequently, the ranking between two alternatives (in our case two EC motors) with two different assignments of the final belief $S(y)_k$ and $S(y)_m$, can be done in the following manner:

1. $S(y)_k \succ S(y)_m$ if and only if $u_{min}(S(y)_k) > u_{max}(S(y)_m)$ and

2. $S(y)_k \equiv S(y)_m$ if and only if $u_{max}(S(y)_k) = u_{max}(S(y)_m)$,

where "$\succ$" means *is strictly preferred to* and "$\equiv$" means *is equivalent to*. It should be noted that for cases when the assessment is complete, i.e. $\beta_H = 0$, the relations (8.9) reduce to (8.8).

8.3 Application of EFDI to the end-quality assessment of EC motors

Quality assessment is based on calculated set of features. Having features, one can create, the aggregation structure that will define the general attributes as well as the inputs for the calculation of the overall utility value. In this section we will first briefly summarise the feature extraction part of the EFDI system. Then we will turn focus to the quality assessment part.

8.3.1 Feature set

In this application we focus on mechanical faults of which the most frequent are rotor and bearing faults. The faults contained in each of these two subsets are listed in Table 8.1.

Feature group	Features (Amplitude of spectral components at)
Rotor faults	f_{rot}
	$2 \times f_{rot}$
Bearing faults	$n \times f_{BPFO}$
	$n \times f_{BPFI}$
	$n \times f_{BSF}$
	$n \times f_{FTF}$
	where $n = \{1, 2\}$

Table 8.1: Features related to the mechanical faults (see Table 5.1 for characteristic bearing frequencies)

Having defined the required feature set, the feature extraction module was built according to the structure is shown in Figure 8.3.

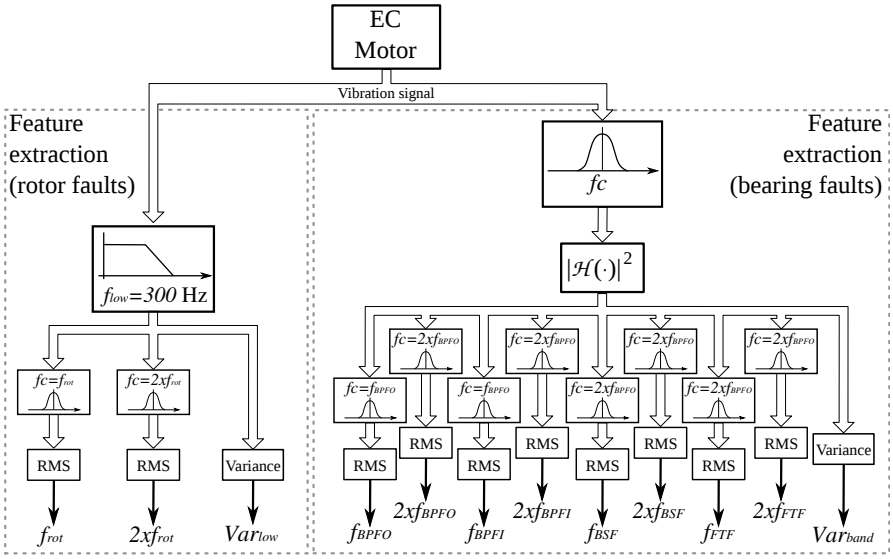


Figure 8.3: Structure of the feature extraction module

The procedure for rotor faults starts with low-pass filtering of the acquired vibration signal. Based on that, three features are extracted: spectral components at f_{rot} and $2 \times f_{rot}$, and the signal variance. Amplitudes of both spectral components are extracted by calculating the RMS value of vibration signal filtered by band-pass filters centred at f_{rot} and $2 \times f_{rot}$ respectively.

Similar procedure was applied for bearing faults. First the vibration signal is band-pass filtered with central frequency f_c located near the excited eigenfrequency. Such a filtering fulfils the conditions needed for calculation of the signal's envelope. The final feature set (extracted from the envelope) consists of the envelope variance, and spectral components located at the bearing fault frequencies. These components are extracted by filtering the envelope signal by means of a set of narrow band-pass filters. Each filter is centred at the corresponding bearing fault frequency.

8.3.2 Hierarchical aggregation structure

The problem of overall motor quality is split into two sub-problems, i.e. assessment of the quality of the electrical part and mechanical part separately. The assessment of the mechanical components can be done by aggregating the estimates for the quality of the rotor shaft and the motor bearings. Finally, the quality of these two components can be directly inferred from information contained in the measured feature set (see Figure 8.3. Such a decomposition process leads to the hierarchical structure shown in Figure 8.4. Consequently, with aggregation of each node, quality assessment for each motor part can be obtained. The aggregation process continues up to the point where the overall motor

quality can be obtained, thus providing a general overview of the examined motor.

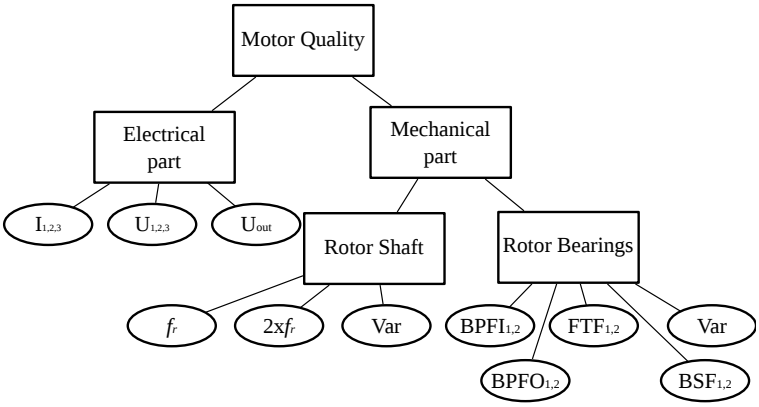


Figure 8.4: Dependancy of the overall motor quality on quality of constituent parts

Besides improvements in the fault isolation task, the application of ER provides an additional mechanism, in the form of *utility function*, that can be applied to overall quality assessment. The value of the utility is calculated for each candidate and it can be regarded as a measure of the motor quality. Consequently, ranking of motors can be performed based on this value.

Therefore, the enhancements in the processes of fault isolation and quality assessment offered by the ER approach, makes this method a liable candidate for the implementation of MADM in the cases of end-quality assessment systems.

8.4 Experimental results of EFDI of EC motors

As we focus on mechanical faults, the previously described ER approach was applied only for assessment of the quality of the motor's mechanical parts (see Figure 8.4). The quality of the EC motors' mechanical part was evaluated using 5 distinct evaluation grades:

$$H = \{NotSatisfactory(NS), Good(G), VeryGood(VG), Excellent(E), Top(T)\}. \quad (8.10)$$

These evaluation grades are applied for both rotor and bearing quality assessment.

The set of basic attributes for the assessment of the bearing quality consists of 8 features (see Figure 8.3):

$$e_k \in \{e_{bpfo}, e_{2\times bpfo}, e_{ftf}, e_{bsf}, e_{2\times bsf}, e_{bpfi}, e_{2\times bpfi}, e_{var}\}.$$

Since the elements of e_k are positive reals, we need to transform them into an equivalent degree of belief, that will be afterwards used in the aggregation process. Such a transformation should provide a relation between the quantitative value and the particular

evaluation grade. Note that the attributes represent RMS values of a particular spectral component. So, they can be treated as "cost" attributes, i.e. lower value is preferred rather than a larger one.

The first step in defining such a transformation between a quantitative value and particular evaluation grade, is to establish intervals that will correspond to a particular evaluation grade. Consequently, any value acquired by a basic attribute e_k can be mapped into a particular evaluation grade H_n. For the case of EC motors, the first logical step is to define the maximal allowed value for each basic attribute. By doing so we implicitly define the interval corresponding to the lowest evaluation grade $H_1 = NS$, which becomes $[h_{1,k}, +\infty]$, where $h_{1,k}$ represents the maximal allowed value. The intervals for the remaining four evaluation grades were determined by dividing the interval $[0, h_{1,k}]$ using the following relation:

$$h_{n,k} = \frac{h_{1,k}}{2n},\ n = 2, \cdots, 5 \text{ and } k = 1, \cdots, L_k, \tag{8.11}$$

where the limits from $h_{2,k}$ to $h_{5,k}$ correspond to the evaluation grades $H_2 = G$ to $H_5 = T$ respectively. Using these intervals, the transformation of an arbitrary value v_k of a quantitative attribute e_k into degree of belief was performed using the following relation:

$$\begin{gathered} \beta_{n,k} = \frac{v_k - h_{n+1,k}}{h_{n,k} - h_{n+1,k}}, \quad \beta_{n+1,k} = 1 - \beta_{n,k} \\ \text{if } h_{n+1,k} \le v_k \le h_{n,k},\ n = 1, \cdots, 5 \text{ and } k = 1, \cdots, L_k. \end{gathered} \tag{8.12}$$

The intervals (8.11) and the corresponding transformation functions (8.12) result in the transformation function shown in Figure 8.5.

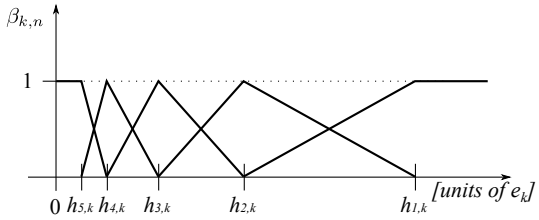


Figure 8.5: Degree of belief functions

From the values obtained by applying the relation (8.12), we can calculate the degree of belief originating from each attribute e_k. As each attribute was considered equally important for the overall bearing quality, the weights w_k in (8.3) are all set to 1/8. In the final step, the evaluation of the bearing quality is completed by aggregating the attributes beliefs using the relations (8.5)–(8.7).

The same approach is used for estimating the rotor quality. The same set of five evaluation grades (8.10) is applied. The only difference is in the number of attributes,

which is now three:

$$e_k \in \{e_{f_{rot}}, e_{2\times f_{rot}}, e_{var}\}. \tag{8.13}$$

Consequently, the weights w_i in (8.3) are all set to 1/3.

In the final step we need to calculate the motor's overall mechanical quality. Since both parts are equally important for the mechanical quality, the weights are set to 1/2. The state of the overall mechanical quality is then calculated using the obtained belief masses and the aggregation relations (8.5)–(8.7).

For the final ranking of the motor quality we employ the utility function (8.9). Since the belief assignment relation (8.12), guarantees complete assignment, the remaining unassigned belief $\beta_{H,i} = 0$. Then for (8.9) the following holds:

$$u_{max} = u_{min}.$$

As a result of this equality the utility of each EC motor was evaluated using the relation (8.8). For each evaluation grade, $u(H_n)$ was obtained by setting $u(H_{Top}) = 1$ and $u(H_{NotSatisfactory}) = 0$ and the utilities of all other grades were set as equidistant points in the interval $[u(H_{NotSatisfactory}), u(H_{Top})]$.

8.4.1 Detailed results for a selected EC motor

Bearing quality In order to see how the ER approach works in case of motor quality assessment, we present detailed analysis of one examined motor. The measured values v_i for all attributes are given in Table 8.2.

Attribute	*FTF*	*BPFO*		*BPFI*		*BSF*		*Var*
$[mm/s^2]$		1×	2×	1×	2×	1×	2×	
Value of v_i	0.2307	1.3110	1.8377	2.0329	1.3870	0.7350	1.6365	8.2545
h_1 limit	0.35	2	3.5	5	5	5	2.6	14

Table 8.2: Measured values for the bearing attributes

The degrees of belief for each attribute e_i were calculated using the relation (8.12). The associated degrees of belief with the corresponding evaluation grades are given in Table 8.3.

Aggregating these attribute values using the relations (8.5) and (8.6) we obtain the overall assessment of the bearing quality for this particular motor as (8.7):

$$S(y_{bearing}(bearing_1)) = \{(NS, 0); (G, 0.2726); (VG, 0.4657); (E, 0.2172); (T, 0.0445)\}, \tag{8.14}$$

	NotSatisfactory	*Good*	*VeryGood*	*Excellent*	*Top*
FTF	0	0.6369	0.3631	0	0
BPFO	0	0.6220	0.3780	0	0
2×BPFO	0	0.1002	0.8998	0	0
BPFI	0	0	0.6263	0.3737	0
2×BPFI	0	0	0.1096	0.8904	0
BSF	0	0	0	0.5880	0.4120
2×BSF	0	0.5177	0.4823	0	0
Var	0	0.3584	0.6416	0	0

Table 8.3: Degrees of belief for the bearing attributes

where the label $bearing_1$ is just a notation to indicate the name of the examined alternative i.e. the bearing of the first examined motor.

Rotor quality Following the same approach the values of the three measured attributes for the rotor quality (8.13) are presented in Table 8.4.

Attribute	f_r	$2 \times f_r$	*Var*
Value of v_i $[mm/s^2]$	0.0689	0.0017	0.0056
h_1 limit $[mm/s^2]$	0.8	0.03	0.5

Table 8.4: Measured values for the rotor attributes

The degrees of belief obtained thereupon are shown in Table 8.5. Consequently, the

f_r	$2 \times f_r$	*Var*
{(E,0.3446), (T,0.3446)}	{(E,0.2222), (T,0.7778)}	{(E,0.0446), (T,0.9554)}

Table 8.5: Degrees of belief for the rotor attributes

overall rotor quality is calculated by aggregating the degrees of belief using the relations (8.3), (8.5) and (8.6) with $w_i = 1/3$, yielding the following result:

$$S(y_{rotor}(rotor_1)) = \{(NS, 0); (G, 0); (VG, 0); (E, 0.1612); (T, 0.8388)\}. \tag{8.15}$$

Overall quality and utility assessment The final step is the assessment of the overall mechanical quality of the examined EC motor. According to the attribute relationship, shown in Figure 8.4, the overall mechanical quality can be obtained by aggregating the bearing and rotor quality. The aggregation process uses the already determined degrees of belief for the bearing quality (8.14) and for the rotor quality (8.15). Since both compound elements equally influence the overall motor quality, the weight factors w_i take value of 1/2. Finally, the aggregated motor quality has been assessed as:

$$S(y_{motor}(motor_1)) = \{(NS, 0); (G, 0.1315); (VG, 0.2247); (E, 0.1995); (T, 0.4442)\}.$$

Using the relation (8.8), the utility of the examined motor was determined to be

$$u[S(y_{motor}(motor_1))] = 0.7391.$$

8.4.2 Results

The proposed approach is applied on a batch of 130 EC motors. Motors with different quality as well as motors with mechanical damage were included. The decision whether the motor was acceptable or not was done by examining two criteria:

1. the value of the utility function (8.8), and
2. the belief assigned to the "*NotSatisfactory*" evaluation grades (8.10).

The overall expected utility values for the tested motors are shown in Figure 8.6. There are three distinctive groups marked with different shades of grey: the first group contains motors with acceptable quality; the second group contains motors that have non-zero belief for the evaluation grade "*NotSatisfactory*"; and the third group consists of only two motors that represent special cases for which the efficiency of the algorithm is the most expressed.

The motors belonging to the group with the highest utility values are the ones that have belief assigned to the top evaluation grades H. Consequently, these motors can be regarded as units with acceptable quality.

The second group of motors include units that have some of the belief assigned to the "*NotSatisfactory*" evaluation grade. Since utility for this evaluation grade is zero $U(H_{NS}) = 0$, these motors have lower utility value. In order to examine, why these motors have some overall belief assigned to the "*NotSatisfactory*" evaluation grade, we analysed the assigned belief to the particular mechanical parts, i.e. "Rotor shaft" and "Rotor Bearings" (see Figure 8.4). By doing so we actually executed the fault isolation task. The combined process of quality assessment and fault isolation is most clearly expressed in the two special cases: the 65^{th} and the 122^{nd} test motor.

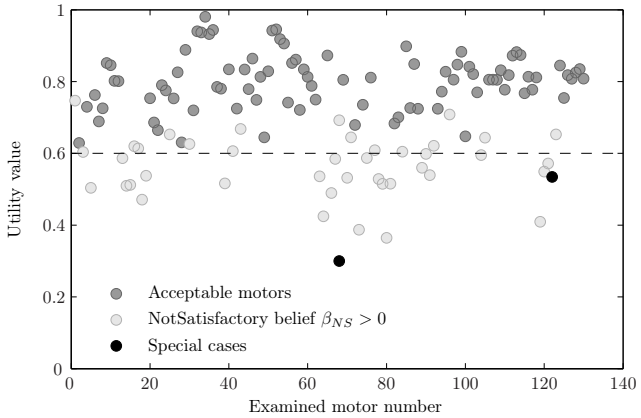


Figure 8.6: Expected Utility values for the examined test motors

The 65^{th} test motor has the lowest utility value. The mechanical condition of this motor was evaluated as:

$$S(y_{motor}(a_{65}) = \{(NS, 0.4911); (G, 0.3351); (VG, 0.1566); (E, 0.0172); (T, 0)\}\,.$$

From the result, it is visible that the majority of the belief assigned to the two lowest evaluation grades, i.e. *NS* and *G*, hence such low expected utility value. The reason for such a high value of belief assigned to the "*NotSatisfactory*" evaluation grade can be determined by analysing the assessments of the "Rotor shaft" and "Rotor Bearings" general attributes. For this particular motor they take the following values:

$$\begin{aligned} S(y_{bearings}(a_{65}) &= \{(NS, 0.0066); (G, 0.3253); (VG, 0.5740); (E, 0.0940); (T, 0)\}\,, \\ S(y_{shaft}(a_{65}) &= \{(NS, 0.8068); (G, 0.1932); (VG, 0); (E, 0); (T, 0)\}\,. \end{aligned} \tag{8.16}$$

From the belief assignments shown in (8.16), we can conclude that the main reason for such a low overall quality is a fault in the rotor shaft, since it has most of the belief assigned to the "*NotSatisfactory*" evaluation grade.

The second case, the 122^{nd} test motor, has zero belief assigned to the "*NotSatisfactory*" evaluation grade. Despite this, the low expected utility value classifies the motor in the group with unsatisfactory mechanical quality. In order to explain such a behaviour we have to examine the belief assignment to the aggregated motor quality:

$$S(y_{motor}(a_{122}) = \{(NS, 0); (G, 0.8111); (VG, 0.1889); (E, 0); (T, 0)\}\,.$$

Most of the belief is assigned to the evaluation grade "*Good*". By applying the relation (8.8) the expected utility becomes

$$u[S(y_{motor}(motor_{122}))] = 0.5342.$$

Although the motor has no belief assigned to the evaluation grade *NotSatisfactory*, low value of the utility is assigned as a consequence of the decision to favour top quality products. Such a strategy concentrates the majority of the utility to the higher evaluation grades, hence favouring them over the others.

Similarly to the previous case, we can go one step further and analyse the aggregated belief assignment for the "Rotor shaft" and "Rotor Bearings" general attributes, in order to isolate the possible cause for such a low utility value. These assignments for the 122^{nd} test motor are:

$$S(y_{bearings}(a_{122}) = \{(NS, 0); (G, 0.6408); (VG, 0.3592); (E, 0); (T, 0)\},$$
$$S(y_{shaft}(a_{122}) = \{(NS, 0); (G, 0.9075); (VG, 0.0925); (E, 0); (T, 0)\}.$$

By observing the assigned belief, we can conclude that both "Rotor shaft" and "Rotor Bearings" have most of the belief assigned to the evaluation grade "*Good*". Therefore, corrections in both areas are necessary so that this particular test motor can meet the required quality standard.

By using the presented results, we can conclude that the decision whether the quality of mechanical part should be accepted or not, can be based on the information obtained from the state of the general aggregated attribute as well as the calculated expected utility value. The results reveal the method's capability of detecting motors with both severe damages as well as motors whose mechanical quality is unacceptable in spite of the desirable low attribute values.

9 Conclusions

This dissertation presents three main results in the context of fault detection and identification of mechanical drives under variable operating conditions. The first result is a conceptual model of vibrations, generated by bearings with surface faults, defined in the framework of temporal point processes. The second result, concerning the feature extraction process, specifies a set of features whose values predominantly depend on the condition of the monitored drive and can be treated as sufficiently robust to variations in the operating conditions. Finally, the third results addresses the issues of fault isolation and identification in the context of end-quality assessment systems.

The proposed bearing vibration model is defined as temporal point process governed by Inverse Gaussian interevent distribution. Such a model properly describes bearing vibrations as periodic impulses under constant rotational speed and as quasi-cyclostationary random process for cases of variable rotational speed. Therefore, such a model is applicable regardless of the statistical characteristics of random process governing the rotational speed. Finally, an effective bearing fault detection can be performed using a feature set based on the statistical properties of such a point process and the corresponding counting process.

In addition to the bearing fault model, this thesis proposes a set of features capable of quantifying the most commonly occurring mechanical faults. The feature set exploits the fact that the vibrations produced by mechanical drives can be modelled as a sum of random sine waves. Thus, the features set quantifies the changes in the PDF of the envelope of the generated vibrations. Features are based on Rényi entropy and α-Jensen divergence and are estimated using wavelet packet coefficients. Defined in such a way, the feature set has two main assets:

- features are shown to be robust to changes in the operating conditions and in the same time sufficiently sensitive to changes in the drive's condition and
- the process of calculating the feature values relies neither on any information regarding the operating conditions nor on data describing physical characteristics regarding the monitored drive, such as bearing dimensions, gear type etc.

The calculation process impose no limitations on the statistical properties of the analysed

signals. Therefore, the proposed feature set is applicable to both stationary and variable operating conditions.

Apart from the proposed improvements in the feature extraction segment, this dissertations proposes a solution for fault isolation and identification based on multi-attribute decision making employing evidential reasoning. The method fuses the information contained in the feature set into a ranked list of possible faults as well as an single overall assessment of the utility of the monitored system. The evidential reasoning methodology mimics the reasoning process of the maintenance personnel thus making it intuitively understandable and open for seamless integration of experts knowledge.

Future work

The presented results covering the segment of condition monitoring of mechanical drives under variable operating conditions indicate several key improvement points. The first, and the probably the most important one, is upgrading the proposed methods beyond the stage of fault detection i.e. introducing techniques that can perform fault isolation and identification. Secondly, the performance of the methods was evaluated mainly on mechanical drives with low mechanical complexity. Finally, the time evolution of the calculated features should be investigated in order to examine the possibility of using these features for the purpose of prognostics. Completing these improvements will result into a feature set containing all needed information for each segment of the CBM/PHM paradigm.

Initially, the Rényi entropy based features were applied solely for the purpose of fault detection. However, from the present results it is apparent that different faults tend to group into clusters, which poses the question of the using these features for the purpose of fault diagnosis. However, at the present stage, the diagnostic resolution of the proposed feature set insufficient, and consequently several faults become indistinguishable. Therefore, the application of such a feature set for the purpose of fault diagnosis would require detailed analysis how different conditions affect the feature values, hence increasing the diagnostic resolution to a satisfactory level.

Probably the biggest challenge would be determining a link between the evolution of the feature values and the remaining useful life of the monitored system. This might be achieved by exploiting the fact that the entropy like features can be regarded as “arrow” in time i.e. their evolution can be linked with the time evolution of the monitored system. Thus, by analysing the trends of the statistical complexity of the acquired signals one might provide the needed information for the estimation of the remaining useful life.

10 Acknowledgements

I would like to pay special thanks to a number of people around me without whom this wonderful journey would have been impossible.

Firstly, I would like to express my deeply-felt thanks to my supervisor, Professor Dr. Đani Juričić, for his guidance, infinite patience and perpetual encouragement. His support transformed my doctoral studies into a magnificent experience.

Additionally, I wish to thank my co-supervisor Professor Dr. Mile Stankovski, for his trust and support throughout my studies. Also I would like to express my gratitude to Professor Dr. Stanislav Strmčnik and Professor Dr. Jožef Vižintin for accepting me as a member of their teams, hence providing me with all the necessary means for conducting my research.

Furthermore, I would like to thank my colleagues at the Department of Systems and Control at Jožef Stefan Institute for accepting me as co-worker and friend. Their support and friendship helped me feel Slovenia as my home.

Last but not least, I thank my wife Biljana for giving me the necessary and sufficient impulse to start this endeavour. Thank you for standing by my side despite all transients, local minima and state changes.

Finally, I would like to acknowledge the support of Ad futura Programme of the Slovene Human Resources and Scholarship Fund and the support of the Slovenian Research Agency through Research Programme P2-0001 and L2-2110.

11 References

A. Abdi, H. Hashemi, and S. Nader-Esfahani. On the pdf of the sum of random vectors. *IEEE Transactions on Communications*, 48(1):7 –12, jan 2000.

R. Aini, H. Rahnejat, and R. Gohar. Vibration modeling of rotating spindles supported by lubricated bearings. *Journal of Tribology*, 124(1):158–165, 2002.

P. F. Albrecht, J. C. Appiarius, and D. K. Shrama. Assesment of the reliability of motors in utility applications. *IEEE Transactions of Energy Conversion*, EC-1:39–46, 1986.

P. O. Amblard, M. Gaeta, and J. L. Lacoume. Statistics for complex variables and signals - part i: Variables. *Signal Processing*, 53:1–13, 1996.

T. Anagnos and A. S. Kiremidjian. Stochastic time-predictable model for earthquake occurrences. *Bulletin of the Seismological Society of America*, 74(6):2593–2611, 1984.

J. Antoni. Cyclic spectral analysis in practice. *Mechanical Systems and Signal Processing*, 21:597 – 630, 2007a.

J. Antoni. Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *Journal of Sound and Vibration*, 304:497 – 529, 2007b.

J. Antoni. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, 21:108–124, 2007c.

J. Antoni. Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, 23:987–1036, 2009.

J. Antoni and R. B. Randall. Differential diagnosis of gear and bearing faults. *Journal of Vibration and Acoustics*, 124(2):165–171, 2002.

J. Antoni and R. B. Randall. A stochastic model for simulation and diagnostics of rolling element bearings with localized faults. *Journal of Vibration and Acoustics*, 125 (3):282–289, 2003.

J. Antoni and R.B. Randall. The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing*, 20:282–307, 2006.

J. Antoni, F. Bonnardot, A. Raad, and M. El Badaoui. Cyclostationary modelling of rotating machine vibration signals. *Mechanical Systems and Signal Processing*, 18:1285 – 1314, 2004.

W. Bartelmus and R. Zimroz. A new feature for monitoring the condition of gearboxes in non–stationary operating conditions. *Mechanical Systems and Signal Processing*, 23 (5):1528–1534, 2009.

M. Basseville. Divergence measures for statistical data processing. Technical report, IRISA, 2010.

M. Basseville and J.-F. Cardoso. On entropies, divergences, and mean values. In *Information Theory, 1995. Proceedings., 1995 IEEE International Symposium on*, page 330, sep 1995.

N. Baydar and A. Ball. Detection of gear deterioration under variying load conditions using the instantaneous power spectrum. *Mechanical Systems and Signal Processing*, 14 (6):907–921, 2000.

V. Belton and T. J. Steward. *Multiple Criteria Decision Analysis, An Integrated Approach*. Kluwer Academic Publishers, 2002.

A. Ben Hamza, Yun He, and H. Krim. An information divergence measure for isar image registration. In *Statistical Signal Processing, 2001. Proceedings of the 11th IEEE Signal Processing Workshop on*, pages 130 –133, 2001.

S. Blanco, A. Figliola, R. Quian Quiroga, O. A. Rosso, and E. Serrano. Time-frequency analysis of electroencephalogram series. iii. wavelet packets and information cost function. *Phys. Rev. E*, 57(1):932–940, 1998.

B. Boashash. Estimating and interpreting the instantaneous frequency of a signal part i: Fundamentals. *Proceedings of the IEEE*, 80(4):540 –568, Apr 1992.

P. Boškoski and Đ. Juričić. Point processes for bearing fault detection under non-stationary operating conditions. In *Annual Conference of the Prognostics and Health Management Society*, pages 427–434, Montreal, QC, Canada, September 2011.

P. Boškoski, Đ. Juričić, and M. Stankovski. Gear and bearing fault detection under variable operating conditions. In *Annual Conference of the Prognostics and Health Management Society*, page 8, Portland, OR, USA, October 2010a. URL `http://www.phmsociety.org/sites/phmsociety.org/files/phm_submission/2010/phmc_10_017.pdf`.

P. Boškoski, J. Petrovčič, B. Musizza, and Đ. Juričić. Detection of lubrication starved bearings in electrical motors by means of vibration analysis. *Tribology International*, 43 (9):1683 – 1692, 2010b.

P. Boškoski, J. Petrovčič, B. Musizza, and Đ. Juričić. An end-quality assessment system for electronically commutated motors based on evidential reasoning. *Expert Systems with Applications*, 38(11):13816 – 13826, 2011.

J. Briët and P. Harremoës. Properties of classical and quantum jensen-shannon divergence. *Phys. Rev. A*, 79(5):052311, 2009.

C. S. Burrus, R. A.. Gopinath, and H. Guo. *Introduction to Wavelets and Wavelet Transforms: A Primer*. Prentice Hall, New Jersey, 1994.

W. Caesarendra, A. Widodo, and B. Yang. Application of relevance vector machine and logistic regression for machine degradation assessment. *Mechanical Systems and Signal Processing*, 24(4):1161 – 1171, 2010.

C. Castejón, O. Lara, and J.C. García-Prada. Automated diagnosis of rolling bearings using mra and neural networks. *Mechanical Systems and Signal Processing*, 24(1):289 – 299, 2010.

P. Chen, M. Taniguchi, T. Toyota, and Z. He. Fault diagnosis method for machinery in unsteady operating condition by instantaneous power spectrum and genetic programming. *Mechanical Systems and Signal Processing*, 19(1):175 – 194, 2005.

A. Cichocki and S. Amari. Families of alpha- beta- and gamma- divergences: Flexible and robust measures of similarities. *Entropy*, 12(6):1532–1568, 2010.

D. R. Cox and V. Isham. *Point Processes*. Chapman and Hall, Cambidge, 1980.

C. J. Crabtree. Survey of commercially available condition monitoring systems for wind turbines. Technical report, Durham University, School of Engineering and Computing Science, 2010.

D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Elementary Theory and Methods*, volume 1. Springer-Verlag, New York, second edition edition, 2003.

I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

F. M. Dekking, C. K., H. P. Lopuhaä, and Ludolf Erwin Meester. *A Modern Introduction to Probability and Statistics*. Springer, London, 2005.

D. Dowson. Elastohydrodynamic and micro-elastohydrodynamic lubrication. *Wear*, 190 (2):125 – 138, 1995.

R. Dwyer. Detection of non-gaussian signals by frequency domain kurtosis estimation. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP*, 8:607–610, 1983.

J. Eriksson, E. Ollila, and V. Koivunen. Statistics for complex random variables revisited. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 3565 –3568, april 2009.

R. Esposito and L. Wilson. Statistical properties of two sine waves in gaussian noise. *IEEE Transactions on Information Theory*, 19(2):176 – 183, mar 1973.

X. Fan and M. J. Zuo. Gearbox fault detection using hilbert and wavelet packet transform. *Mechanical Systems and Signal Processing*, 20:966 – 982, 2006.

Y. Feng and F. S. Schlindwein. Normalized wavelet packets quantifiers for condition monitoring. *Mechanical Systems and Signal Processing*, 23(3):712–723, 2009.

A. Figliola and E. Serrano. Analysis of physiological time series using wavelet transforms. *IEEE Engineering in Medicine and Biology*, 16(3):74–79, 1997.

J. L. Folks and R. S. Chhikara. The inverse gaussian distribution and its statistical application–a review. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(3):263–289, 1978.

W. A. Gardner, A. Napolitano, and L. Paura. Cyclostationarity: Half a century of research. *Signal Processing*, 86:639 – 697, 2006.

C. Gargour, M. Gabrea, V. Ramachandran, and J.-M. Lina. A short introduction to wavelets and their applications. *Circuits and Systems Magazine, IEEE*, 9(2):57 –68, quarter 2009.

M. Ge, R. Du, and Y. Xu. Hidden markov model based fault diagnosis for stamping processes. *Mechanical Systems and Signal Processing*, 18(2):391 – 408, 2004.

J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. CRC Press, New York, 1998.

GfT. Arbeitsblatt 7 tribologie definitionen, begriffe, prüfung (in german). Technical Report 7, GfT, 2002.

A. Grossmann and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. 15(4):723–736, 1984.

A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69: 331–371, 1910.

C.W. Helstrom. Distribution of the envelope of a sum of random sine waves and gaussian noise. *IEEE Transactions on Aerospace and Electronic Systems*, 35(2):594 –601, apr 1999.

A. O. Hero, B. Ma, O. Michel, and J. Gorman. Alpha-divergence for classification, indexing and retrieval. Technical Report CSPL-328, Communications and Signal Processing Laboratory, The University of Michigan, 2002.

D. Ho and R. B. Randall. Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mechanical Systems and Signal Processing*, 14(5):763–788, 2000.

I. Howard. *A Review of Rolling Element Bearing Vibration "Detection, Diagnosis and Prognosis"*. DSTO Aeronautical and Maritime Research Laboratory, 1994.

I. Howard, S. Jia, and J. Wang. The dynamic modelling of a spur gear in mesh including friction and crack. *Mechanical Systems and Signal Processing*, 15:831–853, 2001.

N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Journal of Proceedings of the Royal Society of London Series A*, 454:903 – 995, 1998.

R. Huang, L. Xi, X. Li, C. R. Liu, H. Qiu, and J. Lee. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical Systems and Signal Processing*, 21(1):193 – 207, 2007.

T. Igarashi. Noise of ball bearing in electric motor. *The Japan Society of Mechanical Engineers*, 7(25):200–208, 1964.

R. Isermann and P. Ballé. Trends in the application of mode-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 5:709–719, 1997.

L. B. Jack and A. K. Nandi. Fault detection using support vector machines and artificial neural networks augmented by genetic algorithms. *Mechanical Systems and Signal Processing*, 16:373–390, 2001.

A.K.S. Jardine, D. Lin, and D. Banjevič. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(6):1483–1510, 2006.

C. Kar and A. R. Mohanty. Monitoring gear vibrations through motor current signature analysis and wavelet transform. *Mechanical Systems and Signal Processing*, 20(1):158–187, 2006.

C. Karlsson, J. Arriagada, and M. Genrup. Detection and interactive isolation of faults in steam turbines to support maintenance decisions. *Simulation Modelling Practice and Theory*, 16(10):1689 – 1703, 2008.

J.-H. Kuang and A.-D. Li. Theoretical aspects of torque responses in spur gearing due to mesh stiffness variation. *Mechanical Systems and Signal Processing*, 17:255–271, 2003.

S. J. Louitridis. Damage detection in gear systems using empirical mode decomposition. *Engineering Structures*, 26:1833 – 1841, 2004.

N. Lynagh, H. Rahnejat, M. Ebrahimi, and R. Aini. Bearing induced vibration in precision high speed routing spindles. *International Journal of Machine Tools and Manufacture*, 40(4):561 – 577, 2000.

S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, Burlington, MA, 3rd edition, 2008.

M. V. Matthews, W. L. Ellsworth, and P. A. Reasenberg. A brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America*, 92(6):2233–2250, 2002.

J. M. Mendel. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proceedings of the IEEE*, 79 (3):278–305, 1991.

Q. Miao and V. Makis. Condition monitoring and classification of rotating machinery using wavelets and hidden markov models. *Mechanical Systems and Signal Processing*, 21(2):840 – 855, 2007.

G. Michaelov, S. Sarkani, and L.D. Lutes. Spectral characteristics of nonstationary random processes – a critical review. *Structural Safety*, 21(3):223–244, 1999.

A. Muller, M. Suhner, and B. Iung. Maintenance alternative integration to prognosis process engineering. *Journal of Quality in Maintenance Engineering*, 13(2):198–211, 2007.

F. Nielsen and S. Boltz. The burbea-rao and bhattacharyya centroids. *IEEE Transactions on Information Theory*, 57(8):5455 –5466, aug. 2011.

H. Ochiai. Exact and approximate distributions of instantaneous power for pulse-shaped single-carrier signals. *IEEE Transactions on Wireless Communications*, 10(2):682 –692, february 2011.

Y. Ogata. On lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23 – 31, January 1981.

H. N. Özguven and D. R. Houser. Mathematical models used in gear dynamics: A review. *Journal of Sound and Vibration*, 121(3):383–411, 1988.

L.R. Padovese. Hybrid time–frequency methods for non-stationary mechanical signal analysis. *Mechanical Systems and Signal Processing*, 18(4):1047—-1064, 2004.

A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1991.

B. E. Parker, H. A. Ware, D. P. Wipf, W. R. Tompkins, B. R. Clark, and E. C. Larson. Fault diagnosis using statistical change detection in the bispectral domains. *Mechanical Systems and Signal Processing*, 14(4):561–570, 2000.

Z. Peng and F. Chu. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical Systems and Signal Processing*, 18:199–211, 2004.

Z. K. Peng, P. W. Tse, and F. L. Chu. An improved hilbert – huang transform and its application in vibration signal analysis. *Journal of Sound and Vibration*, 286:187 – 205, 2005.

W. D. Penny. *Signal Processing Course*. 2000.

D. B. Percival and A. T. Walden. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge, 2000.

PHM. Prognostics and health managment society 2009 data challenge. http://www.phmsociety.org/competition/09, 2009.

B. Picinbono. On circularity. *IEEE Transactions on Signal Processing*, 42(12):3473 –3482, dec 1994.

A. Ponce-Alvarez, B. Kilavik, and A. Riehle. Comparison of local measures of spike time irregularity and relating variability to firing rate in motor cortical neurons. *Journal of Computational Neuroscience*, 29(1–2):351–365, 2010.

A.G. Poulimenos and S.D. Fassois. Parametric time-domain methods for non-stationary random vibration modelling and analysis — a critical survey and comparison. *Mechanical Systems and Signal Processing*, 20(4):763—-816, 2006.

M.B. Priestley. *Spectral analyses and Time Series*. Academic Press, London, 1981.

J. Rafiee, F. Arvani, A. Harifi, and M.H. Sadeghi. Intelligent condition monitoring of a gearbox using artificial neural network. *Mechanical Systems and Signal Processing*, 21 (4):1746 – 1754, 2007.

A. Rakar, Đ. Juričić, and P. Ballé. Transferable belief model in fault diagnosis. *Engineering Applications of Artifcial Intelligence*, 12(5):555–567, 1999.

R. B. Randall. State of the art in monitoring rotating machinery. In *Proceedings of ISMA 2002*, volume IV, pages 1457–1478, September 16-18 2002.

R. B. Randall, J. 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*, 15:945 – 962, 2001.

R.B. Randall. A new method of modeling gear faults. *Journal of Mechanical Design*, 104:259–267, 1982.

Motor reliability working group MRWG. Report of large motor reliability survey of industrial and commercial installations, part i. *IEEE Transactions of Industry Applications*, IA-21:853–864, 1985a.

Motor reliability working group MRWG. Report of large motor reliability survey of industrial and commercial installations, part ii. *IEEE Transactions of Industry Applications*, IA-21:865–872, 1985b.

Motor reliability working group MRWG. Report of large motor reliability survey of industrial and commercial installations, part iii. *IEEE Transactions of Industry Applications*, IA-23:153–158, 1985c.

A. Rényi. On measures of information and entropy. In *4th Berkeley Symposium on Mathematics, Statistics and Probability*, 1960.

R. Ricci and P. Pennacchi. Diagnostics of gear faults based on emd and automatic selection of intrinsic mode functions. *Mechanical Systems and Signal Processing*, 25(3): 821 – 838, 2011.

O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, and E. Basar. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods*, 105(1):65–75, 2001.

O.A. Rosso, M.T. Martin, A. Figliola, K. Keller, and A. Plastino. Eeg analysis using wavelet-based information tools. *Journal of Neuroscience Methods*, 153:163–182, 2006.

R. Rubini and U. Meneghetti. Application of the envelope and wavelet transform and analyses for the diagnosis of incipient faults in ball bearings. *Mechanical Systems and Signal Processing*, 15(2):287–302, 2001.

S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition, 2009. ISBN 0136042597, 9780136042594.

B. Samanta. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical Systems and Signal Processing*, 18:625–644, 2004.

N. Sawalhi and R. Randall. Simulating gear and bearing interactions in the presence of faults part i. the combined gear bearing dynamic model and the simulation of localised bearing faults. *Mechanical Systems and Signal Processing*, 22:1924–1951, 2008.

N. Sawalhi, R.B. Randall, and H. Endo. The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. *Mechanical Systems and Signal Processing*, 21:2616–2633, 2007.

M. Schlechtingen and I. F. Santos. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical Systems and Signal Processing*, 25(5):1849 – 1875, 2011.

F. P. Schoenberg. Point processes. Technical report, UCLA Department of Statistics, 2001.

M. Simon. On the probability density function of the squared envelope of a sum of random phase vectors. *IEEE Transactions on Communications*, 33(9):993 – 996, sep 1985.

P. D. Spanos and G. Failla. Evolutionary spectra estimation using wavelets. *Journal of Engineering Mechanics*, 130(8):952–960, 2004.

C.J. Stander and P.S. Heyns. Instantaneous angular speed monitoring of gearboxes under non-cyclic stationary load conditions. *Mechanical Systems and Signal Processing*, 19(4): 817—-835, 2005.

W. J. Staszewski. Wavelet based compression and feature selection for vibration analysis. *Journal of Sound and Vibration*, 211:735 – 760, 1998.

V. Sugumaran and K.I. Ramachandran. Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. *Mechanical Systems and Signal Processing*, 21(5):2237 – 2247, 2007.

T.A.Harris. *Rolling Bearing Analysis*. Wiley-Interscience, New York, 2000.

N. Tandon and A. Choudhury. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International*, 32:469–480, 1999.

M. C. K. Tweedie. Statistical properties of inverse gaussian distributions. i. *The Annals of Mathematical Statistics*, 28(2):362–377, 1957.

M. Unser and P. Tafti. Stochastic models for sparse and piecewise-smooth signals. *IEEE Transactions on Signal Processing*, PP(99):1–1, 2010.

G. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess, and B. Wu. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Wiley, 2006.

C. van Vreeswijk. *Analysis of Parallel Spike Trains*, volume 7 of *Springer Series in Computational Neuroscience*, chapter Stochastic Models of Spike Trains, pages 3–20. Springer, 2010.

V. D. Vrabie, P. Granjon, and C. Serviere. Spectra kurtosis: From definition to application. In *6th IEEE International Workshop on Nonlinear Signal and Image Processing (NSIP 2003), Grado Trieste*, 2003.

W. Wang. Early detection of gear tooth cracking using the resonance demodulation technique. *Mechanical Systems and Signal Processing*, 15:887–903, 2001.

W. Wang and D. Kanneg. An integrated classifier for gear system monitoring. *Mechanical Systems and Signal Processing*, 23(4):1298 – 1312, 2009.

X. Wang, V. Makis, and M. Yang. A wavelet approach to fault diagnosis of a gearbox under varying load conditions. *Journal of Sound and Vibration*, 329(9):1570 – 1585, 2010.

Y. Wang and T. M.S. Elhag. Evidential reasoning approach for bridge condition assessment. *Expert Systems with Applications*, 34:689–699, 2008.

G. A. Whitmore and V. Seshadri. A Heuristic Derivation of the Inverse Gaussian Distribution. *The American Statistician*, 41(4):280–281, 1987.

A. Widodo and B. Yang. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 21(6):2560–2574, 2007.

Y. H. Wijnant, J. A. Wensing, and G. C. van Nijen. The influence of lubrication on the dynamic behaviour of ball bearings. *Journal of Sound and Vibration*, 222:579–596, 1999.

K. Worden, W. J. Staszewski, and J. J. Hensman. Natural computing for mechanical systems research: A tutorial overview. *Mechanical Systems and Signal Processing*, 25 (1):4 – 111, 2011.

D. Xu and D. Erdogmuns. *Information Theoretic Learning*, chapter Rényi's Entropy, Divergence and Their Nonparametric Estimators, pages 47–102. Springer, 2010.

J. Yang. Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties. *European Journal of Operational Research*, 131:31–61, 2001.

J. Yang and Pratyush Sen. A general multi-level evaluation process for hybrid madm with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics*, 24:1458–1473, 1994.

J. Yang and Madan G. Singh. An evidential reasoning approach for multiple-attribute decision making with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics*, 24:1–18, 1994.

J. Yang and Dong Ling Xu. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics*, 32:289–304, 2002.

K. Yoon and C. Hwang. *Multiple attribute decision making: An introduction.* Sage Publications, 1995.

Y. Yu, Y. Dejie, and C. Junsheng. A roller bearing fault diagnosis method based on emd energy entropy and ann. *Journal of Sound and Vibration*, 294:269 – 277, 2006.

Y. Zhan, V. Makis, and A. K.S. Jardine. Adaptive state detection of gearboxes under varying load conditions based on parametric modelling. *Mechanical Systems and Signal Processing*, 20(1):188–221, 2006.

L. Zunino, D.G. Pérez, M. Garavaglia, and O.A. Rosso. Wavelet entropy of stochastic processes. *Physica A: Statistical Mechanics and its Applications*, 379(2):503–512, 2007.

List of Figures

List of Tables

Publications related to the dissertation

P. Boškoski and Đ. Juričić. Point processes for bearing fault detection under non-stationary operating conditions. In *Annual Conference of the Prognostics and Health Management Society*, pages 427–434, Montreal, QC, Canada, September 2011.

P. Boškoski and A. Urevc. Bearing fault detection with application to phm data challenge. *International Journal of Prognostics and Health Managment*, 2(1):10, 2011. URL `http://www.phmsociety.org/sites/phmsociety.org/files/phm_submission/2009/ijPHM_11_003.pdf`.

P. Boškoski, Đ. Juričić, and M. Stankovski. Gear and bearing fault detection under variable operating conditions. In *Annual Conference of the Prognostics and Health Management Society*, pages 8, Portland, OR, USA, October 2010a. URL `http://www.phmsociety.org/sites/phmsociety.org/files/phm_submission/2010/phmc_10_017.pdf`.

P. Boškoski, J. Petrovčič, B. Musizza, and Đ. Juričić. Detection of lubrication starved bearings in electrical motors by means of vibration analysis. *Tribology International*, 43 (9):1683 – 1692, 2010b.

P. Boškoski, J. Petrovčič, B. Musizza, and Đ. Juričić. An end-quality assessment system for electronically commutated motors based on evidential reasoning. *Expert Systems with Applications*, 38(11):13816 – 13826, 2011a.

P. Boškoski, J. Petrovčič, B. Musizza, Đ. Juričić, and A. Biček. End-quality assessment of electrical motors based on the concept of virtual sensors. *Ventil*, 17(2):148–153, 2011b.

M. Gašperin, Đ. Juričić, P. Boškoski, and J. Vižintin. Model-based prognostics of gear health using stochastic dynamical models. *Mechanical Systems and Signal Processing*, 25(2):537–548, 2011.

Other results

- Second ranked on the competition for Innovation/Innovative project at the Technology Transfer Conference 2010.
- Best result achieved on the PHM 2009 Data Challenge for bearing fault detection