University of Southampton
Faculty of Engineering and Applied Science
Institute of Sound and Vibration Research

HIGHER ORDER SPECTRA AND THEIR APPLICATION TO NONLINEAR MECHANICAL SYSTEMS

William Beningfield Collis

A thesis submitted for the degree of

Doctor of Philosophy

February 1996

University of Southampton

Abstract

Faculty of Engineering and Applied Science
Institute of Sound and Vibration Research

Doctor of Philosophy

Higher Order Spectra and their Application to Nonlinear Mechanical Systems

by William Beningfield Collis

This thesis is concerned with the development of useful engineering techniques to detect and analyse nonlinearities in mechanical systems. The methods developed are based on the concepts of higher order spectra, in particular the bispectrum and trispectrum, and the Volterra series. The study of higher order statistics has been dominated by work on the bispectrum. The bispectrum can be viewed as a decomposition of the third moment (skewness) of a signal over frequency and as such is blind to symmetric nonlinearities. To study such phenomena one has to go a stage further and resort to the trispectrum, or decomposition of kurtosis over frequency. Techniques are presented here that enable one to estimate and display both auto and cross, bispectra and trispectra.

Initially auto higher order spectra are studied in detail with particular attention being paid to normalisation methods. Two traditional methods based on the bicoherence and skewness function are studied and these are expanded to their fourth order equivalents, the tricoherence and kurtosis functions. Under certain conditions, notably narrow band signals, the above normalisation methods are shown to fail and so a new technique based on pre whitening the signal in the time domain is developed. Examples of these functions are given both for memoryless and dynamic systems. The Volterra series is presented and discussed in some detail. Techniques for calculating a system's Volterra kernels from cross higher order spectra are derived. New methods are presented for the estimation of higher order Volterra kernels which are shown to produce better results than traditional approaches. These are then applied to some simple analytic systems which include the Duffing oscillator. Some discussion is then given to determine under what circumstances these Volterra models are suitable for the modelling of a particular nonlinear system.

Finally, the application of these techniques to data from some actual mechanical systems is performed. The mechanical systems concerned consist of two beams. The first is a simply supported beam, driven by an electromagnetic shaker, with pairs of repelling magnets placed at its tip. By varying the position of the magnets the nature and strength of the nonlinearity can be altered. The second is also a simply supported beam attached to a shaker. By loosening the attaching bolt to the shaker a rattling nonlinearity can be introduced into the system. Useful results, that are of practical interest to an engineer, were obtained from both the auto higher order spectral techniques and Volterra analysis.

Contents

Acknowledgements

My trio of supervisors, Paul White, Joe Hammond and Stuart Dyne are a continual source of guidance and encouragement. I would like to thank them for all of the help they have given.

In addition, I would like to thank the following; Justin Fackrell for solving many a problem, both by regular e-mail and at the various conferences we have attended, Peter Studzinski for providing expert help on data visualisation, Gill Jewell and Jon Hall for their computing support, Tim Leighton for his help in reading this thesis, and Maureen Strickland for sorting all manner of problems and always having an endless supply of biscuits.

Finally I would like to thank Mum and Dad for their support and encouragement and Uncle Antony for inspiring my interest in engineering.

This work was funded by the EPSRC.

List of Principle Symbols Used

A	arbitrary constant
$b^2(f_1,f_2)$	bicoherence function
$cum(x_1 \ldots x_n)$	n^{th} order cumulant operator
$C_{xx}(t)$	second order auto cumulant function
$C_{xxx}(t_1,t_2)$	third order auto cumulant function
$C_{xxxx}(t_1,t_2,t_3)$	fourth order auto cumulant function
$C_{xy}(t)$	second order cross cumulant function
$C_{xxy}(t_1,t_2)$	third order cross cumulant function
$C_{xxxy}(t_1,t_2,t_3)$	fourth order cross cumulant function
$e(n)$	mean square error signal
E	Young's modulus
$E[\,]$	expectation operator
$E_{xx}(f)$	deterministic spectrum
$E_{xxx}(f_1,f_2)$	deterministic bispectrum
$E_{xxxx}(f_1,f_2,f_3)$	deterministic trispectrum
f_r	resonant frequency
f_s	sample frequency
f_L	lower frequency of bandlimited signal
f_U	upper frequency of bandlimited signal
$\mathcal{F}[\,]$	Fourier operator
$G(f)$	arbitrary linear system
$\mathbf{h}$	impulse response vector
$h(t)$	impulse response
$h_1(t)$	linear impulse response
$h_2(t)$	quadratic polynomial impulse response
$h_3(t)$	cubic polynomial impulse response
$h_2(t_1,t_2)$	quadratic impulse response
$h_3(t_1,t_2,t_3)$	cubic impulse response
$H(f)$	transfer function
$H_1(f)$	linear frequency response

$H_2(f)$	quadratic polynomial frequency response
$H_3(f)$	cubic polynomial frequency response
$H_2(f_1,f_2)$	quadratic Volterra kernel
$H_{2(mod)}(f_1,f_2)$	modified quadratic Volterra kernel
$H_3(f_1,f_2,f_3)$	cubic Volterra kernel
$H_n[\,]$	n^{th} order Volterra operator
i(t)	impulse train
I(f)	Fourier transform of impulse train
I	second moment of area
$k^2(f_1,f_2,f_3)$	kurtosis function
k	number of blocks used in estimation
$K_1(f)$	linear Wiener kernel
$K_2(f_1,f_2)$	quadratic Wiener kernel
$K_3(f_1,f_2,f_3)$	cubic Wiener kernel
$K_{1(3)}(f)$	derived Wiener kernel
$K(\mathbf{v})$	characteristic function
l	length
L_i	residual spectra filters
L	number of taps in a digital filter
m	mass
M	segment size
N	data length
p_i	frequency of sinusoid
p	cross correlation vector
P(x)	probability density function
R	auto correlation matrix
$R_{XX}(t)$	auto correlation function
$R_{XXX}(t_1,t_2)$	auto bi-correlation function
$R_{XXXX}(t_1,t_2,t_3)$	auto tri-correlation function
$R_{XY}(t)$	cross correlation function
$R_{XXY}(t_1,t_2)$	cross bi-correlation function
$R_{XXXY}(t_1,t_2,t_3)$	cross tri-correlation function
$s^2(f_1,f_2)$	skewness function

$S_{xx}(f)$	power spectrum
$S_{xxx}(f_1,f_2)$	bispectrum
$S_{xxxx}(f_1,f_2,f_3)$	cumulant trispectrum
$S_{xy}(f)$	cross spectrum
$S_{xxy}(f_1,f_2)$	cross bispectrum
$S_{xxxx}(f_1,f_2,f_3)$	cumulant cross trispectrum
S_{ij}	cross spectrum of x_i and x_j
$T_{xxxx}(f_1,f_2,f_3)$	moment trispectrum
$T_{xxxy}(f_1,f_2,f_3)$	moment cross trispectrum
var	variance
$\dot{x}$	first derivative of x
$\ddot{x}$	second derivative of x
x(t)	input signal
X(f)	Fourier transform of x(t)
$X_{i,j}$	Fourier transform of process x_i conditioned with respect to x_j
y(t)	output signal
Y(f)	Fourier transform of y(t)
y_{lin}	linear component of y(t)
y_{quad}	quadratic component of y(t)
y_{cub}	cubic component of y(t)
*	conjugate
*	linear convolution
α	quadratic coefficient
α_h	constant relating to $h_2(t)$
β	cubic coefficient
$\gamma^2_{lin}(f)$	linear coherence
$\gamma^2_{quad}(f)$	quadratic coherence
δ	delta function
δ_k	Kronecka delta function
Δ	spacing between delta functions
$\varepsilon(t)$	error signal
θ_n	estimate of cross correlation vector

μ_r	r^{th} order broad band moment
ξ	damping coefficient
σ_x^2	variance of x
Φ_n	estimate of auto correlation matrix
Ψ	average output power
ω_0	fundamental angular frequency

Chapter 1

Introduction

1.1 General introduction

A fundamental approach to the classification of a random signal is to characterise the process by its joint probability density functions. The full set of probability density functions gives a complete description of any stochastic process. From the joint probability density function it is possible to form the joint moments of a process and from these the joint cumulants, further details of which can be found in Appendix A. Higher order spectra are defined as multi-dimensional Fourier transforms of the joint cumulant functions. This thesis is concerned with the use of these higher order spectra to study nonlinear systems.

A Gaussian process can be completely described by its second order joint probability density function, i.e. all joint cumulants of order higher than two vanish for a Gaussian process. This leads to one of the most useful properties of higher order spectra, that is, if the process is Gaussian all the spectra of order higher than two are equal to zero. If a Gaussian signal is operated on by a nonlinear system the resulting signal will be non-Gaussian. By studying this non-Gausian signal it is possible to obtain information about any non linearity in the system.

One of the most fundamental and useful tools in signal processing has been the estimation of the power spectrum. In power spectral estimation, the signal under consideration is processed in such a way that the distribution of power among its frequency components is estimated. The information present in the power spectrum is exactly that which is present in the autocorrelation function and is all that is needed for the complete description of a Gaussian signal. However, there are many practical situations where the power spectrum

cannot provide all the information and it is necessary to look beyond the power spectrum to extract information regarding deviations from Gaussianity.

Higher order spectra, which are defined in terms of the joint cumulants of a process, do contain such information. Particular cases of higher order spectra are the third order spectrum, also known as the bispectrum, which is defined as the Fourier transform of the third order joint cumulant, and the fourth order spectrum, or trispectrum, which is defined in terms of the Fourier transform of the fourth order joint cumulant. The power spectrum, bispectrum and trispectrum are just particular examples of the generalised concept of polyspectra which were introduced by Tukey [78] and Brillinger [9, 10, 11] in the early 1960's.

The vast majority of papers published concerning higher order statistics involve the bispectrum. The bispectrum is a decomposition of the third order moment or skewness of a process and as such is only able to analyse non-symmetric nonlinearities. In practical applications, many systems will contain symmetric nonlinear terms or, more likely, a mixture of skewed and symmetric nonlinearities. It is therefore necessary to consider not only the bispectrum but also the trispectrum, and to use the two as complementary tools in the analysis of a system. The aim of this work is to extend much of the work done on the bispectrum to the trispectrum and then use higher order spectra to analyse practical, nonlinear systems to produce information useful to an engineer.

In many applications, only a single signal is available, such as from an accelerometer mounted on a machine in a condition monitoring situation. In these cases auto higher order spectra must be used. The bispectra and trispectra can then give useful insight into the nonlinearities occurring in the system which are often related to the onset of faults in the machine. However, there are other applications where both an input signal and output signal are available, such as when a mechanical structure is being excited, and in these cases cross higher order spectra can be used to estimate system properties.

The main techniques used in this work are first, the bicoherence and tricoherence which are normalised versions of the auto bispectrum and auto trispectrum respectively. These are

used to detect quadratic and cubic phase coupling, that is, the interactions that can occur when a signal is passed through a system containing quadratic or cubic nonlinearities. The second technique involves cross higher order spectra and is based on the Volterra series. From cross higher order spectra, higher order transfer functions and impulse response functions can be calculated. By summing these and forming a partial Volterra series the response of a nonlinear system to a range of inputs can be approximated.

Being of lower order, the bispectrum is easier to compute than the trispectrum, which has many problems associated with its estimation. A fundamental difficulty when dealing with the trispectrum is simply in displaying it. The bispectrum is a function of two frequency variables, and so can easily be plotted using a mesh type plot with the bispectral magnitude rising out of the f_1, f_2 plane. The trispectrum is a function of three frequency variables and so requires four dimensional space to display it. Previous work on the trispectrum has always examined slices or planes through the f_1, f_2, f_3 frequency space which have been very difficult to interpret. It was therefore felt important that a method for displaying the trispectrum in its entirety should be developed in order to be able to easily distinguish the different frequency interactions.

Although higher order spectra have been actively researched for the last quarter of a century, very few useful practical applications have been found and the vast majority of the work published has concerned theoretical aspects. The aim of this work was not only to obtain a sound understanding of the theory of the bispectrum and trispectrum but to apply them to practical situations and obtain information that could be of use to an engineer studying the particular system.

To this end, two practical systems have been chosen to demonstrate higher order spectra. They are both mechanical structures and consist of beams driven by electromagnetic shakers onto which nonlinear forces can be exerted. One has a magnetic nonlinear restoring force, produced by pairs of repelling magnets placed at its tip. This is used because of the ease in which both the nature and magnitude of the nonlinearity can be controlled. The other has a rattling type nonlinearity which can be introduced into the

system so as to be akin to a fault occurring in a machine. Both auto and cross higher order spectral techniques are used to try to detect and model the nonlinearities.

1.2 Historical perspective and scope of the thesis

A brief summary of the contents of each of the chapters is now given, together with references to any similar work published by other authors, emphasising the contributions made here.

In chapter two, the concept of higher order spectra is introduced, first from an intuitive frequency domain point of view, and then more mathematically in the time domain. A number of authors, Nikias et al [58, 59, 60], Mendel [52], Priestley [65] and Subba Rao [73] have written review articles or books on higher order spectra and their approach is typically a mathematical one. The aim here was to get a 'feeling' for the frequency interactions that can occur in higher order spectra before using the more mathematical approach of cumulant and moment functions and this is done with simple examples using sine waves. Many papers have been published on the bispectrum, particularly on its theoretical aspects but few authors have studied the trispectrum and no one, to the author's knowledge, has displayed the trispectrum in its entirety. Dalle Molle [25, 26, 27] is the most widely quoted author for the trispectrum but his work is very theoretical; other authors include, Chandran et al [12, 13, 14], Kravtchenko-Berejnoi et al [47], Lutes and Chen [49] and Walden and Williams [79]. Some of the practical issues concerning the calculation of higher order spectra are then examined, such as the methods of estimation, windowing, symmetry, and aliasing. There are two approaches to higher order spectra estimation: parametric and non parametric. Parametric methods are based on assuming the signal is the output of a model and estimating the parameters of the model, e.g. AR, MA. These methods are not studied in this work but more information can be found in [66, 59]. The two main non parametric methods of estimation of higher order spectra are the direct and indirect methods. These are well understood and detailed in many places, such as Nikias and Petropulu [59]. The symmetries of the bispectrum have now been well documented by, for example, Subba Rao [73]. However the symmetries of the trispectrum

are more complicated. Dalle Molle [27] and Williams [80] have both produced papers suggesting principal domains for the trispectrum and recently this has been extended to a general procedure for the derivation of the principal domain of an n^{th} order spectra by Chandran and Elgar [15]. Components of higher order spectra outside the principal domain have been the basis of a well established test for aliasing by Hinich [39]. Recent work by Parsons and Williams [61], Le Roux et al [48], and Stogioglou and McLaughlin [72] have used this to show that a continuous stationary signal sampled at its Nyquist rate cannot be modelled by a discrete process based on passing independent, identically distributed (IID) noise through a linear filter. The novel contributions in this chapter are in extending much of the work on the bispectrum to the trispectrum, and especially in the display of the trispectrum.

The variance of estimates of the bispectrum and trispectrum are dependent on the second order statistics of the signal and so it is common practice to normalise the bispectrum and trispectrum to remove these effects. Possible methods are discussed in chapter three. These include the well known skewness function, first introduced by Hinich [40] and the bicoherence function of Kim and Powers [45]. These are extended to their fourth order equivalents, the kurtosis function and the tricoherence. Examples are given of the normalisation functions particularly concentrating on the bicoherence and tricoherence and drawing on the work of Fackrell et al [29, 33, 34] who have studied in detail, some of the problems associated with the practical use of the bicoherence. Only a small amount of literature has been published on the tricoherence, notably on its statistics by Chandran et al [12] and so effort is made here to concentrate on the tricoherence, using examples wherever possible. These traditional methods of normalisation sometimes fail when narrow band signals or signals with sharp resonant peaks are considered. Williams [80] has previously worked on the higher order spectra of band limited signals. A new method is proposed, based on a pre-whitening technique which has similarities with work recently presented on phase only spectra (POS) by Lyons et al [50] and is demonstrated with the use of an amplitude modulated (AM) signal. A brief mention is given to periodic signals, extending Fackrell's [32] work on periodic signals in the bispectrum, or the 'bed of nails', to the fourth order. Finally, a detailed example of a cubic nonlinear dynamic system, the Duffing oscillator, is given. Such a system has previously been modelled using the

trispectrum by Lutes and Chen [49]. This chapter, as in chapter two, mainly deals with extending concepts from the bispectrum to the trispectrum. The main novel contribution concerns the work with narrow band signals and the development of the pre-whitening technique as a normalisation method for the trispectrum.

Chapter four expands the work on auto higher order spectra to study cross higher order spectra and the Volterra series. A number of methods are proposed, all of which use higher order spectra, and are based on determining the first three Volterra kernels or parts of them. Wiener first noticed how the Volterra series could analyse nonlinear systems using higher order spectra as detailed in Schetzen [70]. Tick [76] then demonstrated how, in the frequency domain, the linear and quadratic Volterra kernels could be directly estimated in terms of the input and output characteristics of a system when the input was assumed to be a zero mean, Gaussian signal. This was extended to the cubic case by Hong et al [41] but based on a time domain approach. Before studying the full Volterra series model a similar series based on linear filters acting on a polynomial input is examined. This is shown to be akin to just examining the main diagonal of the Volterra kernels. Ralston and Zoubir [67] have recently used this type of series, which they have termed a Hammerstein series, to study engine knock. This type of model may also be viewed as a multiple input, single output system. Much work has been done on the concepts of residual spectra and partial and multiple coherence for this type of system notably by Bendat et al [3], Clarkson and Hammond [18] and Fitzpatrick et al [35] who have compared this technique to higher order spectra for modelling such applications as squeeze film dynamics [28]. It is shown how similar results can be obtained using both higher order spectra and residual spectra and these two techniques are compared and contrasted. The full Volterra model, based on the assumption of a Gaussian input, is then described in detail both for the quadratic and cubic cases. Many authors have used the simple quadratic Volterra model to analyse skewed type nonlinearities, notably Bondon [5], Zoubir [83], Hinich [37], and Perrochaud [62], but little use has been made of the cubic model. As with auto higher order spectra, problems occur in the estimation process due to the variance of the cross bispectrum and trispectrum containing second order effects. An alternative technique based on successively subtracting off lower order components is proposed, which is shown to minimise the variance of the quadratic kernel. Examples are given for both the quadratic and cubic cases. In this work a

block estimation approach is used, however, a possible alternative philosophy, as detailed in [51], is based on adaptive methods where the estimate is updated for every sample. The main original contribution in this chapter is the frequency domain technique used for estimating the Volterra kernels.

In chapter five, before applying the Volterra series to experimental results, a number of possible pitfalls are examined. These include the convergence, causality, and optimality of the Volterra models developed in the previous chapter, together with the difficulties that can occur when the input cannot be assumed to be Gaussian. Boyd and Chua [7] have given detailed conditions for the convergence of the Volterra series, as has Schetzen [70] who likened it to a Taylor series with memory. Wright [81] has clarified their conditions and demonstrated them with examples. In this chapter some more simple examples of converging and diverging Volterra series are given. During the estimation of the Volterra kernels in chapter four, the kernels of a related series, the Wiener series [1, 2, 70], are estimated. The relationship between Wiener and Volterra kernels is detailed in Schetzen [70] but is also described here as it is closely linked with the way the Volterra kernels are estimated. Throughout this work, it is always required that the input to the system is Gaussian. In many cases this assumption may be invalid, and so a method, based on the work of Powers et al [16, 43, 44, 55, 63, 77] is presented for the quadratic Volterra model that removes this restriction. Different techniques are described, together with their advantages and disadvantages, and these are categorised together with their respective authors. The frequency domain technique developed in the previous chapter to estimate the Volterra kernels is compared with the optimum least squares time domain technique. It is shown that the new technique, although not optimum, produces much better results than traditional methods. Finally the topic of the causality of the estimated Volterra kernels is briefly mentioned.

In the final chapter, both auto and cross higher order spectra are applied to some real mechanical systems. Over the past two decades higher order statistics have been applied to may different types of physical phenomena. Applications have been found in fields as diverse as speech processing [31], chaos [75], underwater acoustics [68], medical engineering [71], share prices [39] and condition monitoring. Condition monitoring of

mechanical systems has been one of the largest areas of application of higher order statistics. Broad band higher order moments, such as kurtosis, are well established as a condition monitoring tool, as described by Braun [8], so it is likely that higher order statistics will also be able to offer useful information. Much of the recent work has concerned the condition monitoring of rotating machines, notably by Fackrell et al [30], Nandi et al [56, 57], Howard [42], Sato et al [69], and Zhou et al [82] and has tended to concentrate on the bispectrum. In this work, both the auto bispectrum and trispectrum are used to detect the nonlinearities in two systems, first for a beam excited by an electromagnetic shaker with pairs of repelling magnets at its tip producing a nonlinear restoring force [21], and second for a beam loosely attached to an electromagnetic shaker so as to produce a rattling type nonlinearity. The auto bispectrum of this last system has previously been studied by Fackrell et al [30]. Cross higher order spectra are then used to identify the systems by calculating the cubic Volterra model of the two beams. The novel contribution in this chapter is in applying the above methods to produce useful results from real systems.

Chapter 2

Auto higher order spectra: the bispectrum and trispectrum

2.1 Introduction

This chapter begins by examining the probability density function and moments of a signal. The concepts of higher order spectra (HOS) are then introduced, initially from a frequency domain point of view, for both stochastic and deterministic processes. However, the majority of the work in this thesis assumes that the signal is zero mean, stationary, and random. The importance of these assumptions are discussed in detail for the second order case in Appendix B.

To help introduce higher order spectra, a simple example using deterministic signals based on sine waves is given. This example helps to show some of the frequency interactions that can occur in higher order spectra and the methods used to display them. Next, a more mathematical time domain approach is used, to present higher order statistics, deriving them from joint cumulant and moment functions.

Some of the practical issues concerning the calculation of higher order spectra are examined, including the different methods of estimation, either in the frequency or time domain, and the effects of windowing. The symmetrical properties of the bispectrum and trispectrum are discussed and their principal domains defined. Finally, the implications of components of higher order spectra appearing in different regions of the bispectrum and trispectrum are considered with particular attention to aliasing.

2.2 Probability density functions and moments

If x(t) is a stationary, random, signal, the r^{th} moment of x(t), denoted μ_r, is defined as,

$$\mu_r = E[x^r(t)] \qquad 2.1$$

where E denotes the expectation operator. Note that $\mu_1 = E[x(t)] = \mu_x$, the mean of x(t).

Higher order moments [53] are usually calculated as central moments about the mean. That is:

$$\mu_r = E[(x(t)-\mu_x)^r] \qquad 2.2$$

The second central moment is the variance of a signal,

$$var[x(t)] = \mu_2 = E[(x(t)-\mu_x)^2] = \sigma_x^2 \qquad 2.3$$

This gives a measure of the spread of a signal about the mean. The probability density function of a signal with a Gaussian or normal distribution (see figure 2.1) is completely described by its mean and variance. Higher order moments are often used to describe the properties of more complex signals.

The third moment about the mean, μ_3, is sometimes called skewness and is a measure of asymmetry of the probability density function. For symmetric distributions $\mu_3=0$. A probability density function similar to that shown by the solid line in figure 2.2 is said to be skewed to the left and has a negative skewness, whilst one similar to that shown by the dotted line is said to be skewed to the right and has a positive skewness. The ratio μ_3/σ_x^3, which is dimensionless, is called the coefficient of skewness and gives a measure of the degree to which a distribution is skewed.

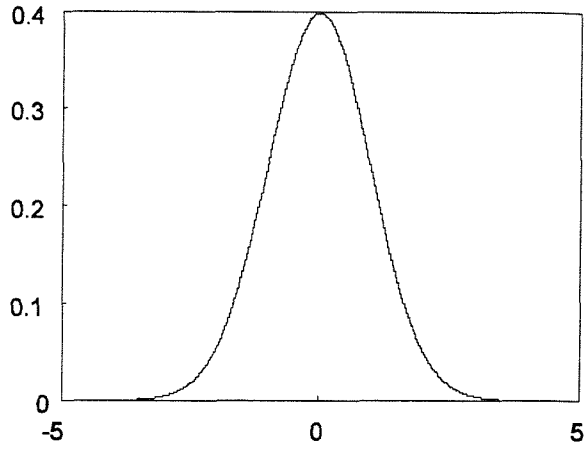

Figure 2.1: Gaussian pdf, $\mu_3=0$, $\mu_4=3$

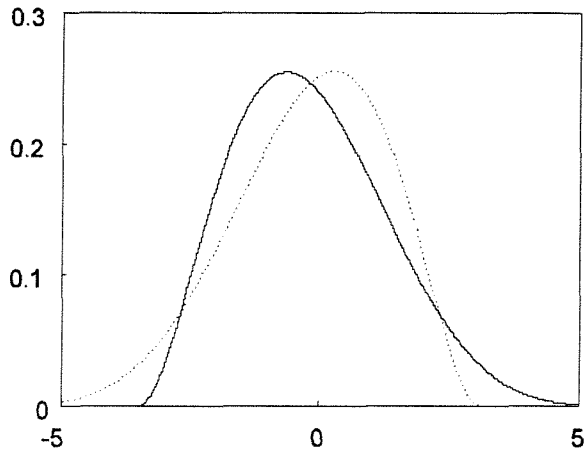

Figure 2.2: Negative skewed pdf, $\mu_3 < 0$, (solid line) and positive skewed pdf, $\mu_3 > 0$, (dotted line)

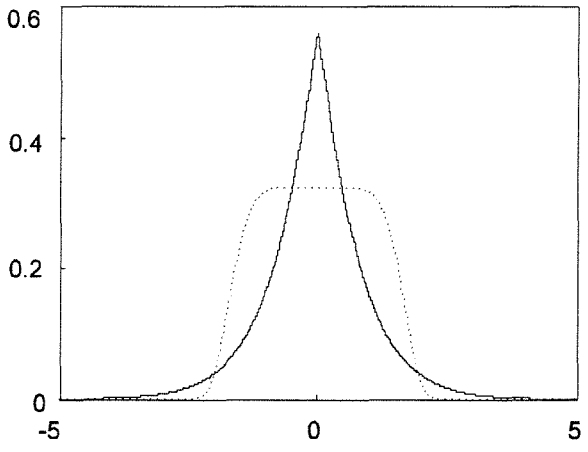

Figure 2.3: Leptokurtic pdf, $\mu_4 < 3$, (solid line) and platykurtic pdf, $\mu_4 > 3$, (dotted line)

The fourth moment about the mean, μ_4, is used as a measure of kurtosis which is related to the degree of flatness of a distribution near its centre. The ratio μ_4/σ_x^4 is known as the coefficient of kurtosis. A Gaussian signal has a kurtosis of three. Values of kurtosis greater than three indicate that a probability density function is more peaky (as shown by the solid line in figure 2.3) around its centre than the density function of a normal distribution and are known as leptokurtic [6]. Kurtosis values of less than three indicate that the probability density function is flat around its centre (dashed line, figure 2.3) as compared with a normal distribution and are known as platykurtic.

Within signal processing, higher order moments can give a basic description of the properties of a signal. In particular the fourth order moment, or kurtosis of a signal, has successfully been used in the field of condition monitoring [8], for example, to detect bearing wear in a rotating machine.

2.3 An introduction to auto higher order spectra - the frequency domain

The power spectrum, bispectrum and trispectrum are just particular examples of the generalised concept of polyspectra [78]. Just as the power spectrum is able to give a decomposition of power over frequency, it is possible to use higher order spectra to obtain a decomposition of skewness and kurtosis over frequency and so obtain more information about the higher order statistics of a signal. The second order polyspectra is the conventional power spectrum and so this section begins with a brief discussion of the power spectrum.

2.3.1 The power spectrum

The power spectrum is the main tool of signal analysis and a huge body of literature has been published concerning its use and properties, for example [64]. It is the most commonly used of the polyspectra for, being of the lowest order, it is the simplest to calculate and easiest to interpret. The power spectrum is concerned with the second order

statistics of a signal and will now be defined both in the context of deterministic and stochastic processes.

The energy in a signal is:

$$\overline{x^2(t)} = \int_{-\infty}^{\infty} x^2(t)dt \qquad 2.4$$

Substituting $x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}df$ into equation 2.4 gives,

$$\overline{x^2(t)} = \int\int\int_{-\infty}^{\infty} X(f_1)X(f_2)e^{j2\pi t(f_1+f_2)}dtdf_1df_2 \qquad 2.5$$

Integrating equation 2.5 with respect to t and using the shifting property of the δ function results in,

$$\begin{aligned}\overline{x^2(t)} &= \int\int_{-\infty}^{\infty} X(f_1)X(f_2)\delta(f_1+f_2)df_1df_2 \\ &= \int_{-\infty}^{\infty} X(f_1)X(-f_1)df_1 \qquad 2.6\end{aligned}$$

From this the energy spectrum can be defined as,

$$E_{XX}(f) = X(f)X(-f) \qquad 2.7$$

For a stationary stochastic process it is possible to use a similar method, which is detailed in Appendix B, to obtain the power spectrum which is defined as:

$$S_{XX}(f_1,f_2) = E[X(f_1)X(-f_2)] \qquad 2.8$$

This simplified notation, which absorbs the time interval into the definition, is explained in detail in Appendix B. For a stationary process it can be shown [64] that $S_{XX}(f_1,f_2)$ is equal

to zero except along $f_1 = -f_2$. This results in the following, more usual, definition for the power spectrum of a stationary stochastic process:

$$S_{XX}(f) = E[X(f)X^*(f)] \qquad 2.9$$

where * denotes the complex conjugate. Note that this is a partial description of the power spectrum; the full definition can be seen in Appendix B. The power spectrum treats each frequency component as independent from all others and measures the power of the signal at each frequency. It is a real quantity and contains no phase information and as such is said to be phase blind.

2.3.2 The auto bispectrum

Rather than decomposing the energy of a signal to produce the energy spectrum, it is possible to conduct similar analysis on a cubed signal,

$$\overline{x^3(t)} = \int_{-\infty}^{\infty} x^3(t)dt \qquad 2.10$$

Substituting $x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}df$ into equation 2.10 gives,

$$\begin{aligned}\overline{x^3(t)} &= \int\int\int\int_{-\infty}^{\infty} X(f_1)X(f_2)X(f_3)e^{j2\pi t(f_1+f_2+f_3)}dtdf_1df_2df_3 \\ &= \int\int\int_{-\infty}^{\infty} X(f_1)X(f_2)X(f_3)\delta(f_1+f_2+f_3)df_1df_2df_3 \\ &= \int\int_{-\infty}^{\infty} X(f_1)X(f_2)X(-f_1-f_2)df_1df_2 \qquad 2.11\end{aligned}$$

From this, the bispectrum of a deterministic signal can be defined as,

$$E_{XXX}(f_1,f_2) = X(f_1)X(f_2)X(-f_1-f_2) \qquad 2.12$$

For a stochastic process, using the same method as for the power spectrum, the bispectrum is defined as:

$$S_{xxx}(f_1, f_2, f_3) = E[X(f_1)X(f_2)X(f_3)] \qquad 2.13$$

If the process is stationary, it has been shown [80] that $S_{XXX}(f_1,f_2,f_3)$ is equal to zero except on the plane $f_3 = -f_1-f_2$. Therefore, the bispectrum of a stationary stochastic process is defined as:

$$S_{xxx}(f_1, f_2) = E[X(f_1)X(f_2)X^*(f_1 + f_2)] \qquad 2.14$$

In the same way that the power spectrum is concerned with the power of a signal, or second order moment, the bispectrum is concerned with the skewness, or third order moment. The bispectrum is a function of two frequency variables, f_1 and f_2, and while the power spectrum considers each frequency component independently, the bispectrum analyses the frequency interactions between the frequency components f_1, f_2, and f_1+f_2. It is a complex quantity containing both real and imaginary parts. However, throughout this work only the magnitude of the bispectrum is considered.

Two simple examples, using sine waves, are now given demonstrating some of the possible frequency interactions that can occur in the bispectrum. Sine waves are used as an example because they produce easily understood results despite the fact that they do not conform to the assumption of being stationary random signals.

Consider a complex sine wave of frequency p_1. A complex sine wave is used in order to suppress unwanted cross terms between the positive and negative frequency components.

$$x(t) = e^{j2\pi p_1 t} \qquad 2.15$$

This has a Fourier transform,

$$X(f) = \delta(f-p_1), \qquad 2.16$$

where δ represents the Dirac delta function. This is shown diagrammatically in figure 2.4. If X(f) is substituted from equation 2.16 into equation 2.12, the bispectrum is equal to:

$$E_{xxx}(f_1,f_2) = \delta(f_1 - p_1)\delta(f_2 - p_1)\delta(f_1 + f_2 - p_1) \qquad 2.17$$

This contains the triple product $\delta(f_1-p_1)\delta(f_2-p_1)\delta(f_1+f_2-p_1)$. There will only be a non-zero point in the bispectrum when all three terms in the above product are non-zero. Plotting the three terms in the (f_1,f_2) plane leads to the three lines, $f_1 = p_1$, $f_2 = p_1$ and $f_1+f_2 = p_1$, as shown in figure 2.5. For $p_1 \neq 0$ there is no point of intersection of all three lines and hence the bispectrum of a complex sine wave is zero.

Next consider a signal consisting of two complex sine waves of frequency p_1 and p_2. The Fourier transform of this signal is,

$$X(f) = \delta(f-p_1) + \delta(f-p_2) \qquad 2.18$$

This is shown in figure 2.6. The deterministic bispectrum is now equal to,

$$E_{xxx}(f_1,f_2) = \{\delta(f_1 - p_1) + \delta(f_1 - p_2)\}\{\delta(f_2 - p_1) + \delta(f_2 - p_2)\} \{\delta(f_1 + f_2 - p_1) + \delta(f_1 + f_2 - p_2)\} \qquad 2.19$$

This can be shown to consist of eight terms, each of which is a triple product. If these are plotted in the (f_1,f_2) plane they appear as the six possible lines $f_1 = p_1$, $f_2 = p_1$, $f_1 = p_2$, $f_2 = p_2$, $f_1 + f_2 = p_1$, and $f_1 + f_2 = p_2$ as shown in figure 2.7. There will be an intersection of the three terms if $p_2 = 2p_1$. The intersection will then occur at (p_1,p_1) as shown by the dot in figure 2.7.

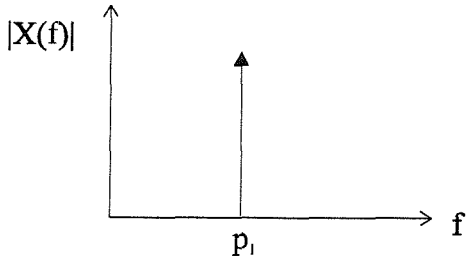

Figure 2.4: Fourier transform of sine wave

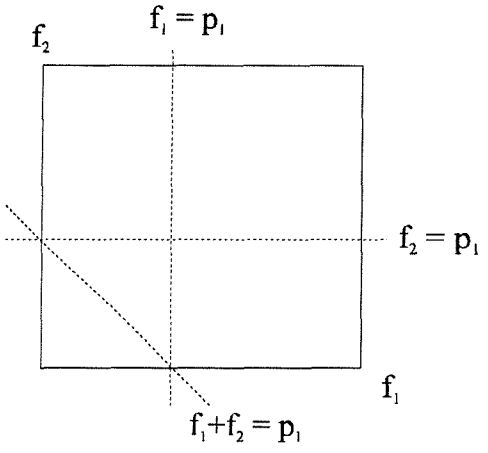

Figure 2.5: Bispectrum of sine wave

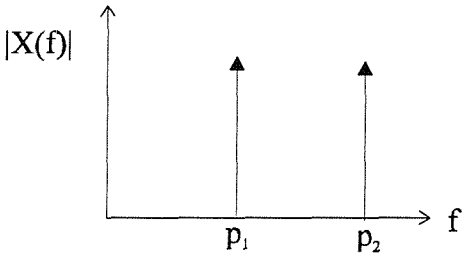

Figure 2.6: Fourier transform of two sine waves ($p_2 = 2p_1$)

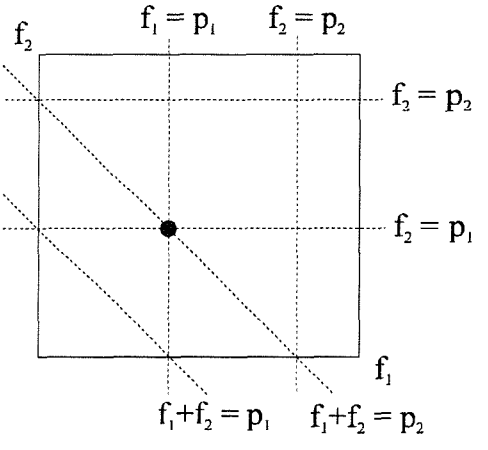

Figure 2.7: Bispectrum of two sine waves ($p_2 = 2p_1$)

An example of the bispectrum of two sine waves of frequencies 50 Hz and 100 Hz is shown in figure 2.8, where it can be clearly seen that there is a peak at (50,50) Hz. As the bispectrum is a function of two frequency variables it is easy to plot it as a three dimensional function with the bispectral content rising out of the (f_1,f_2) plane. Here a 'mesh' type plot is used to show the magnitude of the bispectrum as a three dimensional surface. This is predominantly the method used to display bispectra in this work, although simple 'contour' maps occasionally allow one to interpret the fine detail with more accuracy.

2.3.3 The auto trispectrum

In the previous section, the bispectrum was defined as a decomposition of the average of a signal cubed and as such is concerned with the skewness of a signal. In many of the problems that are considered in this work it will be necessary to consider both skewed and symmetric signals. The trispectrum can be defined as a decomposition of the average of a signal to the fourth power and as such is concerned with the kurtosis of a signal. Using a similar analysis as for the bispectrum it is possible to define the trispectrum for a deterministic process as,

$$E_{xxxx}(f_1,f_2,f_3) = X(f_1)X(f_2)X(f_3)X(-f_1-f_2-f_3) \qquad 2.20$$

and the trispectrum for a stochastic process as[1],

$$T_{xxxx}(f_1,f_2,f_3) = E[X(f_1)X(f_2)X(f_3)X^*(f_1+f_2+f_3)] \qquad 2.21$$

As with the bispectrum a simple example using complex sine waves will now be given in order to show the frequency interactions that can occur in the trispectrum and the methods used in this work to display it. It is easy to show that the trispectrum of a single complex sine wave is zero. This can be done as for the bispectrum but instead of considering lines in

[1] Note, the change in terminology from $S_{xxxx}(f_1,f_2,f_3)$ to $T_{xxxx}(f_1,f_2,f_3)$ is to denote the difference between the cumulant trispectrum and moment trispectrum. This is explained in section 2.4.3.

the (f_1,f_2) plane it is necessary to consider planes in the (f_1,f_2,f_3) space. If the trispectrum of two complex sine waves of frequency p_1 and p_2 are calculated, by considering the interaction of planes in the (f_1,f_2,f_3) space, it can be shown that there will be a peak in the trispectrum at (p_1,p_1,p_1) if $p_2 = 3p_1$. Note there will also be a peak if $p_1 = -p_2$.

An example of the trispectrum of two sine waves of frequency 50 Hz and 150 Hz is shown in figure 2.9. A major problem when calculating the trispectrum is that of deciding how to display it. Whereas, the bispectrum is a function of two frequency variables and can easily be plotted using three dimensional space, the trispectrum is a function of three frequency variables and so requires four dimensional space to display it. All previous work on the trispectrum has used various slices through the (f_1,f_2,f_3) space to display the important features. However, it was felt in this work that it was important to display it in its entirety in order to get a 'feel' for the various frequency interactions.

The method chosen to do this uses the Application Visualisation System (AVS) software on a Silicon Graphics machine. For each (f_1,f_2,f_3) in the trispectral space, a sphere is drawn. The size and colour of the sphere represents the magnitude at that point: large red spheres represent points of large magnitude, reducing across the colour spectrum to small blue spheres which represent points of smallest magnitude. Very small values are not drawn, otherwise the (f_1,f_2,f_3) space would be covered by small, blue spheres representing very small magnitude points. This method results in a cube of spheres of varying sizes and colours representing the frequency interactions. It is possible to rotate the image on the screen and so precisely determine the points of interaction. When the image is printed on to paper, a four dimensional image has to be reduced to two dimensions and so inevitably some clarity is lost. However, it still possible, at a glance, to understand where the dominant points of interest are. More details of the software used, together with some demonstrations of simple signals, are given in Appendix C.

For the case of the trispectrum of two complex sine waves it is clear that there is a point of interaction at (50,50,50) Hz, of magnitude 1, denoted by the red spheres, together with some points of magnitude 0.5 at other frequency triplets. It will be shown later in this chapter that these points are caused by the symmetries of the trispectrum and are reflections of the (50,50,50) Hz interaction.

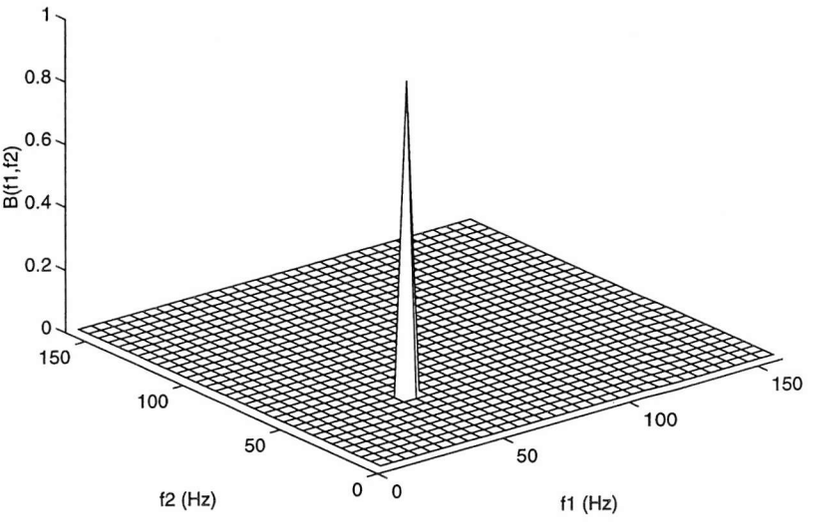

Figure 2.8: The bispectrum of two sine waves of frequency 50Hz and 100Hz

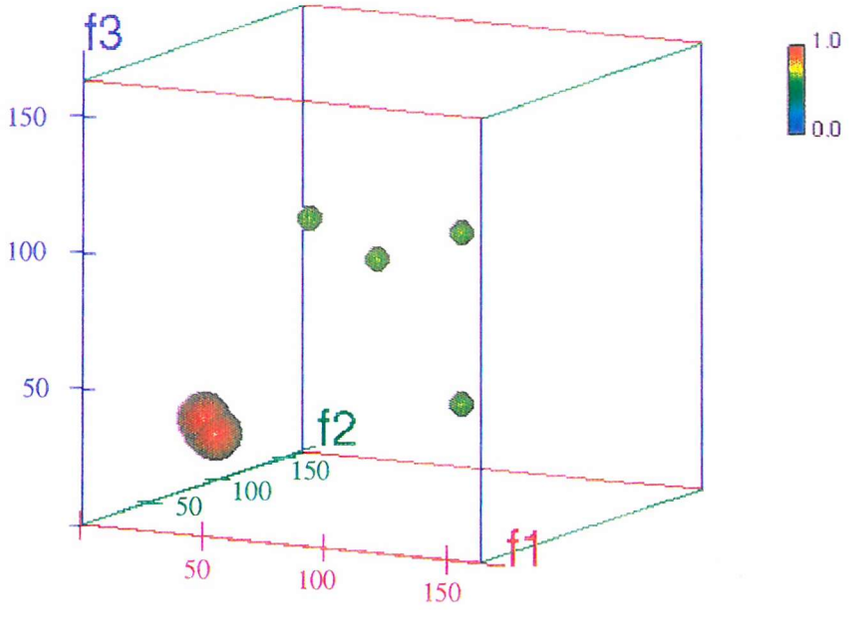

Figure 2.9: The trispectrum of two sine waves of frequency 50Hz and 150Hz

2.4 An introduction to auto higher order spectra - the time domain

The previous section gave an intuitive introduction to higher order spectra starting from the frequency domain and illustrating some of the frequency interactions that can create structure in bispectra and trispectra. In this section a more rigorous, mathematical approach is discussed which arises from considering joint moment and cumulant functions in the time domain.

2.4.1 Moment functions

In section 2.1 the broad band moments of a signal were considered. Here this is extended to look at joint moment functions which describe the moments of a signal at different lags or delays. These moment functions can be used to describe the higher order properties of a signal. The n^{th} order moment function of a stationary random signal, x(t) is defined as,

$$R_{X...X}(\tau_1, \tau_2, ..., \tau_N) = E[x(t)x(t+\tau_1)x(t+\tau_2)...x(t+\tau_N)] \qquad 2.22$$

The second order moment function of a stationary random signal (see Appendix B), $R_{XX}(\tau)$, is the well known autocorrelation function,

$$R_{XX}(\tau) = E[x(t)x(t+\tau)] \qquad 2.23$$

The third and fourth order moment functions, termed the autobicorrelation and autotricorrelation, for stationary signals, are defined as,

$$R_{XXX}(\tau_1, \tau_2) = E[x(t)x(t+\tau_1)x(t+\tau_2)] \qquad 2.24$$

$$R_{XXXX}(\tau_1, \tau_2, \tau_3) = E[x(t)x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)] \qquad 2.25$$

For a zero mean Gaussian process all odd order functions are identically zero. The zero lag component of the second moment or autocorrelation function, $R_{XX}(0)$, is the signal variance which can also be found by integrating the power spectrum over all frequencies. In a

similar way, the zero lag component of the third moment function, $R_{xxx}(0,0)$, is μ_3 and the zero lag component of the fourth moment function, $R_{xxxx}(0,0,0)$, is μ_4. These can also be found by integrating the bispectrum across all frequency pairs and integrating the trispectrum across all frequency triplets, respectively.

2.4.2 Cumulant functions

The N^{th} order cumulant function of a stationary random signal x(t) is defined (for N = 3, 4) as,

$$C_{X...X}(\tau_1,\tau_2,...,\tau_N) = R_{X...X}(\tau_1,\tau_2,...,\tau_N) - R^{G}_{X...X}(\tau_1,\tau_2,...,\tau_N) \qquad 2.26$$

where $R_{X...X}(\tau_1,\tau_2,...,\tau_N)$ is the N^{th} order moment function of x(t), and $R^{G}_{X...X}(\tau_1,\tau_2,...,\tau_N)$ is the N^{th} order moment function of an equivalent Gaussian signal that has the same mean value and autocorrelation function of x(t). For a real stationary random process with zero mean, the following relationships exist between the cumulant functions and moment functions:

$$C_{xx}(\tau) = R_{xx}(\tau) \qquad 2.27$$

$$C_{xxx}(\tau_1,\tau_2) = R_{xxx}(\tau_1,\tau_2) \qquad 2.28$$

The second and third order cumulant functions are identical to the second and third order moment functions. However, to generate the fourth order cumulant, $C_{xxxx}(\tau_1,\tau_2,\tau_3)$, it is necessary to use both the second and fourth order moments.

$$\begin{aligned} C_{xxxx}(\tau_1,\tau_2,\tau_3) = R_{xxxx}(\tau_1,\tau_2,\tau_3) &- R_{xx}(\tau_1).R_{xx}(\tau_3-\tau_2) \\ &- R_{xx}(\tau_2).R_{xx}(\tau_3-\tau_1) \\ &- R_{xx}(\tau_2).R_{xx}(\tau_2-\tau_1) \end{aligned} \qquad 2.29$$

A more detailed description of the properties of moment and cumulant functions and their associated generating functions is given in Appendix A.

2.4.3 The relationship between cumulants and higher order spectra

It is well known that Fourier transforming the autocorrelation function results in the power spectrum,

$$S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j2\pi f\tau} d\tau = \int_{-\infty}^{\infty} C_{xx}(\tau) e^{-j2\pi f\tau} d\tau \qquad 2.30$$

Similar results can be obtained for the bispectrum, $S_{XXX}(f_1,f_2)$, and cumulant trispectrum, $S_{XXXX}(f_1,f_2,f_3)$ by taking the double Fourier transform of the second order cumulant function, and the triple Fourier transform of the third order cumulant function respectively,

$$\begin{aligned} S_{XXX}(f_1,f_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{XXX}(\tau_1,\tau_2) e^{-j2\pi f_1\tau_1} e^{-j2\pi f_2\tau_2} d\tau_1 d\tau_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C_{XXX}(\tau_1,\tau_2) e^{-j2\pi f_1\tau_1} e^{-j2\pi f_2\tau_2} d\tau_1 d\tau_2 \end{aligned} \qquad 2.31$$

$$S_{XXXX}(f_1,f_2,f_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C_{XXXX}(\tau_1,\tau_2,\tau_3) e^{-j2\pi f_1\tau_1} e^{-j2\pi f_2\tau_2} e^{-j2\pi f_3\tau_3} d\tau_1 d\tau_2 d\tau_3 \qquad 2.32$$

$$T_{XXXX}(f_1,f_2,f_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{XXXX}(\tau_1,\tau_2,\tau_3) e^{-j2\pi f_1\tau_1} e^{-j2\pi f_2\tau_2} e^{-j2\pi f_3\tau_3} d\tau_1 d\tau_2 d\tau_3 \qquad 2.33$$

The difference between the cumulant and moment trispectrum is that the moment trispectrum is non-zero along three planes even for a Gaussian process. The relationship between cumulants and higher order spectra is shown in figure 2.10.

If x(t) is Gaussian then $C_{X\ldots X}(\tau_1,\tau_2,\ldots,\tau_N)$ is zero for $N > 2$. This result leads directly to one of the most useful properties of higher order spectra. That is, all higher order spectra of order greater than two vanish when x(t) is a Gaussian process. Hence the bispectrum and cumulant trispectrum can be regarded as a measure of the departure of the process

from Gaussianity and many tests, for example [26, 40, 74], have been developed to this effect.

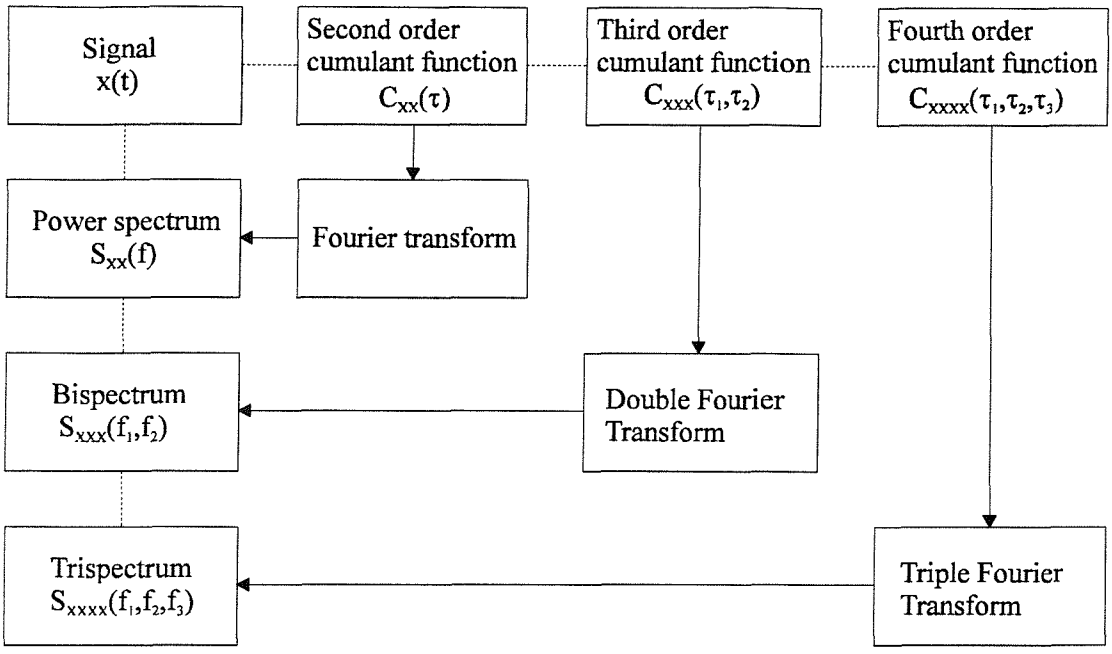

Figure 2.10: The relationship between cumulants and higher order spectra.

As, for a Gaussian process, the second and third order moment functions are identical to the second and third order cumulant functions, no distinction can be drawn between the moment power spectrum and bispectrum and the cumulant power spectrum and bispectrum. However, the fourth order cumulant function is equal to the fourth order moment function minus three terms which consist of second order components. Throughout this work higher order spectra generated from cumulant functions are studied, but estimating a trispectrum using equation 2.21, which is the most commonly used form of estimation, results in a moment trispectrum. It is therefore necessary to separately calculate the three second order terms and subtract them from the moment trispectrum, to form the true cumulant trispectrum.

2.5 Estimation of higher order spectra

The techniques used to estimate higher order spectra are similar to those used in power spectrum estimation. Essentially there are two commonly used approaches:

The indirect method:-	Estimating the cumulant function and then taking a Fourier transform.
The direct method:-	Calculating raw spectra from segments of the signal and then averaging across segments in the frequency domain.

Herein, the direct method is predominantly used, although higher order spectra calculated with the indirect method are sometimes estimated to verify results and so an explanation of both methods is included for completeness.

2.5.1 The indirect method

To calculate the n^{th} order cumulant function the time series is split into k blocks, each of length M points. For each block $R_{X...X}(\tau_1,...,\tau_N)$ is calculated, using equation 2.34.

$$R_{X...X}(m_1,\ldots,m_N) = \frac{1}{m+1}\sum_{n=0}^{M} x(n)x(n+m_1)\ldots x(n+m_{N-1}) \qquad 2.34$$

For the fourth order moment (the trispectrum) and above, $C_{X...X}(\tau_1,...,\tau_N)$ is calculated from $R_{X...X}(\tau_1,...,\tau_N)$ by subtracting off lower order terms, as detailed previously. The k blocks are averaged to give the n^{th} order cumulant function. This method of averaging small blocks is used, as opposed to calculating one large block, as it requires much less computation. However as less averaging is used in the estimation process, it will give results with a larger variance. Finally, the higher order spectrum is calculated by taking the n^{th} order Fourier transform of the cumulant function. As is the case of conventional power spectrum estimation, better estimates can sometimes be obtained using suitable windowing functions. These will be discussed in section 2.5.3.

2.5.2 The direct method

The direct method is based on the Welch periodogram technique. The data is divided into a series of M overlapping blocks. For each block, the FFT is calculated and the product of the spectral coefficients formed to produce the 'raw' higher order spectra. These 'raw' higher order spectra are then averaged to lower the variance of the periodogram estimates.

In order to ensure statistical confidence it is necessary to take care when choosing the block length compared to the data length, N. A larger block size will have a less coarse grid, resulting in a reduced bias, but a higher variance [73]. Dalle Molle and Hinich in [25] state that the block length should be the $(n-1)^{th}$ root of the sample size when working with the n^{th} order spectrum. In the case of the trispectrum when working with a FFT size of 64 the data length should be 64^3 or 262144 samples long. This in itself can cause problems as the data should be stationary over the complete length.

The bispectral estimate obtained from this method can be shown [59] to have a variance of:

$$\mathrm{var}(S_{xxx}(f_1,f_2)) = \frac{1}{M}[1+\delta_k(f_1-f_2)]\, S_{xx}(f_1)S_{xx}(f_2)S_{xx}(f_1+f_2) \qquad 2.35$$

where $\delta_k(0) = 1$ and $\delta_k(f) = 0$ for non-zero f. It can be seen that the variance of the estimate of the bispectrum depends not only on the number of segments but also on the triple product of power spectral terms, $S_{XX}(f_1)S_{XX}(f_2)S_{XX}(f_1+f_2)$. Hence, the variance of the bispectral estimate is heavily dependent on the power spectrum. This fact will continually cause problems throughout this work particularly in the estimation and use of auto higher order spectra in chapter three, where a number of normalisation methods are used to alleviate the problem, and in the estimation of Volterra kernels in chapter four, where some novel methods are used to try and reduce the variance of the higher order spectral estimates.

2.5.3 The use of linear windowing

In power spectral estimation the need for data windows is well understood. As stated in the previous section, the length of the window crucially affects the variance and bias of the estimate. In general, for a fixed length of data, increasing the length of window will increase the variance but decrease the bias, whilst decreasing the length of window will decrease the variance but increase the bias. Further discussion on the variance and bias of higher order spectral estimates is given in the next chapter. The remainder of the section gives a brief insight into the effect of the shape of the window, particularly where periodic signals are concerned.

When estimating power spectra, for periodic signals, if there is not an integer number of signal periods in each FFT frame, sidelobes can become a problem. Similar problems can occur in bispectral estimation and windowing effects can be very pronounced. Fackrell [32] gives a detailed study of the use of differently shaped windows when calculating the higher order spectra of periodic signals and concludes that a rectangular window gives minimal spreading of the main lobe, but causes high sidelobes which show up very strongly in the bispectrum. However, using a Hamming window gives much better results by reducing the side lobes.

This work is predominantly concerned with the higher order spectra of random signals. If the signal is broad band and has a spectrum which is slowly varying the effects of window shape will be greatly reduced and the results produced when using a rectangular window are minimally different from those generated when a Hamming window is used. Therefore, the majority of simulation and experimental work here uses a rectangular window.

2.5.4 Regions of symmetry and the principal domain

Just as the continuous power spectrum has symmetrical properties so do the continuous bispectrum and trispectrum. It is only necessary to evaluate the bispectrum and trispectrum

in the principal domain or non-redundant area, as all other regions can be found by symmetrical transformations of this area. In addition to the symmetrical properties of the continuous bispectrum and trispectrum, the discrete versions, like the discrete power spectrum, are periodic. In this section, the symmetries and principal domains of the bispectrum and trispectrum will be defined.

Using the symmetrical properties of the cumulant function, as detailed in Appendix A, it is easy to show that the continuous bispectrum will have the following symmetries:

$$\begin{aligned} |S_{xxx}(f_1,f_2)| &= |S_{xxx}(f_2,f_1)| \\ &= |S_{xxx}(-f_2,-f_1)| = |S_{xxx}(-f_1-f_2,f_2)| \\ &= |S_{xxx}(f_1,-f_1-f_2)| = |S_{xxx}(f_2,-f_1-f_2)| \end{aligned} \qquad 2.36$$

These are shown in figure 2.11. A star indicates where the symmetries involve a conjugation.

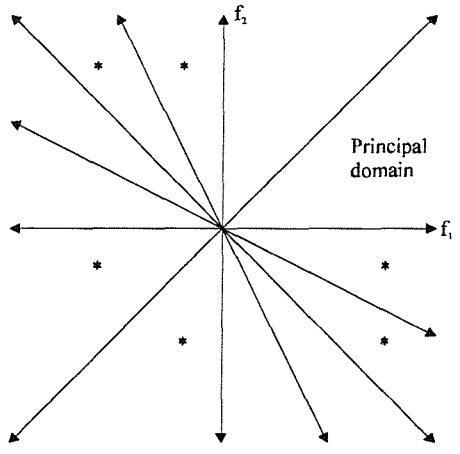

Figure 2.11: Symmetries of the continuous bispectrum

Thus a knowledge of the bispectrum in the triangular region $f_1,f_2 > 0$, and $f_2 < f_1$ is enough for a complete description of the continuous bispectrum. There are in fact twelve regions of symmetry in the bispectrum. As the discrete Fourier transform is periodic, the discrete bispectrum will also be periodic which means that there will be added symmetries for the discrete bispectrum and it can be shown [59, 73] that it is only necessary to estimate the

discrete bispectrum in the region $0 < f_2 < f_1$, and $f_1+f_2 < f_s/2$. This region is known as the principal domain and can be further subdivided into two regions. These are shown by the two triangles in figure 2.12, the inner triangle and the outer triangle. The importance of components of the bispectrum appearing in each of these regions will be discussed in the next section.

The discrete trispectrum has 96 regions of symmetry and has been extensively discussed [15, 25, 26, 80]. The principal domain is shown in figure 2.13. Like the bispectrum, it is divided into two regions: an inner volume (shown in dark grey); and an outer volume (shown in light grey). Referring to figure 2.13, both the inner and outer volume can be subdivided into two further regions: above and below the $f_3 = 0$ plane. Above the plane, all the frequency indices are positive and so the sum of three frequencies is equal and opposite to the fourth. Below the plane, only two of the first three frequencies are positive and so this region contains interactions where the sum of two frequencies is equal and opposite to the sum of the other two. Whereas the bispectrum of a narrow band signal is always zero, these extra frequency interactions allow the trispectrum of a narrow band signal to be non-zero. This will be extensively discussed in the next chapter.

It is of interest to note that if a slice through the principal domain of the trispectrum is taken along the $f_3 = 0$ plane, the principal domain of the bispectrum will be seen. It should however be noted that it is neither possible to obtain the bispectrum by taking a slice of the trispectrum nor the power spectrum by taking a slice of the bispectrum.

2.5.5 Plotting Conventions

In the bispectrum, the whole of the principal domain lies in the first quadrant and so in general when the bispectrum is plotted the whole of the first quadrant is shown. Other authors often display only the principal domain, setting the rest of the bispectral space to zero. However, it is often more pleasing to the eye, and the bispectral features are more easily interpreted, if the whole of the first quadrant is drawn, including all symmetries, and so predominantly this method is used.

The trispectrum splits its principal domain between the first and eighth octant. When there is no extra information in the eighth octant only the first octant is displayed. However, occasionally, as with narrow band signals, all the information in the trispectrum is in the eighth octant and in these cases the entire trispectrum is displayed. Again, except in special cases, all the symmetries are shown, not just the principal domain.

On all of the plots the position of the axes makes it clear exactly which part of the bispectrum or trispectrum is being displayed. If the axes intersect in the bottom left hand corner then only the first octant has been plotted whereas if they intersect in the centre of the trispectrum, all eight octants are plotted.

2.5.6 Sampling Considerations and Aliasing

It has been shown by Hinich [85] that the discrete bispectrum over the outer triangle is zero if the signal satisfies the following three conditions: 1) It is a random signal, 2) It is a stationary signal, 3) It has been sampled without aliasing. A statistical test for aliasing using this information was presented by Hinich [39]. The test proposed by Hinich for the bispectrum can be directly extended to the trispectrum, that is, significant outer volume content in the discrete trispectrum of a 4th order stationary random process identifies the presence of aliasing in the signal. Although aliasing is rarely a problem in electronic signals since it is standard practice to low pass filter signals to ensure $f_0 < f_s/2$ (where f_0 is the highest frequency present in the signal), there are certain circumstances where aliasing can cause problems.

First, some types of data are only available in discrete form and so analogue filtering is not possible. Such an example used by Hinich [39], as an application of his test, is in stock market data. Hinich showed that stock market results sampled at a rate of 1 sample a day were aliased as they had statistically significant outer triangle content. However, it is possible that the stock market data failed the test, not because it was aliased, but because

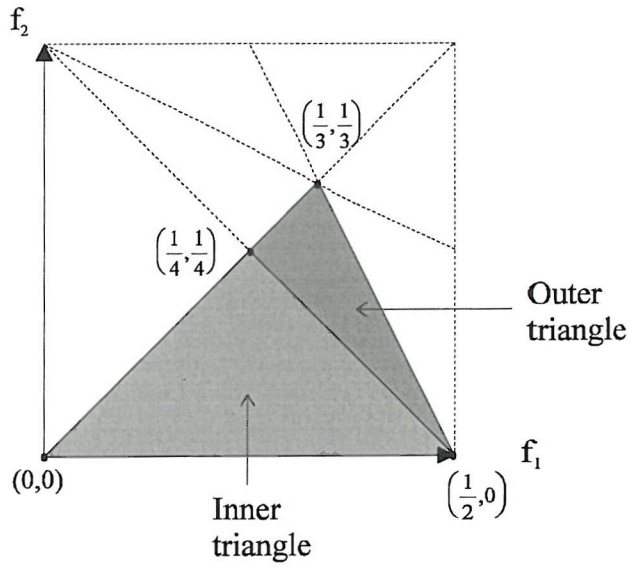

Figure 2.12: Principal domain of bispectrum

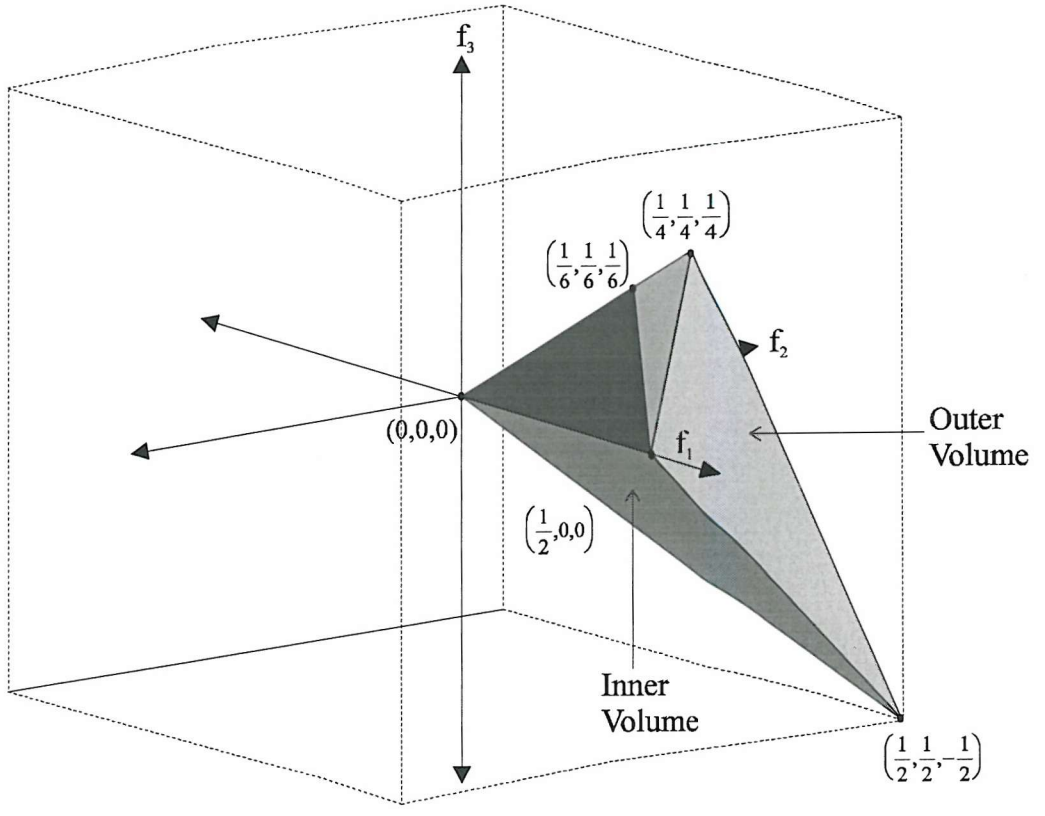

Figure 2.13: Principal domain of trispectrum

the data was non-stationary. An important criterion for the test is that the signal is stationary up to 3rd order for the bispectrum or stationary to 4th order for the trispectrum.

Another important concept involving the bispectral content of the outer triangle was first noted by Parsons and Williams [61] and has subsequently been developed in more detail by Stogioglou and McLaughlin [72]. That is, if a continuous stationary signal is sampled at its Nyquist rate then the higher order properties of the sampled signal cannot be modelled by passing discrete, independently and identically distributed (IID) noise through a linear filter.

Hinich has shown that the bispectrum of a discrete signal, which results from sampling a continuous, stationary signal at its Nyquist rate, with no aliasing, has zero bispectral content in the outer triangle. However, the bispectrum of a discrete stationary signal generated by passing IID noise through a linear filter will, in general, have a non-zero outer triangle. These statements are obviously contradictory, indicating the failure of the model to correctly represent a stationary signal sampled at its Nyquist rate.

Close inspection of Hinich's test shows that it is not the discrete signal but the parent signal from which it is sampled that must be stationary. Obviously if the parent signal is stationary then the discrete signal will also be stationary as it is a subset of its parent signal. However the converse is not true and the above contradiction implies that the parent signal of the discrete linear model is non-stationary at third order. Thus the location of components in the discrete bispectrum is also able to give some indication as to the stationarity of the continuous signal from which the discrete signal was sampled.

In conventional spectra, the presence or absence of aliasing is usually confirmed by making sure that the power spectrum is very low at the folding frequency $f_s/2$. A simple example of detecting 'digital aliasing' with the trispectrum is now given. If a discrete, broad band, stationary, Gaussian white process is passed through a cubic type nonlinearity the resulting signal will be broad band leptokurtic. However the cubing process will have produced components up to $3f_0$ and so the signal may be aliased. The trispectrum detects leptokurtic components across all frequencies as can be seen in figure 2.14 by the interactions at all frequency triplets. However close inspection of figure 2.14 will reveal that, rather than the

trispectrum being uniform across all frequencies as expected, some small almost periodic structure has been introduced by the aliasing.

If a similar Gaussian process is, before being cubed, first passed through a low pass filter with cut-off frequency $f_s/6$, the resulting signal should still be broad band leptokurtic but with no components above $f_s/2$. The trispectrum should have zero outer volume content and this is shown in figure 2.15, where it can be clearly seen that frequency interactions only occur in the inner volume.

From this example it is important to note that aliasing can occur not only when sampling continuous signals but also when working with synthetic discrete signals, and extreme care must be taken not to allow any frequency components of the signal to exceed $f_s/2$ as this can lead to spurious results.

2.6 Conclusions

This chapter has aimed to present some of the underlying theory behind higher order spectra. Initially broad band moments were discussed and their relationship with the probability density function explained. The bispectrum and trispectrum were introduced for both deterministic and stochastic processes. The possible frequency interactions which can occur in higher order spectra were demonstrated using a simple example based on sine waves. It was then shown, in a more traditional, mathematical, way, how higher order spectra could be derived from joint moment and cumulant functions.

The practicalities concerned with the estimation of the bispectrum and trispectrum, both in the frequency domain and time domain, were detailed together with the associated problem of windowing. Finally, the symmetries of both the bispectrum and trispectrum were described and the importance of structure occurring in the different regions of the discrete bispectrum and trispectrum stated. In the next chapter this work will be built upon to see how the bispectrum and trispectrum can be used to detect nonlinearities, and the many pitfalls that can occur in the process.

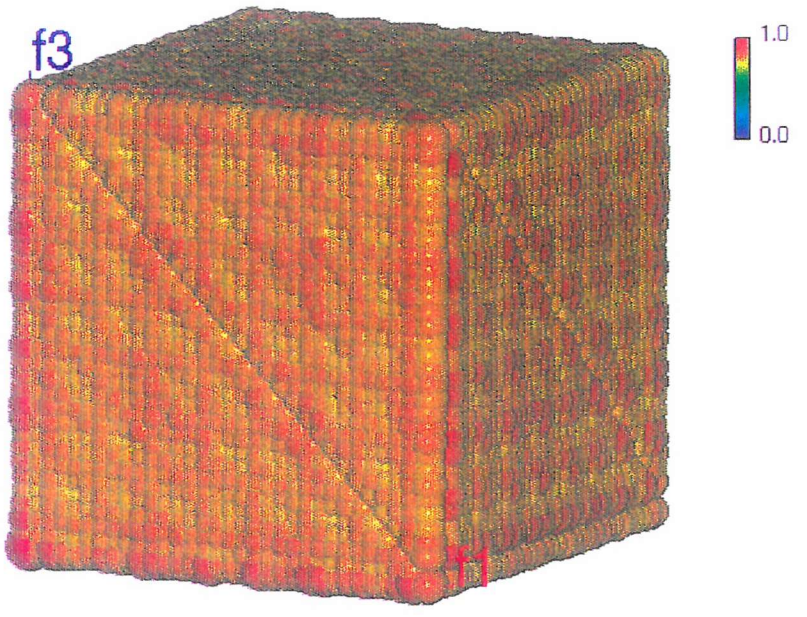

Figure 2.14: The trispectrum of a leptokurtic signal with aliasing

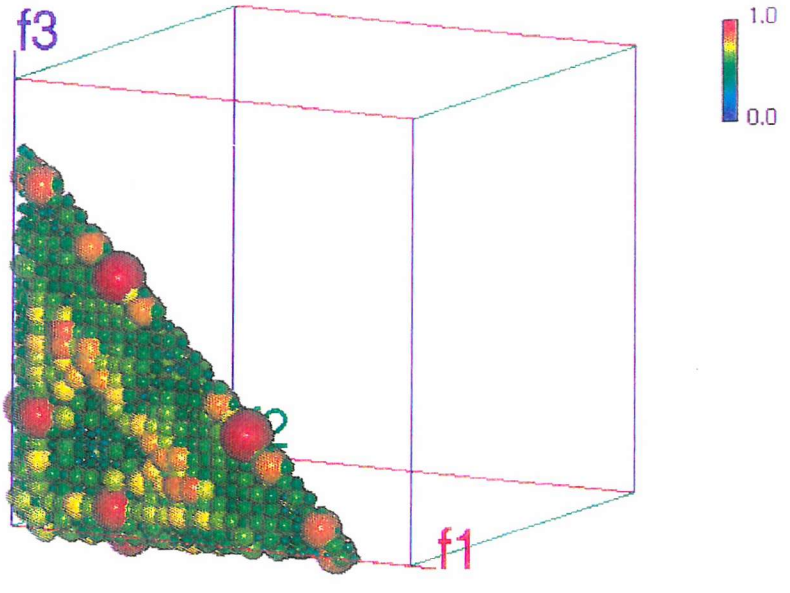

Figure 2.15: The trispectrum of a leptokurtic signal with no aliasing

Chapter 3

Normalisation techniques and examples of auto higher order spectra

3.1 Introduction - the need for normalisation

It was seen in equation 2.35 that bispectral estimates have a variance that is dependent on the power spectral properties of a signal. In this chapter, the different methods used to normalise the bispectrum and trispectrum, to remove these effects, are considered.

Two well developed methods which are based on the skewness function and the bicoherence are presented. The skewness function is used primarily to make decisions, based on statistical tests, about the symmetry, aliasing and linearity of a signal, whereas the bicoherence is used to detect the presence of quadratic phase coupling in a signal. These two third order functions are extended to their fourth order equivalents, the kurtosis function and tricoherence function, respectively. A series of examples is given showing the uses of the bicoherence and tricoherence function and some of the computational problems associated with them.

In section 3.6, narrow band signals are considered and it is shown that the bispectrum of a narrow band signal will always equal zero, but the trispectrum of such a signal is not necessarily zero. Traditional methods of normalisation of the trispectrum are often found to fail when applied to narrow band signals due to bias errors and so a new method is proposed, based on a pre-whitening technique. Two further examples which utilise this technique are given.

3.2 The skewness and kurtosis functions

Equation 2.35 for the variance of the auto bispectrum contains the following product of power spectral terms: $S_{XX}(f_1)S_{XX}(f_2)S_{XX}(f_1+f_2)$. Unless these terms are removed, estimates of the bispectrum may be contaminated by power spectral effects. One technique for removing these terms utilises the skewness function which is defined as:

$$s^2(f_1,f_2)=\frac{\left|S_{XXX}(f_1,f_2)\right|^2}{S_{XX}(f_1)S_{XX}(f_2)S_{XX}(f_1+f_2)} \qquad 3.1$$

This is the most commonly used normalisation method for statistical tests since it has the most useful and well understood statistical properties. Hinich [40], for example, use this function with some scaling modifications in their tests. It is simple to extend this to the 4th order kurtosis function for normalisation of the trispectrum:

$$k^2(f_1,f_2,f_3)=\frac{\left|S_{XXXX}(f_1,f_2,f_3)\right|^2}{S_{XX}(f_1)S_{XX}(f_2)S_{XX}(f_3)S_{XX}(f_1+f_2+f_3)} \qquad 3.2$$

It is noted that some authors [59] call the square root of equation 3.1 the 3rd order coherency function or bicoherency. The major difference of the skewness and kurtosis functions as compared to the bicoherence and tricoherence respectively, which are considered next, is that they have no upper bound.

It should be noted that the bicoherence and bicoherency are normalised auto bispectra and as such should not be confused with the ordinary coherence function. In chapter four, cross higher order spectra are introduced together with the concepts of quadratic and cubic coherence which can be thought of as extensions of the ordinary coherence function.

3.3 The bicoherence function

Another commonly used method of normalisation for the bispectrum is the bicoherence. The bicoherence is defined as,

$$b^2(f_1,f_2)=\frac{\left|S_{xxx}(f_1,f_2)\right|^2}{E\left[\left|X(f_1)X(f_2)\right|^2\right]E\left[\left|X(f_1+f_2)\right|^2\right]} \qquad 3.3$$

This has been shown by Kim and Powers [45] to have a variance, assuming $E[X(f_1)X(f_2)X(f_1+f_2)]=S_{xx}(f_1)S_{xx}(f_2)S_{xx}(f_1+f_2)$, that satisfies,

$$\mathrm{var}\left(b^2(f_1,f_2)\right)\approx\frac{1}{M}[1-b^2(f_1,f_2)] \qquad 3.4$$

where M is defined as the number of segments used in the estimation. Compared with the variance of the bispectrum, equation 2.35, there are no terms due to power spectral effects and so the bicoherence function should be solely dependent on the third order properties of the signal.

A useful feature of the bicoherence function is that it is bounded between 0 and 1. This can simply be shown using the Cauchy inequality [32], by making the substitution $Z(f_1,f_2) = X(f_1)X(f_2)$ into equation 3.3 to give,

$$b^2(f_1,f_2)=\frac{\left|E\left[Z(f_1,f_2)X^*(f_1+f_2)\right]\right|^2}{E\left[\left|Z(f_1,f_2)\right|^2\right]E\left[\left|X(f_1+f_2)\right|^2\right]} \qquad 3.5$$

Equation 3.5 is then of the form of the Cauchy inequality,

$$\left|E\left[Z(f_1,f_2)X^*(f_1+f_2)\right]\right|^2\le E\left[\left|Z(f_1,f_2)\right|^2\right]E\left[\left|X(f_1+f_2)\right|^2\right] \qquad 3.6$$

and so the bicoherence takes values satisfying,

$$0 \leq b^2(f_1, f_2) \leq 1 \qquad 3.7$$

The bicoherence is most often used in the detection of Quadratic Phase Coupling (QPC) which will be described in section 3.3.2 and is the predominantly used method of normalisation throughout this work.

3.3.1 An example of the bicoherence function: the mixed signal

The effectiveness of the bicoherence is easily shown by considering the signal, w(t). This was created, as in figure 3.1, by taking a white Gaussian signal, x(t), and filtering it through a low pass filter with a cut-off frequency of 0.1 to produce y(t). The frequencies are normalised to set f_s, the sampling frequency, equal to one. The signal y(t) was then squared to produce a new signal with a low frequency skewed component. Finally, this signal was added to z(t) which consists of x(t) filtered through a high pass filter with a cut-on frequency of 0.25. The resulting signal, w(t), now consists of a low frequency skewed component and a high frequency Gaussian component. This signal is dubbed a mixed signal and is used throughout the next chapters to demonstrate the properties of both auto and cross higher order spectra. When the mixed signal is used to demonstrate the properties of the trispectrum, y(t) will be cubed rather than squared so as to produce a low frequency leptokurtic component.

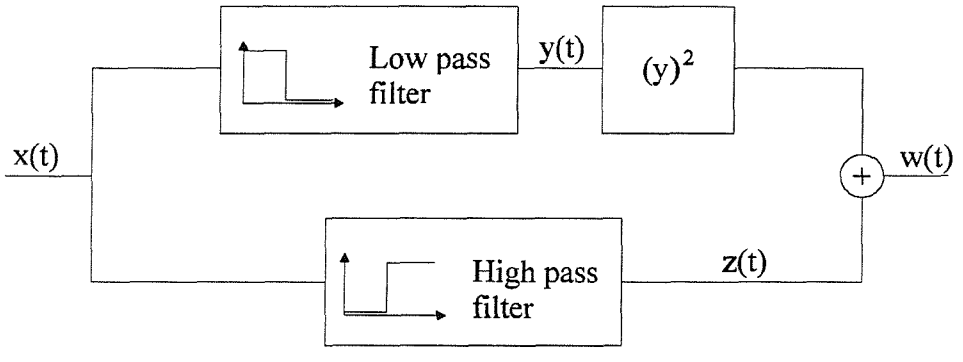

Figure 3.1: The skewed mixed signal

The power spectrum of the mixed signal, figure 3.2, shows the two distinct regions but of course is not able to differentiate the skewed component from the Gaussian component. If the raw bispectrum of the mixed signal, (see figure 3.3), is estimated, the skewed component of the signal is visible between f_1, $f_2 < 0.2$. However the magnitude of the bispectrum is equally large at high frequencies. This is because the variance of the raw bispectrum contains terms due to the power spectral effects of the mixed signal and so both the skewed low pass component and the Gaussian high pass component are detected. Unless the raw bispectrum is averaged over vastly longer data lengths to reduce its variance, which is impractical, it is unable to differentiate between the two components. However, the bicoherence, figure 3.4, normalises the raw bispectrum by removing the power spectral effects and can be seen to correctly detect just the low frequency skewed component.

Unless otherwise stated, all estimates of the bispectrum use a data length of 4096 sample points and a window size of 64 and hence are averaged over 64 segments. Estimates of the trispectrum use a data length of 262144 and the same window size and are averaged over 4096 segments.

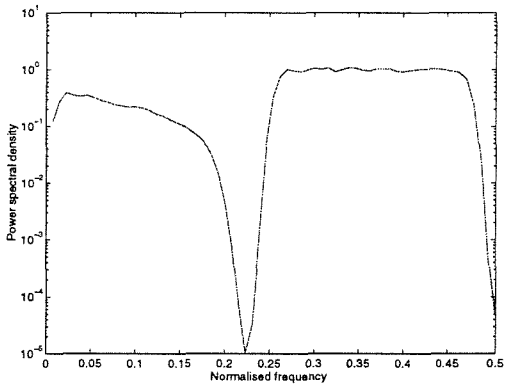

Figure 3.2: Power spectrum of skewed mixed signal

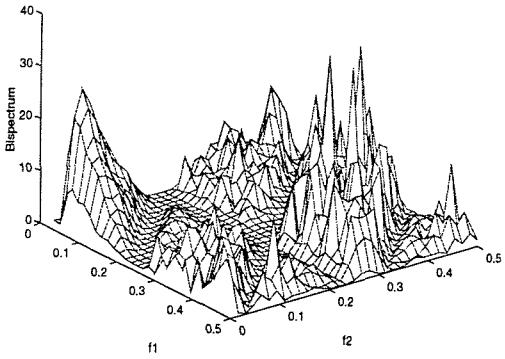

Figure 3.3: Bispectrum of mixed signal

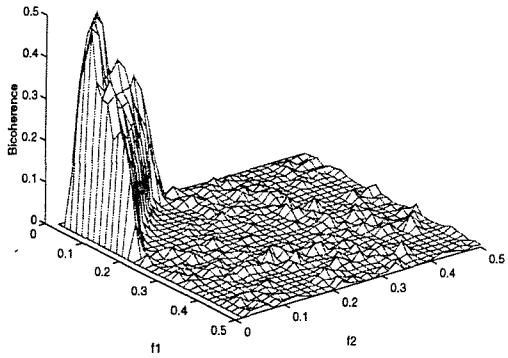

Figure 3.4 Bicoherence of mixed signal

3.3.2 Quadratic Phase Coupling

In a Gaussian signal all of the phases of different frequency components are uniformly distributed between 0 and 2π. When a signal has quadratic phase coupling, the phases of components at frequencies f and 2f are correlated. Phase coupling occurs due to nonlinear interactions between components in a signal and it is sometimes of interest to know whether peaks at harmonically related positions in the power spectrum are caused by this interaction or not. The power spectrum is phase blind, and so is not able to detect the presence of phase coupling. However, the magnitude of the bicoherence gives an indication of the correlation between the phases of the frequency components at f_1 and f_2. Quadratic phase coupling will occur in a signal with a quadratic nonlinear component.

Consider a signal, x(t), which consists of a white, Gaussian signal passed through a narrow band filter, bandlimited between the frequencies 0.1 to 0.2, and then passed through a quadratic, x^2, type nonlinearity. A second filter can be generated that has the same characteristics as the spectrum of x(t). The same Gaussian signal is passed through this filter to produce a signal, y(t) that has a similar spectrum as x(t) but contains no nonlinear components.

As can be seen in figures 3.5 and 3.6, although the power spectra of the two signals are almost identical, x(t) contains quadratic phase coupled components caused by the squaring

operation and so the bicoherence, figure 3.7, has peaks over the frequency range 0.1 to 0.2. However, y(t), contains Gaussian noise with no phase coupling and so the bicoherence, figure 3.8, is zero.

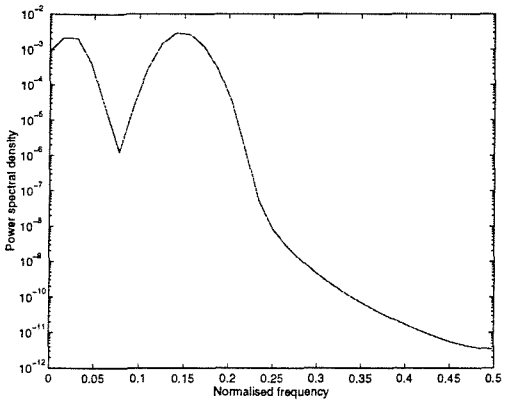

Figure 3.5: Power spectrum of signal with quadratic phase coupling

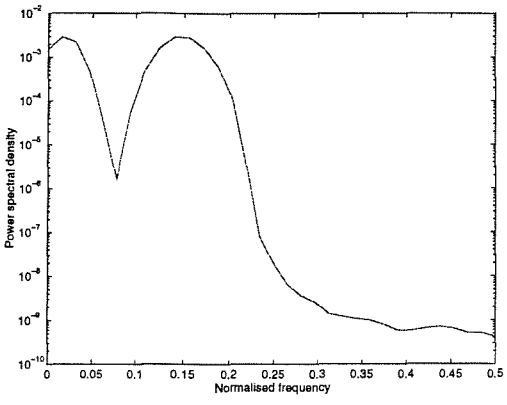

Figure 3.6: Power spectrum of signal with no quadratic phase coupling

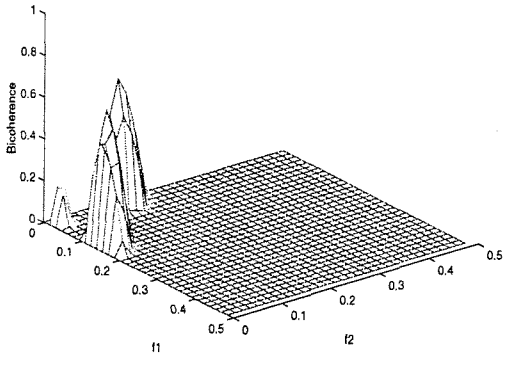

Figure 3.7: Bicoherence of signal with quadratic phase coupling

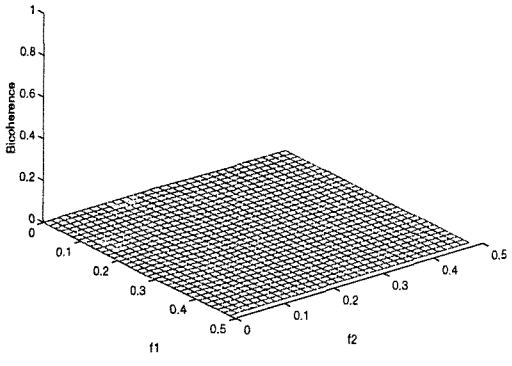

Figure 3.8: Bicoherence of signal with no quadratic phase coupling

3.4 The tricoherence function

The bicoherence can be extended to the fourth order case to form the tricoherence which is defined as:

$$t^2(f_1,f_2,f_3) = \frac{\left|S_{xxxx}(f_1,f_2,f_3)\right|^2}{E\left[\left|X(f_1)X(f_2)X(f_3)\right|^2\right]E\left[\left|X(f_1+f_2+f_3)\right|^2\right]} \qquad 3.8$$

Using a similar proof as for the bispectrum it can be shown that the tricoherence function can only take values from 0 to 1. The effectiveness of the tricoherence as a normalisation process can be demonstrated by considering the mixed signal described in section 3.3.1, except rather than considering a low frequency skewed component, a low frequency leptokurtic component is used. The frequency decomposition of the signal will again have two distinct regions: a low frequency leptokurtic part and a high frequency Gaussian part.

The fourth order spectrum should only contain components due to the leptokurtic signal but the trispectrum, figure 3.9, contains components right up to the folding frequency, $f_s/2$. The tricoherence, figure 3.10, picks out only the components due to the low frequency part of the signal. This confirms that a normalisation process is necessary to separate second order effects from fourth order just as it is necessary to separate third order effects from second order in the bispectrum.

3.4.3 Cubic Phase Coupling

To demonstrate cubic phase coupling in the tricoherence, the example used in the bispectrum is extended to consider a narrow band signal which has been cubed rather than squared, thus creating a signal with cubically phase coupled terms. The resulting tricoherence contains cubic nonlinear terms and shows a peak over the frequency range 0.1 to 0.2 (figure 3.11). The tricoherence of the filtered Gaussian signal (figure 3.12) contains no such structure. The maximum value of the tricoherence of the filtered Gaussian signal is 2×10^{-5} compared with a maximum value of 0.3 for the peak in the tricoherence of the signal containing cubic phase coupling.

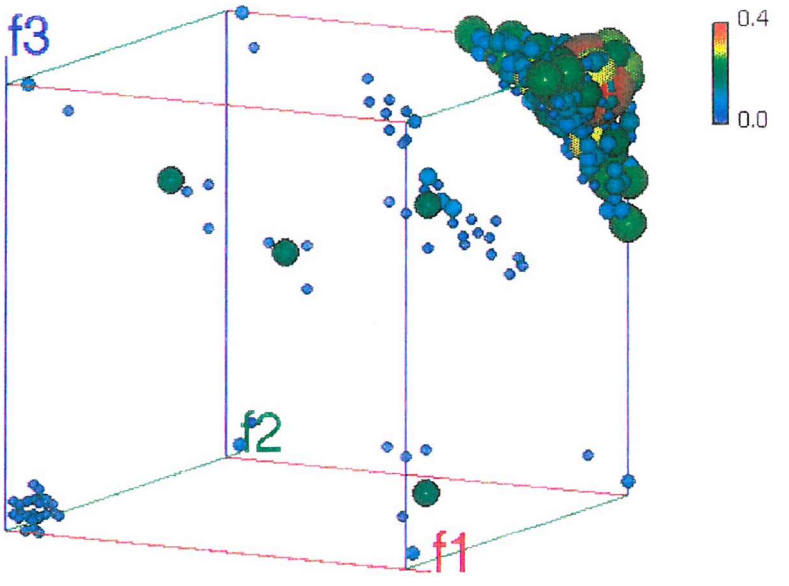

Figure 3.9: Trispectrum of a leptokurtic mixed signal

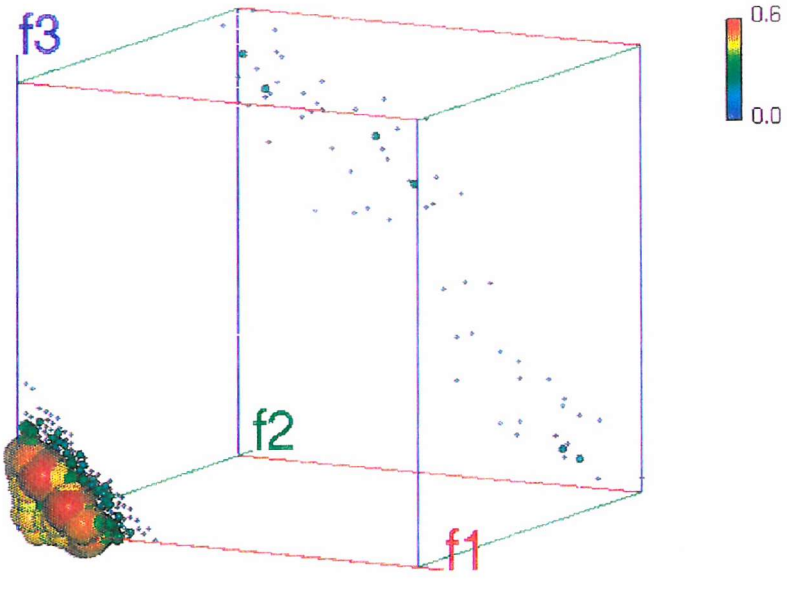

Figure 3.10: Tricoherence of a leptokurtic mixed signal

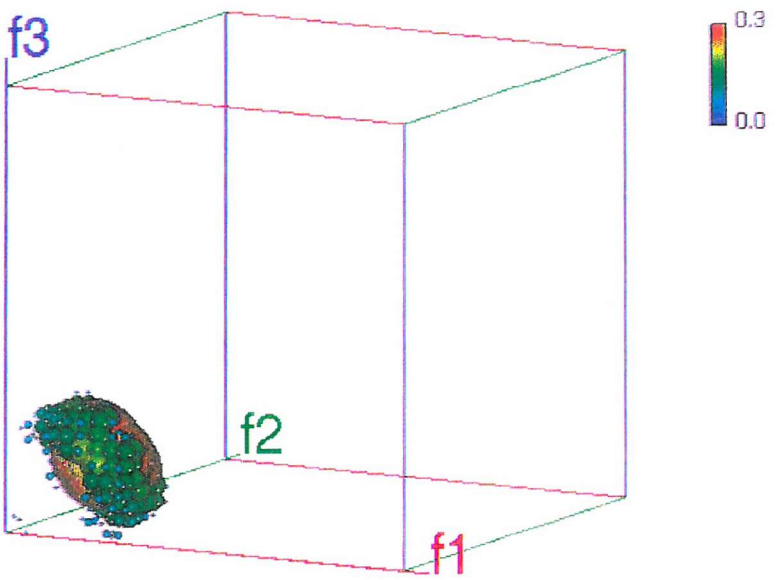

Figure 3.11: Tricoherence of a signal with cubic phase coupling

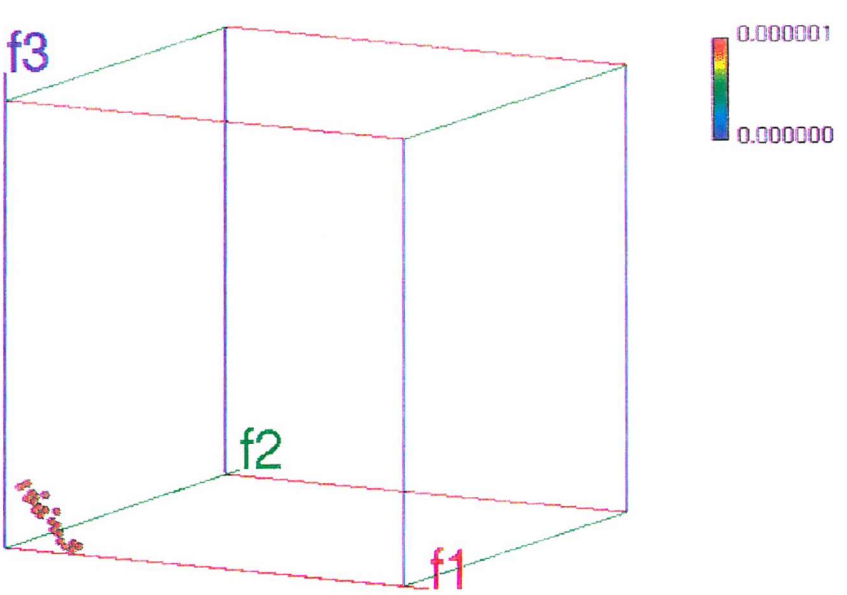

Figure 3.12: Tricoherence of a signal with no cubic phase coupling

3.5 Problems associated with the estimation of the bicoherence and tricoherence

If the bicoherence of a low pass skewed signal is calculated, the low pass filter should remove all energy above the cut-off frequency and the bicoherence should only contain peaks due to the low frequency skewed component. However prominent vertical and horizontal lines are visible in the bicoherence right up to the folding frequency. Fackrell [29] showed that these were due to small values occurring in the denominator of the bicoherence which cause the estimate to blow up. They can be removed by adding a small constant, ε, to the denominator across all frequencies and calculating the bicoherence:

$$b^2(f_1,f_2)=\frac{\left|S_{xxx}(f_1,f_2)\right|^2}{E\left[\left|X(f_1,f_2)\right|^2\right]E\left[\left|X(f_1+f_2)\right|^2\right]+\varepsilon} \qquad 3.9$$

Where there is no energy in the signal, this results in a non-zero denominator and a zero numerator and so the bicoherence is zero. The disadvantage of using this correction technique is that ε introduces a small amount of negative bias in the bicoherence estimator.

A similar effect occurs in the fourth order spectrum where small denominator terms can cause spurious structure in the tricoherence. A low pass leptokurtic signal was created by passing a Gaussian signal through a low pass filter with a normalised cut-off frequency of 0.1. The resulting signal was then cubed. The tricoherence of this signal should contain no energy above three times the cut-off frequency. However it can be seen in figure 3.13 that there is dominant structure right up to $f_s/2$. As with the bicoherence this is an effect of the normalisation process and can be removed by adding a small constant, ε, to the denominator across all frequencies as in equation 3.10. When $\varepsilon = 1\times10^{-14}$ the tricoherence of the above signal, figure 3.14, correctly detects just the leptokurtic component.

$$t^2(f_1,f_2,f_3)=\frac{\left|S_{xxxx}(f_1,f_2,f_3)\right|^2}{E\left[\left|X(f_1)X(f_2)X(f_3)\right|^2\right]E\left[\left|X(f_1+f_2+f_3)\right|^2\right]+\varepsilon} \qquad 3.10$$

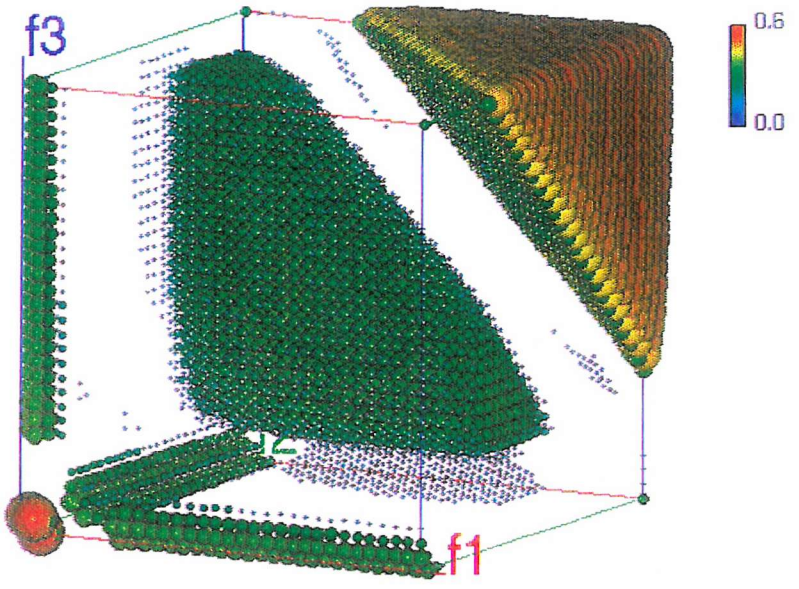

Figure 3.13: Tricoherence of a low pass leptokurtic signal with $\varepsilon = 0$

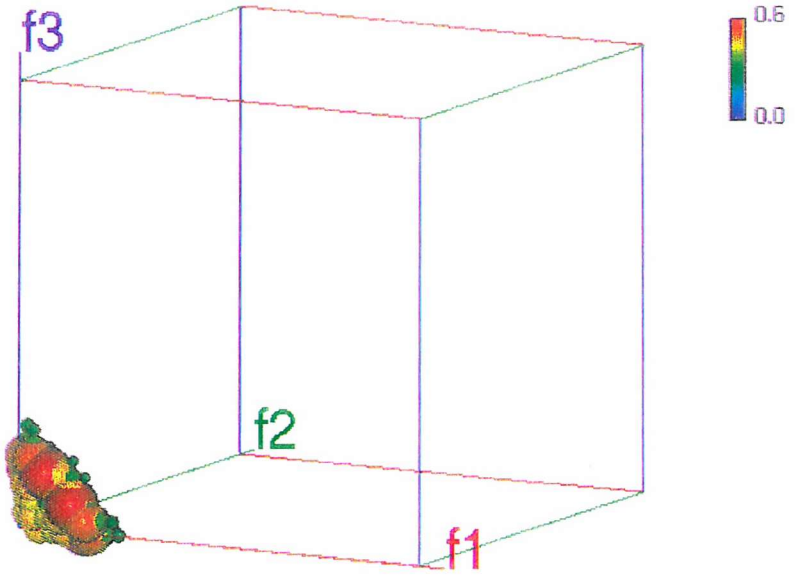

Figure 3.14: Tricoherence of a low pass leptokurtic signal with $\varepsilon = 1 \times 10^{-14}$

3.6 Pre-whitening techniques

The bicoherence and tricoherence approaches are the most common methods used for normalising the bispectrum and trispectrum. However, problems can occur, due to bias errors, if the tricoherence of a very narrow band signal is estimated. The problem has never arisen in the bicoherence as it will be shown that the bicoherence of a signal with a bandwidth of less than one octave is zero, whereas the tricoherence of such a narrow band process is not necessarily zero.

It was explained in section 2.5 that the window length has to be chosen carefully so as to minimise the variance whilst keeping the bias error small. The largest window length that can reasonably be used in the estimation of a trispectrum is 64, owing to the large computational requirement. Therefore for systems with very narrow resonant peaks the tricoherence may be significantly biased. This is due to the coarseness of the spectral estimate and the narrowness of the resonant peak, i.e. it will occur when the width of the frequency cell is wider than the resonance. Subba Rao [73] has developed theoretical expressions for the bias of the bispectrum and these are crucially dependent on the two dimensional curvature of the true bispectrum. Therefore around a sharp peak where there is large curvature the bias error will be large. Similar expressions exist for the bias of the trispectrum which depend on the three dimensional curvature of the true trispectrum.

In this section a possible method of overcoming the large bias errors often encountered in the trispectrum is discussed. The technique used is to pre-whiten the signal before estimating the trispectrum. As this can be done in the time domain, using the full data length, a much greater frequency resolution can be used than for the estimation of the trispectrum. The pre-whitening operation is a linear operation, and it will be shown that linear operations have no effect on the tricoherence.

The pre-whitening technique is first demonstrated on a narrow band, amplitude modulated (AM) type process. It is shown that conventional normalisation techniques fail to differentiate between a Gaussian and a kurtotic signal, but pre-whitening can overcome this.

3.6.1 Bispectrum and trispectrum of a bandlimited signal.

Consider a bandlimited signal that has components with an upper frequency of f_U and lower frequency of f_L. From the definition of the continuous bispectrum,

$$S_{XXX}(f_1,f_2)=E[X(f_1)X(f_2)X^*(f_1+f_2)] \quad 3.11$$

for the bispectrum to have non-zero values the following inequalities must all apply:

$$f_L < f_1 < f_U \quad 3.12$$

$$f_L < f_2 < f_U \quad 3.13$$

$$f_L < (f_1+f_2) < f_U \quad 3.14$$

Adding 3.12 and 3.13 gives,

$$2f_L < (f_1+f_2) < 2f_U \quad 3.15$$

From 3.14 and 3.15,

$$f_U > f_1+f_2 \quad 3.16$$

$$2f_L < f_1+f_2 \quad 3.17$$

Combining equations 3.16 and 3.17 gives the condition:

$$f_U > 2f_L \quad 3.18$$

Therefore, the upper frequency must be greater than twice the lower frequency for the continuous bispectrum to be non-zero and so the bispectrum of a narrow band process, with a bandwidth of less than an octave, is zero. An interesting corollary to this is that the probability density function of a stationary narrow band signal must always be symmetric.

However, the continuous cumulant trispectrum of a bandlimited signal,

$$\begin{aligned} S_{xxxx}(f_1,f_2,f_3) = {} & E[X(f_1)X(f_2)X(f_3)X^*(f_1+f_2+f_3)] \\ & - E\left[|X(f_1)|^2\right]E\left[|X(f_2)|^2\right]\delta(f_2+f_3) \\ & - E\left[|X(f_2)|^2\right]E\left[|X(f_1)|^2\right]\delta(f_1+f_3) \\ & - E\left[|X(f_1)|^2\right]E\left[|X(f_3)|^2\right]\delta(f_1+f_2) \end{aligned} \qquad 3.19$$

can have non-zero values throughout its range because of its three frequency interactions. For example, consider the case where the frequencies f_1, f_2, f_3 take the following values,

$$f_1 = f_L,\ f_2 = -f_L, \text{ and } f_3 = f_U. \qquad 3.20$$

Then,

$$f_1 + f_2 + f_3 = f_U \qquad 3.21$$

and equation 3.19 can be written as,

$$\begin{aligned} S_{xxxx}(f_1,f_2,f_3) &= E[X(f_L)X(-f_L)X(f_U)X(-f_U)] - E\left[|X(f_L)|^2\right]E\left[|X(f_U)|^2\right] \\ &= E\left[|X(f_L)|^2|X(f_U)|^2\right] - E\left[|X(f_L)|^2\right]E\left[|X(f_U)|^2\right] \end{aligned} \qquad 3.22$$

If there is a correlation between the amplitudes of the frequencies at f_U and f_L then the trispectrum is non-zero, as is the case for many amplitude modulated processes.

3.6.2 An example of pre-whitening: an amplitude modulated process

A bandlimited signal that contains third order nonlinearities was created by passing a Gaussian signal through a low pass filter with a normalised cut-off frequency of 0.1. The resulting signal was multiplied by a sine wave to produce an amplitude modulated signal. The power spectrum of this signal is shown in figure 3.15. The apparent power at DC is

caused by cross terms between the positive and negative frequency components. A Gaussian bandlimited signal with a similar spectrum to the leptokurtic signal was also created by passing a Gaussian signal through a filter which had the same shape as the spectrum of the amplitude modulated signal.

As both signals have a bandwidth of less than an octave, they both have zero bispectra. However the Gaussian signal should have zero trispectrum, whereas the amplitude modulated process should have significant structure in the trispectrum.

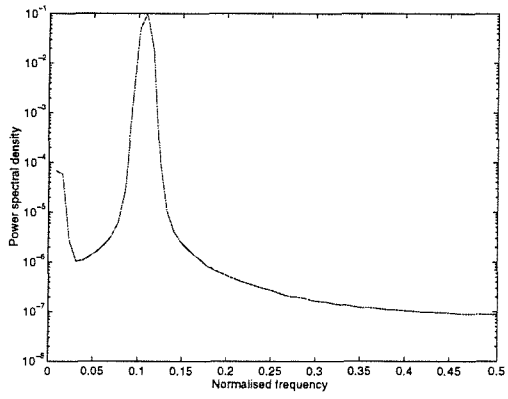

Figure 3.15: Power spectrum of leptokurtic, bandlimited, amplitude modulated signal.

The tricoherence of the two signals is shown in figures 3.16 and 3.17. Both signals have significant tricoherence and the same structure, and both are dominated by the peak in the spectrum. It is expected that the Gaussian signal should have a zero trispectrum, but it can be seen that it still has significant structure (maximum value = 0.08) as compared to the leptokurtic signal (maximum value = 0.23) which can cause confusion. The normalisation of the trispectrum in the frequency domain to form the tricoherence has not been effective, as both the estimate of the tricoherence of the nonlinear signal and the tricoherence of the linear signal are still dominated by their spectra. Owing to the coarseness of the spectral estimates, and the narrowness of the resonant peak, normalisation in the frequency domain has failed. For such signals, which contain peaks narrow in comparison to a frequency cell, the tricoherence produces a biased estimate.

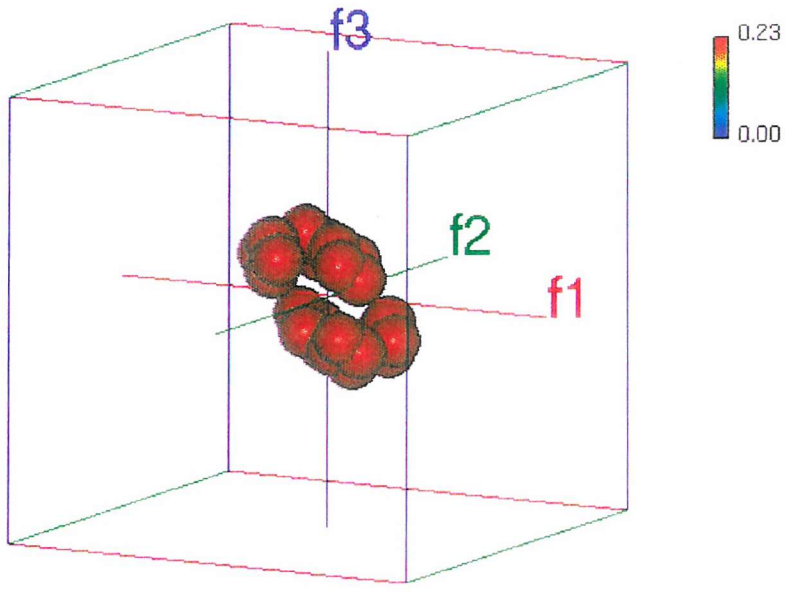

Figure 3.16: Tricoherence of a leptokurtic, narrow band, amplitude modulated process

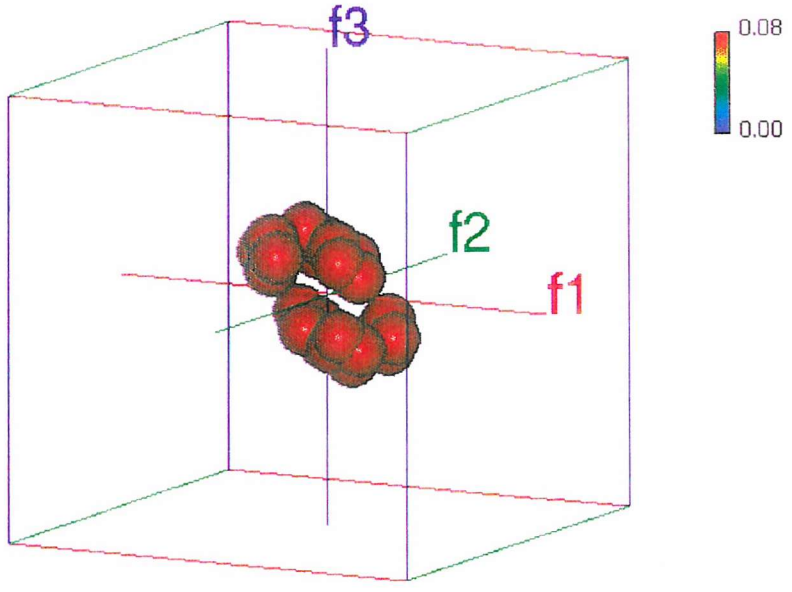

Figure 3.17: Tricoherence of a Gaussian, narrow band process

In order to remove the second order effects, the spectra of the signals were flattened, prior to calculation of the trispectrum. A finite impulse response filter was calculated that has the same frequency response as the square root of the inverse of the spectrum of the signal. The original signal was convolved with this inverse filter to produce a signal with a spectral density function of unity. The inverse filter must be linear phase to ensure that (correlated) frequency components under go the same delay. The phase characteristic of the filter used here is shown in figure 3.18a. As convolution is a linear operation it should have no effect on the tricoherence. It is easy to show that both the bicoherence and tricoherence are unaffected by linear transfer functions, except in the case of the tricoherence along the planes $f_1+f_2 = 0$, $f_1+f_3 = 0$, $f_2+f_3 = 0$.

Consider the bicoherence function,

$$b^2(f_1,f_2) = \frac{\left|E\left[X(f_1)X(f_2)X^*(f_1+f_2)\right]\right|^2}{E\left[\left|X(f_1)X(f_2)\right|^2\right]E\left[\left|X(f_1+f_2)\right|^2\right]} \qquad 3.23$$

Now apply a linear transfer function with impulse response, h(t), to the signal, x(t), to obtain a new signal, y(t). Convolution in the time domain is equivalent to multiplication in the frequency domain. This can be written as,

$$Y(f)=H(f)X(f) \qquad 3.24$$

If the bicoherence of the filtered version of the signal is now calculated, the following expression is obtained,

$$b^2(f_1,f_2) = \frac{\left|E\left[H(f_1)X(f_1)H(f_2)X(f_2)H^*(f_1+f_2)X^*(f_1+f_2)\right]\right|^2}{E\left[\left|H(f_1)X(f_1)H(f_2)X(f_2)\right|^2\right]E\left[\left|H(f_1+f_2)X(f_1+f_2)\right|^2\right]} \qquad 3.25$$

As H(f) is time invariant, it is unaffected by the expectation operator and so it can be taken outside the expectation operation. The numerator transfer functions will then cancel with the denominator transfer functions to leave the original expression for the bicoherence.

The flattening or 'pre-whitening' of a spectrum of a signal occurs in the time domain before the trispectrum is calculated. This means that it can occur at a much finer frequency resolution (fft size = 1024) as the whole of the data length (262144 samples for a trispectrum with a fft size of 64) can be used to calculate the filter. The power spectrum of the flattened amplitude modulated signal is shown in figure 3.18b, plotted against the original signal.

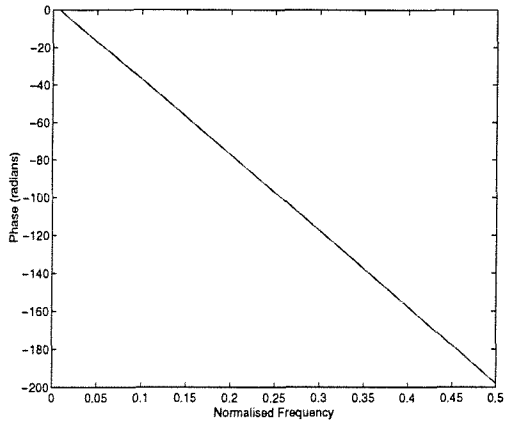

Figure 3.18a: Phase response of inverse filter

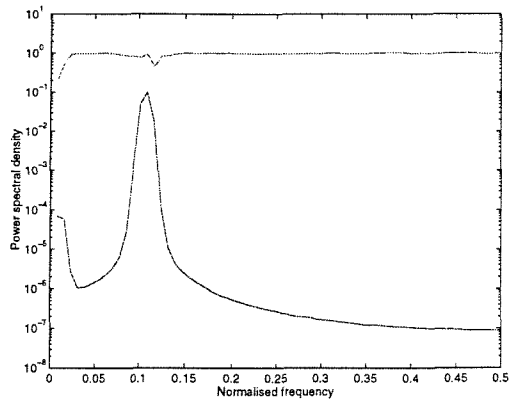

Figure 3.18b: Power spectrum of amplitude modulated signal and flattened amplitude modulated signal

The trispectrum of the flattened Gaussian signal (figure 3.19) now has no structure and a maximum value of 0.003, whereas the flattened nonlinear signal (figure 3.20) has peaks of value 0.18 at frequencies corresponding to the position of the nonlinearity in the original signal. Therefore it has been shown that in order to reduce the bias of the estimate of the trispectrum, for such a narrow band process, whilst still ensuring a low variance, it is necessary to pre-whiten the signal in the time domain.

If the signal has been pre-whitened, the spectral effects have been removed and so there is no longer any need to normalise the trispectrum with respect to the power spectrum to form the tricoherence. However, it is still possible to calculate the tricoherence of a pre-whitened signal and there is the added advantage that the tricoherence gives a normalised measure from 0 to 1, whereas the pre-whitened trispectrum will not have any upper bound.

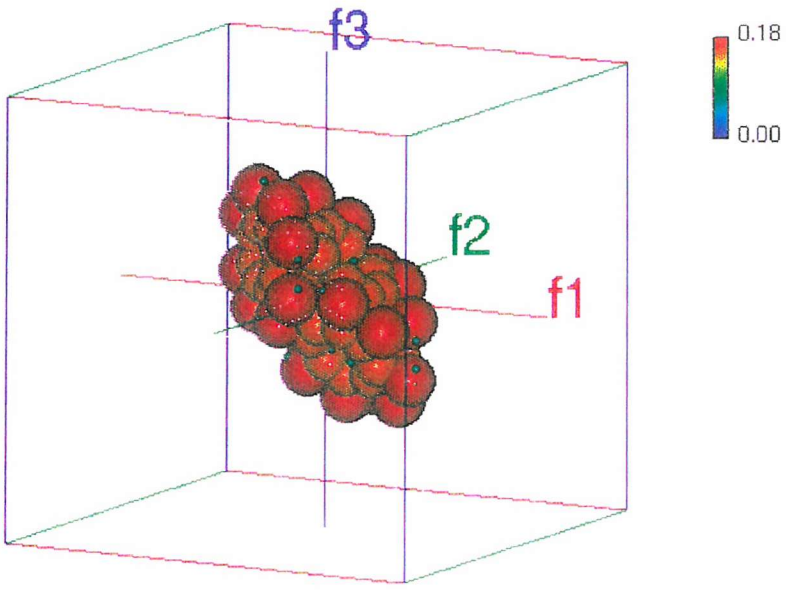

Figure 3.19: Trispectrum of a pre-whitened, leptokurtic, narrow band, amplitude modulated process

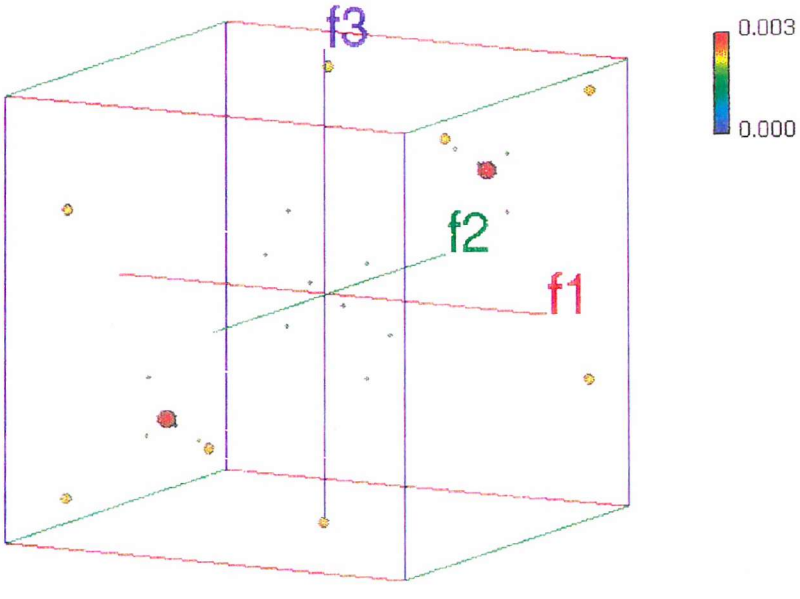

Figure 3.20: Trispectrum of a pre-whitened, Gaussian, narrow band process

3.6.3 The Duffing oscillator

A second, more practical, example is given, using the pre-whitening technique to detect nonlinearity in a dynamic system, the Duffing oscillator. A process, x(t), is classed as the output of a Duffing oscillator if x(t) satisfies the nonlinear differential equation:

$$\ddot{x} + 2\zeta\omega_0\dot{x} + \omega_0{}^2 x + \alpha x^3 = f(t) \qquad 3.26$$

In the following, the input, f(t), is assumed to be a bandlimited, Gaussian white noise signal. This equation can be used to model many practical applications, for example [14, 36, 49].

Figures 3.21-3.24 show the power spectra obtained by simulating the Duffing equation using a fourth order, fixed step, Runge Kutta approximation. The system is lightly damped (ζ=0.1) with a linear undamped natural frequency of 0.025 ad/s. Figure 3.21 shows the linear case, where $\alpha = 0$, and then the degree of nonlinearity is increased; $\alpha = 0.0001$ (figure 3.22), $\alpha = 0.001$ (figure 3.23), and $\alpha = 0.01$ (figure 3.24).

As expected in the linear case there is a single resonance. As the nonlinear component is increased, a number of effects take place. These are explained in more detail in [36].

- A component at three times ω_0 appears. This is as expected as the nonlinearity is predominantly cubic.
- The resonant frequency increases. As α increases, the system becomes more stiff for large displacements and so the frequency of the resonance effect will increase.
- The resonant peak broadens.

When the nonlinearity is introduced, the broad band kurtosis, μ_4, decreases from 3 as shown in table 3.1, and the probability density function becomes platykurtic, i.e. the tails on the probability density function are reduced and the distribution becomes more uniform.

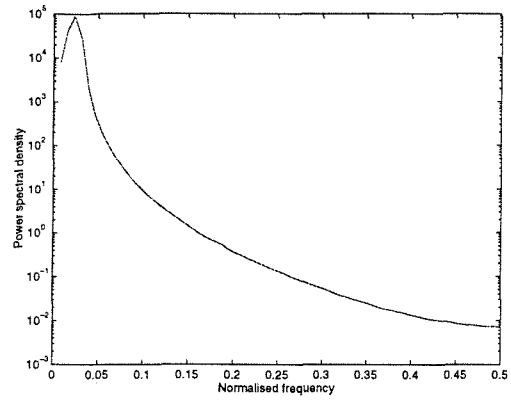

Figure 3.21: Power spectrum of the Duffing oscillator with $\alpha = 0$

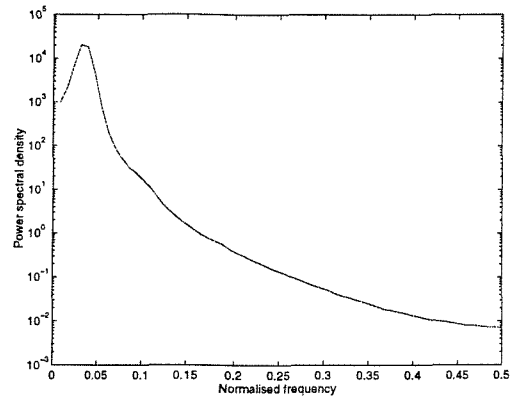

Figure 3.22: Power spectrum of the Duffing oscillator with $\alpha = 0.0001$

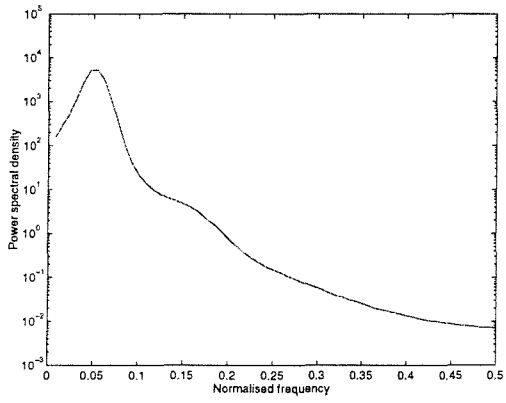

Figure 3.23: Power spectrum of the Duffing oscillator with $\alpha = 0.001$

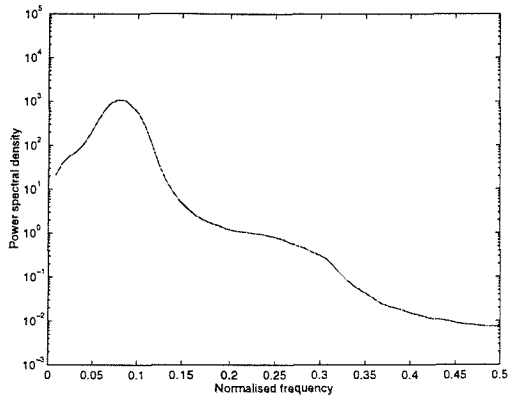

Figure 3.24: Power spectrum of the Duffing oscillator with $\alpha = 0.01$

The power spectrum and kurtosis both give some idea of the nonlinearity in the differential equation. However, a possible method for accurately detecting the magnitude and frequency of the nonlinear terms, is to look at the higher order spectra of the system. As the input, f(t), is symmetric, and the system contains only odd order terms, x(t) will be symmetric and so the bispectral density function of x(t) will be zero. Therefore, in order to analyse the nonlinearities in x(t) it is necessary to study the trispectrum.

α	μ_4
0	3.03
0.0001	2.23
0.001	2.17
0.01	2.33

Table 3.1: Values of broadband kurtosis, μ_4, compared with the magnitude of the nonlinear term, α, in the Duffing oscillator

For each of the four cases, the data was pre-whitened before the trispectra, figures 3.25 - 3.28, were calculated. For the linear case, (figure 3.25), there is no significant structure in the trispectrum as when $\alpha = 0$, the output is Gaussian and so the trispectrum is zero. As the nonlinearity is gradually increased (figures 3.26-3.28), structure begins to appear in the trispectrum at (f_r,f_r,f_r), where f_r is the resonant frequency, and all the symmetrical reflections of this point. The magnitude of this structure is directly related to the size of the nonlinear term (table 3.2). Note, the broadband kurtosis is not directly related to the size of the nonlinear term. These values were calculated for a trispectrum with an FFT size of 64, using a data length of 262144 points.

Value of α	Peak in trispectrum
0	0.001
0.00001	1.269
0.0001	16.114
0.001	25.581

Table 3.2: Value of the magnitude of the peak in the trispectrum compared with the magnitude of the nonlinear term, α, in the Duffing oscillator

This example has shown that the pre-whitened trispectrum is able to detect the magnitude and frequency of the nonlinear terms in the Duffing equation and is sensitive to relatively small digressions from linearity.

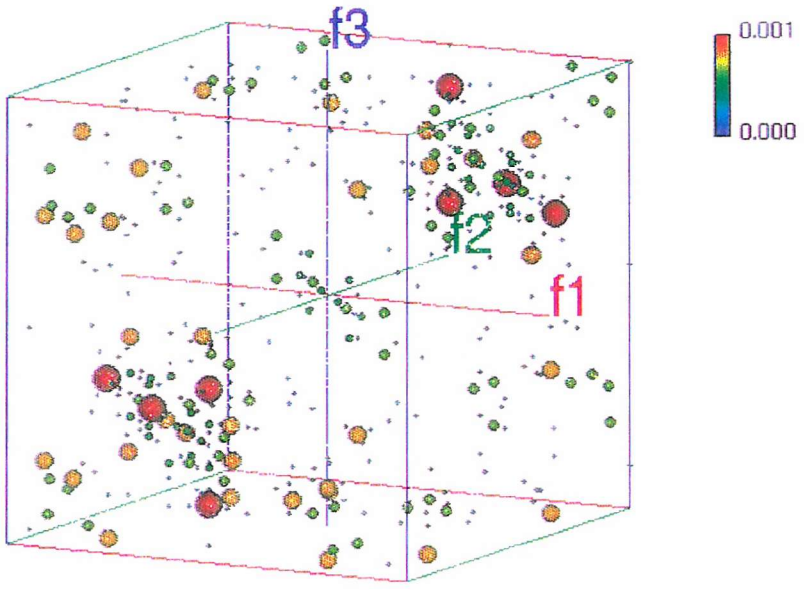

Figure 3.25: Trispectrum of the Duffing oscillator with $\alpha = 0$

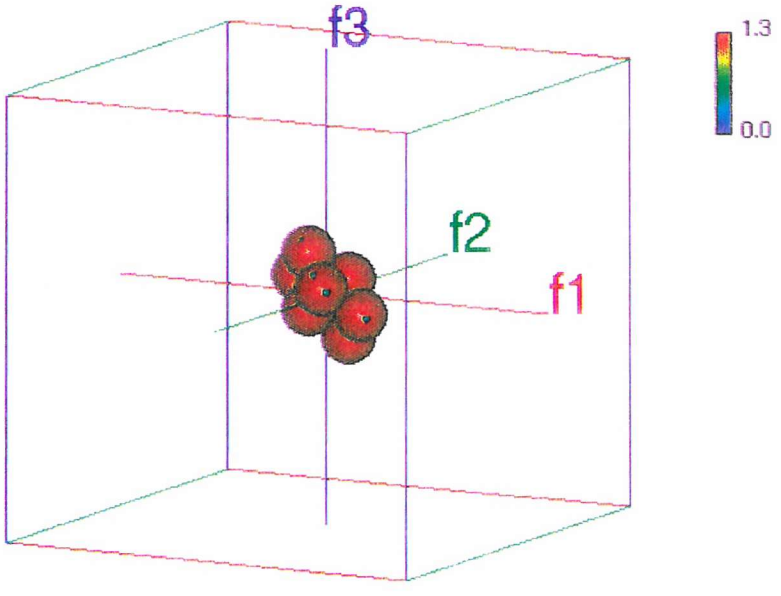

Figure 3.26: Trispectrum of the Duffing oscillator with $\alpha = 0.0001$

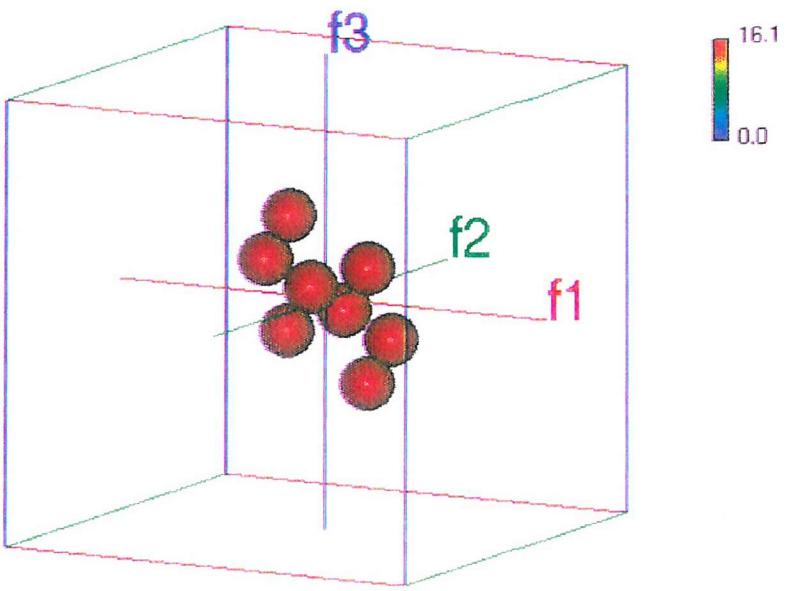

Figure 3.27: Trispectrum of the Duffing oscillator with $\alpha = 0.001$

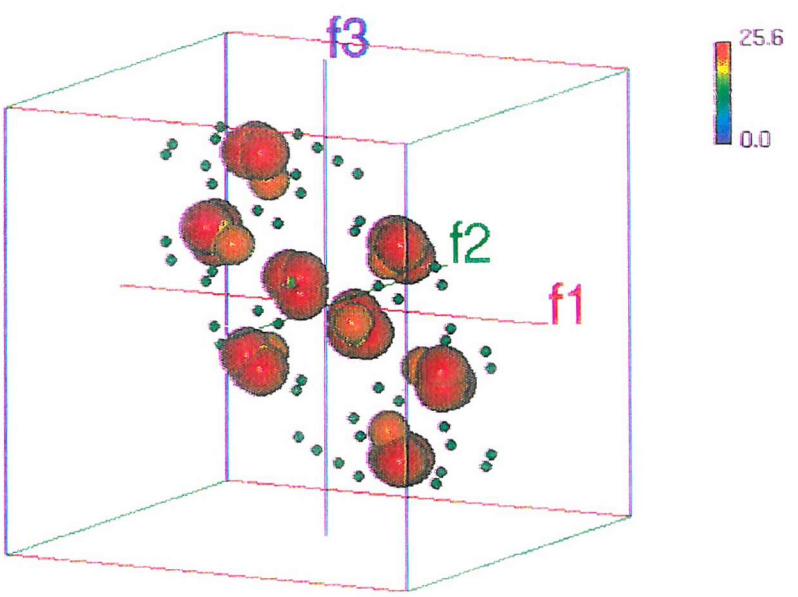

Figure 3.28: Trispectrum of the Duffing oscillator with $\alpha = 0.01$

3.7 Periodic signals

Throughout this chapter examples of either wideband or narrow band, random, stochastic signals have been studied. These are largely the type of signals that are used in the analysis of the experiments in chapter six. However, in the past, higher order spectra have been applied to situations where periodic signals are likely to be encountered [30, 56, 57, 69], for example in some condition monitoring applications where rotating machines are investigated. In this section it will be shown that the higher order spectra of periodic signals take particular forms known as the 'bed of nails' for the bispectrum and 'box of balls' for the trispectrum.

Any periodic signal can be reconstructed by convolving an impulse response function, h(t), with a train of delta functions, i(t). The Fourier transform of the impulse train, I(f), is also an impulse train.

$$I(f) = \frac{1}{\Delta} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{\Delta}\right), \tag{3.27}$$

where Δ is the spacing of the impulses. By substituting this into equation 2.10 for the deterministic bispectrum it can be seen that the bispectrum of an impulse train is a two dimensional impulse train,

$$S_{XXX}(f_1, f_2) = \frac{1}{\Delta^3} \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \delta\left(f_1 - \frac{n_1}{\Delta}, f_2 - \frac{n_2}{\Delta}\right) \tag{3.28}$$

This is known as the 'bed of nails' [32], and is shown in figure 3.29. Substituting equation 3.27 into the cumulant trispectrum, results in a three dimensional impulse train, or the 'box of balls', figure 3.30.

$$S_{XXXX}(f_1, f_2, f_3) = \frac{1}{\Delta^4} \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \sum_{n_3=-\infty}^{\infty} \delta\left(f_1 - \frac{n_1}{\Delta}, f_2 - \frac{n_2}{\Delta}, f_3 - \frac{n_3}{\Delta}\right) \tag{3.29}$$

Using the ideas presented in section 3.6.2 it is possible to show that the bicoherence and tricoherence of any periodic signal results in the 'bed of nails' and 'box of balls'. As convolution in the time domain is equal to multiplication in the frequency domain, any periodic signal can be written as,

$$X(f)=I(f)H(f), \qquad 3.30$$

where H(f) is the Fourier transform of h(t). When X(f) is substituted into equation 3.3 for the bicoherence, the H(f)'s can all be taken outside the expectation operation (as they are time invariant) and cancel and so the bicoherence of any periodic signal is the same as the bicoherence of an impulse train. Care must be taken to ensure ε is included in the denominator term to avoid 0/0 occurring in the bed of nails and leading to spurious results. In the same way, the tricoherence of a periodic signal can always be shown to be the 'box of balls'.

A major limitation on the use of the bicoherence in the detection of quadratic phase coupling has recently been highlighted by Fackrell et al [34] which concerns phase randomisation. For the bicoherence to reliably detect quadratic phase coupling it assumes that the signal component phases are randomised for each segment of the signal. For a random signal this is a valid assumption. However, the bicoherence has been applied in the detection of quadratic phase coupling in signals where such an assumption does not hold, for example, in sine waves. In these cases quadratic phase coupling is an unsuitable measure of nonlinearity as it will arise because the signal is deterministic and not necessarily be due to a nonlinear system.

Although none of the experimental work in chapter six includes periodic signals, it is useful to be able to recognise the 'bed of nails' or 'box of balls' structure if it appears in a higher order spectrum and to understand what type of process could have created it.

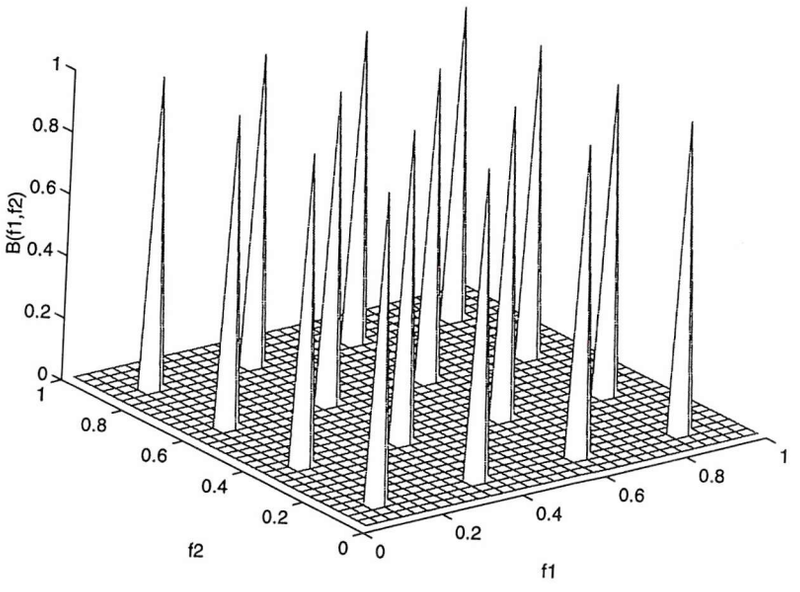

Figure 3.29: The bed of nails

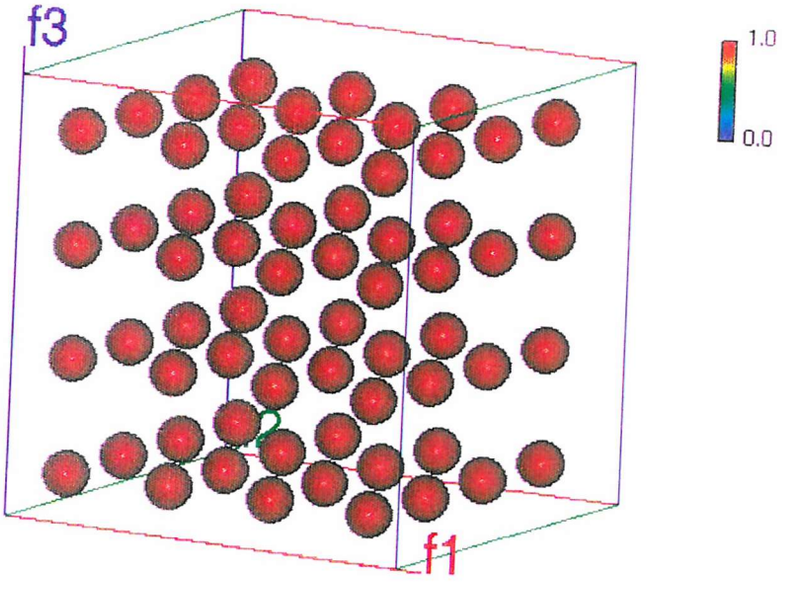

Figure 3.30: The box of balls

3.8 Conclusions

In this chapter a series of normalisation methods for the bispectrum and trispectrum have been presented. Their main uses are shown in table 3.3.

The normalisation processes were then demonstrated with a number of examples using both memoryless and dynamic systems. It was shown how the bicoherence and tricoherence could be used to detect quadratic and cubic nonlinearities respectively. If only an output measurement is available, using auto higher order spectra in this way can give very useful information about the nonlinearities in a signal. However, to find out more about the 'system', both input and output information is needed. It is then possible, with the use of cross higher order spectra to carry out a true system identification, as is discussed in the next chapter.

Method	Uses
Skewness function	third order statistical tests
Kurtosis function	fourth order statistical tests
Bicoherence function	quadratic phase coupling detection
Tricoherence function	cubic phase coupling detection
Pre-whitening technique	cubic phase coupling detection in narrow band or highly resonant systems

Table 3.3: Normalisation methods and their uses

Chapter 4

Cross higher order spectra and the Volterra series

4.1 Introduction

The previous chapters have considered the higher order spectra of a single signal. This could be, for example, the output of a (nonlinear) system. In this chapter, both input and output measurements are used to estimate cross higher order spectra and so identify the nonlinearities in a system. A major assumption made throughout this chapter is that the input to the system is a Gaussian, random signal. A number of methods are proposed, all of which use higher order spectra, and are based on determining the first three Volterra kernels or parts of them. The method of residual spectra is also used and shown to give equivalent results to higher order spectra under certain conditions.

Initially, in section 4.2, cross higher order spectra are introduced. This is followed in section 4.3 by an explanation of the Volterra series. The Volterra series can be thought of as an extension of the linear convolution integral. Rather than calculating the convolution of the linear impulse response of a system with its input, it is possible to consider the infinite sum of higher order impulse responses convolved with interactions of the input. For example, a quadratic Volterra model would consist of the convolution of the linear impulse response of the system with its input, plus the two dimensional convolution of the quadratic impulse response with a quadratic expansion of the input. These higher order impulse response functions are known as Volterra kernels after the mathematician Vito Volterra, who first studied this series. By estimating the Volterra kernels of some nonlinear systems it is possible to characterise their nonlinear response to any input.

The Volterra series is an infinite series which in practice must be truncated. In this work only the first three terms will be considered. The quadratic terms are included to take

account of skewed nonlinearities and the cubic term is included, as it is the first term to take account of any symmetric nonlinearities. If the cubic term was not included, the class of problem which could be solved would be strictly limited to non-symmetric or skewed type nonlinearities. As by truncating the system an approximation is being made, it is important to consider what remains of the signal once the linear, quadratic, and cubic components have been removed. This will consist of nonlinear terms not accounted for by the model and measurement noise. If this remainder is large compared with the linear, quadratic and cubic terms then the approximation is likely to be untrustworthy and it may be necessary to include more terms in the Volterra series expansion.

The aim of this work is to try to find separate expressions for: 1) the quadratic components of a signal using the quadratic transfer function; 2) the cubic component of a signal using the cubic transfer function; 3) the remaining part of the signal which will contain elements due to all higher order nonlinearities and noise. The model can be considered as the parallel connection of a linear, quadratic and cubic system.

The initial model of the system, section 4.4, consists of three linear filters acting on memoryless nonlinear transformations of the input. Specifically, $H_1(f)$, $H_2(f)$ and $H_3(f)$ are convolved with the input signal, $x(t)$, the input signal squared, and the input signal cubed, respectively. This set of inputs is referred to as polynomial inputs. Using higher order cross correlation functions, as in section 4.4.1, expressions for $H_1(f)$, $H_2(f)$ and $H_3(f)$ can be found in terms of the cross higher order spectra.

This model may also be viewed as a multiple input, single output system which, in the past, has been modelled using the concepts of residual spectra and partial and multiple coherence [3, 4, 18, 28]. It is shown in section 4.4.2 that for the polynomial input used above it is possible to obtain expressions for $H_1(f)$, $H_2(f)$ and $H_3(f)$ whether using higher order cross correlation functions and higher order cross spectra or the residual spectra method. These two techniques are compared, in section 4.5, as they each have their own estimation problems associated with them.

Volterra theory is well established in the field of nonlinear system analysis [70] and the polynomial input to the linear filters model is shown to be equivalent to simply considering the main diagonal of the Volterra kernel. This model is extended to a full Volterra type model. Initially, in section 4.6, a quadratic model is assumed, and the Volterra series is truncated at the third order and only the first two kernels are calculated in terms of the auto spectrum, cross spectrum and cross bispectrum. The problems of estimating the linear transfer function and in particular the quadratic transfer function are discussed. The linear kernel is easily estimated using traditional methods, but the variance of the estimate of the second order kernel contains elements due to the signal's spectrum which can corrupt the estimation process. In order to make the variance more uniform across all frequencies a technique is used that removes the linear component from the signal before the quadratic transfer function is calculated. For a symmetric input process, the problem of estimating the linear and quadratic kernels decouples and so subtracting the linear component will not bias the quadratic kernel. A simple example of a signal that contains a linear and quadratic component is used to demonstrate these problems. Finally, the quadratic Volterra model is then extended to a cubic Volterra model (section 4.7). This not only increases the number of terms but also increases the complexity of the system as the estimate of the linear kernel is affected by the third order kernel.

4.2 Cross higher order spectra

In chapters two and three, the concepts of auto higher order spectra were developed. In traditional spectral estimation, if both the input and output of a system, such as in figure 4.1, are available, the cross power spectrum, $S_{XY}(f)$, often contains useful information. This can be defined as,

$$S_{XY}(f) = E[X(f)Y^*(f)] \qquad 4.1$$

This is a partial description of the cross spectrum. A full description is similar to that of the power spectrum as in Appendix B. This is not the standard engineering definition of the

cross spectrum but is the definition most commonly extended to higher order spectra, e.g. [59, 70].

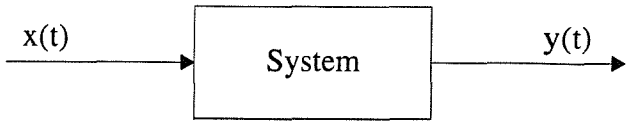

Figure 4.1: A single input single output system

Equation 4.1 can be extended to higher order spectra such that the cross bispectrum, $S_{XXY}(f_1,f_2)$, and moment cross trispectrum, $T_{XXXY}(f_1,f_2,f_3)$, are defined as,

$$S_{XXY}(f_1,f_2) = E[X(f_1)X(f_2)Y^*(f_1+f_2)] \tag{4.2}$$

$$T_{XXXY}(f_1,f_2,f_3) = E[X(f_1)X(f_2)X(f_3)Y^*(f_1+f_2+f_3)] \tag{4.3}$$

All cross higher order spectra are estimated using the natural extensions to the methods described in chapter two, i.e. using the direct method in terms of moments. However, all the theory in this chapter is based on cumulant spectra, therefore care must be taken to ensure that cumulants rather than moments are used when estimating the cross spectra. As with auto cumulants the second and third order cross cumulant functions are the same as the cross moment functions and so this causes no confusion.

$$C_{XY}(\tau) = R_{XY}(\tau) \tag{4.4}$$

$$C_{XXY}(\tau_1,\tau_2) = R_{XXY}(\tau_1,\tau_2) \tag{4.5}$$

However, the fourth order cross cumulant function is equal to the fourth order moment function minus three second order moment functions.

$$\begin{aligned} C_{XXXY}(\tau_1,\tau_2,\tau_3) = R_{XXXY}(\tau_1,\tau_2,\tau_3) - \{&R_{XY}(\tau_1)R_{XY}(\tau_3-\tau_2) + \\ &R_{XY}(\tau_2)R_{XY}(\tau_3-\tau_1) + \\ &R_{XY}(\tau_3)R_{XY}(\tau_2-\tau_1)\} \end{aligned} \tag{4.6}$$

Taking the three dimensional Fourier transform of equation 4.6 gives the cumulant cross trispectrum,

$$S_{xxxy}(f_1,f_2,f_3) = T_{xxxy}(f_1,f_2,f_3) - \{S_{xx}(f_1)S_{xy}(f_2)\delta(f_2+f_3) + S_{xx}(f_2)S_{xy}(f_1)\delta(f_1+f_3) + S_{xx}(f_1)S_{xy}(f_3)\delta(f_1+f_2)\} \quad 4.7$$

To calculate the cumulant cross trispectrum it is necessary to calculate the product of the auto spectrum and cross spectrum along the three planes $f_1+f_2=0$, $f_2+f_3=0$, and $f_1+f_3=0$ and subtract this from the moment trispectrum.

Up to this point conventional definitions of the autocorrelation functions have been used (equations 2.23, 2.24 and 2.25). However from now on it is important to realise that the following definitions are used for the higher order cross moment functions. It is assumed that x(t) and y(t) are zero mean, strictly stationary, random variables.

$$R_{xy}(\tau) = E[x(t-\tau)y(t)] \quad 4.8$$

$$R_{xxy}(\tau_1,\tau_2) = E[x(t-\tau_1)x(t-\tau_2)y(t)] \quad 4.9$$

$$R_{xxxy}(\tau_1,\tau_2,\tau_3) = E[x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)y(t)] \quad 4.10$$

It will be noted that these differ from the more commonly used definitions which are given below (equations 4.11, 4.12 and 4.13), and this in the past has caused confusion [59]. Notice the τ's in equations 4.8, 4.9 and 4.10 represent delays whereas in equations 4.11, 4.12 and 4.13 they are advances. Some authors [59] prefer to use advances which will lead to a conjugate appearing in the final expression for the Volterra kernels. Here, in order to simplify the calculations and remove the conjugate, delays are used. This also, in the context of system identification, has a more intuitive feel as the current output is related to past inputs.

$$\tilde{R}_{XY}(\tau) = E[x(t)y(t+\tau)] \tag{4.11}$$

$$\tilde{R}_{XXY}(\tau_1,\tau_2) = E[x(t)x(t+\tau_1)y(t+\tau_2)] \tag{4.12}$$

$$\tilde{R}_{XXXY}(\tau_1,\tau_2,\tau_3) = E[x(t)x(t+\tau_1)x(t+\tau_2)y(t+\tau_3)] \tag{4.13}$$

Equations 4.9 and 4.10 can be thought of as natural progressions of equation 4.8 and similarly, equations 4.12 and 4.13 are natural progressions of equation 4.11. Although for the second order case, equations 4.8 and 4.11 are equivalent, the same does not hold for higher orders, for example, equation 4.9 is not equivalent to equation 4.12. Fourier transforming equation 4.8, 4.9, and 4.10 leads to the following expressions for the cross spectrum, cross bispectrum, and moment cross trispectrum.

$$S_{XY}(f) = E[X^*(f)Y(f)] \tag{4.14}$$

$$S_{XXY}(f_1,f_2) = E[X^*(f_1)X^*(f_2)Y(f_1+f_2)] \tag{4.15}$$

$$T_{XXXY}(f_1,f_2,f_3) = E[X^*(f_1)X^*(f_2)X^*(f_3)Y(f_1+f_2+f_3)] \tag{4.16}$$

Note, equation 4.14 is a more common definition of the cross spectrum than equation 4.1.

4.3 The Volterra series

For a linear system, the relationship between the input x(t), the system's impulse response h(t) and the output y(t) is given by the convolution integral,

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \tag{4.17}$$

Convolution only gives the relationship between the input and the output for a linear system. For some nonlinear systems this can be extended to the Volterra series [70] which

may be considered as a direct generalisation of the above input/output relationship for a linear system. For a linear system, the output is thought of as being the sum of the response to each elemental input. The Volterra series extends this to include the responses to pairs of elemental inputs, triplets of elemental inputs and so forth. This can be expressed as,

$$\begin{aligned} y(t) = & \int_{-\infty}^{\infty} h_1(\tau_1)\, x(t-\tau_1)d\tau_1 \\ & + \int\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)\, x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 \\ & + \int\int\int_{-\infty}^{\infty} h_3(\tau_1,\tau_2,\tau_3)\, x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)d\tau_1 d\tau_2 d\tau_3 \\ & \qquad \vdots \\ & + \int\cdots\int_{-\infty}^{\infty} h_n(\tau_1,\cdots,\tau_n)x(t-\tau_1)\cdots x(t-\tau_n)d\tau_1\cdots d\tau_n \end{aligned} \qquad 4.18$$

This is often written in the form [70],

$$y(t) = H_1[x(t)] + H_2[x(t)] + \cdots + H_n[x(t)] + \cdots \qquad 4.19$$

where,

$$H_n[x(t)] = \int\cdots\int_{-\infty}^{\infty} h_n(\tau_1,\cdots,\tau_n)x(t-\tau_1)\cdots x(t-\tau_n)\, d\tau_1\cdots d\tau_n \qquad 4.20$$

H_n is called an n^{th} order Volterra operator, and the functions $h_n(\tau_1,...,\tau_n)$ are called the Volterra kernels. Identifying these kernels is analogous to identifying the impulse response for a linear system, although to fully characterise a system one would need to identify a potentially infinite series of kernels. Note that a linear system is just a first order Volterra system and hence only has a first order kernel, $h_1(t_1)$. Fourier transforming the n^{th} order Volterra operator using the n-dimensional Fourier transform gives,

$$H_n(f_1,\cdots,f_n) = \int\cdots\int_{-\infty}^{\infty} h_n(\tau_1,\cdots,\tau_n)e^{-j2\pi(f_1\tau_1+\cdots+f_n\tau_n)}\, d\tau_1\cdots d\tau_n \qquad 4.21$$

which has the inverse transform,

$$h_n(\tau_1,\cdots,\tau_n) = \int\cdots\int_{-\infty}^{\infty} H_n(f_1,\cdots,f_n)e^{j2\pi(f_1\tau_1+\cdots+f_n\tau_n)}\,df_1\cdots df_n \qquad 4.22$$

One can then show that the relationship between the input and the output in the frequency domain for the n^{th} term is,

$$Y_n(f_1,\cdots,f_n) = H_n(f_1,\cdots,f_n)X(f_1)\cdots X(f_n) \qquad 4.23$$

4.4 Linear filters acting on a polynomial input

Initially, rather than considering a full Volterra series model, a more simple idea is used to characterise the system, based on linear filters acting on a polynomial input. Later in the chapter, this model will be shown to be a subset of the Volterra model. Two different methods are presented to solve for the linear, quadratic, and cubic components of the system, one using higher order spectra, and the other based on the ideas of residual spectra [3, 4, 18, 28]

Consider a system with three linear filters $H_1(f)$, $H_2(f)$ and $H_3(f)$, where $H_1(f)$ acts on an input signal, $x(t)$, $H_2(f)$ acts on the input signal squared, $x^2(t)$, and $H_3(f)$ acts on the input signal cubed, $x^3(t)$. This can be viewed as the parallel connection of a linear, quadratic and cubic system (figure 4.2). The output of the system, $y(t)$, is due to the sum of these three components and a term due to noise and terms not included in the nonlinear model.

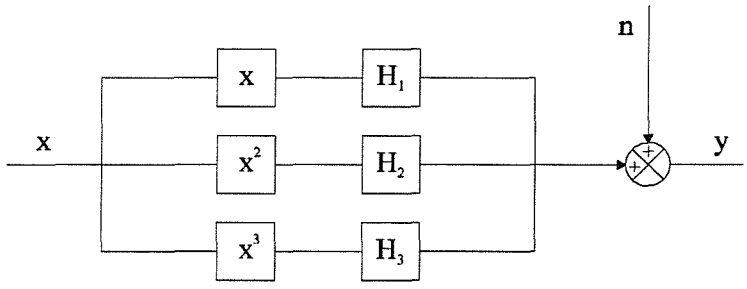

Figure 4.2: Cubic model of linear filters acting on a polynomial input

Assuming a perfect model, the output y(t) can be formed by the convolution of the impulse responses $h_1(t)$, $h_2(t)$ and $h_3(t)$ of the filters with the inputs, x(t), $x^2(t)$ and $x^3(t)$ respectively. Hence,

$$y(t) = \int_{-\infty}^{\infty} h_1(\tau)x(t-\tau)d\tau + \int_{-\infty}^{\infty} h_2(\tau)x^2(t-\tau)d\tau + \int_{-\infty}^{\infty} h_3(\tau)x^3(t-\tau)d\tau \qquad 4.24$$

Equation 4.24 has been dubbed a Hammerstein series by some authors [67].

4.4.1 Higher order spectral method

To characterise the system it is necessary to find expressions for $H_1(f)$, $H_2(f)$, and $H_3(f)$ in terms of higher order spectra. To do this, higher order cross correlation functions are formed between the input and output and then their multiple Fourier transforms taken to form higher order moment spectra. This is done separately for the linear, quadratic and cubic components.

Note, $S_{XY}(f)/S_{XX}(f)$ is often referred to as the estimator $H_1(f)$ which is the optimal linear least squares filter relating x(t) and y(t). This is not to be confused with the $H_1(f)$ used here.

Firstly, y(t) is substituted from equation 4.24 into the cross correlation function,

$$R_{XY}(\tau) = E[x(t-\tau)y(t)] \qquad 4.25$$

to give,

$$\begin{aligned} R_{XY}(\tau) = & \int_{-\infty}^{\infty} h_1(u)\, E[x(t-\tau)x(t-u)]\, du + \\ & \int_{-\infty}^{\infty} h_2(u)\, E[x(t-\tau)x^2(t-u)]\, du + \\ & \int_{-\infty}^{\infty} h_3(u)\, E[x(t-\tau)x^3(t-u)]\, du \end{aligned} \qquad 4.26$$

It is shown in Appendix D and [64, 70, 84] that the expected value of the product of an odd number of Gaussian variables is zero and the product of an even number of Gaussian variables can be decomposed into the sum of all possible combinations of pairs of the products. Specifically the expected value of the product of four Gaussian variables can be written as,

$$E[X_1X_2X_3X_4] = E[X_1X_2]\,E[X_3X_4] + E[X_1X_3]E[X_2X_4] + E[X_1X_4]\,E[X_2X_3] \qquad 4.27$$

The second term of equation 4.26 is zero as it contains the product of three Gaussian variables and the third term can be expressed as the sum of averages of pairs of variables.

$$\begin{aligned} R_{XY}(\tau) = {} & \int_{-\infty}^{\infty} h_1(u)\,E[x(t-\tau)x(t-u)]\,du + \\ & \int_{-\infty}^{\infty} h_3(u)\,\{E[x(t-\tau)x(t-u)].E[x(t-u)x(t-u)] + \\ & \qquad E[x(t-\tau)x(t-u)].E[x(t-u)x(t-u)] + \\ & \qquad E[x(t-\tau)x(t-u)].E[x(t-u)x(t-u)]\}\,du \end{aligned} \qquad 4.28$$

Expressing this in terms of auto correlation functions gives,

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h_1(u)\,R_{XX}(\tau-u)\,du + 3\sigma_x^{\,2} \int_{-\infty}^{\infty} h_3(u) R_{XX}(\tau-u)\,du \qquad 4.29$$

Taking the Fourier transform with respect to τ gives,

$$S_{XY}(f) = H_1(f)S_{XX}(f) + 3\sigma_x^{\,2}\,H_3(f)S_{XX}(f) \qquad 4.30$$

so that,

$$\frac{S_{XY}(f)}{S_{XX}(f)} = H_1(f) + 3\sigma_x^{\,2}\,H_3(f) \qquad 4.31$$

Note the ratio, $S_{XY}(f)/S_{XX}(f)$, contains terms due to both $H_1(f)$ and $H_3(f)$, so that if this ratio was used as the optimal least squares filter linking x(t) and y(t) then clearly $H_3(f)$ will influence the result.

Next, y(t) is substituted from equation 4.24 into the second order cumulant function,

$$R_{XXY}(\tau_1,\tau_2) = E[x(t-\tau_1)x(t-\tau_2)y(t)] \tag{4.32}$$

to give,

$$\begin{aligned} R_{XXY}(\tau_1,\tau_2) = & \int_{-\infty}^{\infty} h_1(u)E[x(t-\tau_1)x(t-\tau_2)x(t-u)]\,du + \\ & \int_{-\infty}^{\infty} h_2(u)E[x(t-\tau_1)x(t-\tau_2)x^2(t-u)]\,du + \\ & \int_{-\infty}^{\infty} h_3(u)E[x(t-\tau_1)x(t-\tau_2)x^3(t-u)]\,du \end{aligned} \tag{4.33}$$

As the expected value of the product of an odd number of Gaussian variables is zero, the first and third order terms are zero. This leaves,

$$\begin{aligned} R_{XXY}(\tau_1,\tau_2) = \int_{-\infty}^{\infty} h_2(u) \{ & E[x(t-\tau_1)x(t-\tau_2)].E[x(t-u)x(t-u)] + \\ & E[x(t-\tau_1)x(t-u)].E[x(t-\tau_2)x(t-u)] + \\ & E[x(t-\tau_2)x(t-u)].E[x(t-\tau_1)x(t-u)]\}\,du \end{aligned} \tag{4.34}$$

which can be written in terms of autocorrelation functions to give,

$$R_{XXY}(\tau_1,\tau_2) = \sigma_x^2 \int_{-\infty}^{\infty} h_2(u)R_{XX}(\tau_1-\tau_2)\,du + 2\int_{-\infty}^{\infty} h_2(u)R_{XX}(\tau_1-u)R_{XX}(\tau_2-u)\,du \tag{4.35}$$

Taking the double Fourier transform with respect to τ_1 and τ_2 gives,

$$S_{XXY}(f_1,f_2) = \sigma_x^2 S_{XX}(f_1)\delta(f_1+f_2)\int_{-\infty}^{\infty} h_2(u)\,du + 2S_{XX}(f_1)S_{XX}(f_2)H_2(f_1+f_2) \tag{4.36}$$

where $S_{XXY}(f_1,f_2)$ is the cross bispectrum. The first term contains the delta function $\delta(f_1+f_2)$ and so if $(f_1+f_2) \neq 0$ then,

$$S_{XXY}(f_1,f_2) = 2S_{XX}(f_1)S_{XX}(f_2)H_2(f_1+f_2) \qquad 4.37$$

Finally, y(t), is substituted from equation 4.24 into the third order cumulant function,

$$R_{XXXY}(\tau_1,\tau_2,\tau_3) = E[x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)y(t)] \qquad 4.38$$

to give,

$$\begin{aligned} R_{XXXY}(\tau_1,\tau_2,\tau_3) = & \int_{-\infty}^{\infty} h_1(u)\, E[x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)x(t-u)]\, du + \\ & \int_{-\infty}^{\infty} h_2(u)\, E[x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)x^2(t-u)]\, du + \\ & \int_{-\infty}^{\infty} h_3(u)\, E[x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)x^3(t-u)]\, du \end{aligned} \qquad 4.39$$

As the expected value of the product of an odd number of Gaussian variables is zero, the second term is zero. This leaves,

$$\begin{aligned} R_{XXXY}(\tau_1,\tau_2,\tau_3) = & \int_{-\infty}^{\infty} h_1(u)\, E[x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)x(t-u)]\, du + \\ & \int_{-\infty}^{\infty} h_3(u)\, E[x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)x^3(t-u)]\, du \end{aligned} \qquad 4.40$$

Expanding the first term of equation 4.40 and rewriting in terms of auto correlation functions gives,

$$\begin{aligned} \int_{-\infty}^{\infty} h_1(u)\, E[x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)x(t-u)]\, du = & \int_{-\infty}^{\infty} h_1(u)\, \{R_{XX}(\tau_1-\tau_2)R_{XX}(\tau_3-u) + \\ & R_{XX}(\tau_1-\tau_3)R_{XX}(\tau_2-u) + \\ & R_{XX}(\tau_2-\tau_3)R_{XX}(\tau_1-u)\}\, du \end{aligned} \qquad 4.41$$

The triple Fourier transform of the first part of 4.41 is as follows,

$$\iiint_{-\infty}^{\infty} h_1(u) R_{xx}(\tau_3 - u) R_{xx}(\tau_1 - \tau_2)\, e^{-j2\pi f_1 \tau_1} e^{-j2\pi f_2 \tau_2} e^{-j2\pi f_3 \tau_3} d\tau_1 d\tau_2 d\tau_3 du$$

Substituting, $\theta = \tau_1 - \tau_2$, $d\tau_1 = d\theta$ gives $\int_{-\infty}^{\infty} R_{xx}(\theta) e^{-j2\pi f_1(\theta+\tau_2)} d\theta = S_{xx}(f_1) e^{-2\pi\tau_2 f_1}$

Substituting, $\vartheta = \tau_3 - u$, $d\tau_3 = d\vartheta$ gives $\int_{-\infty}^{\infty} R_{xx}(\vartheta) e^{-j2\pi f_3(\vartheta+u)} d\vartheta = S_{xx}(f_3) e^{-2\pi f_3 u}$

$$= S_{xx}(f_1) S_{xx}(f_3) \int_{-\infty}^{\infty} h_1(u) e^{-j2\pi f_3 u} du \int_{-\infty}^{\infty} e^{-j2\pi\tau_2(f_1+f_2)} d\tau_2$$

$$= S_{xx}(f_1) S_{xx}(f_3) \delta(f_1 + f_2) \int_{-\infty}^{\infty} h_1(u) e^{-j2\pi f_3 u} du \qquad 4.42$$

Equation 4.42 contains the delta function, $\delta(f_1+f_2)$, and so is equal to zero if $(f_1+f_2) \neq 0$. Similarly the second and third parts of equation 4.41 are equal to zero if $(f_1+f_3) \neq 0$ and $(f_2+f_3) \neq 0$.

The second term of equation 4.40 contains the product of six Gaussian variables. There are fifteen possible ways of arranging three pairs but as three variables are the same there are only four different combinations.

$$\begin{aligned} E[X_1X_2X_3X_4X_4X_4] = {} & 3\,E[X_1X_2]\,E[X_3X_4]\,E[X_4X_4] + \\ & 3\,E[X_1X_3]\,E[X_2X_4]\,E[X_4X_4] + \\ & 3\,E[X_1X_4]\,E[X_2X_3]\,E[X_4X_4] + \\ & 6\,E[X_1X_4]\,E[X_2X_4]\,E[X_3X_4] \end{aligned} \qquad 4.43$$

Therefore, the second term in equation 4.40 can be written as,

$$\begin{aligned} & \int_{-\infty}^{\infty} h_3(u) E[x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)x^3(t-u)]du \\ & = \int_{-\infty}^{\infty} h_3(u)\{(3\sigma_x^2 R_{xx}(\tau_1 - \tau_2)R_{xx}(\tau_3 - u) + \\ & \qquad 3\sigma_x^2 R_{xx}(\tau_1 - \tau_3)R_{xx}(\tau_2 - u) + \\ & \qquad 3\sigma_x^2 R_{xx}(\tau_3 - \tau_2)R_{xx}(\tau_1 - u) + \\ & \qquad 6R_{xx}(\tau_1 - u)R_{xx}(\tau_2 - u)R_{xx}(\tau_3 - u)\}\, du \end{aligned} \qquad 4.44$$

The first three terms are the same integration as equation 4.42 but with $h_1(u)$ replaced by $h_3(u)$ and so contain delta functions which are equal to zero if $(f_1+f_3) \neq 0$, $(f_2+f_3) \neq 0$, and $(f_1+f_2) \neq 0$.

The triple Fourier transform of the last term is:

$$
\begin{aligned}
&6\int\int\int_{-\infty}^{\infty} h_3(u).R_{xx}(\tau_1 - u)\,R_{xx}(\tau_2 - u)R_{xx}(\tau_3 - u)e^{-j2\pi f_1\tau_1}e^{-j2\pi f_2\tau_2}e^{-j2\pi f_3\tau_3}d\tau_1 d\tau_2 d\tau_3 du \\
&= 6S_{xx}(f_1)S_{xx}(f_2)S_{xx}(f_3)\int_{-\infty}^{\infty} h_3(u)e^{-j2\pi u(f_1+f_2+f_3)}du \\
&= 6S_{xx}(f_1)S_{xx}(f_2)S_{xx}(f_3)\,H_3(f_1+f_2+f_3)
\end{aligned}
\qquad 4.45
$$

From equations 4.31, 4.37 and 4.45, $H_1(f)$, $H_2(f)$ and $H_3(f)$ can now be expressed in terms of cross higher order spectra and the input auto spectrum.

$$H_1(f) = \frac{S_{xy}(f)}{S_{xx}(f)} - 3\sigma_x{}^2 H_3(f) \qquad 4.46$$

$$H_2(f) = \frac{S_{xxy}(f_1, f - f_1)}{2S_{xx}(f_1)S_{xx}(f - f_1)} \qquad 4.47$$

$$H_3(f) = \frac{S_{xxxy}(f_1, f_2, f - f_1 - f_2)}{6S_{xx}(f_1)S_{xx}(f_2)S_{xx}(f - f_1 - f_2)} \qquad 4.48$$

Equations 4.47 and 4.48 imply that there must be symmetry in both the bispectrum and trispectrum. In equation 4.47, f_1 can be arbitrarily chosen and so, for example, setting $f_1 = f/2$ gives:

$$H_2(f) = \frac{S_{xxy}\left(\frac{f}{2}, \frac{f}{2}\right)}{2S_{xx}\left(\frac{f}{2}\right)S_{xx}\left(\frac{f}{2}\right)} \qquad 4.49$$

$S_{XXY}\left(\frac{f}{2},\frac{f}{2}\right)$ represents the leading diagonal of the bispectrum in the (f_1,f_2) space. Setting f_1 to a different value and evaluating $S_{XXY}(f_1,f-f_1)$ over another diagonal will result in the same information being obtained as all the other terms in $S_{XXY}(f_1,f-f_1)$ should consist of symmetries of the leading diagonal. A similar result occurs with equation 4.48 except rather than considering diagonal lines, diagonal planes in the (f_1,f_2,f_3) space should be considered.

4.4.2 Residual Spectral Method

Suppose a series of random processes that are thought to be related are measured. When a single component of a random process is influenced by others any distinct relationship connecting it with any one of them is likely to be obscured by the action of the remainder. It is only by removing the effects of the remainder in some way that the connection can be established. For example, if the processes x(t) and y(t) are measured, it is easy to form $x(t)^2$ and $x(t)^3$ and the method of residual spectra [3, 4, 18, 28] can then be used to try and find the connections between x(t), $x(t)^2$, $x(t)^3$ and y(t).

The basic building block for residual spectra is shown in figure 4.3. Assuming x_1 and x_2 are two measured random processes [2], x_1 can be split into two parts: that fully coherent with x_2, that is the part of x_1 that can be accounted for by the optimal linear operation (in the least squares sense) on x_1 through the filter L_1, and the part of x_1 that is not coherent with x_2 termed $x_{2.1}$.

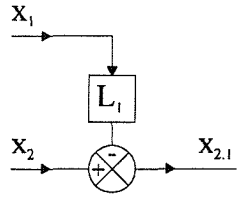

Figure 4.3: The basic building block for residual spectra

[2] To simplify notation, the time and frequency arguments on the signal notation are temporarily suspended.

A number of these simple building blocks can be cascaded together to form more complex systems such as that in figure 4.4 where the four processes, x_1, x_2, x_3, and x_4 are considered. In this case, at the first stage, x_2, x_3 and x_4 are each split into two components: the part accounted for by the optimal linear operation on x_1 through the filters L_1, L_2, and L_3, and the parts that are not coherent with x_1 which are termed $x_{2.1}$ $x_{3.1}$ and $x_{4.1}$. These are the parts of x_2 x_3 and x_4 left over when the linear effect of x_1 is removed and are termed residual random variables.

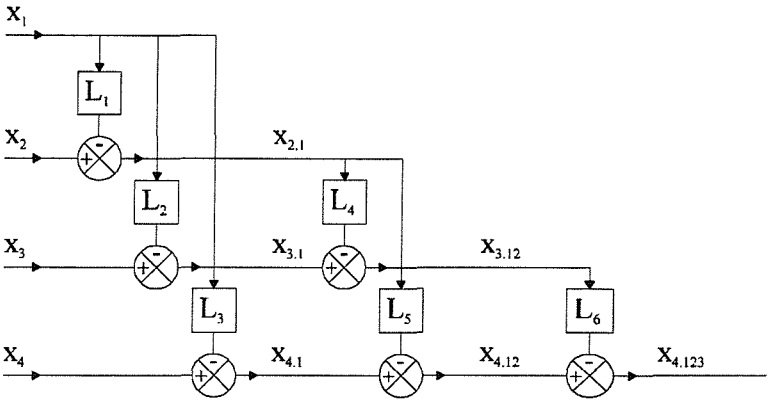

Figure 4.4: A multiple input, single output system

At the second stage $x_{3.1}$ and $x_{4.1}$ can be further conditioned with respect to $x_{2.1}$ through filters L_4 and L_5 to give $x_{3.12}$ and $x_{4.12}$, where $x_{3.12}$ is that part of x_3 that is not coherent with both x_1 and x_2, and, $x_{4.12}$ is that part of x_4 which is not coherent with x_1 and x_2. L_4 is the optimal linear filter relating $x_{2.1}$ to $x_{3.1}$ and L_5 the optimal linear filter that relates $x_{2.1}$ to $x_{4.1}$. At the final stage it is possible to condition $x_{3.12}$ with respect to $x_{4.12}$ to produce $x_{4.123}$, the part of the signal that is not coherent with any of the inputs. L_6 is the optimal linear filter relating $x_{3.12}$ to $x_{4.12}$.

The optimal linear filter L_1 can be shown to be,

$$L_1 = \frac{S_{12}}{S_{11}} \qquad 4.50$$

where $S_{11} = S_{x_1x_1}(f)$ etc. That is, L_1 is equal to the cross spectrum of x_1 and x_2 divided by the auto spectrum of x_1. L_2 to L_6 are defined similarly as follows,

$$L_1 = \frac{S_{12}}{S_{11}}$$

$$L_2 = \frac{S_{13}}{S_{11}} \qquad L_4 = \frac{S_{23.1}}{S_{22.1}} \qquad 4.51$$

$$L_3 = \frac{S_{14}}{S_{11}} \qquad L_5 = \frac{S_{24.1}}{S_{22.1}} \qquad L_6 = \frac{S_{34.12}}{S_{33.12}}$$

In section 4.4.1, expressions for H_1, H_2, and H_3 were obtained from,

$$y(t) = \int_{-\infty}^{\infty} h_1(\tau)x(t-\tau)d\tau + \int_{-\infty}^{\infty} h_2(\tau)x^2(t-\tau)d\tau + \int_{-\infty}^{\infty} h_3(\tau)x^3(t-\tau)d\tau \qquad 4.52$$

which can be expressed in the frequency domain as,

$$Y = H_1X + H_2X^2 + H_3X^3 \qquad 4.53$$

where X^2 is defined as the Fourier transform of the input signal squared and X^3 is defined as the Fourier transform of the input signal cubed. If the four random processes x_1, x_2, x_3 and x_4, are now chosen to be the three inputs to the system x, x^2, x^3 and the output y, it should be possible to find similar expressions for $H_1(f)$, $H_2(f)$, and $H_3(f)$ using residual spectra.

It is important to realise that L_1, L_2 and L_3 are least squares optimal filters and not $H_1(f)$, $H_2(f)$ and $H_3(f)$. In order to find $H_1(f)$, $H_2(f)$ and $H_3(f)$ an expression for the output X_4 in terms of the three inputs X_1, X_2 and X_3 must be found.

From the flow logic of figure 4.4,

$$X_4 = L_3X_1 + L_5X_{2.1} + L_6X_{3.12} + X_{4.123} \qquad 4.54$$

$$X_{2.1} = X_2 - L_1X_1 \qquad 4.55$$

$$X_{3.12} = X_{3.1} - L_4X_{2.1} = X_3 - L_2X_1 - L_4X_2 + L_1L_4X_1 \qquad 4.56$$

Substituting 4.54 and 4.55 into 4.56 gives,

$$X_4 = (L_3 - L_1L_5 - L_2L_6 + L_1L_4L_6)X_1 + (L_5 - L_6L_4)X_2 + (L_6)X_3 + X_{4.123} \quad 4.57$$

So expressions for $H_1(f)$, $H_2(f)$ and $H_3(f)$ can be written in terms of L_1 to L_6 as,

$$H_1 = L_3 - L_1L_5 - L_2L_6 + L_1L_4L_6 \quad 4.58$$

$$H_2 = L_5 - L_6L_4 \quad 4.59$$

$$H_3 = L_6 \quad 4.60$$

By substituting y, x, x^2, x^3 for x_1, x_2, x_3, x_4 and assuming that x is a Gaussian process, expressions for L_1 to L_6 are calculated as follows:

$$L_1 = \frac{S_{12}}{S_{11}} = \frac{\mathcal{F}\{E[x(t)x^2(t+\tau)]\}}{\mathcal{F}\{E[x(t)x(t+\tau)]\}} = 0 \quad 4.61$$

where $\mathcal{F}$ denotes the Fourier transform. The numerator $\mathcal{F}\{E[x(t)x^2(t+\tau)]\}$ is the product of an odd number of Gaussian variables and so equal to zero, hence L_1=0.

$$L_2 = \frac{S_{13}}{S_{11}} = \frac{\mathcal{F}\{E[x(t)x^3(t+\tau)]\}}{S_{xx}} = \frac{\mathcal{F}[3\sigma_x^2 R_{xx}(\tau)]}{S_{xx}} = \frac{3\sigma_x^2 S_{xx}}{S_{xx}} = 3\sigma_x^2 \quad 4.62$$

$$L_3 = \frac{S_{14}}{S_{11}} = \frac{\mathcal{F}\{E[x(t)y(t+\tau)]\}}{S_{xx}(f)} \quad 4.63$$

If equation 4.52 is used to substitute for $y(t+\tau)$ into equation 4.63, the term containing $h_2(t)$ will contain the product of an odd number of Gaussian variables and so be equal to zero. Hence,

$$L_3 = \frac{\mathcal{F}\left[\int_{-\infty}^{\infty} h_1(u)R_{xx}(\tau - u)du + 3\sigma_x^2 \int_{-\infty}^{\infty} h_3(u)R_{xx}(\tau - u)du\right]}{S_{xx}(f)}$$

$$= \frac{H_1(f)S_{xx}(f) + H_3(f)S_{xx}(f)}{S_{xx}(f)} = H_1(f) + 3\sigma_x^2 H_3(f) \qquad 4.64$$

$$L_4 = \frac{S_{23.1}}{S_{22.1}} \qquad 4.65$$

$S_{23.1}$ can be expanded by writing it as the average of the conjugate of $X_{2.1}$ times $X_{3.1}$. Expressions for $X_{2.1}$ and $X_{3.1}$ can then be substituted as follows,

$$\begin{aligned} S_{23.1} &= E[X_{2.1}^* X_{3.1}] = E[(X_2^* - L_1^* X_1^*)(X_3 - L_2 X_1)] \\ &= S_{23} + L_1^* L_2 S_{11} - L_1^* S_{13} - L_2 S_{21} \\ &= S_{23} + \frac{S_{21}}{S_{11}}\frac{S_{13}}{S_{11}}S_{11} - \frac{S_{21}}{S_{11}}S_{13} - \frac{S_{13}}{S_{11}}S_{21} = S_{23} - \frac{S_{21}S_{13}}{S_{11}} \end{aligned} \qquad 4.66$$

$S_{22.1}$ can be found similarly and substituted into equation 4.65. S_{21} and S_{23} contain the product of three Gaussian variables and so are equal to zero. Hence the expression for L_4 is:

$$L_4 = \frac{S_{23.1}}{S_{22.1}} = \frac{S_{23} - \frac{S_{21}S_{13}}{S_{11}}}{S_{22} - \frac{S_{21}S_{12}}{S_{11}}} = \frac{S_{11}S_{23} - S_{21}S_{13}}{S_{22}S_{11} - S_{21}S_{12}} = \frac{S_{23}}{S_{22}} = 0 \qquad 4.67$$

$$L_5 = \frac{S_{24.1}}{S_{22.1}} = \frac{S_{24} - \frac{S_{21}S_{14}}{S_{11}}}{S_{22} - \frac{S_{21}S_{12}}{S_{11}}} = \frac{S_{24}}{S_{22}} = \frac{\mathcal{F}\{E[x^2(t)y(t+\tau)]\}}{\mathcal{F}\{E[x^2(t)x^2(t+\tau)]\}} \qquad 4.68$$

If $y(t+\tau)$ is substituted from equation 4.52, the terms containing $h_1(t)$ and $h_3(t)$ will consist of the product of an odd number of Gaussian variables and so are equal to zero. Hence,

$$L_5 = \frac{\mathcal{F}\left[\int_{-\infty}^{\infty} h_2(u)\,E[x^2(t)x^2(t+\tau-u)]\,du\right]}{\mathcal{F}\{E[x^2(t)x^2(t+\tau)]\}}$$

$$= \frac{\mathcal{F}\left[\sigma_x{}^4\int_{-\infty}^{\infty} h_2(u)du + 2\int_{-\infty}^{\infty} h_2(u)R_{xx}{}^2(\tau-u)du\right]}{\mathcal{F}\left[\sigma_x{}^4 + R_{xx}{}^2(\tau-u)\right]}$$

$$= \frac{\sigma_x{}^4\alpha_h\delta(f) + 2H_2(f)[S_{xx}(f) * S_{xx}(f)]}{\sigma_x{}^4\delta(f) + 2[S_{xx}(f) * S_{xx}(f)]} = H_2(f) \quad \text{if } f \neq 0 \qquad 4.69$$

where $\alpha_h = \int_{-\infty}^{\infty} h_2(u)du$ and is independent of f.

$$L_6 = \frac{S_{34.12}}{S_{33.12}} = \frac{X_{3.12}{}^{*}X_{4.12}}{X_{3.12}{}^{*}X_{3.12}}$$

From the flow logic of figure 4.4 and $L_1 = L_4 = 0$, expressions for $X_{3.12}$ and $X_{4.12}$ follow,

$$X_{3.12} = X_3 - L_2X_1 \qquad 4.70$$

$$X_{4.12} = X_4 - L_3X_1 - L_5X_2 \qquad 4.71$$

Substituting 4.70 and 4.71 into 4.69 gives,

$$L_6 = \frac{(X_3^{*} - L_2{}^{*}X_1^{*})(X_4 - L_3X_1 - L_5X_2)}{(X_3^{*} - L_2{}^{*}X_1^{*})(X_3 - L_2X_1)}$$

$$= \frac{S_{34} - L_3S_{31} - L_2{}^{*}S_{14} + L_2{}^{*}L_3S_{11}}{S_{33} - L_2{}^{*}S_{13} - L_2{}^{*}S_{31} + L_2{}^{*}L_2S_{11}} \qquad 4.72$$

Substituting for L_2 and L_3 gives,

$$L_6 = \frac{S_{34} - \frac{S_{31}S_{14}}{S_{11}}}{S_{33} - \frac{S_{31}S_{13}}{S_{11}}} = \frac{S_{34} - 3\sigma_x{}^2S_{14}}{S_{33} - 3\sigma_x{}^2S_{13}} \qquad 4.73$$

If the spectral terms in equation are expanded and y(t+τ) is substituted from equation 4.52, the term containing $h_2(t)$ will consist of the product of an odd number of Gaussian variables and so is equal to zero. Hence,

$$L_6 = \frac{\mathcal{F}\left[\begin{array}{l}\int_{-\infty}^{\infty} h_1(u)E[x^3(t)x(t+\tau-u)]\,du + \int_{-\infty}^{\infty} h_3(u)E[x^3(t)x^3(t+\tau-u)]\,du \\ \qquad -3\sigma_x^2\int_{-\infty}^{\infty} h_1(u)E[x(t)x(t+\tau-u)]\,du - \int_{-\infty}^{\infty} h_3(u)E[x(t)x^3(t+\tau-u)]\,du\end{array}\right]}{\mathcal{F}\left[E[x^3(t)x^3(t+\tau)] - 3\sigma_x^2 E[x(t)x^3(t+\tau)]\right]}$$

$$= \frac{\mathcal{F}\left[\begin{array}{l}3\sigma_x^2\int_{-\infty}^{\infty} h_1(u)R_{XX}(\tau-u)du + 9\sigma_x^4\int_{-\infty}^{\infty} h_3(u)du + 6\int_{-\infty}^{\infty} h_3(u)R_{XX}^3(\tau-u)\,du \\ \qquad -3\sigma_x^2\int_{-\infty}^{\infty} h_1(u)R_{XX}(\tau-u)du - 9\sigma_x^4\int_{-\infty}^{\infty} h_3(u)R_{XX}(\tau-u)du\end{array}\right]}{\mathcal{F}\left[9\sigma_x^4 R_{XX}(\tau) + 6R_{XX}^3(\tau) - 9\sigma_x^4 R_{XX}(\tau)\right]}$$

$$= \frac{\mathcal{F}\left[6\int_{-\infty}^{\infty} h_3(u)R_{XX}^3(\tau-u)\,du\right]}{\mathcal{F}\left[6R_{XX}^3(\tau)\right]} = H_3(f) \qquad 4.74$$

Thus, the equations for L_1 to L_6 are,

$$L_1 = 0 \qquad 4.75$$

$$L_2 = 3\sigma_x^2 \qquad 4.76$$

$$L_3 = H_1 + 3\sigma_x^2 H_3 \qquad 4.77$$

$$L_4 = 0 \qquad 4.78$$

$$L_5 = H_2 \qquad 4.79$$

$$L_6 = H_3 \qquad 4.80$$

With these expressions for L_1 to L_6 figure 4.4 can be redrawn for a system with a Gaussian input as figure 4.5. For such a system, to find $H_1(f)$, $H_2(f)$ and $H_3(f)$ only L_3, L_5 and L_6 need to be calculated.

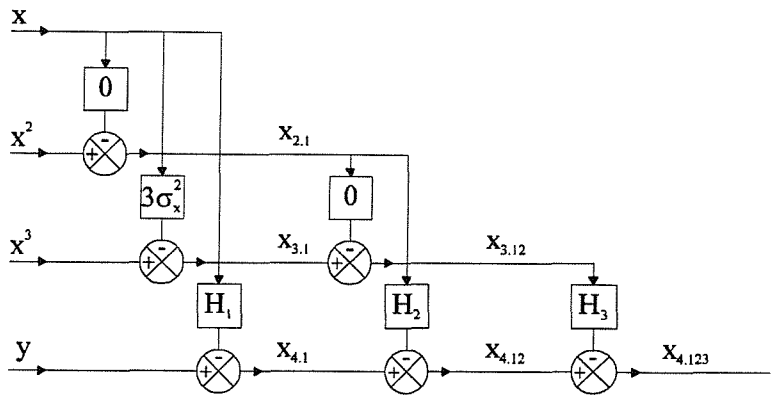

Figure 4.5: Gaussian input to multiple input, single output system

4.5 Relationship between the linear filters acting on a polynomial input and the Volterra series

It has been shown how it is possible to find expressions for $H_1(f)$, $H_2(f)$ and $H_3(f)$ by either using a higher order spectral approach or by using residual spectra and both methods have their advantages and disadvantages.

The main difference between the methods is that the residual spectral method has the advantage of allowing the inputs to be arbitrary functions of x(t) and not strictly limited to simple powers of x(t) as with the higher order spectral method. Hence, with the residual spectral method, if there is prior knowledge of the type of nonlinearity it is possible to construct a model based on that knowledge, as in [28]. For example if $\sqrt{x(t)}$ is known to be important it can be used as an input.

To determine $H_1(f)$, $H_2(f)$ and $H_3(f)$ by the residual spectra method for a Gaussian input only L_3, L_5 and L_6 need to be calculated. These are one dimensional spectra and so can be calculated by standard methods. If the higher order spectral approach is used both the bispectrum and trispectrum must be calculated which requires two and three dimensional matrix calculations which can be very computationally expensive. Also, considerably longer lengths of data may be required when calculating higher order spectra than with ordinary spectra as more averaging is required due to the large variance terms.

However, in this section, the overall problem has been simplified by assuming $H_1(f)$, $H_2(f)$ and $H_3(f)$ are linear filters acting on a polynomial input. It is possible to write the expression for the linear filters model in the form of the Volterra series as,

$$\begin{aligned} y(t) = & \int_{-\infty}^{\infty} h_1(\tau_1)\, x(t-\tau_1)d\tau_1 \\ & + \int\int_{-\infty}^{\infty} h_2(\tau_1)\, \delta(\tau_1 - \tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 \\ & + \int\int\int_{-\infty}^{\infty} h_3(\tau_1)\, \delta(\tau_1 - \tau_2)\delta(\tau_1 - \tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)d\tau_1 d\tau_2 d\tau_3 \end{aligned} \qquad 4.81$$

If this is compared with a cubic Volterra series model,

$$\begin{aligned} y(t) = & \int_{-\infty}^{\infty} h_1(\tau_1)\, x(t-\tau_1)d\tau_1 \\ & + \int\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)\, x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 \\ & + \int\int\int_{-\infty}^{\infty} h_3(\tau_1,\tau_2,\tau_3)\, x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)d\tau_1 d\tau_2 d\tau_3 \end{aligned} \qquad 4.82$$

it can be seen that Volterra kernels are equivalent to the linear filters multiplied by delta functions. That is,

$$h_1(\tau_1) \equiv h_1(\tau_1) \qquad 4.83$$

$$h_2(\tau_1,\tau_2) \equiv h_2(\tau_1)\delta(\tau_1 - \tau_2) \qquad 4.84$$

$$h_3(\tau_1,\tau_2,\tau_3) \equiv h_3(\tau_1)\delta(\tau_1 - \tau_2)\delta(\tau_1 - \tau_3) \qquad 4.85$$

It can now be seen that the linear filters calculated in this section, $H_1(f)$, $H_2(f)$ and $H_3(f)$, are equivalent to considering only the main diagonal of the frequency domain Volterra kernels $H_1(f)$, $H_2(f_1,f_2)$ and $H_3(f_1,f_2,f_3)$. So in using residual spectra to solve this type of nonlinear identification problem, much of the information in the data is being ignored and it is possible that the problem is being over simplified.

4.6 Quadratic Volterra model

Rather than just considering the main diagonal of the Volterra kernels, the model is now extended to try and identify the whole of the Volterra kernels. Suppose a stationary, zero mean signal, x(t), is acted on by a nonlinear system which can be described by a 2nd order Volterra model.

$$y(t) = \int_{-\infty}^{\infty} h_1(\tau_1)\, x(t-\tau_1)d\tau_1 + \int\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)\, x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 \qquad 4.86$$

To characterise the system it is necessary to determine the first order impulse response or kernel, $h_1(\tau)$ and the two dimensional impulse response or second order kernel, $h_2(\tau_1,\tau_2)$. This can be viewed as the parallel connection of a linear and quadratic system (figure 4.6).

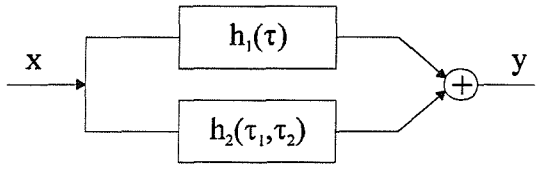

Figure 4.6: Quadratic Volterra model

In all of the following a Gaussian input to the system is assumed. The cross correlation function is formed,

$$R_{XY}(\tau) = E[x(t-\tau)y(t)] \qquad 4.87$$

Substituting for y(t) from equation 4.86 gives,

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h_1(u)E[x(t-\tau)x(t-u)]\,du + \int\int_{-\infty}^{\infty} h_2(u,v)E[x(t-\tau)x(t-u)x(t-v)]\,du\,dv \qquad 4.88$$

As the expected value of the product of an odd number of Gaussian variables is zero, the second order term is zero. This leaves the first order linear term,

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h_1(u)\, R_{XX}(\tau - u)\, du \qquad 4.89$$

Taking the Fourier transform with respect to τ gives,

$$S_{XY}(f) = H_1(f)\, S_{XX}(f) \qquad 4.90$$

Hence, if the input is Gaussian then the linear kernel can be estimated by use of standard cross correlation methods.

To identify the quadratic component of the signal, the second order cross correlation function or second order cumulant function of the signal is formed,

$$R_{XXY}(\tau_1, \tau_2) = E[x(t - \tau_1)x(t - \tau_2)y(t)] \qquad 4.91$$

Substituting for y(t) from equation 4.86 gives,

$$\begin{aligned} R_{XXY}(\tau_1, \tau_2) = & \int_{-\infty}^{\infty} h_1(u)\, E[x(t - \tau_1)x(t - \tau_2)x(t - u)]\, du + \\ & \int\int_{-\infty}^{\infty} h_2(u, v)\, E[x(t - \tau_1)x(t - \tau_2)x(t - u)x(t - v)]\, du\, dv \end{aligned} \qquad 4.92$$

As the expected value of the product of an odd number of Gaussian variables is zero, the first order term is zero. This leaves just the second order term which contains the product of four Gaussian variables and so can be written as the sum of three pairs of products. Hence 4.92 can be written in terms of three pairs of auto correlation functions.

$$\begin{aligned} R_{XXY}(\tau_1, \tau_2) = \int\int_{-\infty}^{\infty} h_2(u, v) \{ & R_{XX}(\tau_1 - \tau_2).R_{XX}(v - u) + \\ & R_{XX}(\tau_1 - u)R_{XX}(\tau_2 - v) + \\ & R_{XX}(\tau_1 - v)R_{XX}(\tau_2 - u)\}\, du\, dv \end{aligned} \qquad 4.93$$

Taking the first term of 4.93 and calculating the double Fourier transform with respect to τ_1 and τ_2 gives,

$$\begin{aligned}
&\int\int\int_{-\infty}^{\infty} h_2(u,v)\, R_{xx}(\tau_1-\tau_2) R_{xx}(v-u)\, e^{-j2\pi f_1\tau_1} e^{-j2\pi f_2\tau_2}\, d\tau_1 d\tau_2 du\, dv \\
&= S_{xx}(f_1)\delta(f_1+f_2)\int_{-\infty}^{\infty} h_2(u,v)\, R_{xx}(v-u)\, du\, dv \qquad 4.94 \\
&= 0 \quad \text{if } (f_1+f_2)\neq 0
\end{aligned}$$

The Fourier Transform of the second term of 4.93 is,

$$\int\int\int_{-\infty}^{\infty} h_2(u,v)\, R_{xx}(\tau_1-u) R_{xx}(\tau_2-v)\, e^{-j2\pi f_1\tau_1} e^{-j2\pi f_2\tau_2}\, d\tau_1 d\tau_2 du\, dv$$

substituting $\theta = \tau_1 - u$, $d\tau_1 = d\theta$, gives $\int_{-\infty}^{\infty} R_{xx}(\theta) e^{-j2\pi f_1(\theta+u)} = S_{xx}(f_1) e^{-j2\pi f_1 u}$

substituting $\vartheta = \tau_2 - v$, $d\tau_2 = d\vartheta$, gives $\int_{-\infty}^{\infty} R_{xx}(\vartheta) e^{-j2\pi f_2(\vartheta+v)} = S_{xx}(f_2) e^{-j2\pi f_2 v}$

$$\begin{aligned}
&= S_{xx}(f_1) S_{xx}(f_2) \int_{-\infty}^{\infty} h(u,v) e^{-j2\pi f_1 u} e^{-j2\pi f_2 v}\, du dv \\
&= S_{xx}(f_1) S_{xx}(f_2) H(f_1,f_2) \qquad 4.95
\end{aligned}$$

The third term is the same as the second but with the roles of u and v reversed. Hence,

$$\begin{aligned}
&\int\int\int_{-\infty}^{\infty} h_2(u,v)\, R_{xx}(\tau_1-v) R_{xx}(\tau_2-u)\, e^{-j2\pi f_1\tau_1} e^{-j2\pi f_2\tau_2}\, d\tau_1 d\tau_2 du\, dv \\
&= S_{xx}(f_1) S_{xx}(f_2) H(f_2,f_1) \qquad 4.96
\end{aligned}$$

Exploiting the (assumed) symmetry of the second order kernel gives $H_2(f_1,f_2)=H_2(f_2,f_1)$ and so adding the three terms 4.94, 4.95 and 4.96 results in,

$$S_{xxy}(f_1,f_2) = 2S_{xx}(f_1) S_{xx}(f_2) H(f_1,f_2) \quad \text{if } (f_1+f_2)\neq 0 \qquad 4.97$$

Hence, even in the presence of $H_1(f)$, an unbiased estimate of the quadratic component of the second order Volterra system is,

$$H_2(f_1,f_2) = \frac{S_{XXY}(f_1,f_2)}{2S_{XX}(f_1)S_{XX}(f_2)} \qquad 4.98$$

as shown by, for example, [59, 62, 76].

4.6.1 Estimation of the quadratic transfer function

A major difficulty when estimating the auto bispectrum, which was discussed extensively in chapter three, is that the variance of the bispectrum contains terms due to power spectral effects. Averaging over longer data lengths will reduce the variance. However, if some form of normalisation is performed on the estimate to down weight the estimate of the bispectrum at frequencies where the second order properties are large, then the variance can be made more uniform across all frequencies. For this reason, the bicoherence or skewness function was usually estimated.

The variance of the real and imaginary parts of the cross bispectrum [38] can be easily obtained from the variance of the auto bispectrum and is given by,

$$\mathrm{var}(S_{XXY}(f_1,f_2)) = \frac{1}{M}[1+\delta_k(f_1-f_2)]\,S_{XX}(f_1)S_{XX}(f_2)S_{YY}(f_1+f_2) \qquad 4.99$$

where M is the number of blocks used in the estimation, $\delta_k(0) = 1$, and $\delta_k(f) = 0$ for non-zero f.

This contains the triple product $S_{XX}(f_1)S_{XX}(f_2)S_{YY}(f_1+f_2)$ and so an estimate of the cross bispectrum will be sensitive to power spectral effects. As $H_2(f_1,f_2)$ contains the cross bispectrum, the variance of $H_2(f_1,f_2)$ will be affected by the spectrum of x(t) which is in turn dependent on $H_1(f)$. Hence whilst the mean of the estimate of the second order kernel is independent of the linear kernel; its variance is not. In order to reduce the variance of $H_2(f_1,f_2)$ it is either necessary to average over more data or to normalise it to down weight the estimate where the linear component is large.

The technique used here to normalise the quadratic transfer function differs from the bicoherence. Taking the quadratic Volterra model, it is possible to place in parallel with it an arbitrary linear system, G(f). The value of G(f) will not affect the mean of the estimate of the quadratic transfer function as it only contains linear components and referring to figure 4.7 it can be shown that,

$$S_{XXY}(f_1,f_2) = S_{XXZ}(f_1,f_2) \qquad 4.100$$

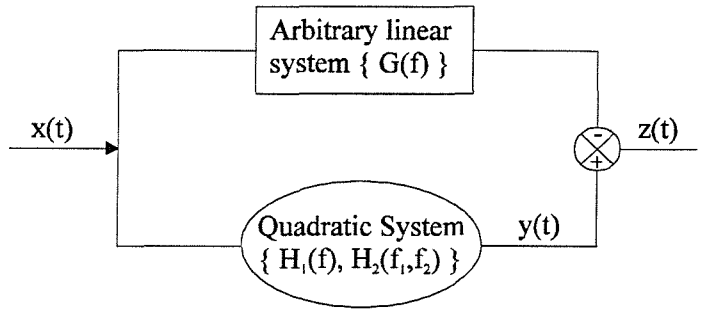

Figure 4.7: Quadratic and linear systems in parallel

A modified transfer function estimate, $H_{2(mod)}(f_1,f_2)$, can be defined as,

$$H_{2(mod)}(f_1,f_2) = \frac{S_{XXZ}(f_1,f_2)}{2S_{XX}(f_1)S_{XX}(f_2)} \qquad 4.101$$

Theoretically, this is the same as $H_2(f_1,f_2)$. However whilst estimates based on 4.98 can be unbiased (assuming sufficient resolution), prudent choice of the linear filter G(f) can yield estimates, based on 4.101, with smaller variances. It will now be shown that the choice $G(f) = H_1(f)$ gives the minimum output power ($E[z^2(t)]$) and hence minimises the variance of the transfer function estimate. From figure 4.7, z(t) can be written as,

$$z(t) = \int_{-\infty}^{\infty}(h_1(\tau_1) - g(\tau_1))x(t-\tau_1)d\tau_1 + \int\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2$$

4.102

To minimise the output power, it is necessary to solve $\frac{d\Psi}{dg} = 0$ where $\Psi = E[z(t)^2]$.

Ψ can be seen to consist of the following three terms:

$$\Psi = E\left[\begin{array}{l} 2\int\int\int_{-\infty}^{\infty}(h_1(\tau_1)-g(\tau_1))h_2(\tau_2,\tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)d\tau_1 d\tau_2 d\tau_3 + \\ \int\int\int\int_{-\infty}^{\infty}h_2(\tau_1,\tau_2)h_2(\tau_3,\tau_4)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)x(t-\tau_4)d\tau_1 d\tau_2 d\tau_3 d\tau_4 \\ \int\int_{-\infty}^{\infty}(h_1(\tau_1)-g(\tau_1))(h_1(\tau_2)-g(\tau_2))x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 \end{array}\right] \quad 4.103$$

The first term contains the product of three Gaussian variables and so will equal zero; the second does not contain g(t) and so $d\Psi/dg$ will equal zero; and intuitively, as Ψ is a squared expression the third term cannot be negative and so will be minimised when it is equal to zero, i.e. when $g(\tau) = h_1(\tau)$.

To implement this, z(t), is created by forming the output of the system, y(t), minus $y_{lin}(t)$. $y_{lin}(t)$ is the convolution of the impulse response, h(t), and the input to the system, x(t).

$$z(t) = y(t) - h(t)*x(t) \quad 4.104$$

This produces a new 'delinearised' signal from which it is possible to calculate the modified version of $H_2(f_1,f_2)$. The modified transfer function, $H_{2(mod)}(f_1,f_2)$, will have a more uniform variance as the second order properties have been removed. Double inverse Fourier transforming $H_{2(mod)}(f_1,f_2)$ with respect to f_1 and f_2 will produce the two dimensional impulse response function $h_2(\tau_1,\tau_2)$. The quadratic component of the signal can be formed by a two dimensional convolution of $h_2(\tau_1,\tau_2)$ with the input to the system, x(t), to produce, $y_{quad}(t)$.

4.6.2 Example of a simple quadratic system

The same skewed, mixed signal, as used in chapter three, is now used to demonstrate the quadratic Volterra model. The system consists of the sum of a Gaussian signal filtered through a low pass filter and squarer and the same Gaussian signal high pass filtered. It therefore contains only linear and quadratic elements and so can be solved with the second order Volterra model. Figure 4.8 shows $S_{YY}(f)$, the power spectrum of the output which shows the two distinct regions of the signal: a low frequency skewed component, and a high frequency Gaussian component. The coherence function, figure 4.9, is approximately unity at higher frequencies due to the linear part of the signal. $H_1(f)$, the linear transfer function, is calculated but not shown here. It is high pass in form as expected.

Figure 4.10 shows the quadratic transfer function calculated using an FFT size of 64 and a sample length of 4096 points. It can be seen that the quadratic component is correctly detected as shown by the data in the low bifrequency region. However the features in the mid bifrequency region are unexpected and are caused by the linear component. Hence when the quadratic transfer function is averaged over 64 segments, the quadratic and linear components are seen to have similar strengths. If the data length is increased to 262144 data points and the cross bispectrum averaged over 4096 segments the variance of the linear component decreases and the quadratic transfer function correctly detects only the low frequency quadratic component of the signal (figure 4.12). However, if the data length is reduced to 4096 data points, and the modified quadratic transfer function calculated, $H_{2(mod)}(f_1,f_2)$, that is with the linear component of the signal removed, the variance is reduced and the quadratic transfer function only detects the quadratic component of the signal (figure 4.11). In this case, the magnitude of the linear component due to the variance of $H_2(f_1,f_2)$ is of similar value as when the signal was averaged over 4032 more realisations. Note that by subtracting the linear component of the signal, the variance of the linear component has been minimised but the variance of the quadratic component has not been altered. This can be seen by comparing the roughness of the low frequency component of the quadratic transfer function in figure 4.11 with that of figure 4.12, where a longer data length was used. Figure 4.13 shows the modified quadratic transfer function calculated using a data length of 262144 sample points. Normalising the signal by removing the linear

component or 'delinearising' allows improved estimation of the quadratic transfer function, suppressing troublesome linear terms.

The linear and quadratic impulse responses, calculated using a data length of 4096 samples, are shown in figures 4.14 and 4.15 respectively. By convolving these with the Gaussian input data, x(t), two separate components are calculated: the signal due to the linear part, and the signal due to the quadratic part. The power spectra of these two signals are shown in figures 4.16 and 4.17 respectively. The original output power spectrum is shown by the dotted line. Figure 4.18 shows the power spectrum of the sum of the two components which is the same as the original signal, thus showing the quadratic Volterra model has correctly accounted for all the terms in the signal. Since in this simulation the system is known to be only second order, this is to be expected.

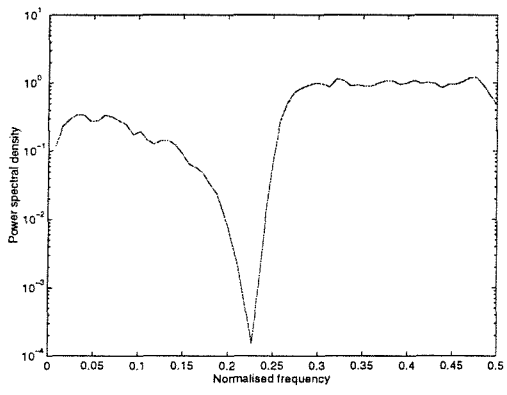

Figure 4.8: Output power spectrum of mixed signal

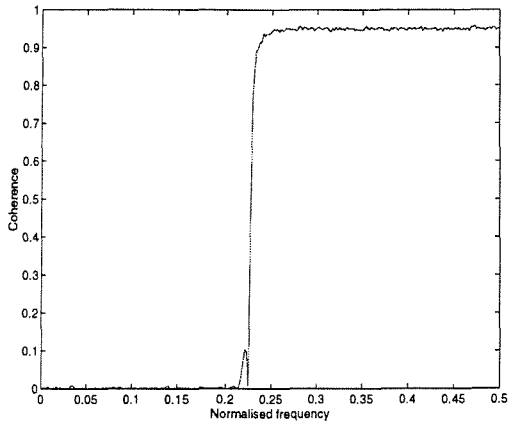

Figure 4.9: Coherence function of mixed signal

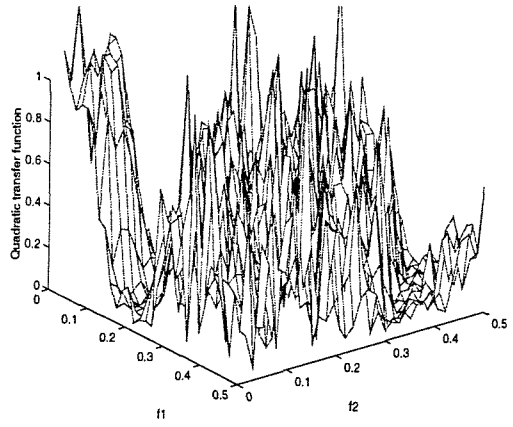

Figure 4.10: Quadratic transfer function of mixed signal

(Data length = 4096, FFT size = 64)

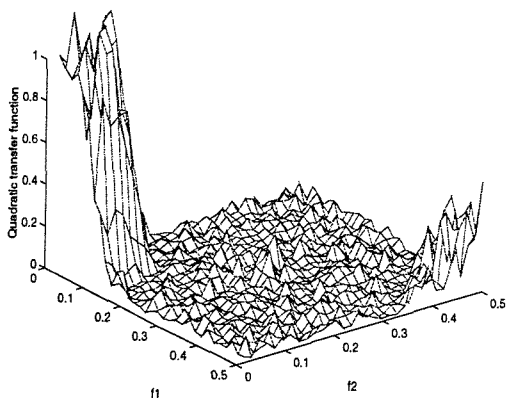

Figure 4.11: Quadratic transfer function of mixed signal with linear component removed
(Data length = 4096, FFT size=64)

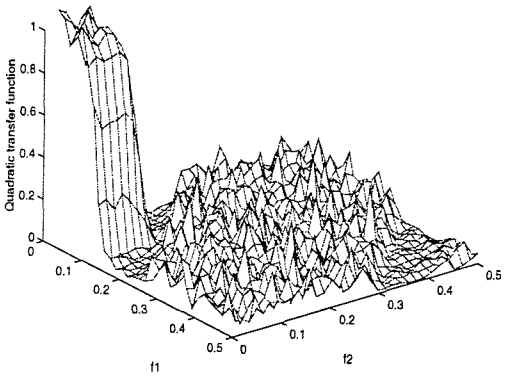

Figure 4.12: Quadratic transfer function of mixed signal

(Data length = 262144, FFT size = 64)

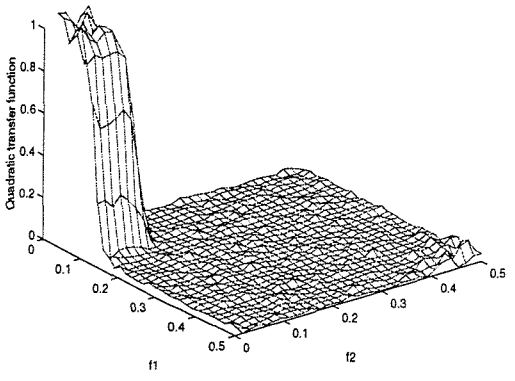

Figure: 4.13 Quadratic transfer function of mixed signal with linear component removed
(Data length = 262144, FFT size=64)

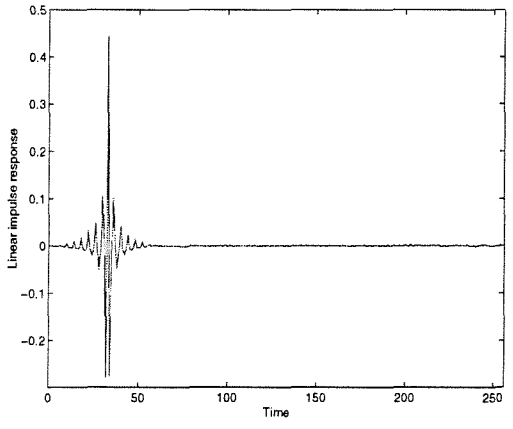

Figure 4.14: Linear impulse response

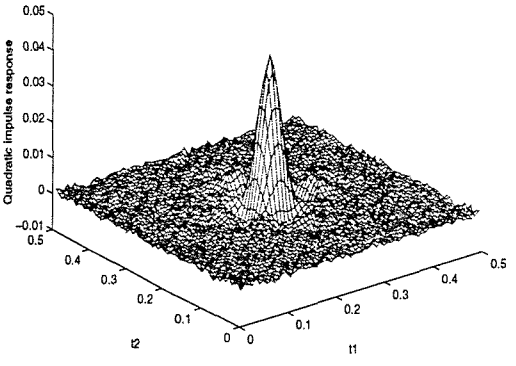

Figure 4.15: Quadratic impulse response

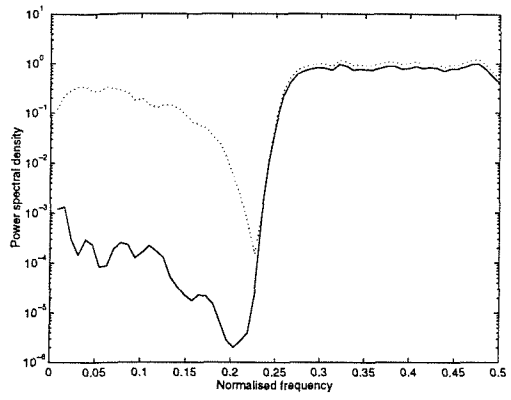

Figure 4.16: Power spectrum of the linear component (solid) and power spectrum of the output (dots).

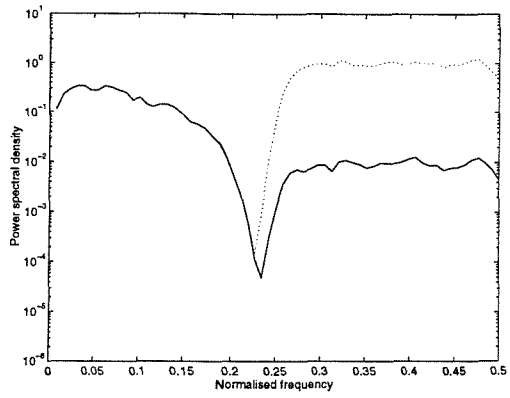

Figure 4.17: Power spectrum of quadratic component (solid) and power spectrum of the output (dots).

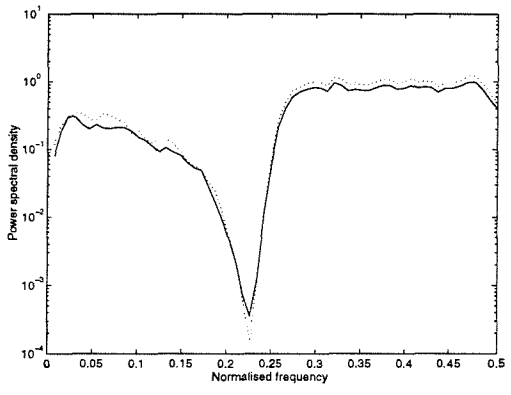

Figure 4.18: Power spectrum of sum of linear and quadratic components (solid) and power spectrum of the output (dots).

4.7 Cubic Volterra Model

The second order Volterra model, equation 4.86, can be extended to a cubic model,

$$y(t) = \int_{-\infty}^{\infty} h_1(\tau_1)\, x(t-\tau_1)d\tau_1 + \int\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)\, x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 + \int\int\int_{-\infty}^{\infty} h_3(\tau_1,\tau_2,\tau_3)\, x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)d\tau_1 d\tau_2 d\tau_3 \qquad 4.105$$

This can be viewed as the parallel connection of a linear, quadratic and cubic system (figure 4.19).

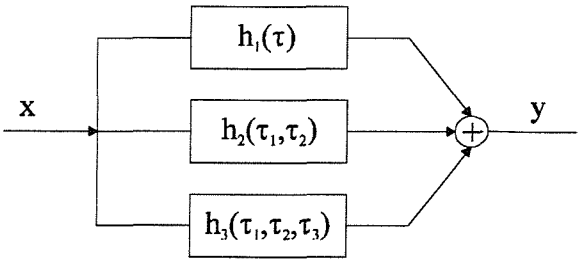

Figure 4.19: Cubic Volterra model

For a Gaussian input it is possible, as in the second order case, to find the following expressions involving $H_1(f)$, $H_2(f_1,f_2)$ an $H_3(f_1,f_2,f_3)$. This is shown in Appendix E.

$$H_1(f) = \frac{S_{XY}(f)}{S_{XX}(f)} - 3\int_{-\infty}^{\infty} S_{XX}(g)H_3(-f,g,-g)\,dg \quad \text{if } f \neq 0 \qquad 4.106$$

$$H_2(f_1,f_2) = \frac{S_{XXY}(f_1,f_2)}{2S_{XX}(f_1)S_{XX}(f_2)} \quad \text{if } (f_1+f_2) \neq 0 \qquad 4.107$$

$$H_3(f_1,f_2,f_3) = \frac{S_{XXXY}(f_1,f_2,f_3)}{6S_{XX}(f_1)S_{XX}(f_2)S_{XX}(f_3)} \quad \text{if } (f_1+f_2) \neq 0,\ (f_2+f_3) \neq 0 \text{ and } (f_1+f_3) \neq 0 \qquad 4.108$$

N.B. The ratio, $S_{XY}(f)/S_{XX}(f)$, now includes a term due to $H_3(f_1,f_2,f_3)$ namely,

$$3\int_{-\infty}^{\infty} S_{XX}(g)\,H_3(-f,g,-g)\,dg \qquad 4.109$$

$H_3(f_1,f_2,f_3)$ is a function of three frequency variables but the linear transfer function is only a function of a single frequency variable. Therefore equation 4.109 represents a projection of a function in three dimensional space onto a function in one dimensional space.

As in the quadratic model, before calculating the quadratic transfer function the linear component should be removed. However, knowledge of $H_3(f_1,f_2,f_3)$ is needed to form an unbiased estimate of the linear transfer function but this is not yet known. One estimate of $H_1(f)$ available is $S_{XY}(f)/S_{XX}(f)$. Such an estimate is biased but this should be corrected at a later stage. The biased estimate of the linear transfer function is used to generate an estimate of the linear component of the signal. Having 'delinearised' the signal an estimate of the quadratic transfer function, $H_2(f_1,f_2)$, can be calculated. From this an estimate of the quadratic impulse response function, and hence of the quadratic component of the system, can be generated.

Equation 4.108 for $H_3(f_1,f_2,f_3)$ contains no terms in either $H_2(f_1,f_2)$ or $H_1(f)$ so before $H_3(f_1,f_2,f_3)$ is calculated the linear component and quadratic component of the signal are removed. An estimate of $H_3(f_1,f_2,f_3)$ is made and by triple inverse Fourier transforming this with respect to f_1, f_2 and f_3, an estimate of the cubic impulse response function can be made. The triple convolution of the input data with the cubic impulse response gives the estimate of the cubic component of the signal.

The residual error is formed by subtracting the linear, quadratic, and cubic components of the signal. This error consists of components of orders higher than three, measurement noise, and components due to inaccurate estimation of the first three terms in the model.

However, as it was assumed that $H_3(f_1,f_2,f_3)$ was equal to zero when the 'delinearising' filter was estimated, there is room to improve the kernel estimates. The estimated value for $H_3(f_1,f_2,f_3)$ can now be substituted into equation 4.106 to form a more accurate delinearising filter. From this a new estimate of the quadratic transfer function and then the cubic transfer function can be calculated. This process can be continued in an iterative manner until the estimates of $H_1(f)$, $H_2(f_1,f_2)$ and $H_3(f_1,f_2,f_3)$ achieve steady state values. The whole process is depicted in figure 4.20.

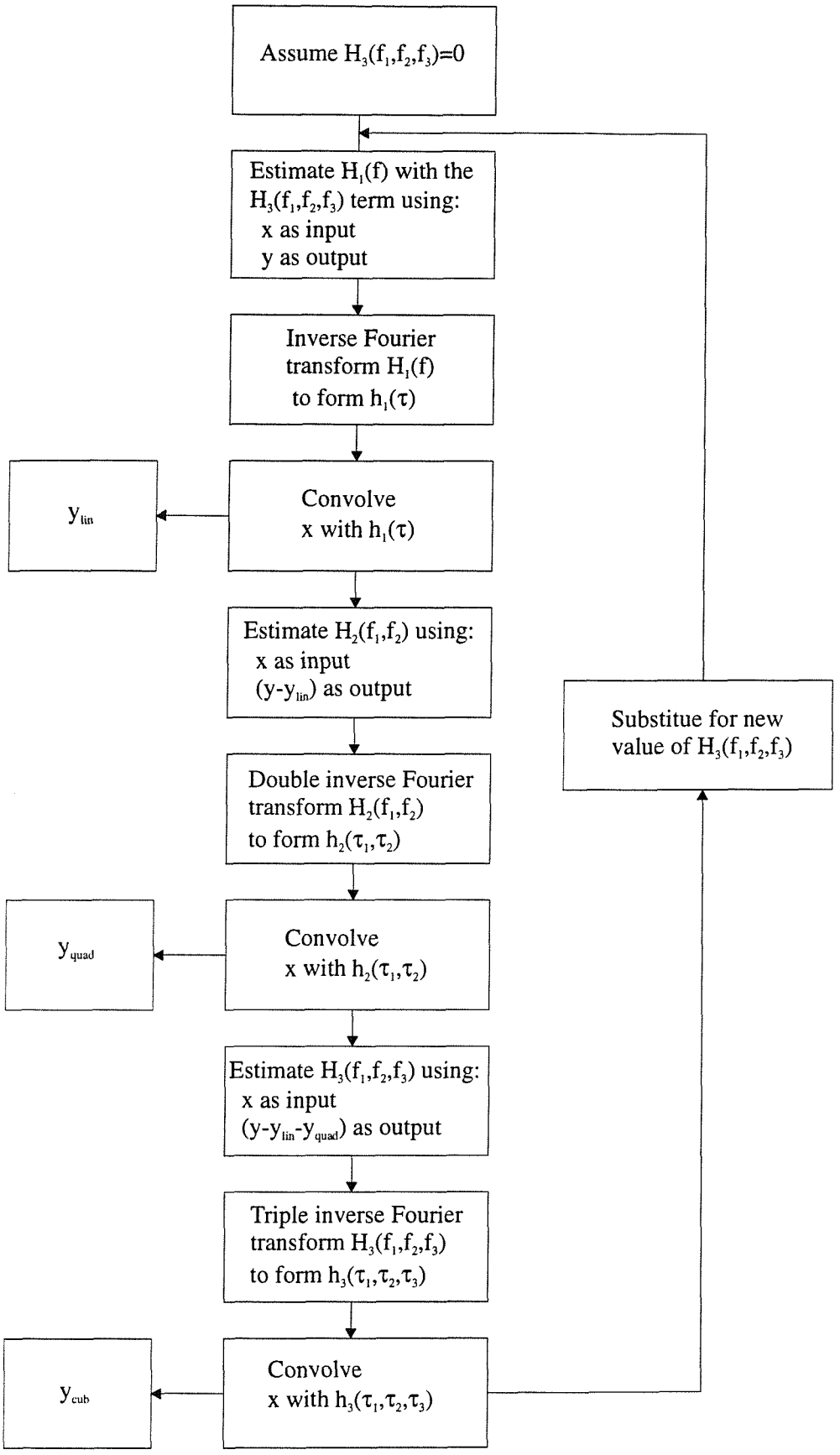

Figure 4.20: Flow diagram of cubic Volterra model

4.7.1 Example of a simple cubic system

An example of a simple system than contains only cubic elements is now given to illustrate some of the difficulties when estimating the cubic transfer function and impulse response. A more complex example will be given in the next chapter.

If a Gaussian white signal is taken and passed through a cube law device (i.e. $y(t) = x(t)^3$) the resulting signal will contain only cubic terms. This simplifies the estimation of the third order Volterra kernel as there are no first or second order terms and so there is no need for an iterative solution.

The third order frequency domain Volterra kernel of this system is shown in figure 4.21 and has an approximately constant value of unity throughout. This has been calculated by taking the cross moment trispectrum and dividing through by the power spectral terms as in equation 4.108. As the moment trispectrum has been calculated it is then necessary to remove the effect of the second order terms to form the cumulant trispectrum. These second order terms appear as three planes, $f_1+f_2 = 0$, $f_2+f_3 = 0$, and $f_1+f_3 = 0$. By calculating the product of the auto and cross spectrum along each of the planes and then subtracting this from the moment trispectrum, the true cumulant trispectrum can be calculated.

Triple Fourier transforming this with respect to τ_1, τ_2 and τ_3 will give the cubic impulse response of the system. This is shown in figure 4.22 and is seen to be a single delta function at the origin. If the original input data is convolved with this three dimensional impulse response the resulting data should be identical to the output signal. Figures 4.23 and 4.24 show the original output signal, and the result of convolving the input data with the impulse response. The difference between these two signals is shown in figure 4.25.

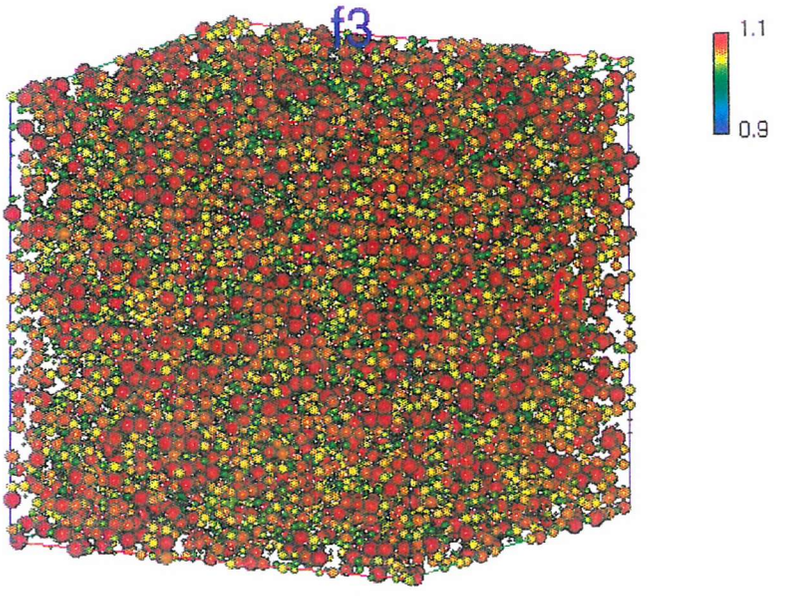

Figure 4.21: Third order frequency domain Volterra kernel of the cubic system

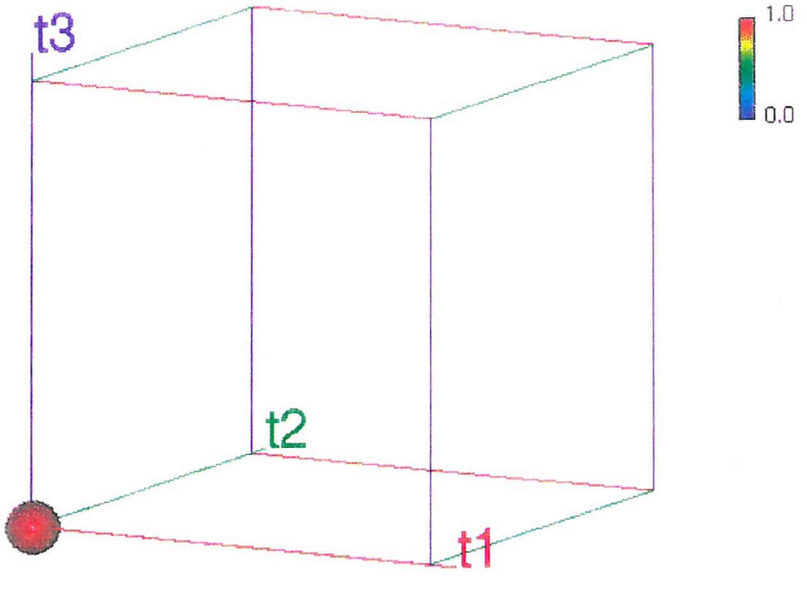

Figure 4.22: Cubic impulse response of the system

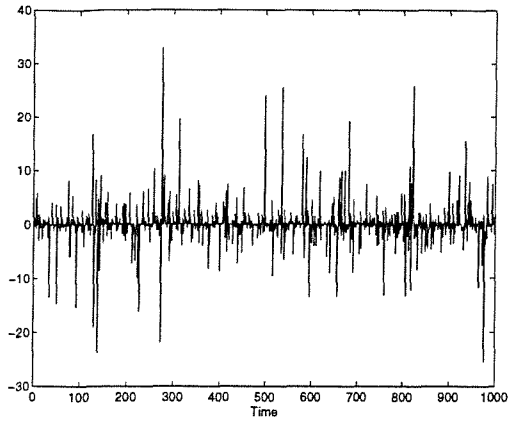

Figure 4.23: Output of the cubic system

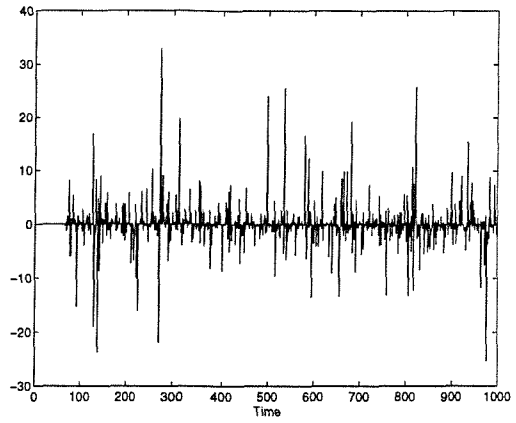

Figure 4.24: Output of the convolution of the cubic impulse response with the input data

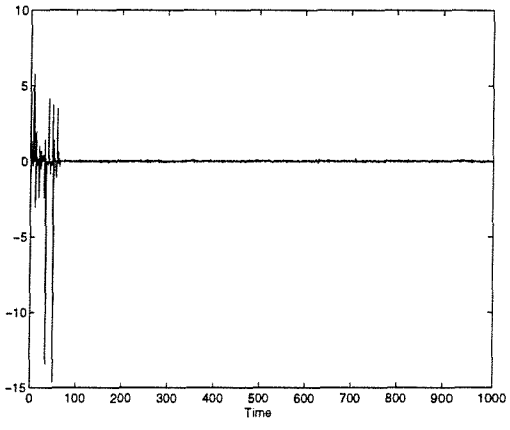

Figure 4.25: Difference between actual output and predicted output of the cubic system

4.8 A note on higher order coherence functions

In section 4.6.2, in the example of the mixed signal, the results are displayed by looking at the power spectra of the linear, and quadratic terms and comparing these with the known output power spectrum of the signal. In this way, from an engineering point of view, it is easy to see what part of the signal accounts for the power at every frequency.

Some authors [63] prefer to take this one stage further and calculate higher order coherence functions to display the same information. For the simple quadratic model the linear and quadratic coherences are defined as:

$$\gamma^2_{lin}(f) = \frac{E\left[\left|Y_{lin}(f)\right|^2\right]}{E\left[\left|Y(f)\right|^2\right]} \qquad 4.110$$

$$\gamma^2_{quad}(f) = \frac{E\left[\left|Y_{quad}(f)\right|^2\right]}{E\left[\left|Y(f)\right|^2\right]} \qquad 4.111$$

In a similar fashion to the traditional second order coherence function these will take values on a scale from zero to unity depending on how well the models fit the data.

Although this may seem a logical extension, many unhelpful contradictions can occur. For example although the coherence functions theoretically should take values between zero and one in some cases it can be either negative or greater than one. This may occur for a number of reasons: if the model does not correctly account for the data it could be possible that the power spectrum of the model is greater than the power spectrum of the actual output. This would lead to a coherence function taking values greater than unity. Also in the case of the cubic model, cross terms would be present which can lead to negative coherence values.

4.9 Conclusions

In this chapter, techniques using cross higher order spectra have been presented for calculating the Volterra kernels of a system. Two methods have been proposed, the first calculates only the main diagonal of the Volterra kernel, which although simpler, has limited applicability. The second calculates the full Volterra kernel.

Both quadratic and cubic Volterra models have been presented. The quadratic model produces a closed form solution and no iteration is needed to solve it. However, an iterative procedure has been proposed to solve the cubic model. There are methods of producing closed form solutions for the cubic model and these will be examined in the next chapter. Although the cubic model is considerably more complex than the quadratic model, for the

experiments considered in chapter six it is found to be necessary. The problems that may occur when too low an order model is used for a particular situation will also be discussed.

As with the auto bispectrum, calculations of the cross bispectrum can easily be corrupted by second order properties. For the auto bispectrum the usual technique to avoid this is to calculate the bicoherence or skewness function. For the cross bispectrum, in the calculation of Volterra kernels, this is not possible, and so a new method has been proposed which involves subtracting off lower order terms. This has been demonstrated to work very well for a simulated quadratic system.

This whole chapter has been based on the assumption that the input to the system is Gaussian. If this assumption is invalid, all the products of an odd number of Gaussian variables can no longer be set to zero and must be included in the solution of the Volterra model. This obviously leads to a far more complex problem. In chapter five, some of the possible methods of calculating the Volterra kernels when the input is non-Gaussian will be discussed.

Chapter 5

Limitations of the Volterra series

5.1 Introduction

In the previous chapter, it was shown how the Volterra series could be used to model some nonlinear systems. Before studying the experimental results in the next chapter, a number of difficulties which can arise with the practical application of Volterra models will be examined and some possible solutions presented.

The first difficulty concerns the convergence of the Volterra series. A brief discussion of some of the types of system that are likely to have converging or diverging Volterra series are given, with references to more analytic work on the subject. The second difficulty, concerning the calculation of the Volterra kernels of a system, is that estimation is only possible if the contributions of each of the system's operators can be separated from the total system response. No exact method of isolating an individual Volterra operator exists, except by truncating the series, as in this work. The techniques derived in the previous chapters, in the process of estimating the Volterra kernels, do in fact estimate the kernels of a related series, the Wiener series, whose kernels can be calculated without truncating the series. Although a detailed description of the Wiener series is not given here, it is of interest to note the relationship between the Volterra and Wiener kernels.

In many situations, it is important to know the response of a system to a particular input. A possible technique to find the response, is to apply a Gaussian input to the system and determine the Volterra kernels which can then be used to predict the response of the system to any input. This method is demonstrated using the example of the Duffing oscillator. The Volterra kernels are calculated using a Gaussian input and then the response of the

Duffing oscillator to a sine wave is predicted and compared with the actual response evaluated using numerical techniques.

Throughout this work, it has always been assumed that the input to the system is Gaussian. In many practical applications this assumption may be invalid, and so a method is presented for the quadratic Volterra model that removes this restriction. The solution of this new quadratic Volterra model differs from the previous one, because it is no longer possible to assume that the product of an odd number of input terms is equal to zero or that the product of an even number of input terms can be decomposed into pairs of terms. These expressions must now be solved explicitly using higher order correlation functions. This vastly increases the complexity of the model and it is shown that the quadratic model requires that higher order spectra up to the fourth order, the trispectrum, are studied. It is noted that a cubic model would require the calculation of sixth order spectra.

The topic of causality is briefly mentioned. In the solution of the Volterra models, the Volterra kernels have not been restricted to be zero in negative time and so could well be non-causal. However, it is seen that for the practical examples studied, all the estimated kernels are causal in nature.

A number of different Volterra models, all calculated using higher order spectra, have now been presented. They can be broadly split into a number of categories such as the order of the model, the assumptions placed on the input, and the highest order spectra required for the solution of the model. Therefore, a summary of the different models, their restrictions, features, and the techniques that different authors have used to solve them is given.

5.2 Convergence of the Volterra Series

The Volterra series has been termed 'a Taylor series with memory' [70] and the convergence properties of the Volterra series are closely linked to the convergence of the Taylor series. For example, consider the memoryless system, with an input, x(t), and output, y(t), defined by,

$$y(t) = A\,\mathrm{sgn}[x(t)]$$
$$= \begin{cases} A & \text{for } y(t) > 0 \\ 0 & \text{for } y(t) = 0 \\ -A & \text{for } y(t) < 0 \end{cases} \qquad 5.1$$

where A is a constant. There is no convergent Taylor series for equation 5.1 about $y(t) = 0$ and so a nonlinear system, which has such characteristics, cannot be represented by a Volterra series model.

For the Volterra series to be convergent, any memory it possesses must be finite. That is, the effect of any input on a Volterra series will die away and become insignificant in finite time. An example, given by Schetzen [70], of a system which does not possess finite memory is a fuse which, after its rated current has been exceeded will never return to its original equilibrium state no matter how long one waits.

In general, any system which has multiple equilibria cannot be modelled by a Volterra series except locally around one equilibrium and with the class of inputs restricted so as to ensure that none of the other equilibrium states will be reached. In the next chapter, a practical example, based on a beam constrained by pairs of repelling magnets is given. This system will have a convergent Volterra series. However, there is a very similar experiment [54], which is often used to demonstrate chaotic vibrations, in which a vibrating beam, rather than being constrained by two pairs of repelling magnets, is attracted by either of two pairs of magnets. This experiment would have multiple equilibria and so it would not be possible to find a convergent Volterra series representation for it, except locally around one of the magnets.

The above three criteria have given very simple guidelines as to the types of system that will have a convergent Volterra series. No attempt has been made to prove any of the statements but further discussion on the convergence of the Volterra series can be found in Boyd and Chua [7] and Wright [81].

5.3 The Wiener series

The two main difficulties associated with the practical application of the Volterra series are, first, the problem of convergence which was discussed in the previous section, and second, the difficulty concerning the measurement of the Volterra kernels of a given system. As the Volterra series is an infinite series, it is only possible to measure a system's Volterra kernels if the contribution of each of the Volterra operators can be separated from the total system response. In this work, this problem has been circumvented by truncating the Volterra series and so making a finite order system. However, no exact method of isolating an individual Volterra operator exists for systems that are not of finite order.

Wiener [70] avoided these problems by forming a new series from the Volterra series. The Wiener series allows the identification of an individual operator without the need for truncation but only when the input is a white, Gaussian signal. Higher order spectral methods are particularly suited for the estimation of the Wiener kernels and it can be seen in the next section how the methods that have been used in the previous chapters closely relate to the conversion between Wiener and Volterra kernels.

5.3.1 Conversion between Wiener and Volterra functions

For a system truncated at third order the relationship between the Volterra kernels and Wiener kernels is [70]:

$$H_1(f)=K_1(f)-K_{1(3)}(f) \tag{5.2}$$

$$H_2(f_1,f_2)=K_2(f_1,f_2) \tag{5.3}$$

$$H_3(f_1,f_2,f_3)=K_3(f_1,f_2,f_3) \tag{5.4}$$

where $K_1(f)$, $K_2(f_1,f_2)$, and $K_3(f_1,f_2,f_3)$ are the systems Wiener kernels and $K_{1(3)}(f)$ is the derived Wiener kernel, defined as:

$$K_{1(3)}(f) = 3A\int_{-\infty}^{\infty} K_3(-f,g,-g)dg \qquad 5.5$$

where A is a constant. If this is compared with the cubic Volterra model from chapter four, strong similarities can be seen.

$$H_1(f) = \frac{S_{XY}(f)}{S_{XX}(f)} - 3\int_{-\infty}^{\infty} S_{XX}(g)H_3(-f,g,-g)\,dg \qquad 5.6$$

$$H_2(f_1,f_2) = \frac{S_{XXY}(f_1,f_2)}{2S_{XX}(f_1)S_{XX}(f_2)} \qquad 5.7$$

$$H_3(f_1,f_2,f_3) = \frac{S_{XXXY}(f_1,f_2,f_3)}{6S_{XX}(f_1)S_{XX}(f_2)S_{XX}(f_3)} \qquad 5.8$$

In order to solve the cubic Volterra model, the H_3 term in the estimation of the linear kernel was set to zero, and the linear, quadratic, and cubic kernels estimated. It can now be seen that at this stage, the Wiener kernels were in fact being estimated. The substitution of H_3 into equation 5.6 can be interpreted as attempting to obtain the Volterra kernels from their Wiener counterparts.

5.4 The use of Volterra models to find the response of a non-Gaussian signal

Although the methods described in chapter four for the calculation of the Volterra kernels assume a Gaussian input, once the Volterra kernels of a system have been calculated, they can be used to predict the response to any input. In this section, the Volterra kernels of the Duffing oscillator will be calculated and then the response of the Duffing oscillator to a sine wave predicted. In chapter three, the auto higher order spectra of the output from a Duffing oscillator was studied. Cross higher order spectral analysis will now be used to study the Duffing oscillator and estimate its Volterra kernels.

5.4.1 The cubic Volterra model of a Duffing oscillator

In chapter three, the Duffing oscillator was defined as,

$$\ddot{x} + 2\zeta\omega_0\dot{x} + \omega_0{}^2 x + \alpha x^3 = f(t) \qquad 5.9$$

where the input, f(t), was assumed to be a Gaussian process. For this work, a small quadratic nonlinearity is included to generalise the Duffing equation to,

$$\ddot{x} + 2\zeta\omega_0\dot{x} + \omega_0{}^2 x + \alpha x^3 + \beta x^2 = f(t) \qquad 5.10$$

The system was simulated using a fourth order Runge Kutta model, with fixed step length, to obtain, x(t), the output. The coefficients of the cubic and quadratic terms were set to 0.25 and 0.1 respectively and the system is lightly damped (ζ=0.1) with a natural frequency (ω_0) of 0.2 rad/s. It is possible, using traditional sinusoidal probing methods [81], to find analytical expressions for the first three Volterra kernels of the Duffing Oscillator. These are quoted as [55];

$$H_1(f) = \frac{1}{-4\pi f^2 + j4\pi\xi f + 1} \qquad 5.11$$

$$H_2(f_1, f_2) = -2\beta H_1(f_1)H_1(f_2)H_1(f_1 + f_2) \qquad 5.12$$

$$H_3(f_1, f_2, f_3) = -6\alpha H_1(f_1)H_1(f_2)H_1(f_3)H_1(f_1 + f_2 + f_3) \qquad 5.13$$

Note that the magnitude of the quadratic and cubic Volterra kernels are directly proportional to β and α respectively. Using the cubic Volterra model developed in chapter four, the first three Volterra kernels can be calculated from the simulated input/output data. These estimated kernels are depicted in figures 5.1-5.6: figures 5.1 and 5.2 show the linear time domain and frequency domain Volterra kernels, respectively; figures 5.3 and 5.4 show

the quadratic time domain and frequency domain Volterra kernels, respectively; and figures 5.5 and 5.6 show the cubic time domain and frequency domain Volterra kernels, respectively.

Figure 5.7, shows the power spectrum of the linear (thick solid line), quadratic (dashed line), and cubic (thin solid line) components of the signal, plotted against the actual output of the Runge Kutta model (dotted line). It can be seen, as expected, that the linear component accounts for the most power in the signal, followed by the cubic component, and then the quadratic component. The nonlinearity is distributed across all frequencies, although it is most dominant at the resonance, which is not unexpected as the resonant frequency is where most of the energy of the signal can be found. The sum of these three components gives the output of the Volterra model, which is shown in figure 5.8 by the solid line, again plotted against the output of the Runge Kutta model (dotted line). The complete Volterra model can be seen to account very well for the vast majority of the power across all frequencies

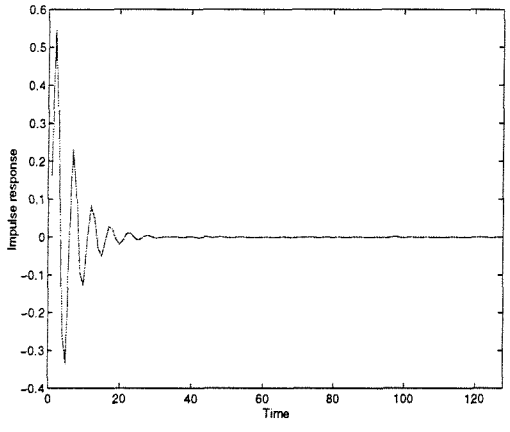

Figure 5.1: Linear impulse response of the Duffing oscillator.

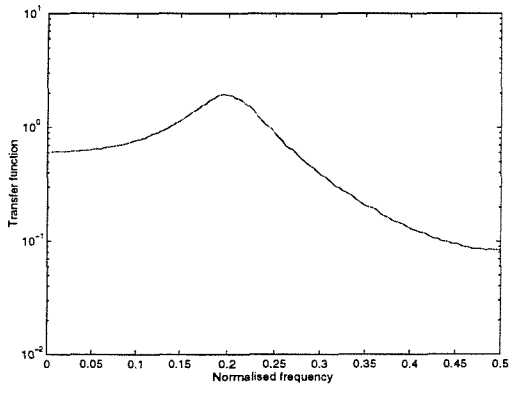

Figure 5.2: Linear transfer function of the Duffing oscillator.

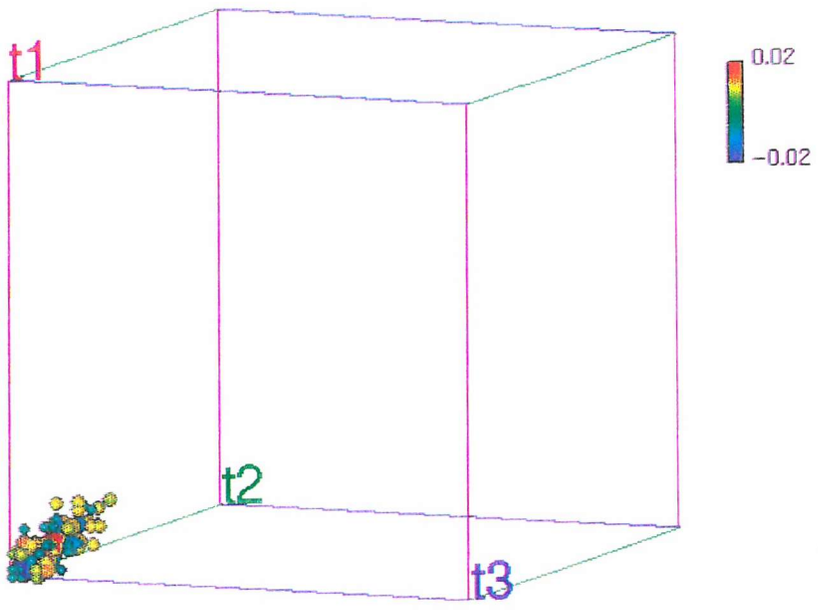

Figure 5.5: Cubic impulse response of the Duffing oscillator

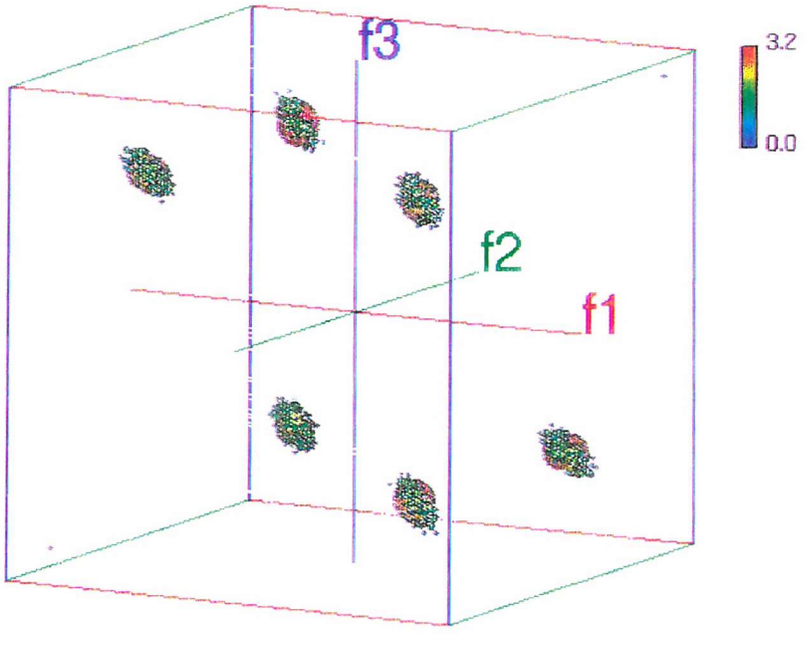

Figure 5.6: Cubic transfer function of the Duffing oscillator

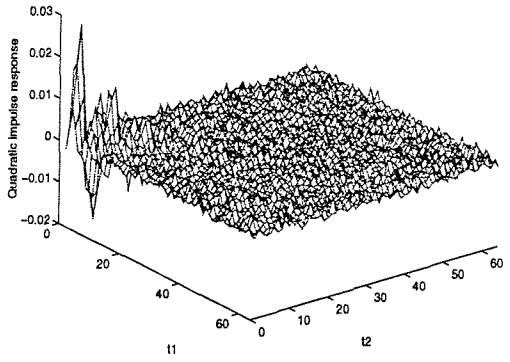

Figure 5.3: Quadratic impulse response of the Duffing oscillator .

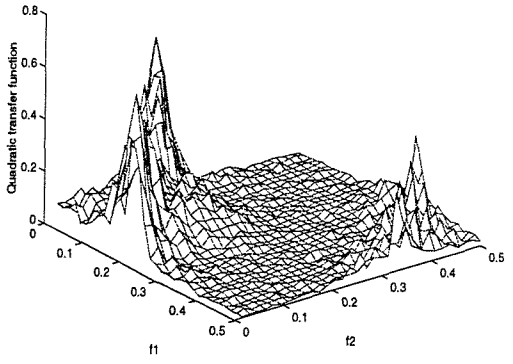

Figure 5.4: Quadratic transfer function of the Duffing oscillator.

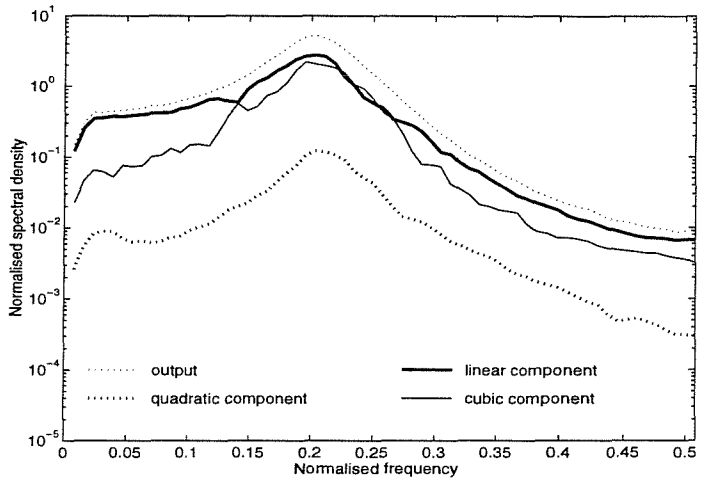

Figure 5.7: Power spectra of the linear, quadratic and cubic components of the Volterra model and the power spectrum of the output of Duffing oscillator.

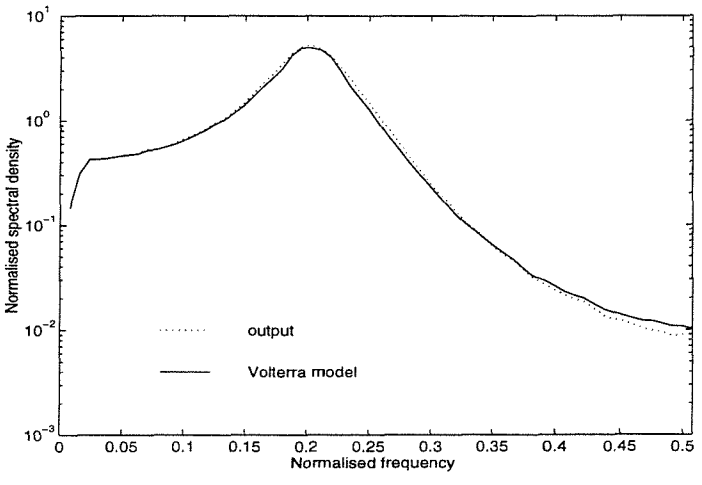

Figure 5.8: Power spectra of the Volterra model and output of the Duffing oscillator

5.4.2 Sine wave input to the Volterra model of a Duffing oscillator

Having used the input and output from the Runge Kutta model to calculate the first three Volterra kernels, these can now be used to predict the response to any other input. In this case the response to a sine wave is calculated, first by simulating the Duffing oscillator with the Runge Kutta method, using the sine wave as an input, and then by using the sine wave as the input to the Volterra kernels, as shown diagrammatically in figure 5.9. The results from both methods can then be compared.

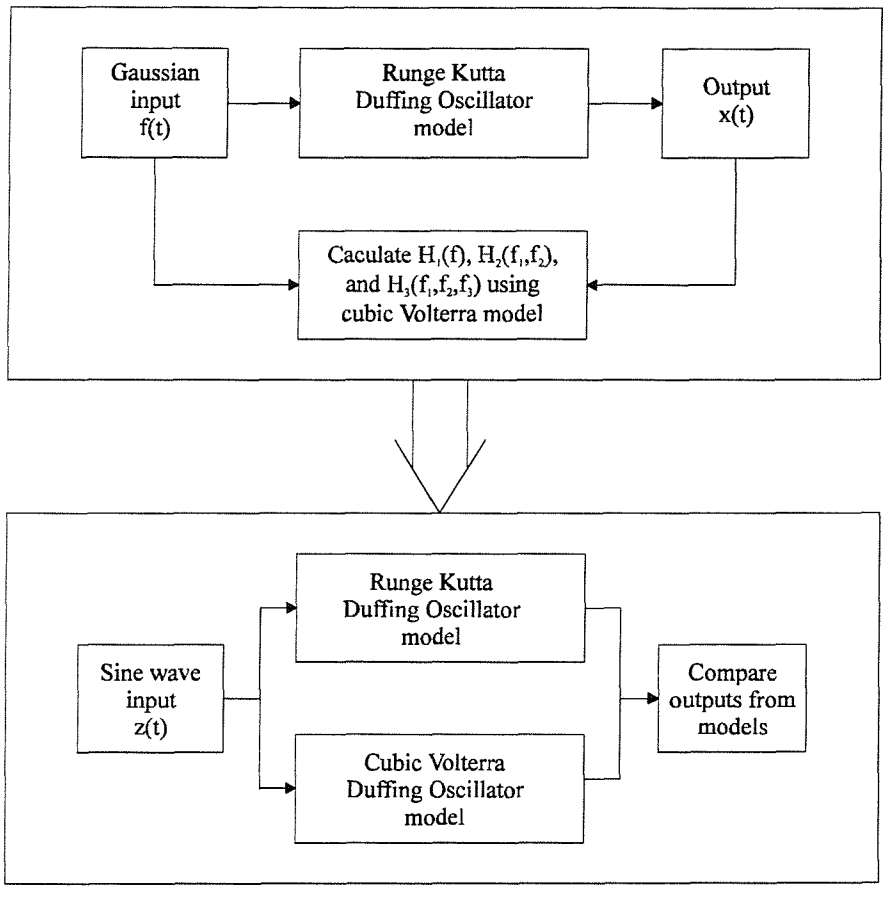

Figure 5.9: Method for comparison of the Runge Kutta model with the Volterra model

If the Runge Kutta simulation is repeated, but with a sine wave as the input, the power spectrum of the output, figure 5.10, dotted line, can be seen to consist of four resonant peaks. The response of the linear, quadratic, and cubic Volterra kernels to the sine wave is also shown in figure 5.10 by the power spectra of the linear (thick solid line), quadratic

(dashed line), and cubic (thin solid line) components. It can be seen that the linear component contributes to the first resonance in the spectrum, the quadratic component to the second resonance, and the cubic component to both the third and first resonant peaks. As only a cubic Volterra model is used, there will be no components at frequencies higher than the third resonance.

By summing the linear, quadratic, and cubic components the total output of the Volterra model can be obtained as shown by the solid line in figure 5.11. If this is compared with the output of the Runge Kutta solution, figure 5.11 (dotted line), it can be seen that it accounts fairly well for the first three resonances. All higher order odd kernels will contribute some power to the first and third resonance, and as only two have been considered, it is unreasonable for the magnitude of the power of the Volterra model to be as high as the Runge Kutta solution. Similarly, all even kernels will contribute some power to the second resonance. However, allowing for the limitations caused by truncation, the Volterra model has performed quite well at predicting the response of the Duffing oscillator to a sine wave.

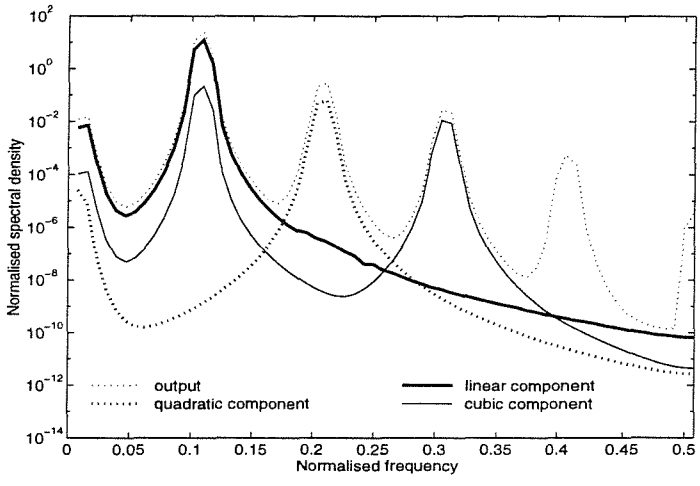

Figure 5.10: Power spectra of the linear, quadratic and cubic components of the Volterra model and the output of the Runge Kutta model for the Duffing oscillator with a sine wave input

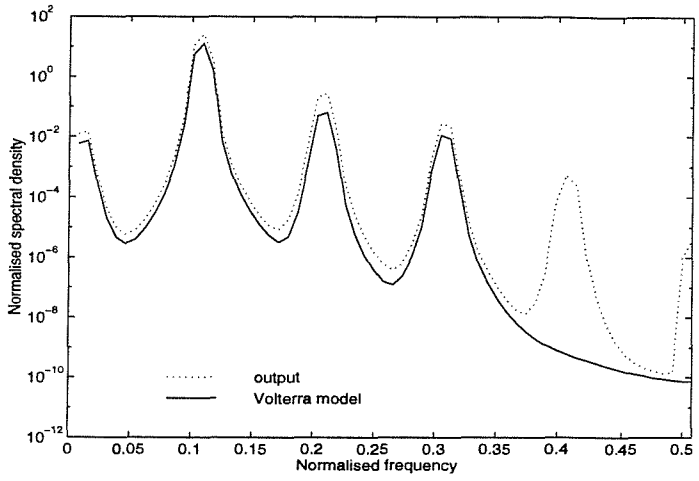

Figure 5.11: Power spectra of the Volterra model and output of the Runge Kutta model for a Duffing oscillator with a sine wave input.

5.5 Solutions for the Volterra models when the input is non-Gaussian

The derivation of estimators for the quadratic and cubic Volterra models in chapter four rely on the properties of the higher statistical moments of a Gaussian process. They are therefore subject to error if used with signals whose higher order moments deviate from those of a true Gaussian signal. In this section, the quadratic Volterra model is solved with no assumptions placed on the statistics of the input.

In section 4.6, by substituting the quadratic Volterra model (equation 4.86) into the cross correlation function (equation 4.87) the following expression is obtained:

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h_1(u)E[x(t-\tau)x(t-u)]\,du + \int\int_{-\infty}^{\infty} h_2(u,v)E[x(t-\tau)x(t-u)x(t-v)]\,du\,dv \qquad 5.14$$

As with the original quadratic Volterra model, the linear component can be found by calculating the Fourier transform of equation 5.14 to give:

$$S_{XY}(\upsilon) = \int_{-\infty}^{\infty} h_1(u)E[x(t-\tau)x(t-u)]e^{-j2\pi\upsilon\tau}\,du\,d\tau + \iint_{-\infty}^{\infty} h_2(u,v)E[x(t-\tau)x(t-u)x(t-v)]e^{-j2\pi\upsilon\tau}\,du\,dv\,d\tau \qquad 5.15$$

Evaluating the first term of equation 5.15 leads to the same expression as found for the previous quadratic model. That is,

$$\int_{-\infty}^{\infty} h_1(u)R_{xx}(\tau-u)e^{-j2\pi\upsilon\tau}d\tau du = H_1(\upsilon)S_{xx}(\upsilon) \qquad 5.16$$

The second term of equation 5.15 would previously have contained the product of an odd number of Gaussian variables and so be set to zero. However, it now must be solved, and so is rewritten in terms of an auto bicorrelation function as,

$$\int_{-\infty}^{\infty} h_2(u,v)E[x(t-\tau)x(t-u)x(t-v)]\,du\,dv = \int_{-\infty}^{\infty} h_2(u,v)\,R_{xxx}(\tau-u,\tau-v)\,du\,dv \qquad 5.17$$

$R_{xxx}(\tau_1,\tau_2)$ can be written in the frequency domain as $\int_{-\infty}^{\infty} S_{xxx}(f_1,f_2)e^{j2\pi(f_1\tau_1+f_2\tau_2)}d\tau_1 d\tau_2$. Substituting this into equation 5.17 and taking a Fourier transform gives,

$$\begin{aligned}
&\int\int\int\int_{-\infty}^{\infty} h_2(u,v)\,S_{xxx}(f_1,f_2)\,e^{j2\pi(f_1(\tau-u)+f_2(\tau-v)}e^{-j2\pi\upsilon\tau}\,d\tau\,du\,dv\,df_1\,df_2 \\
&= \int\int\int_{-\infty}^{\infty} H_2(f_1,f_2)\,S_{xxx}(f_1,f_2)\,e^{-j2\pi\tau(f_1+f_2-\upsilon)}\,d\tau\,df_1\,df_2 \\
&= \int\int_{-\infty}^{\infty} H_2(f_1,f_2)\,S_{xxx}(f_1,f_2)\,\delta(f_1+f_2-\upsilon)\,df_1\,df_2 \\
&= \int_{-\infty}^{\infty} H_2(\upsilon-f_2,f_2)\,S_{xxx}(\upsilon-f_2,f_2)\,df_2 \qquad 5.18
\end{aligned}$$

Therefore, the Fourier transform of equation 5.14 can be written as the sum of equations 5.16 and 5.18 as,

$$S_{XY}(\upsilon) = H_1(\upsilon)S_{XX}(\upsilon) + \int_{-\infty}^{\infty} H_2(\upsilon - f_2, f_2)\, S_{XXX}(\upsilon - f_2, f_2)\ df_2 \qquad 5.19$$

Previously, to identify the quadratic component of the signal, the quadratic Volterra model (equation 4.86) was substituted into the second order cross correlation function (equation 4.91) to give,

$$\begin{aligned} R_{XXY}(\tau_1, \tau_2) = & \int_{-\infty}^{\infty} h_1(u)\, E[x(t-\tau_1)x(t-\tau_2)x(t-u)]\ du\ + \\ & \int\int_{-\infty}^{\infty} h_2(u,v)\, E[x(t-\tau_1)x(t-\tau_2)x(t-u)x(t-v)]\ du\ dv \end{aligned} \qquad 5.20$$

The first term of equation 5.20 can no longer be set to zero and but now be solved explicitly and so is expressed in terms of an auto bicorrelation function as,

$$\int_{-\infty}^{\infty} h_1(u)\, E[x(t-\tau_1)x(t-\tau_2)x(t-u)]\ du = \int_{-\infty}^{\infty} h_1(u)\, R_{XXX}(u-\tau_1, u-\tau_2)\ du \qquad 5.21$$

$R_{XXX}(\tau_1, \tau_2)$ can again be written in the frequency domain as $\int\int_{-\infty}^{\infty} S_{XXX}(f_1, f_2) e^{j2\pi(f_1\tau_1 + f_2\tau_2)} d\tau_1 d\tau_2$

Substituting this into equation 5.21 and taking a two dimensional Fourier transform gives,

$$\begin{aligned} & \int\int\int\int \int_{-\infty}^{\infty} h_1(u) S_{XXX}(f_1, f_2) e^{j2\pi(f_1(u-\tau_1)+f_2(u-\tau_2))} e^{-j2\pi\upsilon_1\tau_1} e^{-j2\pi\upsilon_2\tau_2} df_1 df_2 du\, d\tau_1 d\tau_2 \\ & = \int\int\int \int_{-\infty}^{\infty} H_1^*(f_1 + f_2) S_{XXX}(f_1, f_2) e^{-j2\pi(f_1\tau_1 + f_2\tau_2 + \upsilon_1\tau_1 + \upsilon_2\tau_2)} df_1 df_2\, d\tau_1 d\tau_2 \\ & = \int\int_{-\infty}^{\infty} H_1^*(f_1 + f_2) S_{XXX}(f_1, f_2) \delta(\upsilon_1 + f_1) \delta(\upsilon_2 + f_2) df_1 df_2 \\ & = H_1(\upsilon_1 + \upsilon_2) S_{XXX}^*(\upsilon_1, \upsilon_2) \end{aligned} \qquad 5.22$$

The second term of equation 5.20 can be rewritten in terms of an auto tricorrelation function as,

$$\begin{aligned} & \int\int_{-\infty}^{\infty} h_2(u,v)\, E[x(t-\tau_1)x(t-\tau_2)x(t-u)x(t-v)]\ du\ dv \\ & \quad = \int\int_{-\infty}^{\infty} h_2(u,v)\, R_{XXXX}(\tau_1 - \tau_2, \tau_1 - u, \tau_1 - v) \end{aligned} \qquad 5.23$$

As before, it is possible to write the auto tricorrelation function, $R_{xxxx}(\tau_1,\tau_2,\tau_3)$, in the frequency domain as $\int\int\int_{-\infty}^{\infty} S_{xxx}(f_1,f_2,f_3)e^{j2\pi(f_1\tau_1+f_2\tau_2+f3\tau3)}d\tau_1 d\tau_2 d\tau_3$ Substituting this into equation 5.23 and taking a two dimensional Fourier transform gives,

$$\int\int\int\int\int\int_{-\infty}^{\infty} h_2(u,v)\, S_{xxxx}(f_1,f_2,f_3)e^{j2\pi(f_1(\tau_1-\tau_2)+f_2(\tau_1-u)+f_3(\tau_1-v)}e^{-j2\pi(\upsilon_1\tau_1+\upsilon_2\tau_2)}$$
$$df_1df_2df_3d\tau_1d\tau_2dudv$$
$$=\int\int\int\int\int_{-\infty}^{\infty} H_2(f_2,f_3)\, S_{xxxx}(f_1,f_2,f_3)e^{-j2\pi(\tau_1(\upsilon_1-f_1-f_2-f_3)+\tau_2(\upsilon_2+f_1))}df_1df_2df_3d\tau_1d\tau_2$$
$$=\int\int\int_{-\infty}^{\infty} H_2(f_2,f_3)\, S_{xxxx}(f_1,f_2,f_3)\delta(\upsilon_1-f_1-f_2-f_3)\delta(\upsilon_2+f_1)df_1df_2df_3$$
$$=\int_{-\infty}^{\infty} H_2(\upsilon_1+\upsilon_2-f_3,f_3)\, S_{xxxx}(-\upsilon_2,\upsilon_1+\upsilon_2-f_3,f_3)df_3 \qquad 5.24$$

Hence the two dimensional Fourier transform of equation 5.20 can be written as the sum of equations 5.22 and 5.24 as,

$$S_{xxy}(\upsilon_1,\upsilon_2)=H_1(\upsilon_1+\upsilon_2)S_{xxx}{}^*(\upsilon_1,\upsilon_2)+\int_{-\infty}^{\infty} H_2(\upsilon_1+\upsilon_2-f_3,f_3)\, S_{xxxx}(-\upsilon_2,\upsilon_1+\upsilon_2-f_3,f_3)df_3 \qquad 5.25$$

Therefore, in order to solve the quadratic Volterra model, in the general case, it is necessary to simultaneously solve equations 5.19 and 5.25. Note, that if x(t) is Gaussian the second term of equation 5.19 is zero, the first term of equation 5.25 is zero, and the second term of equation 5.25 can be expressed as the product of two power spectra, to give the familiar result from chapter four.

It is possible to solve equations 5.19 and 5.25 in an iterative manner in order to obtain expressions for $H_1(f_1)$ and $H_2(f_2,f_3)$. However, Kim and Powers [46] have produced a closed form solution using matrix methods. With both methods, care must be taken to ensure that the variance of the bispectrum and trispectrum is minimised. By removing the assumption of a Gaussian input, the complexity of the solution has increased and it can now be seen that there is a term involving the trispectrum in the expression for the quadratic

kernel. The same method can be used to solve a cubic Volterra model, however, the complexity is vastly increased and it is found necessary to use a sixth order spectrum. This is impractical using current technology.

5.6 Optimality of the Volterra models

In the previous chapters, frequency domain techniques for estimating Volterra kernels have been discussed and a modification based on subtracting off lower order terms in order to minimise the variance suggested. In this section these frequency domain methods are now compared with the optimal least squares time domain method.

It is shown in Appendix F that the optimal least squares estimate for a linear filter **h** is given by the solution of,

$$\mathbf{Xh} = \mathbf{y} \tag{5.26}$$

where X and y are as defined in Appendix F. This leads to,

$$\mathbf{h} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{y}) \tag{5.27}$$

The same method can be extended to obtain a solution for the optimum n^{th} order filters for a system that contains nonlinear elements. Here, just the linear and quadratic filters will be considered, and so the matrix, X, will now contain quadratic terms as well as the linear terms and is defined as,

$$\mathbf{X} = \begin{bmatrix} x(L) & x(L-1) & \cdots & x(L-N+1) & x(L)^2 & x(L)x(L-1) & \cdots & x(L)x(L-N+1) & X(L-1)^2 & \cdots & x(L-N+1)^2 \\ x(L-1) & x(L-2) & \cdots & x(L-N) & x(L-1)^2 & x(L-1)x(L-2) & \cdots & x(L-1)x(L-N) & X(L-2)^2 & \cdots & x(L-N)^2 \\ x(L-2) & x(L-3) & \cdots & x(L-N-1) & x(L-2)^2 & x(L-2)x(L-3) & \cdots & x(L-1)x(L-N-1) & X(L-3)^2 & \cdots & x(L-N-1)^2 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ x(3) & x(2) & \cdots & 0 & x(3)^2 & x(3)x(2) & \cdots & 0 & x(2)^2 & \cdots & 0 \\ x(2) & x(1) & \cdots & 0 & x(2)^2 & x(2)x(1) & \cdots & 0 & x(1)^2 & \cdots & 0 \\ x(1) & 0 & \cdots & 0 & x(1)^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

From this X^TX can be formed which can be seen to consist of three distinct components: a linear component which is the same as for the linear solution; a quadratic component; and a component consisting of linear/quadratic cross terms. However, for a system with a Gaussian input, as in this case, the cross terms, on average, will be zero.

$$X^TX = \begin{bmatrix} \text{LIN} & \text{LQ} \\ \text{LQ}^T & \text{QUAD} \end{bmatrix}$$

Multiplying the inverse of this matrix with the cross correlation vector formed by the multiplication of X^T with **y**, as in Appendix F, gives the linear and quadratic filter vector, **h**, which is of the form,

$$\mathbf{h} = \begin{bmatrix} h_{LIN} \\ h_{QUAD} \end{bmatrix}$$

From **h** it is possible to reconstruct in full the linear and quadratic filters.

This least squares optimal time domain method was then used to solve for the linear and quadratic Volterra kernels for the quadratic mixed system as in section 4.6.2. The following results are based on using a data length of 10000 samples and a FFT size of 32. Figure 5.12 shows the quadratic impulse response estimated by this method. These results are compared with the frequency domain methods for estimating the quadratic impulse response: firstly without subtracting of the linear term, figure 5.13; and secondly with the linear term removed, figure 5.14.

It can be seen that the least squares time domain method gives a solution containing less noise than its frequency domain counterparts. This is as expected as it represents an

optimal solution. However, the modified frequency domain method produces reasonable results which are far better than traditional frequency domain methods. The standard deviation of the residual error, e(n), is defined as,

$$\sqrt{\frac{1}{L}\sum_{n=0}^{L-1} e(n)^2} \qquad 5.28$$

(Further explanation can be found in Appendix F.) This was calculated for each of the methods and is shown in table 5.1. The time domain method has the smallest standard deviation (0.03), followed by the modified frequency domain method (0.06). However the traditional frequency domain method can be seen to have a significantly larger error, the standard deviation of which is 0.25.

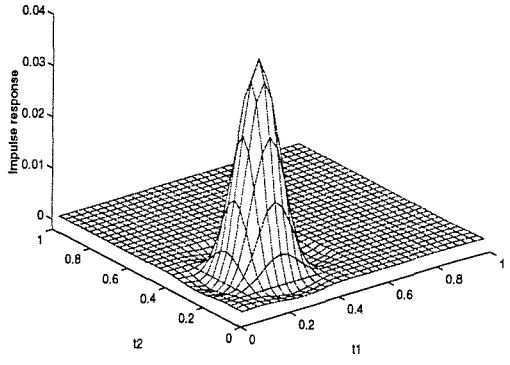

Figure 5.12: Quadratic impulse response of mixed system calculated using the least squares optimum time domain method.

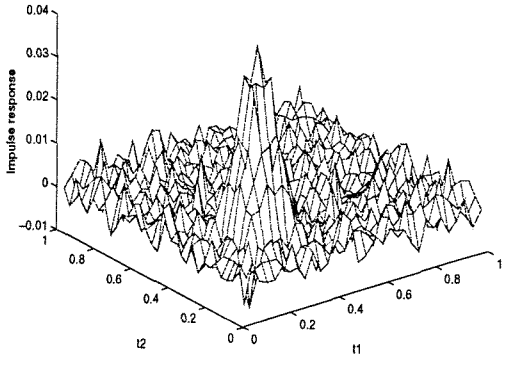

Figure 5.13: Quadratic impulse response of mixed system calculated using the traditional frequency domain method.

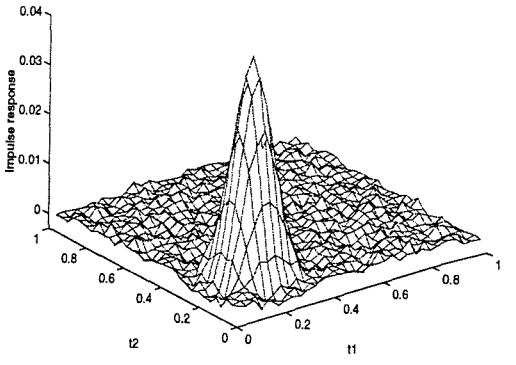

Figure 5.14: Quadratic impulse response of mixed system calculated using the modified traditional frequency domain method.

Although the time domain method gives the best results it is far more computationally intensive. For the above quadratic model with an FFT size of 32 it requires the multiplication of two 1024 by 560 matrices and then the inversion of a 560 by 560 matrix. For a more realistic FFT size of 64 these dimensions are increased to 4096 and 2144. If a cubic model is attempted using this method both these figures are increased by approximately a factor of 64 and the computation becomes unfeasible large, using available compute power. However it can be seen that although the modified frequency domain method is not optimal, it is much closer to optimal than traditional frequency domain methods and has the major advantage that it can be easily calculated.

Method	Standard deviation of error
Time domain least squares	0.032
Frequency domain	0.251
Modified frequency domain	0.060

Table 5.1: Standard deviation of error for the different methods

5.7 Causality and the Volterra series

Consider the linear system identification problem shown in figure 5.15 where x(t) and y(t) are the input and output measurements of the system. The problem is to establish a linear transfer characteristic linking the two. Let $y_0(t)$ be a stationary random process produced by operating on x(t) linearly. That is,

$$y_0(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \qquad 5.29$$

and $e(t) = y(t)-y_0(t)$ denote the error. The system identification problem is to find the transfer function, $h(\tau)$ that minimises $E[e^2(t)]$. This can be easily solved to give the Wiener-Hopf equation:

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} h(\tau)R_{xx}(t-\tau)d\tau \qquad 5.30$$

Fourier transforming equation 5.30, leads to the standard results:

$$H(f) = \frac{S_{xy}(f)}{S_{xx}(f)} \qquad 5.31$$

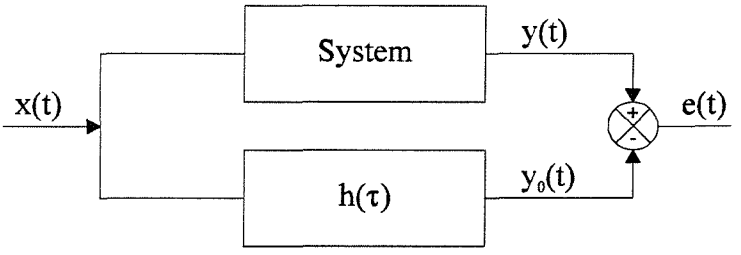

Figure 5.15: Linear system identification

Note however that the limits of the integral in equation 5.30 are from minus infinity to infinity. This means that H(f) will not necessarily be causal. If the restriction that $h(t) = 0$, $t > 0$ is applied, the estimated system will always be causal, but, it is then not possible to

simply Fourier transform equation 5.30 to obtain equation 5.31. In order to obtain a solution in the frequency domain, the method of spectral factorisation [64] has to be used.

The same problem will occur with the Volterra kernels. On all the Volterra models used in this work, the limits on the integral are from minus infinity to infinity. This means that the Volterra kernels estimated will not necessarily be causal. One possible approach to this problem would be to generalise the concepts of spectral factorisation but this is not pursued here. Another approach to avoiding non causality would be to use a parametric method in which the model could be limited to being causal. However, for the practical systems studied in chapter six, all the Volterra kernels are in fact causal, as would be hoped with a real system, and so in these cases there is not a problem.

5.8 Overview of Volterra models

A number of different methods for estimating Volterra kernels have now been presented, each of which require certain assumptions and have certain features. Different techniques have also been used by other authors to solve some of the Volterra models. In this section, a summary of the main models, together with references to other work, is given.

Some of the more important features of estimation schemes for Volterra models are listed below:

- The order of the model - either quadratic (can only analyse skewed type nonlinearities) or cubic (can analyse both skewed and symmetric nonlinearities).

- The conditions imposed on the input - either Gaussian or non-Gaussian.

- The method used to solve the model - either an iterative or closed form solution.

- The highest order spectra needed to calculate the Volterra kernels - realistically it is only possible to calculate the bispectrum and trispectrum.

Figure 5.16 shows some of the possible quadratic Volterra models that can be created. The simplest of all is the quadratic Volterra model, assuming a Gaussian input, which was first proposed by Tick [76] in 1961. It requires only the calculation of the spectrum and bispectrum. However it is strictly limited to skewed type nonlinearities. If the assumption of Gaussianity is relaxed, to solve the quadratic model it is necessary to utilise the trispectrum. The Volterra model must then be solved either in an iterative manner or by a closed from solution as in Kim and Powers [46].

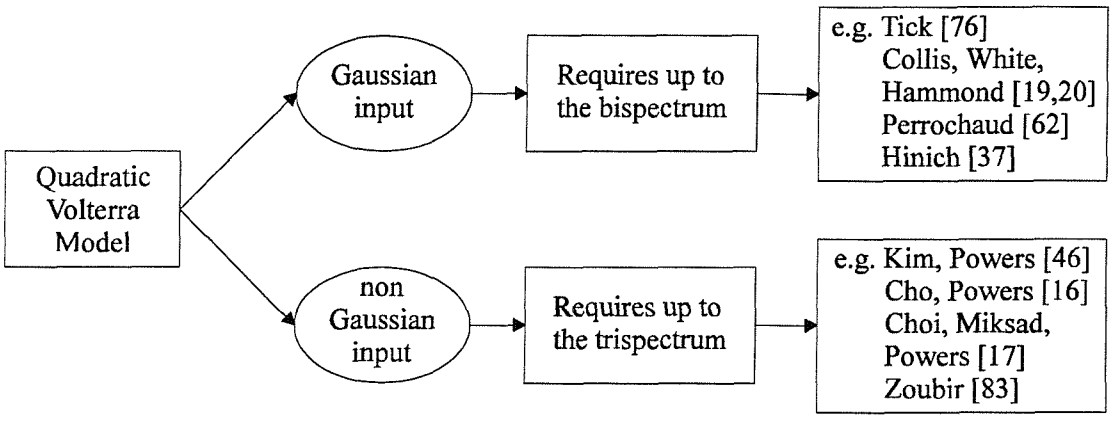

Figure 5.16: Types of quadratic Volterra model

Figure 5.17 shows the possible cubic Volterra models. A general solution, with no restrictions on the statistics of the input has been presented by Nam and Powers [55]. However, this requires the calculation of up to a sixth order spectrum and so is of little practical use. If the input is assumed to be Gaussian, only the power spectrum, bispectrum and trispectrum are required. As this model can handle both skewed and symmetric nonlinearities, it is likely to be of most practical use, and is the model used for the experimental work in chapter six.

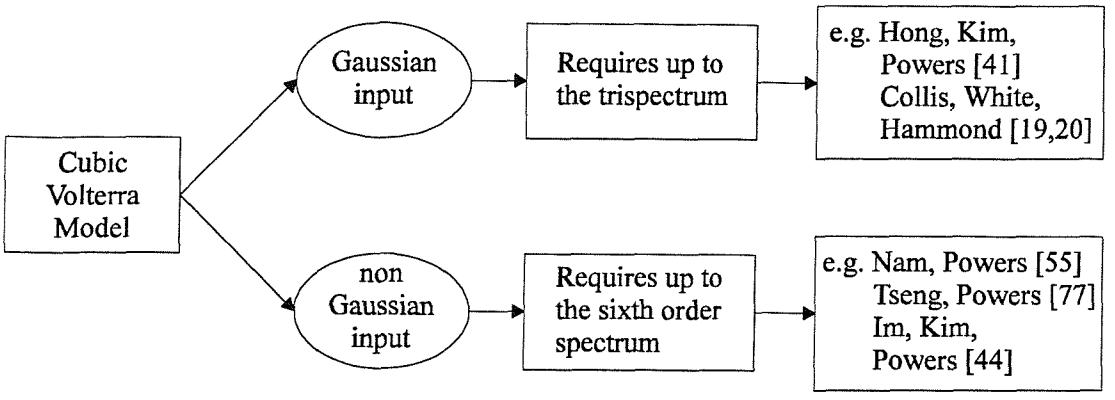

Figure 5.17: Types of cubic Volterra model

5.9 Conclusions

In this chapter it has been shown that care must be taken when estimating the Volterra kernels of the system, firstly to check that the Volterra series is likely to converge for the particular system, and also to check that the calculated Volterra kernels are causal. Having estimated the Volterra kernels of a particular system it has been shown that it is then possible to predict the response of the system to any input.

A quadratic Volterra model has been developed which places no restrictions on the statistics of the input. However, it is far more complex than the Volterra model that assumes a Gaussian input, and requires the calculation of the trispectrum. The quadratic model can of course only analyse skewed nonlinearities and so is fairly limited. If the cubic model, which can analyse both skewed and symmetric type nonlinearities, is extended to any input it is found that spectra up to the sixth order have to be calculated which prohibits its practical use. It is therefore suggested, that the cubic Volterra model that assumes a Gaussian input is likely to be of the most practical use. In the next chapter, it is this model that is used to analyse the experimental results.

Chapter 6

Practical examples of the use of higher order spectra in mechanical systems

6.1 Introduction

In this chapter higher order spectral techniques are applied to some simple mechanical systems. The first system studied is the nonlinear magnetic beam. This is a simply supported beam, driven by an electromagnetic shaker. At the free end, pairs of repelling magnets are placed. By varying the position and number of magnets, the nature of the nonlinearity can be changed, be it skewed or symmetric, and by varying the distance between the magnets the strength of the nonlinearity can also be altered. Using this controllable system, auto higher order spectral methods are applied, assuming only a knowledge of an output signal. These are then compared with the results obtained from cross higher order spectral techniques, where both an input and output are known.

For the magnetic beam, the expected nature of the nonlinearity was known; however, in the second example, there is no prior knowledge of the type of the nonlinearity. This system consists of a beam attached to electromagnetic shaker. By loosening the attachment bolt, a rattling type nonlinearity can be introduced to the system. Varying the degree of tightness will vary the degree of nonlinearity. As in the previous example both auto and cross higher order spectral techniques are compared. In this system the precise nature of the nonlinearity is unclear, and it is not certain that the Volterra series will converge. However, a cubic Volterra model is applied to the system and seems to produce useful results.

6.2 The nonlinear magnetic beam

The experimental set up for the symmetric nonlinear magnetic beam experiment, figure 6.1, consists of a beam, clamped at one end and attached to a shaker which is excited by a Gaussian signal. At its tip two pairs of repelling magnets are placed such that there is a symmetric nonlinear restoring force that tries to centralise the beam between the magnets. The strength of the nonlinearity can be varied by changing the distance between the magnets. By removing the top pair of magnets, as in figure 6.2, this symmetric restoring force becomes skewed as the beam is no longer constrained from above.

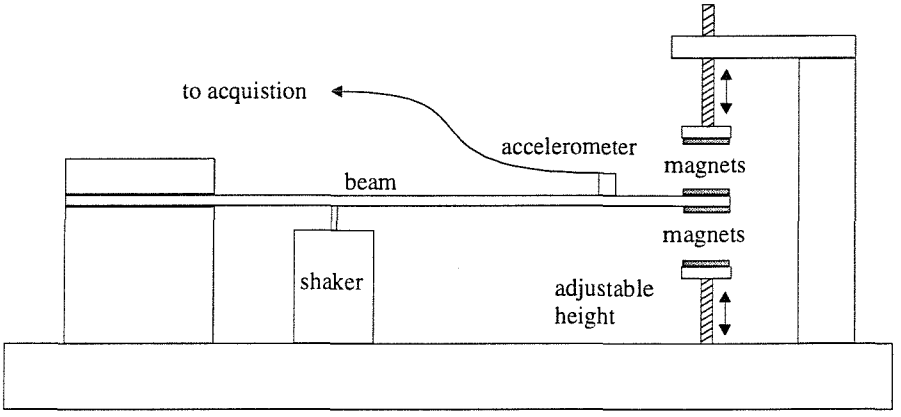

Figure 6.1: Experimental set up for symmetric magnetic beam

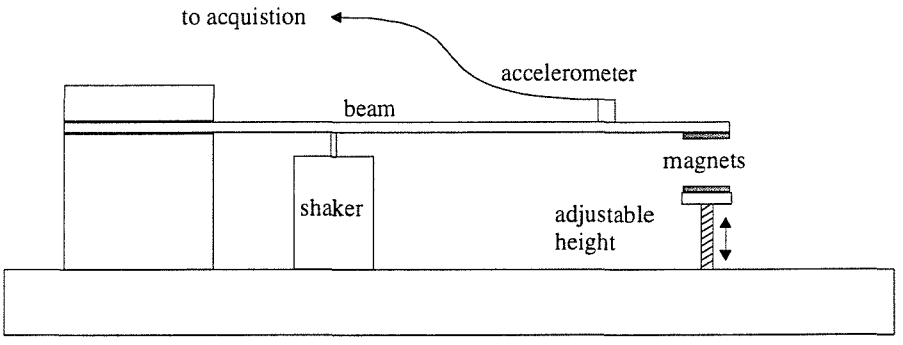

Figure 6.2: Experimental set up for skewed magnetic beam

The input signal to the shaker and the output signal from the accelerometer were both synchronously sampled into a P.C. and stored ready for analysis. The input to the shaker was recorded as opposed to the force on the beam as it is a requirement that the input to the Volterra model is Gaussian. Although the input to the shaker is Gaussian, the force on the beam will not necessarily be as the beam will induce a nonlinear force onto the shaker.

6.2.1 Auto higher order spectral methods

Initially all magnets are removed from the beam to produce a predominantly linear system. A Gaussian signal, with an upper frequency of 50 Hz was passed through a power amplifier and into the shaker. The signal from the accelerometer on the beam was sampled at 100 Hz having been passed through an anti-alias filter with cut-off frequency 40 Hz. Approximately 262144 samples were recorded.

The power spectrum of the response from the beam with no magnets is shown in figure 6.3. It can be seen that there are a number of resonances and anti-resonances in the system and the first resonance is at approximately 5 Hz. This can be predicted using the Bernoulli-Euler equation. For a clamped free beam it can be shown that the frequency of the first resonance is given by [24],

$$f_1 = 0.56\sqrt{\frac{EI}{m}}\frac{1}{l^2} = 4.9\text{Hz} \qquad 6.1$$

where m is the mass (0.102 kg), E is the Young's modulus (68.9×10^{-9} Nm^{-2}), I is the second moment of area (7.03×10^{-12} m^4), and l is the length (0.494 m).

The auto bicoherence is shown in figures 6.4 and 6.5. Figure 6.5 is a mesh plot, used to give an idea of the overall magnitude of the bicoherence, which in this case is approximately zero. Figure 6.4 is a contour plot which shows the precise detail of the bicoherence. The dashed lines on the contour plots of the bicoherence mark the position of the peaks and troughs in the spectrum, as it is often useful to be able to correlate power spectral effects

with peaks in the bicoherence. The same contour levels are used for all the contour plots in this chapter so as to allow for easy comparison between plots. In this case, there is no significant structure in the bicoherence, indicating that there are no quadratic type nonlinearities in the system. The tricoherence is shown in figure 6.6, and again, there is no significant structure, indicating that there are no cubic type nonlinearities. As both the bicoherence and tricoherence are approximately zero, it is reasonable to suppose that the beam can be approximated by a linear system.

The lower set of magnets are replaced to produce a system that should be dominantly skewed, as shown by the configuration in figure 6.2. The power spectrum is shown in figure 6.7. It can be seen that the position of the main resonance has increased which is partly due to the weight of the magnets on the tip of the beam. A number of possible harmonics of the fundamental frequency at approximately 8 Hz have also appeared close to 16 Hz and 23 Hz.

The bicoherence, figures 6.8 and 6.9, show significant quadratic phase coupling (maximum value = 0.7) between the harmonically related peaks in the spectrum, in particular between the 8 Hz and 16 Hz peaks which produce the largest interaction at (8,8) Hz. There is some structure in the tricoherence, figures 6.10 and 6.11, although it is not of particularly large magnitude (maximum value = 0.1) and again it occurs at (8,8,8) Hz. Two different angles are used to display the trispectrum, so as to gain an overall view of where the interactions occur as well as more accurate positioning of their frequencies.

With the use of auto higher order spectra, it has been possible to determine that the nonlinearity has more dominant quadratic terms than cubic terms. It is also easy to tell from the bicoherence that the fundamental frequency and its first two harmonics are related by quadratic phase coupling, something that the power spectrum can give us no information about.

Finally, both sets of magnets are replaced, to obtain a system with a symmetric nonlinearity, the power spectrum of which is shown in figure 6.12. However, because of the strength of the magnets, and the fact they were not perfectly aligned, not all the vibration was in the

vertical plane. For this experiment the sampling frequency was increased to 200 Hz and the signal passed through an anti-alias filter with cut off frequency 90 Hz. This is to account for the fact that the dominant modes have increased in frequency.

The bicoherence (figures 6.13 and 6.14) shows a single peak at (35,35) Hz, indicating the presence of some quadratic phase coupling, although it is half the magnitude of the peak for the skewed system (maximum value = 0.3). The tricoherence (figures 6.15 and 6.16) shows a large peak of magnitude 0.4 indicating a very strong cubic interaction. Although the nonlinearity is dominantly symmetric as seen by the peak in the trispectrum, there still appears to be significant skewed activity, which could be caused by the extra vibration that is not in the vertical plane.

In this experiment, auto higher order spectral methods have been shown to provide useful and interesting information. They can be used to detect whether there is nonlinearity present in a system, what the dominant nature of the nonlinearity is, and can also give information about phase coupling between peaks in the spectrum. In the next section the same experiment will be used to see how much more information about the system can be obtained by assuming knowledge of the input, and using cross higher order spectral techniques.

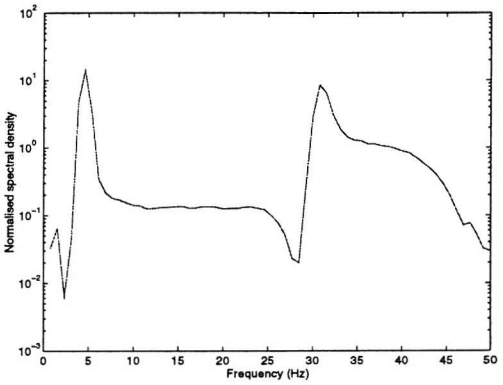

Figure 6.3: Normalised power spectrum of the linear beam

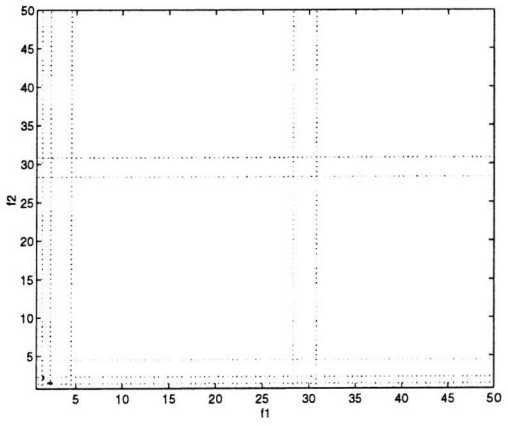

Figure 6.4: Contour plot of the bicoherence of the linear beam

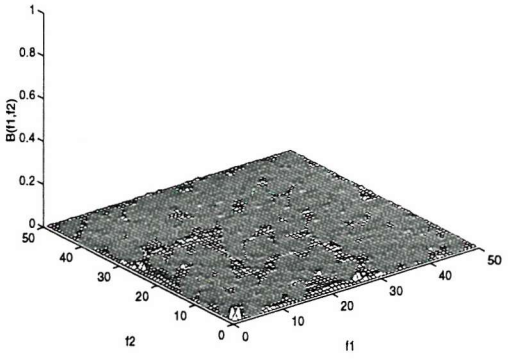

Figure 6.5: Mesh plot of the bicoherence of the linear beam

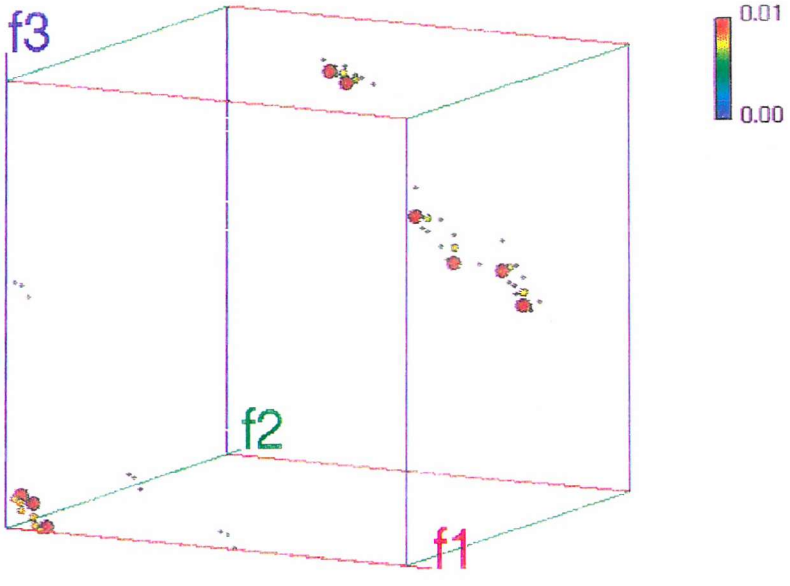

Figure 6.6: Tricoherence of the linear beam

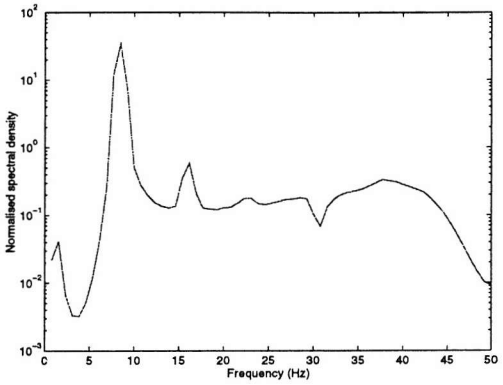

Figure 6.7: Normalised power spectrum of the skewed magnetic beam

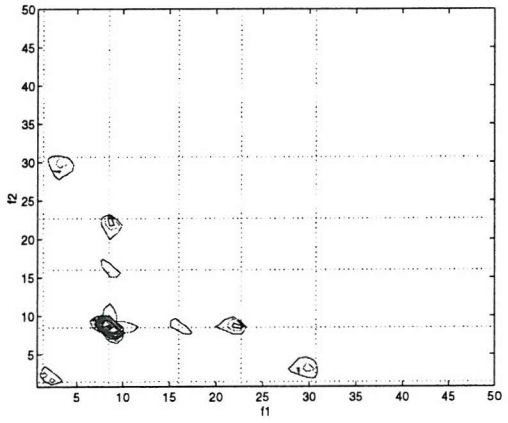

Figure 6.8: Contour plot of the bicoherence of the skewed magnetic beam

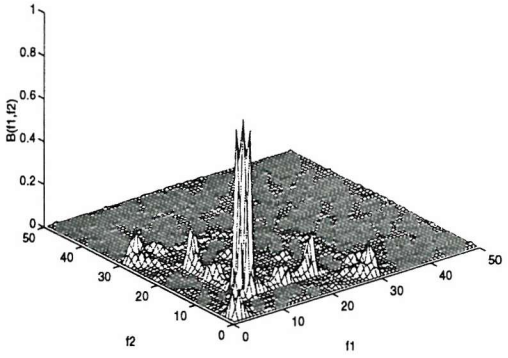

Figure 6.9: Mesh plot of the bicoherence of the skewed magnetic beam

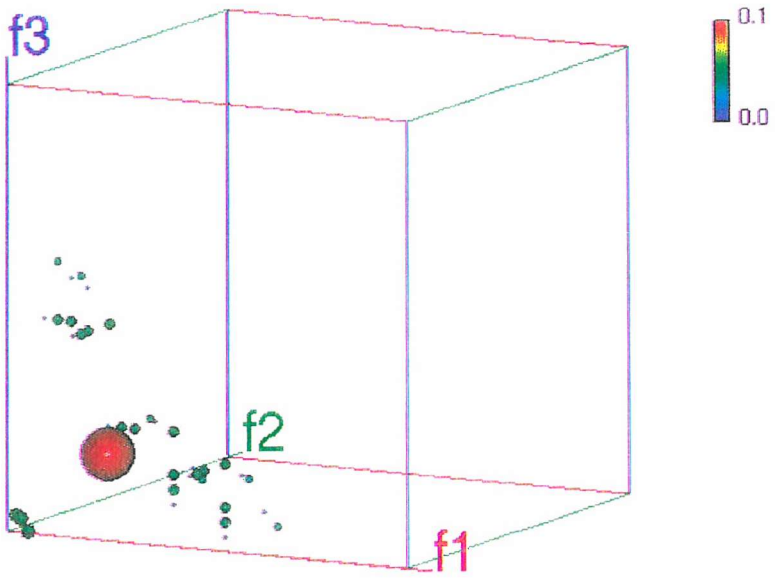

Figure 6.10: Tricoherence of the skewed magnetic beam

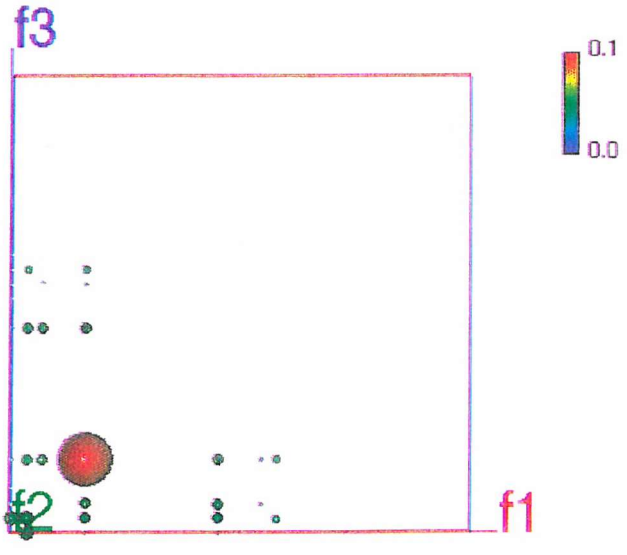

Figure 6.11: Tricoherence of the skewed magnetic beam (end view)

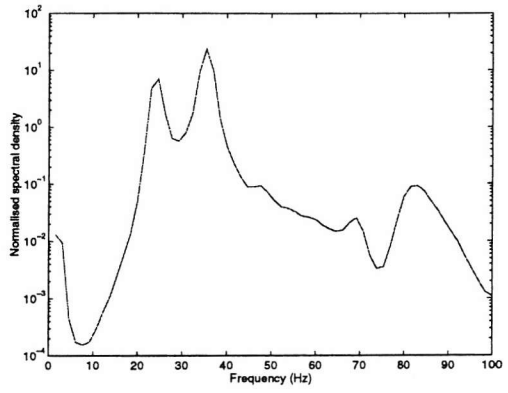

Figure 6.12: Normalised power spectrum of the symmetric magnetic beam

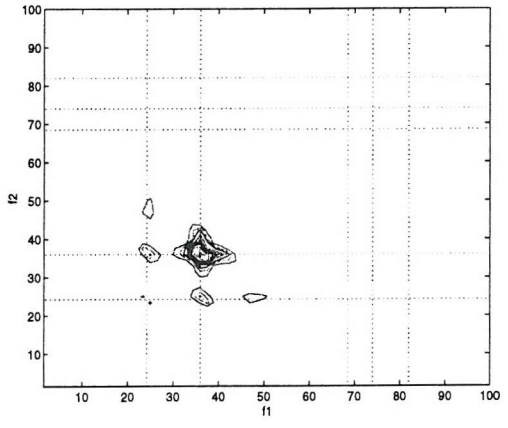

Figure 6.13: Contour plot of the bicoherence of the symmetric magnetic beam

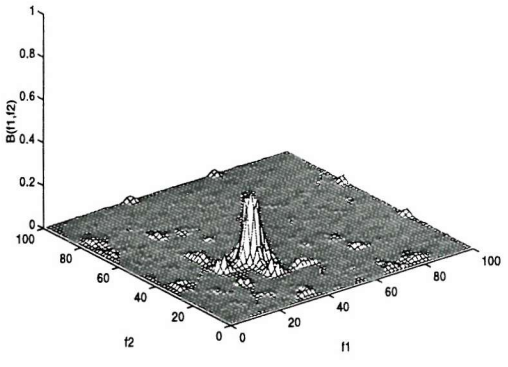

Figure 6.14: Mesh plot of the bicoherence of the symmetric magnetic beam

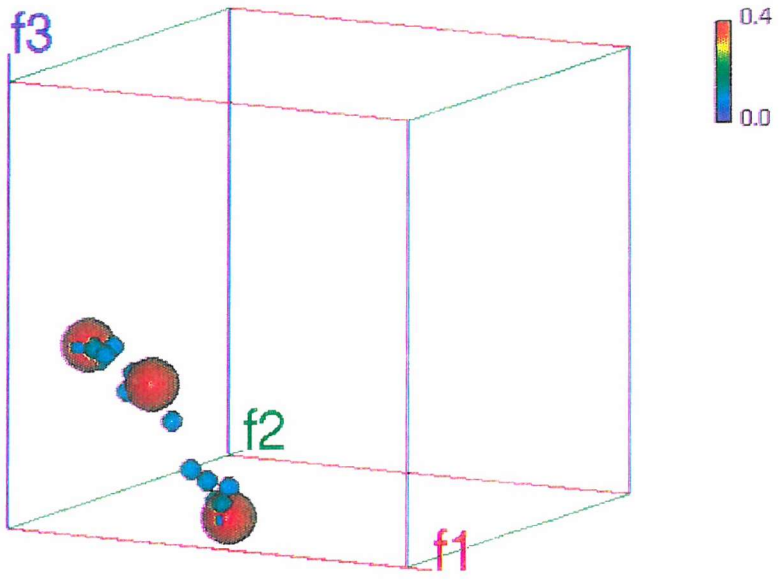

Figure 6.15: Tricoherence of the symmetric magnetic beam

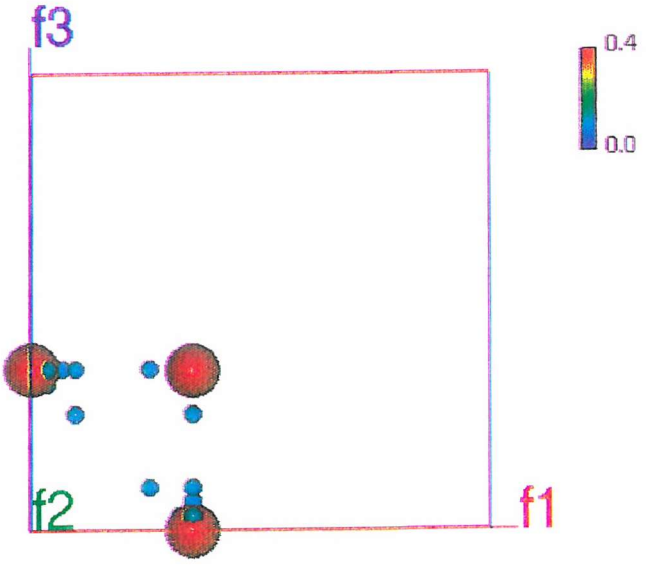

Figure 6.16: Tricoherence of the symmetric magnetic beam (end view)

6.2.2 Cross higher order spectral methods

Using the measurements of the system input and output it is possible to calculate a measure of the ordinary coherence function for the beam in each state. The ordinary coherence function takes a value of unity if there is a linear relationship between the input and output. For the beam with all the magnets removed, the coherence function is shown in figure 6.17. It is close to unity everywhere, indicating a linear transmission path between the input and output. The coherence does dip in the regions where there are resonances in the system but this feature is well understood, being due to the resolution of the estimate, and not necessarily to nonlinearity. The linear transfer function, figure 6.19, and the impulse response, figure 6.18, are estimated, and from these, the output of this linear approximation can be calculated. The power spectrum of the linear component is shown in figure 6.20 by the solid line, together with the actual output spectrum of the beam, shown by the dotted line. It can be seen to account for nearly all the power in the signal, apart from a small amount around the first resonance, thus verifying that with no magnets the beam can be adequately modelled by a linear system.

The coherence is reduced, figure 6.21, for the beam with the lower set of magnets in place, indicating that there is no longer a purely linear relationship between input and output. Using the cubic Volterra model, figure 6.22, it is now hoped to shed more light on what type of nonlinearities are acting in the system and what frequencies they affect most.

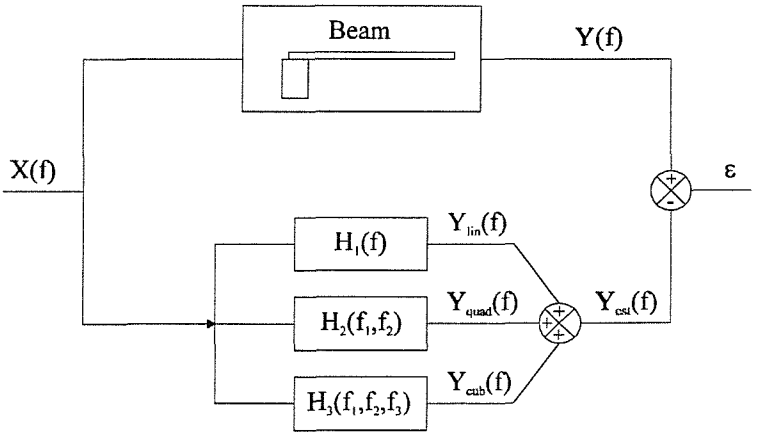

Figure 6.22: Frequency domain cubic Volterra model of beam

With the lower set of magnets in place, the iterative procedure, described in chapter four for calculating the Volterra kernels of a cubic system is carried out. Initially it is assumed that the cubic term is zero. The figures in this chapter, are all shown for the second iteration of the process, although for a system such as this, with no significant cubic term, the differences between the results for the first and second iteration are negligible.

The linear transfer function, $H_1(f)$, figure 6.24, and impulse response, $h_1(t)$, figure 6.23, for the system are calculated. The linear component of the signal is then obtained from the convolution of the impulse response with the original input data. Having removed the linear component, the cross bispectrum and hence the second order frequency domain Volterra kernel, $H_2(f_1,f_2)$, are calculated, figure 6.25. From this, the quadratic component of the signal is estimated by performing a two dimensional convolution of the input signal with the quadratic time domain kernel, figure 6.26. Finally, the trispectrum and hence the third order frequency domain Volterra kernel, $H_3(f_1,f_2,f_3)$, figure 6.27, is calculated and from this the cubic component of the signal is formed by a three dimensional convolution of the cubic time domain kernel, figure 6.28, with the input signal.

Figure 6.29 shows the power spectrum of the linear (thick solid line), quadratic (dashed line) and cubic (thin solid line) components together with the output power spectrum (dotted line) of the system. The linear component can be seen to account for the main resonance at 8 Hz but the peaks at 16 Hz and 23 Hz are accounted for by the quadratic kernel, thus confirming the results obtained from the auto bispectrum. Since the restoring forces are skewed then one anticipates that the quadratic kernel will be more important than the cubic one. The cubic terms caused by symmetric nonlinearities are seen to produce no significant power at any frequency. Figure 6.30 shows the output of the Volterra model (solid line) and the output of the beam (dotted line). The model accounts for the majority of the output power from the beam apart from at low frequencies which is possibly due to the poor signal to noise ratio.

The final experiment involved replacing both the upper and lower magnets, thus producing a system with a symmetric nonlinear restoring force. The coherence function of this system

can be seen in figure 6.31. The same procedure as above was followed and the linear (figure 6.33), quadratic (figure 6.35), and cubic (figure 6.37) frequency domain Volterra kernels estimated from higher order spectra. Multi-dimensional inverse Fourier transforms were then used to create the time domain kernels (figures 6.32, 6.34, and 6.36 respectively) which were then convolved with the original input data to give the linear, quadratic, and cubic components of the signal. These are shown in figure 6.38. It can be seen that the cubic component (thin solid line) is now more significant than the quadratic component (dashed line) thus indicating that the dynamics are now dominated by symmetric rather than skewed nonlinearities. However there are still parts of the signal generated by skewed nonlinearities and it can be seen that the slight peak between 70 Hz and 90 Hz is only present in the quadratic term. Figure 6.39 shows the output of the Volterra model (solid line) together with the output of the beam (dotted line) and again, barring low frequencies, the model has correctly accounted for most of the power across all frequencies.

6.2.3 The interpretation of higher order impulse response functions

A 'typical' linear impulse response function will hopefully oscillate and then die away. The quadratic and cubic impulse response functions behave similarly although a brief explanation of how to interpret this from the plots is needed.

The quadratic impulse response function in figure 6.25 can be seen to have a large peak at the origin and decay away along the main diagonal. It will also fade along other diagonals from the origin, but in this case, these diagonals reduce in magnitude much faster and so cannot be obviously seen.

Interpretation is yet more difficult for the cubic impulse response. This can be seen in figure 6.36 where the impulse response function oscillates and dies away mainly along the main diagonal. This is shown in the plot by the colours changing from red (large positive) through to blue (large negative) and back again. Further examples on the interpretation of this type of plot are given in Appendix C.

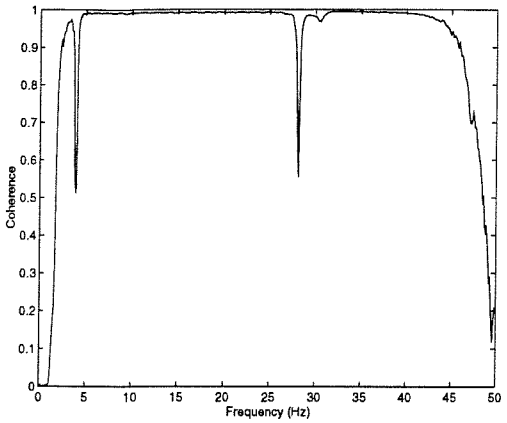

Figure 6.17: Coherence function of the beam with no magnets

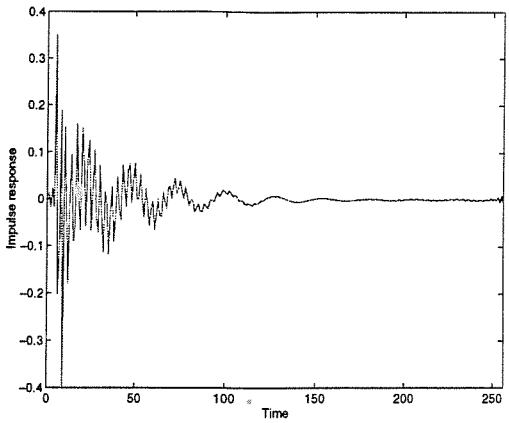

Figure 6.18: Linear impulse response of the beam with no magnets

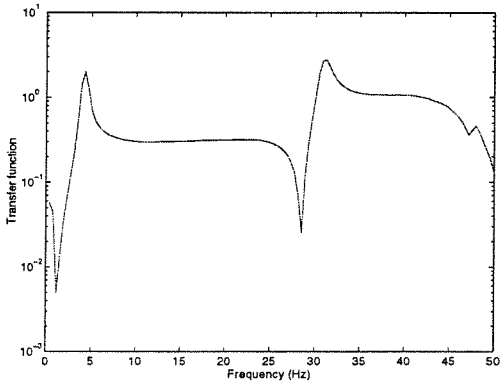

Figure 6.19: Transfer function of the beam with no magnets

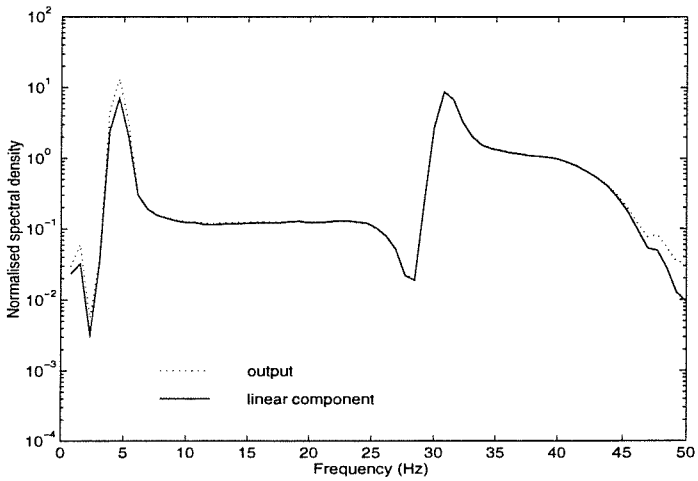

Figure 6.20: Power spectra of the linear component of the beam with no magnets and power spectrum of the output of the beam

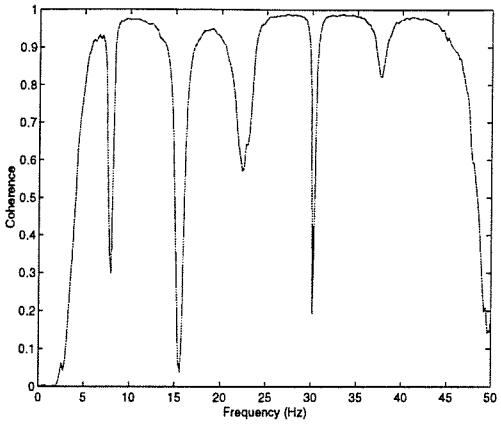

Figure 6.21: Coherence function of the skewed magnetic beam

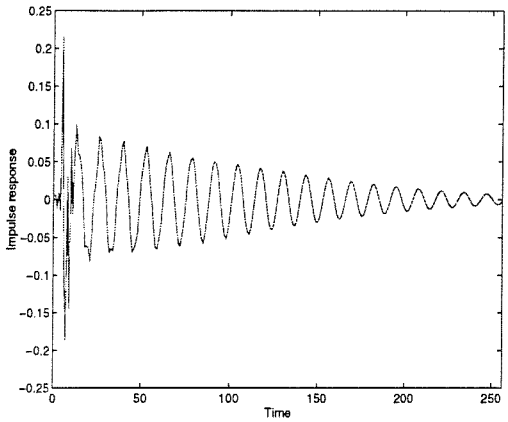

Figure 6.23: Linear impulse response of the skewed magnetic beam.

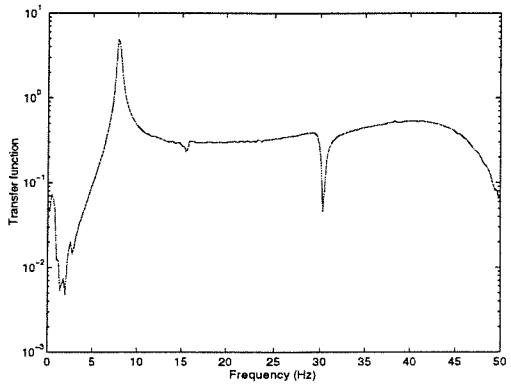

Figure 6.24: Linear transfer function of the skewed magnetic beam

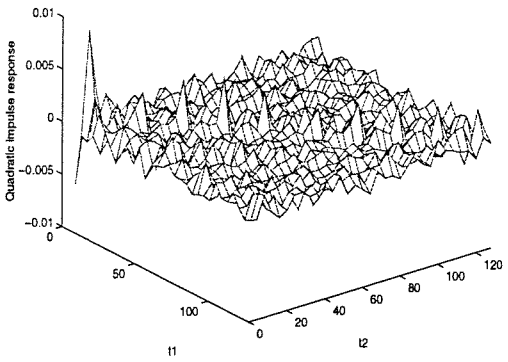

Figure 6.25: Quadratic impulse response of the skewed magnetic beam

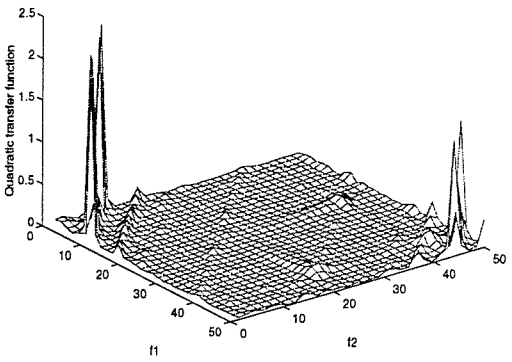

Figure 6.26: Quadratic transfer function of the skewed magnetic beam

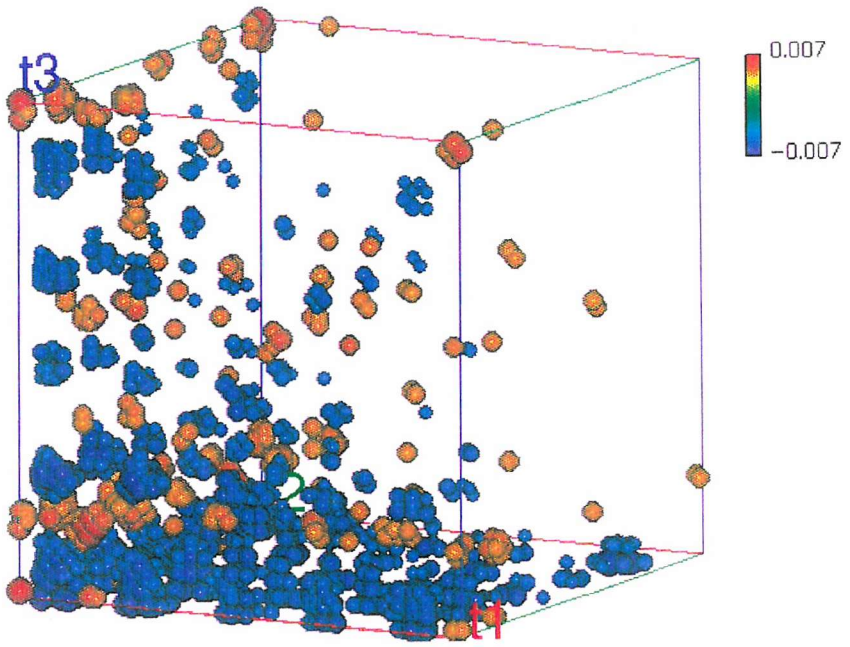

Figure 6.27: Cubic impulse response of the skewed magnetic beam

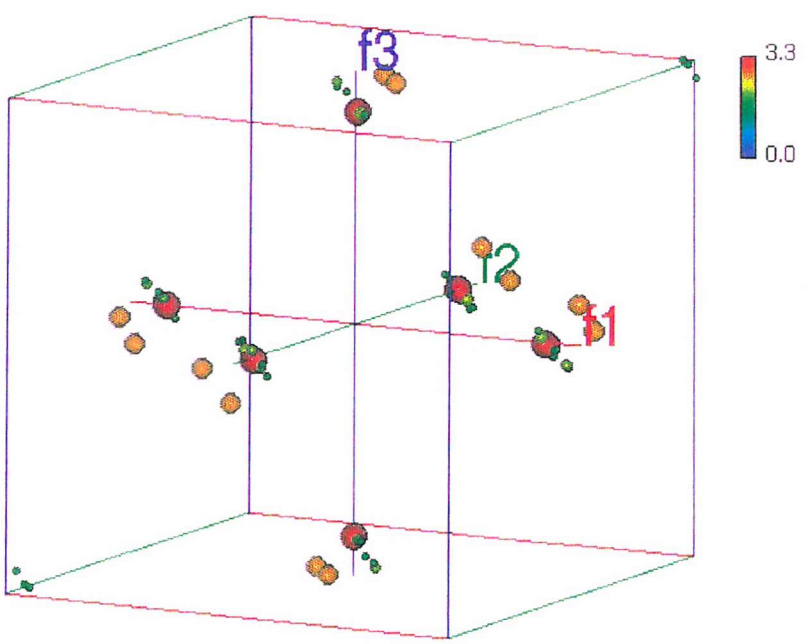

Figure 6.28: Cubic transfer function of the skewed magnetic beam

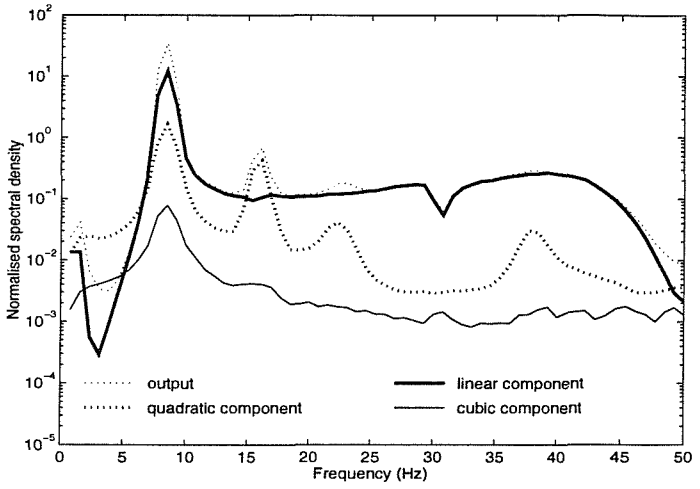

Figure 6.29: Power spectra of the linear, quadratic and cubic components of the Volterra model and the power spectrum of the output of the skewed magnetic beam.

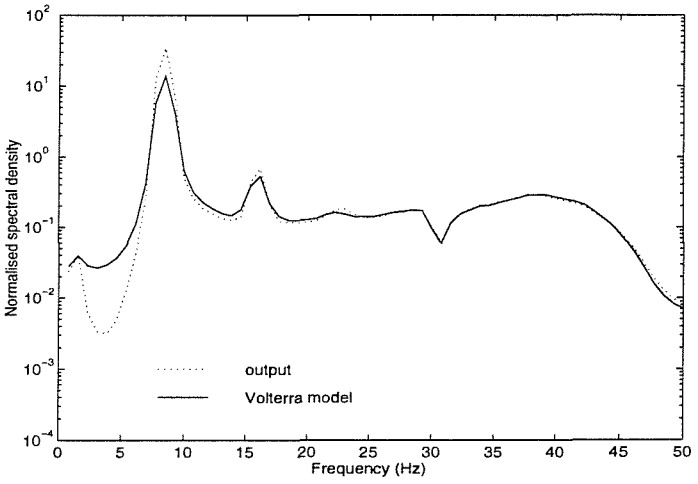

Figure 6.30: Power spectra of the Volterra model and output of the skewed magnetic beam.

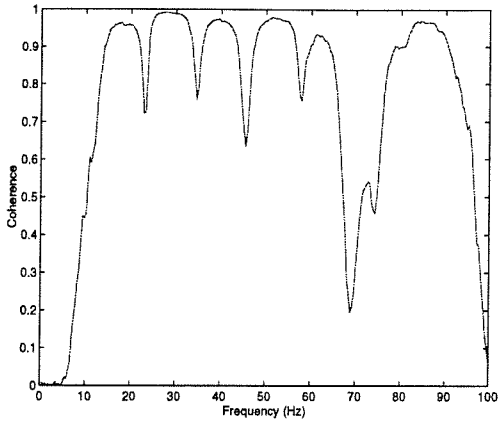

Figure 6.31: Coherence function of the symmetric magnetic beam

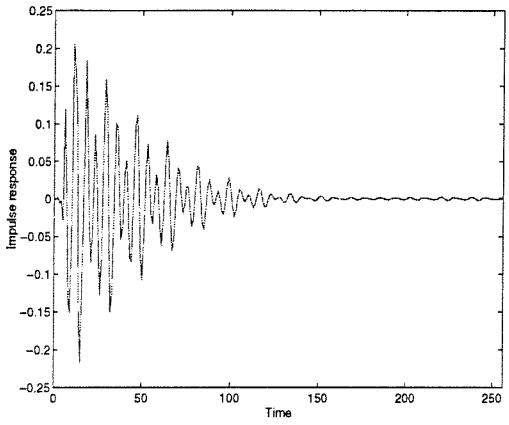

Figure 6.32: Linear impulse response of the symmetric magnetic beam.

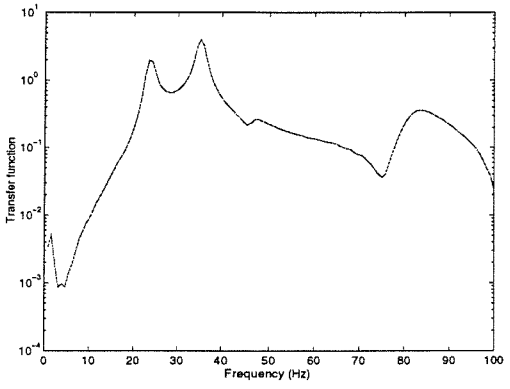

Figure 6.33: Linear transfer function of the symmetric magnetic beam.

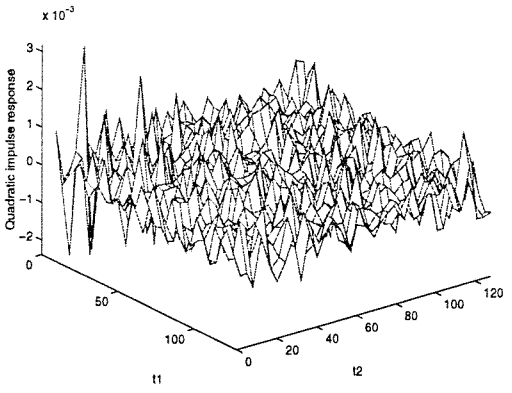

Figure 6.34: Quadratic impulse response of the symmetric magnetic beam.

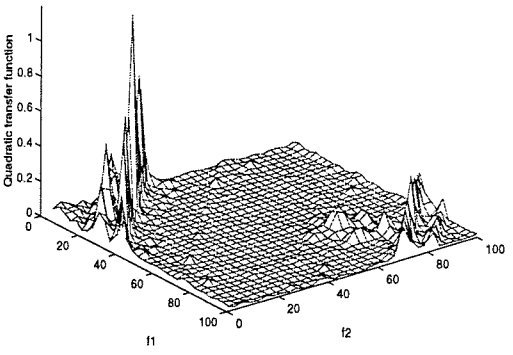

Figure 6.35: Quadratic transfer function of the symmetric magnetic beam.

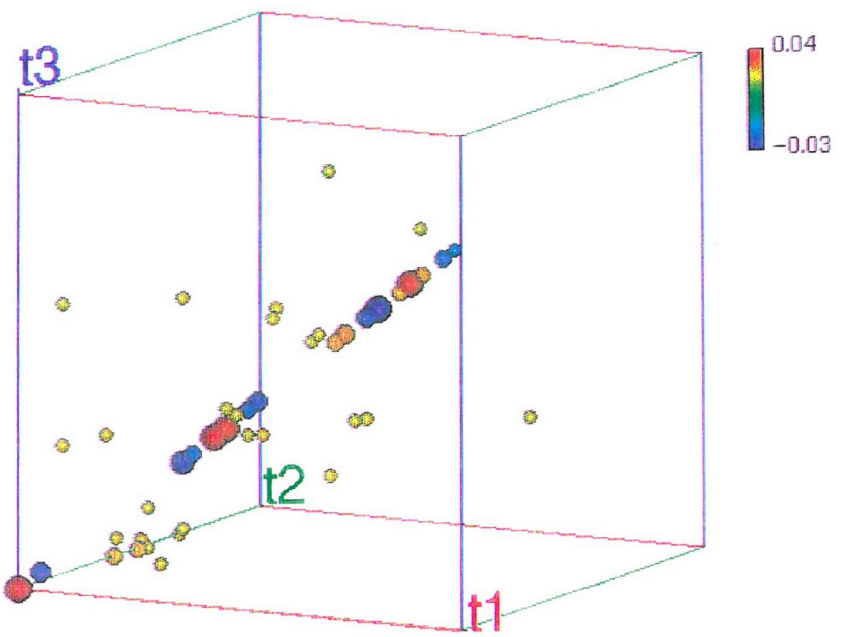

Figure 6.36: Cubic impulse response of the symmetric magnetic beam

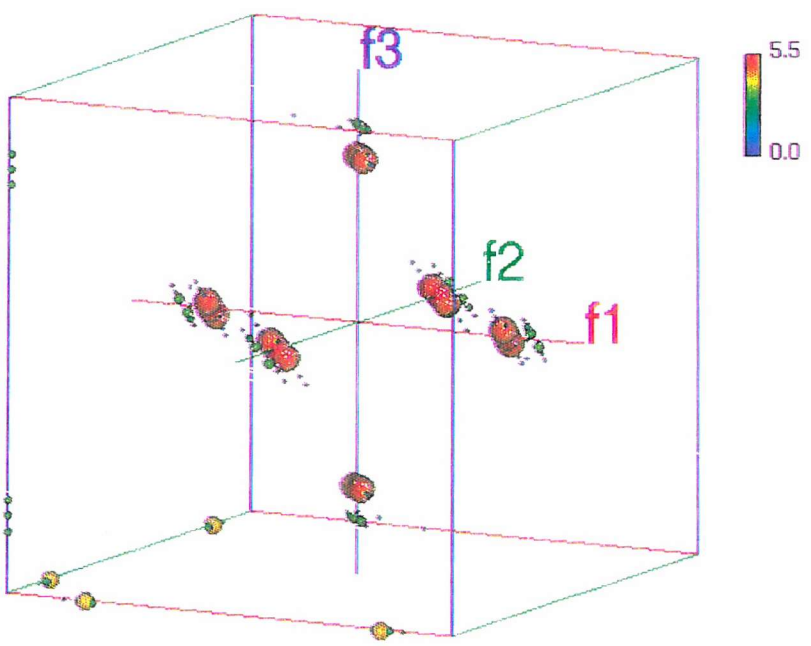

Figure 6.37: Cubic transfer function of the symmetric magnetic beam

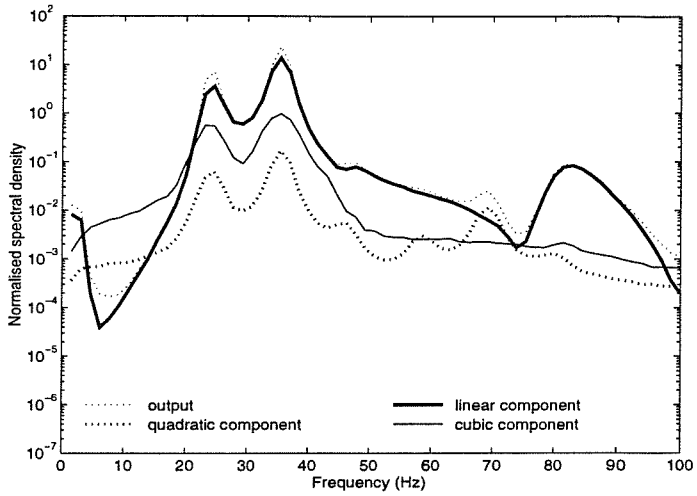

Figure 6.38: Power spectra of the linear, quadratic and cubic components of the Volterra model and the power spectrum of the output of the symmetric magnetic beam.

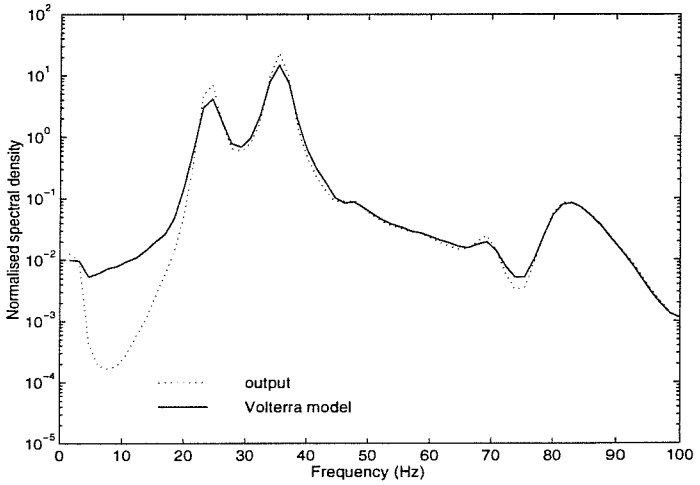

Figure 6.39: Power spectra of the Volterra model and output of the symmetric magnetic beam.

6.3 The nonlinear rattling beam

In the second experiment, a beam was attached to a shaker, and a 'fault' was introduced into the system by loosening the attaching nut. Higher order spectra were then used, as a condition monitoring tool, first to try and detect the fault, and second to determine something about the nonlinearity produced by it.

A 10 cm long aluminium beam was attached at one end to a shaker, and the other end was free to vibrate, figure 6.40. The beam was excited by passing a Gaussian signal through a power amplifier and into the shaker. The signal from the accelerometer and the input signal to the shaker were sampled at 5 kHz having passed through an anti-alias filter with a cut-off frequency of 2.2 kHz. Approximately 262144 samples of each channel were synchronously sampled into a P.C. and stored ready for analysis. It was possible to introduce nonlinearity into the system by loosening the nut that connected the beam to the shaker. Recordings of the data were made both with the nut tightened and loosened.

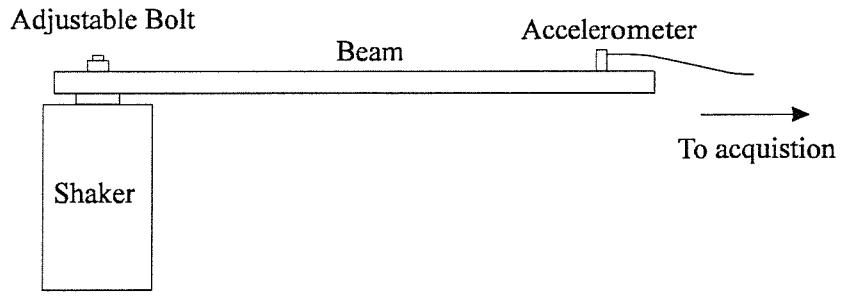

Figure 6.40: Experimental set up for nonlinear beam experiment

The only control over the nonlinearity in this experiment is the degree to which the nut is loosened. The aim is therefore to try and determine what the nature of the nonlinearity is and around what frequencies it is most dominant. As with the magnetic beam, first it will be assumed that only an output measurement is available, and auto higher order spectra used to find out as much as possible about the beam. Cross higher order spectra will then be used to try and obtain more information about the system.

6.3.1 Auto higher order spectral methods

With the nut tight, the power spectrum of the beam, figure 6.41, shows a number of resonances. There is no significant structure anywhere in either the bicoherence, figures 6.42 and 6.43, or the tricoherence, figure 6.44. This indicates that the system is linear up to third order. Figure 6.45 shows the power spectrum of the beam with the nut loosened. Loosening the nut, produces significant structure in both the bicoherence and tricoherence. Figures 6.46 and 6.47 show a large number of frequency interactions, indicating a large amount of quadratic phase coupling between the peaks in the spectrum and similarly, figures 6.48 and 6.49, show significant cubic phase coupling. Using only auto higher order spectra, this experiment has shown that signals measured from mechanical systems with loose fastenings can have much more bispectral and trispectral content than correctly fastened systems.

6.3.2 Cross higher order spectral methods

With the nut in the tight position, the coherence function, figure 6.50 is approximately unity, apart from around regions of resonance. An estimate of the first order transfer function is shown in figure 6.52. Fourier transforming this gives the linear impulse response, figure 6.51. By convolving this with the original input data to the shaker, x(t), it is possible to obtain an estimate of just the linear component of the system. The power spectrum of this linear component is shown by the solid line in figure 6.53 together with the output power spectrum of the system, shown by the dotted line. The linear component can account for nearly all the power across all frequencies, indicating that the beam with the nut tight can be modelled as a linear system.

If the nut is loosened through 90 degrees the coherence function falls dramatically, figure 6.54, indicating that there is no longer a purely linear transmission path between the input and output. As it is not clear what the nature of the nonlinearity is, it is not obvious that it is possible to model this type of nonlinearity using the Volterra series; however, a cubic Volterra model is fitted to the system. The standard procedure for estimating the Volterra

model was carried out and after the second iteration the linear, quadratic, and cubic transfer functions are as shown in figures 6.56, 6.58 and 6.60 respectively. Their time domain equivalents are also shown in figures 6.55, 6.57 and 6.59. By convolving these with the input it was possible to obtain a measure of the linear, quadratic and cubic components of the system. The linear component, shown by the thick solid line in figure 6.61, accounts for a large proportion of the power of the signal but by no means all. In the regions of the resonances and at the high frequencies there are discrepancies which may be due to the nonlinearities in the system. As a quadratic nonlinearity results in a doubling of frequency and a cubic nonlinearity may result in a trebling of frequency, it is not unexpected that that regions of nonlinearity appear at higher frequencies. The power spectrum of the quadratic component is shown by the dashed line in figure 6.61. It is considerably smaller than the linear component and accounts for only a small proportion of the total signal strength. Intuitively it is more likely that the beam will have a symmetric nonlinearity rather than a skewed nonlinearity as the driving force is symmetric. The cubic component is shown by the thin solid line in figure 6.61. It accounts for much of the signal that was not accounted for by the linear or quadratic components. This mainly occurs at the higher frequencies for the reasons discussed previously. It would be expected for the cubic component to account for the nonlinearities in the regions of resonance but, due to the poor resolution of the cubic Volterra kernel, this does not happen.

Figure 6.62 shows the spectrum of the sum of the linear, quadratic and cubic components together with the spectrum of the output of the system. The difference between these two signals is that part of the signal which cannot be accounted for by our cubic Volterra model. These will be due to a combination of a number of reasons such as: nonlinearities of order higher than three, measurement noise, the incorrect assumption of a Gaussian input, or difficulties caused by the signal processing algorithms such as the poor resolution in the cubic term. Overestimates typically occur due to the incorrect assumption that the input is Gaussian. If the input is non-Gaussian, there will be cross terms present in the Volterra analysis which will lead to an overestimate. Also, it was assumed that this rattling nonlinearity could be modelled by the Volterra series and so some of the discrepancies could be due to the unsuitability of a Volterra model for this application. However, over most of the frequency range the model fits the data reasonably well.

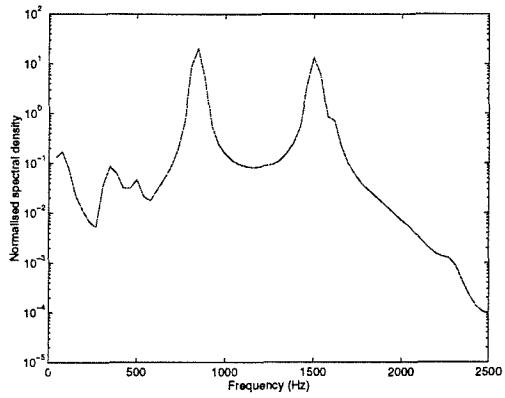

Figure 6.41: Normalised power spectrum of the beam with nut tight

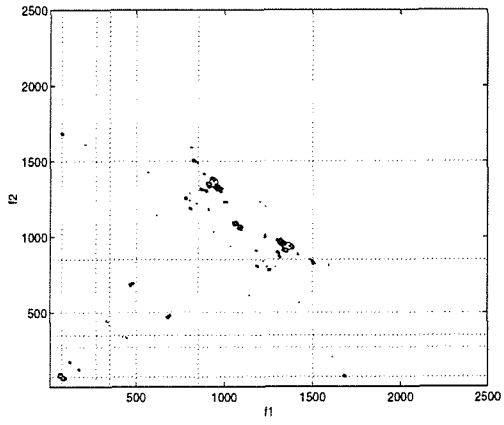

Figure 6.42: Contour plot of the bicoherence of the beam with nut tight

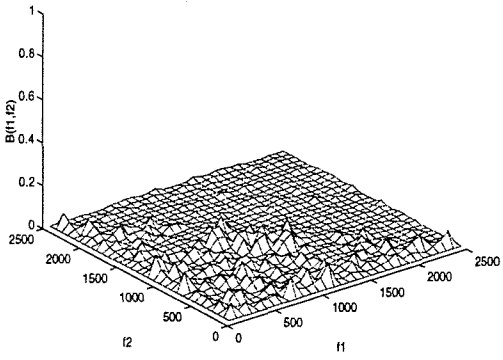

Figure 6.43: Mesh plot of the bicoherence of the beam with nut tight

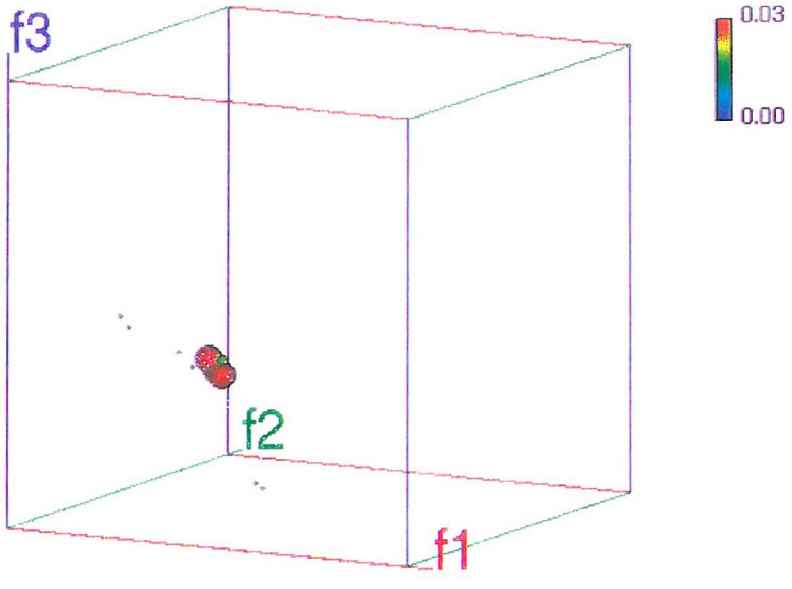

Figure 6.44: Tricoherence of the beam with nut tight

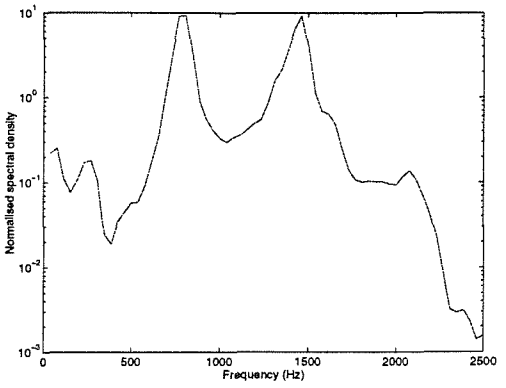

Figure 6.45: Normalised power spectrum of the beam with nut loose

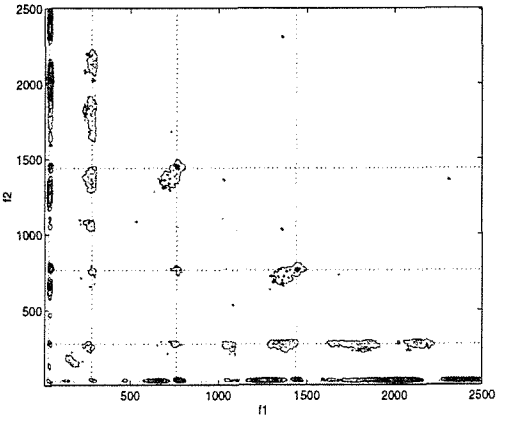

Figure 6.46: Contour plot of the bicoherence of the beam with nut loose

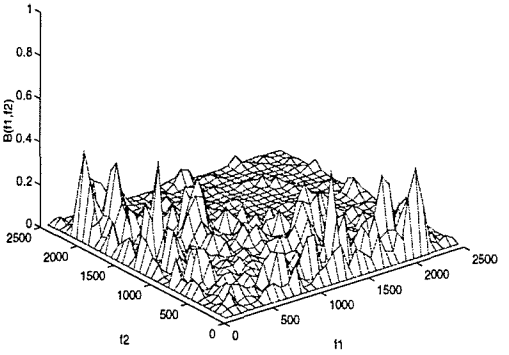

Figure 6.47: Mesh plot of the bicoherence of the beam with nut loose

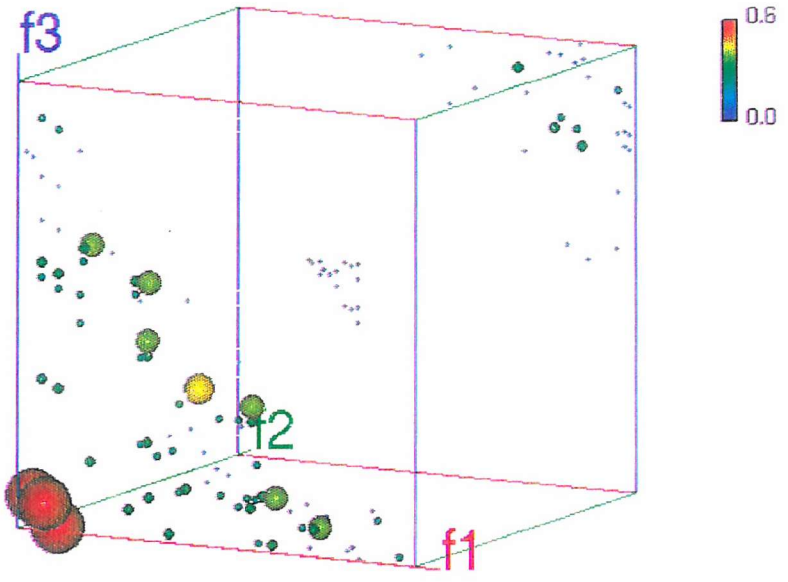

Figure 6.48: Tricoherence of the beam with nut loose

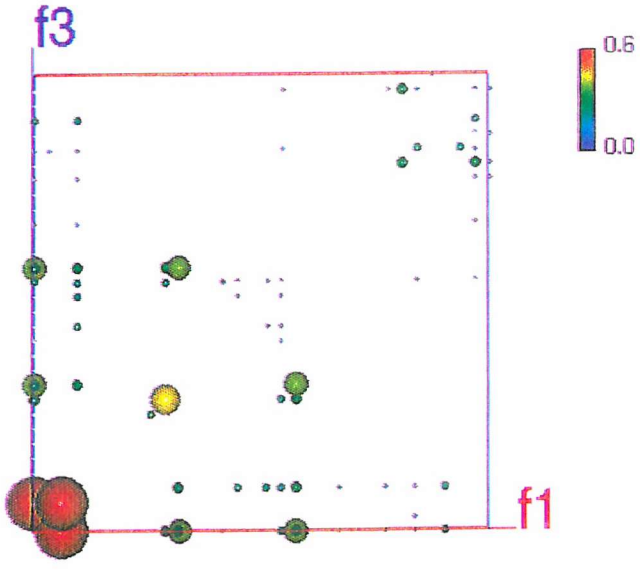

Figure 6.49: Tricoherence of the beam with nut loose (end view)

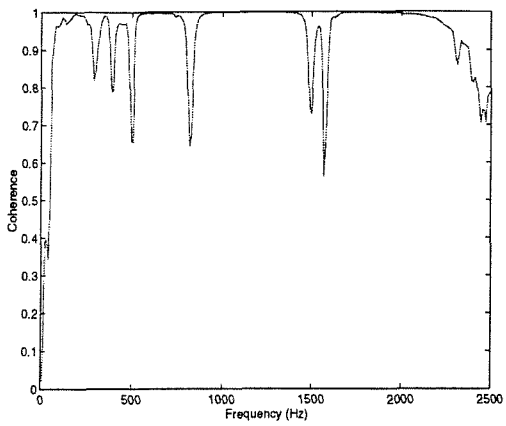

Figure 6.50: Coherence function of the beam with nut tight

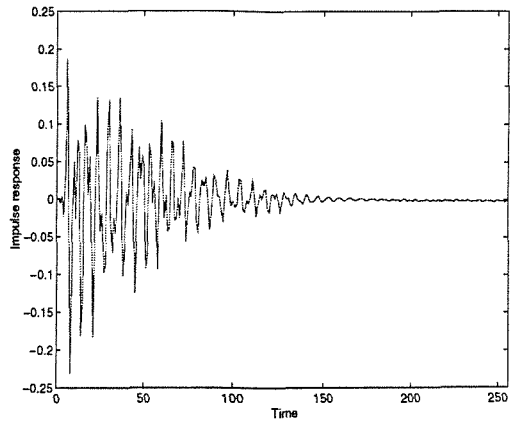

Figure 6.51: Linear impulse response of the beam with nut tight

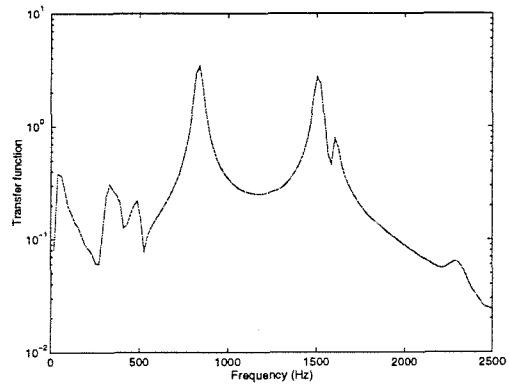

Figure 6.52: Transfer function of the beam with nut tight

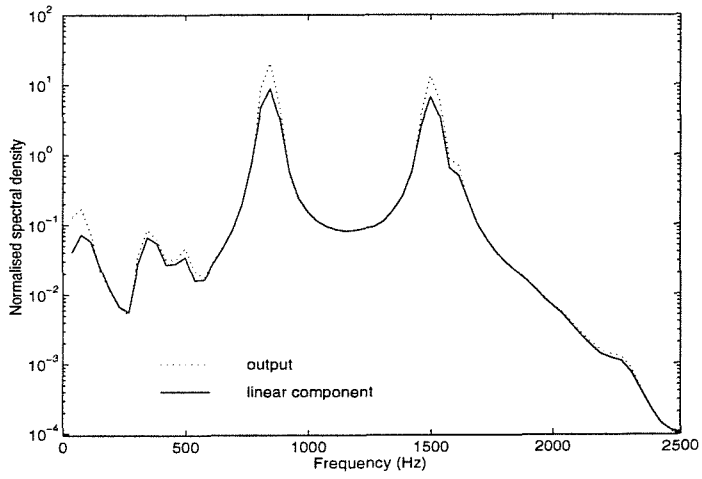

Figure 6.53: Power spectra of the linear component of the beam with nut tight and power spectrum of the output of the beam

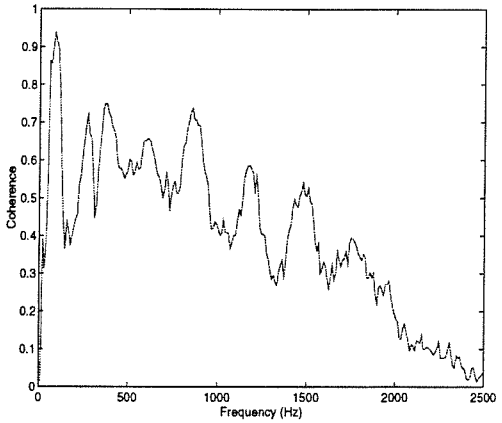

Figure 6.54: Coherence function of the beam with nut loose

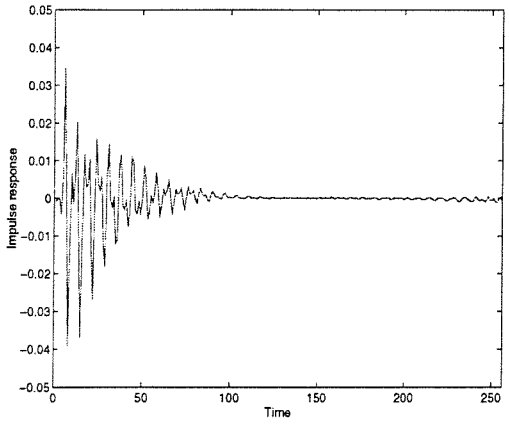

Figure 6.55: Linear impulse response of the beam with nut loose

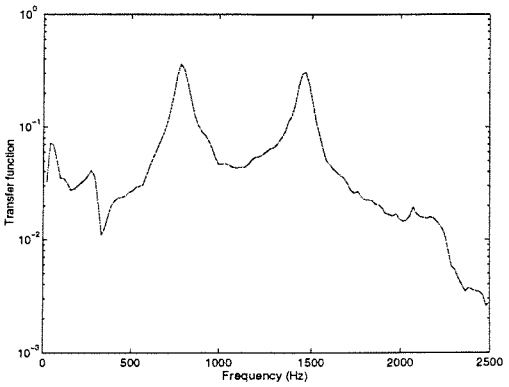

Figure 6.56: Linear transfer function of the beam with nut loose

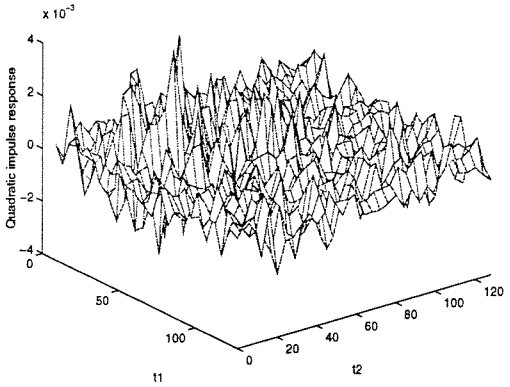

Figure 6.57: Quadratic impulse response of the beam with nut loose

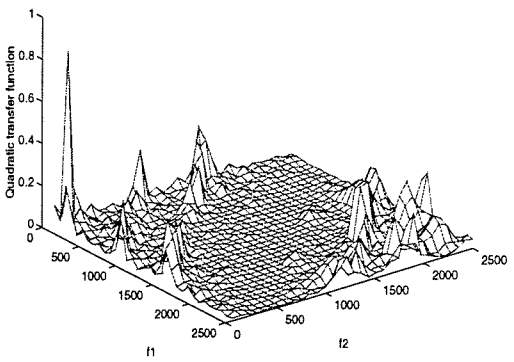

Figure 6.58: Quadratic transfer function of the beam with nut loose

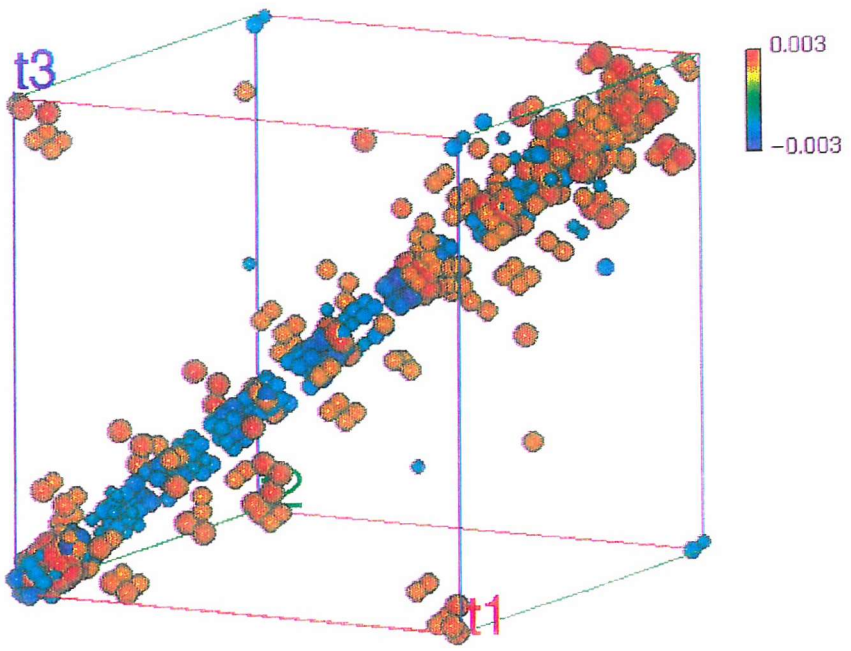

Figure 6.59: Cubic impulse response of the beam with nut loose

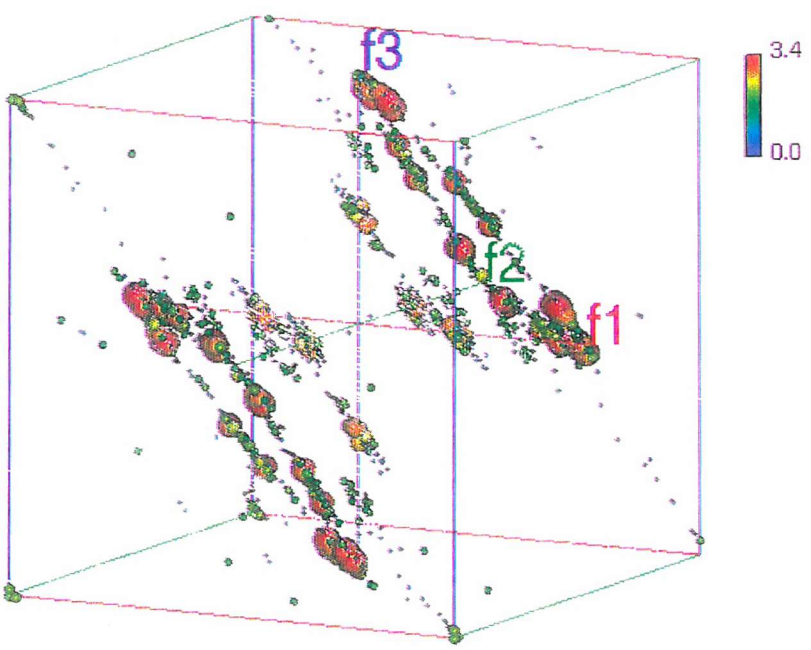

Figure 6.60: Cubic transfer function of the beam with nut loose

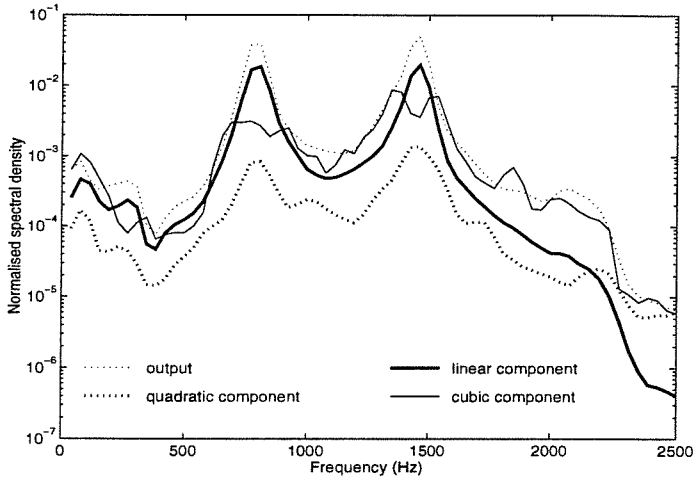

Figure 6.61: Power spectra of the linear, quadratic and cubic components of the Volterra model and the power spectrum of the output of the beam with nut loose

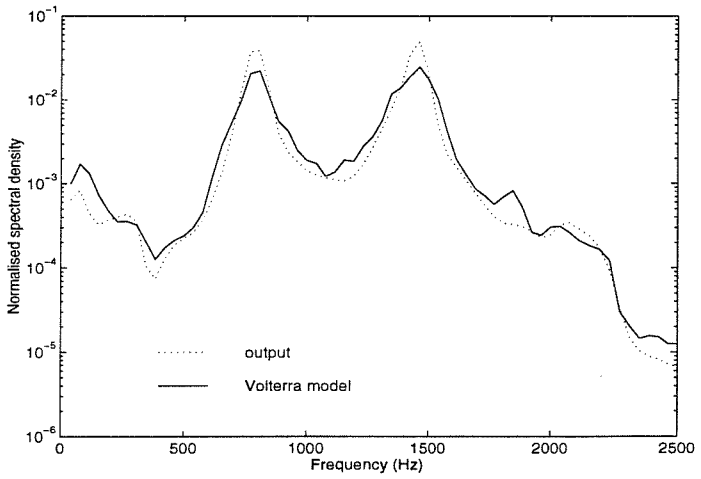

Figure 6.62: Power spectra of the Volterra model and output of the beam with nut loose

6.4 Conclusions

In this chapter two experiments have been successfully analysed with the use of higher order spectra. Although cross higher order spectra and the Volterra series can give a much more complete picture of the system, the importance of the results obtained using auto higher order spectra must not be ignored.

In many applications, particularly in condition monitoring, only a single measurement will be available. In these cases it has been shown that enough information can be obtained to tell an engineer:

1) If there is any nonlinearity in the system. This is often a good indication that there may be a fault in the system.

2) The dominant nature of the nonlinearity and the frequencies at which it occurs. This information is often helpful in determining where, in the system, the fault is.

When the input to the system is available, the cubic Volterra model has been shown to work well with the nonlinear systems studied. Volterra kernels up to third order have been successfully identified for all the systems studied using the methods proposed in chapter four. Using these kernels it is possible for an engineer to take the information which can be obtained from auto higher order spectra a stage further and to predict the response of the system to any input.

Chapter 7

Concluding Remarks

This thesis has discussed some of the issues associated with the use of higher order spectra and the application of such techniques to the detection and classification of nonlinearity in mechanical systems. Techniques have been developed which use both the bispectrum and trispectrum as complimentary tools in the analysis of nonlinear systems. Chapter two introduced higher order spectra and discussed some of the basic theory and difficulties associated with the bispectrum and trispectrum, the key features of which can be summarised as follows:

- A signal can be completely described its joint probability density functions. Higher order spectra are defined as multiple Fourier transforms of joint moment and cumulant functions which can in turn be defined in terms of the joint probability density function. Hence, higher order spectra can be used to describe the properties of a signal.

- The bispectrum and trispectrum can be thought of as decompositions of skewness and kurtosis over frequency respectively, in the same way as the power spectrum gives a decomposition of power over frequency.

- The magnitude of the bispectrum is a function of two frequency variables and so can be easily displayed using three dimensions. The magnitude of the trispectrum, however, is a function of three frequencies and so requires four dimensions to display it. A new method has been proposed where for every frequency triplet, a sphere is drawn; the size and colour of the sphere represent its magnitude.

- Higher order spectra can be estimated via either the time domain (indirect method) or frequency domain (direct method). As with power spectral estimation the length of

window used in the calculation crucially affects the variance and bias of the estimate. Considerably longer data lengths are needed to ensure statistical convergence of higher order spectra than with ordinary power spectral estimation.

- Just as the discrete power spectrum has symmetry about its folding frequency the bispectrum and trispectrum have many regions of symmetry. It is only necessary to evaluate the bispectrum and trispectrum in the principle domain.

- The principle domain can be further subdivided into the inner and outer region. It was shown how the position of components of higher order spectra in these regions could form the basis of tests for aliasing and stationarity.

An important consideration is that estimates of higher order spectra have a variance which is dependent upon the signal's spectrum. In chapter three, the possible methods that can be used to normalise the bispectrum and trispectrum to remove these power spectral effects were explained and a number of examples given to illustrate the processes. The main normalisation methods can be outlined as follows:

- The skewness and kurtosis functions are normalised bispectra and trispectra respectively. They are mainly used to make decisions, based on statistical tests, about the symmetry, aliasing, and linearity of a signal.

- The bicoherence and tricoherence are normalised bispectra and trispectra receptively and are predominantly used to measure quadratic and cubic phase coupling. Their main advantage over the skewness and kurtosis function is that they give results which are bounded between zero and one.

- Under certain circumstances, notably systems with sharp resonant peaks, the above normalisation methods have been shown to fail due to bias problems. An alternative technique has been developed to alleviate this problem based on pre-whitening the signal. The pre-whitening can be carried out using a very much larger window length, and hence greater resolution, than the estimation of the higher order spectra in which

the window length is constrained by computational requirements. It has been successfully demonstrated both with a narrow band amplitude modulated process and a simulation using the Duffing oscillator.

The bispectrum and trispectrum can be used to detect non-Gaussianity in a signal. If a Gaussian signal is operated on by a nonlinear system then the resulting signal will be non-Gaussian. By studying this non-Gaussian signal it is possible to obtain information about possible nonlinearity in the system. However, to fully characterise the system it is necessary to have some knowledge of both the input and the output. In chapter four, the Volterra series was introduced and a detailed analysis of how it could be used to model a nonlinear system, using cross higher order spectra, was given. The key points of this analysis can be summarised as:

- Before analysing a nonlinear system with the Volterra series, a more simple model, based on the polynomial input to linear filters was used. With this model and using both higher order spectra and residual spectra, it is possible to calculate expressions for the linear, quadratic, and cubic components of the system. However, although this model is more simple to analyse, it is equivalent to just calculating the main diagonal of the full Volterra kernels, and for many systems, is an over simplification.

- A quadratic Volterra model has been developed to analyse nonlinear systems with skewed components. However, it was found that, as with auto higher order spectra, estimates of cross higher order spectra have variances which depend on the signals spectrum. A technique has been developed to calculate the Volterra kernels, which minimises the variance of the bispectrum and so minimises the effect of the second order properties. This is based on successively subtracting of lower order terms.

- The quadratic Volterra model can be extended to a cubic Volterra model which is capable of analysing both skewed and symmetric type nonlinearities. Whereas, the quadratic model produced a simple closed form solution, the cubic model is more complex and an iterative procedure is adopted.

Chapter four was based on the assumption that the input to the system is Gaussian. In chapter five the possible approaches that can be used when this assumption is invalid were discussed, together with the following pitfalls that can occur whilst using the Volterra series.

- The Volterra series has been likened to a Taylor series with memory and so as some functions have a diverging Taylor series then some systems will also have a diverging Volterra series. To converge, the system must posses finite memory, that is, the response to any input must die away with time. Also, systems with multiple equilibria will not have convergent Volterra series except if modelled locally around one equilibrium point.

- If the input to the system cannot be assumed to be Gaussian then it is still possible to find solutions to the quadratic and cubic Volterra models, however, the complexity of the algorithms involved is vastly increased. The quadratic Volterra model requires knowledge of up to the fourth order statistics and the cubic model assumes knowledge of up to the sixth order.

- Having calculated the Volterra kernels of a system it is possible to find the response of that system to any input. This has been demonstrated by firstly calculating the Volterra kernels of a Duffing oscillator using a Gaussian input and then using them to predict the response of the system to a sine wave.

- The causality of the Volterra kernels is checked after calculation. Although, the theory developed does not necessarily force the Volterra kernels to be causal, it has been found that for the practical systems analysed, all the kernels estimated, were causal.

- The frequency domain technique developed in chapter four to estimate the Volterra kernels, although not optimal, has been shown to be a great improvement on traditional methods, when compared to the optimum least squares time domain technique.

Finally, in chapter six two mechanical systems were successfully analysed with the use of higher order spectra. It was shown how higher order spectra could provide the following useful information to an engineer:

- If only an output signal is available, as with many condition monitoring applications, auto higher order spectra are able to detect for nonlinearity in a process, so long as the input can be assumed to be Gaussian. It is possible to determine both the nature and the dominant frequencies of the nonlinearity. This information can help diagnose if there is a fault in a system and, if so, the nature and likely position of the fault.

- If both input and output signals are available, it is possible to use cross higher order spectra and the Volterra series to model the system.

There are a number of possible extensions to this work some of which are listed below.

- It would be of interest, for real data, to compare the approach used in this thesis with other techniques e.g. a model based parametric approach.

- Further work could be carried out on the time domain methods discussed in chapter five in order to produce leaner algorithms that are capable of coping with greater resolution and higher orders. These could be based around a sparse Volterra model as suggested in [63].

- Having estimated the Volterra kernels of the beams in chapter six it would be of interest to excite them with a sine wave and compare the actual output measured from the beam with the output of the Volterra models. In might be expected that these results would be similar for the case of the magnetic beam, where the nonlinearity is analytic, but not for the rattling beam.

- Further work needs to be done to fully address the topic of the causality of the estimated kernels. One possible approach would be to generalise the concepts of spectral factorisation to higher order spectral factorisation.

Chapter 8

Appendix A

Properties of cumulant functions

The following appendix gives a brief explanation of cumulant functions, how they are derived from the joint probability density function, and some of their important properties. A more detailed discussion can be found in [26]. The appendix is written in terms of auto cumulants although it is all equally applicable to cross cumulants.

Let $\mathbf{v} = (v_1,v_2,...v_k)^T$ and $\mathbf{x} = (x_1,x_2,...x_k)^T$ where $(x_1,x_2,...,x_k)$ denotes a collection of random variables. The k^{th} order cumulant of these random variables is defined as the coefficient of $(v_1,v_2,...v_k)$ in the Taylor series expansion of the logarithm of the k^{th} order joint characteristic function. The k^{th} order characteristic function, $\mathbf{K}(\mathbf{v})$, is defined as the k^{th} order Fourier transform of the k^{th} order joint probability density function.

$$K(\mathbf{v}) = \ln E[\exp(j\mathbf{v}'\,\mathbf{x})] \qquad \text{A.1}$$

The second, third, and fourth order cumulants of a real, zero mean, random variable are thus defined as,

$$\mathrm{cum}(x_1,x_2) = E[x_1x_2] \qquad \text{A.2}$$

$$\mathrm{cum}(x_1,x_2,x_3) = E[x_1x_2x_3] \qquad \text{A.3}$$

$$\begin{aligned} cum(x_1,x_2,x_3,x_4) = E[x_1x_2x_3x_4] &- E[x_1x_2]E[x_3x_4] \\ &- E[x_1x_3]E[x_2x_4] \\ &- E[x_1x_4]E[x_2x_3] \end{aligned} \quad \text{A.4}$$

Let x(t) be a zero mean, strictly stationary, random process. The k^{th} order cumulant of this process, denoted $C_{X...X}(\tau_1,\tau_2,...,\tau_k)$, is defined as the joint k^{th} order cumulant of the random variables x(t), $x(t+\tau_1)$, $x(t+\tau_2)$, ..., $x(t+\tau_k)$. Hence, the second, third, and fourth order cumulants of x(t), are:

$$C_{xx}(\tau) = E[x(t)x(t+\tau)] \quad \text{A.5}$$

$$C_{xxx}(\tau_1,\tau_2) = E[x(t)x(t+\tau_1)x(t+\tau_2)] \quad \text{A.6}$$

$$\begin{aligned} C_{xxxx}(\tau_1,\tau_2,\tau_3) = E[x(t)x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)] &- C_{xx}(\tau_1)C_{xx}(\tau_2-\tau_3) \\ &- C_{xx}(\tau_2)C_{xx}(\tau_3-\tau_1) \\ &- C_{xx}(\tau_3)C_{xx}(\tau_1-\tau_2) \end{aligned} \quad \text{A.7}$$

For a zero mean stationary random process and for k = 3,4 the k^{th} order cumulant of x(t) can also be defined as,

$$C_{X...X}(\tau_1,\tau_2,\ldots,\tau_{k-1}) = E[x(\tau_1)\ldots x(\tau_{k-1})] - E[g(\tau_1)\ldots g(\tau_{k-1})] \quad \text{A.8}$$

where g(t) is a Gaussian random process with the same second order statistics as x(t). Cumulants therefore, not only display the amount of higher order correlation, but also provide a measure of the distance of the random process from Gaussianity.

The following are important properties of cumulants, some of which are used in this work:

- Cumulants are symmetric in their arguments, i.e.

$$cum(x_1,\ldots,x_k) = cum(x_{i_1},\ldots,x_{i_k}) \quad \text{A.9}$$

where $(i_1,...,i_k)$ is a permutation of (1,...k). Hence for the third order cumulants, the following symmetries exist:

$$\begin{aligned} C_{xxx}(\tau_1,\tau_2) &= C_{xxx}(\tau_2,\tau_1) \\ &= C_{xxx}(-\tau_2,\tau_1-\tau_2) \\ &= C_{xxx}(\tau_2-\tau_1,-\tau_1) \\ &= C_{xxx}(\tau_1-\tau_2,-\tau_2) \\ &= C_{xxx}(-\tau_1,\tau_2-\tau_1) \end{aligned} \qquad \text{A.10}$$

- If α is a constant then,

$$\text{cum}(\alpha + x_1,\ldots,x_k) = \text{cum}(x_1,\ldots,x_k) \qquad \text{A.11}$$

- Cumulants are additive in their arguments:

$$\text{cum}(x_0 + y_0, z_1,\ldots,z_k) = \text{cum}(x_0, z_1,\ldots,z_k) + \text{cum}(y_0, z_1,\ldots z_k) \qquad \text{A.12}$$

- If the random variables $\{x_i\}$ are independent of the random variables $\{y_i\}$, then,

$$\text{cum}(x_1 + y_1,\ldots,x_k + y_k) = \text{cum}(x_1,\ldots,x_k) + \text{cum}(y_1,\ldots y_k) \qquad \text{A.13}$$

- If λ_i are constants then,

$$\text{cum}(\lambda_1 x_1,\ldots,\lambda_k x_k) = \left(\prod_{i=1}^{k}\lambda_k\right)\text{cum}(x_1,\ldots,x_k) \qquad \text{A.14}$$

Appendix B

Second order analysis of stationary random signals

In this appendix, some details of second order analysis are discussed. Three topics are briefly mentioned, the stationarity of a signal, the auto correlation function, and the power spectrum. Further explanation can be found in Priestley [64].

In general, the properties of a stochastic process are time dependent but to simplify matters it is often assumed that a sort of 'steady state' has been reached in the sense that the statistical properties are unchanged under a shift in time and the following two properties hold:

$$P(x,t) = P(x) \qquad \text{B.1}$$

$P(x_1,t_1;x_2,t_2)$ is a function of t_2-t_1 and not both t_1 and t_2 B.2

where P(x,t) denotes the probability density function at time, t.

If the above conditions hold, the process is said to be weakly stationary. If similar conditions hold for all the higher orders then the process is said to be completely stationary.

A widely used statistic for a random signal, x(t), is the autocorrelation function, which, for a zero mean signal, is defined as:

$$R_{xx}(t_1,t_2) = E[x(t_1)x(t_2)] \qquad \text{B.3}$$

If x(t) is stationary, its statistical properties are unchanged under a time shift so that,

$$R_{XX}(t_1, t_2) = R_{XX}(t_2 - t_1) \qquad \text{B.4}$$

Equation B.4 can then be written in its more common form in terms of a lag, τ, as:

$$R_{XX}(\tau) = E[x(t)x(t+\tau)] \qquad \text{B.5}$$

By assuming that the process x(t) is ergodic with respect to its mean and auto correlation function it is possible to replace the ensemble average in equation B.5 with a time average.

$$R_{XX}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t+\tau)dt \qquad \text{B.6}$$

Fourier transforming equation B.3 leads to an expression for the power spectrum,

$$S_{XX}(f_1, f_2) = \lim_{T\to\infty} \frac{E[X_T(f_1)X_T(f_2)]}{T} \qquad \text{B.7}$$

where $X_T(f)$ is the Fourier transform of x(t) from -T/2 to T/2. This is written in this form as it is not strictly possible to Fourier transform x(t). This is because x(t) is stationary and extends from $-\infty$ to ∞ and as such has unbounded energy, i.e. it is not a function in L^2. However, this is often of no practical importance and it is common to write equation B.7 as:

$$S_{XX}(f_1, f_2) = E[X(f_1)X(f_2)] \qquad \text{B.8}$$

For a stationary signal it can be shown, as in Priestley [64], that equation B.8 is equal to zero except along $f_1 = -f_2$. Hence, for a stationary signal the power spectrum is defined as,

$$S_{XX}(f) = E[X(f)X^*(f)] \qquad \text{B.9}$$

Appendix C

Visualisation of the trispectrum

The methods used to display the trispectrum were introduced in chapter two. In this appendix some examples of simple, analytic signals will be used to help familiarise the reader with this visualisation technique.

The signals used are the three dimensional versions of an exponential, a sine wave, and an exponentially decaying sine wave. They are defined as:

$$x_1 = e^{-0.05(i+j+k)} \qquad \text{C.1}$$

$$x_2 = \sin(0.5(i+j+k)) \qquad \text{C.2}$$

$$x_3 = \sin(0.5(i+j+k)).e^{-0.05(i+j+k)} \qquad \text{C.3}$$

where i, j, k are the three axes scaled from 0 to 31.

Figures C1, C2, and C3 show the three dimensional representations of x_1, x_2, and x_3 respectively. The method used was stated in detail in chapter two but a few key points will be noted here: For every (i,j,k) a sphere is plotted and the size and colour of the sphere represent its magnitude. Spheres of very small magnitudes are not plotted. Note, a sphere of small magnitude means one that is approximately zero, not a large negative value. It is always important to look at the key along side the graph as the colour scales are not the same between different plots.

The exponential signal, figure C1, can be seen to decay away from the origin, (0,0,0), in all three dimensions. Points whose magnitudes are close to zero, which would be displayed in dark blue, have zero radius and so do not appear in the plot. The sine wave, figure C2, can be seen to oscillate from -1 to 1, going away from the origin. Again, points of magnitude that are approximately equal to zero, in this case coloured green, are omitted. The decaying exponential, figure C3, oscillates and dies away from the origin. This particular signal is very typical of a three dimensional impulse response function, examples of which were seen in the experiments in chapter 6.

The software used for this visualisation technique was Application Visualisation System (AVS). AVS is a system which allows users to visualise their data by constructing applications from a series of software components called modules. Modules are assembled into networks where each module performs a specific task. For completeness, a diagram of the network used for the trispectral visualisation, figure C4, is included. For more information on AVS see [22].

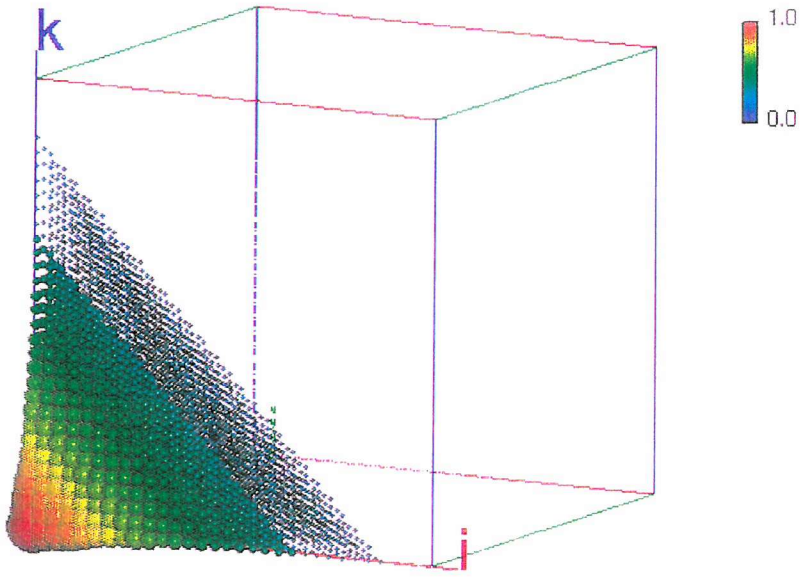

Figure C1: Three dimensional exponential

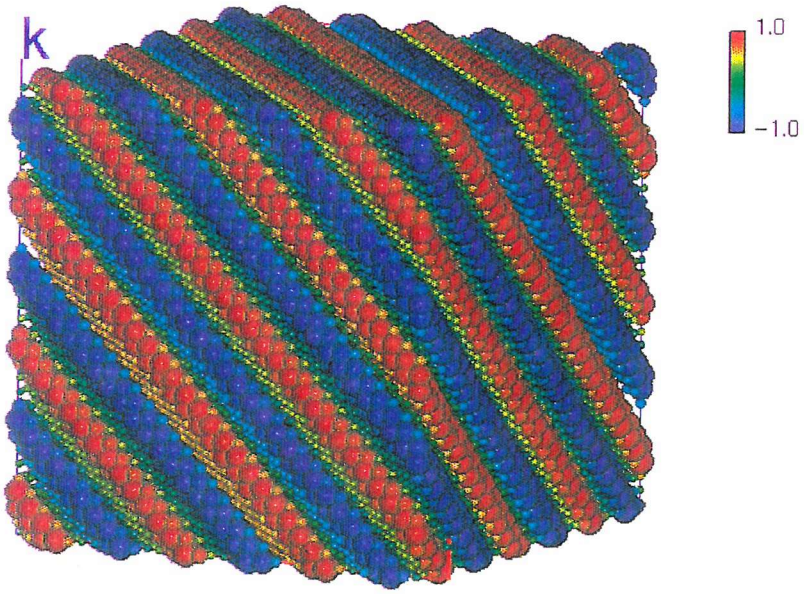

Figure C2: Three dimensional sine wave

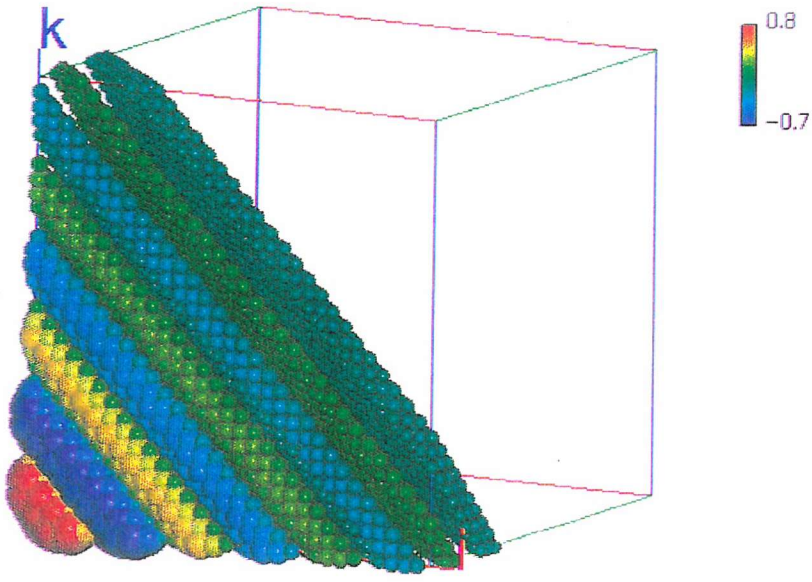

Figure C3: Three dimensional decaying sine wave

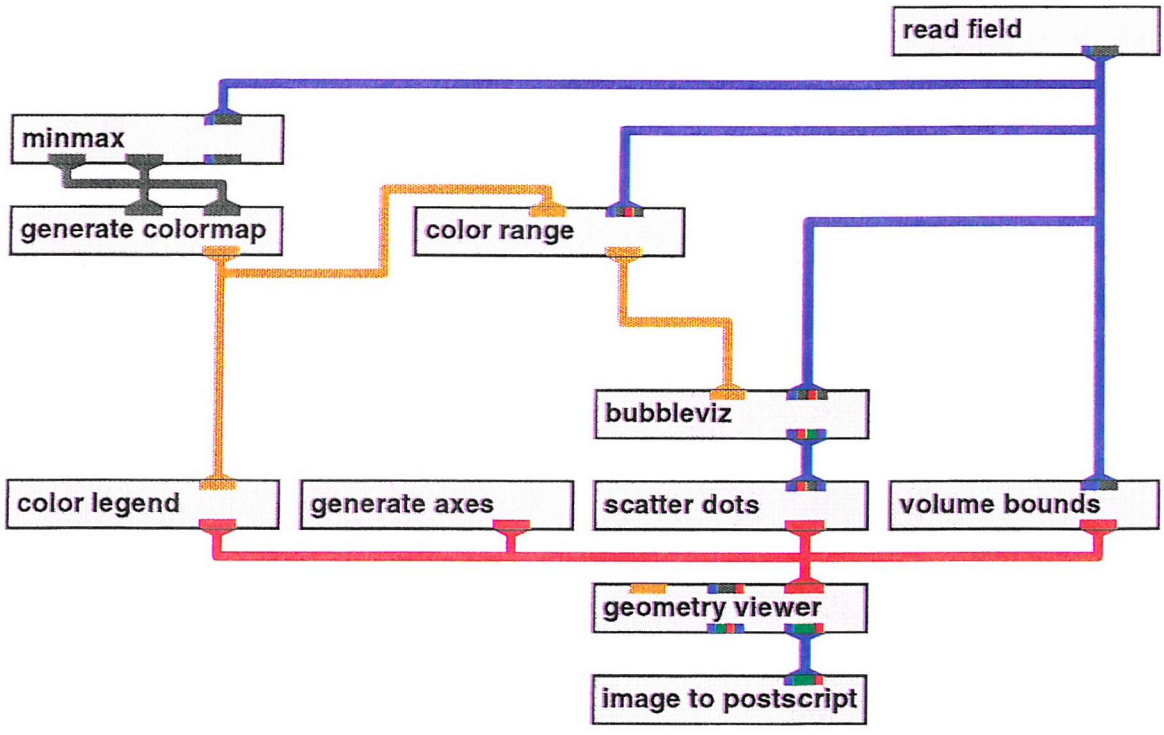

Figure C4: AVS network used to display the trispectrum

Appendix D

Average of the Product of Gaussian Variables

If x is a Gaussian variable with a zero mean and a standard deviation of one then the probability distribution is,

$$P(x) = \frac{1}{\sqrt{2p}} e^{-\frac{1}{2}x^2} \qquad \text{D.1}$$

The expected value of a product of such Gaussian variables, obeys the following law,

$$E[x_1 x_2 \cdots x_{2N}] = \Sigma \prod E[x_i x_j] \qquad \text{D.2}$$

$$E[x_1 x_2 \cdots x_{2N+1}] = 0 \qquad \text{D.3}$$

where $\Sigma \Pi$ means the sum of all completely distinct ways of arranging x_1, x_2, ... x_{2N} into pairs. The number of ways of doing this is $(2N)! / N!2^N$.

For example for N=2,

$$\frac{(2N)!}{(N)!2^N} = \frac{4!}{2!2^2} = 3 \qquad \text{D.4}$$

and,

$$E[x_1 x_2 x_3 x_4] = E[x_1 x_2] E[x_3 x_4] + E[x_1 x_3] E[x_2 x_4] + E[x_1 x_4] E[x_2 x_3] \qquad \text{D.5}$$

To prove this result, as in Schetzen [70], some results will be used based on the characteristic function of a Gaussian variable. The characteristic function for a Gaussian variable, x, is defined as:

$$M_x(a) = E[e^{jax}] = \int_{-\infty}^{\infty} e^{jax} P(x)\, dx \qquad \text{D.6}$$

where P(x) is the probability density function of x. This can be extended to the characteristic function of the joint probability density function of N random variables, $P(x_1, x_2, \ldots x_N)$ as,

$$M_x(a_1, a_2, \cdots, a_N) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} P(x_1, x_2, \cdots, x_N)\, e^{j(a_1x_1,\, a_2x_2,\, \cdots,\, a_Nx_N)}\, dx_1 dx_2 \cdots dx_N$$

D.7

Equation D.7 can be expanded to give,

$$M_x(a_1, a_2, \cdots, a_N) = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_N=0}^{\infty} c_{k_1k_2\cdots k_N}\, a_1^{k_1} a_2^{k_2} \cdots a_N^{k_N} \qquad \text{D.8}$$

where

$$c_{k_1k_2\cdots k_N} = \frac{1}{k_1!k_2!\cdots k_n!}\, \frac{\partial^{k_1}}{\partial a_1^{k_1}} \frac{\partial^{k_2}}{\partial a_2^{k_2}} \cdots \frac{\partial^{k_N}}{\partial a_N^{k_N}} M_x(a_1, a_2, \cdots, a_N)\Big|_{a_1=a_2=\cdots=a_N=0}$$

D.9

Differentiating equation D.7 gives

$$\frac{\partial^{k_1}}{\partial a_1^{k_1}} \frac{\partial^{k_2}}{\partial a_2^{k_2}} \cdots \frac{\partial^{k_N}}{\partial a_N^{k_N}} M_x(a_1, a_2, \cdots, a_N)\Big|_{a_1=a_2=\cdots=a_N=0} = E[x_1^{k_1} x_2^{k_2} \cdots x_N^{k_N}]\, j^{k_1} j^{k_2} \cdots j^{k_N}$$

D.10

Substituting equation D.10 and D.9 into D.8 produces,

$$M_x(a_1,a_2,\cdots,a_N)=\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_N=0}^{\infty}E[x_1^{k_1}x_2^{k_2}\cdots x_N^{k_N}]\frac{(ja_1)^{k_1}(ja_2)^{k_2}\cdots(ja_N)^{k_N}}{k_1!k_2!\cdots k_N!} \qquad \text{D.11}$$

The only term in equation D.11 that will contain the term $E[x_1x_2\cdots x_N]$ is if $k_1 = k_2 = \ldots = k_N = 1$. So taking only this term gives,

$$\begin{aligned} M_x(a_1,a_2,\cdots,a_N) &= E[x_1x_2\cdots x_N](ja_1)(ja_2)\cdots(ja_N) \\ &= E[x_1x_2\cdots x_N](j)^N(a_1a_2\cdots a_N) \end{aligned} \qquad \text{D.12}$$

Another expression for the characteristic function of N normalised jointly Gaussian random variables $x_1,x_2\ldots x_N$ is given by Cramer [23] as,

$$M_x(a_1,a_2,\cdots,a_N)=e^{-\frac{1}{2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}E[x_ix_j]a_ia_j\right)} \qquad \text{D.13}$$

It is possible to express e^x as the following series,

$$e^x=\sum_{p=0}^{\infty}\frac{x^p}{p!} \qquad \text{D.14}$$

Writing equation D.13 in this format gives

$$M_x(a_1,a_2,\cdots,a_N)=\sum_{p=0}^{\infty}\frac{1}{p!}\left(-\frac{1}{2}\right)^p\left[\sum_{i=1}^{N}\sum_{j=1}^{N}E[x_ix_j]a_ia_j\right]^p \qquad \text{D.15}$$

Equation D.15 can then be expressed as,

$$M_x(a_1,a_2,\cdots,a_N) = 1 + \left(-\frac{1}{2}\right)\sum_{k_1=1}^{N}\sum_{k_2=1}^{N} E[x_{k_1}x_{k_2}]a_{k_1}a_{k_2} + \frac{1}{2!}\left(-\frac{1}{2}\right)^2 \sum_{k_1=1}^{N}\sum_{k_2=1}^{N}\sum_{k_3=1}^{N}\sum_{k_4=1}^{N} E[x_{k_1}x_{k_2}]E[x_{k_3}x_{k_4}]a_{k_1}a_{k_2}a_{k_3}a_{k_4} + \cdots \quad \text{D.16}$$

Equation D.16 contains no odd products of a, so equating this with equation D.12 means that all odd terms in equation D.12 must be zero. Hence,

$$x_1x_2\cdots x_{2M+1} = 0 \quad \text{D.17}$$

Thus proving equation D.2.

Taking equation D.15 and setting N=2M, P=M, and using the expansion,

$$\left(\sum_{k=1}^{N} x_k y_k\right)^M = \sum_{k_1=1}^{N}\sum_{k_2=1}^{N}\cdots\sum_{k_M=1}^{N}(x_{k_1}x_{k_2}\cdots x_{k_M})(y_{k_1}y_{k_2}\cdots y_{k_M}) \quad \text{D.18}$$

will give the term of equation D.15 that contains only the product $(a_{k_1}a_{k_2}\cdots a_{k_{2M}})$ as in equation D.12. Hence,

$$M_x(a_1,a_2,\cdots a_{2M}) = \frac{1}{M!}\left(-\frac{1}{2}\right)^M \sum_{k_1=1}^{2M}\sum_{k_2=1}^{2M}\cdots\sum_{k_{2M}=1}^{2M} E[x_{k_1}x_{k_2}]\cdots E[x_{k_{2M-1}}x_{k_{2M}}](a_{k_1}a_{k_2}\cdots a_{k_{2M}}) \quad \text{D.19}$$

Taking the term for which $k_1 \neq k_2 \neq \ldots \neq k_{2M}$ produces an equation of the form of equation D.12. That is,

$$M_x(a_1,a_2,\cdots,a_{2M}) = \frac{1}{M!}\left(-\frac{1}{2}\right)^M (a_1a_2\cdots a_{2M}) \sum E[x_{k_1}x_{k_2}]\cdots E[x_{k_{2M-1}}x_{k_{2M}}] \quad \text{D.20}$$

where the sum is over all terms where $k_1 \neq k_2 \neq \ldots \neq k_{2M}$. However there are many terms in equation D.20 which are the same. There are 2^M of the form $x_i x_j = x_j x_i$ and M! of the form $E[x_i x_j]\,E[x_k x_l] = E[x_k x_l]\,E[x_i x_j]$. Rewriting equation D.20 summing all these terms gives,

$$M_x(a_1, a_2, \cdots a_{2M}) = (-1)^M (a_1 a_2 \cdots a_{2M}) \sum \prod E[x_{k_i} x_{k_j}] \qquad \text{D.21}$$

where $\prod E[x_{k_i} x_{k_j}]$ means $E[x_{k_1} x_{k_2}]\,E[x_{k_3} x_{k_4}] \cdots E[x_{k_{2M-1}} x_{k_{2M}}]$ and the sum is over all distinct ways of forming the product.

Equating equation D.12 with equation D.21,

$$E[x_1 x_2 \cdots x_N](j)^N (a_1 a_2 \cdots a_N) = (-1)^M (a_1 a_2 \cdots a_{2M}) \sum \prod E[x_{k_i} x_{k_j}] \qquad \text{D.22}$$

Therefore,

$$E[x_1 x_2 \cdots x_{2M}] = \sum \prod E[x_i x_j] \qquad \text{D.23}$$

Thus proving equation D.2.

Appendix E

Solution of the Cubic Volterra Model

The solution to the third order Volterra model is a straight extension to the second order model. The third order Volterra model is,

$$y(t) = \int_{-\infty}^{\infty} h_1(u)x(t-u)du + \int\int_{-\infty}^{\infty} h_2(u,v)x(t-u)x(t-v)du\,dv + \int\int\int_{-\infty}^{\infty} h_3(u,v,w)x(t-u)x(t-v)x(t-w)du\,dv\,dw \quad \text{E.1}$$

Using the cross correlation function,

$$R_{XY}(\tau) = E[x(t-\tau)y(t)] \quad \text{E.2}$$

and substituting from equation E.1 for y(t) gives,

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h_1(u)\,E[x(t-\tau)x(t-u)]\,du + \int\int_{-\infty}^{\infty} h_2(u,v)\,E[x(t-\tau)x(t-u)x(t-v)]\,du\,dv + \int\int\int_{-\infty}^{\infty} h_3(u,v,w)\,E[x(t-\tau)x(t-u)x(t-v)x(t-w)]\,du\,dv\,dw \quad \text{E.3}$$

As the expected value of an odd number of Gaussian variables is zero, the quadratic term is zero. The cubic term is the product of four Gaussian variables and so can be written as the sum of three pairs of products. Hence,

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h_1(u)\, R_{XX}(\tau - u)\, du + \iiint_{-\infty}^{\infty} h_3(u,v,w)\{R_{XX}(\tau - u)R_{XX}(v - w) + R_{XX}(\tau - v)R_{XX}(u - w) + R_{XX}(\tau - w)R_{XX}(u - v)\}du\, dv\, dw \qquad \text{E.4}$$

Taking the Fourier transform with respect to τ gives,

$$\begin{aligned} S_{XY}(f) = H_1(f)S_{XX}(f) &+ S_{XX}(f) \iiint_{-\infty}^{\infty} h_3(u,v,w)\, R_{XX}(v - w)e^{j2\pi fu}du\, dv\, dw \\ &+ S_{XX}(f) \iiint_{-\infty}^{\infty} h_3(u,v,w)\, R_{XX}(u - w)e^{j2\pi fv}du\, dv\, dw \\ &+ S_{XX}(f) \iiint_{-\infty}^{\infty} h_3(u,v,w)\, R_{XX}(u - v)e^{j2\pi fw}du\, dv\, dw \end{aligned} \qquad \text{E.5}$$

Substituting $R_{XX}(v - w) = \int_{-\infty}^{\infty} S_{XX}(g)e^{j2\pi g(v-w)}\, dg$ into the first part of the cubic term,

$$S_{XX}(f) \iiint_{-\infty}^{\infty} h_3(u,v,w)\, R_{XX}(v - w)e^{j2\pi fu}du\, dv\, dw \qquad \text{E.6}$$

gives,

$$\begin{aligned} &S_{XX}(f) \iiint_{-\infty}^{\infty} S_{XX}(g)\, h_3(u,v,w)\, e^{j2\pi gu}e^{j2\pi gv}e^{-j2\pi fw}du\, dv\, dw\, dg \\ &= S_{XX}(f) \int_{-\infty}^{\infty} S_{XX}(g)\, H_3(-f,g,-g)\, dg \end{aligned} \qquad \text{E.7}$$

The other parts can be treated similarly giving,

$$\begin{aligned} \frac{S_{XY}(f)}{S_{XX}(f)} = H_1(f) &+ \int_{-\infty}^{\infty} S_{XX}(g)H_3(-f,g,-g)\, dg \\ &+ \int_{-\infty}^{\infty} S_{XX}(g)H_3(-g,-f,g)\, dg \\ &+ \int_{-\infty}^{\infty} S_{XX}(g)H_3(g,-g,-f)\, dg \end{aligned} \qquad \text{E.8}$$

The symmetrical properties of the third order kernel follow on from those of the trispectrum, as described in chapter two. Using these properties equation E.8 can be written as,

$$\frac{S_{XY}(f)}{S_{XX}(f)} = H_1(f) + 3\int_{-\infty}^{\infty} S_{XX}(g)H_3(-f,g,-g)\,dg \qquad \text{E.9}$$

Using the 2nd order cumulant function,

$$R_{XXY}(\tau_1,\tau_2) = E[x(t-\tau_1)x(t-\tau_2)y(t)] \qquad \text{E.10}$$

and substituting from equation E.1 for y(t) gives,

$$\begin{aligned} R_{XXY}(\tau_1,\tau_2) = &\int_{-\infty}^{\infty} h_1(u)\,E[x(t-\tau_1)x(t-\tau_2)x(t-u)]\,du + \\ &\int\int_{-\infty}^{\infty} h_2(u,v)\,E[x(t-\tau_1)x(t-\tau_2)x(t-u)x(t-v)]\,du\,dv + \\ &\int\int\int_{-\infty}^{\infty} h_3(u,v,w)\,E[x(t-\tau_1)x(t-\tau_2)x(t-u)x(t-v)x(t-w)]\,du\,dv\,dw \end{aligned} \qquad \text{E.11}$$

The linear and cubic terms contain the product of an odd number of Gaussian variables and so are equal to zero. Hence,

$$\begin{aligned} R_{XXY}(\tau_1,\tau_2) = \int\int_{-\infty}^{\infty} h_2(u,v)\,\{&R_{XX}(\tau_1-\tau_2).R_{XX}(u-v) + \\ &R_{XX}(\tau_1-u)R_{XX}(\tau_2-v) + \\ &R_{XX}(\tau_1-v)R_{XX}(\tau_2-u)\}\,du\,dv \end{aligned} \qquad \text{E.12}$$

The double Fourier transform of the first term of equation E.12 is;

$$\begin{aligned} &\int\int\int\int_{-\infty}^{\infty} h_2(u,v)\,R_{XX}(\tau_1-\tau_2)R_{XX}(u-v)\,e^{-j2\pi f_1\tau_1}e^{-j2\pi f_2\tau_2}\,d\tau_1 d\tau_2 du\,dv \\ &= S_{XX}(f_1)\delta(f_1+f_2)\int_{-\infty}^{\infty} h_2(u,v)\,R_{XX}(u-v)\,du\,dv \\ &= 0 \quad \text{if } (f_1+f_2)\neq 0 \end{aligned} \qquad \text{E.13}$$

The double Fourier transform of the second term of equation E.12 is,

$$\iiint_{-\infty}^{\infty} h_2(u,v)\, R_{xx}(\tau_1 - u) R_{xx}(\tau_2 - v)\, e^{-j2\pi f_1 \tau_1} e^{-j2\pi f_2 \tau_2} d\tau_1 d\tau_2 du\, dv$$

substituting $\theta = \tau_1 - u$, $d\tau_1 = d\theta$ gives $\int_{-\infty}^{\infty} R_{xx}(\theta) e^{-j2\pi f_1(\theta+u)} = S_{xx}(f_1) e^{-j2\pi f_1 u}$

substituting $\vartheta = \tau_2 - v$, $d\tau_2 = d\vartheta$ gives $\int_{-\infty}^{\infty} R_{xx}(\vartheta) e^{-j2\pi f_2(\vartheta+v)} = S_{xx}(f_2) e^{-j2\pi f_2 v}$

$$= S_{xx}(f_1) S_{xx}(f_2) \int_{-\infty}^{\infty} h(u,v) e^{-j2\pi f_1 u} e^{-j2\pi f_2 v} du dv$$

$$= S_{xx}(f_1) S_{xx}(f_2) H(f_1, f_2) \qquad \text{E.14}$$

The third term of equation E.12 is the same as the second but with the roles of u and v reversed. Hence,

$$\iiint_{-\infty}^{\infty} h_2(u,v)\, R_{xx}(\tau_1 - v) R_{xx}(\tau_2 - u)\, e^{-j2\pi f_1 \tau_1} e^{-j2\pi f_2 \tau_2} d\tau_1 d\tau_2 du\, dv$$

$$= S_{xx}(f_1) S_{xx}(f_2) H(f_1, f_2) \qquad \text{E.15}$$

Therefore adding the three terms gives,

$$S_{xxy}(f_1, f_2) = 2 S_{xx}(f_1) S_{xx}(f_2) H(f_1, f_2) \qquad \text{if } (f_1 + f_2) \neq 0 \qquad \text{E.16}$$

Using the third order moment function,

$$R_{xxxy}(\tau_1, \tau_2, \tau_3) = E[x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) y(t)] \qquad \text{E.17}$$

and substituting from equation E.1 for y(t) gives,

$$\begin{aligned} R_{xxxy}(\tau_1, \tau_2, \tau_3) = & \int_{-\infty}^{\infty} h_1(u)\, E[x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) x(t - u)]\, du + \\ & \int_{-\infty}^{\infty} \int h_2(u,v)\, E[x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) x(t - u) x(t - v)]\, du\, dv + \\ & \int_{-\infty}^{\infty} \iint h_3(u,v,w)\, E[x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) x(t - u) x(t - v) x(t - w)]\, du\, dv\, dw \end{aligned}$$

E.18

The linear term of equation E.18 is,

$$\int_{-\infty}^{\infty} h_1(u)E[x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)x(t-u)]\,du = \int_{-\infty}^{\infty} h_1(u)\{R_{xx}(\tau_1-\tau_2)R_{xx}(\tau_3-u) + R_{xx}(\tau_1-\tau_3)R_{xx}(\tau_2-u) + R_{xx}(\tau_2-\tau_3)R_{xx}(\tau_1-u)\}\,du \qquad \text{E.19}$$

Taking the triple Fourier transform of the first term of equation E.19 gives:

$$\iiint \int_{-\infty}^{\infty} h_1(u).R_{xx}(\tau_1-\tau_2)R_{xx}(\tau_3-u)e^{-j2\pi f_1\tau_1}e^{-j2\pi f_2\tau_2}e^{-j2\pi f_3\tau_3}d\tau_1 d\tau_2 d\tau_3\,du$$

$$= S_{xx}(f_2)S_{xx}(f_3)\delta(f_1+f_2)H_1(f_3) = 0 \quad \text{if } (f_1+f_2)\neq 0 \qquad \text{E.20}$$

Similarly the other two terms of equation of equation E.19 are equal to zero if $(f_2+f_3) \neq 0$ and $(f_1+f_3) \neq 0$.

As the expectation of the product of an odd number of Gaussian variables is zero, the quadratic term of equation E.18 is zero. This leaves just the cubic term,

$$\iint \int_{-\infty}^{\infty} h_3(u,v,w)\,E[x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)x(t-u)x(t-v)x(t-w)]\,du\,dv\,dw \qquad \text{E.21}$$

There are 15 possible ways of arranging 3 pairs of six variables. Here they can be arranged into two classes.

There are nine of the form,

$$R_{xx}(\tau_x-\tau_y)R_{xx}(\tau_z-a)R_{xx}(b-c)$$

where $\{x,y,z\} \in \{1,2,3\}$ and $\{a,b,c\} \in \{u,v,w\}$.

The Fourier transform of these combinations contain delta functions and so equal zero if $(f_1+f_2) \neq 0$, $(f_2+f_3) \neq 0$, and $(f_1+f_3) \neq 0$.

There are six of the form,

$$R_{xx}(\tau_x + a)R_{xx}(\tau_y + b)R_{xx}(\tau_x + c)$$

where $\{x,y,z\} \in \{1,2,3\}$ and $\{a,b,c\} \in \{u,v,w\}$

The Fourier transform of these six terms are:

$$6\, S_{xx}(f_1)S_{xx}(f_2)S_{xx}(f_3)H_3(f_1,f_2,f_3) \qquad \text{E.22}$$

Therefore adding all the terms together gives,

$$S_{xxxy}(f_1,f_2,f_3) = 6\, S_{xx}(f_1)S_{xx}(f_2)S_{xx}(f_3)H_3(f_1,f_2,f_3) \qquad \text{E.23}$$

if $(f_1+f_2) \neq 0$, $(f_2+f_3) \neq 0$, and $(f_1+f_3) \neq 0$

Hence the three equations for $H_1(f)$, $H_2(f_1,f_2)$ and $H_3(f_1,f_2,f_3)$ are,

$$\frac{S_{xy}(f)}{S_{xx}(f)} = H_1(f) + 3\int_{-\infty}^{\infty} S_{xx}(g)H_3(-f,g,-g)\,dg \qquad \text{E.24}$$

$$\frac{S_{xxy}(f_1,f_2)}{S_{xx}(f_1)S_{xx}(f_2)} = 2H_2(f_1,f_2) \qquad \text{E.25}$$

$$\frac{S_{xxxy}(f_1,f_2,f_3)}{S_{xx}(f_1)S_{xx}(f_2)S_{xx}(f_3)} = 6H_3(f_1,f_2,f_3) \qquad \text{E.26}$$

Appendix F

Optimal linear filtering

Consider a linear system as in figure F.1. The object is to design an L point digital finite impulse response filter, h, to modify the input, x(n), in such a way as to minimise the mean square error, e(n), between the filter output and the desired signal, y(n). It is assumed that the process x(n) is ergodic.

The mean squared error solution is given by:

$$\min\{ E[e(n)^2] \} = \min\{ E[(y(n)-\mathbf{h}^T\mathbf{x}_n)^2] \} \qquad \text{F.1}$$

where $\mathbf{h} = [h(0)\ h(1)\ ...h(L-1)]^T$ and $\mathbf{x}_n = [x(n)\ x(n-1)\ ...\ x(n-L+1)]^T$.

An expression for the minimum of equation F.1 can be found by differentiating it with respect to **h** and equating to zero to yield:

$$-2(E[y(n)\mathbf{x}_n]-E[\mathbf{x}_n\mathbf{x}_n^T]\mathbf{h}_n) = 0 \qquad \text{F.2}$$

By defining the auto correlation matrix for x(n) as $\mathbf{R} = E[\mathbf{x}_n\ \mathbf{x}_n^T]$,and the cross correlation vector between x(n) and y(n) as $\mathbf{p} = E[y(n)\mathbf{x}_n]$ then equation F.2 can be solved for **h** to give the optimal filter vector as:

$$\mathbf{h} = \mathbf{R}^{-1}\mathbf{p} \qquad \text{F.3}$$

This is referred to as the Wiener filter or the minimum mean squared error solution.

In practise it is not possible to evaluate F.3 as only a finite data length is available and so given L samples of a time history, by replacing the expectations in the Wiener solution by time averages, the optimal least squares solution arises by solving:

$$\min\left\{\sum_{n=0}^{L-1} e(n)^2\right\} \text{ where } e(n)=y(n)-\mathbf{h}^T\mathbf{x}_n$$

This differs from the Wiener solution by virtue of the fact that it is data dependent. The optimum least squares solution is the solution which minimises the sum of the past n squared errors. The least squares filter is given by the solution of:

$$\mathbf{Xh} = \mathbf{y} \qquad \text{F.4}$$

where

$$X = \begin{bmatrix} x(L) & x(L-1) & \cdots & x(L-n+1) \\ x(L-1) & x(L-2) & \cdots & x(L-n) \\ x(L-2) & x(L-3) & \cdots & x(L-n-1) \\ \vdots & \vdots & & \vdots \\ x(3) & x(2) & \cdots & 0 \\ x(2) & x(1) & \cdots & 0 \\ x(1) & 0 & \cdots & 0 \end{bmatrix}$$

and

$$\mathbf{y} = [\, y(L)\ y(L\text{-}1)\ y(L\text{-}2)\ \ldots\ y(3)\ y(2)\ y(1)\,]^T$$

Assuming $n \geq L$ the least squares solution to this system of equations is given by:

$$\mathbf{h} = (\Phi_n)^{-1}(\theta_n) \qquad \text{F.5}$$

where

$$\Phi_n = \frac{1}{L}\sum_{n=0}^{L-1} \mathbf{x}_n \mathbf{x}_n^T = X^T X; \qquad \theta_n = \frac{1}{L}\sum_{n=0}^{L-1} y(n)\mathbf{x}_n = X^T \mathbf{y} \qquad \text{F.6}$$

From this is can be seen that the matrix Φ_n is an estimate of the auto correlation matrix R, to within a scaling factor. Similarly θ_n is an estimate of the cross correlation vector **p**. The scaling factors in both estimates are the same and in the calculation of **h** these cancel out.

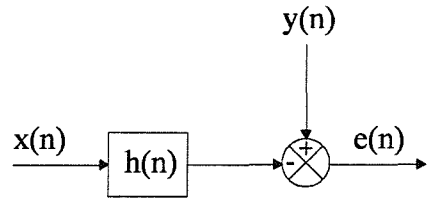

Figure F.1: Linear filtering

Chapter 9

References

[1] J F Barrett, 'Formula for output autocorrelation and spectrum of a Volterra system with stationary Gaussian input', IEE Proceedings, Vol. 127, Pt. D, No. 6, pp 286-289, (November 1980)

[2] J F Barrett, 'Functional Series Representation of Nonlinear Systems - Some Theoretical Comments', 6th IFAC Symp. Identif., Arlington, Va., USA, (1982)

[3] J S Bendat, A G Piersol, 'Spectral Analysis of nonlinear systems involving square law operations', Journal of Sound and Vibration, Vol. 81, pp 199-213, (1982)

[4] J S Bendat, A G Piersol, 'Decomposition of Wave Forces into Linear and Nonlinear components', Journal of Sound and Vibration, Vol. 106, Pt. 3, pp 391-408, (1986)

[5] P Bondon, 'On the identifiability of a quadratic stochastic system', IEEE Signal Processing Workshop on Higher Order Statistics, Lake Tahoe, USA, pp 46-50, (June 1993)

[6] G E P Box, G C Tiao, **'Bayesian Inference in Statistical Analysis'**, Addison Wesley, (1973)

[7] S Boyd, L O Chua, 'Fading Memory and the Problem of Approximating Nonlinear Operators with Volterra Series', IEEE Transactions on Circuits and Systems, Vol. 32, No.11, pp 1150-1161, (1985)

[8] S Braun, '**Mechanical Signature Analysis: Theory and Applications**', Academic Press, (1986)

[9] D R Brillinger, 'The Identification of Polynomial Systems by means of Higher Order Spectra', Applications and Methods of Random Data Analysis, A Lecture Series, Institute of Sound and Vibration Research, University of Southampton, pp M1-M21, (July 1969)

[10] D R Brillinger, '**Time Series, Data Analysis and Theory**', New York: Holt, Rinehart and Winston, expanded ed., (1981)

[11] D R Brillinger, 'An Introduction to Polyspectra', Ann Math. Statist., Vol. 63, pp 1352-1374, (1965)

[12] V Chandran, S Elgar, B Vanhoff, 'Statistics of Tricoherence', IEEE Transactions on Signal Processing, Vol. 42, No. 12, pp 3430-3440, (1994)

[13] V Chandran, 'On the Computation and Interpretation of Auto and Cross Trispectra', International Conference on Acoustics Speech and Signal Processing, Adelaide, Vol. IV, pp 445-448, (1994)

[14] V Chandran, S Elgar, C Pezeshki, 'Bispectral and Trispectral Characterisation of Transition to Chaos in the Duffing Oscillator', International Journal of Bifurcation and Chaos, Vol. 3, No. 3, pp 551-557, (1993)

[15] V Chandran, S Elgar, 'A General Procedure for the Derivation of Principle Domains of Higher Order Spectra', IEEE Transactions on Signal Processing, Vol. 42, No. 1, pp 229-233, (January 1994)

[16] Y S Cho, E J Powers, 'Quadratic System Identification Using Higher Order Spectra of IID Signals', IEEE Transactions on Signal Processing, Vol. 42, No. 5, pp 1268-1271, (May 1994)

[17] DooWhan Choi, R W Miksad, E J Powers, 'Application of Digital Cross Bispectral Analysis techniques to model the nonlinear response of a moored vessel system in random seas', Journal of Sound and Vibration, Vol. 99 No. 3, pp 309-326, (1985)

[18] B L Clarkson, J K Hammond, 'Random vibration' in '**Noise and Vibration**', edited by R G White, J G Walker, chapter 5, pp 111-135, Ellis Horwood Publishers, (1982)

[19] W B Collis, J K Hammond, P R White, 'Higher Order Spectra and the Trispectrum', Sixth IEEE Digital Signal Processing Workshop, Yosemite, California, pp 85-88, (1994)

[20] W B Collis, J K Hammond, P R White, 'Bispectrum and Trispectrum of mechanical systems', IEEE Workshop on Higher Order Statistics, Begur, Spain, pp 124-129, (June 1995)

[21] W B Collis, J K Hammond, P R White, 'Higher order spectra and condition monitoring', Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques, Senlis, France, pp 351-360, (October 1995)

[22] U Craig, T Faulhaber, D Kamins, D Laidlaw, D Schlegel, J Vroom, R Gurwitz, A van Dam, "The Application Visualisation System: A Computational Environment for Scientific Visualisation.", IEEE Computer Graphics and Applications, Vol. 9, No. 4, pp 30-42, (July 1989)

[23] H Cramer, '**Methods of Statistics**', Princeton University Press, (1946)

[24] L Cremer, M Heckl, E E Ungar, '**Structure Borne Sound**', Springer-Verlag, (1973)

[25] J W Dalle Molle, M J Hinich, 'The Trispectrum', Proceedings of the Workshop Higher Order Spectral Analysis, Vail, CO, pp 68-72, (1989)

[26] J W Dalle Molle, 'Higher Order Spectral Analysis and the Trispectrum', PhD thesis, The University of Texas at Austin, (1992)

[27] J W Dalle Molle, M J Hinich, 'Trispectral analysis of stationary random time series', Journal of the Acoustical Society of America', Vol. 97(5), Pt. 1, pp 2963-2978, (May 1995)

[28] H Esmonde, J A Fitzpatrick, H J Rice, F Axisa, 'Modelling and Identification of Nonlinear Squeeze Film Dynamics', Journal of Fluids and Structures, Vol. 6, pp 223-248, (1992)

[29] J W A Fackrell, P R White, J K Hammond, R J Pinnington, A T Parsons, 'The Interpretation of the Bispectra of Vibration Signals: Part 1 - Theory', Mechanical Systems and Signal Processing, Vol. 9, Pt. 3, pp 257-266, (1995)

[30] J W A Fackrell, P R White, J K Hammond, R J Pinnington, A T Parsons, 'The Interpretation of the Bispectra of Vibration Signals: Part 2 - Experimental Results and Applications', Mechanical Systems and Signal Processing, Vol. 9, Pt. 3, pp 267-274, (1995)

[31] J W A Fackrell, S McLaughlin, 'The Higher Order Statistics of Speech Signals', IEE Colloquium on Techniques in Speech Processing', London, Digest no 1994/138, pp 7/1-7/6, (June 1994)

[32] J W A Fackrell, 'Higher Order Spectral Content of Mechanical Systems' ISVR Contract Report, (September 1993)

[33] J W A Fackrell, S McLaughlin, 'Quadratic phase coupling detection using higher order statistics', IEE Colloquium on Higher Order Statistics in Signal Processing - are they of any use?, London, Digest no 1995/111, pp 9/1-9/6, (May 1995)

[34] J W A Fackrell, S McLaughlin, P R White, 'Practical issues concerning the use of the bicoherence for the detection of quadratic phase coupling', IEEE Workshop on Higher Order Statistics, Begur, Spain, pp 310-314, (June 1995)

[35] J A Fitzpatrick, H J Rice, 'Equivalent Time Domain Procedures for Bispectral Analysis', Journal of Sound and Vibration, Vol. 137 No. 1, pp 131-134, (1990)

[36] R F Harrison, 'The non-stationary response of vehicles on rough ground', PhD Thesis, Institute of Sound and Vibration Research, University of Southampton, (1983)

[37] M J Hinich, D M Patterson, 'Identification of the coefficients in a nonlinear time series of the quadratic type', Journal of Econometrics, Vol. 30, pp 269-288, (1985)

[38] M J Hinich, G R Wilson, 'Time delay estimation using the cross bispectrum', IEEE Transactions on Signal Processing, Vol. 40, No. 1, pp 106-113, (1992)

[39] M J Hinich, M A Wolinsky, 'A test for aliasing using bispectral components', Journal of the American Statistical Association Theory and Methods, Vol. 83, No. 402, pp 449-502, (1988)

[40] M J Hinich, 'Testing for Gaussianity and linearity of a stationary time series', Journal of Time Series Analysis', Vol. 3, No. 3, pp. 169-176, (1982)

[41] J Y Hong, Y C Kim, E J Powers, ' On Modelling the Nonlinear Relationship Between Fluctuations with Nonlinear transfer functions', Proceedings of the IEEE, Vol. 68, No. 8. pp 1026-1027, (August 1980)

[42] I Howard, 'A Review of Rolling Element Bearing Vibration " Detection Diagnosis and Prognosis"', Department of Defence, Defence Science and Technology Organisation, Aeronautical and Maritime Research Laboratory, Airframes and Engines Division, Report DSTO-RR-0013

[43] S Im, E J Powers, 'Extended Principle Domain for Volterra Models', IEEE Workshop on Higher Order Statistics, Begur, Spain, pp 381-385, June (1995)

[44] S Im, S B Kim, E J Powers, 'Utilisation of orthogonal higher order coherence functions for cubic Volterra model identification', IEEE Signal Processing Workshop on Higher Order Statistics, Lake Tahoe, USA, pp 116-120, (June 1993)

[45] Y C Kim, E J Powers, 'Digital Bispectral Analysis and Its Applications to Nonlinear Wave Interactions', IEEE Transactions on Plasma Science, Vol. PS-7, No. 2, pp 120-131, (1979)

[46] Y C Kim, E J Powers, 'A Digital Method of Modelling Quadratically Nonlinear Systems with a General Random Input', IEEE Transactions on Acoustics Speech and Signal Processing, Vol. 36, No. 11, pp 1758-1769, (1988)

[47] V Kravtchenko-Berejnoi, F Lefeuvre, V Kransnossel'skikh, D Lagoutte, 'On the use of tricoherent analysis to detect nonlinear wave-wave interactions', Signal Processing, 42, pp 291-309, (1995)

[48] J Le Roux, C Coroyer, D Rossille, 'Illustration of the effects of sampling on higher order spectra', Signal Processing, Vol. 36, pp 375-390, (1994)

[49] L D Lutes, D C K Chen, 'Trispectrum for the response of a Nonlinear Oscillator', International Journal of Nonlinear Mechanics, Vol. 26, No. 6, pp 893-909, (1991)

[50] A R Lyons, T J Newton, N J Goddard, A T Parsons, 'Can passive sonar signals be classified on the basis of their higher order statistics?', IEE Colloquium on Higher Order Statistics in Signal Processing - are they of any use?, London, Digest no 1995/111, pp 6/1-6/6, (May 1995)

[51] V J Mathews, 'Adaptive Polynomial Filters', IEEE Signal Processing Magazine, pp 11-26, (July 1991)

[52] J M Mendel, 'Tutorial on Higher Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications', Proceedings of the IEEE, Vol. 79, No. 3, pp 278-305, (March 1991)

[53] A Mood, F A Graybill, D C Boes, '**Introduction to the theory of statistics**', McGraw-Hill, (1974)

[54] F C Moon, '**Chaotic Vibrations - An Introduction for Applied Scientists and Engineers**', John Wiley & Sons, (1987)

[55] S W Nam, E J Powers, 'Application of Higher Order Spectral Analysis to Cubically Nonlinear System Identification', IEEE Transactions on Signal Processing, Vol. 42, No. 7, pp 1746-1765, (1994)

[56] A K Nandi, K Tutschku, 'Machine condition monitoring based on higher order spectra and statistics', ATHOS Workshop on Higher Order Statistics in Signal Processing, Edinburgh, (September 1994)

[57] A K Nandi, J R Dickie, J A Smith, K Tutschku, 'Classification of rotating machines using higher order statistics', IEE Colloquium on Higher Order Statistics in Signal Processing - are they of any use?, London, Digest No. 1995/111, pp 3/1-3/6, (May 1995)

[58] C L Nikias, J M Mendel, 'Signal Processing with Higher-Order Spectra', IEEE Signal Processing Magazine, pp 11-37, (July 1993)

[59] C L Nikias, A P Petropula, '**Higher order spectra analysis. A nonlinear signal processing framework**', Prentice Hall, (1993)

[60] C L Nikias, M R Raghuveer, 'Bispectrum Estimation: A Digital Signal Processing Framework', Proceedings of the IEEE, Vol. 75, No. 7, pp 869-891, (July 1987)

[61] A T Parsons, M L Williams, 'Limitations on the Use of Discrete Linear Models of Continuous Random Processes', Circuit Systems Signal Processing, Vol. 13, No. 4, pp 403-410, (1994)

[62] J B Perrochaud, 'Bispectral Analysis of Nonlinear Systems', M.Sc. Thesis, ISVR, University of Southampton, (1982)

[63] E J Powers, S Im, A Duggal, P Johnson, 'Applications of higher order statistics to nonlinar hydrodynamics', IEEE Workshop on Higher Order Statistics, Begur, Spain, pp 414-418, (June 1995)

[64] M B Priestley, '**Spectral Analysis and Time Series**', Academic Press, (1981)

[65] M B Priestley, '**Nonlinear and non-stationary time series analysis**', Academic Press, (1988)

[66] M R Raghuveer, C L Nikias, 'Bispectrum Estimation: A Parametric Approach', IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 33, No. 4, (October 1985)

[67] J Ralston, A M Zoubir, 'Identification of quadratically nonlinear systems under stationary non-Gaussian excitation', IEEE Workshop on Higher Order Statistics, Begur, Spain, pp 419-423, (June 1995)

[68] A M Richardson, W S Hodgkiss, 'Bispectral analysis of underwater acoustic data', Journal Acoustic Society of America, Vol. 96, No. 2, pp 828-837, (1994)

[69] T Sato, K Sasaki, Y Nakamura, 'Real-time Bispectral Analysis of Gear Noise and Its Application to Contactless Diagnosis', Journal of Acoustic Society of America, Vol. 62, No. 2, pp 382-387, (1977)

[70] M Schetzen, '**The Volterra and Wiener theories of nonlinear systems**', John Wiley & Sons, (1980)

[71] C A Speirs, J J Soraghan, R W Stewart, M J R Polson, 'Bispectral Analysis for the Detection of Ventricular Late Potentials', Computers in Cardiology, pp 427-430, (1993)

[72] A G Stogioglou, S McLaughlin, 'Overcoming the effects of sampling of band limited stationary analogue signals in HOC based MA modelling', IEE Colloquium on Higher Order Statistics in Signal Processing - are they of any use?, London, Digest No. 1995/111, pp 4/1-4/6, (May 1995)

[73] T Subba Rao, M M Gabr, '**An Introduction to Bispectral Analysis and Bilinear Time Series Models**', Lecture Notes in Statistics, 24, Springer-Verlag, (1984)

[74] T Subba Rao, 'Nonlinear (non-Gaussian) Time series and higher order spectra', ATHOS Workshop on Higher Order Statistics in Signal Processing, Edinburgh, (September 1994)

[75] T Subba Rao, 'Analysis of Nonlinear Time Series (and Chaos) by Bispectral Methods', Nonlinear Modelling and Forecasting, SFI Studies in the Sciences of Complexity, 199-225, Vol. XII, edited by M Casdalgi and S Eubank, Addison-Wesley, (1992)

[76] L J Tick, 'The estimation of "Transfer Functions" of Quadratic Systems', Technometrics, Vol. 3, No. 4, pp 563-567, (1961)

[77] Ching-Hsiang Tseng, E J Powers, 'Identification of cubic systems using higher order moments of IID signals.", IEEE Transactions on Signal Processing, Vol. 53, No. 7, pp 1733-1735, (July 1995)

[78] J W Tukey, '**The Collected Works of John W. Tukey**', Volumes 1and 2, edited by D R Brillinger, Monterey, California, Wadsworth, (1984)

[79] A T Walden, M L Williams, 'Deconvolution, Bandwidth, and the Trispectrum', Journal of the American Statistical Association, Vol. 88, No. 424, pp 1323-1329, (December 1993)

[80] M L Williams, 'The use of the Bispectrum and Other Higher Order Statistics in the Analysis of One Dimensional Signals', PhD thesis, Dept of Physics, Imperial College, (July 1992)

[81] M C M Wright, 'Volterra Series Characterisation and Identification of Nonlinear Bioacoustic systems', PhD thesis, ISVR, University of Southampton, (April 1993)

[82] G Zhou, S Rantala, G B Giannakis, 'Condition monitoring of rotating machinery using higher order spectral slices', IEEE Workshop on Higher Order Statistics, Begur, Spain, pp 129-133, (June 1995)

[83] A M Zoubir, 'Identification of Quadratic Volterra Systems Driven by Non-Gaussian Processes', IEEE Transactions on Signal Processing, Vol. 43, No. 5, pp 1302-1306, (May 1995)

[84] L Isserlis, 'On a formula for the product moment coefficient of any order of a normal frequency distribution in any number of variables, Biometricka, Vol. 12, pp 134-139 (1918)

[85] M J Hinich, H Messer, 'On the principal domain of the discrete bispectrum of a stationary signal', IEEE Transactions on Signal Processing, Vol. 43, No. 9, 2130-2134, (September 1995)