

Blind fault detection and source identification
using Higher Order Statistics for impacting systems

Jong-Soo Seo

A thesis submitted for the degree of
Doctor of Philosophy

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

November 2000

Acknowledgements

I would like to thank Professor Joe Hammond, for his continual guidance and encouragement since I started my study in ISVR. He is not only an excellent supervisor but also is a good friend to open my eyes in this **blind world**.

I am also grateful to Professor Steve Elliott and Dr. Paul White for their constructive advice.

In particular, I would like to thank Maureen Strickland who was always there to offer help in various ways.

Many thanks go to all the members of the Signal Processing and Control Group for their friendly environment and encouragement throughout this study.

In addition I would like to thank Joyce Shotter who has never hesitate to offer help.

Also I am deeply indebted to Korean student members in Southampton University who have helped me during these days.

Special thanks go to Young-Sup Lee and his wife for their warm tenderness toward me from the first living in UK.

Many thanks also go to Dr. R. J. Green and his team who have helped me to continue my study.

I would like to thank the encouragement and precious sacrifices from my family members, my wife Suk-Kyung You, elder daughter Eun-Joo, younger daughter Min-Jee, and son Min-Sung.

With all my heart and the greatest thanks, I am definitely obliged to say that my achievements are originated from the financial supports, sacrifices and forgiveness of my mother who has been left alone since I decided to leave my country to study in this place. Of course there has been enormous tenderness from my relatives and friends for her and me for which I appreciate so much.

Thank you for all.

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE
INSTITUTE OF SOUND AND VIBRATION RESEARCH

Doctor of Philosophy

**Blind fault detection and source identification
using Higher Order Statistics for impacting systems**

By Jong-Soo SEO

Classical deconvolution methods for source identification can only be used if the transfer function of the system is known. For many practical situations, however, this information is not accessible and/or is time varying. The problem addressed is that of reconstruction of the original input from only the measured signal. This is known as 'blind deconvolution'. By using Higher Order Statistics (HOS), the restoration of the input signal is established through the maximisation of higher order moments (cumulants) with respect to the characteristics of the signals concerned.

This paper demonstrates the restoration of input signals that have a pulse-like form. From only the measured signal (an output of the unknown system), its normalised cumulant is constructed and employed to calculate the coefficients of the inverse filter through both a Wiener approach and global optimisation. This filter is then convolved with the measured signal to give the restored signal.

The inverse filter is determined iteratively and aspects affecting convergence and performance that are investigated include: The choice of the initial inverse filter and, order determination of the filter for both nonrecursive and recursive deconvolution operators. An experimental verification is carried out for the restoration of our impacting signal arising in the response of a cantilever beam with an end stop when randomly excited.

Technique for the detection of non-Gaussian impacting signal from the observed signal through the higher order (>2) cumulant tensor (known as 'Higher Order Singular Value Decomposition, HOSVD') are introduced and discussed.

CONTENTS

Glossary of Symbols

PART I	Introduction and basic statistical foundations	1
---------------	---	----------

CHAPTER 1	Introduction	1
------------------	---------------------	----------

1.1	The aim of the research	1
1.2	Review of theoretical analysis and practical applications	2
1.2.1	Theoretical background to blind deconvolution	2
1.2.2	Practical applications of blind deconvolution	5
1.3	Structure of the thesis	8

CHAPTER 2	Analysis of Random signals	11
------------------	-----------------------------------	-----------

2.1	Introduction	11
2.2	Moments and cumulants	12
2.3	Linear filtering and statistical properties of random signals	20
2.3.1	Relationship between input and output cumulants in linear filtering	20
2.3.2	Normalisation of cumulants in blind deconvolution problems	24
2.3.3	Properties of higher order statistical parameters (computational review)	27
2.4	Concluding Remarks	42

PART II Fundamental considerations on Higher Order Statistics (HOS) 44

CHAPTER 3 The deconvolution problem and Higher Order Statistics 44

3.1	Introduction	44
3.2	Optimal deconvolution (The “classic” Wiener optimisation problem)	46
3.3	Higher order deconvolution (Blind Deconvolution, BD)	49
3.3.1	Blind Deconvolution process with a constraint	50
3.3.2	Blind Deconvolution process with normalisation	53
3.3.3	Examples of signal restoration using the constrained and normalised objective functions	54
3.3.4	IIR deconvolution operator	59
3.3.5	Comparison of MA, AR and ARMA deconvolution operators	60
3.4	Concluding remarks	71

CHAPTER 4 Higher Order Singular Value Decomposition (HOSVD) (Application to detection, classification and reconstruction)	73
4.1 Introduction	73
4.2 SVD and HOSVD	75
4.2.1 Basic Definitions of Tensor Algebra	76
4.2.2 Properties of Moment and Cumulant Tensors	81
4.2.3 Singular Value Decomposition (SVD) and Higher-Order Singular Value Decomposition (HOSVD)	83
4.3 HOSVD and determination of the presence of a non-Gaussian impacting signal	90
4.3.1 Construction of the higher order tensor	91
4.3.2 Preliminaries for simulations	94
4.3.3 Detection and classification of input signal	99
4.3.4 Reconstructability assessment from HOSVD	107
4.4 Summary and conclusions	115

PART III Practical considerations for blind deconvolution 117

CHAPTER 5 Single Channel Blind Deconvolution	117
5.1 Introduction	117
5.2 Descriptions of the model and performance measures	118
5.2.1 Model (single input single output system)	118
5.2.2 Performance measures for the blind deconvolution process	119
5.3 Effect of initial filter impulse response on deconvolution	124
5.4 Determination of the inverse filter length	131
5.4.1 Inverse performance index based length determinator	132
5.4.2 Shape parameter (α coefficient) based length determinator	132
5.4.3 Equivalent spikiness (entropy of the signal) based length determinator	133
5.4.4 Simulation results and discussions	133
5.4.5 Observation parameter based predictive determinator	136
5.5 Summary and conclusion	147

CHAPTER 6 Global optimisation in Blind Deconvolution	148
6.1 Introduction	148
6.2 Differential Evolution	149
6.2.1 Global optimisation scheme using Differential Evolution	150
6.2.2 Performance test for Differential Evolution method	155
6.3 Application of DE to Blind Deconvolution	161
6.3.1 Preliminaries for simulations	161
6.3.2 Global maximum searching and restoration of signals	163
6.3.3 Computational efficiency of DE optimisation method	167
6.4 Summary and conclusion	170

CHAPTER 9 Conclusions	218
9.1 General conclusion	218
9.2 Future research	221
References	224
Appendix A	Central limit theorem and partial order
Appendix B	Numerical expansion of the higher order Deconvolution operator (part of Chapter 3)
Appendix C	Scheme of Independent Component Analysis (Blind Source Separation, BSS)
Appendix D	Detection, classification and reconstruction of Input signals using HOSVD (summary of simulation results of Chapter 4)
Appendix E	Determination of the inverse filter length (simulation results for Chapter 5)
Appendix F	Independence and source separation structure (part of Chapter 7)

CHAPTER 9 Conclusions	218
9.1 General conclusion	218
9.2 Future research	221
References	224
Appendix A	Central limit theorem and partial order
Appendix B	Numerical expansion of the higher order Deconvolution operator (part of Chapter 3)
Appendix C	Scheme of Independent Component Analysis (Blind Source Separation, BSS)
Appendix D	Detection, classification and reconstruction of Input signals using HOSVD (summary of simulation results of Chapter 4)
Appendix E	Determination of the inverse filter length (simulation results for Chapter 5)
Appendix F	Independence and source separation structure (part of Chapter 7)

Glossary of Symbols

α	Shape parameter for Generalised Gaussian
β	Scale parameter for Generalised Gaussian
δ	Kroneckor delta function
ϕ_x	Moment generating function of signal x
γ_x^p	p -th order normalized cumulant of signal x
μ	Adaptation gain
μ_k^x	k -th order central moment of signal x
λ	Eigen value, Lagrange multiplier
Λ	Singular values of covariance matrix
σ	Singular values
σ_x^2	Varianace of signal x
τ_i	i -th time lag
ψ_x	Cumulant generating function of signal x
\mathcal{A}, \mathcal{B}	N -th order tensors ($\in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$)
c_x^p	p -th order cumulant of signal x
$\mathcal{C}_x^{(N)}$	N -th order cumulant tensor of signal x
$d(n)$	Desired signal
$e(n)$	Error signal

$E\{ \}$	Expectation
E	Entropy of signal, singular values of covariance matrix
$f(k)$	Impulse response of inverse system
$g(k)$	Impulse response of cascade system (convolution-deconvolution)
$h(k)$	Impulse response of unknown system
J	Cost function
K	Order integer (>0)
$K_x(p, q)$	p, q order normalized cumulant of signal x
L	Length of FIR inverse filter
m	Number of data channels
m_k^x	k -th order moment of signal x
$\mathcal{M}_x^{(N)}$	N -th order moment tensor of signal x
n	Number of data points, integer (>0)
N	Total number of data points
O	Objective function
p	Order integer (>0)
$p(x)$	Probability density function of signal x
q	Order integer (>0)
r	Order integer (>0)
\mathbf{R}_{vv}	Autocorrelation function of observed signal (toeplitz matrix)
s	Order integer (>0), singular values
\mathcal{S}	Singular value tensor

u	Left singular vector
v	Right singular vector
\mathbf{v}	Observed (measured) signal vector
\mathbf{w}	Noise signal vector
\mathbf{x}	Input signal vector
X	Random variable y
\mathbf{y}	Restored signal vector
Y	Random variable y
\mathbf{z}	Output signal vector (noise free)

PART I Introduction and basic statistical foundations

Chapter 1

Introduction

1.1 The aim of the research

For many mechanical systems undergoing normal operation, indications of malfunctions and advance warning of system failure may be contained in measurements of physical characteristics. Similar situations arise in seismic information, characterising earthquake symptoms or monitoring of ECG (electrocardiograph) and other signals for heart conditions. The problems addressed here relate to obtaining more reliable and consistent detection of the so-called “hidden” signals which are the causes of system malfunctioning. These hidden signals are not directly measurable. The determination of these ‘causes’ from output variables is an inverse problem. These problems can sometimes be straightforward when the system through which the causes pass is known. For many physical situations, however, where it is impractical to assume the availability of the system characteristics we require restoration of the original input signal solely from the measured (observed) signal. In this case, the restoration is called *blind inversion*.

As can be seen in Figure 1.1.1, suppose a process consists of an input signal (which cannot be observed) and an unknown linear system producing a measured signal which is often noise corrupted.

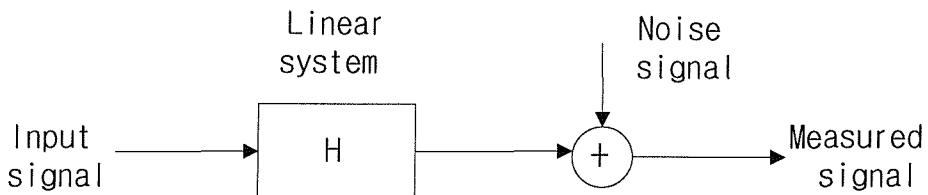


Figure 1.1.1 Linear Time-Invariant (LTI) system and signals

The aim here is to find the *input signal* from the *measured signal* alone. One practical representative example of the above situation can be found in condition monitoring, which may require the identification of the ‘cause’ of a mechanical imbalance or impacting phenomenon which can arise in rotating machinery. For the retrieval of this cause, the statistical properties of the measured signal and their relationships with the unknown linear system are considered; Specifically, higher order statistics (more than second order) are the key to possible solutions and are the focus of this work.

1.2 Review of theoretical analysis and practical applications

1.2.1 Theoretical background to blind deconvolution

If the input (original) signal cannot be observed, we may be able to utilise its statistical characteristics as the basis of its restoration. In this work, the structure of system responding to the input is assumed to be linear and time invariant (LTI) characterised by the time-domain impulse response sequence. Recently, Higher-Order (≥ 3) Statistics (HOS) [Mendel, 1991; Nikias and Petropulu, 1993] have been considered in various signal processing areas, and used to find optimum inverse filters to restore the original input signal. In particular, cumulants display the degree of higher-order correlation and also provide a measure of the

“departure” from the Gaussianity. The advantages of HOS are due to their ability to carry the phase information of a signal or a system and to suppress any (white or coloured) Gaussian additive noise [Nikias and Raghubeer, 1987].

Under certain conditions, such as for non-Gaussian, independent, identically, distributed (*i.i.d.*) signals, Donoho [Donoho, 1981] has shown that the probability distribution of a linear combination of these signals tend to become ‘closer’ to Gaussian (this is sometimes referred to as *partial order*) than that of the individual components before the linear combination (e.g., input signals). Based on this, the idea of blind deconvolution is approached by selecting an inverse system that can *decrease* the Gaussianity of the output of the inverse system. Thus, maximising an appropriately selected function (which can represent the degree of the Gaussianity) with respect to the parameters (coefficients of the linear inverse filter) of the inverse system achieves blind deconvolution. Concerning this ‘appropriate’ function, Wiggins [Wiggins, 1978] has proposed an *objective function* which consists of two cumulants (i.e., the fourth-order cumulant divided by the squared second-order cumulant), which is called ‘Minimum Entropy Deconvolution (MED)’. This objective function is called the *kurtosis* and can be related to the partial order described by Donoho. When the kurtosis of any signal is greater than 3 (or greater than zero, according to another widespread definition of kurtosis), this is referred to the ‘super-Gaussianity’ or when smaller than 3 it is known as ‘sub-Gaussianity’. For both cases, *maximisation* of the absolute value of the objective function has been used to yield the reconstruction of the input signal. Other criteria have been proposed. For example, another objective function, namely entropy has been suggested by Claerbout [Claerbout, 1977] leading to ‘parsimonious deconvolution’. He considered that MED is excessively biased towards the larger events and a method which ‘sees more’ of the data would result in a better deconvolution. It was essentially a generalisation of Wiggins’ work [Wiggins, 1978]. Ooe and Ulrych [Ooe and Ulrych , 1978] proposed maximising a modified ratio with the application of an exponential transformation of the measured data, which lead to faster convergence relative to MED. These have been a variety of attempts on the problem and Figure 1.2.1 illustrates the various

objective functions that have been used.

For any of the methods described above, by pushing the objective functions toward their maximum (by adjusting the inverse filter coefficients) one attains blind deconvolution (sometimes referred to as equalisation). The procedure of obtaining the coefficients of the inverse filter from which we can reconstruct the unknown input signal may be achieved by (i) (nonlinear) iterative methods using a matrix equation (e.g., [Nandi and Mampel, 1997]) or (ii) a stochastic gradient method using an updating parameter to maximise/minimise the objective function (e.g., [Cadzow, 1996]).

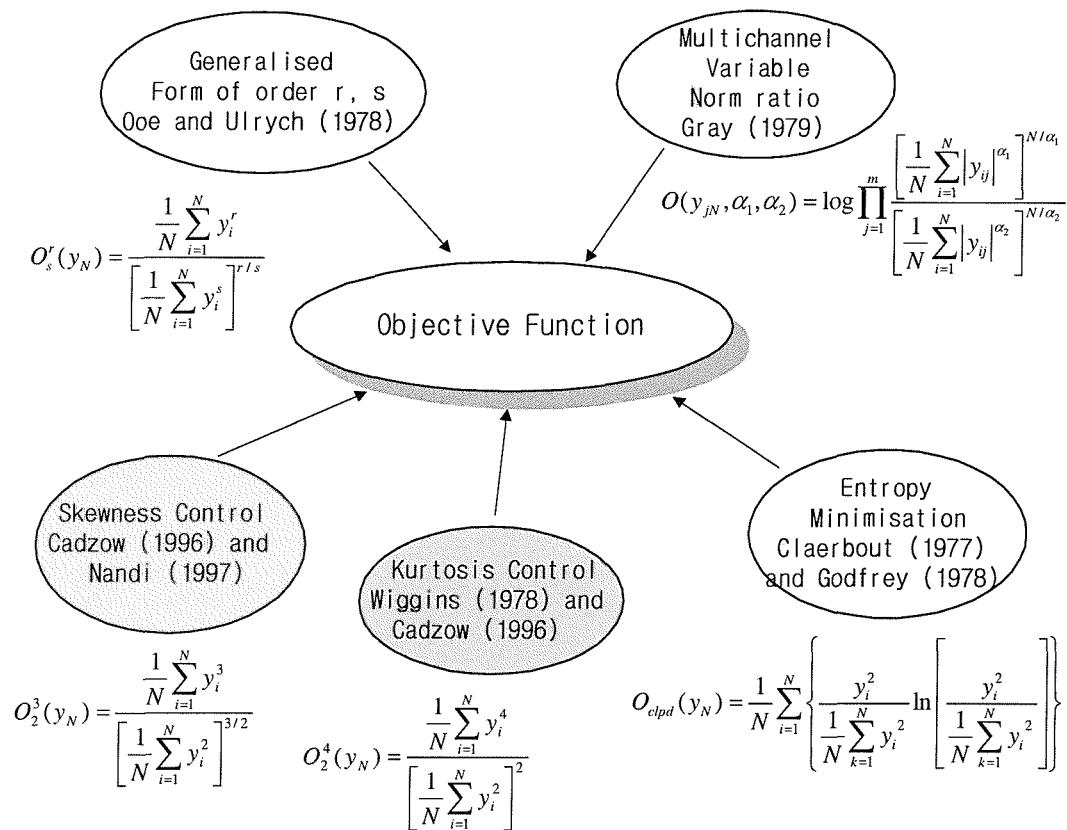


Figure 1.2.1 The objective functions

1.2.2 Practical applications of blind deconvolution

As practical motivation for blind deconvolution (and equalisation), these methods have been applied in different fields. The name blind deconvolution was first used by Stockham [Stockham et al, 1975] for the restoration of old recordings based on a model of the signal. We note a few of these applications.

A high speed data transmission over a communication channel (e.g., telephone channel) relies on the use of adaptive equalisation. In its traditional form, adaptive equalisation requires the transmission of a training sequence, the exact form of which is known at the receiver. The training sequence provides the “desired response” for the adjustment of the tap weights (filter coefficients) of a linear transversal filter (i.e., Finite Impulse Response, FIR filter), so as to minimise the mean-square value of the error signal. There are, however, practical situations where it is not feasible to use a training sequence. For example, in *a digital radio system* the received signal suffers from a phenomenon known as *multipath*, which arises from the fact of the transmitted signal reaches the receiver via a multiplicity of paths. This presence of multipath can produce severe channel fading [Pozidis and Petropulu, 1997; Tsatsanis and Giannakis, 1997] and therefore system outage, characterised by a significant reduction in the received signal power. If the outage occurs during the training process, the adaptive equaliser in the receiver is deprived of its desired response, and the adaptive filtering process is thereby seriously impaired. In such a situation, one is compelled to use some form of blind equalisation which does not require the use of a training sequence for the adjustment of the equaliser’s tap weights. In multilevel digital transmission systems, blind deconvolution techniques are required to equalise dispersive channels (systems) when equaliser convergence is not guaranteed due to high channel distortion, and the simple solution of transmitting a training sequence known to the receiver is not feasible. This may occur, for instance, in multipoint networks for computer communication [Godard, 1980], in microwave digital radio after deep fades [Foschini, 1985].

In a *speech quality enhancement problem*, elimination of acoustic reverberation components produced by the surrounding environment of a talker (adaptive cancellation of this reverberation) is a blind deconvolution problem [Juang et al, 1995]. In a typical situation, the original speech is unobservable and the multipath channel responsible for the generation of the reverberation is unknown.

In *seismic deconvolution*, the usual procedure is to assume a layered earth model, and the requirement is to use the received signal to estimate the sequence of reflection coefficients corresponding to the various layers of the model. The received signal is itself made up of echoes produced at the different layers of the model in response to the excitation which is ordinarily in the form of a short-duration pulse. The equally spaced time sequence of reflection coefficients may be viewed as the impulse response of the layered earth model. In this case, the deconvolution problem is complicated by the fact that the exact waveform of the excitation responsible (called a wavelet) for the generation of the received signal is usually unknown. Wiggins [Wiggins, 1977] introduced *minimum entropy deconvolution* (MED) in seismic data analysis, seeking the phase and amplitude of that transfer function of the inverse channel (system) that maximises the kurtosis of the deconvolved data. The restoration of seismic reflectivity series is often achieved by multichannel deconvolution. Gray [Gray, 1979] has introduced a *variable norm deconvolution* from which appropriate objective functions are selected depending on the statistical characteristics of each channel by assuming one of a generalised Gaussian function family. As shown in Figure 1.2.1, the coefficients α_1 and α_2 are adjustable to characterise the data. Ideally, the setting of these coefficients can be made in each calculation by inspecting the probability distribution of the currently estimated reflectivity series on each channel.

For *Non-destructive Evaluation* (NDE) of materials problems, which aims to restore the impulsive spikes leaving the material under examination [Nandi and Mampel, 1997], the scheme of the objective function is applied as for the problem of seismic deconvolution.

A similar problem arises in *image restoration*. In this application, an unknown system represent the blurring effects [Kundur and Hatzinakos, 1996] caused by

photographic or electronic imperfections or both. An original image or scene of interest constitutes the system input. The system output is a blurred version of the original image. Given the blurred image, the requirement is to restore the original image.

Among the above applications of the deconvolution problems, *mechanical system diagnosis and condition monitoring* is an important issue in engineering. The identification of defects inducing periodic [Pachaud et al, 1997] or non-periodic impulsive forces from measured signals can be achieved through the utilisation of the higher order statistical values (e.g. skewness and kurtosis) of the available data. It is this notion that is investigated in what follows.

1.3 Structure of the thesis

This thesis is presented in three parts:

Part I (Chapters 1, 2); This is a general introduction to blind deconvolution techniques and higher order statistical properties of random signals, Part II (Chapters 3, 4); These describe fundamental concepts of blind deconvolution problems using higher order (normalised) cumulants and the application of higher order singular value decomposition, and Part III (Chapters 5, 6, 7 and 8); This includes the practical application of higher order statistics with detailed descriptions of blind deconvolution and includes experimental verification. The contents of the chapters is summarised below, emphasising the novel contribution contained in each.

Chapter 2: This is a theoretical background to Higher Order Statistics (HOS), and defines moments and cumulants. The merits of higher-order cumulants are explained and the justification for the use of normalised higher order cumulants in deconvolution problem is provided. To verify the theoretical properties of independent, identically distributed (*i.i.d.*) signals in higher order statistics and to validate the use of higher order cumulants for impacting signals restoration, some computational simulations are carried out using two different types of random signals namely exponentially distributed *i.i.d.* and a random impulse train signal. This chapter focuses on the theme of this thesis emphasising the validity of the application of higher order statistics to the source signal reconstruction problem.

Chapter 3 deals with the formulations and procedures of blind deconvolution using constrained and normalised cumulants as the objective functions and their use in blind deconvolution problems. Discussions of three different inverse filter types (MA, AR, and ARMA systems) are provided and their performances in restoring source signals are compared. The behaviour of the objective functions for each type of inverse filter is tested. This chapter justifies a fundamental procedure for blind deconvolution based on a local optimisation.

Chapter 4 introduces the concept of Singular Value Decomposition (SVD) of random signals and Higher-Order Singular Value Decomposition (HOSVD) via the construction of higher-order tensors of measured signals. The singular values from this tensorial decomposition become essential parameters for checking *the existence* of non-Gaussian impacting signals and also provide an approximate guide to the classification of measured signals which are appropriate for impacting signal reconstruction. The ability of HOSVD for detection, classification and reconstructability of non-Gaussian signals through various simulations are provided. The aim of this contribution is to put HOSVD/tensors into a practical context. This has included computational experiments and from this the deduction of empirical criteria for detection.

Chapter 5 describes a blind deconvolution procedure to restore an input signal using the third- and fourth-order cumulants in which the effect of initial inverse filter selection and determination of the appropriate length of the FIR inverse filter is investigated. Since the blind deconvolution problem inherently suffers from a lack of information, this chapter gives a guide to practical signal reconstruction, especially with respect to the selection of inverse filter type and inverse filter length determination.

In Chapter 6, a generalised Wiener optimisation scheme is described and contrasted with an optimisation method known as Differential Evolution (DE) (similar to the Genetic Algorithm) used for seeking a global optimum. Signal restoration is demonstrated using both the generalised Wiener approach and DE method.

Chapter 7 deals with multichannel signal processing for source signal identification problems; including Blind Source Separation (BSS) for convolutive channels and a multichannel blind deconvolution process. The results provide practical tools for impacting signal reconstruction.

Chapter 8 concludes the research with an experimental verification of blind deconvolution. We demonstrate a reconstruction of an impacting signal under controlled experimental, but reasonably realistic conditions.

Conclusions and future research paths are described in Chapter 9.

Chapter 2

Analysis of Random signals

2.1 Introduction

In signal processing, signals may be broadly classified into two types – deterministic and random. A *deterministic* signal is one that may be reproduced exactly in repeated measurements. The unit impulse response of a linear time invariant filter is an example of a deterministic signal. A *random* signal, or random process, is a signal that is not exactly repeatable. Tape hiss or background clutter in radar images, speckle noise in synthetic aperture radar (SAR) images, and engine noise in speech transmission from the cockpit of an aircraft are examples. In general practical signals are a mixture of the two. This thesis emphasises the modelling of signals as random process.

In this chapter, the background that is necessary to understand how a random process may be described and how its statistical properties are affected by a linear time-invariant system is presented. This includes the introduction of cumulants and their relationship to moments, how changes occur as a result of linear filtering and noise corruption, leading to justification of the use of normalised higher order cumulants for the deconvolution problem (in Chapter 3).

2.2 Moments and cumulants

In signal processing, we often express a random variable as $X(n)$ and its realisation as $x(n)$, where the indexing variable, n , takes on integer values and is frequently associated with time. This discrete random variable sequence is often referred to as a *random time series*. These random variables are described by the probability density function $p_X(x)$. For convenience, the simple notation $p(x)$ will be employed as a probability density function (p.d.f.) of a random variable X . The probability density function is analogous to a unit mass distributed along the x -axis. If this mass is distributed in a continuous fashion, then X is said to be a continuous random variable and $p(x)$ is a continuous function of x . On the other hand, if the mass is located at only a finite or a countably infinite number of points on the x -axis, X is said to be a discrete random variable and $p(x)$ is composed of a sum of weighted displaced Dirac delta functions [Bendat and Piersol, 1986]. When the unit mass is distributed in both a continuous and a discrete fashion, the random variable is mixed and the associated probability density function contains both continuous and Dirac delta components.

Moments and the moment generating function

Moments are introduced to summarise the manner in which the unit mass is distributed in terms of a set of discrete parameters. In particular, the n th order moment of random variable X is specified by

$$E\{X^n\} = \int_{-\infty}^{\infty} x^n p(x) dx, \quad \text{for } n=1, 2, \dots \quad (2.2.1)$$

where the symbol E denotes the expected value operator. If the n th order moment exists, it then follows that all moments of order smaller than n also exist. The first-order moment is commonly referred to as the *mean* value of random variable X and corresponds to the centre of gravity of the unit mass distribution. To emphasise its importance, the symbol $m_1^x = E\{X\}$ denotes the mean value.

The *Fourier transform* serves as an important analysis and synthesis tool in

mathematically based disciplines. It is therefore quite natural that its use in the study of random variables be considered. In particular, the Fourier transform of the probability density function of the random variable X is

$$\begin{aligned}\phi_X(\omega) &= \int_{-\infty}^{\infty} e^{j\omega x} p(x) dx \\ &= E\{e^{j\omega X}\}\end{aligned}\quad (2.2.2)$$

This Fourier transform is referred to as the *moment generating function* of random variable X [Cadzow, 1996].

For two random variables X and Y related as below, properties of the moment generating functions include;

Translation

$$Y = X + a \Leftrightarrow \phi_Y(\omega) = e^{j\omega a} \phi_X(\omega) \quad (2.2.3)$$

Scaling

$$Y = aX \Leftrightarrow \phi_Y(\omega) = \phi_X(a\omega) \quad (2.2.4)$$

Addition

$$Y = X_1 + X_2 \Leftrightarrow \phi_Y(\omega) = \phi_{X_1}(\omega) \phi_{X_2}(\omega) \quad (2.2.5)$$

where X_1 and X_2 are mutually independent random variables and a is a scalar. Since a probability density function has unit area and is real, the moment generating functions properties follow as;

$$\phi_X(0) = 1, |\phi_X(\omega)| \leq 1 \text{ for all } \omega$$

$$\phi_X^*(\omega) = \phi_X(-\omega), \text{ where } '*' \text{ means complex conjugate}$$

We now explain why this function is called the moment generating function.

Applying Taylor series to the equation (2.2.2), the term $E\{e^{j\omega X}\}$ can be expanded

$$E\{e^{j\omega X}\} = E\left\{\sum_{k=0}^{\infty} \frac{(j\omega X)^k}{k!}\right\} = \sum_{k=0}^{\infty} \left[\frac{(j\omega)^k}{k!} E\{X^k\} \right] \quad (2.2.6)$$

If we denote $E\{X^k\} \equiv m_k^X$, the moment generating function and moments are related as follows (omitting the symbol 'X' in m_k^X for brevity);

$$\phi_X(\omega) = 1 + j\omega m_1 + \frac{(j\omega)^2}{2!} m_2 + \dots + \frac{(j\omega)^k}{k!} m_k + \dots \quad (2.2.7)$$

The k th coefficient of this expansion is obtained by evaluating k th derivative of the moment generating function and equating $\omega = 0$, which corresponds to the k th moment of the random variable,

$$\left. \frac{d^k \phi_X(\omega)}{d\omega^k} \right|_{\omega=0} = (j)^k E\{X^k\} \text{ for } k = 1, 2, 3, \dots, \quad (2.2.8)$$

Thus, the above equation (2.2.8) finally yields the *moments* up to k th order. The moments up to fourth order ($k=1,2,3,4$) are written as

$$\begin{aligned} m_1 &= \text{Mom}[X] = E\{X\}, & m_2 &= \text{Mom}[X, X] = E\{X^2\} \\ m_3 &= \text{Mom}[X, X, X] = E\{X^3\}, & m_4 &= \text{Mom}[X, X, X, X] = E\{X^4\} \end{aligned} \quad (2.2.9)$$

The *central moments* can be considered as a random variable's unit-mass distribution about its mean value. The n th order central moment of random variable X is defined by

$$\begin{aligned} \mu_k^X &= E\{[X - m_1^X]^k\} \\ &= \int_{-\infty}^{\infty} [X - m_1^X]^k p(x) dx \quad \text{for } k = 1, 2, 3, \dots \end{aligned} \quad (2.2.10)$$

Clearly, a random variable's moments and central moments are identical when its mean value is zero. The second-order central moment is commonly referred to as the *variance* of random variable X and specially denoted as $\sigma_x^2 = \mu_2^X$. Variance provides a measure of how dispersed the mass is about its centre of gravity (mean). The third-order central moment is typically used to measure the skewness of the density function about its mean value and denoted as μ_3^X i.e.,

$$\mu_3^X = E\{[X - m_1^X]^3\} \quad (2.2.11)$$

For example, the skewness measure is zero if the density function is symmetric about its mean value (and further all odd central moments other than 3 are zero as well in this case). The fourth-order central moment is often used to measure the *excess* [Cadzow, 1996] or *flatness* (i.e., kurtosis) [Braun and Hammond, 1986] of the probability density function about its mean. This is expressed as

$$\mu_4^X = E\{[X - m_1^X]^4\} \quad (2.2.12)$$

Cumulant generating function and cumulants

By taking the natural logarithm of the moment generating function defined in equation (2.2.2), another characteristic function is introduced. This logarithm is commonly referred to as the *cumulant generating function* (also called the *second characteristic function*) [Nikias and Petropulu, 1993; Barrett, 1964; Rice, 1944] and is formally specified as

$$\psi_X(\omega) = \ln[\phi_X(\omega)] = \ln[E\{e^{j\omega X}\}] \quad (2.2.13)$$

This possesses the following properties;

Translation

$$Y = X + a \Leftrightarrow \psi_Y(\omega) = e^{j\omega a} + \psi_X(\omega) \quad (2.2.14)$$

Scaling

$$Y = aX \Leftrightarrow \psi_Y(\omega) = \psi_X(a\omega) \quad (2.2.15)$$

Addition

$$Y = X_1 + X_2 \Leftrightarrow \psi_Y(\omega) = \psi_{X_1}(\omega) + \psi_{X_2}(\omega) \quad (2.2.16)$$

where X_1 and X_2 are mutually independent random variables and a is a scalar.

Note that comparing (2.2.16) to (2.2.5), the addition of the two random variables turns out to be simple *addition* of the cumulant generating function rather than *multiplication* as for moments.

From the relationship (2.2.7) and (2.2.13), a Taylor series expansion of the cumulant generating function can be written as

$$\psi_x(\omega) = \ln[\phi_x(\omega)] = \ln \left[1 + j\omega m_1 + \frac{(j\omega)^2}{2!} m_2 + \dots + \frac{(j\omega)^k}{k!} m_k + \dots \right] \quad (2.2.17)$$

and as with the same relationship between the moment generating function and moment, we define the k th order *cumulant* of the random variable by differentiating the cumulant generating function with respect to ω k times and equating $\omega=0$ as

$$c_k^x = (-j)^k \left. \frac{d^k \psi_x(\omega)}{d\omega^k} \right|_{\omega=0} \quad \text{for } k = 1, 2, 3, \dots, \quad (2.2.18)$$

from which the cumulants are expressed in terms of the moments as,

$$\begin{aligned} c_1 &= \text{Cum}[X] = m_1, \\ c_2 &= \text{Cum}[X, X] = m_2 - m_1^2, \\ c_3 &= \text{Cum}[X, X, X] = m_3 - 3m_2 m_1 + 2m_1^3, \\ c_4 &= \text{Cum}[X, X, X, X] = m_4 - 4m_3 m_1 - 3m_2^2 + 12m_2 m_1^2 - 6m_1^4. \end{aligned} \quad (2.2.19)$$

Also, the cumulants of the random variable are functions of the moments and central moments can be expressed by using (2.2.18). The first eight cumulants as functions of the central moments are

$$\left. \begin{aligned} c_1 &= m_1 \\ c_2 &= \mu_2 = \sigma^2 \\ c_3 &= \mu_3 \\ c_4 &= \mu_4 - 3\mu_2^2 \\ c_5 &= \mu_5 - 10\mu_3\mu_2 \\ c_6 &= \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3 \\ c_7 &= \mu_7 - 21\mu_5\mu_2 - 35\mu_4\mu_3 + 210\mu_3\mu_2^2 \\ c_8 &= \mu_8 - 28\mu_6\mu_2 - 56\mu_5\mu_3 - 35\mu_4^2 + 420\mu_4\mu_2^2 \\ &\quad + 560\mu_3^2\mu_2 - 630\mu_2^4 \end{aligned} \right\} \quad (2.2.20)$$

Note that,

(i) For the symmetric probability density functions, all central moments μ_k and

cumulants c_k for k odd are zero.

(ii) For the Gaussian case, all cumulants c_k of order greater than *second* ($k>2$) are *zero* [Nikias and Petropulu, 1993].

This is the key for recognising that cumulants are indications of non-Gaussianity of signals.

Moment generating function and cumulant generating function for a multivariate process

Given a set of n real random variables $\{X_1, X_2, \dots, X_n\}$, their joint moments of order $r = k_1 + k_2 + \dots + k_n$ are given by [Papoulis, 1991]

$$\begin{aligned} \text{Mom}[X_1^{k_1}, X_2^{k_2}, \dots, X_n^{k_n}] &\equiv E\{X_1^{k_1} \cdot X_2^{k_2} \cdots X_n^{k_n}\} \\ &= (-j)^r \left. \frac{\partial^r \phi_X(\omega_1, \omega_2, \dots, \omega_n)}{\partial \omega_1^{k_1} \partial \omega_2^{k_2}, \dots, \partial \omega_n^{k_n}} \right|_{\omega_1=\omega_2=\dots=\omega_n=0} \end{aligned} \quad (2.2.21)$$

where

$$\phi_X(\omega_1, \omega_2, \dots, \omega_n) \equiv E\{\exp(j(\omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n))\} \quad (2.2.22)$$

is their joint characteristic function. In the same way, the joint cumulants of order r are defined as the differentiation of the natural logarithm of joint moments i.e.,

$$\text{Cum}[X_1^{k_1}, X_2^{k_2}, \dots, X_n^{k_n}] \equiv (-j)^r \left. \frac{\partial^r \psi_X(\omega_1, \omega_2, \dots, \omega_n)}{\partial \omega_1^{k_1} \partial \omega_2^{k_2}, \dots, \partial \omega_n^{k_n}} \right|_{\omega_1=\omega_2=\dots=\omega_n=0} \quad (2.2.23)$$

where

$$\psi_X(\omega_1, \omega_2, \dots, \omega_n) \equiv \ln[\phi_X(\omega_1, \omega_2, \dots, \omega_n)] \quad (2.2.24)$$

The above may be applied to time series. If $\{X_1\}$ is a real stationary random process, cumulant sequences are described using the relations with the moments (if their moments exist up to n th order)

1st-order cumulants ;

$$c_1^{X_1} = \text{Cum}[X_1] = m_1^{X_1} = E\{X_1(k)\}, \quad (\text{mean value}) \quad (2.2.25)$$

2nd-order cumulants ;

$$c_2^{X_1}(\tau_1) = \text{Cum}[X_1, X_1] = m_2^{X_1}(\tau_1) - (m_1^{X_1})^2, \quad (\text{covariance sequence}) \quad (2.2.26)$$

where $m_2^{X_1}(\tau_1)$ is the autocorrelation sequence and τ_1 is the time lag along the sequence X_1 (assuming X_1 is a discrete time series, then τ_1 is an integer).

3rd-order cumulants ;

$$\begin{aligned} c_3^{X_1}(\tau_1, \tau_2) &= \text{Cum}[X_1, X_1, X_1] \\ &= m_3^{X_1}(\tau_1, \tau_2) - m_1^{X_1} \left[m_2^{X_1}(\tau_1) + m_2^{X_1}(\tau_2) + m_2^{X_1}(\tau_2 - \tau_1) \right] + 2(m_1^{X_1})^3 \end{aligned} \quad (2.2.27)$$

4th-order cumulants ;

$$\begin{aligned} c_4^{X_1}(\tau_1, \tau_2, \tau_3) &= \text{Cum}[X_1, X_1, X_1, X_1] \\ &= m_4^{X_1}(\tau_1, \tau_2, \tau_3) - m_2^{X_1}(\tau_1)m_2^{X_1}(\tau_3 - \tau_2) - m_2^{X_1}(\tau_2)m_2^{X_1}(\tau_3 - \tau_2) \\ &\quad + m_2^{X_1}(\tau_3)m_2^{X_1}(\tau_2 - \tau_1) - m_1^{X_1} [m_3^{X_1}(\tau_2 - \tau_1, \tau_3 - \tau_1) + m_3^{X_1}(\tau_1, \tau_3) \\ &\quad + m_3^{X_1}(\tau_2, \tau_3) + m_3^{X_1}(\tau_1, \tau_2)] + (m_1^{X_1})^2 [m_2^{X_1}(\tau_1) + m_2^{X_1}(\tau_2) + m_2^{X_1}(\tau_3)] \\ &\quad + m_2^{X_1}(\tau_3 - \tau_1)m_2^{X_1}(\tau_3 - \tau_2) + m_2^{X_1}(\tau_2 - \tau_1) - 6(m_1^{X_1})^4 \end{aligned} \quad (2.2.28)$$

which are similar expressions to (2.2.19) except now including the time lags.

Properties of cumulants

The following are some important properties of cumulants ;

[CP1] If λ_i , $i = 1, 2, \dots, k$, are constants, and X_i , $i = 1, 2, \dots, k$, are random variables, then

$$\text{Cum}(\lambda_1 X_1, \lambda_2 X_2, \dots, \lambda_k X_k) = \left(\prod_{i=1}^k \lambda_i \right) \text{Cum}(X_1, X_2, \dots, X_k).$$

[CP2] Cumulants are symmetric in their arguments, i.e.,

$$\text{Cum}(X_1, X_2, \dots, X_k) = \text{Cum}(X_{i_1}, X_{i_2}, \dots, X_{i_k}),$$

where (i_1, i_2, \dots, i_k) is a permutation of $(1, 2, \dots, k)$.

[CP3] Cumulants are additive in their arguments, i.e.,

$$\text{Cum}(X_0 + Y_0, Z_1, Z_2, \dots, Z_k) = \text{Cum}(X_0, Z_1, Z_2, \dots, Z_k) + \text{Cum}(Y_0, Z_1, Z_2, \dots, Z_k)$$

This means that cumulants of sums equal sums of cumulants (hence, the name “cumulant”).

[CP4] If α is a constant, then

$$Cum(\alpha + Z_1, Z_2, \dots, Z_k) = Cum(Z_1, Z_2, \dots, Z_k).$$

[CP5] If the random variables $\{X_i\}$ are independent of the random variables $\{Y_i\}$, $i = 1, 2, \dots, k$, then

$$Cum(X_1 + Y_1, X_2 + Y_2, \dots, X_k + Y_k) = Cum(X_1, X_2, \dots, X_k) + Cum(Y_1, Y_2, \dots, Y_k).$$

[CP6] If a subset of the k random variables $\{X_i\}$ is independent of the rest, then

$$Cum(X_1, X_2, \dots, X_k) = 0.$$

Suppose $v(n) = z(n) + w(n)$ where $z(n)$ and $w(n)$ are independent ; then from [CP5],

$c^v_k(\tau_1, \tau_2, \dots, \tau_{k-1}) = c^z_k(\tau_1, \tau_2, \dots, \tau_{k-1}) + c^w_k(\tau_1, \tau_2, \dots, \tau_{k-1})$. If $w(n)$ is Gaussian (coloured or white) and $k \geq 3$ then, $c^v_k(\tau_1, \tau_2, \dots, \tau_{k-1}) = c^z_k(\tau_1, \tau_2, \dots, \tau_{k-1})$, whereas $c^v_2(\tau) = c^z_2(\tau) + c^w_2(\tau)$. In essence, cumulants can draw non-Gaussian signals out of Gaussian noise, thereby boosting their signal-to-noise ratios.

Cumulants of an independent, identically distributed (i.i.d.) random sequence are delta (Kronecker) functions (the same is not true for joint moments), i.e., if $w(n)$ is an i.i.d. process, then [Brillinger, 1965; Giannakis and Mendel, 1989; Nikias and Petropulu, 1993]

$$c^w_k(\tau_1, \tau_2, \dots, \tau_{k-1}) = \gamma^w_k \delta(\tau_1) \delta(\tau_2) \cdots \delta(\tau_{k-1}),$$

where γ^w_k is the k th-order cumulant of the stationary random sequence $w(n)$.

The n th-order cumulant function of a non-Gaussian stationary random process $\{X(k)\}$ can be written as (for $n=3,4$ only) ;

$$c^x_n(\tau_1, \tau_2, \dots, \tau_{n-1}) = m^x_n(\tau_1, \tau_2, \dots, \tau_{n-1}) - m^G_n(\tau_1, \tau_2, \dots, \tau_{n-1}) \quad (2.2.29)$$

where $m^x_n(\tau_1, \tau_2, \dots, \tau_{n-1})$ is the n th-order moment function of $\{X(k)\}$ and

$m^G_n(\tau_1, \tau_2, \dots, \tau_{n-1})$ is the n th-order moment function of an equivalent Gaussian process that has the same mean value and autocorrelation sequence as $\{X(k)\}$. Clearly, if $\{X(k)\}$ is Gaussian, $m^x_n(\tau_1, \tau_2, \dots, \tau_{n-1}) = m^G_n(\tau_1, \tau_2, \dots, \tau_{n-1})$ and thus $c^x_n(\tau_1, \tau_2, \dots, \tau_{n-1}) = 0$. Note, however, that this is only true of orders $n=3$ and 4 .

By putting $\tau_1 = \tau_2 = \tau_3 = 0$ and for zero mean ($m^x_1 = 0$) we get

$$\begin{aligned}\gamma_2^x &= E\{x(k)^2\} = c_2^x(0) && \text{(variance)} \\ \gamma_3^x &= E\{x(k)^3\} = c_3^x(0, 0) && \text{(skewness)} \\ \gamma_4^x &= E\{x(k)^4\} - 3[\gamma_2^x]^2 = c_4^x(0, 0, 0) && \text{(kurtosis)}\end{aligned}\quad (2.2.30)$$

The normalised kurtosis is defined as $\gamma_4^x / [\gamma_2^x]^2$, which is widely used in machine condition monitoring [Braun and Hammond, 1986; Dyer and Stewart, 1978].

2.3 Linear filtering and statistical properties of random signals

This section explains the properties of the higher order statistics during the linear filtering process, leading to the use of normalised cumulants in the blind deconvolution problem.

2.3.1 Relationship between input and output cumulants in linear filtering

A familiar starting point for many problems in signal processing and system theory is the single-input single-output (SISO) linear and time-invariant (LTI) model depicted in the Figure 2.3.1.

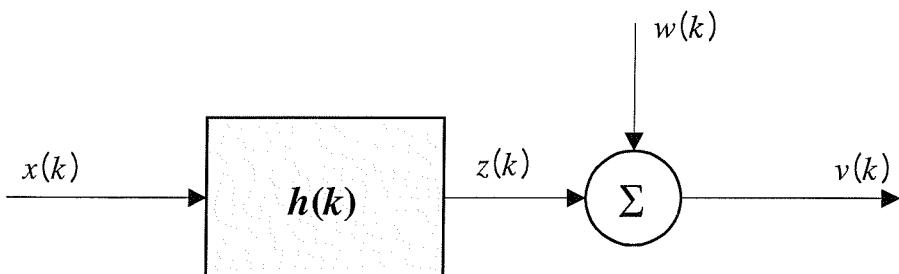


Figure 2.3.1 Single-channel system

Second order properties;

In Figure 2.3.1, in which $x(k)$ is a white sequence input with finite variance σ_x^2 ; $H(z)(=Z\{h(k)\})$ is the transfer function of a causal stable system (channel) having impulse response sequence $h(k)$; $w(k)$ is white Gaussian noise with variance σ_w^2 ; $x(k)$ and $w(k)$ are statistically independent and so therefore are $z(k)$ and $w(k)$; $z(k)$ is the output of the system that is assumed noise free; and, $v(k)$ is the output of the system corrupted by Gaussian noise. Letting $r(\bullet)$ and $S(\bullet)$ denote the correlation function and Fourier transform of the correlation function (i.e., the spectrum), respectively, then

$$r_v(k) = r_z(k) + r_w(k) = \sigma_x^2 \sum_{i=0}^{\infty} h(i)h(i+k) + \sigma_w^2 \delta(k) \quad (2.3.1)$$

$$S_v(\omega) = \sigma_x^2 |H(\omega)|^2 + \sigma_w^2 \quad (2.3.2)$$

$$r_{xv}(k) \equiv E\{x(n)v(n+k)\} = \sigma_x^2 h(k) \quad (2.3.3)$$

From equation (2.3.2) we see that all phase information has been lost in the spectrum (or in the autocorrelation); hence, we say that correlation or power spectrum is *phase blind*.

Higher order properties;

We now construct the higher order statistics of the input x and the output v .

For input x ;

The system in Figure 2.3.1 is assumed to be causal and stable, and $\{x(i)\}$ is assumed to be independent, identically, distributed (i.i.d.), and *non-Gaussian*, i.e.,

$$c_x^k(\tau_1, \tau_2, \dots, \tau_{k-1}) = \gamma_x^k \quad \text{if} \quad \tau_1 = \tau_2 = \dots = \tau_{k-1} = 0 \quad (2.3.4)$$

$$c_x^k(\tau_1, \tau_2, \dots, \tau_{k-1}) = 0 \quad \text{otherwise} \quad (2.3.5)$$

where γ_x^k denotes the k th-order cumulant of $x(i)$. The additive noise $w(k)$ is

assumed to be Gaussian.

To justify the equation (2.3.4), we note that

$$\begin{aligned}
 \text{Cum}[x(i_0), x(i_1), \dots, x(i_{k-1})] &= \text{Cum}[x(i_0), x(i_0 + i_1 - i_0), \dots, x(i_0 + i_{k-1} - i_0)] \\
 &= c^z_k(i_1 - i_0, i_2 - i_0, \dots, i_{k-1} - i_0) \\
 &= \gamma^x_k \quad \text{only if } i_1 = i_0, \dots, i_{k-1} = i_0
 \end{aligned} \tag{2.3.6}$$

To show equation (2.3.5), we have made a substitution of variables and invoked the stationarity of $z(k)$ and the causality of $h(k)$. The former tells us that $c^z_k(\tau_1, \dots, \tau_{k-1}) = c^v_k(\tau_1, \dots, \tau_{k-1})$ will not depend upon time k ; the latter tells us that $h(k) = 0$ for $k < 0$.

For output v ;

Using the property of the higher order cumulant of a Gaussian process, the k th-order cumulant of $v(k)$ equals the k th-order cumulant of $z(k)$, as $w(k)$ is assumed to be a white (or coloured) Gaussian process.

Note that the following derivation is made easier by working with the more general convolution form

$$z(k) = \sum_{i=-\infty}^{\infty} x(i)h(k-i) \tag{2.3.7}$$

where i ranges from $-\infty$ to ∞ , instead of the form associated with a causal IR (Impulse Response) for which i ranges from 0 to k ($k \geq 3$).

$$\begin{aligned}
 c^v_k(\tau_1, \tau_2, \dots, \tau_{k-1}) &= c^z_k(\tau_1, \tau_2, \dots, \tau_{k-1}) \\
 &= \text{Cum}[z(l), z(l + \tau_1), \dots, z(l + \tau_{k-1})] \\
 &= \text{Cum}[\sum_{i_0} x(i_0)h(l - i_0), \sum_{i_1} x(i_1)h(l - i_1 + \tau_1), \dots, \sum_{i_{k-1}} x(i_{k-1})h(l - i_{k-1} + \tau_{k-1})] \\
 &= \sum_{i_0} \sum_{i_1} \dots \sum_{i_{k-1}} \text{Cum}[x(i_0)h(l - i_0), x(i_1)h(l - i_1 + \tau_1), \dots, x(i_{k-1})h(l - i_{k-1} + \tau_{k-1})] \\
 &= \sum_{i_0} \sum_{i_1} \dots \sum_{i_{k-1}} h(l - i_0)h(l - i_1 + \tau_1) \dots h(l - i_{k-1} + \tau_{k-1}) \\
 &\quad \cdot \text{Cum}[x(i_0), x(i_1), \dots, x(i_{k-1})]
 \end{aligned} \tag{2.3.8}$$

To arrive at the first line of this derivation we have used cumulant property [CP3]; and, to arrive at the last line we have used cumulant property [CP1].

In the case of a white sequence input, equation (2.3.8) simplifies considerably to

$$\begin{aligned}
 c^v_k(\tau_1, \dots, \tau_{k-1}) &= \gamma^x_k \sum_{i_0} h(l-i_0) h(l-i_1 + \tau_1) \cdots h(l-i_{k-1} + \tau_{k-1}) \\
 &= \gamma^x_k \sum_{n=0}^{\infty} h(n) h(n+\tau_1) \cdots h(n+\tau_{k-1})
 \end{aligned} \tag{2.3.9}$$

Observe that when $k=2$, equation (2.3.9) reduces to equation (2.3.1) [subject to the addition of $\sigma_w^2 \delta(k)$]. The generalization of equation (2.3.9) to the case of *coloured non-Gaussian* input $x(i)$ (we will only consider the $k=3,4$ cases) we first make the substitution of variables in equation (2.3.8): $j_0 = l-i_0, j_1 = l-i_1+\tau_1, \dots, j_{k-1} = l-i_{k-1}+\tau_{k-1}$, so that equation (2.3.8) becomes

$$\begin{aligned}
 c^v_k(\tau_1, \dots, \tau_{k-1}) &= \sum_{j_0} \sum_{j_1} \cdots \sum_{j_{k-1}} h(j_0) h(j_1) \cdots h(j_{k-1}) \\
 &\quad \cdot \text{Cum}[x(l-j_0), x(l+\tau_1-j_1), \dots, x(l+\tau_{k-1}-j_{k-1})]
 \end{aligned} \tag{2.3.10}$$

Using the stationarity of $x(l)$, this equation can be expressed as

$$\begin{aligned}
 c^v_k(\tau_1, \dots, \tau_{k-1}) &= \sum_{j_0} \sum_{j_1} \cdots \sum_{j_{k-1}} h(j_0) h(j_1) \cdots h(j_{k-1}) \\
 &\quad \cdot c^x_k(j_0 - j_1 + \tau_1, j_0 - j_2 + \tau_2, \dots, j_0 - j_{k-1} + \tau_{k-1})
 \end{aligned} \tag{2.3.11}$$

Finally, making a second transformation of variables, $m_1 = j_1 - j_0, m_2 = j_2 - j_0, \dots, m_{k-1} = j_{k-1} - j_0$, we obtain the result

$$\begin{aligned}
 c^v_k(\tau_1, \tau_2, \dots, \tau_{k-1}) &= \sum_{m_1} \sum_{m_2} \cdots \sum_{m_{k-1}} c^x_k(\tau_1 - m_1, \tau_2 - m_2, \dots, \tau_{k-1} - m_{k-1}) \cdot c^h_k(m_1, m_2, \dots, m_{k-1})
 \end{aligned} \tag{2.3.12}$$

where $c^h_k(m_1, m_2, \dots, m_{k-1}) = \sum_{j_0} h(j_0) h(j_0 + m_1) \cdots h(j_0 + m_{k-1})$.

For non-Gaussian i.i.d. input signal $x(n)$, by setting $\tau_1 = \tau_2 = \dots = \tau_{k-1} = 0$ and $m_1 = m_2 = \dots = m_{k-1} = 0$, we obtain a result similar to (2.3.9)

$$c_p^v = c_p^x \sum_k h(k)^p \tag{2.3.13}$$

i.e., the zero lag response cumulant of order p is seen to be the product of the excitation cumulant of order p with the sum of the linear operator's unit-impulse response elements raised to the p th power.

2.3.2 Normalisation of cumulants in blind deconvolution problems

Having established expressions for higher order statistics for stationary inputs and outputs of linear filters, we will now introduce the deconvolution problem and in particular show the concept of normalised higher order statistics are needed.

In the blind deconvolution approach that will be developed, a desirable feature is that it be invariant to signal. To explain this, Figure 2.3.2 depicts a single input single output system in which $x(n)$ is the original input sequence (assumed *i.i.d.*), $v(n)$ is measured sequence, and $y(n)$ is output of the cascade system g_n composed of h_n and f_n (i.e. combined convolution-deconvolution operation). From this structure, the relationship of the signals in convolutive terms and cumulants are again written as

$$y = g * x \Rightarrow c_k^y = c_k^x \sum_{i=0}^{\infty} g(i)^k \quad (2.3.14)$$

where c_k^y and c_k^x are the k th-order cumulant of output and input, respectively. The right hand side of the equation (2.3.14) can be explained by the properties of cumulant.

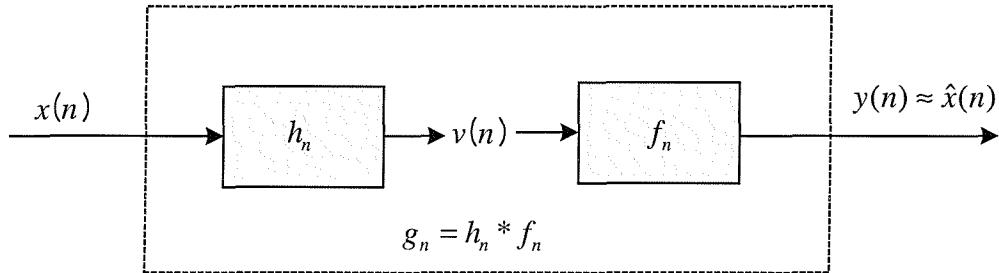


Figure 2.3.2 Single input and single output system with the convolution-deconvolution form.

In order to make the concept of the cumulant invariant with respect to scalar multiplication, it is necessary to provide a normalisation of the cumulant. Consider the *normalised cumulant* of order (p, q) associated with random variables $\{x_i\}$, $i = 1, 2, \dots, n$, as defined by

$$K_x(p, q) = \frac{c_p^x}{|c_q^x|^{p/q}}, \quad \text{for } p > q, q \text{ is even positive integer} \quad (2.3.15)$$

in which it is assumed that both cumulant c_q^x and c_p^x are nonzero (this may reflect the hypothesis that the input signal is a non-Gaussian process). Typically, the integer parameters are selected so that $p > q$, although this definition is applicable for any choice of those parameters using the definition of a cumulant. It directly follows that this normalised cumulant is scalar multiplication invariant in the sense that $K_{ax}(p, q) = K_x(p, q)$ for any positive scalar a . In many applications, the specific selection of $q=2$ provides a logical choice, since $c_2^x = \sigma_x^2$ (variance) is always nonzero for any non-trivial random variables. This particular selection yields the normalised cumulant relationship

$$K_x(p, 2) = \frac{c_p^x}{(\sigma_x^2)^{p/2}} \quad (2.3.16)$$

An important inequality condition relating normalised cumulants between the input and output signal is now described.

Denote the p and q th-order normalised cumulant of the random variables $\{y_i\}$, $i = 1, 2, \dots, N$, as $K_y(p, q)$, is expressed from equation (2.3.13)

$$K_y(p, q) = \frac{\sum_k (g_k)^p}{\left| \sum_k (g_k)^q \right|^{p/q}} K_x(p, q) \quad (2.3.17)$$

(see Figure 2.3.2 to define g_k)

For a positive even integer q ,

$$|K_y(p, q)| \leq |K_x(p, q)| \quad \text{for all even and odd } p > q \quad (2.3.18)$$

To prove equation (2.3.18), the denominator in equation (2.3.17) can be expressed by the standard l_q norm to power p , i.e.,

$$\left| \sum_k (g_k)^q \right|^{p/q} = [\|\mathbf{g}\|_q]^p \quad (2.3.19)$$

$$\text{where } \|\mathbf{g}\|_q = \left| \sum_k (g_k)^q \right|^{1/q}.$$

Hence, by taking absolute value of equation (2.3.17),

$$|K_y(p, q)| = |K_x(p, q)| \left| \sum_k (g_k)^p \right| / [\|\mathbf{g}\|_q]^p \quad (2.3.20)$$

From the inequality of the rational term for q an even integer and $p > q$ case,

$$\frac{\left| \sum_k (g_k)^p \right|}{\left\{ \|\mathbf{g}\|_q \right\}^p} \leq 1 \quad (2.3.21)$$

in which the equality holds if and only if only one of the components of \mathbf{g} is nonzero.

Using the inequality of (2.3.21), equation (2.3.20) has a relation as

$$|K_y(p, q)| \leq |K_x(p, q)| \quad (2.3.22)$$

This is a very important relationship which essentially says that the normalised cumulant of the response of a linearly filtered *i.i.d.* process is smaller than that of the input. In respect to Figure 2.3.2 in which we are trying to make y look like x then since h^*x is no longer *i.i.d.* Then the actual inverse filter \mathbf{f} (its components are expressed by f_k) is adjusted to make the effect as small as possible, i.e., adjust \mathbf{f} to maximise the normalised cumulant of the output v .

In this manner, cumulant based deconvolution problem has been based on the inequality relationship given in the equation (2.3.22) and can be said that “the required deconvolution operator must generate a response whose normalised cumulants have magnitude that are the largest over the class of all linear operators for all even $q < p$ and smallest magnitude for all even $p < q$ ” [Cadzow, 1996].

This normalised cumulant corresponds to the concept of ‘partial order’ [Donoho, 1981] and the maximisation (or minimisation) matches the inequality condition in

the blind deconvolution problems, which is introduced in Appendix A.

For brevity, we shall exclusively consider the case in which $q < p$ with q even, whereby it is desired to select \mathbf{g} so as to maximise an estimate of $|K_y(p, q)|$, that is to say, the ideal deconvolution operation is performed by maximising the magnitude of the normalised response cumulant $K_y(p, q; \mathbf{g})$, where q is any positive even integer less than p for which cumulant of order q , c_q^x is nonzero (e.g., $q=2$). This maximisation is to be made with respect to the unit-impulse response $\{g_n\}$ of the combined convolution-deconvolution operation as shown in Figure 2.3.2. Since the unit-impulse response of the unknown linear convolution operator $\{h_n\}$ is implicitly contained within the observed (measured) data $\{v(n)\}$, this maximisation must be made with respect to the deconvolution operator's unit-impulse response $\{f_n\}$. The required maximisation therefore takes the form

$$\max_{\mathbf{f}} |K_y(p, q)| = \max_{\mathbf{f}} [K_y(p, q) \operatorname{sgn}[K_y(p, q)]] \quad (2.3.23)$$

where \mathbf{f} is an appropriately dimensioned vector whose components are the elements of the unit-impulse response of the deconvolving operator. In words, we desire to find a global maximum of $|K_y(p, q)|$. Often, the necessary condition which gives a local maximum of this value is found by differentiating and equating to zero with respect to the filter \mathbf{f} , which will be discussed in Chapter 3. In addition to the *local* optimisation approach based on the above maximisation scheme, a *global* optimisation method using the *differential evolution* method is introduced in Chapter 6.

2.3.3 Properties of higher order statistical parameters (computational review)

Starting from the definitions of moments and cumulants, the explicit relationship of cumulants between the input and output signal has been justified. However, the justification given in equation (2.3.13) is strictly based on the *i.i.d.* case, which will in general not match the problems dealt with in this thesis. Consequently, this

mismatch requires further exploration to justify the use of cumulants in such problems.

This subsection provides computational simulations to observe the statistical (higher order) changes of signals along with the convolution and noise interference. We now refer to Figure 2.3.1.

Input signals (x);

Two types of non-Gaussian input sequences are considered. The first is an i.i.d. random signal with exponential distribution denoted as x_1 and referred to as the ‘zero mean i.i.d. Non-Gaussian signal’.

The second input sequence chosen to be a signal which contains a series of impulses, i.e.,

$$\begin{aligned} \{x(k)\} &= a_k \cdot \delta(k - k_0), \\ &= 0, \quad \text{for all other } k \in [1, N] \end{aligned} \quad (2.3.24)$$

where a_k is randomly selected non-zero constants, k_0 is a vector of sample numbers (uniformly distributed random integer selected from numbers spanning $[1, N]$) on which the impulses, $[0, 1, \dots, s-1]$ exist, N is length of the input sequence $x(k)$, and s is the chosen total number of impulses. This impulse sequence is chosen as it can represent mechanical effects or results of a test material in non-destructive ultrasonic examination [Nandi et al, 1997] or impacting signal from machines.

Linear system (H);

A simple MA(2) minimum phase system is selected to simplify the calculation of statistical values in this simulation described as

$$H(z) = 1 - 1.3z^{-1} + 0.4z^{-2} \quad (2.3.25)$$

which is shown in Figure 2.3.3.

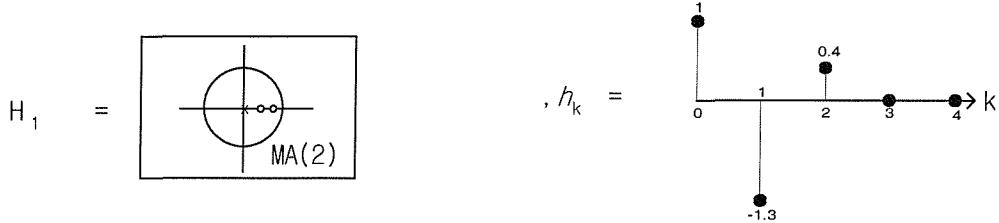


Figure 2.3.3 The pole-zero map and impulse response of the MA(2) minimum phase system used in computational simulation.

Noise signals (w);

Two randomly generated (white) Gaussian signals are used as interference signals both of which has zero dB signal to noise ratio (SNR), which is expressed as

$$\text{SNR} = 10 \log_{10} \left(\frac{\sigma_z^2}{\sigma_w^2} \right), \text{ dB} \quad (2.3.26)$$

where σ_z^2 is the variance of the output signal (z) of the linear system and σ_w^2 represents that of the noise signal (w).

Observed signals (v);

The observed signal is expressed using the non-Gaussian signal (with zero mean and unit variance), the linear system and the noise signal as

$$v(n) = \sum_{k=0}^{\infty} h(k) x(n-k) + w(n) \quad (2.3.27)$$

where k represents time sequence of the impulse response of the system.

Probability density function of each signals (p);

As an example, the probability density function (pdf) of signal x is estimated as

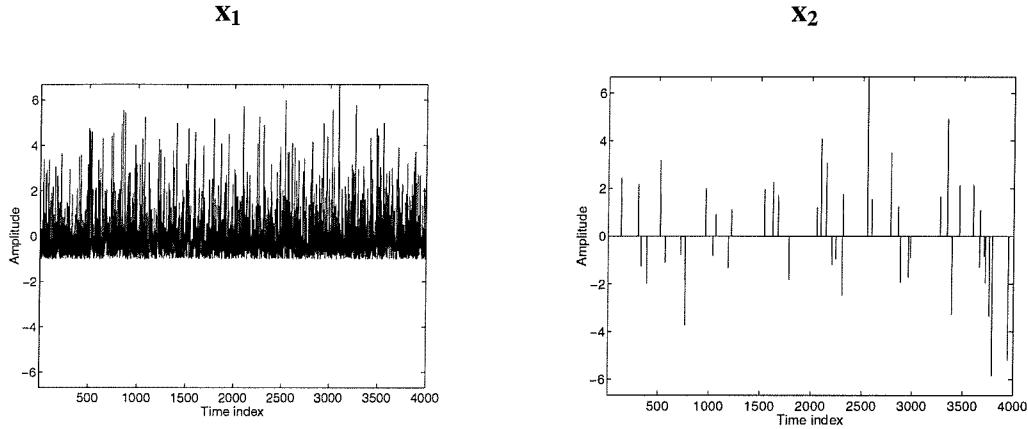
$$p(x) = \frac{\text{number of data in } \{x_i < x < x_i + \Delta x\}}{\text{number of total data length } (N)}, \quad 0 \leq \Delta x \leq \max |x| \quad (2.3.28)$$

When x has mean and variance as

$$\begin{aligned}\mu &= E\{\mathbf{x}\} \\ \sigma_x^2 &= E\{(\mathbf{x} - \mu)^2\}\end{aligned}\quad (2.3.29)$$

The shapes of each signal and statistical distributions (bar graph) are illustrated in comparison to Gaussian distribution (solid line) having the same mean and variance in the following figures;

(a) Time series of input signals



(b) Probability density of input signals

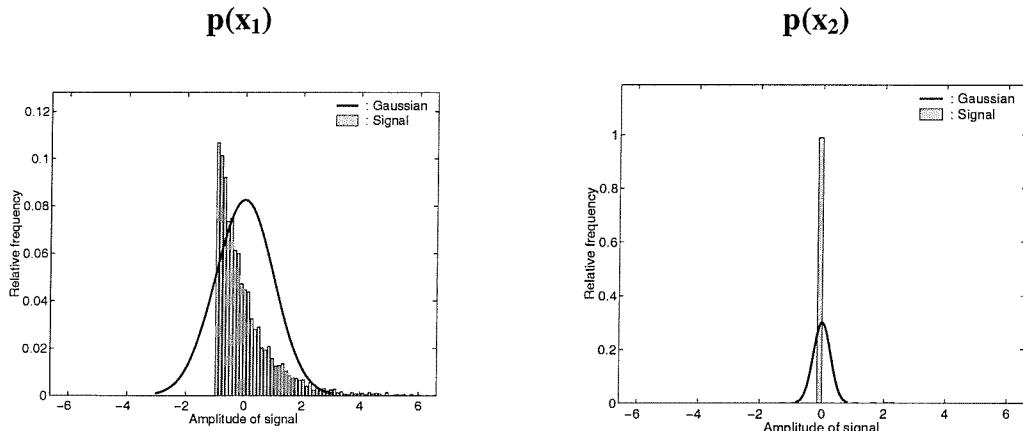
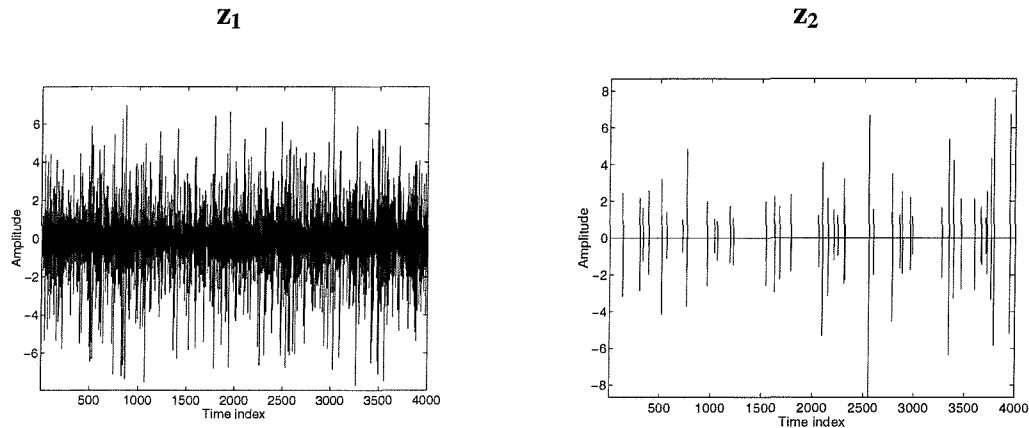


Figure 2.3.4 The input signals and their statistical characteristics; (a): signals in time domain and (b): their probability distribution (bar graph) with equivalent Gaussian distribution (solid line).

(a) Time series of output signals (noise free)



(b) Probability density of output signals (noise free)

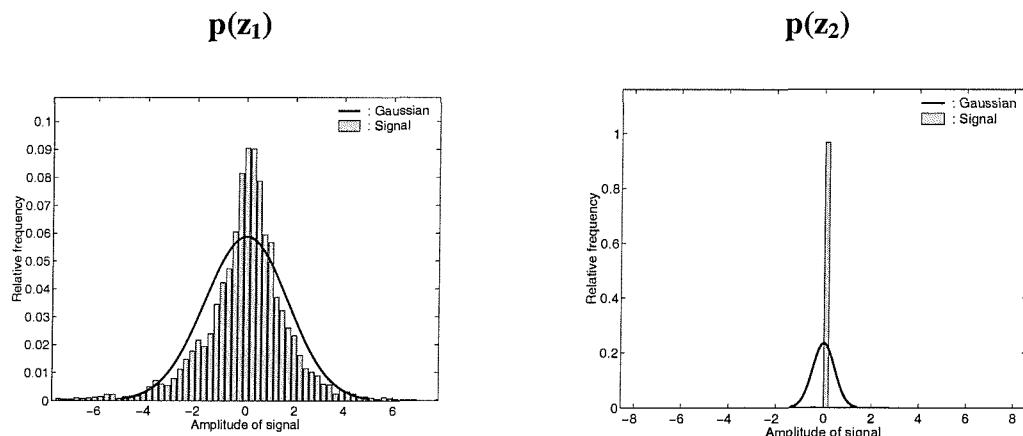
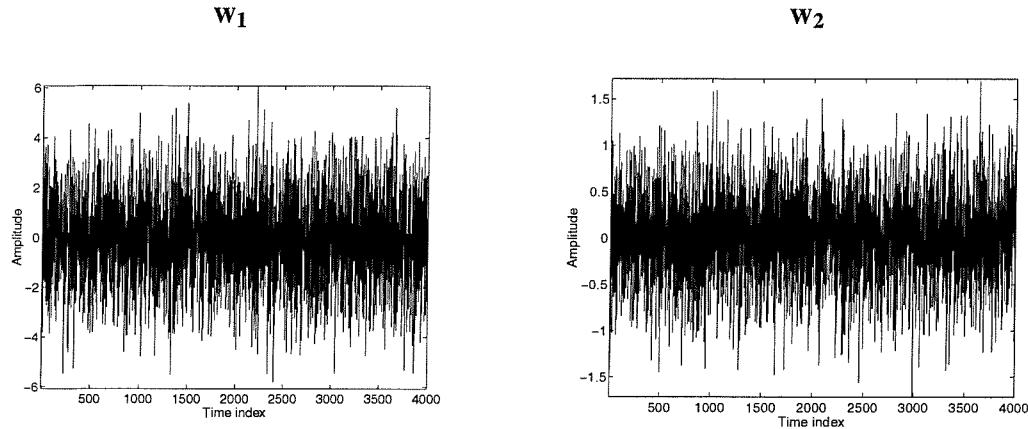


Figure 2.3.5 The output signals (noise free) and their statistical characteristics; (a): signals in time domain and (b): their probability distribution (bar graph) with equivalent Gaussian distribution (solid line).

(a) Time series of noise signals (Gaussian)



(b) Probability density of noise signals (Gaussian)

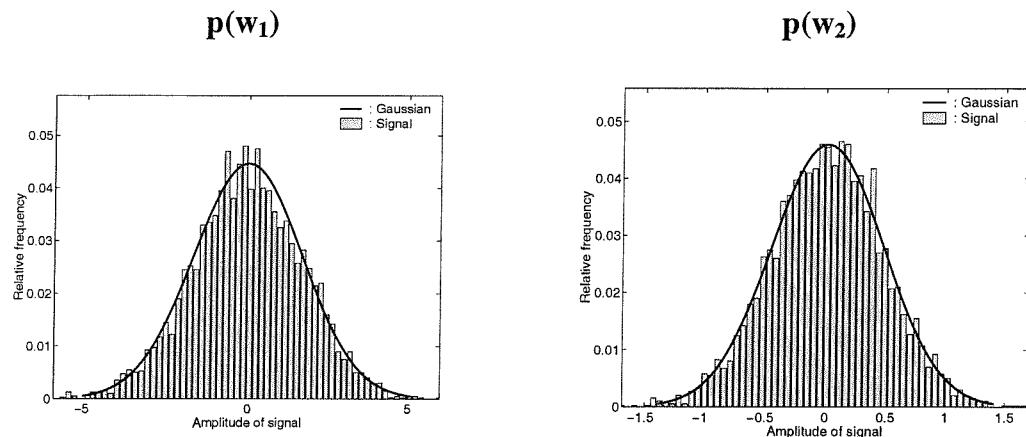
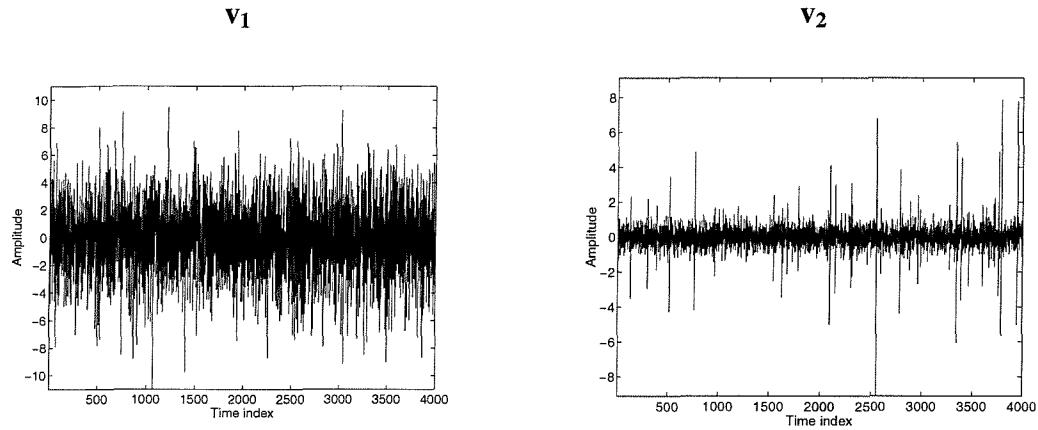


Figure 2.3.6 The Gaussian additive noise signals and their statistical characteristics; (a): signals in time domain and (b): their probability distribution (bar graph) with equivalent Gaussian distribution (solid line).

(a) Time series of measured signals



(b) Probability density of measured signals

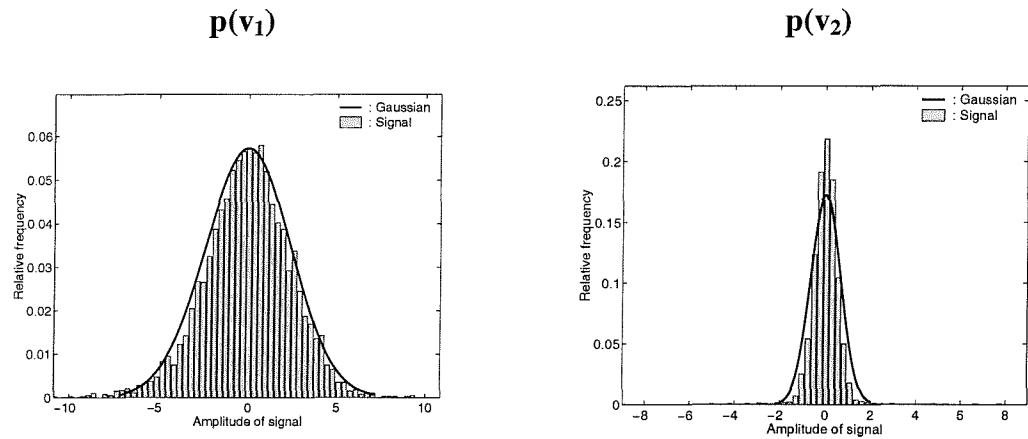


Figure 2.3.7 The measured signals and their statistical characteristics; (a): signals in time domain and (b): their probability distribution (bar graph) with equivalent Gaussian distribution (solid line).

Discussions of the simulation results;

For each of the two cases, 1000 realisation of the time histories were generated, and each time history contains 4000 points. For each realisation the cumulants etc were computed. As examples (i.e., the results of averaged values of 1000 realisations), Figure 2.3.8 depicts the results.

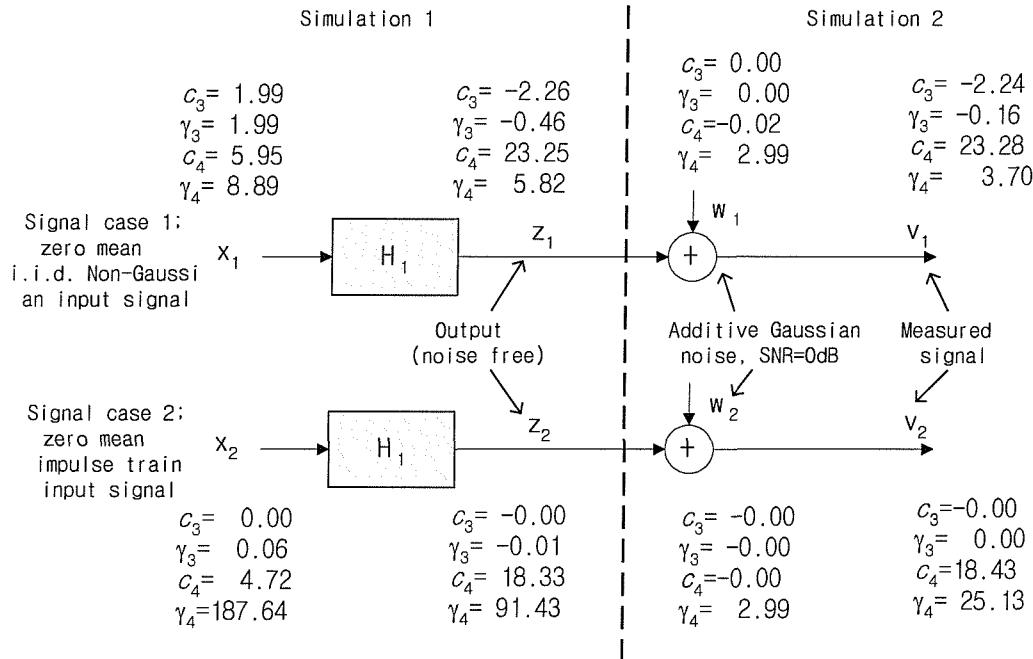


Figure 2.3.8 Statistical changes of signals along the processing. $c_k (k = 3, 4)$ denotes the k -th order cumulant of each signal and $\gamma_k (k = 3, 4)$ denotes the third and fourth order normalised cumulant ('skewness' and 'kurtosis', respectively) of each signal.

From the results of simulation, we focus on the changes in the higher order statistical values of each signal considering their closeness to the theoretical values and also the effect of data length on the estimation.

In order to do this, we will define two measures that approximately represent offset and variability of the estimated cumulant relative to the values obtained from equation (2.3.13). These are not the 'usual' definition of bias and variability but are modified for our purpose.

Offset error (OE);

This is expressed as the magnitude of the averaged absolute difference between the value given by the theory and the estimation as (the magnitude is chosen as we are indifferent to sign)

$$\begin{aligned} \text{OE} &= E\{|\hat{\Phi} - \phi|\} \\ &= |E\{\hat{\Phi}\} - E\{\phi\}| \end{aligned} \quad (2.3.30)$$

where $\hat{\Phi}$ and ϕ denote the calculated value (higher order statistical value) and theoretical estimation (equation (2.3.13)) for each realisation, respectively. Thus the bias error justified above can represent the average offset of the computationally calculated value from the theoretical value (for *i.i.d.* signals).

For example, for two realisations, if the calculated third order cumulant of signal z (output of system) is 3.02 and 3.12 in each realisation, respectively and their corresponding theoretical estimates from equation (2.3.13) are 2.71 and 2.69, then the bias error of this case is

$$\begin{aligned} \text{OE} &= \frac{1}{2} [|3.02 - 2.71| + |3.12 - 2.69|] \\ &= 0.37 \end{aligned}$$

Variability error (VE);

This is the mean of squared difference between the theoretical estimate and the computational calculation as

$$\text{VE} = E\left\{\left(\hat{\Phi} - \phi\right)^2\right\} \quad (2.3.31)$$

This error can represent the spread of the calculated value centred from the theoretical estimation. Using the same example above, the VE is

$$\begin{aligned} \text{VE} &= \frac{1}{2} \left\{ (3.02 - 2.71)^2 + (3.12 - 2.69)^2 \right\} \\ &= 0.1405 \end{aligned}$$

Simulation 1: Higher order cumulants (c_3, c_4) of input (x) and output (z) of the system;

Equation (2.3.13) derived for an *i.i.d.* input which shows that the response cumulant of order p is the product of the excitation cumulant of order p with the sum of the linear operator's unit-impulse response elements raised to the p th power has been tested.

The difference between the theoretically estimated value of output signal and calculated value is estimated consisting of the average offset (bias from the theoretical value) and degree of dispersion (variability of the differences). The result is shown in Table 2.3.1 for 4000 points time histories with 1000 realisations. It is noticeable that even though the non-Gaussian impulse train signal (x_2) may not belong to the category of *i.i.d* signals, the simulation result is consistent with the theoretical prediction (estimation).

Table 2.3.1 Cumulant and linear filtering (input-output relationship) ($N=4000$ samples, 1000 simulation)

Signals and order		Theoretical value, $c_{\text{order}}^{\text{input signal}} \times \sum_k h(k)^{\text{order}}$	Calculated value, $c_{\text{order}}^{\text{output signal}}$	Difference between theory and calculation	
				OE	VE
Signal 1	Third order	$1.99 \times (-1.13) = -2.26$	-2.26	0.00	0.03
	Fourth order	$5.95 \times 3.88 = 23.12$	23.25	0.12	4.59
Signal 2	Third order	$1.191e-5 \times (-1.13) = 0.00$	0.00	0.00	0.00
	Fourth order	$4.72 \times 3.88 = 18.32$	18.33	0.00	4.06

The effect of data length;

Since the use of higher order statistics generally requires more data samples

[Nikias and Petropulu, 1993] than the second order case, the effect of data length has been considered. This is shown empirically as follows;

Figure 2.3.9 demonstrates the decrease of VE along with the increase in the data length.

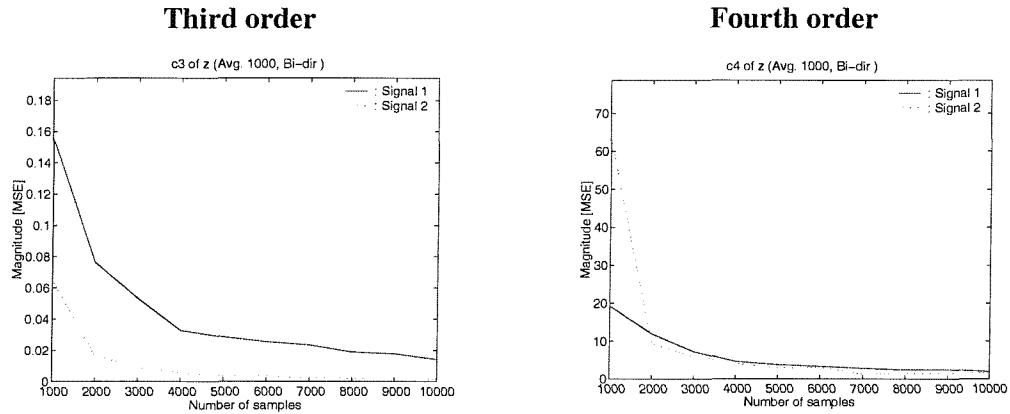


Figure 2.3.9 Effect of data length on the higher order cumulants of output signals (noise free signals).

Thus, the VE of each cumulant of each signal decreases monotonically as the data length increases, which reflects the fact that longer data lengths lead to more consistent results for higher order statistics.

Simulation 2: Higher order cumulants (c_3, c_4) of output signal (z) and observed signal (v) (noise effect);

The effect of noise corruption has been considered. One of the most important properties of the higher order cumulants is their ability to suppress Gaussian additive noise. Since the additive noise w (zero mean) is assumed Gaussian, the second order moment (autocorrelation, ACR), the fourth order moment, and the third order cumulant corresponding to the case of Figure 2.3.8 are described by the following relationship (indices of each signal are omitted):

$$\left. \begin{aligned} r_2^v &= r_2^z + r_2^w, & r_2^w &\neq 0 \\ c_p^v &= c_p^z + c_p^w, & c_p^w &= 0, \quad p = 3, 4 \end{aligned} \right\} \quad (2.3.32)$$

The graphical comparison of the autocorrelation and third order cumulant sequence of the Gaussian noise is shown in Figure 2.3.10, in which the zero-lag second order (correlation) value of Gaussian noise signal retains the value of its variance whereas that of third order vanishes.

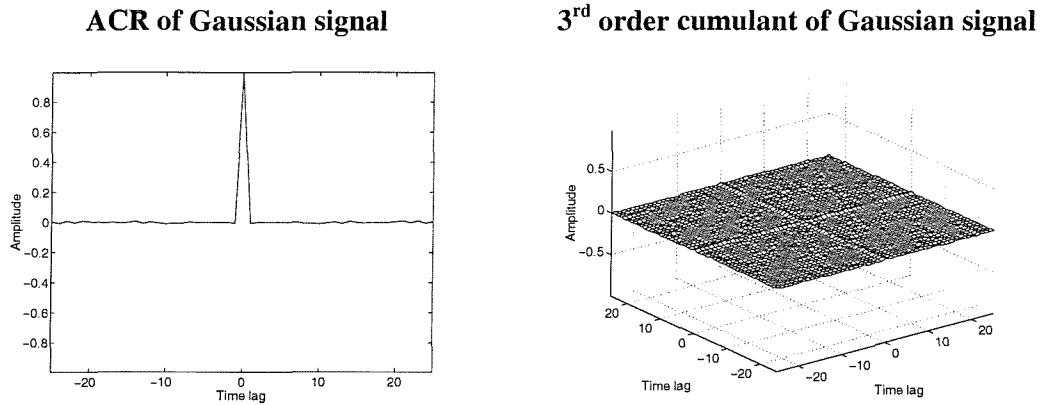


Figure 2.3.10 The second order moment and third order cumulant sequence of a white Gaussian noise

Based on the property of higher order cumulants of Gaussian noise, the theoretical relationship expressed in the equation (2.3.32) for the output of the system (noise free signal) and observed signal (noise corrupted) is tested. Similar to the simulation 1, the difference between the theoretical value of output signal and calculated value is estimated consisting of the average offset (bias from theoretical value) and degree of dispersion (VE of the differences). The result is shown in Table 2.3.2 for 4000 point time histories with 1000 realisations.

Table 2.3.2 Cumulant and Gaussian noise interference (N=4000 samples, 1000 simulation)

Signals and order		Theoretical value, $c_{\text{order}}^z + c_{\text{order}}^w = c_{\text{order}}^z$ $= c_{\text{order}}^v$	Calculated value, $c_{\text{order}}^{\text{observed signal}}$	Difference between theory and calculation	
				OE	VE
Signal 1	Third order	-2.26	-2.24	0.01	0.38
	Fourth order	23.25	23.28	0.02	24.62
Signal 2	Third order	0.00	0.00	0.00	0.02
	Fourth order	18.33	18.43	0.09	4.24

As in the case of Table 2.3.1 (input-output cumulant relationship), the theoretical relationship for the *i.i.d.* non-Gaussian signal \mathbf{x}_1 and Gaussian noise \mathbf{w}_1 is also applicable to the case of the non-Gaussian impulse train signal \mathbf{x}_2 and Gaussian noise \mathbf{w}_2 in Figure 2.3.8.

The effect of data length;

In Figure 2.3.11, the variability of the difference between the theoretical and calculated higher order statistical values along with the increase of the data length as shown.

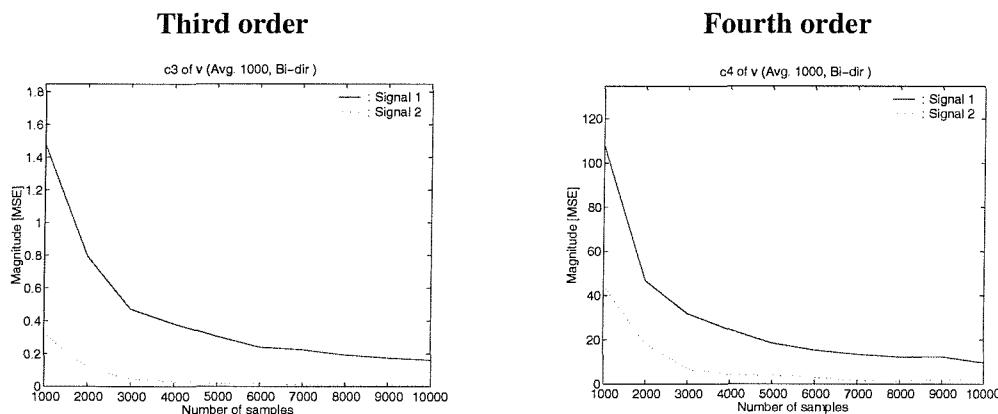


Figure 2.3.11 Effect of data length on the higher order cumulants of observed signals (noise corrupted signals).

Similar to the simulation 1, from Figure 2.3.11, the VE of each cumulant of each signal decreases monotonically as the data length increases confirming consistency.

Higher order normalised cumulants (γ_3, γ_4) relationships for input, output (noise free) and measured (noise corrupted) signals;

Unlike the case of higher order cumulants, the higher order normalised cumulants do not necessarily follow the same properties of cumulants of independent *i.i.d.* signals (refer to section 2.2). Instead, from extensive simulations, for the non-Gaussian *i.i.d.* input case, we have confirmed that there is a consistent trend among the normalised higher order cumulants (γ_3 or γ_4) of input, output (noise free signal, output of linear system alone) and measured signal (Gaussian noise corrupted). The magnitude of the normalised cumulants of each signal gradually decrease as the signal is first filtered and then suffers noise interference as shown in Figure 2.3.1. To be more specific, the linear filtering (convolution) and Gaussian noise addition drive the normalised higher order cumulants of the signal to smaller magnitudes, i.e., the inequality condition [Cadzow, 1996].

A numerical example for this trend is given in Figure 2.3.8 and graphically displayed for each realisation (1000 realisations with 4000 points samples) in the Figure 2.3.12.

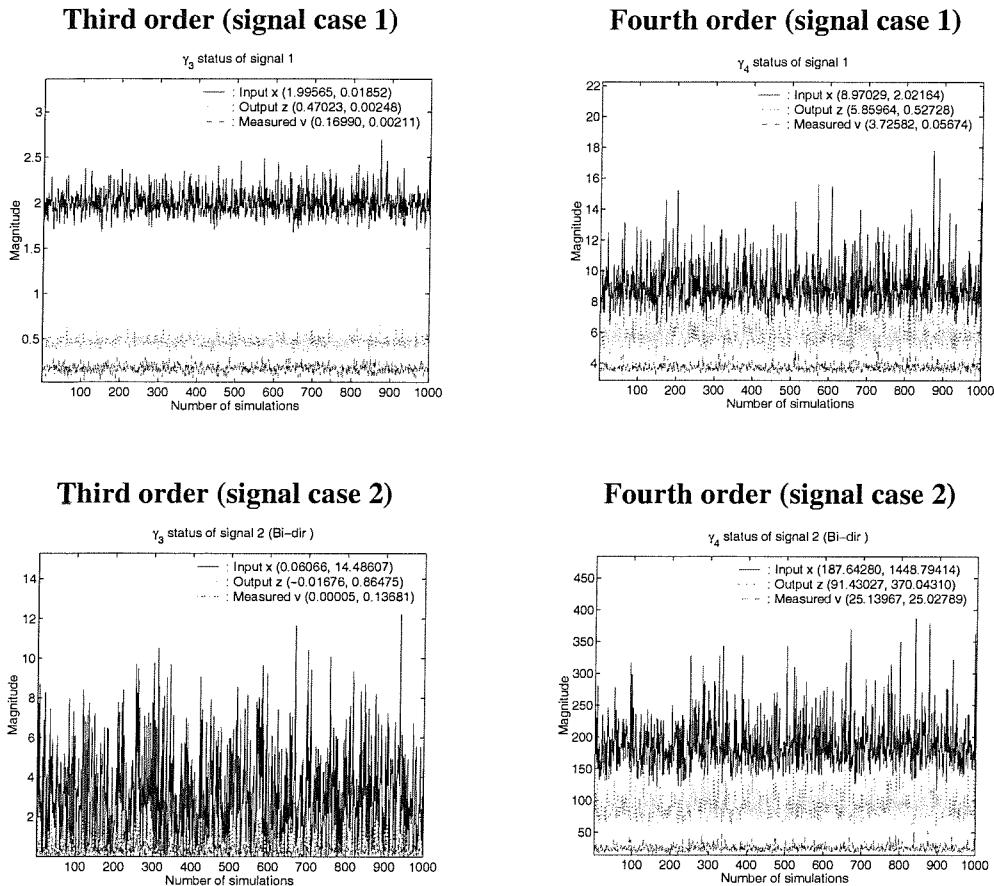


Figure 2.3.12 Change of normalised cumulants of each signal for subsequent simulations. Values in the () designate the average value and variance of each normalised cumulants.

As can be seen in the above figure, the magnitude of the third order normalised cumulant (skewness) and fourth order normalised cumulant (kurtosis) of the measured signals (v, signal case 1 and 2) are smaller than those of noise free signals (z) and those of z are also smaller than those of x.

Concerning the deconvolution problem, it can be said that any linear filtering of input signal (*i.i.d.* or impulsive) with or without noise interference makes the output signal possess small normalised cumulants. Thus, the deconvolution can be applied by incorporating a linear inverse system which can increase those values. This is the key to the application of the normalised cumulants in the blind deconvolution problem considered in this thesis (Chapter 3).

2.4 Concluding Remarks

The characteristics of higher order cumulants of signals and their properties through linear convolution have been explored. Based upon theory and simulation, this chapter provides evidence of the validity of the application of higher order statistics to the source signal reconstruction problem. A summary of results is as follows;

(a) The effect of convolution

1. The response cumulant of order p is the product of the excitation cumulant of order p with the sum of the linear operator's unit-impulse response elements raised to the p th power as $c_p^y = c_p^x \sum_k (h_k)^p$ for *i.i.d.* non-Gaussian signal. This

relationship has been demonstrated (by simulation) as being valid for an impulsive signal.

2. Any linear system tends to make the magnitude of the normalised cumulants of output sequence smaller than that of input's when the order of denominator is less than that of numerator (refer to the equation (2.3.22), in which $q < p$ with q even). This has been proved theoretically for *i.i.d.* non-Gaussian processes, but we have demonstrated further applicability in section 2.3.3. As a result, this is used as the basis for design of inversion (see (c) below).

(b) The effect of noise addition (Gaussian noise)

The higher order cumulants (order greater than 3) are blind to Gaussian signals, hence the observed signal's higher order cumulants are those of the non-Gaussian signal alone.

$$c_p^v = c_p^z + c_p^w = c_p^z, \quad p = 3, 4 \quad (c_p^w = 0).$$

(c) Use of higher-order cumulants

Selecting the value of the normalised cumulant which incorporates the second order ($q=2$) and higher (third or fourth) order ($p=3$ or 4) cumulant, the inequality condition (equation (2.3.22)) is established. This condition now provides the key motivation in the blind deconvolution problems and hence these *normalised cumulants will be extensively employed for the blind reconstruction* of source signal in subsequent chapters.

PART II Fundamental considerations of Higher Order Statistics (HOS)

Chapter 3

The deconvolution problem and Higher Order Statistics

3.1 Introduction

In the previous chapter, the basic characteristics of higher-order moments, cumulants, and their relation to a linear system's impulse response function were introduced. The subject considered here is the restoration of the unknown input signal which gives rise to the observed signal. If the distorting system is known it is a deconvolution problem for which many approaches exist. For example, in the context of communications, system identification, and inverse filtering relating to channel equalisation is treated in the book by Proakis [Proakis, 1995]. Related work in seismic signal processing is reported in [Peacock and Treitel, 1969; Wood and Treitel, 1975; Robinson and Treitel, 1978, 1980], predictive deconvolution for seismic signal processing is in classic work [Robinson, 1967] and homomorphic deconvolution and its application to speech processing is covered in [Oppenheim and Schafer, 1989].

The term ‘blind’ implies that knowledge of the system is not available, and except for homomorphic deconvolution (cepstral analysis) these methods are inapplicable.

However, in our situation, the term ‘blind’ is tempered by noting that we will make some prior assumptions about the signal structure. These assumptions include (i) the input signal (that is to be restored) is ‘significantly’ non-Gaussian and (ideally) independently identically distributed (*i.i.d.*). (ii) the additive noise has a Gaussian distribution and is independent of the input signal. These two conditions lead us to the principle that will be employed for blind signal recovery. It is known that an input signal (*i.i.d.*) which is convolved with the unknown LTI (Linear Time Invariant) system tends to become closer to a Gaussian distribution [Donoho, 1981]. So, to restore the input signal from the measured signal, one approach is to process the measured signals through a (constrained) filter so that the response is strongly ‘non-Gaussian’. In other words, for example, through monitoring the cumulant we may adjust the inverse filter coefficients to make the output signals of this filter be as “far” from Gaussianity as possible.

This chapter follows this approach using an *objective function* [Wiggins, 1978; Nandi, 1997] that is optimised to restore the input signal. This function is composed of higher order cumulants and moments. The deconvolution is achieved through the maximisation of the objective function of the filtered measured signal with respect to the corresponding coefficients of the inverse system.

The major aspects in this chapter are;

- The definition of deconvolution problem as a (generalised) Wiener optimisation problem.
- Maximisation of the objective function which is performed with respect to the coefficients of a linear inverse filter including MA, AR, and ARMA forms and their performances are compared.
- The justification of the preference of the inverse filter as an MA system.

3.2 Optimal deconvolution (The “classic” Wiener optimisation problem)

Given a time series $v(n)$ which is a filtered version of an input sequence $x(n)$ as shown in Figure 3.2.1,

$$v = h * x \quad (3.2.1)$$

where ‘*’ represents the convolution operation, the deconvolution problem is to find a filter $f(n)$ which recovers (as closely as possible) signal $x(n)$ from $v(n)$, i.e.,

$$\hat{x} = f * v \quad (3.2.2)$$

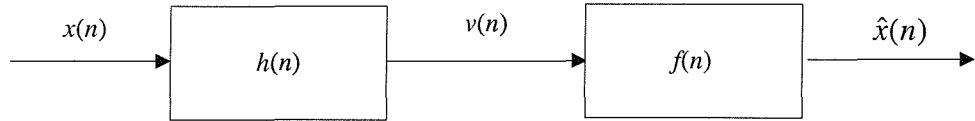


Figure 3.2.1 Deconvolution model

As a step toward addressing the inverse problem, when $h(n)$ is unknown, we summarise the minimum mean square error (MMSE) solution for the case when $h(n)$ is known and $f(n)$ is an inverse filter of length L . The optimal filter coefficients may be obtained on the basis of the least squared error between $\hat{x}(n)$ and a desired signal (in this case this is the input signal). Thus, the deconvolved signal is obtained from the estimated FIR inverse filter. A model is shown in Figure 3.2.2, from which the optimal deconvolution problem can be stated.

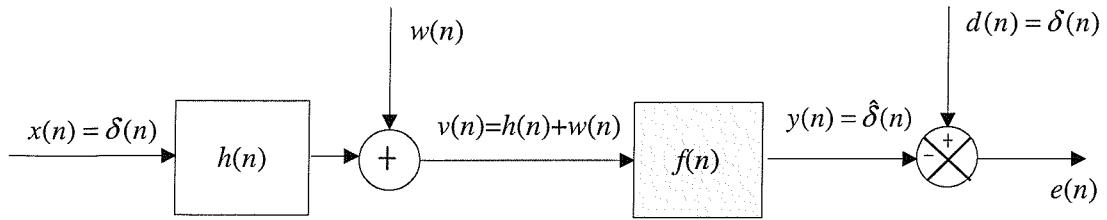


Figure 3.2.2 Deconvolution model for least-square method; $x(n) (= \delta(n))$ is the Original input signal to be recovered, $h(n)$ is the impulse response function of system, $w(n)$ is the noise interference, $v(n)$ is the measured signal, $f(n)$ is the impulse response of the inverse filter, $y(n) (= \hat{\delta}(n))$ is the restored signal (an estimate of the original signal), $d(n)$ is the desired signal ($= \delta(n)$), and $e(n)$ is the error signal

The sum of squared errors between the desired signal and inverse filter output signal is defined as a cost function,

$$J = \sum_{n=0}^{N-1} e(n)^2 \quad (3.2.3)$$

where

$$e(n) = d(n) - \sum_{k=0}^{L-1} f(k)v(n-k) \quad (3.2.4)$$

Minimising the cost function with respect to the filter coefficient $f(m)$, the necessary condition is

$$\frac{\partial J}{\partial f(m)} = 0, \quad m = 0, 1, \dots, L-1 \quad (3.2.5)$$

which leads to

$$2 \sum_{n=0}^{N-1} e(n) \cdot \frac{\partial e(n)}{\partial f(m)} = 0 \quad (3.2.6)$$

where

$$\frac{\partial e(n)}{\partial f(m)} = -v(n-m) \quad (3.2.7)$$

Thus, from equation (3.2.6) and equation (3.2.7) the well known orthogonality property between the error signal and the measured signal is

$$\sum_{n=0}^{N-1} e(n) \cdot v(n-m) = 0 \quad (3.2.8)$$

Substituting equation (3.2.4) into the equation (3.2.8) gives

$$\sum_{n=0}^{N-1} d(n) \cdot v(n-m) = \sum_{k=0}^{L-1} f(k) \sum_{n=0}^{N-1} v(n-k) v(n-m) \quad (3.2.9)$$

which can be expressed in matrix form for $n=0, 1, \dots, N-1$ and $m=0, 1, \dots, L-1$,

$$\begin{bmatrix} \sum_{n=0}^{N-1} v(n-0)v(n-0) & \sum_{n=0}^{N-1} v(n-1)v(n-0) & \sum_{n=0}^{N-1} v(n-L-1)v(n-0) \\ \sum_{n=0}^{N-1} v(n-0)v(n-1) & \sum_{n=0}^{N-1} v(n-1)v(n-1) & \sum_{n=0}^{N-1} v(n-L-1)v(n-1) \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \sum_{n=0}^{N-1} v(n-0)v(n-L-1) & \sum_{n=0}^{N-1} v(n-1)v(n-L-1) & \sum_{n=0}^{N-1} v(n-L-1)v(n-L-1) \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \\ \vdots \\ f(L-1) \end{bmatrix} = \begin{bmatrix} h(0) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (3.2.10)$$

i.e.,

$$\mathbf{R} \cdot \mathbf{f} = \mathbf{g} \quad (3.2.11)$$

where \mathbf{g} is a vector composed of a first element which is the system's first impulse response value and zeros elsewhere ($L \times 1$ column vector), \mathbf{R} is an auto correlation matrix of the measured signal ($L \times L$ symmetry matrix), and \mathbf{f} is the inverse filter ($L \times 1$ column vector).

The required deconvolution which restores the input signal in a least squares sense is

$$y(n) = \sum_{m=0}^{L-1} f_m v(n-m) \quad (3.2.12)$$

Note: Since the cost function is quadratic in the coefficients, the necessary condition (3.2.5) locates the global minimum.

3.3 Higher order deconvolution (Blind Deconvolution, BD)

Blind deconvolution is the problem of restoring an input signal from the measured (observed) signal alone. The measured signal is assumed to be the output of an unknown linear system possibly corrupted by noise. A further key assumption is that (i) the input signal to be restored is non-Gaussian and (ii) the corrupting signal is Gaussian and is independent of the input signal.

Starting from these assumptions, the restoration of the input signal can be addressed by utilising the higher order statistical properties of the signals. The method introduced here is based on the maximisation of higher order cumulants of the restored signal. This blind deconvolution (BD) process is illustrated in the figure below:

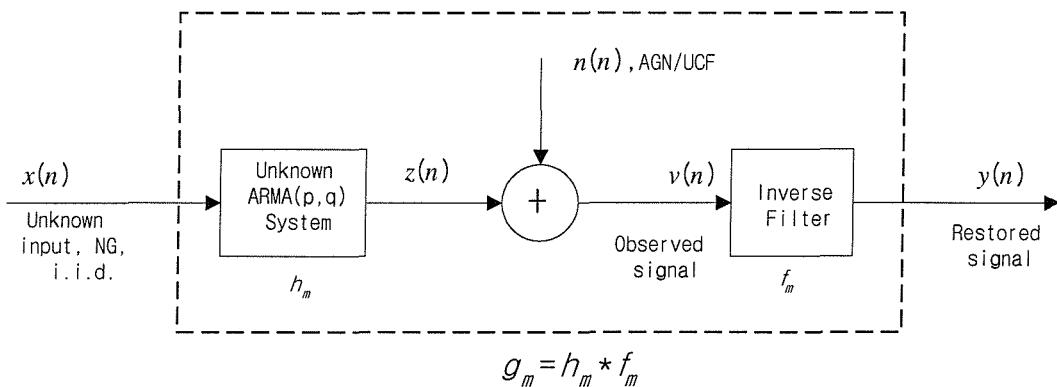


Figure 3.3.1 The process of convolution-deconvolution in blind source reconstruction problem; NG: Non-Gaussian; AGN: Additive Gaussian Noise; UCF: Unknown Covariance Function.

The inverse filter coefficients f_m (denoted \mathbf{f} in vector notation) is estimated from the maximisation of the higher order cumulant of the output $y(n)$. This results in a constrained non-linear optimisation problem which is solved numerically.

The deconvolution process based on higher order statistics which utilises order >2 cumulants as discussed in Chapter 2. Also the concept of partial order has been used and is summarised in Appendix A.

We describe the Blind Deconvolution process using two approaches, namely (i) constrained higher order cumulants and (ii) normalised higher order cumulants.

3.3.1 Blind Deconvolution process with a constraint

Unconstrained maximisation of a function can lead to unhelpful conclusions, i.e., unbounded filters. Constraints limit the behaviour of the inverse filter while optimising the appropriated objective function with respect to the filter coefficients. The chosen constraint is to keep the Frobenius norm of the FIR inverse filter coefficient unity.

$$\|\mathbf{f}\|_2 = 1 \quad (3.3.1)$$

which means the variance of the input and output of the inverse system are the same. A general form of objective function is expressed as

$$\begin{aligned} O_y(\mathbf{f}) &= \text{Cum}[\underbrace{y, y, \dots, y}_{r \text{ times}}] \\ &= C_r^y \underbrace{(0, 0, \dots, 0)}_{r \text{ times}} \end{aligned} \quad (3.3.2)$$

The integers r designate the order indices whose value is normally taken as the integer 3 or 4.

Incorporating the constraint function as

$$g(\mathbf{f}) = 1 - \frac{\frac{1}{N} \sum_{i=0}^{N-1} (y_i - m_y)^2}{\frac{1}{N} \sum_{i=0}^{N-1} (v_i - m_v)^2} \quad (3.3.3)$$

in which m_y and m_v is the mean of the signal y and v , respectively and combining the above two equations (3.3.6) and (3.3.3), the final cost function for the BD process takes the form

$$J(\mathbf{f}) = O_y(\mathbf{f}) + \lambda g(\mathbf{f}) \quad (3.3.4)$$

where λ is a Lagrange multiplier.

The necessary condition for a local optimum with respect to the inverse filter coefficient $f(k)$ is

$$\frac{\partial J(\mathbf{f})}{\partial f_m} = \frac{\partial O_y(\mathbf{f})}{\partial f_m} + \lambda \frac{\partial g(\mathbf{f})}{\partial f_m} = 0 \quad , \quad m = 0, 1, \dots, L-1 \quad (3.3.5)$$

As an example, taking $r=4$ and zero mean signals, we have a fourth order objective function as

$$\begin{aligned} O_y(\mathbf{f}) &= E\{y^4\} - 3E\{y^2\}^{4/2} \\ &= \frac{1}{N} \sum_{i=0}^{N-1} (y_i - m_y)^4 - 3 \left[\frac{1}{N} \sum_{i=0}^{N-1} (y_i - m_y)^2 \right]^{4/2} \end{aligned} \quad (3.3.6)$$

where $E\{\bullet\}$ denotes expectation, m_y is the mean value of y , N is the number of data points.

The first term on the right hand side of equation (3.3.5) becomes

$$\frac{\partial O_y(\mathbf{f})}{\partial f_m} = \frac{4}{N} \sum_{n=0}^{N-1} \{y(n)^3 \cdot v(n-m)\} - \frac{12}{N^2} \sum_{n=0}^{N-1} y(n)^2 \cdot \sum_{k=0}^{L-1} f(k) \sum_{n=0}^{N-1} \{v(n-k) \cdot v(n-m)\} \quad (3.3.7)$$

The second term can be expanded as

$$\lambda \frac{\partial g(\mathbf{f})}{\partial f_m} = -2\lambda \left\{ \sum_{n=0}^{N-1} v(n)^2 \right\}^{-1} \cdot \sum_{k=0}^{L-1} f(k) \sum_{n=0}^{N-1} \{v(n-k) \cdot v(n-m)\} \quad (3.3.8)$$

Combining (3.3.7) and (3.3.8) yields

$$\begin{aligned} \frac{\partial J(\mathbf{f})}{\partial f_m} &= \frac{4}{N} \sum_{n=0}^{N-1} \{y(n)^3 \cdot v(n-m)\} - \frac{12}{N^2} \sum_{n=0}^{N-1} y(n)^2 \cdot \sum_{k=0}^{L-1} f(k) \sum_{n=0}^{N-1} \{v(n-k) \cdot v(n-m)\} \\ &\quad - 2\lambda \left\{ \sum_{n=0}^{N-1} v(n)^2 \right\}^{-1} \cdot \sum_{k=0}^{L-1} f(k) \sum_{n=0}^{N-1} \{v(n-k) \cdot v(n-m)\} \\ &= 0 \end{aligned} \quad (3.3.9)$$

For notational simplicity, let us denote the auto/cross-correlation terms of equation (3.3.9) (for $m=0, 1, \dots, L-1$) as

$$\sum_{n=0}^{N-1} \{y(n)^3 \cdot v(n-m)\} \equiv \mathbf{g}_r$$

$$\sum_{n=0}^{N-1} \{v(n-k) \cdot v(n-m)\} \equiv \mathbf{R}_{vv}$$

Then, the inverse filter coefficient equation becomes

$$\mathbf{f} = \frac{2N \cdot \sum_{n=0}^{N-1} v(n)^2 \cdot \mathbf{R}_{vv}^{-1} \cdot \mathbf{g}_r}{6 \sum_{n=0}^{N-1} y(n)^2 \cdot \sum_{n=0}^{N-1} v(n)^2 + N^2 \cdot \lambda} \quad (3.3.10)$$

To determine the Lagrange multiplier λ , restructure the constraint condition (3.3.1) as

$$\sum_{m=0}^{L-1} f(m)^2 \sum_{n=0}^{N-1} v(n-m)^2 = \sum_{n=0}^{N-1} y(n)^2 \quad (3.3.11)$$

which means

$$\langle \mathbf{f}, \mathbf{f} \rangle = \mathbf{f}^T \cdot \mathbf{f} = 1 \quad (3.3.12)$$

i.e., $\langle \cdot \rangle$ denotes the inner product of the inverse filter coefficient vector.

Substituting (3.3.10) into (3.3.12) yields the Lagrange multiplier

$$\lambda = \frac{2}{N} \sum_{n=0}^{N-1} v(n)^2 \left[\mathbf{g}_r^T \cdot (\mathbf{R}_{vv}^{-1})^T \cdot \mathbf{R}_{vv}^{-1} \cdot \mathbf{g}_r \right]^{1/2} - \frac{6 \cdot \sum_{n=0}^{N-1} y(n)^2 \cdot \sum_{n=0}^{N-1} v(n)^2}{N^2} \quad (3.3.13)$$

Using the Lagrange multiplier and restructuring the inverse filter coefficient equation (3.3.10) leads to

$$\mathbf{f} = \frac{1}{\left[\mathbf{g}_r^T \cdot (\mathbf{R}_{vv}^{-1})^T \cdot \mathbf{R}_{vv}^{-1} \cdot \mathbf{g}_r \right]^{1/2}} \cdot \mathbf{R}_{vv}^{-1} \cdot \mathbf{g}_r \quad (3.3.14)$$

3.3.2 Blind Deconvolution process with normalisation

With reference to Figure 3.3.1, we construct the inverse filters using a normalised cumulant (objective function). This objective function takes the form of the r -th order cumulant of the output signal $y(n)$ divided by the s -th order (s -th moment) of $y(n)$ raised to power r/s

$$O_y(\mathbf{f}) = \frac{E[(y - m_y)^r]}{\{E[(y - m_y)^s]\}^{r/s}} \simeq \frac{\frac{1}{N} \sum_{i=0}^{N-1} (y_i - m_y)^r}{\left[\frac{1}{N} \sum_{i=0}^{N-1} (y_i - m_y)^s \right]^{r/s}} \quad (3.3.15)$$

where m_y is the mean of y . Similar to the constrained case, the normalising order s is given by 2 which corresponds to the variance of y (σ_y^2). Commonly, the integer r is either 3 or 4.

The maximisation of the objective function is achieved through the differentiation of this objective function with respect to the inverse filter coefficients and equating to zero.

$$\partial O_y(\mathbf{f}) / \partial f_m = 0 \quad (3.3.16)$$

The above equation yields inverse filter coefficients as shown in the following equation (3.3.17) which yields multiple solutions.

The maximisation yields an equation relating the inverse filter coefficient vector, the autocorrelation of the measured signal and the higher order cross zero-lag cumulant.

$$\begin{bmatrix} \sum_{n=0}^{N-1} v(n)v(n) & \sum_{n=0}^{N-1} v(n)v(n-1) & \cdots & \sum_{n=1}^{N-1} v(n)v(n-L+1) \\ \sum_{n=0}^{N-1} v(n-1)v(n) & \sum_{n=0}^{N-1} v(n-1)v(n-1) & \cdots & \sum_{n=0}^{N-1} v(n-1)v(n-L+1) \\ \vdots & \vdots & & \vdots \\ \sum_{n=0}^{N-1} v(n-m)v(n) & \sum_{n=0}^{N-1} v(n-m)v(n-1) & \cdots & \sum_{n=1}^{N-1} v(n-m)v(n-L+1) \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_{L-1} \end{bmatrix} = \begin{bmatrix} A \cdot \sum_{n=0}^{N-1} y^{r-1}(n)v(n) \\ A \cdot \sum_{n=0}^{N-1} y^{r-1}(n)v(n-1) \\ \vdots \\ A \cdot \sum_{n=0}^{N-1} y^{r-1}(k)v(k-m) \end{bmatrix} \quad (3.3.17)$$

where $m=L$ and the term $A = \sum_{n=0}^{N-1} y^2(n) / \sum_{n=0}^{N-1} y^r(n)$. For the constrained case (where λ can be calculated from equation (3.3.13)), we select the multiplier

$$A = \frac{2N \cdot \sum_{n=0}^{N-1} v(n)^2}{\frac{3r}{2} \sum_{n=0}^{N-1} y(n)^2 \cdot \sum_{n=0}^{N-1} v(n)^2 + N^2 \cdot \lambda} . \text{ The equation (3.3.17) can be written}$$

equivalently as (details can be found in Appendix B)

$$\mathbf{R}_{vv} \cdot \mathbf{f} = \mathbf{g} \quad (3.3.18)$$

where \mathbf{R}_{vv} denotes the symmetry $L \times L$ autocorrelation matrix of the observed signal, \mathbf{f} is the $L \times 1$ inverse filter coefficient vector, and \mathbf{g} is the $L \times 1$ cross-correlation vector between the observed signal and the output of the inverse filter at each iteration. Note that equation (3.3.18) takes a form similar to equation (3.2.11) for the ‘conventional’ deconvolution problem.

Using either the constrained or normalised objective function maximisation process, the output signal $y(n)$ can yield the input signal restoration through the convolution of the measured signal $v(n)$ and the inverse filter \mathbf{f} with length L

$$y(n) = \sum_{k=0}^{L-1} f(k)v(n-k) \quad (3.3.19)$$

3.3.3 Example of signal restoration using the constrained and normalised objective functions

With a given impulse signal and a simple system, an example of the restored signals the constrained and normalised objective function maximisation are compared:

Input signal

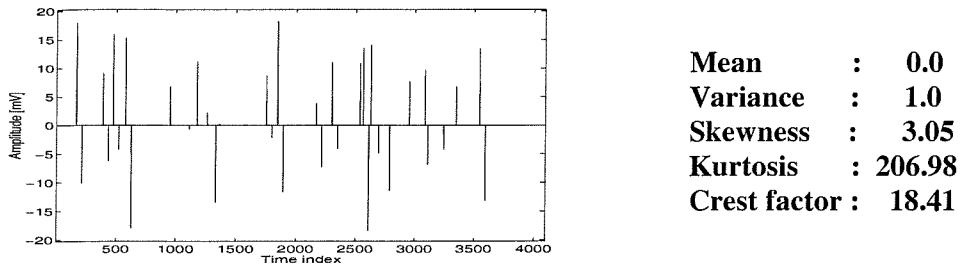


Figure 3.3.2 Input signal used in simulation for restoration of impulse signal via two different objective function maximisation methods.

Unknown system

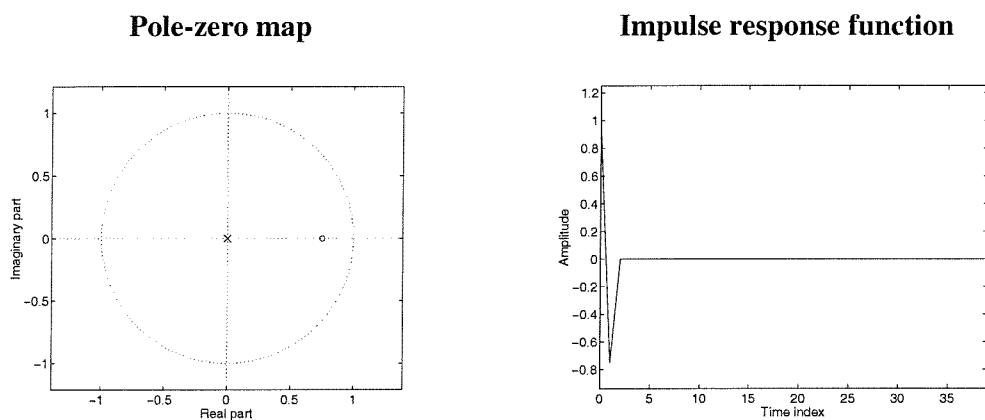


Figure 3.3.3 Unknown system's pole-zero map and its impulse response function

Observed signal (SNR= 10 dB)

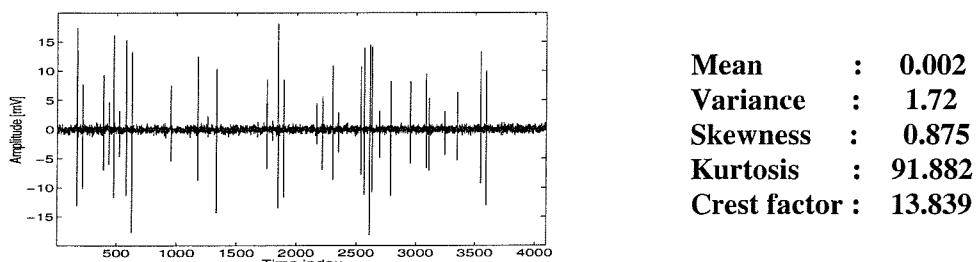


Figure 3.3.4 Observed signal with Gaussian noise corrupted (SNR = 10 dB)

Reconstruction by the constrained objective function (fourth order MA(4) inverse filter)

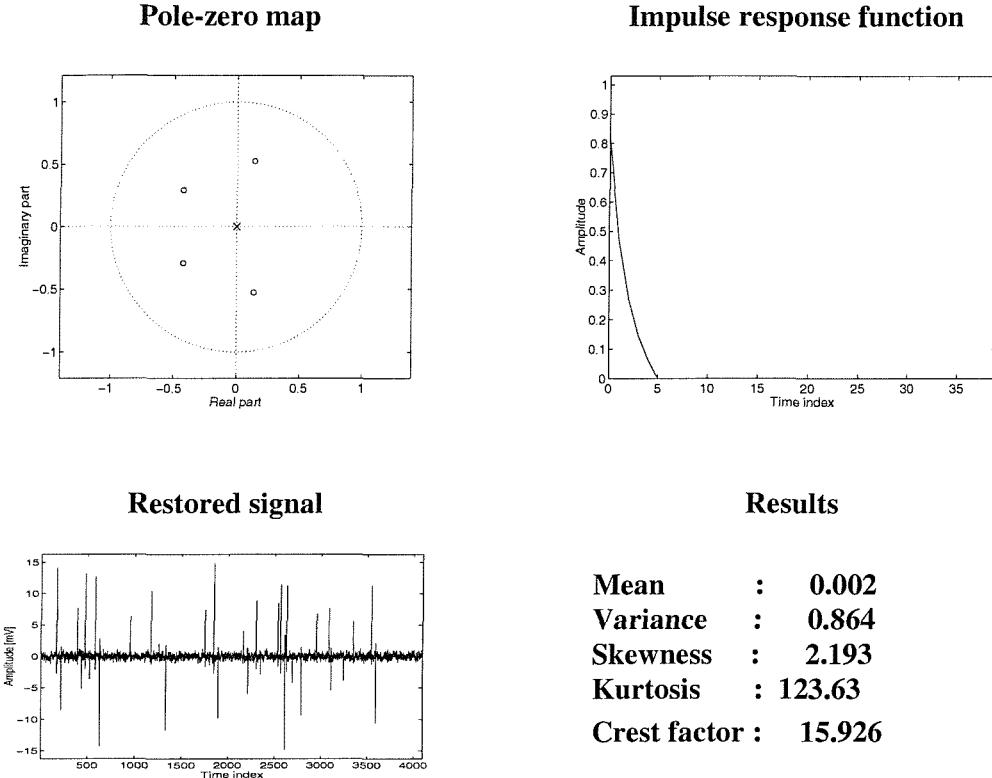


Figure 3.3.5 Inverse system and restored signal from the constrained objective function maximisation

Note: Comparing the observed signal (in Figure 3.3.4) and restored signal shown in Figure 3.3.5, we can see that the true shapes of the impulses of the input signal (shown in Figure 3.3.2) are successfully recovered (i.e., see the smearing direction of the impulses in the observed signal are correctly aligned).

Reconstruction by the normalised objective function (fourth order MA(4) inverse filter)

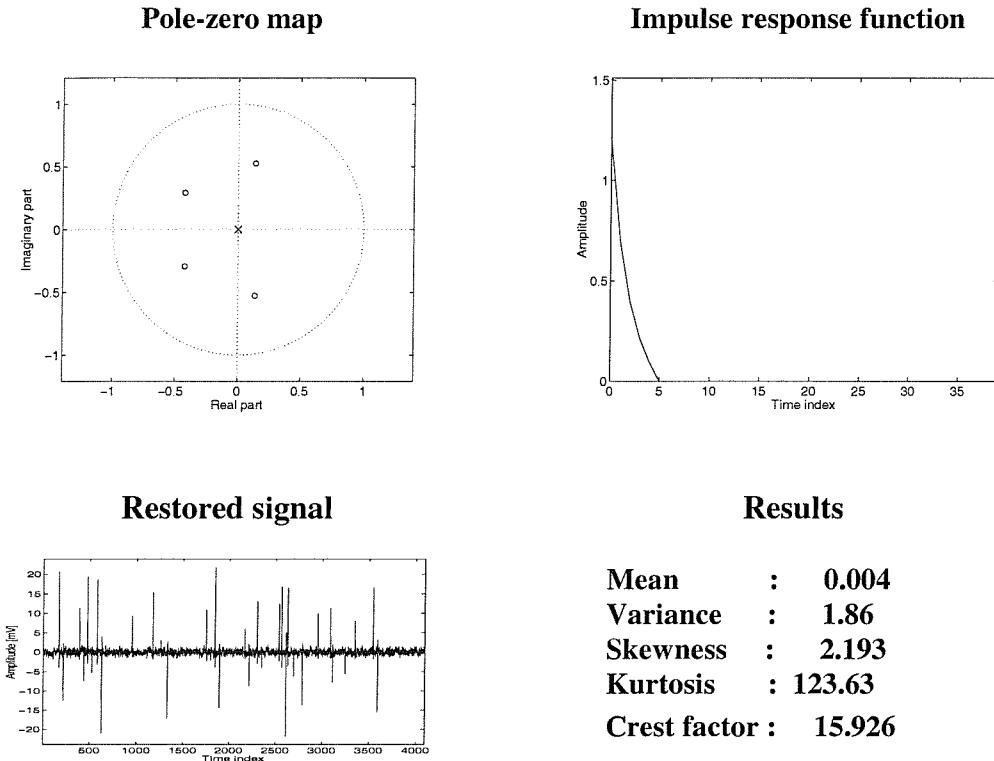


Figure 3.3.6 Inverse system and restored signal from normalised objective function maximisation

Note: Similar to the constrained inverse filtering case, the restored signal is closer to the input signal.

Detailed comparisons of results from two different inverse systems are given in the following.

Discussion

Whilst the restored signals from two different inverse filters may not appear to be very different (from Figure 3.3.5 and Figure 3.3.6) for this example, we compare their performance on the basis of the statistics. The inverse filter coefficient and

the variance of the restored signals from constrained and normalised objective function are summarised in Table 3.3.1.

Table 3.3.1 Comparison of the inverse filter and restored signal from two different optimisation methods

	Constrained case	Normalised case	Remarks
Frobenius norm of the inverse filter	$\ \mathbf{f}\ _2 = 1$	$\ \mathbf{f}\ _2 = 1.47$	$L=4$, FIR filter
Variance of the restored signal	0.86(1.72)	1.86(1.72)	values in () are the variance of the measured signal

From these results, it is clear that signal reconstruction from both types of objective function give similar (satisfactory) results. In fact, the normalised higher order cumulant gives identical results (to the order of accuracy of the computation) as for the constrained cumulant when the signal has unit variance and zero mean. In fact, can be seen in the table, the inverse filter coefficients calculated either from equation (3.3.14) or equation (3.3.18) only differ in the magnitude of these values. Specifically:

The FIR inverse filter coefficients for the constrained objective function is

$$\mathbf{f}_{\text{constraint}} = [0.824 \ 0.472 \ 0.270 \ 0.146 \ 0.065].$$

and the FIR inverse filter coefficients for the normalised objective function is

$$\mathbf{f}_{\text{normalised}} = [1.210 \ 0.693 \ 0.397 \ 0.215 \ 0.095]$$

which is $\mathbf{f}_{\text{constraint}}$ multiplied by the norm ratio (=1.47).

3.3.4 IIR deconvolution operator

As noted in the previous section, although the nonrecursive filter ensures a satisfactory result for many problems, it is an inappropriate choice when the convolving operator's transfer function $H(z)$ is rational (for AR or ARMA processes) and in particular has zeros that lie close to the unit-circle. In this case, the desired deconvolving operator also has a rational transfer function $F(z)=H(z)^{-1}$, which has poles located close to the unit-circle, thus giving long impulse responses. To solve this problem with a nonrecursive deconvolving operator, the inverse system's order would need to be long to give an adequate approximation. This leads to a large computational burden. Hence, in this section, we introduce a recursive deconvolving operator to alleviate this deficiency. The relationship between the output $\{y(n)\}$ of an ARMA inverse system (order p, q) with the observed data $\{v(n)\}$, $n=0,1,\dots,N-1$ can be expressed as the linear constant-coefficient difference equation

$$\sum_{k=0}^{p-1} a_k y(n-k) = \sum_{k=0}^{q-1} b_k v(n-k) \quad (3.3.20)$$

Similar to the case of MA operation, the normalised cumulant of order (r,s) and $r>s$, (integers) may be used [Nandi et al, 1997]

$$O_y(r,s) = \frac{1}{N} \sum_{k=0}^{N-1} y^r(k) / \left[\frac{1}{N} \sum_{k=0}^{N-1} y^s(k) \right]^{r/s} \quad (3.3.21)$$

By optimising this objective function with respect to the filter coefficient b_i ($i=0,1,\dots,q-1$) and a_j ($j=0,1,\dots,p-1$), the necessary condition is

$$\begin{aligned} \partial O_y(r,s) / \partial b_i &= 0 \text{ and} \\ \partial O_y(r,s) / \partial a_j &= 0 \end{aligned} \quad (3.3.22)$$

The matrix formulation of ARMA inverse filter can be written [Cadzow, 1996] as (detailed expressions are found in the Appendix B)

$$\begin{bmatrix}
 \sum_{k=0}^{N-1} s_a(k-0)y(k-0) & \sum_{k=0}^{N-1} s_a(k-0)y(k-1) & \cdots & \sum_{k=0}^{N-1} s_a(k-0)y(k-p+1) \\
 \sum_{k=0}^{N-1} s_a(k-1)y(k-0) & \sum_{k=0}^{N-1} s_a(k-1)y(k-1) & \cdots & \sum_{k=0}^{N-1} s_a(k-1)y(k-p+1) \\
 \vdots & \vdots & \ddots & \vdots \\
 \sum_{k=0}^{N-1} s_a(k-m)y(k-0) & \sum_{k=0}^{N-1} s_a(k-m)y(k-1) & \cdots & \sum_{k=0}^{N-1} s_a(k-m)y(k-p+1)
 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \\
 \\
 \begin{bmatrix}
 \sum_{k=0}^{N-1} s_a(k-0)v(k) & \sum_{k=0}^{N-1} s_a(k-0)v(k-1) & \cdots & \sum_{k=0}^{N-1} s_a(k-0)v(k-q+1) \\
 \sum_{k=0}^{N-1} s_a(k-1)v(k) & \sum_{k=0}^{N-1} s_a(k-1)v(k-1) & \cdots & \sum_{k=0}^{N-1} s_a(k-1)v(k-q+1) \\
 \vdots & \vdots & \ddots & \vdots \\
 \sum_{k=0}^{N-1} s_a(k-m)v(k) & \sum_{k=0}^{N-1} s_a(k-m)v(k-1) & \cdots & \sum_{k=0}^{N-1} s_a(k-m)v(k-q+1)
 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{q-1} \end{bmatrix} + \begin{bmatrix}
 A \cdot \sum_{k=0}^{N-1} y^{r-1}(k)s_a(k-0) \\
 A \cdot \sum_{k=0}^{N-1} y^{r-1}(k)s_a(k-1) \\
 \vdots \\
 A \cdot \sum_{k=0}^{N-1} y^{r-1}(k)s_a(k-m)
 \end{bmatrix} \quad (3.3.23)$$

where $m=p-1$ and $A = \sum_{k=0}^{N-1} y^s(k) / \sum_{k=0}^{N-1} y^r(k)$.

Or, in a compact form,

$$\mathbf{R}_{s_a y} \cdot \mathbf{a} = \mathbf{R}_{s_a v} \cdot \mathbf{b} + \mathbf{r}_{y s_a} \quad (3.3.24)$$

3.3.5 Comparison of MA, AR and ARMA deconvolution operators

The blind deconvolution operators are compared to measure their efficiencies depending on the system employed. The MA, AR or ARMA blind deconvolution operators described in the previous section will now be demonstrated. Each has different characteristics in recovering the source signal. The simulation model is shown below.

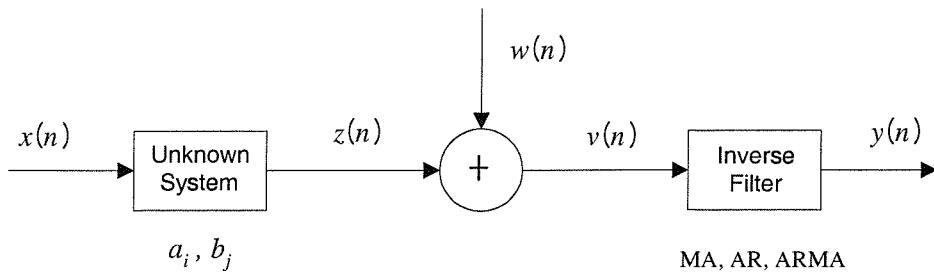


Figure 3.3.7 The signals and deconvolution model

This section demonstrates the validity of the deconvolution procedure with reference to the objective function and the coefficients of three different inverse filters. Using three different inverse systems such as MA(4), AR(2) and ARMA(1,1), the observed signal (which is shown in Figure 3.3.4) is deconvolved with these systems inverse systems.

BD process using an MA(4) inverse filter :

In Figure 3.3.8, the shape and contours of the objective function (kurtosis) are shown. To plot this figure, a mesh of objective function values are calculated by varying the first two coefficients of \mathbf{f}_b in 60 steps of 0.1 and keeping the other MA coefficients (f_{b_2} , f_{b_3} , f_{b_4}) fixed. Thus, the inverse filter coefficient becomes;

$$\mathbf{f}_b = \begin{bmatrix} f_{b_0} & : \text{varying} \\ f_{b_1} & : \text{varying} \\ f_{b_2} & : \text{fixed} \\ f_{b_3} & : \text{fixed} \\ f_{b_4} & : \text{fixed} \end{bmatrix}, \quad \mathbf{f}_a = 1$$

$f_{b_0_opt}$ and $f_{b_1_opt}$ are the first and second MA filter coefficient which are optimally calculated and their co-ordinates are marked by 'opt' while a pair of coefficients denoted by $f_{b_0_max}$ and $f_{b_1_max}$ that gives the maximum value of the objective function among f_{b_0} and f_{b_1} is marked by 'max' in the figure.

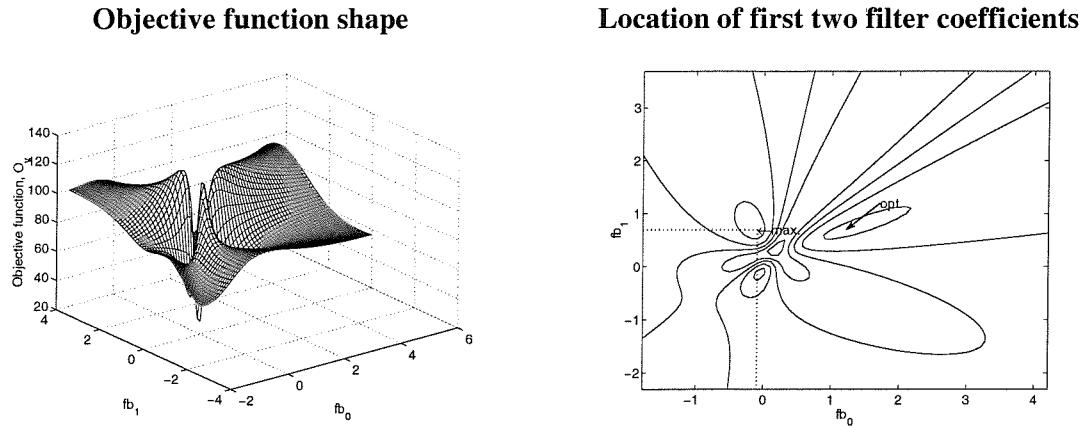


Figure 3.3.8 The Objective function of the 4th order statistics for MA(4) inverse filter. Dotted line point: b_0 and b_1 value which gives maximum Objective value, arrow point: calculated f_{b0} and f_{b1} value.

Thus, the inverse filter from the ‘opt’ point consists of two filter coefficients ($f_{b_0_opt}$, $f_{b_1_opt}$) at the ‘opt’ point and the three other coefficients are (f_{b_2} , f_{b_3} , f_{b_4}) so that

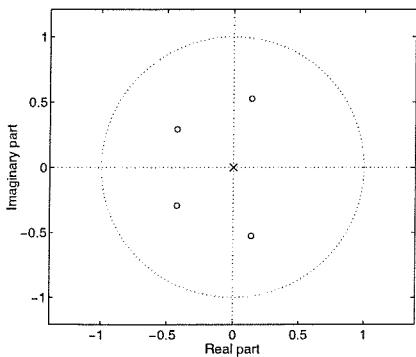
$$\mathbf{f}_{b_opt} = [f_{b_0_opt} \quad f_{b_1_opt} \quad f_{b_2} \quad f_{b_3} \quad f_{b_4}]^T, \quad \mathbf{f}_a = 1$$

and the inverse filter from ‘max’ point becomes

$$\mathbf{f}_{b_max} = [f_{b_0_max} \quad f_{b_1_max} \quad f_{b_2} \quad f_{b_3} \quad f_{b_4}]^T, \quad \mathbf{f}_a = 1$$

The pole-zero maps are shown in the figure below;

Pole-zero map of ‘opt.’ point



Pole-zero map of ‘max.’ point

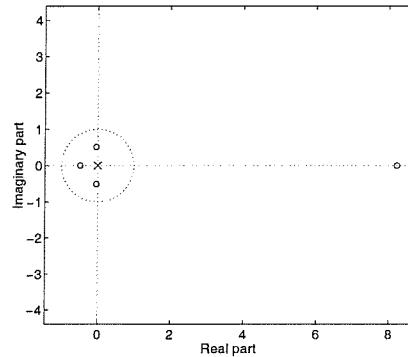
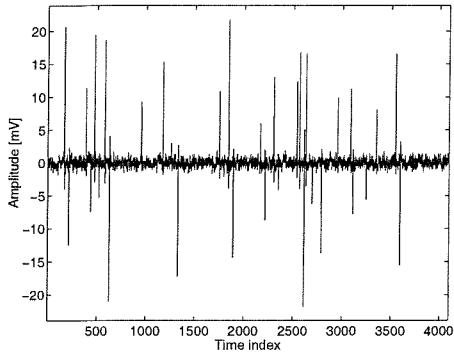


Figure 3.3.9 The MA(4) inverse system obtained by 4th order statistics.

We can obtain two different restored signals by convolving these two inverse systems with the measured signal $v(n)$, which are plotted in the figure below;

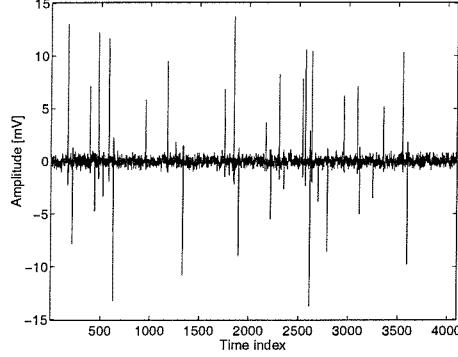
Restored signal by ‘opt.’ point filter



Skewness : 2.193

Kurtosis : 123.643

Restored signal by ‘max.’ point filter



Skewness : 2.226

Kurtosis : 128.319

Figure 3.3.10 The restored signals from 4th order MA(4) inverse filter.

It is noticeable from Figure 3.3.8 that there can be more than one maximum in the blind deconvolution problem when using the MA inverse filter. Thus, it is evident that the blind deconvolution based on the higher-order statistics could end up at a local maximum (see that the intersection of the dotted line does not coincide with the arrow point). The inverse systems shown in Figure 3.3.9 differ in the phase (e.g., inverse system from the ‘opt’ point becomes a ‘minimum’ phase system whereas the inverse system from the ‘max’ point becomes ‘non-minimum’ phase system). For the MA(4) inverse system, the restored signals from the two inverse filters including different coefficients, nevertheless, produces similar satisfactory impacting signal reconstruction as shown in Figure 3.3.10.

BD process using AR(2) inverse filter :

Like the MA(4) inverse system case, the shape and contours of the objective function (kurtosis) are plotted in Figure 3.3.11. Varying the last two coefficients of \mathbf{f}_a in 0.1 steps and keeping the first AR coefficients (f_{a_0}) fixed as 1, the inverse filter coefficients become;

$$\mathbf{f}_a = \begin{bmatrix} f_{a_0} & : \text{fixed} \\ f_{a_1} & : \text{varying} \\ f_{a_2} & : \text{varying} \end{bmatrix}, \quad \mathbf{f}_b = 1$$

$f_{a_1_opt}$ and $f_{a_2_opt}$ are the second and third AR filter coefficient which are optimally calculated and their co-ordinates are marked by 'opt' while a pair of coefficients denoted by $f_{a_1_max}$ and $f_{a_2_max}$ that give the maximum value of the objective function are marked by 'max' in the figure.

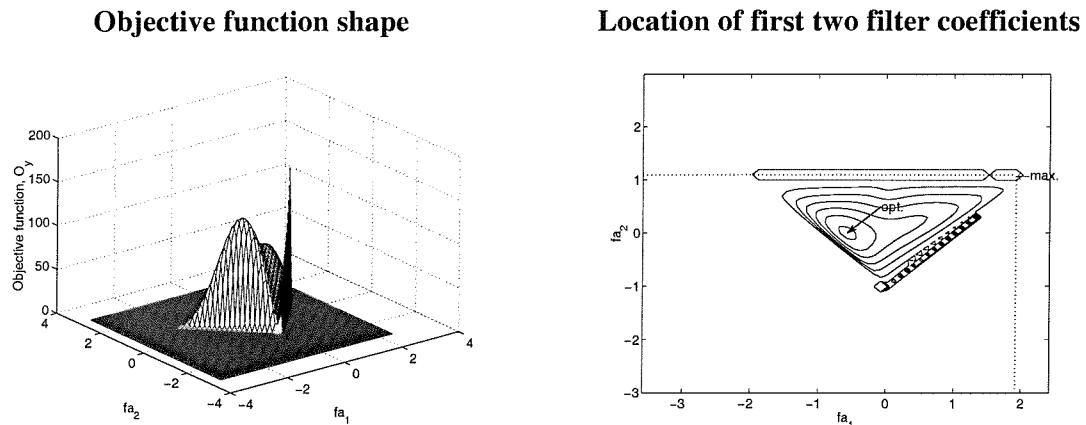


Figure 3.3.11 The Objective function of the 4th order AR(2) inverse filter ($f_{a_0} = 1$).
 Dotted line point: f_{a_1} and f_{a_2} value which gives maximum Objective value, arrow point: calculated f_{a_1} and f_{a_2} value.

Thus, the inverse filter from 'opt' point consists of two filter coefficients ($f_{a_1_opt}$, $f_{a_2_opt}$) at the 'opt' point and the first AR coefficient are ($f_{a_0} = 1$), so

$$\mathbf{f}_{a_opt} = \begin{bmatrix} f_{a_0} & f_{a_1_opt} & f_{a_2_opt} \end{bmatrix}^T, \quad \mathbf{f}_b = 1$$

and the inverse filter from ‘max’ point becomes

$$\mathbf{f}_{a_{-max}} = \begin{bmatrix} f_{a_0} & f_{a_1_{-max}} & f_{a_2_{-max}} \end{bmatrix}^T, \quad \mathbf{f}_b = 1$$

The pole-zero maps are shown in the figure below;

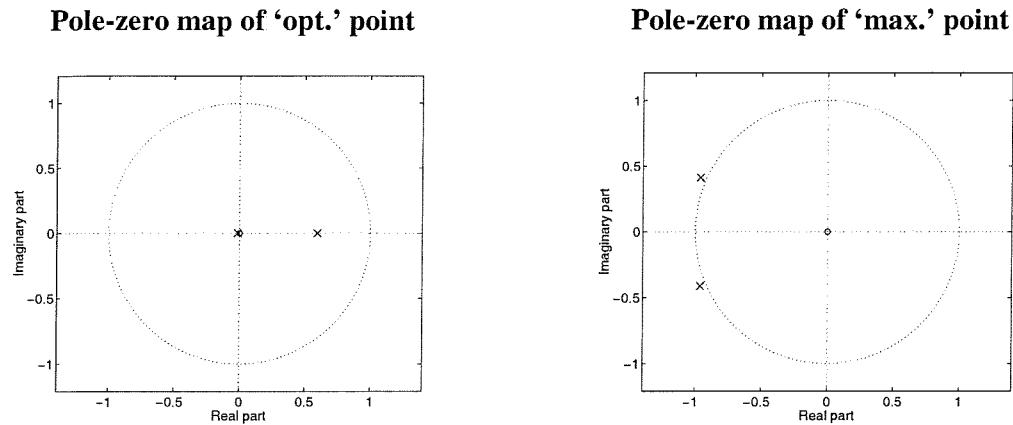


Figure 3.3.12 The AR(2) inverse system obtained by 4th order statistics.

As for the MA(4) case, we obtain two different restored signals which are plotted in Figure 3.3.13.

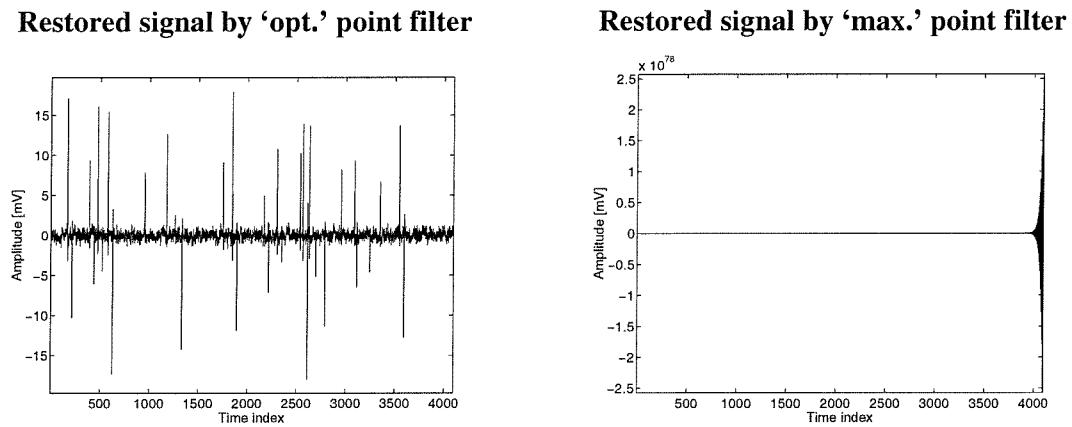


Figure 3.3.13 The restored signals from 4th order AR(2) inverse filter.

As for the MA inverse filter case, the AR filter coefficients that give the maximum point of the objective function value are different from the calculated coefficients. In Figure 3.3.11, the point of the maximum objective function's value (intersection of the dotted line) does not match the calculated maximum (arrowed). The AR inverse filter coefficients which give massive kurtosis values are $f_{a0}=1.0$, $f_{a1}=1.9206$, and $f_{a2}=1.0916$ (filter coefficients denoted 'max.'), whereas the calculated coefficients are $f_{a0}=1.0$, $f_{a1}=-0.5794$, and $f_{a2}=-0.0083$ (filter coefficients at 'opt.' point) giving the kurtosis as 124.504. As expected, as shown in Figure 3.3.13, the result of signal restoration from this 'max.' positioned filter coefficients is totally unsuccessful. On the other hand, the restored signal from that 'opt.' point filter coefficients is slightly better than the case of MA(4) inverse filter (since the unknown system shown in Figure 3.3.3 is an MA system, the inverse system would preferably select an AR system). It is clear that this inverse filter also possesses local maxima.

BD process using ARMA(1,1) inverse filter :

The three dimensional shape and contour of the objective function (kurtosis) is plotted in Figure 3.3.14 varying the last coefficient of \mathbf{f}_a and \mathbf{f}_b in 60 steps of 0.1 and keeping the first AR coefficient (f_{a_0}) fixed as 1 while the first MA coefficient (f_{b_0}) is also kept constant, the inverse filter coefficients are

$$\mathbf{f}_a = \begin{bmatrix} f_{a_0} & : \text{fixed} \\ f_{a_1} & : \text{varying} \end{bmatrix}, \quad \mathbf{f}_b = \begin{bmatrix} f_{b_0} & : \text{fixed} \\ f_{b_1} & : \text{varying} \end{bmatrix}$$

$f_{a_1_opt}$ and $f_{b_1_opt}$ are the second coefficient of AR and MA, respectively which are optimally calculated and their co-ordinates are marked by ‘opt’ while a pair of coefficients denoted by $f_{a_1_max}$ and $f_{b_1_max}$ that give the maximum value of the objective function marked by ‘max’ in the figure.

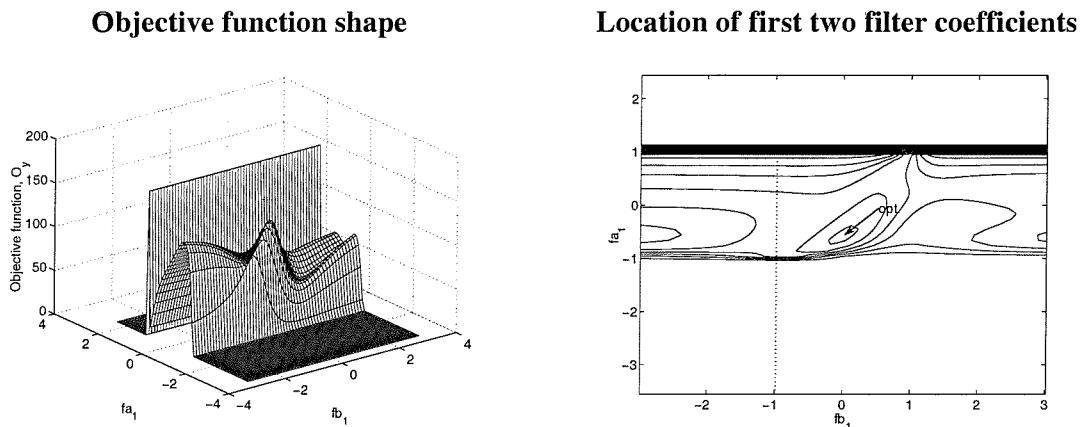


Figure 3.3.14 The objective function of the 4th order ARMA(1,1) inverse filter. Dotted line point: f_{a_1} and f_{b_1} value which gives maximum objective value, arrow point: calculated f_{a_1} and f_{b_1} value.

Thus, the inverse filter from ‘opt’ point consists of two filter coefficients (f_{a_1-opt} , f_{b_1-opt}) at the ‘opt’ point and the first AR coefficient ($f_{a_0}=1$) and MA coefficient (f_{b_0}) as

$$\mathbf{f}_{a_opt} = \begin{bmatrix} f_{a_0} & f_{a_1-opt} \end{bmatrix}^T, \quad \mathbf{f}_b = \begin{bmatrix} f_{b_0} & f_{b_1-opt} \end{bmatrix}^T$$

and the inverse filter from ‘max’ point becomes

$$\mathbf{f}_{a_max} = \begin{bmatrix} f_{a_0} & f_{a_1-max} \end{bmatrix}^T, \quad \mathbf{f}_b = \begin{bmatrix} f_{b_0} & f_{b_1-max} \end{bmatrix}^T$$

whose pole-zero maps are shown in the figure below;

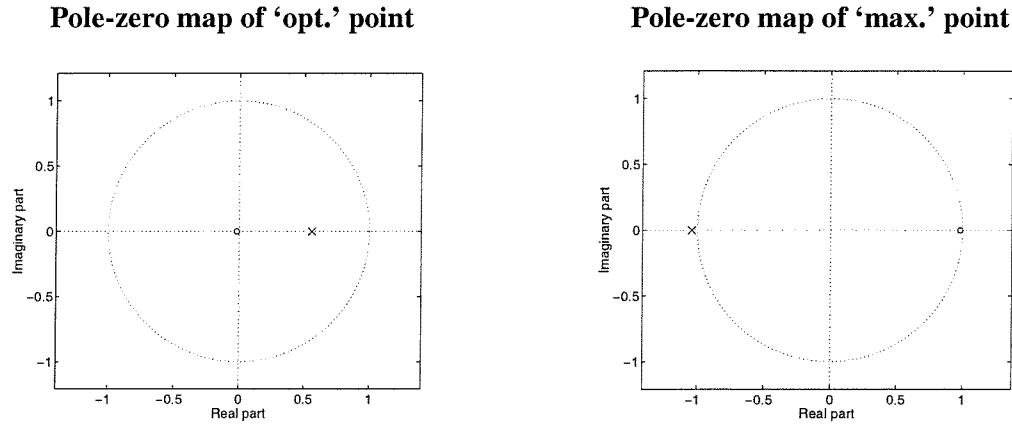


Figure 3.3.15 The ARMA(1,1) inverse system obtained by 4th order statistics.

Similarly, we can obtain two different restored signals which are plotted in Figure 3.3.16.

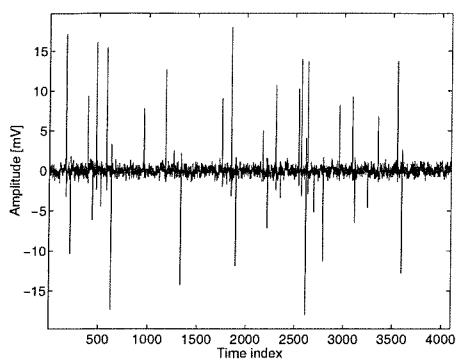
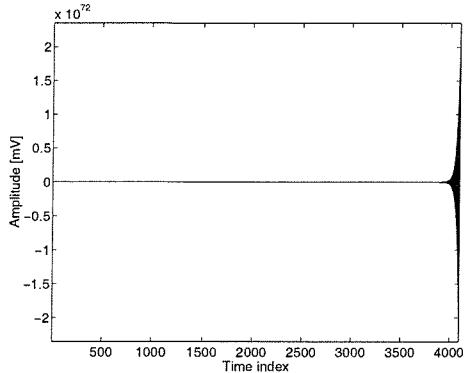
Restored signal by ‘opt.’ point filter

Skewness : 2.214
Kurtosis : 124.375
Restored signal by ‘max.’ point filter

Skewness : -0.728
Kurtosis : 165.562

Figure 3.3.16 The restored signals from 4th order ARMA(1,1) inverse filter.

As can be seen in Figure 3.3.14, similar to the AR(2) case, the point of the maximum objective function’s value (intersection of the dotted line) does not match with that of calculated (arrow point). The AR inverse filter coefficients which give maximum kurtosis value (165.562) are $fa_0=1.0$, $fa_1= 1.0412$, and $fb_0=0.9998$, $fb_1= -0.9814$ (filter coefficients on ‘max.’ point) whereas, the calculated coefficients are $fa_0=1.0$, $fa_1= -0.558$, and $fb_0=0.9998$, $fb_1= 0.0185$ (filter coefficients on ‘opt.’ point) giving the kurtosis value as 124.375. The difference in the signal restoration is compared in Figure 3.3.16 and as similar to the AR case, the restored signal from the ‘max.’ point inverse filter coefficient looks unsuccessful.

The reason for this highly unsatisfactory signal restoration arises from the fact that the recursive filters becomes unstable.

Following Table 3.3.2 depicts the results of the restored signal by above three different inverse filters. It is noticeable that the variances of the incorrectly deconvolved signals reveal large values.

Table 3.3.2 The comparison of restored signal and system from MA, AR and ARMA inverse filters (summary of the results from Figure 3.3.8 to Figure 3.3.16).

Inverse filter type (order)	Points	Kurtosis	Variance	Coefficients of inverse filter	Remarks
MA(4)	Opt.	123.643	1.862	$fb=[1.209, 0.693, 0.397, 0.215, 0.095]$ $fa=1$	Input signal is reconstructed
	Max.	128.319	0.725	$fb=[-0.091, 0.693, 0.397, 0.215, 0.095]$ $fa=1$	Input signal is reconstructed
AR(2)	Opt.	124.504	1.276	$fb=1$ $fa=[1, -0.5794, -0.0083]$	Input signal is reconstructed
	Max.	∞	1.23e+154	$fb=1$ $fa=[1, 1.9206, 1.0916]$	Unsuccessful input signal reconstruction
ARMA (1,1)	Opt.	124.375	1.270	$fb=[0.9998, 0.0185]$ $fa=[1, -0.558]$	Input signal is reconstructed
	Max.	165.562	1.56e+142	$fb=[0.9998, -0.9814]$ $fa=[1, 1.0412]$	Unsuccessful input signal reconstruction

It is almost impossible to point out the global maximum of the objective function as the observed signal has already been noise corrupted with unknown variance. Thus, the ‘max.’ point in objective function shapes in each figure does not guarantee the global maximum. However, from above observations, the MA inverse system have a wider range of acceptable filter coefficients resulting in a stable inverse system.

Concerning the instability of the signal restoration for AR, ARMA systems, the recursive structure of the inverse filter is mainly influenced by the inversion process which is described in the equation (3.3.24). A few methods (Singular Value Discarding, regularisation, etc) [Yoon and Nelson, 1995, 1997; Kirkeby et al, 1996; Tikhonov and Arsenin, 1977; Lamm, 1993] were considered but the results were not satisfactory. Another simple method was considered namely

reflecting every pole positions outside of the unit circle into their reciprocal position [Hammond and Clarkson, 1989] in each iteration process thus retaining stability. This method guarantees the stability of the AR, ARMA inverse system in BD process and yields slightly improved restoration results but incurs an increase in complexity and computational inefficiency.

3.4 Concluding remarks

This section has been devoted to the fundamental consideration of the utilisation of the Higher Order Statistics in order to reconstruct an unknown impacting signal from only a measured signal.

Starting from the basic Wiener optimisation approach for the FIR system, the blind deconvolution procedure has been justified utilising an objective function and its correspondence to ‘partial order’.

In restoring an impacting signal, two different objective functions (constrained and normalised higher order cumulant) have been justified from which FIR (non-recursive, MA inverse system) filter coefficients are calculated. As a result, both the constrained and normalised objective function maximisation procedure has yielded satisfactory restoration (Table 3.3.1).

For completeness, the inverse system has been selected as having a recursive nature (i.e., AR or ARMA system) with the expectation of improved performance over that of the non-recursive system. These three different inverse systems are compared with the shapes of each objective function with respect to the filter coefficients. The result of signal restoration from each system has been also compared. Even though all the systems possess more than one maximum, the MA inverse system turned out to be more robust than the recursive systems. Accordingly FIR inverse filter will be employed in this study.

Further practical aspects on the MA inverse system for signal reconstruction including the initial inverse filter type and filter length determination will be considered in Chapter 5.

Since blind deconvolution based on the higher order statistics utilises Wiener optimisation (zero gradient search), this process is liable to end up at a local maximum. A global optimisation method will be introduced and its performance is discussed in Chapter 6.

CHAPTER 4

Higher Order Singular Value Decomposition (HOSVD) (Application to detection, classification and reconstruction)

4.1 Introduction

An increasing number of signal processing problems involve the manipulation of multidimensional matrices in which the vectors and matrices are termed the first and the second order tensors, respectively. For many applications involving higher-order tensors, the existing framework of vector and matrix algebra is insufficient and/or inappropriate. In this chapter, a generalisation of Singular Value Decomposition (SVD) extended to the higher-order case is presented.

The importance of higher-order tensors is largely due to the expansion of interest in the field of Higher-Order Statistics (HOS). For the multivariate case the basic quantities (higher-order moments, cumulants, spectra and cepstra) are symmetric higher-order tensors, just as the covariance of a stochastic vector is a symmetric (Hermitian) matrix. Statistical descriptions of random processes are more complete when, in addition to first-and second-order statistics, HOS are taken into account. In statistical non-linear system theory, HOS are unavoidable (e.g. the autocorrelation of x^2 is a fourth-order moment). Moreover higher-order cumulants and spectra of a random variable are insensitive to additive Gaussian perturbation of the variable, thus blind to Gaussian noise. The idea of using tensorial decompositions as a basic tool in higher-order signal processing was introduced in the work of Cardoso [Cardoso, 1990, 1991, 1992; Cardoso and Comon, 1990], Comon [Comon, 1994], and Lathauwer [Lathauwer et al 1994, 1995, 1996a, 1996b, 1999].

Applications of these approaches include the following; The tensorial decomposition of multichannel system leads to a generalisation of Principal

Component Analysis (PCA) [Otte et al 1988; Dunteman, 1989] called **Independent Component Analysis** (ICA) relevant to medical signal monitoring [Cardoso and Comon, 1990; De Lathauwer and De Moor, 1996].

HOS makes it possible to solve the **Source Separation** (SS) problem by exploitation of the statistical independence of the sources without knowledge of the mixing matrix (e.g., multi-channel system, path, etc.). A brief summary for source separation from instantaneous mixture processes is given in Appendix C.

In higher order array processing, **blind identification** based on a super-symmetric decomposition of the fourth order cumulant tensor has been developed [Cardoso, 1991] and robust identification has been achieved from low rank estimation of the third order cumulant tensor [Bradaric and Petropulu, 1999].

In this chapter, a pre-diagnosis procedure is introduced which can help the signal processor and/or mechanical engineer assess the status of the measured signal even from a single channel. To do this, a multilinear generalisation of Singular Value Decomposition (SVD), named Higher-Order Singular Value Decomposition (HOSVD) is presented. Using HOSVD, blind detection of non-Gaussian signals is carried out. This detection procedure includes a sub division of single observed signal into multiple channels. The matrix constructed from the subdivision can be found in the embedding process itself by the ‘method of delays’ [Takens, 1981]. This subdivided matrix is known as a trajectory matrix and has been used for example in AR order selection through SVD [Shin, 1996].

By constructing the covariance matrix of these multiple, delayed channels (trajectory matrix), the second order singular value decomposition (SOSVD) is performed. Sequentially, by the construction of the third or fourth order cumulant tensors, higher order (third or fourth) singular value decomposition is carried out. The magnitudes of each singular value (from SOSVD and HOSVD) are compared. The detection of a non-Gaussian signal is based upon computed statistics measured against proposed threshold levels. These statistics are estimates of higher order (third and fourth) singular values. Classification of inputs and systems through which they pass is then relies on properties of these statistics.

Examples of the detection procedure using three different input signals (Gaussian or non-Gaussian) with two selected systems (highly damped and highly resonant) are demonstrated.

Also, based on HOSVD, procedures for signal classification which can determine the status of the signal and the possibility of reconstruction of the non-Gaussian impacting signal are proposed and illustrated.

4.2 SVD and HOSVD

One of the most significant developments in linear algebra and signal processing problems is the concept of the Singular Value Decomposition (SVD) of matrices and has a long history of steadily growing success (for a detailed historical survey, refer to [Stewart, 1993] and [Horn and Johnson, 1991]). For the second-order case, the matrix \mathbf{A} ($I_1 \times I_2$) can be decomposed in terms of elements of matrices

$$a_{i_1 i_2} = \sum_{j_1}^{I_1} \sum_{j_2}^{I_2} u_{i_1 j_1} s_{i_1 i_2} v_{i_2 j_2} \quad (4.2.1)$$

where $s_{i_1 i_2}$ denotes the element of singular value matrix, $u_{i_1 j_1}$ is the left singular vector corresponding to the i_1 -th column and $v_{i_2 j_2}$ is the right singular vector corresponding to the i_2 -th row. Similarly to the second-order case, a decomposition of a real ($I_1 \times I_2 \times I_3$)-tensor \mathcal{A} becomes

$$a_{i_1 i_2 i_3} = \sum_{j_1}^{I_1} \sum_{j_2}^{I_2} \sum_{j_3}^{I_3} s_{i_1 i_2 i_3} u_{i_1 j_1}^{(1)} u_{i_2 j_2}^{(2)} u_{i_3 j_3}^{(3)} \quad (4.2.2)$$

in which $u_{i_1 j_1}^{(1)}$, $u_{i_2 j_2}^{(2)}$, $u_{i_3 j_3}^{(3)}$ are entries of orthogonal matrices (the superscript notation indicates a ‘mode’ to be defined in section 4.2.3), and \mathbf{S} is a real ($I_1 \times I_2 \times I_3$)-tensor with the property of “all-orthogonality”. As will be discussed later, this orthogonality satisfies $\sum_{i_2} s_{i_1 i_2 \alpha} s_{i_1 i_2 \beta} = \sum_{i_3} s_{i_1 \alpha i_3} s_{i_1 \beta i_3} = \sum_{i_3} s_{\alpha i_2 i_3} s_{\beta i_2 i_3} = 0$, whenever $\alpha \neq \beta$. It is the aim of this chapter to derive the tensorial decomposition in an algebraic context, using SVD-terminology and an SVD-based computation scheme, and to present it as a valuable tool relevant to the field of

numerical algebra and signal processing. This is called *Higher-Order Singular Value Decomposition (HOSVD)*.

As a preliminary, definitions and basic tensor algebra are introduced followed by the higher order cumulant tensor and their properties. A singular value decomposition model is proposed for N th-order tensors [De Lathauwer, 1997] and comparison to the structure of matrix SVD show the complete analogy between both cases is presented.

4.2.1 Basic Definitions of Tensor Algebra

In this section, notation and definitions dealing with some basic concepts of multilinear algebra is introduced.

Definition 1 (Nth order tensor)

The term ‘ N th order’ tensor is the generalisation of the dimensions of vector, matrix, and tensor, i.e., a vector is a one dimensional signal and equivalent to first order tensor, a matrix is a two dimensional signal which is the second order tensor. Going further, the N th order tensor is an N dimensional signal space and the general form of N dimensional matrix.

Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$, be N Euclidean vector space with finite dimensions I_1, I_2, \dots, I_N . Consider N vectors $\mathbf{U}_1 \in \mathbf{v}_1, \mathbf{U}_2 \in \mathbf{v}_2, \dots, \mathbf{U}_N \in \mathbf{v}_N$. Then $\mathbf{U}_1 \circ \mathbf{U}_2 \circ \dots \circ \mathbf{U}_N$ can be denoted by the multilinear mapping on $\mathbf{v}_1 \times \mathbf{v}_2 \times \dots \times \mathbf{v}_N$, in which ‘ \circ ’ means the tensor outer product and vector outer product, respectively. The relationship is defined by

$$(\mathbf{U}_1 \circ \mathbf{U}_2 \circ \dots \circ \mathbf{U}_N)(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N) = \langle \mathbf{U}_1, \mathbf{X}_1 \rangle_{\mathbf{v}_1} \langle \mathbf{U}_2, \mathbf{X}_2 \rangle_{\mathbf{v}_2} \dots \langle \mathbf{U}_N, \mathbf{X}_N \rangle_{\mathbf{v}_N} \quad (4.2.3)$$

in which $\langle \mathbf{U}_N, \mathbf{X}_N \rangle_{\mathbf{v}_N}$ denotes the scalar product in \mathbf{v}_N , and \mathbf{X}_N is an arbitrary vector in \mathbf{v}_N ($1 \leq n \leq N$).

The space generated by all the elements $\mathbf{U}_1 \circ \mathbf{U}_2 \circ \dots \circ \mathbf{U}_N$ is called the *tensor product space* of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$. An element of the tensor product space is called as an *N -th order tensor* (over $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$).

For $\mathbf{v}_N = \mathbb{R}^{I_N}$ ($1 \leq n \leq N$), the tensor product space is called the space of real-valued $(I_1 \times I_2 \times \dots \times I_N)$ -tensors, denoted by $\mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. Its complex equivalent is represented by $\mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$.

Definition 2 (Outer product of tensors)

The definition of an outer product generalises expressions of the type $\mathbf{a}\mathbf{b}^T$ in which \mathbf{a} and \mathbf{b} are vectors.

The outer product $\mathcal{A} \circ \mathcal{B}$ of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_p}$ and a tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_q}$, is defined by

$$(\mathcal{A} \circ \mathcal{B})_{i_1 i_2 \dots i_p j_1 j_2 \dots j_q} \stackrel{\text{def}}{=} a_{i_1 i_2 \dots i_p} b_{j_1 j_2 \dots j_q} \quad (4.2.4)$$

for all values of the indices.

For example, the entries of an N th-order tensor \mathcal{A} , equal to the outer product of N vectors $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(N)}$, are given by $a_{i_1 i_2 \dots i_N} = u_{i_1}^{(1)} u_{i_2}^{(2)} \dots u_{i_N}^{(N)}$.

Definition 3 (Scalar product of tensors)

The definition of an scalar product of two vectors \mathbf{a}, \mathbf{b} are expressed in well known form as $\mathbf{a}^T \mathbf{b}$.

The scalar product $\langle \mathcal{A}, \mathcal{B} \rangle$ of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is defined as

$$\langle \mathcal{A}, \mathcal{B} \rangle \stackrel{\text{def}}{=} \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} b_{i_1 i_2 \dots i_N} a_{i_1 i_2 \dots i_N} \quad (4.2.5)$$

Definition 4 (Orthogonality)

Tensors of which the scalar product equals 0, are mutually orthogonal.

$$\langle \mathcal{A}, \mathcal{B} \rangle = 0 \Rightarrow \text{Tensors } \mathcal{A} \text{ and } \mathcal{B} \text{ are mutually orthogonal} \quad (4.2.6)$$

Definition 5 (Frobenius-norm)

The Frobenius-norm of a tensor \mathcal{A} is given by

$$\|\mathcal{A}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} \quad (4.2.7)$$

In tensor terminology, column vectors, row vectors, third-directional (orthogonal to the column and row vectors), etc will be called mode-1 vectors, mode-2 vectors, and mode-3 vectors, etc. In general, the mode- n vectors of an N th-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ are the I_n -dimensional vectors obtained from \mathcal{A} by varying the index i_n and keeping the other indices fixed.

Definition 6 (Multiplication of a higher order tensor and matrix)

The mode- n product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is an $(I_1 \times I_2 \times \dots \times I_{n-1} \times J_n \times I_{n+1} \times \dots \times I_N)$ -tensor defined by

$$\langle \mathcal{A} \times_n \mathbf{U} \rangle_{i_1 i_2 \dots j_n \dots i_N} = \sum_{i_n} a_{i_1 i_2 \dots i_n \dots i_N} u_{j_n i_n} \quad (4.2.8)$$

for all index values.

The mode- n product allows one to express the effect of a basis transformation in \mathbb{R}^{I_n} on the tensor \mathcal{A} .

As an example, for the matrix product $\mathbf{A} = \mathbf{U}^{(1)} \cdot \mathbf{B} \cdot \mathbf{U}^{(2)^T}$, which involves matrices, $\mathbf{B} \in \mathbb{R}^{I_1 \times I_2}$, $\mathbf{U}^{(1)} \in \mathbb{R}^{J_1 \times I_1}$, $\mathbf{U}^{(2)} \in \mathbb{R}^{J_2 \times I_2}$ and $\mathbf{A} \in \mathbb{R}^{J_1 \times J_2}$. Working with the “generalised transposes”, in the multilinear case (in which the fact that mode-1 vectors are transpose-free, would not have an inherent meaning), can be avoided by observing that the relationships between $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ (not $\mathbf{U}^{(2)^T}$) with \mathbf{B} are in fact completely similar: in the same way as $\mathbf{U}^{(1)}$ makes linear combinations of the rows of \mathbf{B} , $\mathbf{U}^{(2)}$ makes linear combinations of the columns; in the same way as the columns of \mathbf{B} are multiplied by $\mathbf{U}^{(1)}$, its rows are multiplied by $\mathbf{U}^{(2)}$; in the same way as the columns of $\mathbf{U}^{(1)}$ are associated with the column space of \mathbf{A} , the columns of $\mathbf{U}^{(2)}$ are associated with the row space. This typical relationship is denoted by means of the \times_n -symbol: $\mathbf{A} = \mathbf{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)}$.

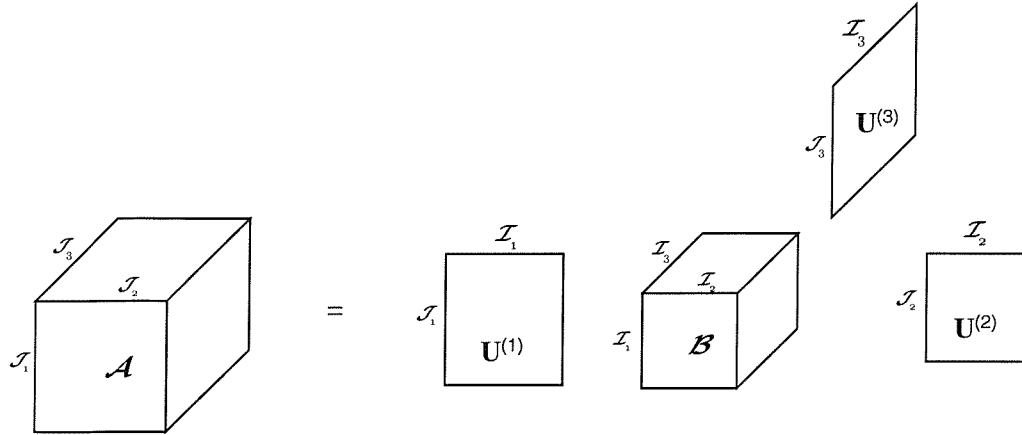


Figure 4.2.1 Visualisation of the multiplication of a third-order $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, with matrices $\mathbf{U}^{(1)} \in \mathbb{R}^{J_1 \times I_1}$, $\mathbf{U}^{(2)} \in \mathbb{R}^{J_2 \times I_2}$ and $\mathbf{U}^{(3)} \in \mathbb{R}^{J_3 \times I_3}$.

Figure 4.2.1 visualises the equation $\mathcal{A} = \mathcal{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}$ for third-order tensors $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. Unlike the conventional way of visualising second-order matrix products, $\mathbf{U}^{(2)}$ has not been transposed, for reasons of symmetry. Multiplication with $\mathbf{U}^{(1)}$ involves linear combinations of the “horizontal matrices” (index i_1 fixed) in \mathcal{B} . Stated otherwise, multiplication of \mathcal{B} with $\mathbf{U}^{(1)}$ means that every column of \mathcal{B} (indices i_2 and i_3 fixed) has to be multiplied from the left with $\mathbf{U}^{(1)}$. Multiplication with $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ can be explained in a similar way.

Definition 7 (Higher order moment tensor)

The N th-order moment tensor $\mathcal{M}_x^{(N)}$ of a real stochastic vector \mathbf{x} is defined by the element-wise equation:

$$\mathcal{M}_x^{(N)} = \text{Mom}(x_{i_1}, x_{i_2}, \dots, x_{i_N}) \quad \underline{\text{def}} \quad E\{x_{i_1} x_{i_2} \cdots x_{i_N}\} \quad (4.2.9)$$

The first-order moment is the mean of the stochastic vector. The second-order moment is the correlation matrix.

Definition 8 (Higher order cumulant tensor)

The N th-order cumulant tensor $\mathcal{C}_x^{(N)}$ of a real stochastic vector x is defined by the element-wise equation:

$$\mathcal{C}_x^{(N)} = \text{Cum}(x_{i1}, x_{i2}, \dots, x_{iN}) \underset{\text{def}}{=} \sum (-1)^N (N-1)! E\left\{ \prod_{i \in A_1} x_i \right\} E\left\{ \prod_{i \in A_2} x_i \right\} \dots E\left\{ \prod_{i \in A_K} x_i \right\} \quad (4.2.10)$$

where the summation involves all possible partitions $\{A_1, A_2, \dots, A_K\}$, $(1 \leq K \leq N)$ of the integers $\{i_1, i_2, \dots, i_N\}$. For a real zero-mean stochastic vector x the cumulants up to order 4 are explicitly given by:

$$(\mathcal{C}_x^{(1)})_i = \text{Cum}(x_i)$$

$$\underset{\text{def}}{=} E\{x_i\}$$

$$(\mathcal{C}_x^{(2)})_{i_1 i_2} = \text{Cum}(x_{i_1}, x_{i_2})$$

$$\underset{\text{def}}{=} E\{x_{i_1} x_{i_2}\}$$

$$(\mathcal{C}_x^{(3)})_{i_1 i_2 i_3} = \text{Cum}(x_{i_1}, x_{i_2}, x_{i_3})$$

$$\underset{\text{def}}{=} E\{x_{i_1} x_{i_2} x_{i_3}\}$$

$$(\mathcal{C}_x^{(4)})_{i_1 i_2 i_3 i_4} = \text{Cum}(x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4})$$

$$\underset{\text{def}}{=} E\{x_{i_1} x_{i_2} x_{i_3} x_{i_4}\}$$

$$- E\{x_{i_1} x_{i_2}\} E\{x_{i_3} x_{i_4}\}$$

$$- E\{x_{i_1} x_{i_3}\} E\{x_{i_2} x_{i_4}\}$$

$$- E\{x_{i_1} x_{i_4}\} E\{x_{i_2} x_{i_3}\}$$

Except for the first-order cumulant case, every component x_i of x that has a non-zero-mean, x_i has to be replaced above formulas by $x_i - E\{x_i\}$.

The first-order cumulant is the mean of the stochastic vector. The second-order cumulant is the covariance matrix.

The particular form of definitions (4.2.9) and (4.2.10) follows from the representation of the probability density function of \mathbf{x} by means of its first and second characteristic function, of which the Taylor series coefficients are proportional to the moments and cumulants, respectively [Barrett, 1964; McCullagh, 1987].

4.2.2 Properties of Moment and Cumulant Tensors

The higher-order moments in above formulas have a simpler structure than those of the cumulant tensor. However, cumulants have a number of important properties, that are not shared with higher order moments, such that in practice cumulants are more frequently used. In this section, some of the most interesting properties of the higher order moment and cumulant tensors [Nikias and Mendel, 1993; De Lathauwer, 1997] have been introduced.

- Super-symmetry: moments and cumulants are symmetric in their arguments, i.e.,

$$(\mathcal{M}_{\mathbf{x}}^{(N)})_{i_1 i_2 \dots i_N} = (\mathcal{M}_{\mathbf{x}}^{(N)})_{P(i_1 i_2 \dots i_N)} \quad (4.2.12)$$

$$(\mathcal{C}_{\mathbf{x}}^{(N)})_{i_1 i_2 \dots i_N} = (\mathcal{C}_{\mathbf{x}}^{(N)})_{P(i_1 i_2 \dots i_N)} \quad (4.2.13)$$

in which P is an arbitrary permutation of the indices.

- Multilinearity: if a real stochastic vector \mathbf{x} is transposed into a stochastic vector $\tilde{\mathbf{x}}$ by a matrix multiplication $\tilde{\mathbf{x}} = \mathbf{A} \cdot \mathbf{x}$, with $\mathbf{A} \in \mathbb{R}^{J \times I}$, then we have:

$$\mathcal{M}_{\tilde{\mathbf{x}}}^{(N)} = \mathcal{M}_{\mathbf{x}}^{(N)} \times_1 \mathbf{A} \times_2 \mathbf{A} \times_3 \dots \times_N \mathbf{A} \quad (4.2.14)$$

$$\mathcal{C}_{\tilde{\mathbf{x}}}^{(N)} = \mathcal{C}_{\mathbf{x}}^{(N)} \times_1 \mathbf{A} \times_2 \mathbf{A} \times_3 \dots \times_N \mathbf{A} \quad (4.2.15)$$

which represents the homogeneity of the moment and cumulant tensor.

- Even distribution: if a real random variable x has an even probability density function $p_x(x)$, i.e. $p_x(x)$ is symmetric about the origin, then the odd moments and cumulants of x vanish.
- Partitioning of independent variables: if a subset of I stochastic variables x_1, x_2, \dots, x_I are independent of the other variables, then we have:

$$\text{Cum}(x_1, x_2, \dots, x_I) = 0 \quad (4.2.16)$$

This property does not hold in general for moments. A consequence of the property is that a higher-order cumulant of a stochastic vector having mutually independent components, is a *diagonal tensor*, i.e. only the entries of which all the indices are equal can be different from zero.

- Sum of independent variables: if the stochastic variables x_1, x_2, \dots, x_I are mutually independent from the stochastic variables y_1, y_2, \dots, y_I , then:

$$\text{Cum}(x_1 + y_1, x_2 + y_2, \dots, x_k + y_k) = \text{Cum}(x_1, x_2, \dots, x_k) + \text{Cum}(y_1, y_2, \dots, y_k) \quad (4.2.17)$$

The cumulant tensor of a sum of independent random vectors is the sum of the individual cumulants. Also, this property does not hold in general for moments, hence it explains the term “cumulant”.

- Non-Gaussianity: if y is a Gaussian variable with the same mean and variance as a given stochastic variable x , then for order, $N \geq 3$:

$$\mathcal{C}_x^{(N)} = \mathcal{M}_x^{(N)} - \mathcal{M}_y^{(N)} \quad (4.2.18)$$

represents the distance from the Gaussian, thus higher-order cumulants of a Gaussian variable are 0. With the multilinearity property, the higher-order cumulants are blind to an additive Gaussian noise which has been already mentioned in Chapter 2.

Generally speaking, it becomes harder to obtain reliable estimates of HOS from sample data as the order increases, i.e. longer data sets are required to retain the

same accuracy as for the second order case [Cardoso, 1991; Comon, 1994]. Hence in practice the use of HOS is usually restricted to third- and fourth-order cumulants. For symmetric distributions, fourth-order cumulants are commonly used.

4.2.3 Singular Value Decomposition (SVD) and Higher-Order Singular Value Decomposition (HOSVD)

The HOSVD is derived by formulating it in terms of matrix/vector-sub problems. Hence, it will be useful to represent higher-order tensors in a matrix format. The tensor-matrix relationship is algebraically and geometrically related through “matrix unfolding” of a given tensor \mathcal{A} from which the process of the matrix SVD and higher-order SVD is explained, which shows the analogy between them. The unfolding of the tensor into the matrix form can yield the conventional SVD process from which the higher-order singular values are estimated. To assist understanding, visualisation of the HOSVD and unfolding is presented. This is followed by examples of HOSVD from a theoretical and computational point of view.

Matrix Singular Value Decomposition

Every complex $(I_1 \times I_2)$ -matrix \mathbf{A} can be written as the product

$$\mathbf{A} = \mathbf{U}^{(1)} \cdot \mathbf{S} \cdot \mathbf{V}^{(2)^H} = \mathbf{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{V}^{(2)^*} = \mathbf{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \quad (4.2.19)$$

in which:

- $\mathbf{U}^{(1)} = \begin{bmatrix} U_1^{(1)} & U_2^{(1)} & \dots & U_{I_1}^{(1)} \end{bmatrix}$ is a unitary $(I_1 \times I_1)$ -matrix,
- $\mathbf{U}^{(2)} = \begin{bmatrix} U_1^{(2)} & U_2^{(2)} & \dots & U_{I_2}^{(2)} \end{bmatrix}$ ($= \mathbf{V}^{(2)^*}$) is a unitary $(I_2 \times I_2)$ -matrix,
- \mathbf{S} is an $(I_1 \times I_2)$ -matrix with the property of:

- *pseudodiagonality*:

$$\mathbf{S} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{\min(I_1, I_2)}) \quad (4.2.20)$$

- ordering:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min(I_1, I_2)} \geq 0 \quad (4.2.21)$$

The σ_i are singular values of A and the vectors $U_i^{(1)}$ and $U_i^{(2)}$ are respectively an i th left and an i th right singular vector. The decomposition is visualised in Figure 4.2.2.

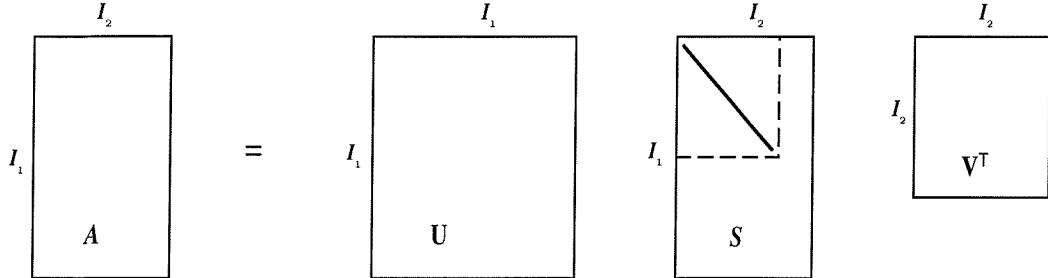


Figure 4.2.2 Visualisation of the matrix SVD

Nth-Order Singular Value Decomposition

Every complex $(I_1 \times I_2 \times \dots \times I_N)$ -tensor \mathcal{A} can be written as the product

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_N \mathbf{U}^{(N)} \quad (4.2.22)$$

in which:

- $\mathbf{U}^{(n)} = [U_1^{(n)} U_2^{(n)} \dots U_{I_n}^{(n)}]$ is a unitary $(I_n \times I_n)$ -matrix,
- \mathcal{S} is a complex $(I_1 \times I_2 \times \dots \times I_N)$ -tensor of which the subtensors $\mathcal{S}_{i_n=\alpha}$, obtained

by fixing the n th index to α , have the property of:

- *all-orthogonality*:

two subtensors $\mathcal{S}_{i_n=\alpha}$ and $\mathcal{S}_{i_n=\beta}$ are orthogonal for all possible values of n , α and β subject to $\alpha \neq \beta$:

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \quad \text{when} \quad \alpha \neq \beta \quad (4.2.23)$$

- *ordering*:

$$\|\mathcal{S}_{i_n=1}\| \geq \|\mathcal{S}_{i_n=2}\| \geq \dots \geq \|\mathcal{S}_{i_n=I_n}\| \geq 0 \quad (4.2.24)$$

for all possible values of n .

The Frobenius-norm $\|\mathcal{S}_{i_n=i}\|$, symbolised by $\sigma_i^{(n)}$ for matrix case, is the i th n -mode singular value of \mathcal{A} and the vector $U_i^{(n)}$ is the i th n -mode singular vector. The decomposition is visualised for third-order tensors in the Figure 4.2.3.

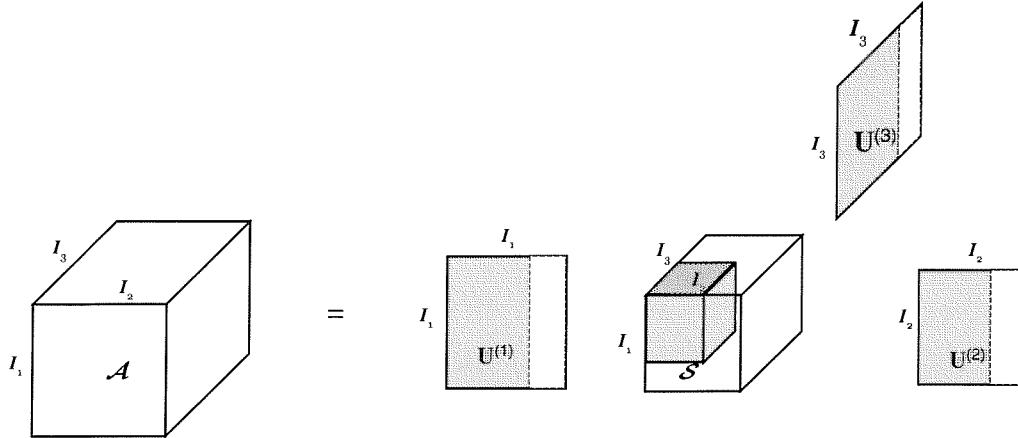


Figure 4.2.3 Visualisation of the HOSVD for a third-order tensor

From a comparison of the matrix and tensor theorem it is clear that in the higher-order case the “core matrix” (singular value matrix) S is replaced by the “core tensor” \mathcal{S} . Instead of being pseudo-diagonal, which would mean that non-zero elements could only occur when the indices $i_1 = i_2 = \dots = i_N$, \mathcal{S} is in general a full tensor, but it obeys the weaker condition of all-orthogonality instead. The actual role of the singular values is taken over by the Frobenius-norms of the $(N-1)$ th-order subtensors of the core tensor. By definition the n -mode singular values are positive and real, as in the matrix case. On the other hand the entries of \mathcal{S} are not necessarily positive in general; they can even be complex, when \mathcal{A} is a complex-valued tensor.

Matrix representation of a higher-order tensor (unfolding of tensor)

The HOSVD will be derived by formulating it in terms of matrix/vector-sub problems. Hence it will be useful to represent higher-order tensors in a matrix form. This tensor-matrix relationship can be geometrically related through “matrix unfolding” of a given tensor \mathcal{A} .

The matrix unfolding $\mathbf{A}_{(I_{P(1)} \times I_{P(2)} \cdot I_{P(3)} \cdots I_{P(N)})} \in \mathbb{C}^{(I_{P(1)} \times I_{P(2)} \cdot I_{P(3)} \cdots I_{P(N)})}$, associated with a permutation $(P(1), P(2), \dots, P(N))$ of $(1, 2, \dots, N)$ of a N th-order tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ contains the element $a_{i_{P(1)} i_{P(2)} \cdots i_{P(N)}}$ at the position with row number $i_{P(1)}$ and column number $(i_{P(2)} - 1)I_{P(3)} I_{P(4)} \cdots I_{P(N)} + (i_{P(3)} - 1)I_{P(4)} I_{P(5)} \cdots I_{P(N)} + \cdots + i_{P(N)}$. An example of this unfolding is given in later part of this section.

The unfolding operation involved with the construction of $\mathbf{A}_{(1)}$, $\mathbf{A}_{(2)}$ and $\mathbf{A}_{(3)}$ is visualised for the third order case in the following figure.

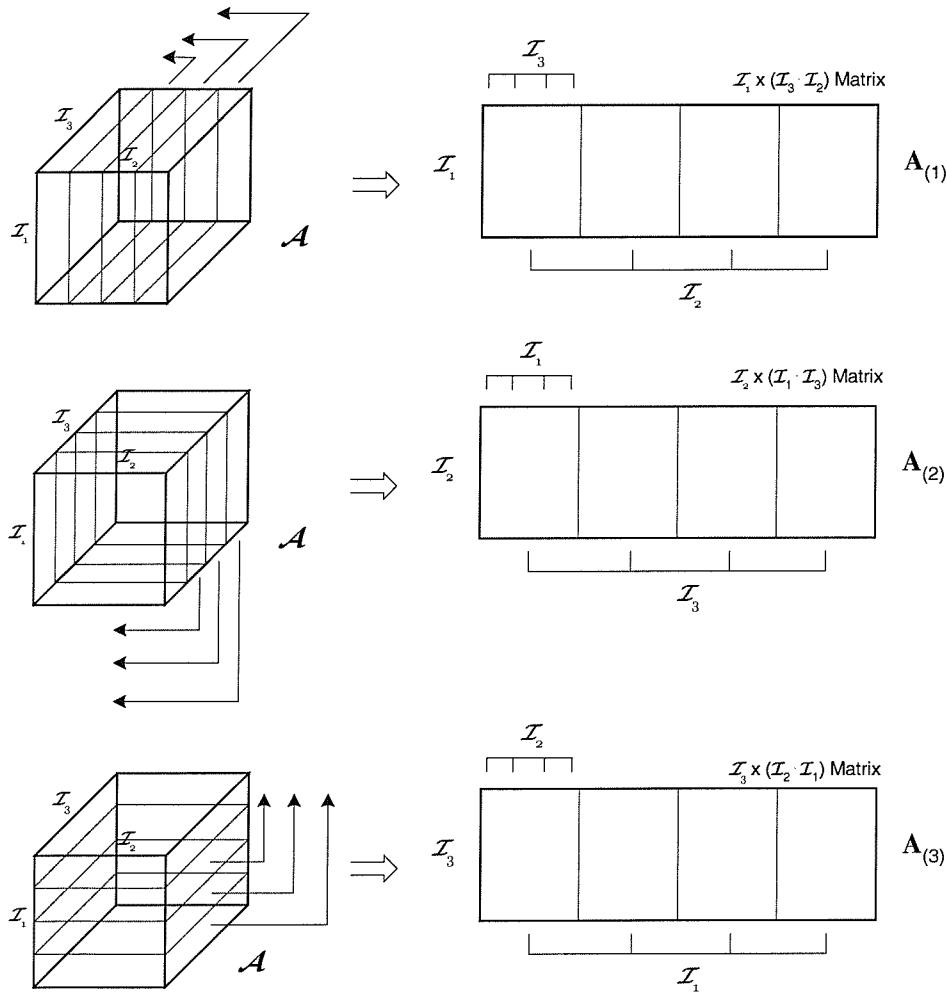


Figure 4.2.4 Unfolding of the $(I_1 \times I_2 \times I_3)$ -tensor to the $(I_1 \times I_3 \cdot I_2)$ -matrix $\mathbf{A}_{(1)}$ (Mode-1 matrix form), the $(I_2 \times I_1 \cdot I_3)$ -matrix $\mathbf{A}_{(2)}$ (Mode-2 matrix form) and the $(I_3 \times I_2 \cdot I_1)$ -matrix $\mathbf{A}_{(3)}$ (Mode-3 matrix form). Note that $(I_1 = I_2 = I_3 = 4)$.

The matrix representation of the HOSVD can be obtained by unfolding \mathcal{A} and \mathcal{S} in equation (4.2.22):

$$\mathbf{A}_{(I_{P(1)} \times I_{P(2)} \times I_{P(3)} \dots \times I_{P(N)})} = \mathbf{U}^{(p(1))} \cdot \mathbf{S}_{(I_{P(1)} \times I_{P(2)} \times I_{P(3)} \dots \times I_{P(N)})} \cdot \left[\mathbf{U}^{(p(2))} \otimes \mathbf{U}^{(p(3))} \otimes \dots \otimes \mathbf{U}^{(p(N))} \right]^T \quad (4.2.25)$$

in which ‘ \otimes ’ denotes the Kronecker (tensor) product and P represents a permutation.

Derivation of the core tensor \mathcal{S}

The derivation establishes the connection between the HOSVD of a tensor \mathcal{A} and the matrix SVD of a matrix unfolding of \mathcal{A} . It is given in terms of real-valued tensors; the complex case is completely analogous but more cumbersome from a notational point of view.

Consider two $(I_1 \times I_2 \times \dots \times I_N)$ -tensors \mathcal{A} and \mathcal{S} , related by

$$\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \dots \times_N \mathbf{U}^{(N)^T} \quad (4.2.26)$$

in which $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}$ are orthogonal matrices. Equation (4.2.26) can be expressed in matrix format as

$$\mathbf{A}_{(n)} = \mathbf{U}^{(n)} \cdot \mathbf{S}_{(n)} \cdot \left[\mathbf{U}^{(1)} \otimes \mathbf{U}^{(2)} \otimes \dots \otimes \mathbf{U}^{(n-1)} \otimes \mathbf{U}^{(n+1)} \otimes \dots \otimes \mathbf{U}^{(N)} \right]^T \quad (4.2.27)$$

Now consider the particular case where $\mathbf{U}^{(n)}$ is obtained from the SVD of $\mathbf{A}_{(n)}$ as

$$\mathbf{A}_{(n)} = \mathbf{U}^{(n)} \cdot \Sigma^{(n)} \cdot \mathbf{V}^{(n)^T} \quad (4.2.28)$$

in which $\mathbf{V}^{(n)}$ is orthogonal and $\Sigma^{(n)} = \text{diag}\{\sigma_1^{(n)}, \sigma_2^{(n)}, \dots, \sigma_{I_n}^{(n)}\}$, where

$$\sigma_1^{(n)} \geq \sigma_2^{(n)} \geq \dots \geq \sigma_{I_n}^{(n)} \geq 0 \quad (4.2.29)$$

The highest index for which $\sigma_{r_n}^{(n)} > 0$ is called rank r_n . Taking into account that the Kronecker factor in equation (4.2.27) is orthogonal, comparing (4.2.27) and (4.2.28) yields

$$\mathbf{S}_{(n)} = \Sigma^{(n)} \cdot \mathbf{V}^{(n)^T} \cdot \left[\mathbf{U}^{(1)} \otimes \mathbf{U}^{(2)} \otimes \dots \otimes \mathbf{U}^{(n-1)} \otimes \mathbf{U}^{(n+1)} \otimes \dots \otimes \mathbf{U}^{(N)} \right] \quad (4.2.30)$$

Using equation (4.2.25), (4.2.28) and (4.2.30), the n -mode core tensor thus, can be calculated as

$$\mathbf{S}_{(I_{P(1)} \times I_{P(2)} \times I_{P(3)} \dots \times I_{P(N)})} = \mathbf{U}^{(p(1))^T} \cdot \mathbf{A}_{(I_{P(1)} \times I_{P(2)} \times I_{P(3)} \dots \times I_{P(N)})} \cdot \left[\mathbf{U}^{(p(2))} \otimes \mathbf{U}^{(p(3))} \otimes \dots \otimes \mathbf{U}^{(p(N))} \right] \quad (4.2.31)$$

The equations (4.2.30) and (4.2.31) imply, for arbitrary orthogonal matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(n-1)}, \mathbf{U}^{(n+1)}, \dots, \mathbf{U}^{(N)}$, that

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \quad \text{when } \alpha \neq \beta \quad (4.2.32)$$

and

$$\|\mathcal{S}_{i_n=1}\| = \sigma_1^{(n)} \geq \|\mathcal{S}_{i_n=2}\| = \sigma_2^{(n)} \geq \dots \geq \|\mathcal{S}_{i_n=I_n}\| = \sigma_{I_n}^{(n)} \geq 0 \quad (4.2.33)$$

and, if $r_n < I_n$,

$$\|\mathcal{S}_{i_n=r_n+1}\| = \sigma_{r_n+1}^{(n)} = \|\mathcal{S}_{i_n=r_n+2}\| = \sigma_{r_n+2}^{(n)} = \dots = \|\mathcal{S}_{i_n=I_n}\| = \sigma_{I_n}^{(n)} = 0 \quad (4.2.34)$$

By constructing $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(n-1)}, \mathbf{U}^{(n+1)}, \dots, \mathbf{U}^{(N)}$ in a similar way as $\mathbf{U}^{(n)}$, \mathcal{S} can be made satisfy all the conditions of the HOSVD-theorem. On the other hand, as can be seen from (4.2.27)-(4.2.34), all matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}$ and tensors \mathcal{S} satisfying the HOSVD-theorem can be found by the SVD of $\mathbf{A}_{(1)}, \mathbf{A}_{(2)}, \dots, \mathbf{A}_{(N)}$, where \mathcal{S} follows from (4.2.26).

Computation of the core tensor \mathcal{S} from a rank 1 tensor \mathcal{A} (Example)

Suppose for a certain tensor $\mathcal{A} \in \mathbb{R}^{3 \times 2 \times 3}$ with $a_{111} = a_{112} = a_{113} = 1$, $a_{211} = a_{123} = -a_{212} = -a_{313} = 1$, $a_{311} = a_{122} = a_{113} = a_{321} = a_{322} = a_{323} = -a_{213} = -a_{222} = 2$, $a_{312} = 0$, $a_{121} = a_{221} = 4$ and $a_{223} = -5$. The three modes of matrix unfolding $\mathbf{A}_{(1)}$, $\mathbf{A}_{(2)}$ and $\mathbf{A}_{(3)}$ are given as (graphical illustration can be found in the Figure 4.2.4)

$$\mathbf{A}_{(1)} = \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 4 & 2 & 1 \\ 1 & -1 & -2 & 4 & -2 & -5 \\ 2 & 0 & -1 & 2 & 2 & 2 \end{array} \right),$$

$$\mathbf{A}_{(2)} = \begin{pmatrix} 1 & 1 & 2 & | & 1 & -1 & 0 & | & 1 & -2 & -1 \\ 4 & 4 & 2 & | & 2 & -2 & 2 & | & 1 & -5 & 2 \end{pmatrix},$$

$$\mathbf{A}_{(3)} = \begin{pmatrix} 1 & 4 & | & 1 & 4 & | & 2 & 2 \\ 1 & 2 & | & -1 & -2 & | & 0 & 2 \\ 1 & 1 & | & -2 & -5 & | & -1 & 2 \end{pmatrix}.$$

The 1-mode singular vectors $\mathbf{U}^{(1)}$ are the columns of the left singular matrix of $\mathbf{A}_{(1)}$. In the same way $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ can be obtained from SVD of $\mathbf{A}_{(2)}$ and $\mathbf{A}_{(3)}$.

$$\mathbf{U}^{(1)} = \begin{pmatrix} 0.1893 & -0.7572 & 0.6252 \\ 0.9816 & 0.1622 & -0.1008 \\ 0.0251 & -0.6327 & -0.7740 \end{pmatrix},$$

$$\mathbf{U}^{(2)} = \begin{pmatrix} 0.3256 & 0.9455 \\ 0.9455 & -0.3256 \end{pmatrix},$$

$$\mathbf{U}^{(3)} = \begin{pmatrix} -0.6564 & 0.7055 & -0.2673 \\ 0.2546 & 0.5406 & 0.8018 \\ 0.7102 & 0.4582 & -0.5345 \end{pmatrix}.$$

The core tensor of the HOSVD then follows from application of equation (4.2.31); unfolded in each mode equal to:

$$\mathbf{S}_{(1)} = \begin{pmatrix} -7.1846 & 0.2235 & 0.0000 & | & 0.0615 & -0.4906 & 0.0000 \\ -0.1703 & -5.9024 & 0.0000 & | & 0.7654 & -0.0991 & 0.0000 \\ 0.0035 & 0.2675 & 0.0000 & | & 2.1071 & 0.3350 & 0.0000 \end{pmatrix},$$

$$\mathbf{S}_{(2)} = \begin{pmatrix} -7.1846 & -0.1703 & 0.0035 & | & 0.2235 & -5.9024 & 0.2675 & | & 0.0000 & 0.0000 & 0.0000 \\ 0.0615 & 0.7654 & 2.1071 & | & -0.4906 & -0.0991 & 0.3350 & | & 0.0000 & 0.0000 & 0.0000 \end{pmatrix}$$

$$\mathbf{S}_{(3)} = \begin{pmatrix} -7.1846 & 0.0615 & | & -0.1703 & 0.7654 & | & 0.0035 & 2.1071 \\ 0.2235 & -0.4906 & | & -5.9024 & -0.0991 & | & 0.2675 & 0.3350 \\ 0.0000 & 0.0000 & | & 0.0000 & 0.0000 & | & 0.0000 & 0.0000 \end{pmatrix}.$$

From above obtained unfolded core matrices $\mathbf{S}_{(1)}$, $\mathbf{S}_{(2)}$ and $\mathbf{S}_{(3)}$ it is observed the fact that the core tensor is all-orthogonal. For example, the rows of $\mathbf{S}_{(1)}$ are mutually orthogonal and also the matrices formed by the first and the last three columns, as well as the three matrices formed by columns 1 and 4, 2 and 5, 3 and 6.

The core tensors are ordered: its matrices are put in order of descending Frobenius-norm. The Frobenius-norms give the singular values of \mathcal{A} :

mode 1: 7.2051, 5.9551 and 2.1502

mode 2: 9.3063 and 2.3221

mode 3: 7.5284, 5.9433 and 0

The sums of the squared n-mode singular values are all equal to 92, which is the squared Frobenius-norm of \mathcal{A} . The third 3-mode singular value equals 0, since the 3-mode singular vectors (the rows of the left and right sub-matrix of $\mathbf{A}_{(1)}$) only span a two-dimensional space.

4.3 HOSVD and determination of the presence of a non-Gaussian impacting signal

Since the higher-order cumulant of random signals is blind to Gaussian signals, non-Gaussian signals may be detected by constructing the higher-order cumulant of the measured signal. The determination of the presence of a non-Gaussian impacting signal from the observed signals can be achieved through the use of the HOSVD and checking the Frobenius-norm of the n-mode core tensor from the constructed higher order cumulant tensor. In this case, the constructed higher-order tensor is symmetric hence any mode of the Frobenius-norm will suffice to extract the information. Two methods will be discussed and their performances demonstrated.

The major assumptions behind this detection procedure are as follows;

1. The observed signal is stationary up to fourth order.
2. The sample size is ‘sufficiently large’.
3. The unknown impacting signal is non Gaussian and has non zero cumulant at least to fourth order.
4. The system through which the impacting signal passes is linear, time invariant and stable.
5. The additive noise is Gaussian and independent of the input signal.

4.3.1 Construction of the higher order tensor

Suppose we have single observed signal $v(n)$, $n=0,1,\dots,N$, which takes the form of

$$v(n) = \sum_{k=0}^L h(k)x(n-k) + w(n) \quad (4.3.1)$$

in which $h(k)$ is the impulse response of an unknown system, $x(n)$ is the unknown input signal (Gaussian or non-Gaussian), and $w(n)$ is additive Gaussian noise. The observed signal is assumed to have zero mean and is stationary. The k -th order cumulant tensor of the data matrix \mathbf{v} can be constructed from a shifted data matrix construction. One sample sequentially shifted data sets are prepared as follows;

$$v_m(n) = v(n) \cdot \delta(n-m), \quad m = 0, 1, \dots, \tau \quad (4.3.2)$$

where the time lag index m denotes the delay.

From (4.3.2), the data matrix is constructed as

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_\tau \end{bmatrix}, \quad (\tau+1) \times (N-(\tau+1)) \text{ data matrix} \quad (4.3.3)$$

Graphically, the construction of the tensor based on the above formula can be illustrated in the following figure;

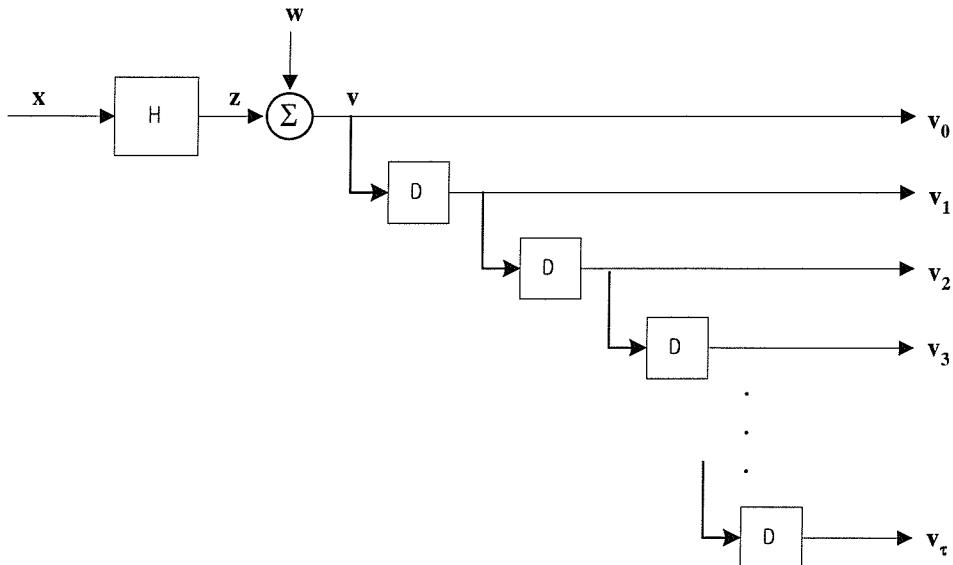


Figure 4.3.1 Construction of the higher order tensor from one sample delayed signals from a single measurement

From this we can obtain the k -th order moment tensor as

$$\mathbf{m}_{k,v} = E\{\mathbf{v}^{(k)}\}, \quad \mathbf{v}^{(k)} = \underbrace{\mathbf{v} \otimes \mathbf{v} \otimes \dots \otimes \mathbf{v}}_{k \text{ times}} \quad (4.3.4)^+$$

or in matrix form as,

$$\mathbf{M}_{k,v} = E\{\mathbf{v}^{(k-1)} \otimes \mathbf{v}^T\} \quad (4.3.5)$$

Assuming the signals are zero mean, the construction of the third-order cumulant matrix of (4.3.5) takes the form of a three dimensional tensor form, which equals the matrix denoted as,

$$\mathcal{C}_{3,v} = \mathbf{M}_{3,v} \quad (4.3.6)$$

Similarly, the fourth-order cumulant tensor $\mathcal{C}_{4,v}$ can be constructed from equation (4.2.11).

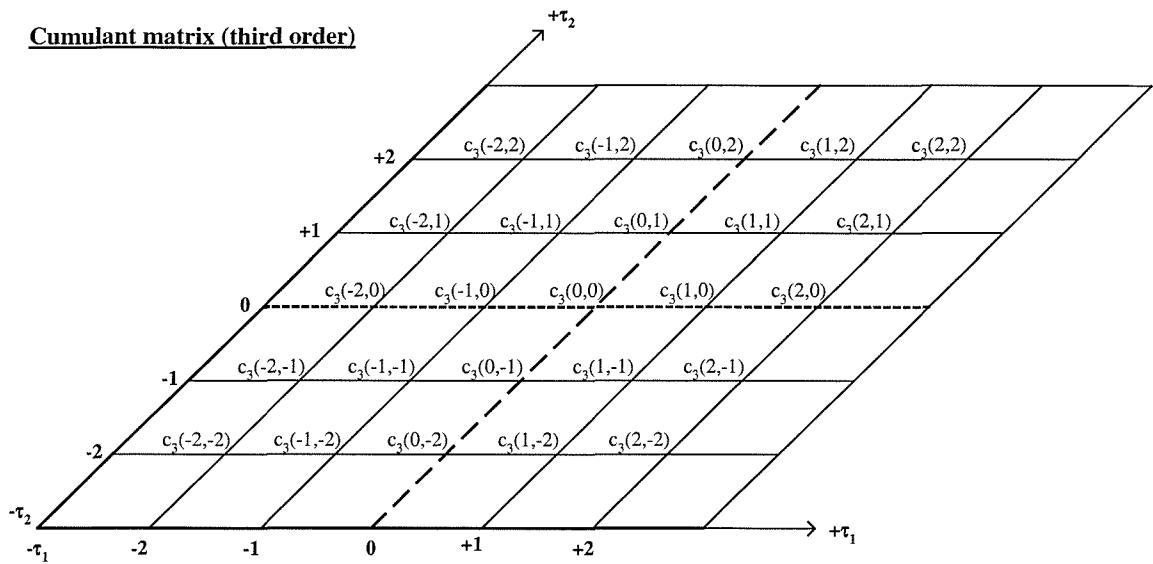
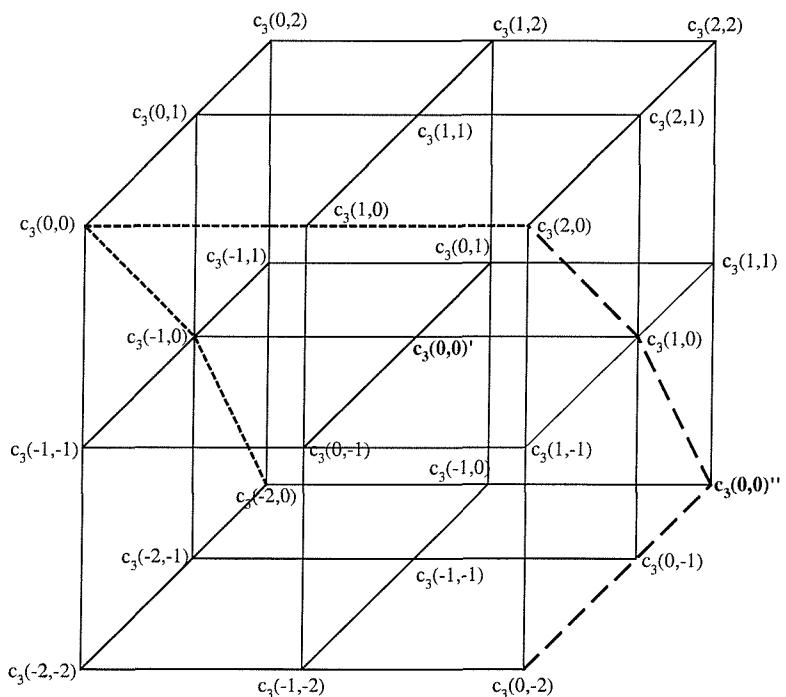
From these constructed higher order cumulant tensors, the higher order singular values $\mathbf{S}_{(n)}^v$ are calculated by the method described in the previous section and the higher order singular values for any mode (symmetric) can be calculated using equation (4.2.30).

A key aspect of detection considers the variation of the statistical parameters of the third and/or fourth order cumulants (before and after the higher order singular value decomposition) by considering the variations of the second order singular values.

The tensor constructed from Figure 4.3.1 takes a similar structure to the higher order non-zero lag cumulant sequences. The tensor, however, possesses more information up to $\tau(\tau^2 - \tau - 1)$, $\tau \geq 2$ than the matrix form of the cumulant sequence. The mapping of the information from the third order cumulant matrix to the third order tensor is shown in the Figure 4.3.2 (the bold faced $c_3(0,0)'$ and $c_3(0,0)''$ for $\tau = 2$ correspond to the additional information).

+ Since the preparation of this thesis the author has been aware of a publication by Rao and Wong where a closely related expansion was presented.

T. Subba Rao and W. K. Wong, *Assymptotics, nonparametrics, and time series*, Chapter 8, Some contributions to multivariate nonlinear time series and bilinear models, edited by S Ghosh, Marcele Dekker, NY (1999).

Cumulant matrix (third order)

Cumulant tensor (third order)

 Figure 4.3.2 Third order cumulant matrix and tensor ($\tau = 2$)

Moreover, the singular value decomposition of the third order cumulant matrix with lag τ is insufficient to reveal significant variation even when there is a non-Gaussian impacting signal. That is to say, the variation of the singular values of the cumulant matrix **cannot discriminate** the difference between a Gaussian and non-Gaussian signal whereas **the third order tensor does**. To illustrate this, the singular values constructed from the cumulant matrix and tensor (from the observed signals) for Gaussian and non-Gaussian input cases are plotted in the following figure;

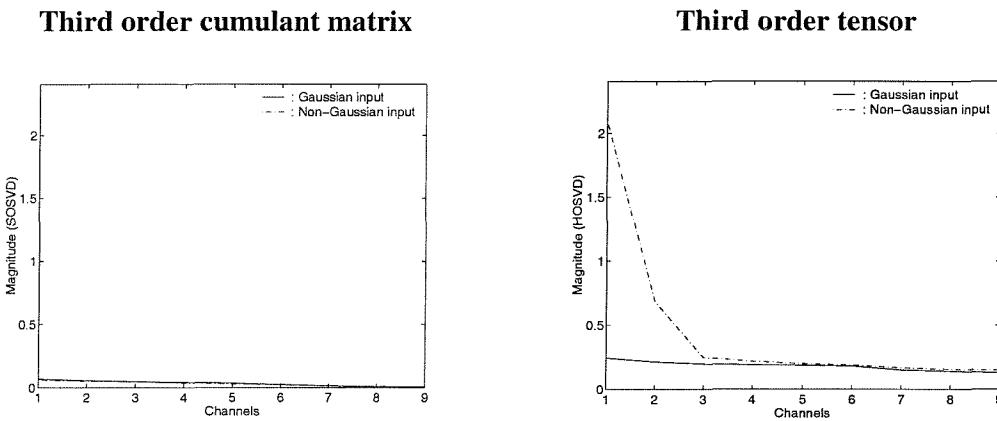


Figure 4.3.3 The singular values from third order cumulant matrix and tensor

In the following subsection, the ability of HOSVD for signal detection, classification and reconstructability problem is demonstrated through simulations.

4.3.2 Preliminaries for simulations

The system for this simulation is shown in the following figure.

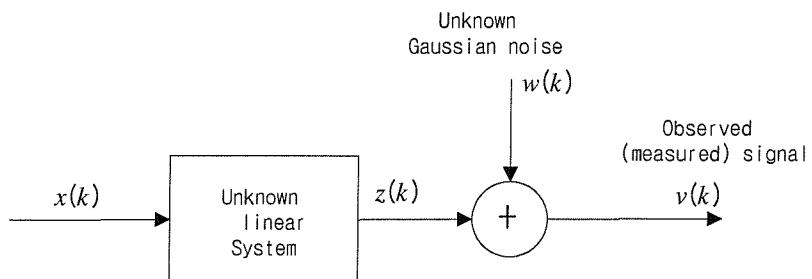


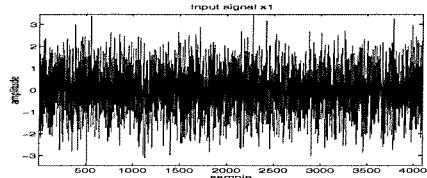
Figure 4.3.4 The signals and system used for non-Gaussian impacting signal detection problem with Gaussian measurement noise

From only the single observed signal $v(k)$ shown in Figure 4.3.4, the construction of higher order (third or fourth) tensor and covariance matrix is carried out using delayed signals (channels). The higher order singular value decomposition and singular value decomposition of covariance matrix (SOSVD) is then carried out to detect non-Gaussian impacting signals, following which the reconstructability of the input signal is considered.

Excitation $x(k)$

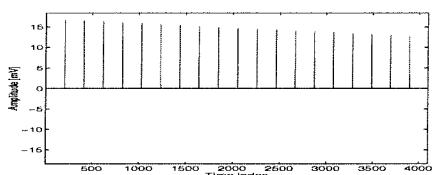
A Gaussian signal and non-Gaussian impacting signals used in this simulation are shown in the Figure 4.3.5.

(A) Input signal type 1 (Gaussian)



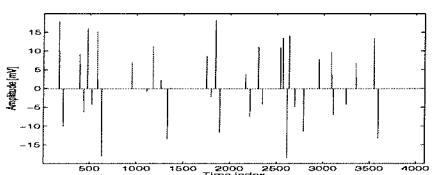
Mean : 0.017
 Variance : 1.0
 Skewness : 0.009
 Kurtosis : 2.94
 Crest factor : 3.86

(B) Input signal type 2 (Uni-directional impacting signal)



Mean : 0.068
 Variance : 0.999
 Skewness : 14.72
 Kurtosis : 219.98
 Crest factor : 16.69

(C) Input signal type 3 (Bi-directional impacting signal)



Mean : 1.4e-16
 Variance : 0.999
 Skewness : 3.05
 Kurtosis : 206.98
 Crest factor : 18.40

Figure 4.3.5 The input signals

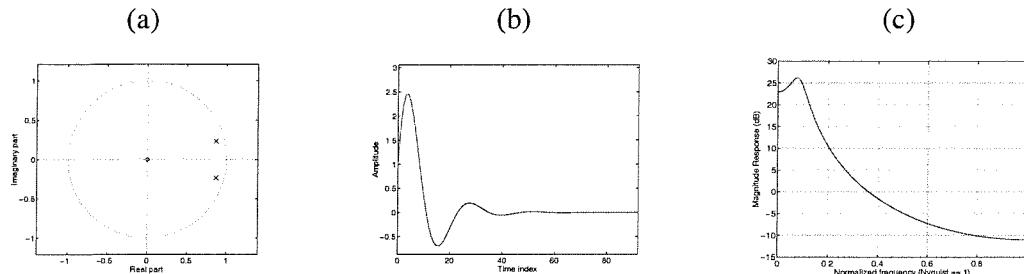
The above input signals are thus labelled as white input (Gaussian, type 1), uni-directional impacting signal (non-Gaussian, type 2), and bi-directional impacting signal (non-Gaussian, type 3) with 4096 sample points and unit variance. The type 2 signal (uni-directional impacting signal) is an idealised model for a mechanical system containing repetitive faults in its operation and the type 3 signal (bi-

directional impacting signal) is typical of signals arising in ultra-sonic examination (the same signal used in Chapter 2).

System

The ‘unknown’ systems selected for detailed discussion here are as follows;

(1) System type A (well damped system)



(2) System type B (highly resonant system)

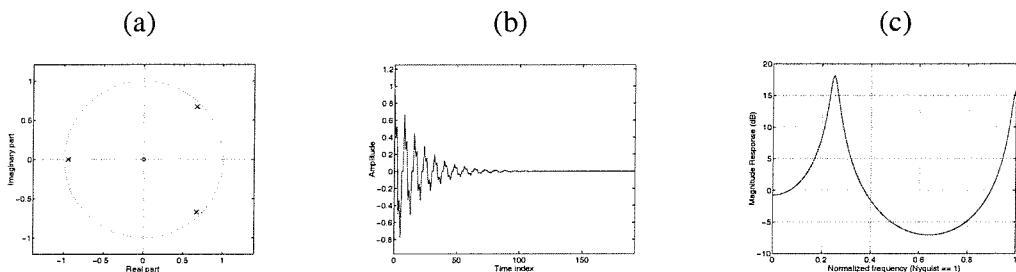


Figure 4.3.6 The systems used in the simulation. (a): System’s pole zero map, (b): impulse response, and (c): Frequency Response Function.

The type A system has a short impulse response function whereas the type B system represents a more complex oscillation with a longer impulse response function.

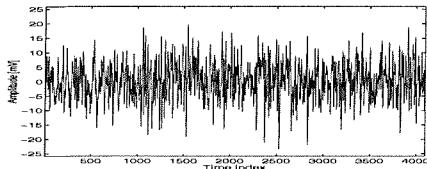
We note that an additional 8 different systems are considered and summarised in the Appendix D.

Noise free output signal $z(k)$

The output of systems A and B for each input signal in the absence of noise is shown in Figure 4.3.7.

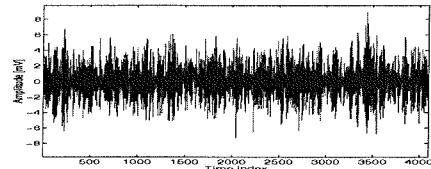
(1) Output signals (noise free, input type 1)

From system A



Mean : 0.151
 Variance : 42.430
 Skewness : -0.082
 Kurtosis : 2.951
 Crest factor : 3.611

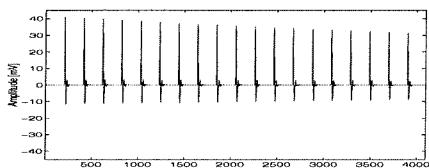
From system B



Mean : 0.009
 Variance : 4.189
 Skewness : 0.020
 Kurtosis : 3.053
 Crest factor : 4.389

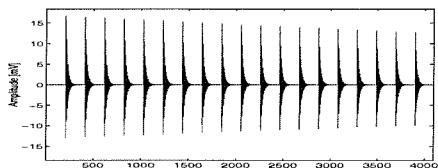
(2) Output signals (noise free, input type 2)

From system A



Mean : 0.954
 Variance : 36.897
 Skewness : 4.227
 Kurtosis : 23.101
 Crest factor : 6.663

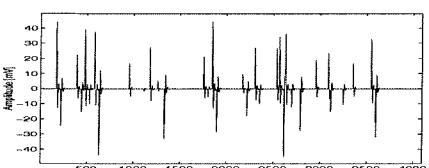
From system B



Mean : 0.062
 Variance : 4.301
 Skewness : 1.28
 Kurtosis : 22.865
 Crest factor : 8.064

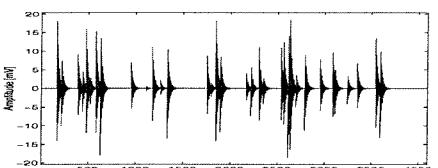
(3) Output signals (noise free, input type 3)

From system A



Mean : 8.4e-5
 Variance : 37.585
 Skewness : 0.928
 Kurtosis : 23.190
 Crest factor : 7.345

From system B



Mean : 0.00
 Variance : 4.308
 Skewness : 0.365
 Kurtosis : 23.420
 Crest factor : 8.935

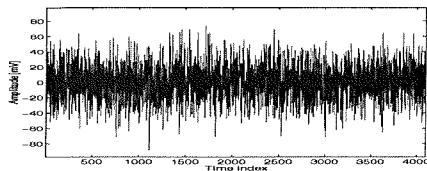
Figure 4.3.7 The output signals

Observed signal $v(k)$

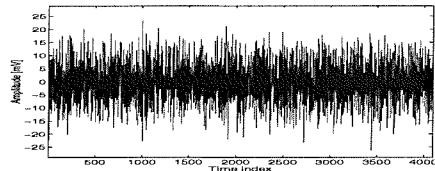
The observed signals are obtained from the output signals including Gaussian noise (SNR=-10 dB) interference in Figure 4.3.8.

(1) Observed signals (SNR=-10dB noise corrupted, input type 1)

From system A



From system B



Mean : 0.00

Variance : 462.88

Skewness : -0.0182

Kurtosis : 3.046

Crest factor : 4.110

Mean : -0.052

Variance : 45.885

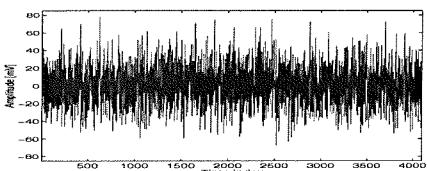
Skewness : -0.013

Kurtosis : 2.981

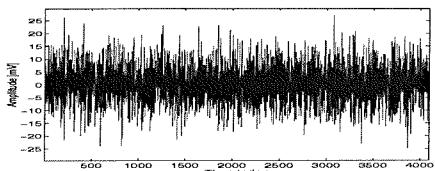
Crest factor : 3.879

(2) Observed signals (SNR=-10dB noise corrupted, input type 2)

From system A



From system B



Mean : 1.155

Variance : 403.698

Skewness : 0.115

Kurtosis : 3.228

Crest factor : 3.839

Mean : 0.386

Variance : 47.398

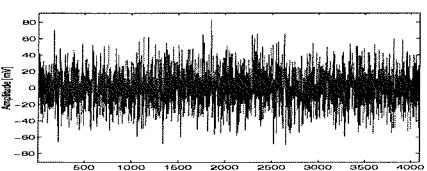
Skewness : 0.025

Kurtosis : 3.102

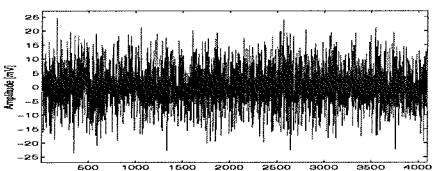
Crest factor : 3.914

(3) Observed signals (SNR=-10dB noise corrupted, input type 3)

From system A



From system B



Mean : 0.204

Variance : 4.12e+2

Skewness : 0.025

Kurtosis : 3.151

Crest factor : 4.069

Mean : 0.0795

Variance : 46.666

Skewness : 0.022

Kurtosis : 3.106

Crest factor : 3.599

Figure 4.3.8 The observed signals

4.3.3 Detection and classification of input signal

Using the three different input signals and two systems, the applicability of HOSVD to the signal detection and classification is tested.

Detection

From the results of extensive simulation, we now propose an empirical detection method. We first compute a threshold level (tr) based on SOSVD.

$$tr = (\text{Max}(\Lambda) - \text{Min}(\Lambda)) \cdot \text{Var}(\overline{\Lambda^2}) \quad (4.3.7)$$

where $\overline{\Lambda^2} = \frac{1}{ch} \sum_{i=1}^{ch} \Lambda(i)^2$ in which 'ch' represents the total number of one sample

delayed signals ($=\tau+1$ in Figure 4.3.1), and Λ denotes the singular values of the covariance matrix.

Next, we compute a detection measure (detector) D_k based on HOSVD where

$$D_k = (\text{Max}(HS_k) - \text{Min}(HS_k)) \cdot \text{Var}(HS_k), \quad k=3 \text{ or } 4 \quad (4.3.8)$$

where $\text{Var}(SK_k) = \frac{1}{ch} \sum_{i=1}^{ch} SK_k(i)^2$ and HS_k are the singular values of the higher

(third or fourth) order tensor.

Finally, D_k is compared to tr to decide the presence or otherwise of a non-Gaussian signal (e.g., the magnitude of D_k in comparison of the tr provides us with the information concerning the existence of a non-Gaussian signal).

Classification

As will be shown very consistent results arise from the previous detection scheme. Building on this we propose a classification based on HOSVD to distinguish between three types of input (Gaussian, non-Gaussian uni-directional impacting, and non-Gaussian bi-directional impacting) and several classes of systems. This follows below;

Class 1: Only Gaussian input signals exist if

$$D_k < tr, \quad k=3 \text{ and } 4$$

Class 2: The input signal is non-Gaussian and uni-directional impulse (type 2 signal) if

$$D_k \geq tr, \quad k=3 \text{ and } 4$$

Class 3: The input signal is non-Gaussian but either the input signal is bi-directional impulse (type 3 signal) or the system is highly resonant if

$$\begin{aligned} D_3 &< tr, \\ D_4 &\geq tr \end{aligned}$$

Simulations and results

The following three examples demonstrate the result of non-Gaussian signal detection and classification using this method. For every case, the observed signal ‘looks’ white and Gaussian. However, HOSVD will be shown to be able to indicate the hidden nature of the input signals.

In the figures that follow, the dotted line in the figure cell labelled ‘detection result’ represents the threshold (tr) calculated from equation(4.3.7) and detection measure (detector) D_k for both third and fourth order is estimated from equation (4.3.8).

Example 1; Gaussian and non-Gaussian impacting signal (uni-directional) with well damped system

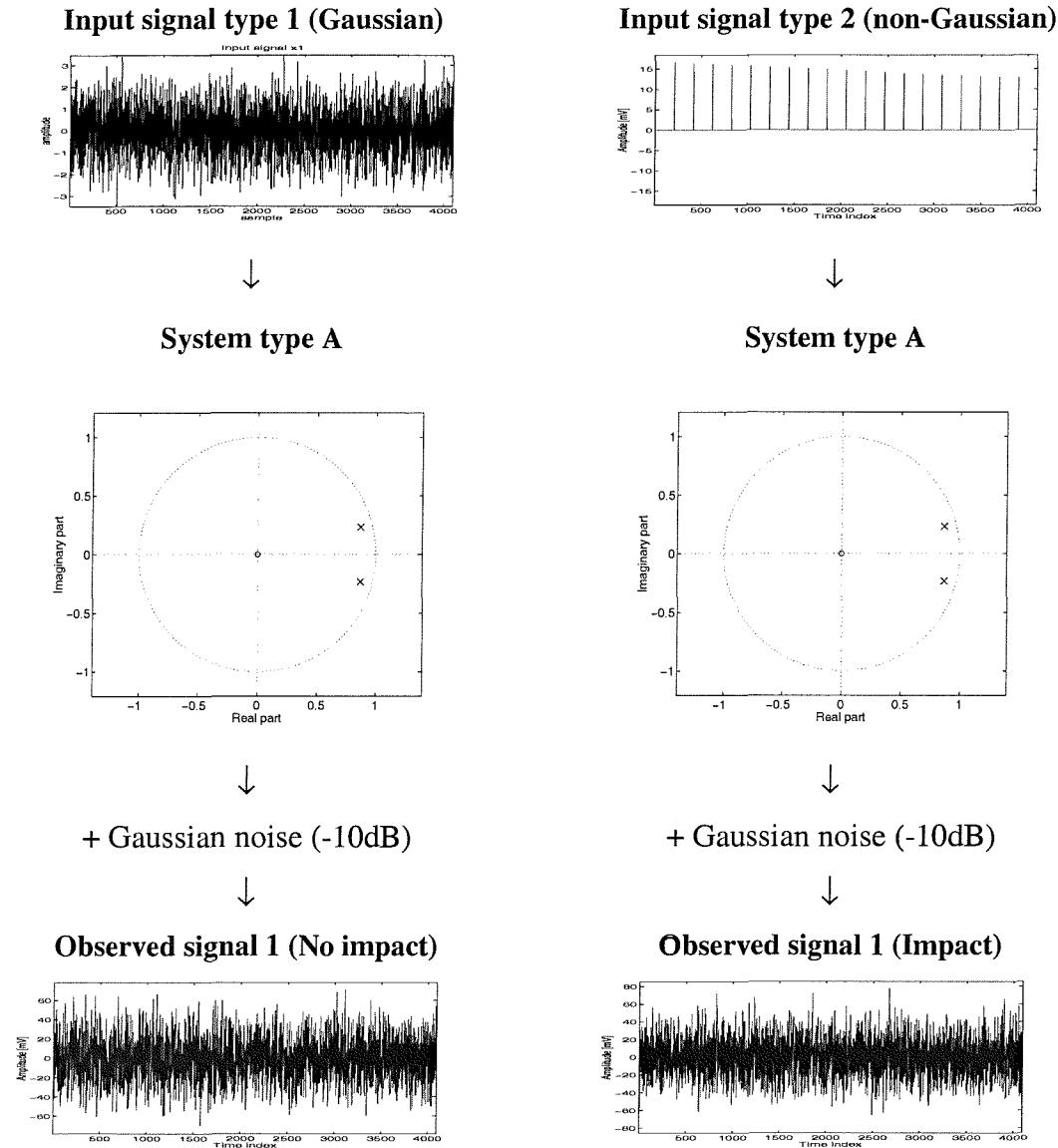


Figure 4.3.9 Signals and system for example 1 (type 1 and type 2 signal with well damped system)

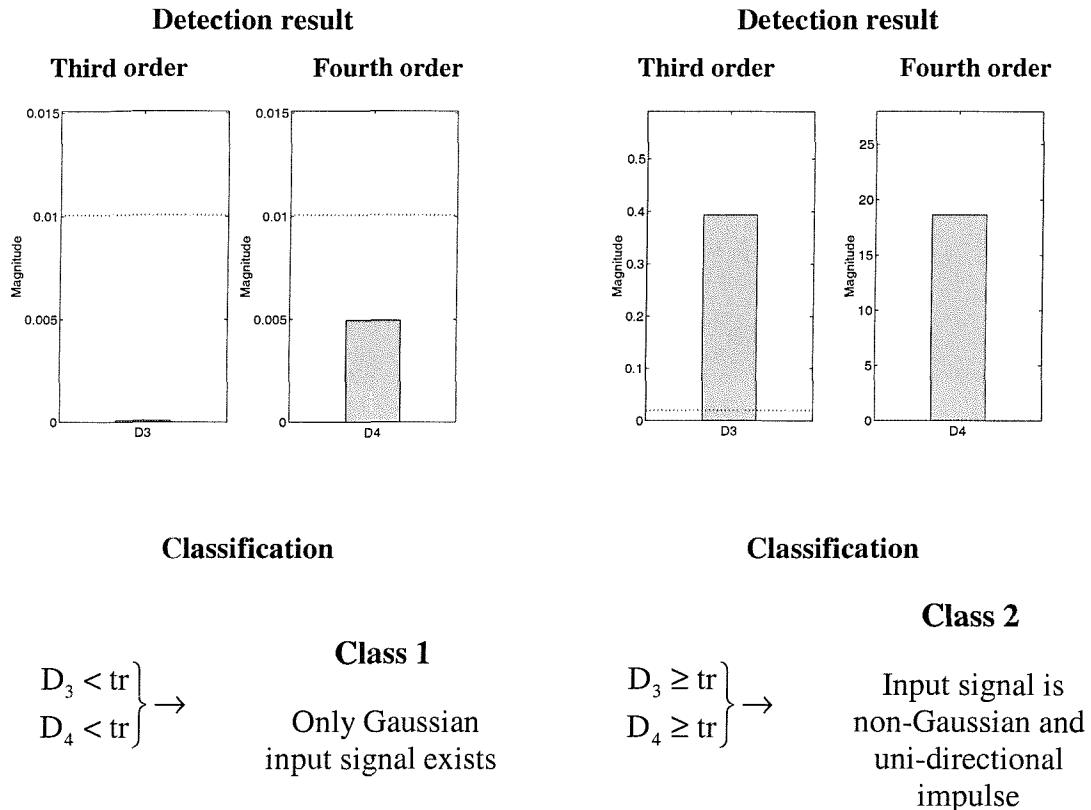


Figure 4.3.10 The detection and classification of signal from HOSVD (type 1 and type 2 signal with well damped system)

In the detection result graph (bar graph), the threshold of detectability calculated from the equation (4.3.7) is displayed by the dotted line. Thus, if the height of each bar is higher than the dotted line, according to our criterion, a certain type of non-Gaussian impacting signal is present in the observed signal. It is clear that for a Gaussian input, none of the indicators reach the dotted line (left side of Figure 4.3.10), thus it belongs to the class 1. On the other hand, when both the third and fourth order detector exceed the dotted line (right side of Figure 4.3.10), the non-Gaussian signal is detected and we can classify the non-Gaussian input signal as a uni-directional impacting signal i.e., class 2.

Example 2; Gaussian and non-Gaussian impacting signal (uni-directional) with highly resonant system

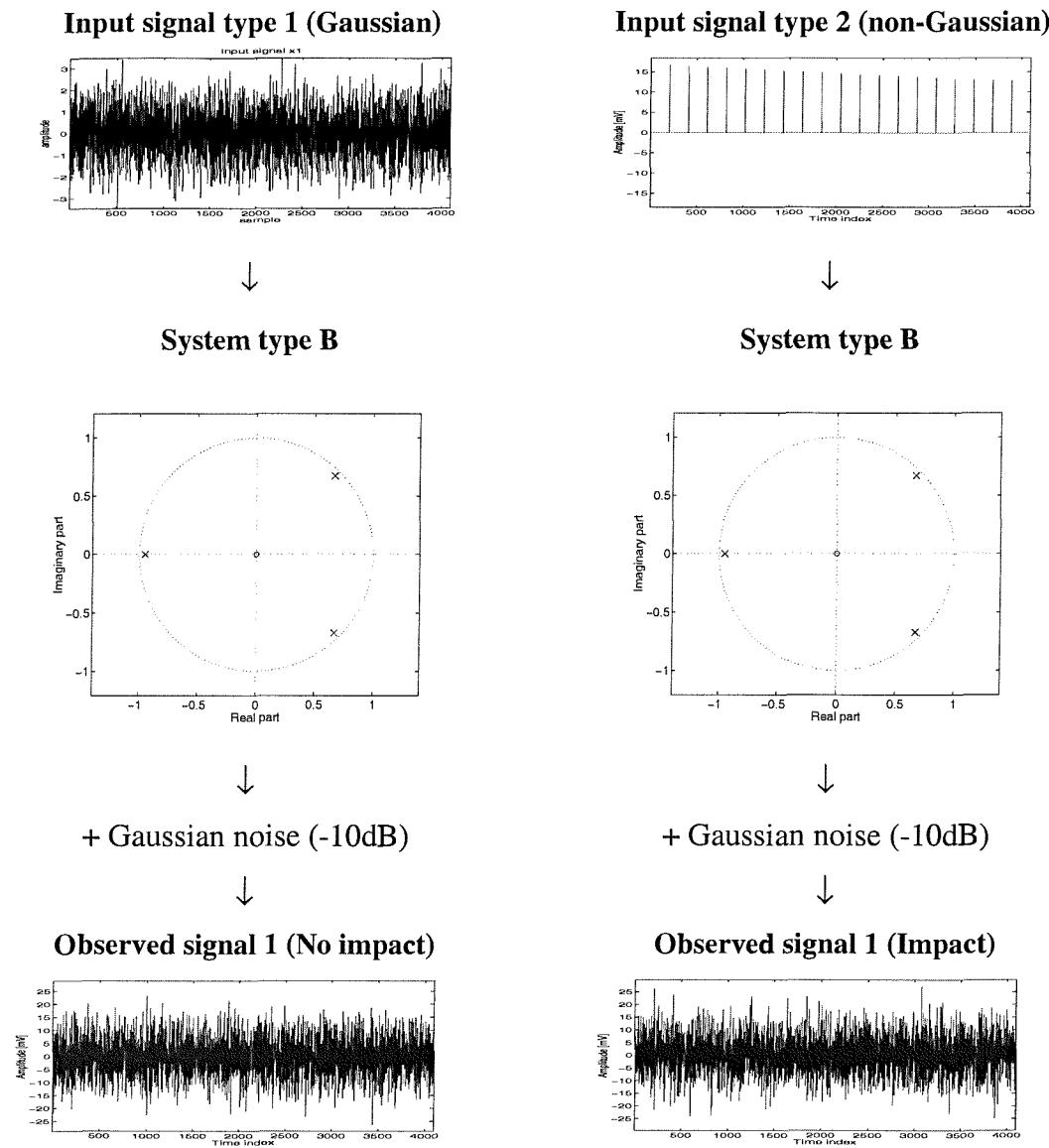


Figure 4.3.11 Signals and system for example 2 (type 1 and type 2 signal with highly resonant system)

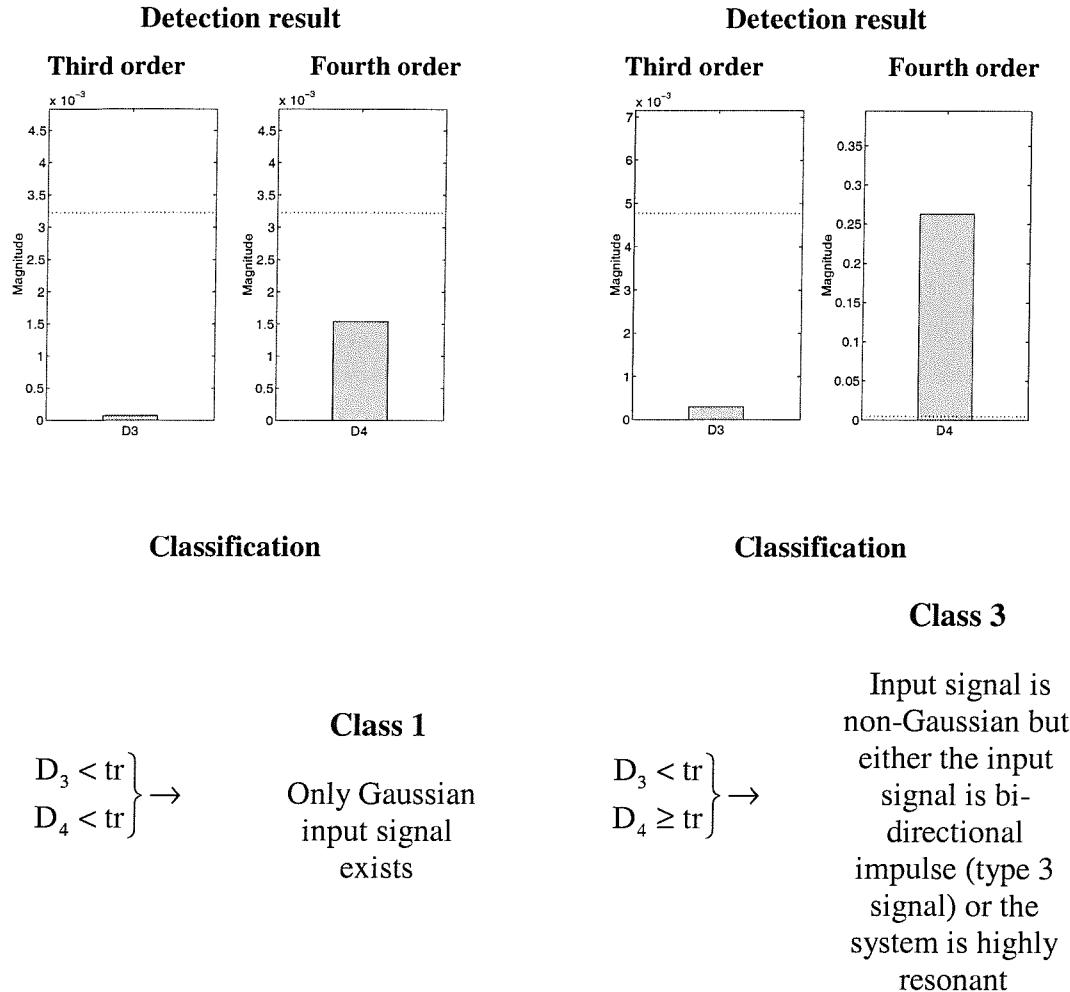
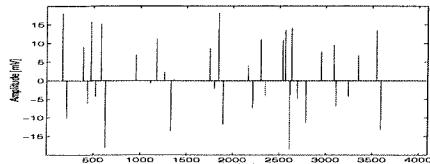


Figure 4.3.12 The detection and classification of signal from HOSVD (type 1 and type 2 signal with highly resonant system)

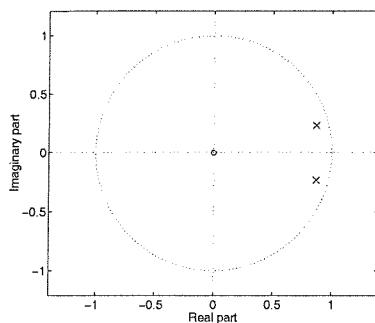
Even for the non-Gaussian impacting signal, the third order detector cannot spot the non-Gaussianity whereas the fourth order does. Since the non-Gaussian impacting signal is uni-directional, the reason for the fact that the third order HOSVD cannot detect the non-Gaussian component is found from the characteristics of unknown system. When a system is highly resonant (type B), then the output of the system becomes close to the symmetric distribution, and in turn the odd order statistical values tends to be closer to zero (see the shape of the output of the system for impacting case in the right column of the Figure 4.3.7). Thus, we obtain two different classes; class 1 as before (left figure) and the new case is class 3 (right figure).

Example 3; Non-Gaussian impacting signal (bi-directional) with well damped and highly resonant systems

Input signal type 3 (non-Gaussian)



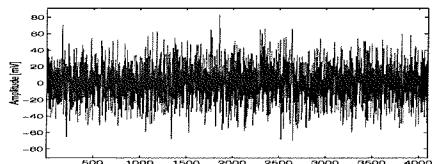
System type A



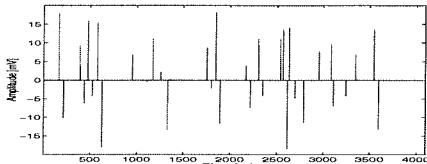
+ Gaussian noise (-10dB)



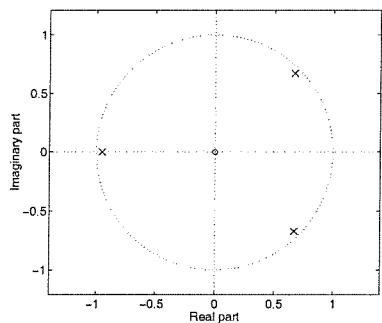
Observed signal 1 (Impact)



Input signal type 3 (non-Gaussian)



System type B



+ Gaussian noise (-10dB)



Observed signal 1 (Impact)

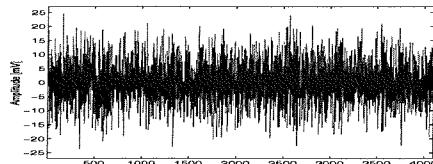


Figure 4.3.13 Signal and systems for example 3 (type 3 signal with well damped and highly resonant systems)

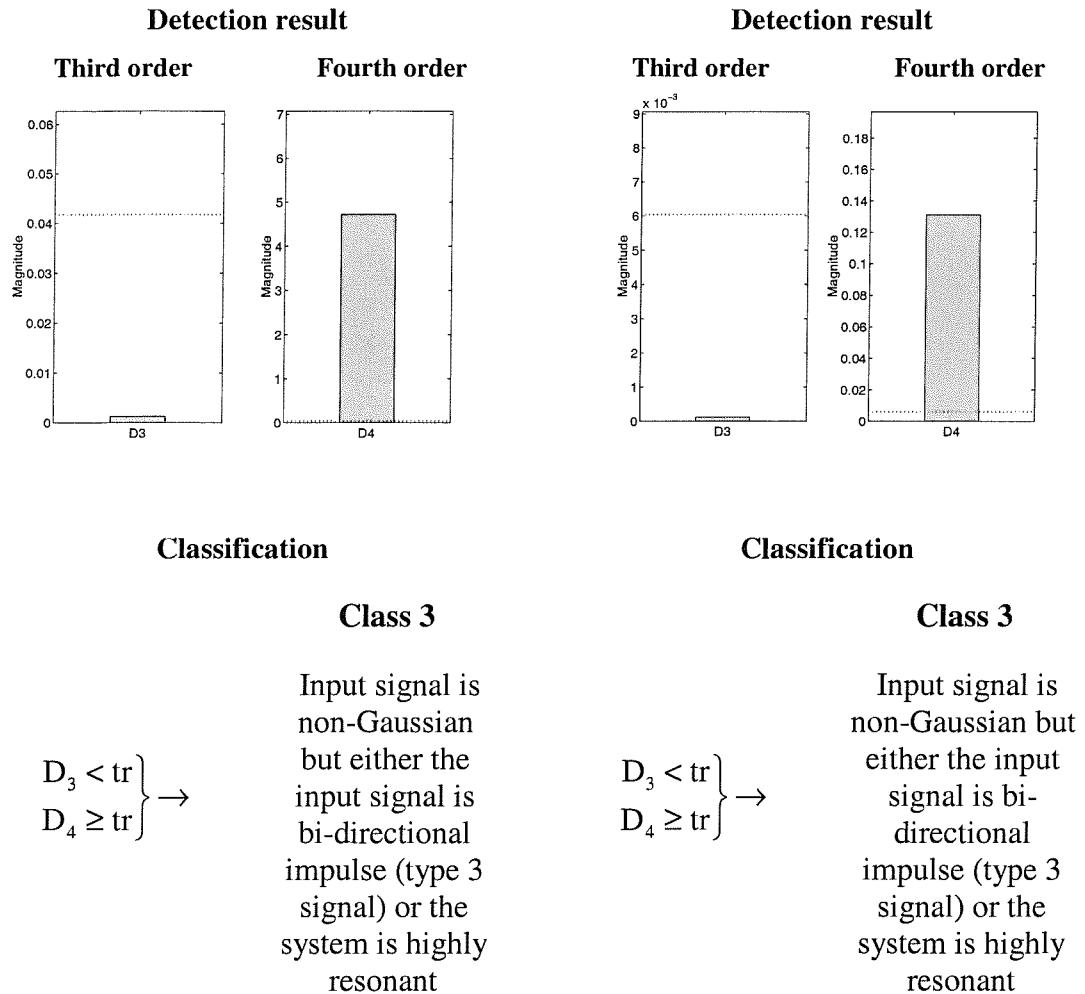


Figure 4.3.14 The detection and classification of signal from HOSVD (type 3 input signal with type A and type B systems)

For the type 3 input signal (bi-directional) case, the third order detector cannot determine the impulsive nature of the input signal regardless of the system. This is because the noise-free output of the system is symmetric (see the signal shapes in the third row of the Figure 4.3.7). However, we do detect the non-Gaussian signal through the fourth order detector. The classification is thus given as class 3 (Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant). As a result, the HOSVD based on the fourth order statistics can provide more consistent detection than the third order.

The above examples demonstrate the significant difference between third and fourth order singular value decomposition in detecting and classifying input signals. Since the third order is blind to symmetric distributions, then when the non-Gaussian input signal has a symmetric distribution or the unknown system is highly resonant (narrow band), the differences of the higher order singular values as compared to the third order tensor may not provide correct results.

4.3.4 Reconstructability assessment from HOSVD

A method for the detection and classification of unknown input signals has been presented. This section is concerned with the next stage of HOSVD application which is the reconstruction of the non-Gaussian signal after it has been detected.

In Chapter 3, we discussed both the Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) blind deconvolution operator (inverse filter). However the choice of FIR is generally enough for practical application as an IIR system can be well approximated by FIR provided the model order L is large enough [Abed-Meraim et al, 1997]. Moreover, as the FIR often leads to simple development of BD process, we restrict ourselves to using FIR inverse filtering leaving the detailed discussions of the filter length to Chapter 5.

Reconstructability factor (R) estimation

1) Highly acceptable signal reconstruction region

From extensive simulations of calculating the higher order singular values, we deduce a relationship between the number of a series of one sample delayed signals (channels) and the maximum value of the singular value ($\text{Max}(\text{HS}_k)$, $k=3$ or 4) for the non-Gaussian input case. That is to say, when a non-Gaussian impacting signal exists, the maximum higher order singular value (from the higher order tensor) increases as the number of channels (number of one sample sequentially delayed signals) increases. To demonstrate this trend, we construct and decompose higher order tensors (third and fourth order) from 2 channels ($\tau=1$) to 9 channels ($\tau=8$) and plot the maximum singular values in the Figure 4.3.15.

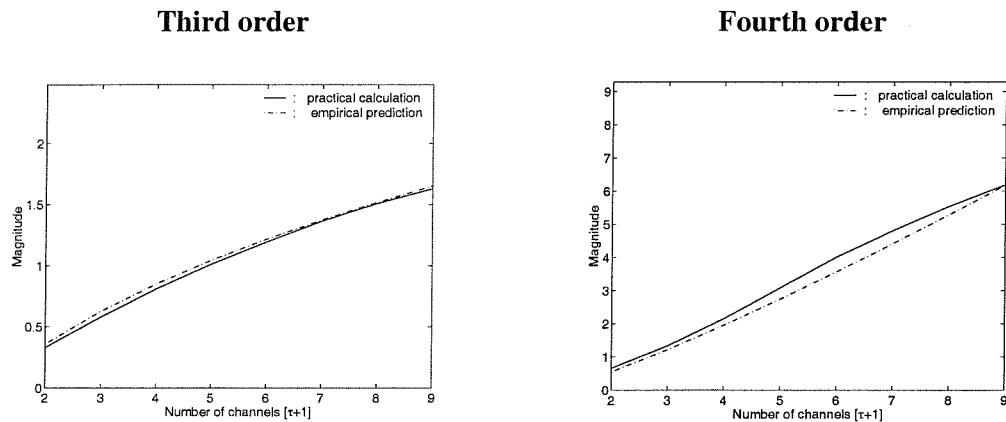


Figure 4.3.15 The relationship between the maximum higher order singular value and number of channels

The solid line of the figure represents the maximum value of the higher order singular values which increases logarithmically as the number of channels increases. The maximum higher order singular value is described approximately by

$$\text{Max}(\text{HS}_k) \propto \log[\tau], \quad k=3 \text{ or } 4 \quad (4.3.9)$$

where τ corresponds to the number of channels.

N.B. note that computational constraints limit the number of channels for the fourth order case to 9. Considerable computational processing power is required to go beyond this.

An empirical result (based on equation (4.3.9)), indicates a ‘sufficient’ condition for reconstructability is deduced as

$$CHS_k = \tau^{m_k} \cdot \log[\tau], \quad k=3 \text{ or } 4$$

where τ is the total number of delayed signals (i.e., the number of channels $= \tau + 1$) and the constants are empirically ‘fitted’ as $m_3 = 0.25$ and $m_4 = 0.85$.

2) Suppression of fluctuation factor (SF)

This is a parameter that reveals the status of the signal from HOSVD and indicate the possibility of signal reconstruction

$$SF_k = \left| \overline{HS_k} - \overline{\gamma_k^v} \right| \cdot [\text{Max}(HS_k) - \text{Min}(\gamma_k^v)], \quad k=3 \text{ or } 4$$

where $\overline{HS_k}$ is the mean of the higher order singular values and $\overline{\gamma_k^v}$ is the skewness or kurtosis of each delayed signal.

3) Reconstructability factor (R)

This is a parameter that can predict the ‘degree’ of signal reconstruction using the FIR inverse filter of the same length as the channel and includes the current status of the observed signal as

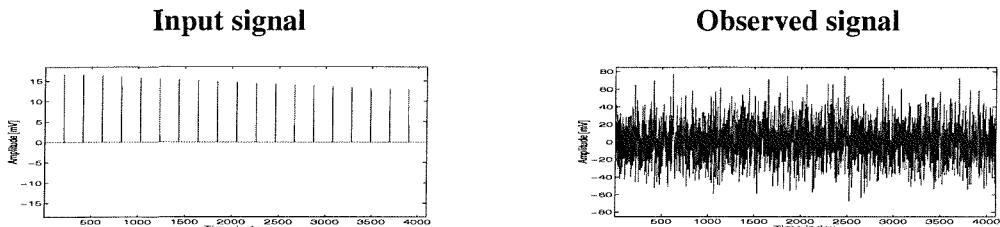
$$R_k = SF_k + \gamma_k^v, \quad k=3 \text{ or } 4$$

where γ_k^v denotes the skewness or kurtosis of the observed signal.

Based on this procedure, some simulations have been carried out and the results and discussions are given as follows;

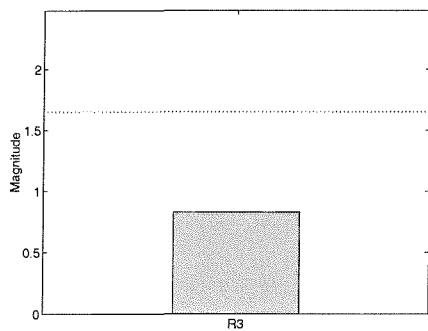
Example 1; Uni-directional impacting signal input with well damped system

Input and observed signals

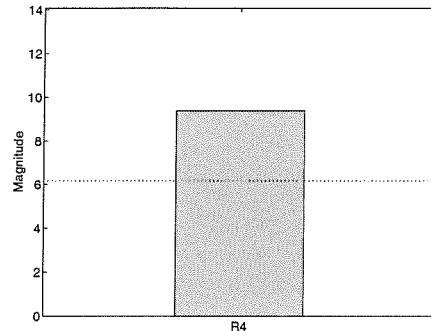


Reconstructability assessment from

Third order

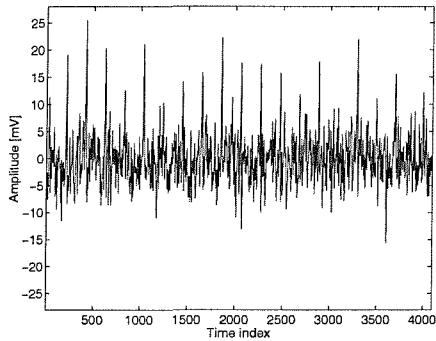


Fourth order



Restored signals from

Third order



Fourth order

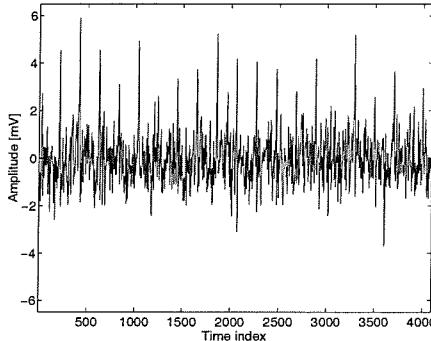
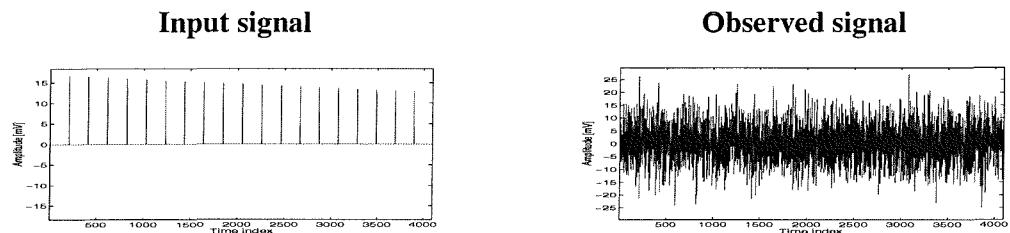


Figure 4.3.16 The reconstructability assessment and restored signals (type 2 input signal with type A system)

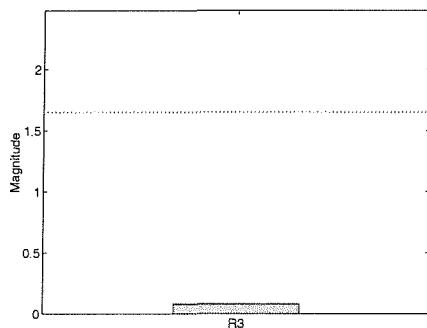
Example 2; Uni-directional impacting signal input with highly resonant system

Input and observed signals

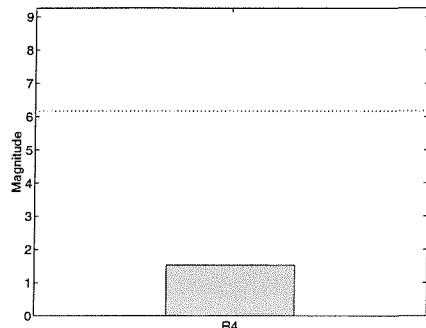


Reconstructability assessment from

Third order

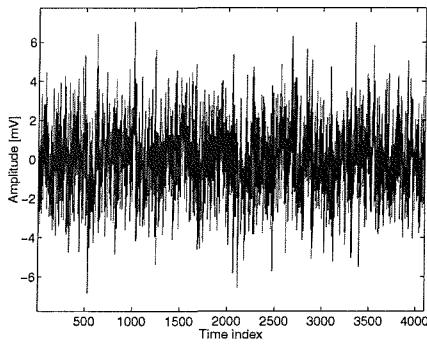


Fourth order



Restored signals from

Third order



Fourth order

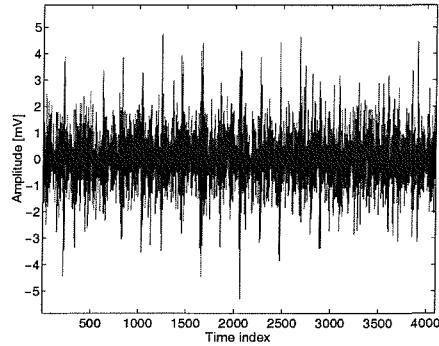
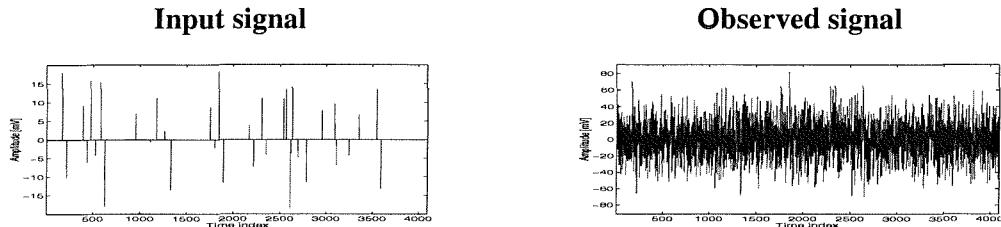


Figure 4.3.17 The reconstructability assessment and restored signals (type 2 input signal with type B system)

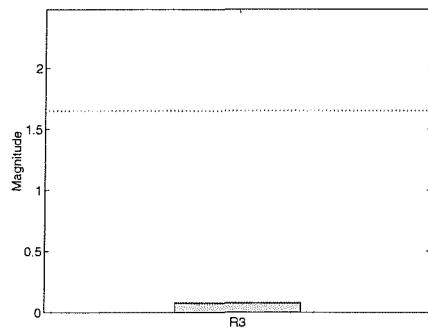
Example 3: Bi-directional impacting signal input with well damped system

Input and observed signals

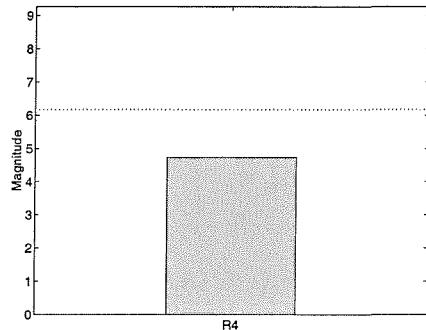


Reconstructability assessment from

Third order

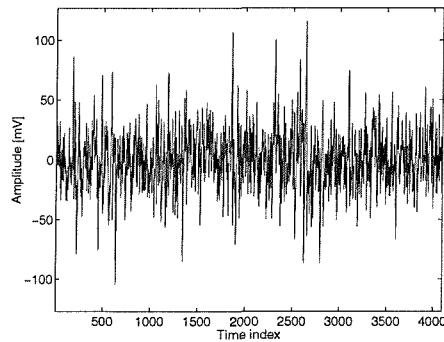


Fourth order



Restored signals from

Third order



Fourth order

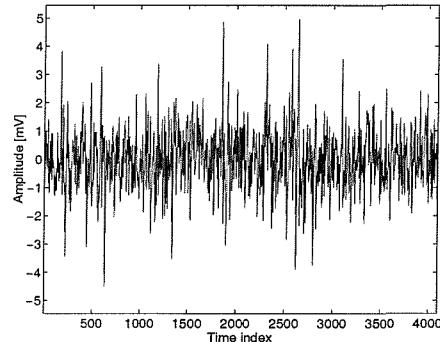
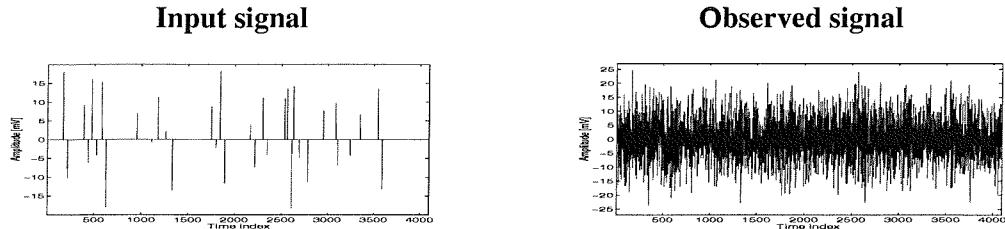


Figure 4.3.18 The reconstructability assessment and restored signals (type 3 input signal with type A system)

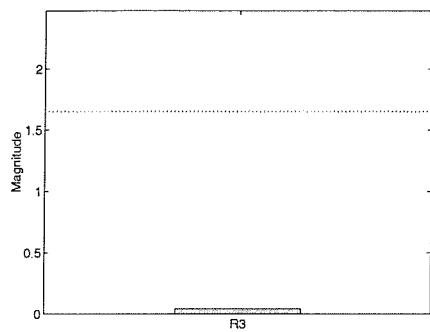
Example 4; Bi-directional impacting signal input with highly resonant system

Input and observed signals

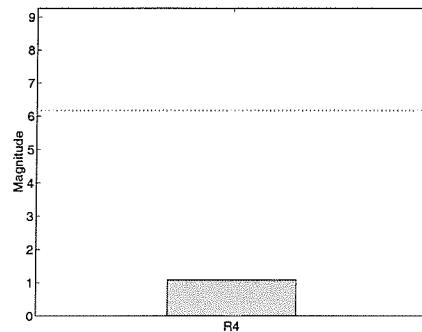


Reconstructability assessment from

Third order

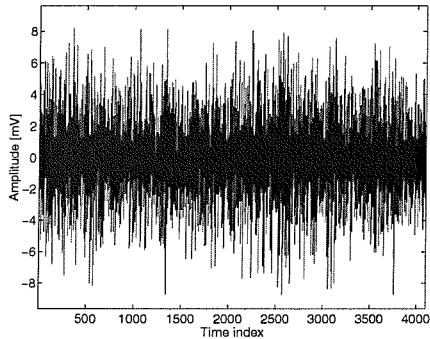


Fourth order



Restored signals from

Third order



Fourth order

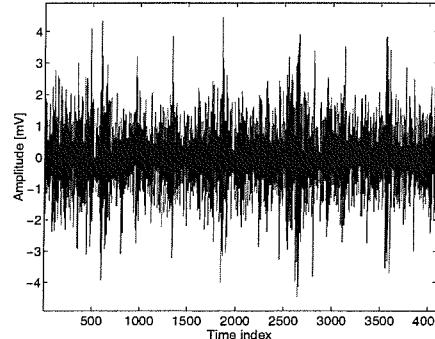


Figure 4.3.19 The reconstructability assessment and restored signals (type 3 input signal with type B system)

Discussion of the results

1) General descriptions; (Figure 4.3.16 - Figure 4.3.19)

In the top row of each figure, a non-Gaussian impacting signal and the observed signal (SNR=-10 dB, Gaussian noise corruption) are displayed. In the middle row, the dotted line in each figure indicates the sufficient condition for the reconstructability of the non-Gaussian impacting input signal when the FIR inverse filter has the same length as the number of delayed signals (channel). Relative to the dotted line, the bar graph represents the degree of reconstructability of the input signal through blind deconvolution. The larger the magnitude of the bar the 'better' the reconstruction of the input signal. To confirm this, each restored signal using both the third and fourth order BD process is plotted in the bottom of each figure.

2) Example 1; (Figure 4.3.16)

When the input signal is an uni-directional impacting signal (type 2 signal in Figure 4.3.5) and the system is well damped (type A system in Figure 4.3.6), the reconstructability assessment is (relatively) high for both the third and fourth order cases. As can be in the bottom row of the figure, both the restored signals certainly indicate the impacting and so support this assessment.

3) Example 2; (Figure 4.3.17)

When the input signal is an uni-directional impacting signal (type 2 signal in Figure 4.3.5) and the system is highly resonant (type B system in Figure 4.3.6), the reconstructability assessment is similar to the detection case. That is to say, the highly resonant system degrades the potential reconstructability using the third order BD process. The assessment and restored signal by the fourth order in contrast is less affected and gives reasonably clear indication of impacting.

4) Example 3; (Figure 4.3.18)

Unlike the detection case, when the input signal is a bi-directional impacting signal (type 3 signal in Figure 4.3.5) and the system is well damped (type A system in Figure 4.3.6), the reconstructability criterion for the third order indicates

difficulties in reconstruction. However, application of the BD process produced encouraging results. In fact, the outcome from the third order is similar to that of the fourth order case. This indicates that the reconstructability criterion is too severe.

5) Example 4; (Figure 4.3.19)

When the input signal is a bi-directional impacting signal (type 3 signal in Figure 4.3.5) and the system is highly resonant (type B system in Figure 4.3.6), the reconstructability assessment for the third order is low and the restored signal is unacceptable as compared to the fourth order which does give an indication of impacting.

Observation

From this simulation study, we conclude (i) that the empirically derived criteria for input reconstruction perhaps overly restrictive and (ii) that detection, classification and reconstruction using HOS is more affected by the system characteristics than the nature of non-Gaussian impacting signals.

4.4 Summary and conclusions

This chapter describes the application of higher order statistics through the construction of higher order tensors and their singular value decomposition. The higher order tensor is a multi-dimensional extension of a matrix retaining the same properties of a matrix (e.g. multilinearity and symmetry) and additionally possesses the merit of suppressing Gaussian signals.

We have tested the ability of HOSVD for detection, classification and reconstructability of non-Gaussian signals through various simulations. The aim of this contribution is to put HOSVD/tensors into a practical context. This has included computational experiments and the deduction of empirical criteria. Specifically, the methods used and results are summarised as follows;

- From a single measured signal, the sequentially delayed signals (channels) are used to construct the higher order tensor (third or fourth order). From the constructed tensors, the higher order singular values are estimated. The essence of non-Gaussian signal detection is based on the comparison of the second order singular values and higher order singular values.
- A threshold and variance comparison of higher order singular values (up to fourth order) enables us to detect non-Gaussian signal under various system characteristics. Also, the threshold can be useful in implementing the on-line automated detection of a non-Gaussian (impacting) signal.
- By comparing the detectability from the third and fourth order, a classification of signals and systems is achieved.
- Following this a formulation based on HOSVD provides a guide to input signal reconstructability.
- From the simulations with various systems, we conclude that the detection, classification and reconstructability assessment using HOSVD can be a useful tool for blind processing of impacting process.

In conclusion, this chapter has demonstrated, through experimental simulation, that HOSVD has considerable promise for detecting impacts in high Gaussian noise levels. A cautionary note, however, is that the methods are complicated and if the presence of impacting can be determined by other means (even visually in low noise environments) obviously this should be exploited. Furthermore the robustness of this method to, for example, non-stationary and other form of non-Gaussianity needs further investigation.

PART III Practical considerations for blind deconvolution

Chapter 5

Single Channel Blind Deconvolution

5.1 Introduction

A single input single output (SISO) Blind Deconvolution (BD) process based on cumulant maximisation requires the input signal to be non-Gaussian. The term 'blind' considerably restricts the 'quality' of the reconstructed impacting signal. This results in ambiguities relating to scale and time delay. Another difficulty encountered is the selection of the acceptable length of the inverse filter and the form of initial inverse filter in the iterative process. The length of inverse filter can be very different depending on whether its impulse response is finite (FIR system) or infinite (IIR system). However, as already discussed in Chapter 3, there are many advantages to using a FIR system, and this is employed here.

In this chapter, we reconsider FIR blind deconvolution introduced in Chapter 3 and two specific aspects are considered in detail namely (i) initialisation of the filter and (ii) criteria for selection of filter length.

5.2 Descriptions of model and performance measures

This section defines the model used for the study, i.e., the non-Gaussian input signal, the unknown system and the noise.

5.2.1 Model (single input single output system)

The situation is depicted below:

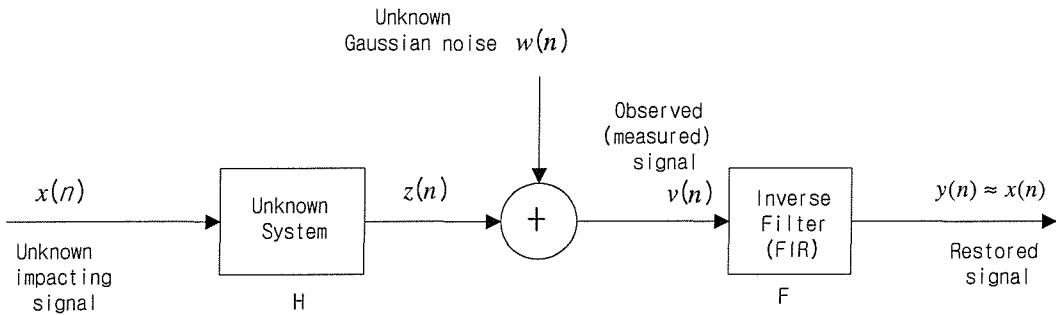


Figure 5.2.1 The model of blind source reconstruction problem (single input single output case)

The observation $v(n)$ is described by

$$v(n) = \sum_{k=0}^{\infty} h(k)x(n-k) + w(n) \quad (5.2.1)$$

where the $h(k)$ designates the impulse response of the unknown system (MA, AR, or ARMA), $x(n)$ is the non-Gaussian input and $w(n)$ is Gaussian noise. The inversion is achieved by an FIR filter yielding the ‘restored’ signal $y(n)$

$$\begin{aligned}
 y(n) &= \sum_{k=0}^L f(k)v(n-k) \\
 &= \sum_{k=0}^L f(k) \sum_{m=0}^{\infty} h(m)x(n-m) + \sum_{k=0}^L f(k)w(n-k) \\
 &\approx x(n)
 \end{aligned} \quad (5.2.2)$$

where $f(k)$ represents the impulse response (length L) of the inverse system (called as ‘BD operator’). As expressed in the equation (5.2.2), the restored signal $y(n)$ inevitably contains filtered noise components and these unwanted noise

components cannot be completely eliminated, even if the noise is Gaussian.

5.2.2 Performance measures for the blind deconvolution process

In order to assess the results of blind deconvolution (signal reconstruction) we define functions related to the status of the restored signals as well as the objective function itself. We introduce five performance measures. These are referred to as; The Sum of Squared Deviations (SSD), the Spikiness Index (SI), Inverted performance index (I_p), Shape parameter (α coefficient), and Equivalent spikiness (Entropy, E) of signal. The last three parameters are also used in the determination of the inverse filter length. Each of them is described below.

Sum of Squared Deviations (SSD)

Measures which can represent the performance of blind deconvolution process can be introduced in several ways. A simple and obvious measure is Sum of Squared Deviations (SSD) which compares the performances of each deconvolution method and is expressed as

$$SSD = \sum_{k=1}^N [y'(k) - x(k)]^2 \quad (5.2.3)$$

where $y'(k)$ is the delay compensated restored signal which can be obtained by estimating the cross-correlation of the restored signal $y(n)$ and input signal $x(n)$. However, this requires knowledge of the input signal, and hence is restricted to the simulation case only.

Inverse Performance (I_p) index

Blind deconvolution is based on the maximisation of an appropriately selected objective function of the output $y(n)$ of an inverse system \mathbf{f} . This objective function used in this study takes the form of either the normalised third order cumulant (skewness) or the fourth order cumulant (kurtosis) expressed as

$$O_y(r, 2) = \frac{\sum_{i=0}^{N-1} y(n)^r}{\left[\sum_{i=0}^{N-1} y(n)^2 \right]^{r/2}} \quad (5.2.4)$$

and r is either 3 or 4. In the same manner, the normalised higher order cumulant of the measured signal $v(n)$ is expressed

$$O_v(r, 2) = \frac{\sum_{i=0}^{N-1} v(n)^r}{\left[\sum_{i=0}^{N-1} v(n)^2 \right]^{r/2}} \quad (5.2.5)$$

For simplicity these values are denoted as O_y and O_v .

The *inverse* performance index is calculated by inverting the absolute difference of the above values

$$I_p = |O_y - O_v|^{-1} \quad (5.2.6)$$

This index indicates the performance of blind impacting signal reconstruction in accordance with the SSD (i.e., a lower I_p implies a better reconstruction result).

Shape parameter (α coefficient)

Another measure that can represent the statistical status of a signal is considered by using the *Generalised Gaussian Distribution* [Miller and Thomas, 1972].

This Generalised Gaussian Distribution (GGD) of a random signal $y(n)$ is expressed via two parameters such as α and β which are defined below

$$f(y, \beta, \alpha) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} e^{-(\frac{|y|}{\beta})^\alpha} \quad (5.2.7)^+$$

where

$-\infty < y < \infty$ is a random signal

$\Gamma(\bullet)$ is the gamma function

$\beta > 0$ is the scale parameter

$\alpha > 0$ is the shape parameter.

This GGD includes a wide range of distributions, e.g.,

$$\alpha = 1 : f(y, \beta, 1) = \frac{1}{2\beta} e^{-\frac{|y|}{\beta}}, \quad -\infty < y < \infty \quad ; \text{Laplacian}$$

$$\alpha = 2 : f(y, \beta, 2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}, \quad -\infty < y < \infty \quad ; \text{Gaussian}$$

$$\alpha \rightarrow \infty : f(y, \beta, \infty) = \frac{1}{2\beta}, \quad -\beta < y < \beta \quad ; \text{Uniform}$$

$\alpha \rightarrow 0$: a certain event, i.e., the chances of finding an event in a finite samples goes to zero ; highly spiky signal case

Like the performance index, the effectiveness of the deconvolution for a spiky signal reconstruction problem can be measured by examining the α value.

+ The reference, *The advanced theory of statistics*, Volume 1 Distribution theory, by Kendal, M. G. and Stuart, A. give a more comprehensive classification scheme.

In fact, the higher order statistical value and the α coefficient of a the random signal y , has a close relationship [Gray, 1979] specifically;

$$\frac{\frac{1}{N} \sum_{i=1}^N |y_i|^{2\rho}}{\left| \frac{1}{N} \sum_{i=1}^N |y_i|^\rho \right|^2} \cong \frac{\Gamma\left[\frac{2\rho+1}{\alpha}\right] \Gamma\left[\frac{1}{\alpha}\right]}{\Gamma\left[\frac{\rho+1}{\alpha}\right]^2} \quad (5.2.8)$$

$\rho = 2$ corresponds to the kurtosis of the random signal y .

The relationship between the kurtosis and the α coefficient expressed in the equation (5.2.8) is plotted in the following figure;

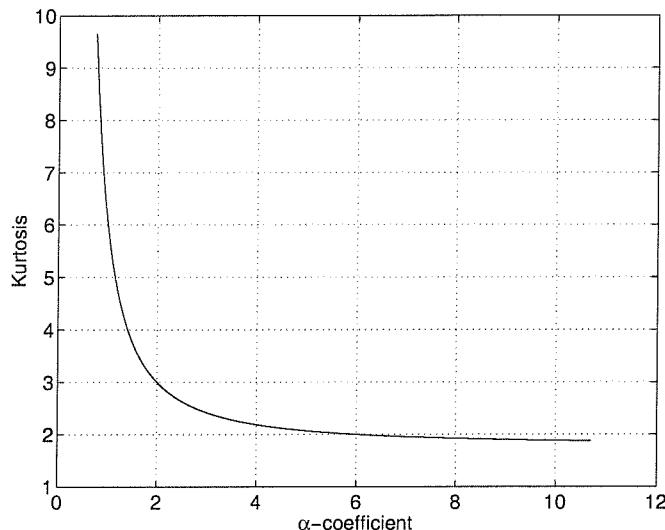


Figure 5.2.2 The relationship between the kurtosis and α coefficient of random signal y .

Equivalent spikiness (entropy of the signal, E)

The observed signal is considered as being linearly filtered through an unknown system and is corrupted by Gaussian noise. When seeking an impacting signal through an iterative inverse filter calculation, the status of the deconvolved signal (the output of the inverse filter) can be monitored by checking the probabilistic characteristics of the signal at each iteration. The impacting signal has several

randomly located large events separated by many near-zero events that can be interpreted as noise. Under the constraint such that the variance of the observed signal and its estimated impacting signal (deconvolved signal) are equal, then the probability distribution of the deconvolved signal will have less entropy [Gray, 1979]. The constraint required above can be achieved by setting the variance of the measured and restored signal to be unity.

$$\frac{\sum_{k=0}^{N-1} y(k)^2}{\sum_{k=0}^{N-1} v(k)^2} = 1 \quad (5.2.9)$$

where $v(k)$ and $y(k)$ designates the measured and restored signals, respectively.

The entropy of the signal y is calculated from its probability function as,

$$E = - \int_{-\infty}^{\infty} f(y) \ln f(y) dy \quad (5.2.10)$$

For our purpose, we calculate the entropy for a discrete time signal. This calculation will be commonly used throughout this thesis.

Assuming the expected data range is between $-3 \cdot \sigma_y + \mu_y \leq y \leq 3 \cdot \sigma_y + \mu_y$, the pdf range is split into equal steps of $0.1 \cdot \sigma_y$. The total number of data falling into each step is counted and divided by the number of data N to give

$$E = - \sum [f(y) \ln f(y)] \quad (5.2.11)$$

This value decreases as the deconvolution process approaches the true impacting signal, and thus, possesses similar characteristics to that of the α curve.

The I_p index, α -coefficient and Entropy (E) are used as major elements of criteria for the inverse filter length determination together with the performance measures of the blind reconstruction of impacting signals.

We first consider the effect of initialisation of the inverse filter coefficients in its iterative calculation procedure, and follow this with consideration of inverse filter length determination.

5.3 Effect of initial filter impulse response on deconvolution

As already discussed in Chapter 3, the inverse filter is calculated from

$$\mathbf{R}_{vv} \cdot \mathbf{f} = \mathbf{g} \quad (5.3.1)$$

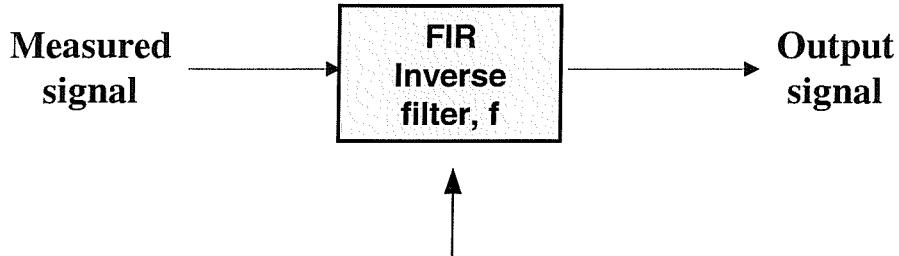
where \mathbf{R}_{vv} denotes the symmetry $L \times L$ autocorrelation matrix of the observed signal, \mathbf{f} is $L \times 1$ inverse filter coefficient vector, and \mathbf{g} is $L \times 1$ cross-correlation vector between the observed signal and the output of the inverse filter. Using either the constrained or normalised objective function maximisation process, the output signal $y(n)$ can yield input signal restoration through the convolution of the measured signal $v(n)$ and the inverse filter \mathbf{f} with length L

$$y(n) = \sum_{m=0}^{L-1} f_m v(n-m) \quad (5.3.2)$$

The equation (5.3.1) is solved in an iterative manner as the equation is non-linear. Thus, at the first stage of maximising the objective function for the blind deconvolution procedure, an initial inverse filter has to be selected.

The aim of this study is to see the effect of this initial inverse filter selection.

Following figure suggests three possible initial inverse filter types (FIR) used in the iterative calculation.



- (a) $f_{ini} = [\text{Random set of filter coefficients}]$
 $= [0.75 \ -0.1 \ 0.3 \ \dots \ -0.12 \ \dots \ 0.25]$
- (b) $f_{ini} = [\text{Impulse type initial filter coefficient}]$
 $= [1 \ 0 \ 0 \ \dots \ 0]$
- (c) $f_{ini} = [\text{Centred impulse initial filter type}]$
 $= [0 \ 0 \dots 1 \ \dots 0 \ 0]$

Figure 5.3.1 Different types of initial inverse filter for deconvolution

According to the central limit theorem and assuming the measured signal is not highly non-Gaussian, any arbitrarily selected inverse filter may make the output signal closer to Gaussian. Thus, from Figure 5.3.1, if an initial inverse filter is chosen randomly, the statistical property of the output of the inverse filter is liable to be *closer to Gaussian* whereas the other initial inverse filters (b) and (c) will not change the statistical properties of the output signal at the first stage of iteration.

We now perform simulations to support the above statements.

Simulation with three different initial inverse filter types

To study the effect of filter initialisation, we select an unknown input signal and an ARMA(2,3) system (unknown) as shown in Figure 5.3.2.

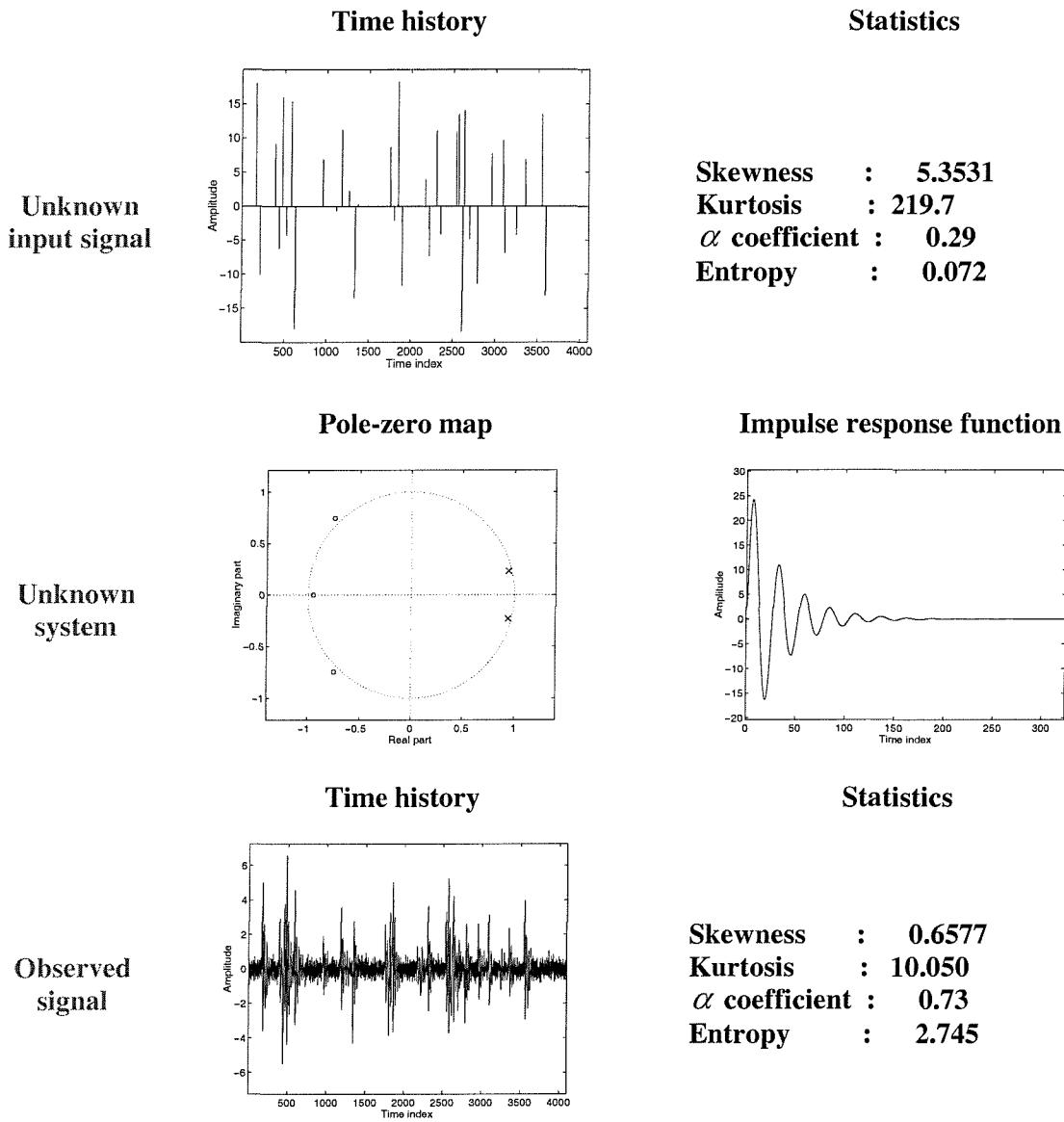


Figure 5.3.2 Sample problem of BD process (unknown input signal, system and observed signal)

From above observed signal, blind deconvolution was carried out using three different initial inverse filter types (the filter length was kept the same). During

the iterative calculation of the inverse filters the values of objective function were monitored and compared to each other.

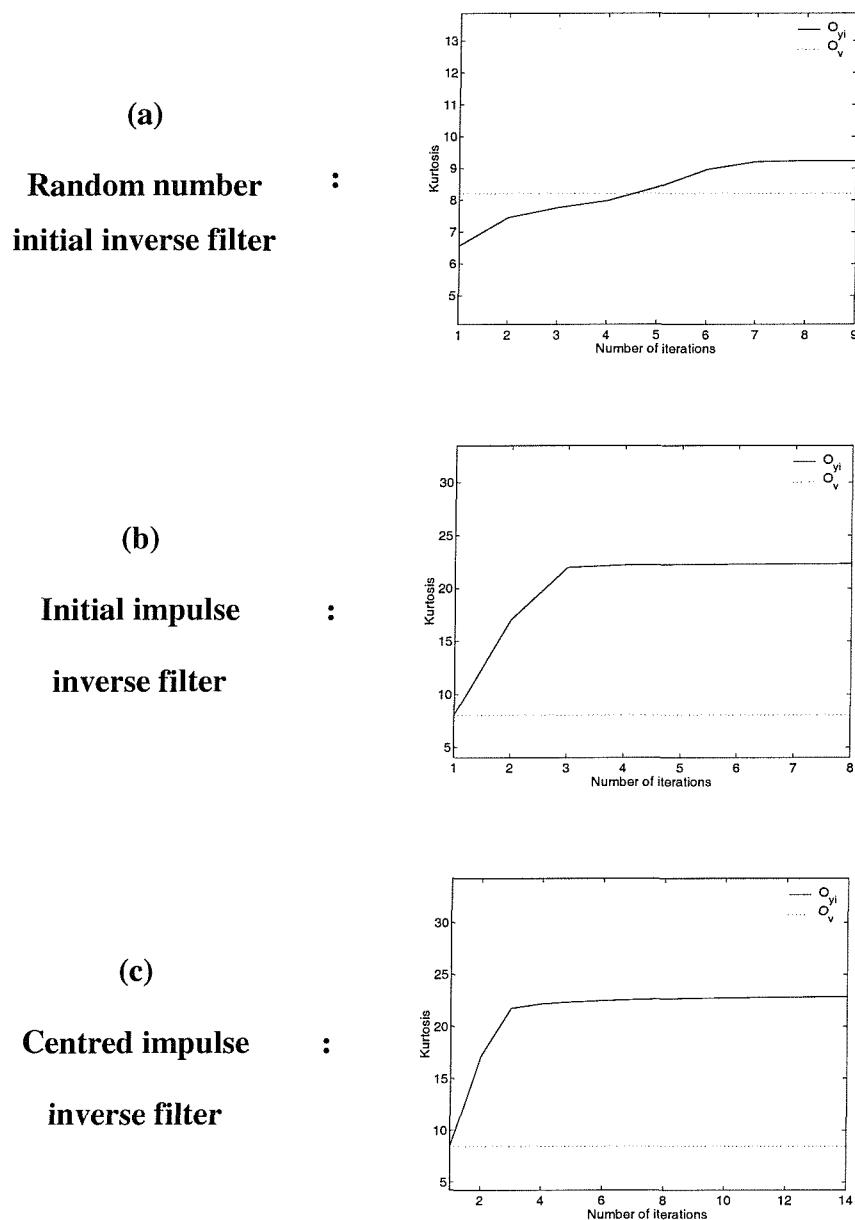


Figure 5.3.3 A comparison of the objective function values from three different initial inverse type. O_{yi} : Value of the objective function of output of the inverse filter at each iteration, O_v : Value of the objective function of the measured signal.

As shown in Figure 5.3.3 (a), the objective function value on the first iteration (first point of the solid line) is smaller than that of the measured signal (dotted line) for the randomly selected initial inverse filter case. Alternatively, if the initial inverse filter is chosen as an impulse type (b) or centred impulse type (c) [Gray, 1979], the output of the initial inverse filter can be at least not closer to the Gaussianity than the measured signal. Hence, choosing the initial inverse filter randomly may result in incorrect restoration of the signal or need more computational time to achieve the same result as the other types of initial inverse filter.

Simulation with three different initial inverse filter types and length

Using the same observed signal as in the previous simulation, the values of objective function (kurtosis) of the restored signals from three different filter types are compared for each length of these inverse filters.

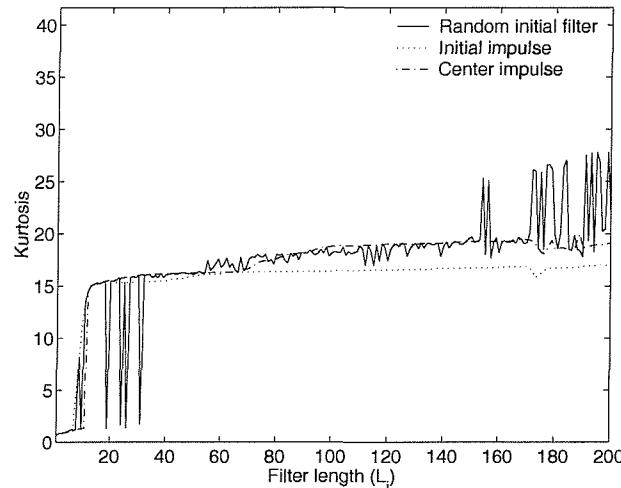


Figure 5.3.4 Comparison of the performance of deconvolution for three different initial inverse filter in each filter length

In Figure 5.3.4, the kurtosis of restored signal using the random initial inverse filter is not always consistent with the increase of the filter length. As an example, the restored signals and inverse systems for each initial inverse filter type are compared in the following figure;

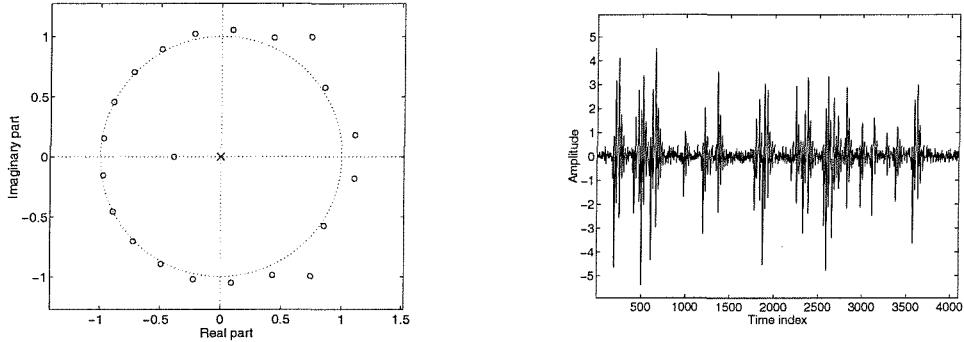
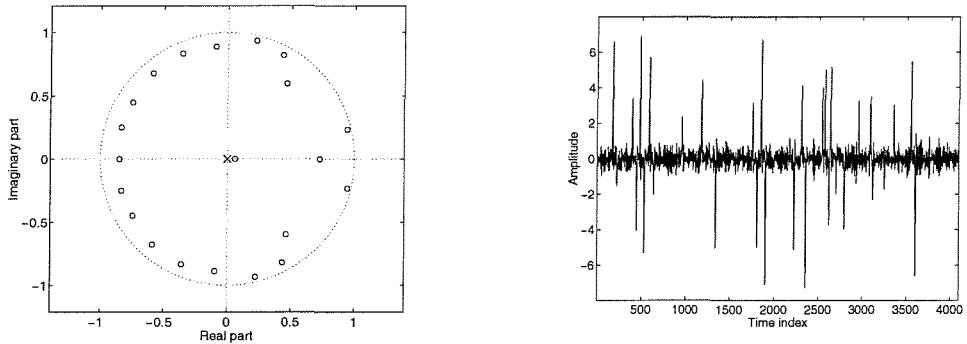
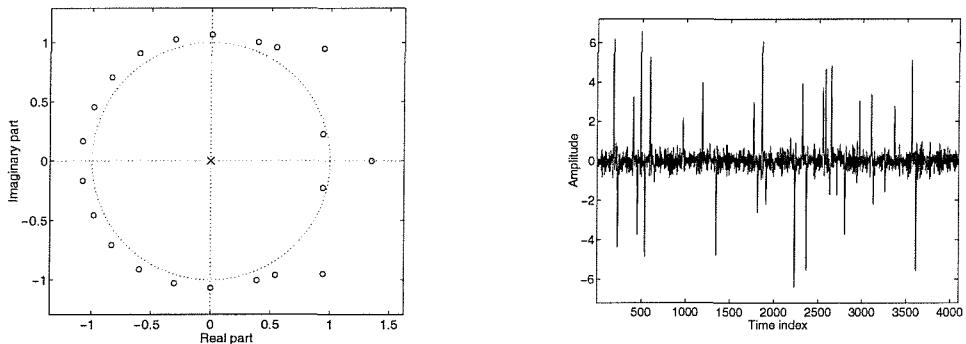
(a) Random initial inverse filter**(b) Initial impulse inverse filter****(c) Centred impulse inverse filter**

Figure 5.3.5 Pole-zero map of inverse filters and the shape of the restored signals from three different initial inverse filter types.

As can be seen in Figure 5.3.5, the restored signals from the initial impulse type inverse filter (middle, right of figure) and the centred impulsive inverse filter (bottom, right of figure) give significantly better results compared to the result

from the random number initial inverse filter (top, right of figure). The pole-zero map of the inverse filter from a randomly selected initial inverse filter type turns out to be a non-minimum phase system. The zeros are spread outside the unit circle without cancelling out the original system's pole position. In contrast, the zeros of the inverse filters obtained from both initial impulse and centred impulse initial inverse filters cancel out the original system's pole position. The resulting restored signals reflect the important effect of selecting the initial inverse filter type.

The results are summarised in the following table.

Table 5.3.1 Comparison of restored signal from three different initial inverse filters (numerical results of Figure 5.3.5)

	Restored signal by MA(21)		
	Random initial inverse filter	Initial impulse inverse filter	Centred impulse inverse filter
α coefficient	0.77	0.52	0.52
Entropy	2.743	2.2443	2.1846
Skewness	-0.43702	0.2462	0.76524
Kurtosis	9.2378	22.3232	22.847
SSD	1.05e+4	4.708e+3	4.7017e+3
Delay	25	7	15

5.4 Determination of the inverse filter length

The inversion formula defined in Chapter 3 (section 3.3) yields the inverse filter coefficient vector in a non-linear iterative manner. To solve this, an initial inverse filter coefficient vector with a chosen length is selected. This section addresses determination of the length of the inverse filter.

Methods for determining the length based on the statistical parameters of the observed and restored signal are introduced and their performances are compared based on the restored signals.

Three performance measures (I_p , α and E) which were introduced in section 5.2 are used to observe the effect of inverse filter length. As an example, the changes of these three performance measures for restored signals from different inverse filter length (2 – 200) are plotted in the following figure.

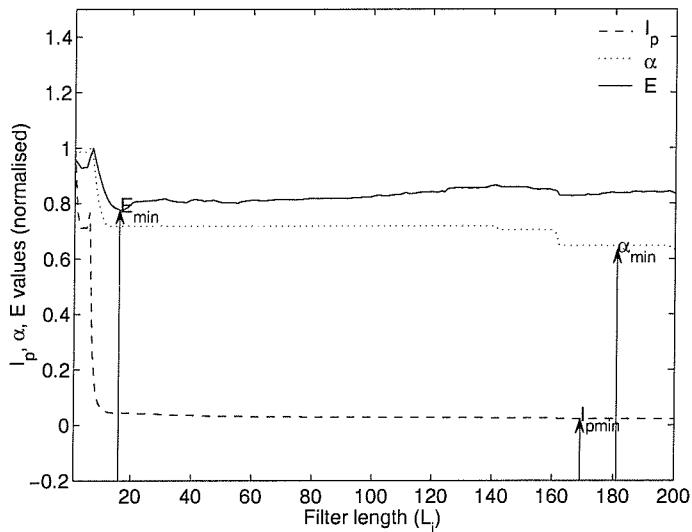


Figure 5.4.1 The shape of the inverse performance value (I_p for skewness), α coefficient, and entropy curve of restored signals produced from different inverse filter length.

Figure 5.4.1 demonstrates the trend of the three performance indices along with

the length of inverse filter (simulations and results are listed in Appendix E). The shapes of the performance indices indicate that a certain ‘limited’ length of inverse filter would be sufficient to restore the impacting signal and can avoid inefficient computational loads caused from an unnecessarily long inverse filter length selection. This is discussed in the following.

5.4.1 Inverse performance index based length determinator

From the shape of I_p curve in Figure 5.4.1, we resort to the spirit of the Akaike Information Criteria (AIC) which determines the orders of an unknown system incorporating the performance (improvement) and a penalty term [Akaike, 1974]. We note that this has not been arrived at from statistical/probabilistic arguments but is empirical. We consider the I_p curve as the performance and select the penalty term arising from the increased filter length as,

$$d_L = 2 \cdot L \cdot \log N / N \quad (5.4.1)$$

where N designates the number of data points. The optimal filter length selection criterion denoted as P_L is defined as,

$$P_L = I_p + d_L \quad (5.4.2)$$

Hence, the optimal length of the inverse filter selected from the point where the P_L curve reaches its minimum point, which is illustrated in Figure 5.4.3.

5.4.2 Shape parameter (α coefficient) based length determinator

The filter length criterion from this parameter is expressed using the same penalty term (d_L) expressed in the equation (5.4.1) as,

$$\alpha_L = \alpha + d_L \quad (5.4.3)$$

The optimal length of the inverse filter selected from above equation and the shape of the restored signal are illustrated in Figure 5.4.4.

5.4.3 Equivalent spikiness (entropy of the signal) based length determinator

The filter length determination criterion from this parameter takes a similar form as the previous methods as,

$$E_L = E + d_L \quad (5.4.4)$$

By this, the optimal length of the inverse filter selected from the point where the E_L curve reaches its minimum point. The shapes of the optimal length selection and restored signals from this length of filter are shown in Figure 5.4.5.

5.4.4 Simulation results and discussions

This section demonstrates the performance of four FIR inverse filter length determinators by comparing the status of the filter length selection, their restored signals, and the statistical values of the signals.

The input signal, unknown system and observed signal used in this simulation are shown in the following figure;

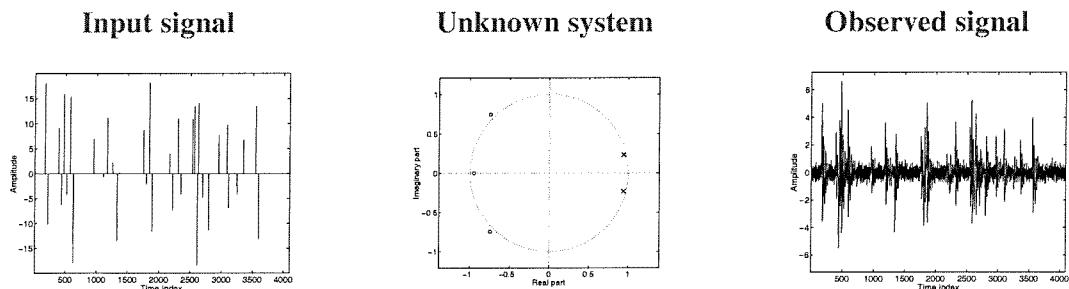
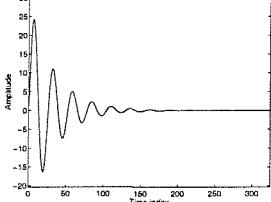


Figure 5.4.2 The shape of the unknown input signal, system and observed signal

The numerical values of the input and observed signal as well as the shape of the impulse response (IR) of unknown system are summarised in the following table.

Table 5.4.1 Numerical representation of the input signal, unknown system and observed signals of Figure 5.4.2.

	Input signal	Unknown system	Observed signal
α coefficient	0.29	 IR length=325	0.73
Entropy	0.072		2.745
Skewness	5.351		0.6577
Kurtosis	219.7		10.050
SSD	-		5.896e3
Delay	-		6

The optimal inverse filter length estimated from the inverse performance (I_p) index

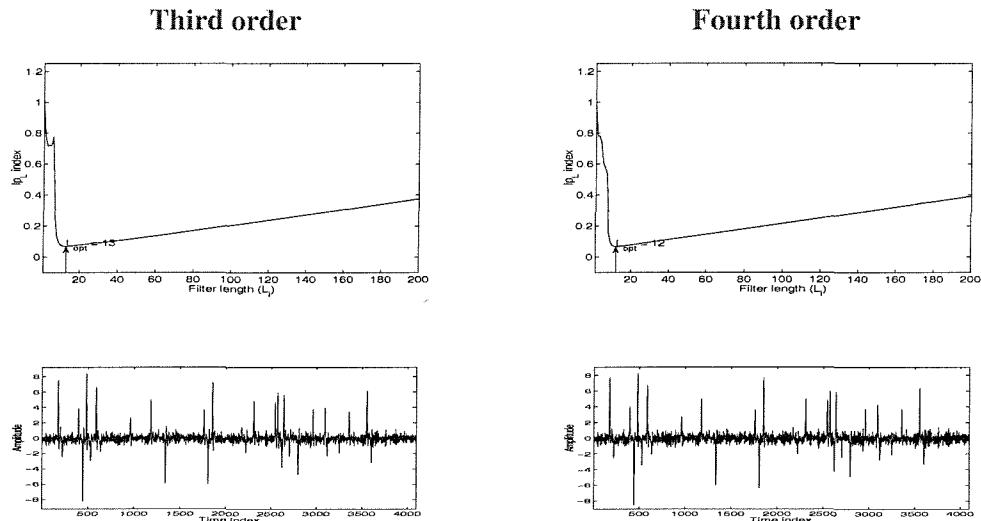


Figure 5.4.3 The optimal length of inverse filter determined from the inverse performance index and the restored signals via 3rd (left column), 4th (right column) order deconvolution methods.

The optimal inverse filter length estimated from the α -coefficient

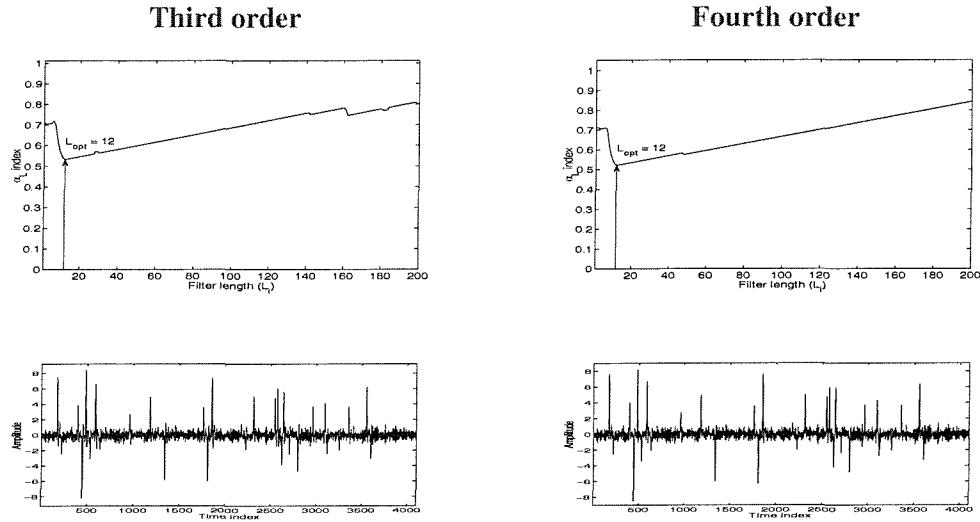


Figure 5.4.4 The optimal length of inverse filter determined from α -coefficient estimation and the restored signals via 3rd (left column), 4th (right column) order deconvolution methods.

The optimal inverse filter length estimated from the entropy

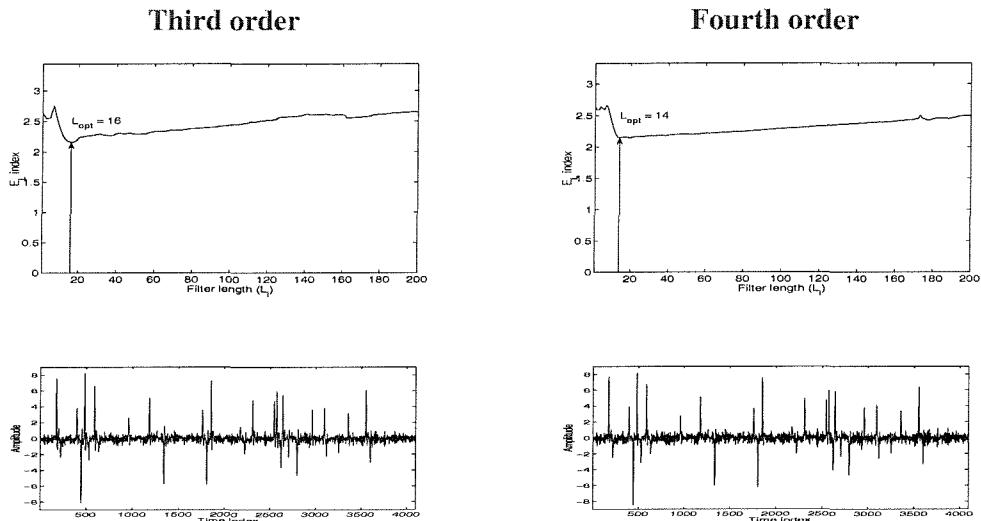


Figure 5.4.5 The optimal length of inverse filter determined from entropy estimation and the restored signals via 3rd (left column), 4th (right column) order deconvolution methods.

For this blind impacting signal reconstruction problem, there is a consistent abrupt change of the inverse performance index (and α -coefficient or entropy) at a certain length of inverse filter. As an example, the observed signal shown in Figure 5.4.2, the length that the abrupt change occurs is approximately 12 (when the initial inverse filter type was selected as impulsive). This implies that for this length the results of signal restoration can be acceptable and a longer length inverse filter is computationally inefficient. This is confirmed by the shape of I_p , the α -coefficient or entropy in Figure 5.4.1 with the estimated filter lengths in Figure 5.4.3 - Figure 5.4.5.

It is hard to say that these methods are an unique way of determining the length of the inverse filter. However, these length determinators are proposed as methods that can help to eliminate one of the ambiguities of the BD process.

5.4.5 Observation parameter based predictive determinator

This sub-section proposes an alternative approach for the optimal inverse filter length determination. Since the methods employed so far require a search procedure spanning a range of lengths of the filter, they are not necessarily an effective approach.

Ideally we need a guide to filter length not including deconvolution activities. We now seek such a guide.

As already mentioned, the three parameters of the restored signals along with their corresponding length of the inverse filters have been proved to possess a consistent trend (e.g., the longer inverse filter length assures the lower I_p index, as do the α coefficient and entropy). Based on this, Figure 5.4.6 depicts the graphical illustration of an optimal inverse filter selection scheme, which is aimed at a predictive inverse filter length determinator.

In the following figure, we propose two curves which we refer to as the *p-curve* and *d-curve*. For the blind impacting signal restoration problem, these two curves are designed to behave monotonically; The p-curve which is the objective function increases, whilst the d-curve (α coefficient and/or entropy) decreases as the length of the inverse filter increases.

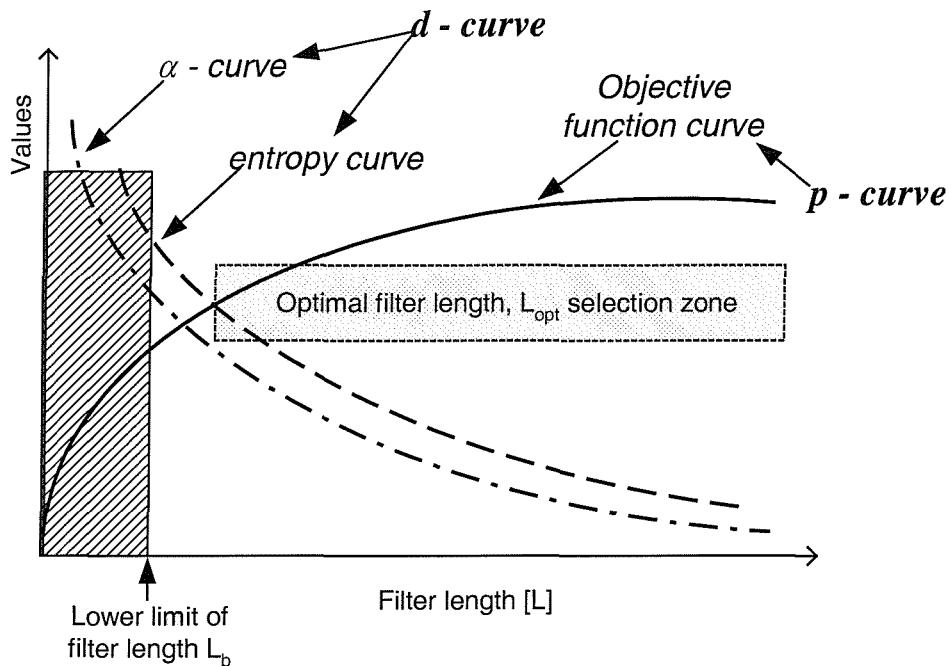


Figure 5.4.6 Graphical illustration of optimal inverse filter length selection.

As shown in Figure 5.4.6, the optimal inverse filter selection range (dotted area) is suggested based on the intersection point of the two curves with a lower limit for filter length indicated by the hatched area.

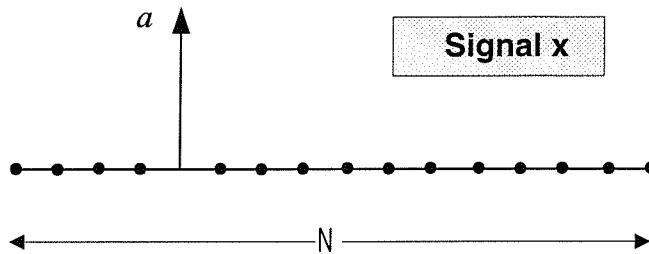
Justification for this predictive approach to filter length determination is now presented. The lower limit for inverse filter length is based on the non-zero lag higher order cumulant sequence and the result of Higher Order Singular Value Decomposition (HOSVD) dealt with in Chapter 4.

Reference parameter justification

To describe the status of the observed signal, the parameters chosen are the objective function value (i.e. skewness or kurtosis), shape parameter (α coefficient) and entropy of the observed signal. In order to obtain a reference parameter value for the objective function, the objective function value of the observed signal is compared to an extreme example of a spike sequence, whereas for the shape parameter (α coefficient) or entropy values, the reference values of these parameters are those of a Gaussian signal. These are explained below;

1) Reference of the objective function value

The objective function of an extreme non-Gaussian signal is chosen as the reference value. This is the normalised higher order cumulant of a single impulse of magnitude a in a signal with N samples as shown in the figure below



The discrete probabilistic density function can be written

$$p(x) = \frac{N-1}{N} \delta(x-0) + \frac{1}{N} \delta(x-a)$$

For $a \ll N$ and N is sufficiently large, the second order and higher order cumulants are estimated through the order of data length N as;

for covariance of \mathbf{x}

$$\begin{aligned}
 \sigma_x^2 &= \int_{-\infty}^{\infty} x^2 p(x) dx \\
 &= \int_{-\infty}^{\infty} \left(x - \frac{a}{N} \right)^2 p(x) dx \\
 &= \frac{a^2}{N^3} \left\{ N - 1 + (N - 1)^2 \right\} \\
 &\Rightarrow \mathcal{O}(N^{-3}), \quad \text{order of } N^{-3}
 \end{aligned}$$

for r th order ($r > 2$) cumulant of \mathbf{x} (symbols of cumulants are defined in Chapter 2, equation (2.2.20))

$$\begin{aligned}
 \mu_r &= \int_{-\infty}^{\infty} x^r p(x) dx \\
 &= \int_{-\infty}^{\infty} \left(x - \frac{a}{N} \right)^r p(x) dx \\
 &= \frac{a^r}{N^{r+1}} \left\{ N - 1 + (N - 1)^r \right\} \\
 &\Rightarrow \mathcal{O}(N^{-(r+1)}), \quad \text{order of } N^{-(r+1)}
 \end{aligned}$$

Thus, the r th order normalised cumulant takes the form of

$$\begin{aligned}
 K_x(r, 2) &= \frac{\mu_r}{(\sigma_x^2)^{r/2}} \\
 &\Rightarrow \mathcal{O}(N^{(r/2-1)}), \quad \text{for } r=4, \text{ subtract 3}
 \end{aligned}$$

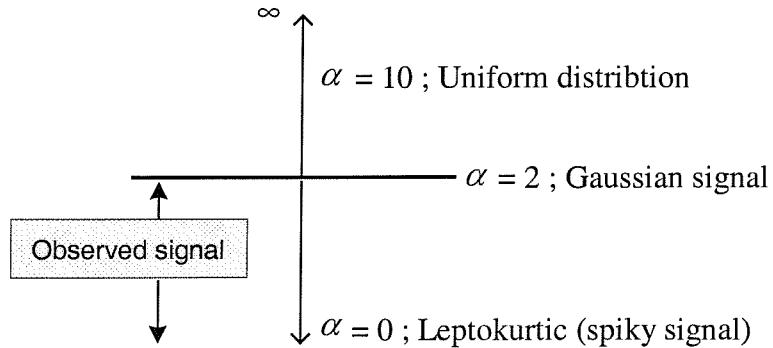
From this, an approximate normalised third- and fourth-order cumulants are calculated as;

Third-order reference objective function value $\approx \sqrt{N}$,

Fourth-order reference objective function value $\approx N - 3$

2) Reference for the α coefficient

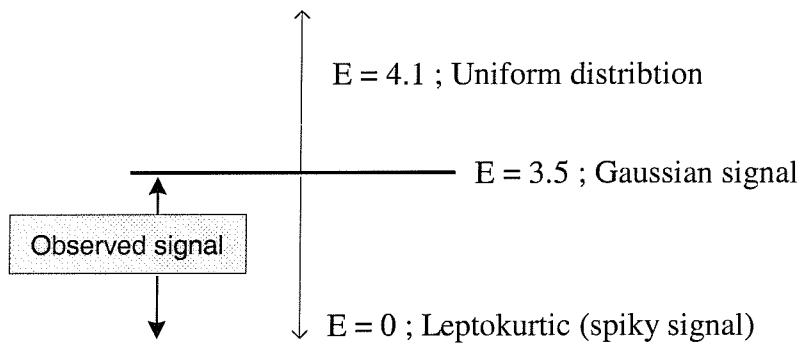
The shape parameter for signals of various distributions is categorised in the following figure;



In this study, the signals considered are assumed to belong to the lower part of the α region (i.e., $0 < \alpha < 2$) and so the reference α coefficient is chosen as 2.

3) Reference for the entropy (for calculation see equation (5.2.11))

The entropy region for signals of various distributions is categorised in the following figure;



As with the α coefficient reference, the entropy region (E region) for the observed signal is assumed to belong to the lower part of the region (i.e., $0 < E < 3.5$) and the reference entropy is chosen to be 3.5.

In summary, the reference parameters are selected as;

$$O_{ref} = \sqrt{N}; \quad \text{skewness (third order)}$$

$$O_{ref} = N - 3; \quad \text{kurtosis (fourth order)}$$

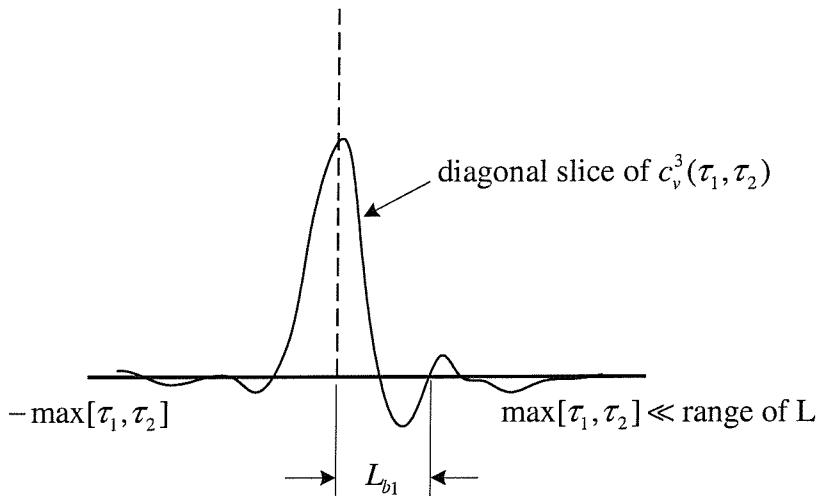
$$\alpha_{ref} = 2.0;$$

$$E_{ref} = 3.5$$

Lower limit of filter length L_b determination

This is done by considering two features; (i) the third order cumulant sequence and (ii) the reconstructability assessment described in Chapter 4.

(i) The first estimation comes from a modification of MA order estimation using the higher order cumulants [Chow, 1972; Chan and Wood, 1984; Giannakis, 1986; Giannakis and Mendel, 1990; Kim, 1998]. The use of the diagonal slice of the third-order cumulant sequence allows us to obtain an approximate lower bound of the inverse filter length. A brief example of the first lower bound L_{b1} is given in the following figure;



We take L_{b1} as the effective range of unknown system's impulse response which has more than 90% energy of diagonal slice of $c_v^3(\tau_1, \tau_2)$ in the figure.

(ii) For the second approach, the estimated lower bound L_b is adjusted by using the parameters involved in the reconstructability assessment namely CHS_k and R_k

(detailed explanations for these parameters are given in Chapter 4, section 4.3.4). Since the impulsive nature estimated from the third order cumulant sequence may be degraded by (severe) noise interference, the reconstructability parameter R_k and its reference CHS_k are used to adjust the length as follows;

Compare R_k and CHS_k and select the lower limit of inverse filter length L_b as

$$L_b = L_{b1}, \quad R_k \geq CHS_k$$

$$L_b = L_{b1} + \left\lceil \frac{CHS_k - R_k}{CHS_k} \cdot L_{b1} \right\rceil, \quad R_k < CHS_k$$

d-curve design

Using the above parameters and the lower bound, the first curve (called the *d-curve*) is determined which can reflect the status of the observed signal. For a given range of the inverse filter length L (e.g., $L = 1 \sim 200$);

$$d = \frac{1}{L^m} \log \left(\frac{L}{L_b} \right)^2,$$

where L_b is the lower bound of the inverse filter length and m is defined as $m = e^{|\alpha - \alpha_{ref}|}$ (for α coefficient case).

Using this, the slope of the d-curve reflects the shape parameter of the observed signal. For example, when the observed signal is close to Gaussian (large α coefficient), then m is smaller. In this case, the roll-off of the curve is slow and thus the d-curve extends to the longer filter length region and vice versa.

p-curve design

Similar to the d-curve, we introduce the *p-curve* which describes the status of the objective function value of the observed signal as,

$$p = \sum_{l=1}^{L_{max}} e^{-l^{(h/3)}},$$

where L_{max} is the maximum range of the inverse filter length under consideration

and h is calculated as $h = e^{\left(\frac{-O_{ref} - |O| + 1}{O_{ref}} \right)}$, $h \in [1, 2.718]$.

Hence, as the objective function value of the observed signal gets closer to Gaussian, then h becomes 1, thus the slope of the ascending curve decreases resulting in a longer inverse filter length selection, and vice versa.

Length determination and examples (the predictive determinator)

The optimal length of the FIR inverse filter is now predictively determined by choosing the intersection point of the two curves (d- and p-curve), which has been already illustrated in Figure 5.4.6.

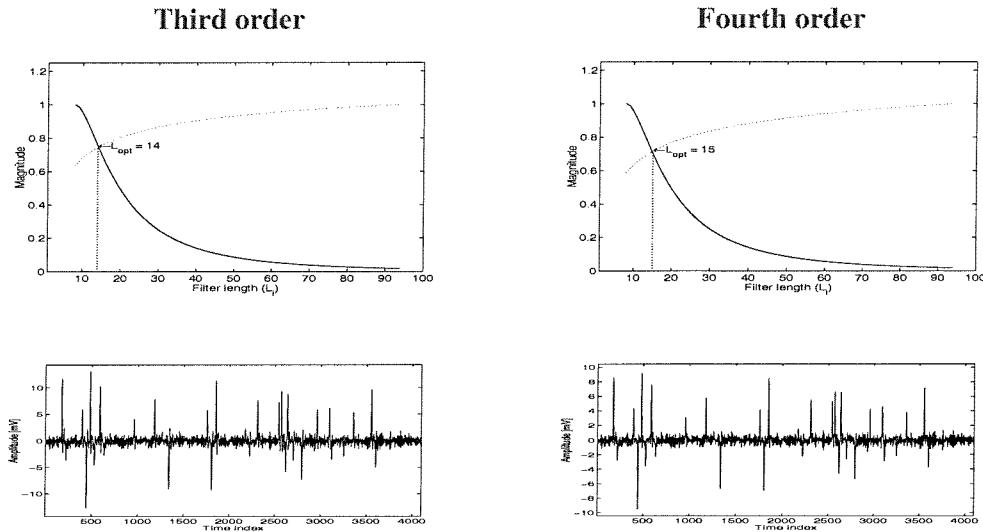


Figure 5.4.7 The optimal length of inverse filter determined from observation parameter based predictive estimation and the restored signals via 3rd (left column), 4th (right column) order deconvolution methods.

The numerical results for four length determinators and their corresponding restored signals are summarised in Table 5.4.2.

Table 5.4.2 Numerical comparison of the restored signals from four different inverse filter length determination criteria)

	Third-order			
	I_p	α	E	Predictive
L_opt	13	12	16	14
α coefficient	0.51	0.51	0.51	0.51
Entropy	2.182	2.230	2.219	2.20
Skewness	1.701	1.648	1.739	1.729
Kurtosis	24.020	23.516	24.116	24.011
SSD	4.85e3	4.85e3	4.89e3	4.89e3
	Fourth-order			
	I_p	α	E	Predictive
L_opt	12	12	14	15
α coefficient	0.5	0.5	0.5	0.49
Entropy	2.186	2.186	2.121	2.122
Skewness	1.578	1.578	1.643	1.859
Kurtosis	24.584	24.584	25.235	25.926
SSD	4.76e3	4.76e3	4.76e3	4.78e3

As a result, the filter length determined by this procedure tends to be a little longer than any of the other methods mentioned previously. This seems an acceptable ‘price to pay’ considering the low risk of poor deconvolution by choice of a shorter inverse filter length. Furthermore, the significant merits of this procedure are:

- Unlike the other length determinators, this method can give a reasonable guide without doing any deconvolution (which is a major task considering the wide range of the candidate inverse filter lengths).
- The length determinator is rarely affected by the choice of either third or fourth order as this procedure includes the additional parameters (α coefficient and/or entropy) and the choice of reference spike signal.

Further examples of the signal restoration by the optimal length of the inverse filter using this predictive length determinator are given in Figure 5.4.8.

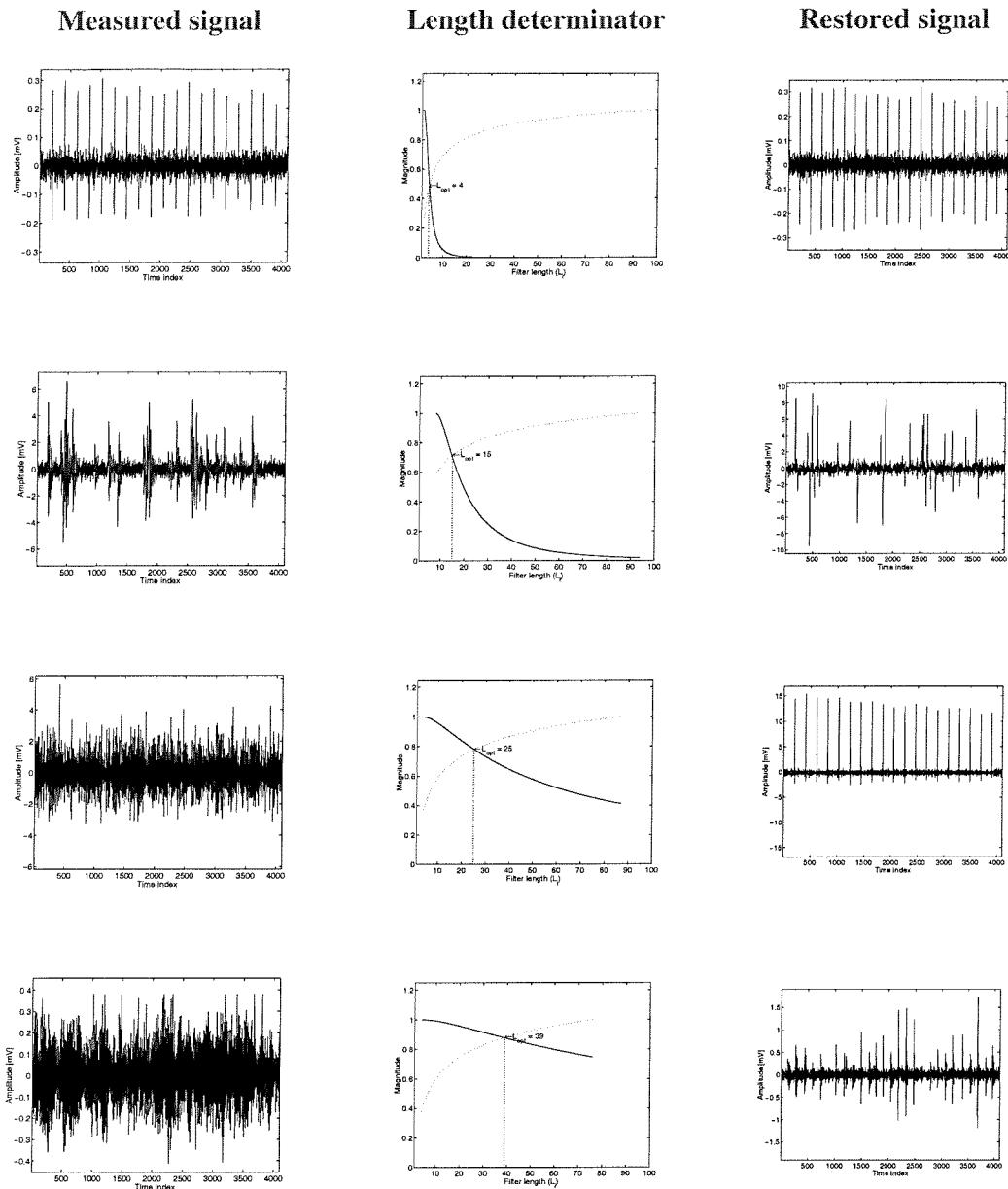


Figure 5.4.8 The performance of the predictive length determinator with different status of observed signals (fourth order case).

The left column shows four different types of observed signals and the middle column gives each estimated FIR inverse filter length. The restored signals constructed by the third-order BD process with the estimated filter length are displayed in the right column of the figure.

It is noticeable that as the observed signals become closer to Gaussian, the inverse filter length estimated by the predictive determinator also tends to be longer. Particularly, comparing the third and fourth row of the Figure 5.4.8, those two signals look very similar and close to the Gaussian signal but the estimated inverse filter length differs considerably ($L_{\text{opt}} = 25$ and 39). This is due to the difference of the generating system (unknown system); the unknown system for the fourth row is in fact more complicated (experimental data from a real cantilever beam), thus the effect of the system has been significantly included in estimating the inverse filter length.

5.5 Summary and conclusion

This chapter has mainly considered two aspects of the BD process.

Since the blind deconvolution problem ends up with iterative solution of the non-linear normal equation as described in the Chapter 3, we need (i) initial filter coefficient vector and (ii) we need to determine the length of the inverse filter.

The initial inverse filter showed generally be chosen so as to make the output of the first filter to be at least equal to or further from Gaussian. Based on this, it is demonstrated that it becomes natural to take the *initial inverse filter as impulsive* in form.

Criteria for the selection of optimal inverse filter length have been proposed. Four different length determinators have been considered and their performances have been compared. Based on this a *predictive determinator* which utilises an objective function (skewness or kurtosis), α -coefficient and/or entropy of the observed signal has been shown to be an *effective length determinator*. The advantage of the predictive approach is that time consuming extensive computation is avoided.

Chapter 6

Global optimisation in Blind Deconvolution

6.1 Introduction

In Chapter 3, calculation of the inverse filter coefficients in Blind Deconvolution (BD) was carried out through solution of a (generalised) Wiener type equation. This approach has been used and developed by many authors for many years [Claerbout, 1977; Wiggins, 1977; Gray, 1979; Donoho, 1981; Cadzow, 1996; Nandi, 1997]. However, these techniques can only guarantee that global minima or maxima will be found if the performance index is convex. If it is non-convex a local optimisation algorithm may not locate the global optimum and, even when it does, there is no indication that the solution is global. In Chapter 3, we saw that the selected objective functions (skewness or kurtosis) lead to *multimodal* behaviour with respect to the inverse filter coefficients. Accordingly, it was decided to investigate this aspect further.

This chapter considers a global optimisation scheme with a view to comparing the results with the generalised Wiener approach. There are many approaches to optimisation and this is not a study of all the various alternatives. We have selected a modification of the Genetic Algorithms (GAs) and known as a ‘Differential Evolution (DE)’ process. This choice was based on recent literature that indicated that DE is relative easy to apply.

6.2 Differential Evolution

This section describes the Differential Evolution (DE) algorithm which uses evolving populations of solutions in much the same way as the Genetic Algorithms. This algorithm works by making a given number of random guesses (called the ‘population’ and denoted ‘P’) of D parameters and then imposing improvements on the ‘cost function’ (denoted as ‘ $f(\mathbf{x})$ ’ in Figure 6.2.1) considered. A number of best guesses from the previous population survive as the process continues. This algorithm is a parallel direct search method which utilises P D-dimensional parameter vectors \mathbf{x}_i (denoted as ‘ \mathbf{x} ’ in Figure 6.2.1).

$$\mathbf{x}_{i,G} \quad , \quad i = 1, 2, \dots, P, \text{ and } \mathbf{x} = [x_1 \ x_2 \ \dots \ x_D]^T \quad (6.2.1)$$

as a population for each generation G. The size P does not change during the minimisation/maximisation process. The initial vector population is chosen randomly and should cover the entire parameter space. As a rule, a uniform probability distribution for all random decisions will be employed unless otherwise stated. Evolution is carried out by forming a child population by mating pairs of parents, followed by mutation. Differential evolution is basically very similar to conventional Genetic Algorithms (GA) [Goldberg, 1989; Ingber and Rosen, 1992]. The differences are in the way the mechanisms of mutation and crossover are performed using real floating point numbers instead of long strings of zeros and ones. In particular, the concept of perturbing a vector with the difference of two other parameter vectors is borrowed from the reflection, expansion and contraction processes of the Nelder and Mead’s downhill simplex optimisation algorithm [Nelder and Mead, 1965]. In comparison with GA, the basic difference lies in the scheme for generating trial vectors [Storn and Price, 1997]. The importance of the term trial parameter vector will be evident in the following description of the differential evolution method.

6.2.1 Global optimisation scheme using Differential Evolution

Figure 6.2.1 shows schematically the basic operations of DE working on optimisation of a hypothetical multi-parameter system. The parameter vectors considered here are composed of five parameters x_{Di} ($D=1,2,\dots,5$ and i is the index of population up to P).

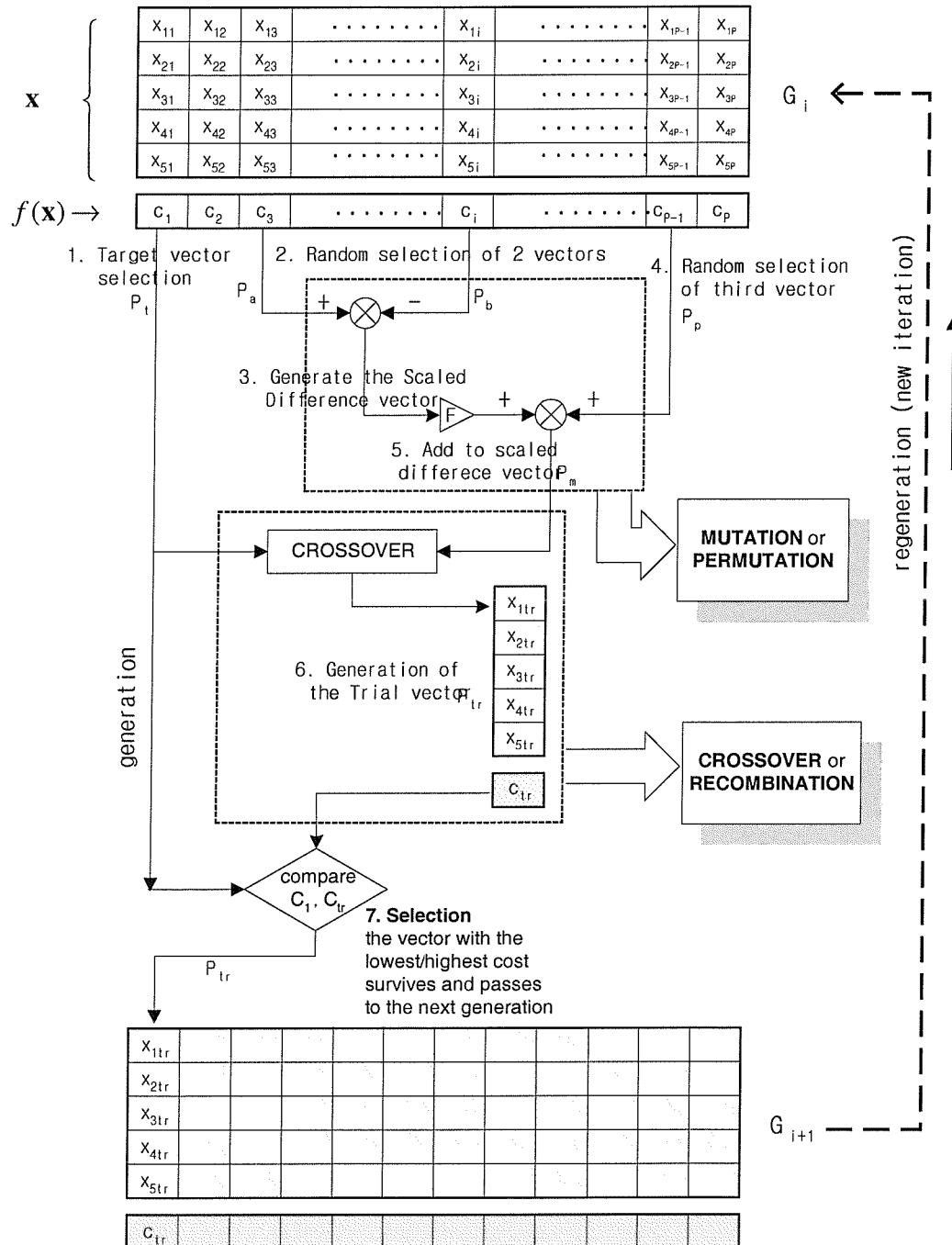


Figure 6.2.1 Schematic representation of differential evolution operations

In Figure 6.2.1, there are two $D \times p$ celled tables (in this case, $D=5$) at the top and bottom of the figure which represent the generations of the parameters. This means the algorithm starts with an initial pool of 5 dimensional parameter vectors with population P (candidates or ancestors) drawn from a particular probability distributions. Normally a uniform probability distribution ensures that the parameter vectors generated will span the parameter space equally. The initial pool acts as the first **generation** from which the whole evolution operation starts. Thus, the top table represents the generation whose members will evolve according to the mechanisms of the differential evolution.

In the bottom table each newly evolved member will be placed according to the positions which its predecessor was holding in the top table. The bottom table is called a **new generation**.

As can be seen in Figure 6.2.1, the (vector) member $[x_{11}, x_{21}, \dots, x_{51}]^T$ of the first table (top) is evolved to the descendant $[x_{1tr}, x_{2tr}, \dots, x_{5tr}]^T$ that holds the leftmost position in the bottom table. The cost function (performance index) for each parameter vector is displayed in the corresponding rows of the two tables. The detailed processes between the top and bottom table are described as follows;

Step 1: Target vector selection

The first left parameter vector of the top table is selected and denoted as the **Target vector P_t** . Note that the steps illustrated in Figure 6.2.1 are applied in parallel to all the other vectors of the top table during a single run (one iteration). All the (parallel) target vectors form the basis of ancestors that are to be compared to the generated descendants and vacate their positions to the new generation that fill the bottom table. At the end of each run the improved generation in the bottom table is passed to the top table in order to play a role of new ancestors.

Step 2: Random selection of two vectors

In the search scheme, a mutation is created by utilising a combination of three parameter vectors. Firstly, two parameter vectors denoted P_a , P_b are selected

among the top table which should not be the same as the target vector \mathbf{P}_t .

Step 3: Generate the scaled difference vector

From the randomly selected parameter vectors \mathbf{P}_a and \mathbf{P}_b in the previous step, the difference of two parameter vectors are multiplied by a user defined constant F ($F \in [0,1]$) to create the **scaled difference vector**.

Step 4: Random selection of third vector

The third parameter vector denoted \mathbf{P}_b is selected randomly. This vector also should not be either the target vector \mathbf{P}_t or two randomly selected parameter vectors $\mathbf{P}_a, \mathbf{P}_b$.

Step 5: Add to scaled difference vector (mutation)

Differential evolution mutates (perturbs) by using two randomly selected and one separately selected vector (i.e. the third vector). This process creates a new mutated parameter vector \mathbf{P}_m . While in Figure 6.2.1 the perturbation process is portrayed in the top dashed box, the visual interpretation of this process is shown in Figure 6.2.2.

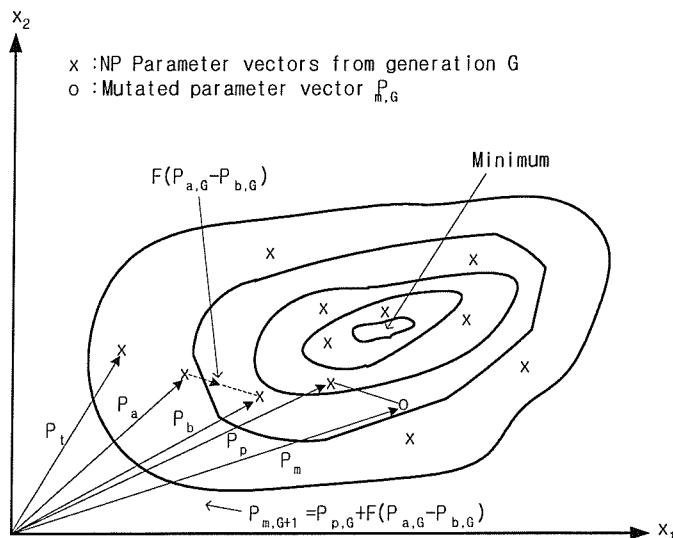


Figure 6.2.2 Perturbation (mutation) process of differential evolution for two dimensional cost function and the process for generating mutated parameter $P_{m,G+1}$

Step 6: Generation of the Trial vector (crossover)

The next operation of the differential evolution is crossover. The mutated parameter vector \mathbf{P}_m recombines with the target vector \mathbf{P}_t in order to generate a new trial vector \mathbf{P}_{tr} . The \mathbf{P}_{tr} consists of parameters obtained from both \mathbf{P}_m and \mathbf{P}_t . In the Genetic algorithm scheme, \mathbf{P}_m and \mathbf{P}_t are known as the parents and the newly generated parameter vector \mathbf{P}_{tr} is known as their child. The actual parameter components of every vector \mathbf{P} are known as genes. Differential evolution implements recombination by using the chance of occurrence applied to the selection of the genes from the parents to the child. The likelihood that governs the inheritance of the parameter vector \mathbf{P}_{tr} is determined by a constant parameter designated as the **crossover ratio** (CR). The detailed operation of this crossover is illustrated and explained in the following;

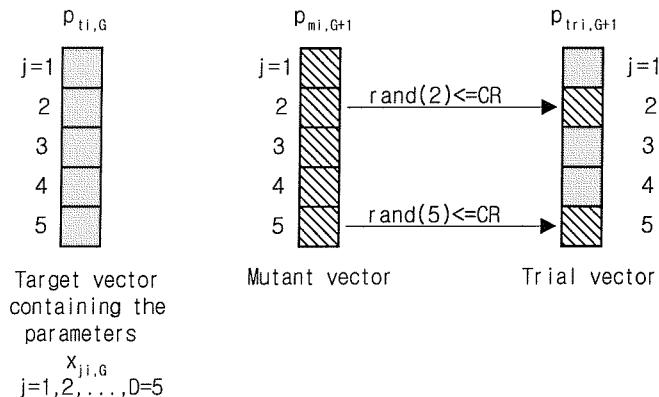


Figure 6.2.3 Illustration of the crossover process for $D=5$ parameters

As can be seen in Figure 6.2.3, the process of crossover creates a new parameter vector named the trial vector (right) by exchanging the elements (genes) of the target parameter vector (left) and the mutant vector (middle). This genetic recombination based on the random search from which exploration of creating improved generations evolves through the following rules.

$$P_{tr,ij,G+1} = \begin{cases} P_{m,ij,G+1} & \text{if } \text{rand}(j)|_{at_j} \leq CR \quad \text{or} \quad j = \text{perm}(1,2,\dots,D)|_{at_j} \\ P_{t,ij,G} & \text{if } \text{rand}(j)|_{at_j} > CR \quad \text{and} \quad j \neq \text{perm}(1,2,\dots,D)|_{at_j} \end{cases} \quad (6.2.2)$$

where $\text{rand}(j)|_{\text{at}_j}$ is the j -th evaluation of a uniform random number generator with outcome $\in [0,1]$, CR is the crossover constant $\in [0,1]$, and $\text{perm}(1,2,\dots,D)|_{\text{at}_j}$ is a randomly chosen index between 0,1 which ensures that $P_{\text{tr},ij,G+1}$ gets at least one parameter from $P_{m,ij,G+1}$.

Step 7: Selection

By comparing the cost value of the target parameter vector and that of the trial vector obtained from the crossover process, the lowest or highest cost value of parameter vector survives and is passed to the next generation depending on the optimisation scheme (e.g., minimisation or maximisation).

The chosen parameter vector and its cost value is then placed in the same position of the bottom table as the target vector in the top table.

Repeating the steps for every target vector in the top table creates a new generation which fills the bottom table and takes the role of the previous generation for each iteration. Since the selection of the target vector is done from the leftmost to the right direction, this procedure as called the parallel search method. The iteration halts when a stopping criterion is satisfied. Usually this criterion is set as either a user defined cost value or a maximum number of iterations or both.

6.2.2 Performance test for Differential Evolution method

To demonstrate the performance of DE in seeking a global minimum/maximum, we have used two different multi-modal functions as examples.

Test function and global optimisation:

Test function 1 (Rosenbrock's saddle) [Ingber and Rosen, 1992]

$$f(\mathbf{x}) = 0.1(x_1^2 - x_2)^2 + 0.001(1 - x_1)^2; \quad (6.2.3)$$

IPR : $x_1 \in [-1.248, 1.248]$,
 IPR : $x_2 \in [-0.848, 1.248]$

where IPR implies the initial parameter range.

The shape of this function is shown in the following figure.

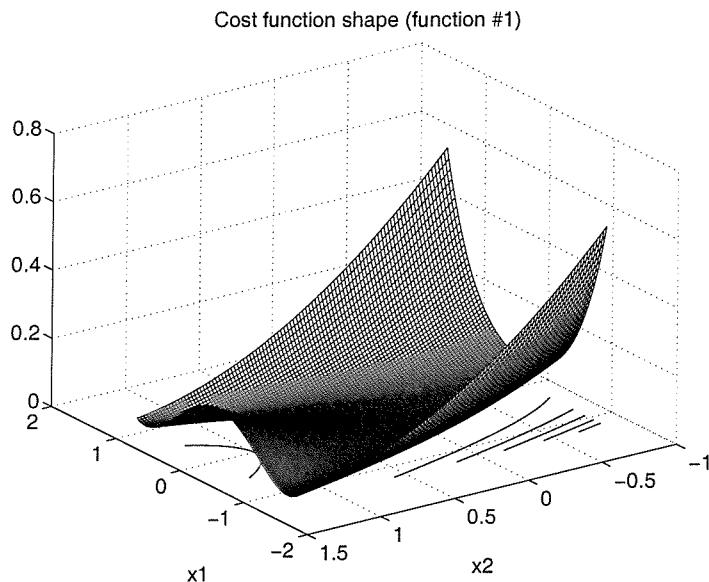


Figure 6.2.4 The shape of test function # 1

Although $f(\mathbf{x})$ has just two parameters, it has the reputation of being a difficult minimisation problem as the global minimum of this function is $f(\mathbf{x}_1) = 0$ at $x_j=1.0$ ($j=1,2$) while another comparable minimum $f(\mathbf{x}_2) \approx 0$, $f(\mathbf{x}_2) > f(\mathbf{x}_1)$ exists at the opposite side of x_1 ($x_{1_opp} = -1.0$, $x_2 = 1.0$).

Optimisation

The initial population (the number of candidate parameter vectors) size is taken as three times bigger than the dimension of parameter vector (i.e., the number of parameters). Thus, from an initially constructed 2×10 matrix, the DE search is tried.

Following figures demonstrate the performance of DE.

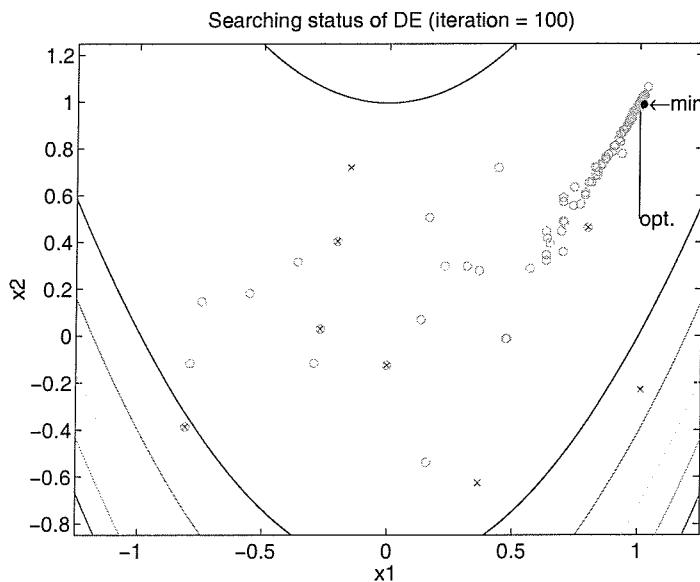


Figure 6.2.5 Searching status of DE process for Test function 1.

A 100 stage iteration is given with scale factor (constant) $F = 0.7$ and cross-over ratio $CR = 0.5$. In the Figure 6.2.5, we can see the result of the iteration in which 'x' designates the initial population and 'o' identifies the optima in every run. As already mentioned, the initial populations of each parameter 'x' are drawn from the uniform distribution to cover the entire landscape (co-ordinates x_1 and x_2). The identified optima 'o' at each iteration then move toward the minimum value position which is marked by '•'.

This track is represented in a 'probabilistic' way in Figure 6.2.6 where the performance of DE is demonstrated through a plot of relative frequency of the minima that arise from DE. It is noticeable that in Figure 6.2.6 the occurrence of parameter values (x_1 and x_2) at each iteration rapidly accumulate in the vicinity of

the true minimum position ($x_1 = 1.0$ and $x_2=1.0$). The behaviour of this bar graph also tells us the speed of convergence of DE to the true minimum point.

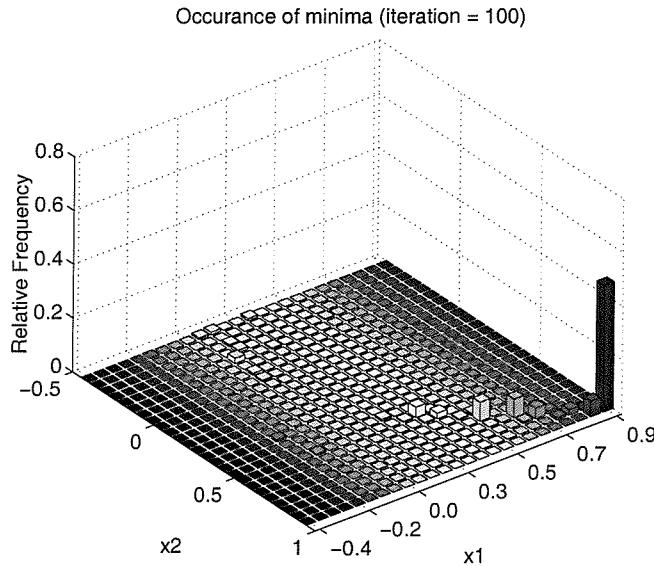


Figure 6.2.6 Relative frequency of each minima appearing by DE process for Test function 1.

Summary of the parameters and results of the test:

Scale constant, $F= 0.7$

Cross-over Ratio, $CR= 0.5$

Number of iterations =100

$x_1_{\text{opt}}=0.949$ ($x_1_{\text{true}} =1.0$)

$x_2_{\text{opt}}=0.901$ ($x_2_{\text{true}} =1.0$)

Opt. value=2.5e-6 (Min. value=0)

() is the theoretical minimum points and minimum value.

Test function 2 (selected by the author)

$$f(\mathbf{x}) = -0.05(x_1 - 0.25)^2 - 0.02 \cdot |x_2 + 0.01|^{1.5} - x_1 \cdot x_2 \cdot \exp(-x_1^2 - 1.15x_2^2) + 1; \quad (6.2.4)$$

IPR : $x_j \in [-2.0, 2.0]$, $j = 1, 2$

This function has been chosen because it has a similar shape to the objective function that we discussed in Chapter 3.

The shape of this function is shown in the following figure.

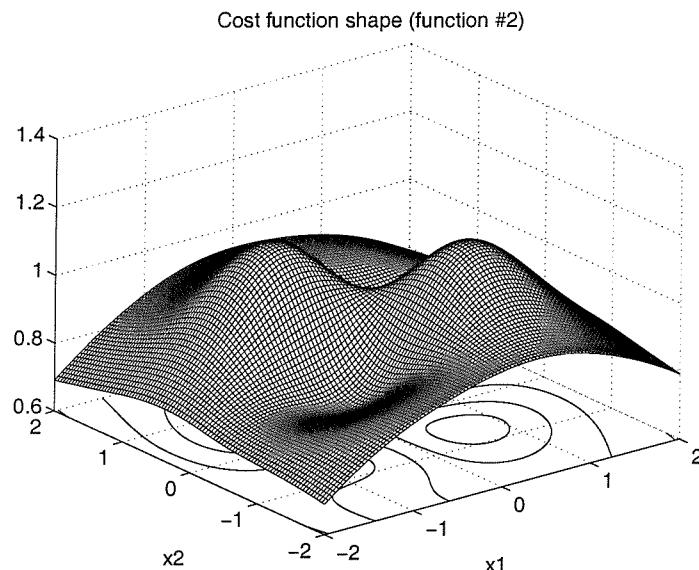


Figure 6.2.7 The shape of test function # 2

The global maximum of this function is $f(\mathbf{x}) = 1.15$ at $x_1=0.66, x_2= - 0.63$,

Optimisation

The initial population is taken the same size as the case of test function 1. The DE method is applied to search for the ‘maximum’ point of the test function 2. Thus, from an initially constructed 2×10 matrix, the DE search yields results shown in the following figures.

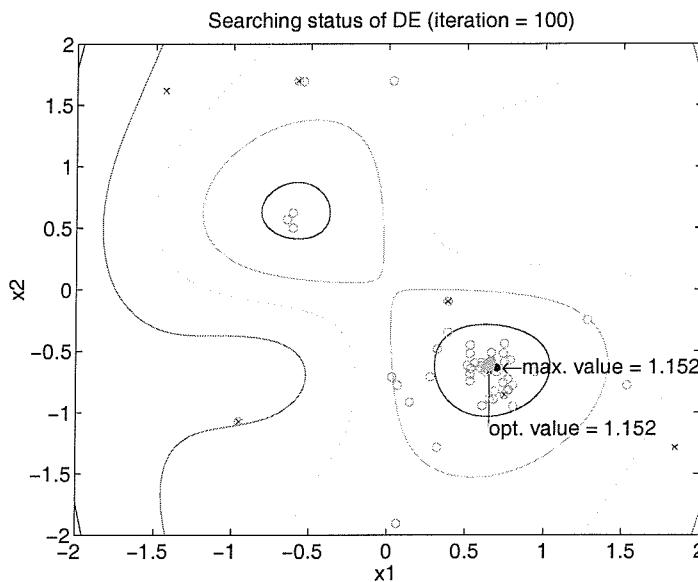


Figure 6.2.8 Searching status of DE process for Test function 2.

During 100 iterative parallel searching process with the same F and CR as were used in the previous test function 1 case, we can observe that some parameter vectors fall into a ‘local maximum’ point (see the left upper corner of Figure 6.2.8) and then promptly head to the ‘global maximum’ point. Their occurrence in a ‘relative frequency’ form is shown in the following figure.

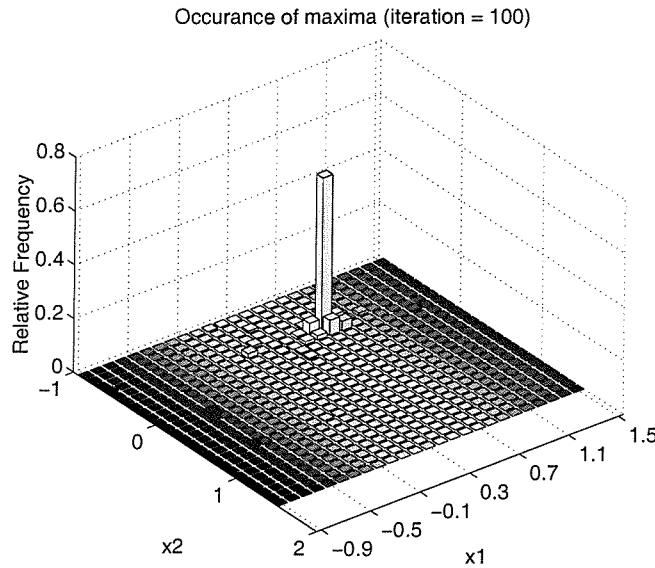


Figure 6.2.9 Relative frequency of each minima appearing by DE process for Test function 2.

Similar to the previous test, most of the search results during 100 iteration fall into the global maximum. Thus, the occurrence probability is dense around the true maximum point.

Summary of the parameters and results of the test:

Scale constant, $F= 0.7$

Cross-over Ratio, $CR= 0.5$

Number of iterations =100

$x_1_{\text{opt}}= 0.651$ ($x_1_{\text{true}} = 0.66$)

$x_2_{\text{opt}}= -0.629$ ($x_2_{\text{true}} = -0.63$)

Opt. value=1.151 (Max. value=1.152)

() is the theoretical maximum points and maximum value.

Remarks:

From two test function tests (one for minimisation and the other for maximisation), we can conclude that DE is a potentially useful search tool.

The speed of convergence to the global minimum/maximum can be affected by the selection of F and CR. As a rule of thumb, normally the scale constant (F) ranges from 0.5 to 0.9 and the cross-over ratio (CR) is 0.1 – 0.6.

6.3 Application of DE to Blind Deconvolution

In this section, the performance of DE for blind signal reconstruction is demonstrated through simulations.

6.3.1 Preliminaries for simulations

The system for this simulation is shown in the following figure.

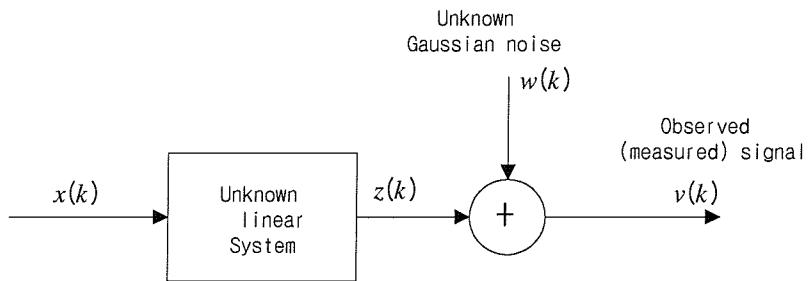


Figure 6.3.1 The signals and system used for non-Gaussian impacting signal detection problem with Gaussian measurement noise

Excitation $x(k)$

A non-Gaussian impacting signal used in this simulation is shown in Figure 6.3.2.

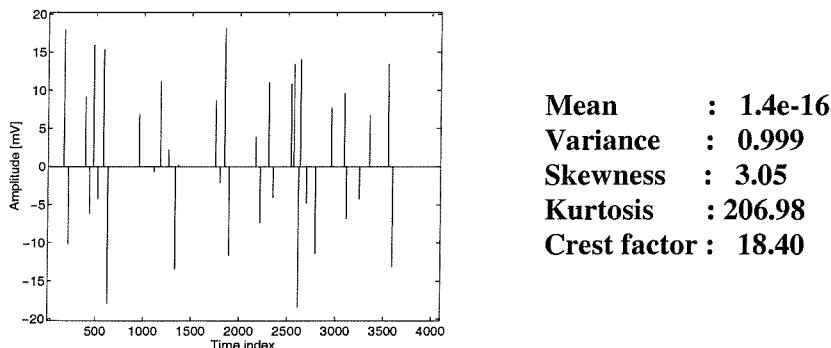


Figure 6.3.2 The input signal

System

The ‘unknown’ system selected is shown in the following figure;

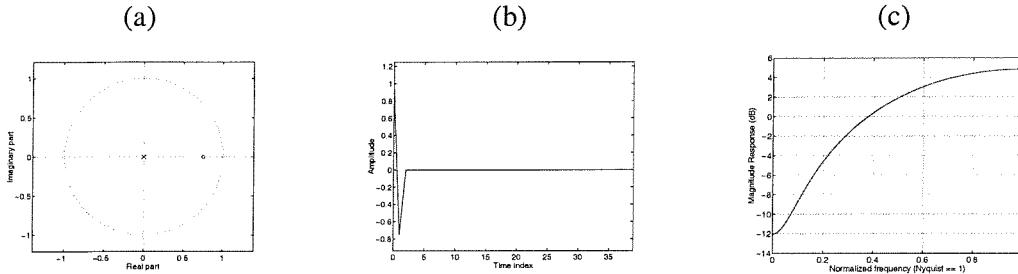


Figure 6.3.3 The system used in the simulation. (a): System’s pole zero map, (b): impulse response, and (c): Frequency Response Function.

This system has a very ‘short’ impulse response function.

The above input signal and system have already been used in Chapter 3, section 3.3. They are used again so as to provide a basis for comparison of the Wiener approach and DE.

Observed signal $v(k)$

The observed signal is obtained from the output signal including Gaussian noise (SNR= 10 dB) interference and shown in Figure 6.3.4.

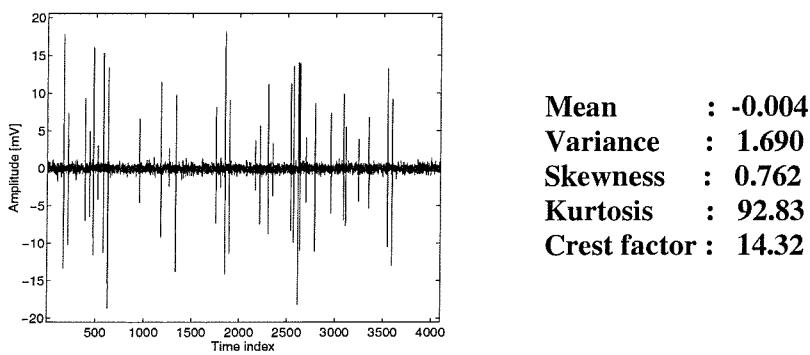


Figure 6.3.4 The observed signal (SNR= 10 dB)

6.3.2 Global maximum searching and restoration of signals

From only an observed signal v , an inverse system (filter) f_k is to be estimated to make the output signal y be the unknown input signal x , which is illustrated as

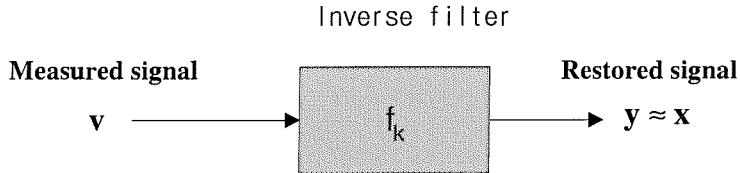


Figure 6.3.5 The deconvolution system

The estimation of the inverse system and signal restoration processes are extensively discussed in Chapter 3. As a reminder, the objective function (cost function) for this estimation takes a higher order normalised value (e.g., kurtosis of the restored signal) as

$$O_y = \frac{\frac{1}{N} \sum_{i=1}^N y(i)^4}{\left[\frac{1}{N} \sum_{i=1}^N y(i)^2 \right]^2} \quad (6.3.1)$$

which is to be maximised with respect to the parameters (inverse filter coefficients). The final signal restoration is achieved from the convolution between the optimally estimated inverse filter coefficients and the observed signal as

$$y(n) = \sum_{k=0}^{L-1} f_k \cdot v(n-k) \quad (6.3.2)$$

where $f_k = [f_0 \ f_1 \ \dots \ f_{L-1}]^T$. The DE process for this problem is described in the following;

From the measured (observed) signal of Figure 6.3.4, the fourth order normalised cumulant (i.e., kurtosis) is chosen as a cost function from which the coefficients of f_k are estimated that can give the maximum kurtosis of the output of the inverse filter. The shape of the objective function (cost function) is shown in Figure 6.3.6.

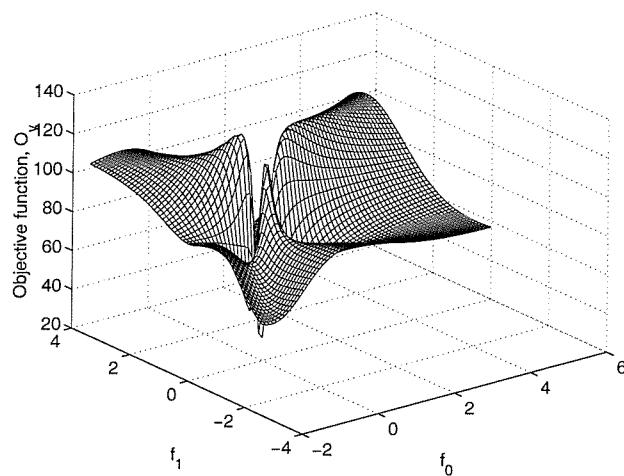


Figure 6.3.6 The shape of cost function (kurtosis)

The global optimality of the DE process is compared to the Wiener Optimisation (denoted by 'WO') using a MA(4) inverse filter in the following figure:

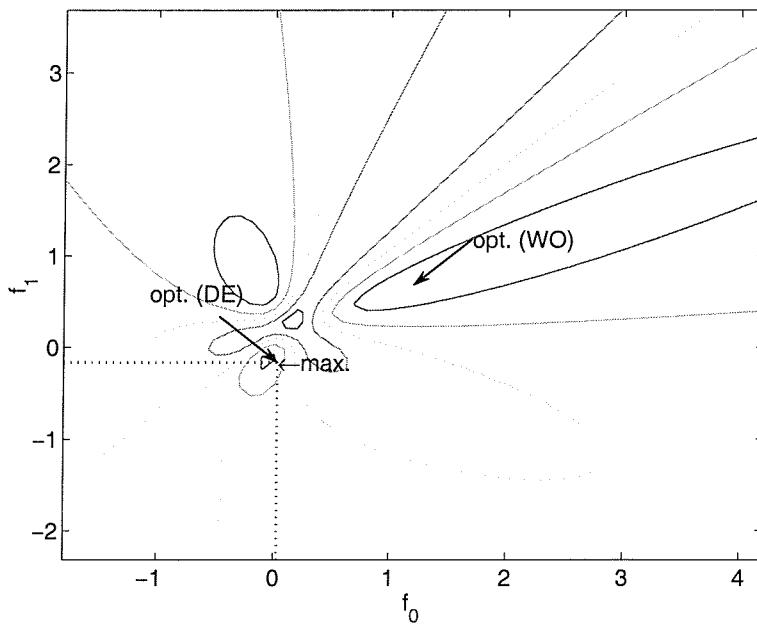


Figure 6.3.7 Result of the maximum point search using Wiener optimisation and DE

As can be seen in Figure 6.3.7, the optimal point of DE marked by ‘opt. (DE)’ coincides with the ‘max’ point whereas the optimal point for the Wiener optimisation marked by ‘opt. (WO)’ is different and is identified as a local maximum.

The restored impacting signals (input signal of Figure 6.3.2, unknown system of Figure 6.3.3, and observed signal of Figure 6.3.4) using both Wiener Optimisation and Differential Evolution method with MA(4) inverse filter are compared below:

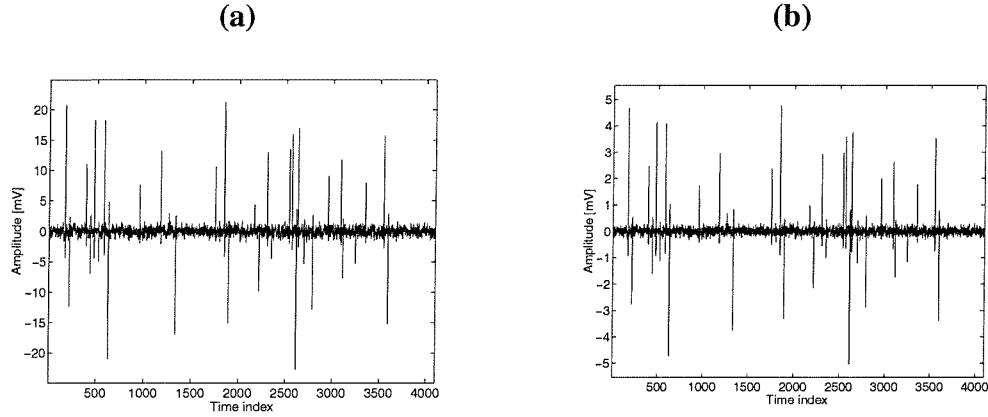


Figure 6.3.8 Restoration of signal using the Wiener solution and the DE method using a MA(4) inverse filter. (a): restored signal by using Wiener optimisation, (b): restored signal by using DE optimisation.

This rather ‘simple’ example using only an MA(4) filter is used for illustrative purposes. The signal restoration results, however, are similar (with a different magnitude). We conclude that the Wiener optimisation does indeed seem to offer a reasonable and feasible method for blind signal restoration provided that the previously considered inverse filter’s requirements (i.e. selection of appropriate initial filter) are satisfied. This is reassuring considering that the computational loads in employing DE become nontrivial for large order filters.

For completeness a second example is considered; specifically an ARMA(4,2) inverse filter is used to restore the input signal using both Wiener Optimisation and Differential Evolution.

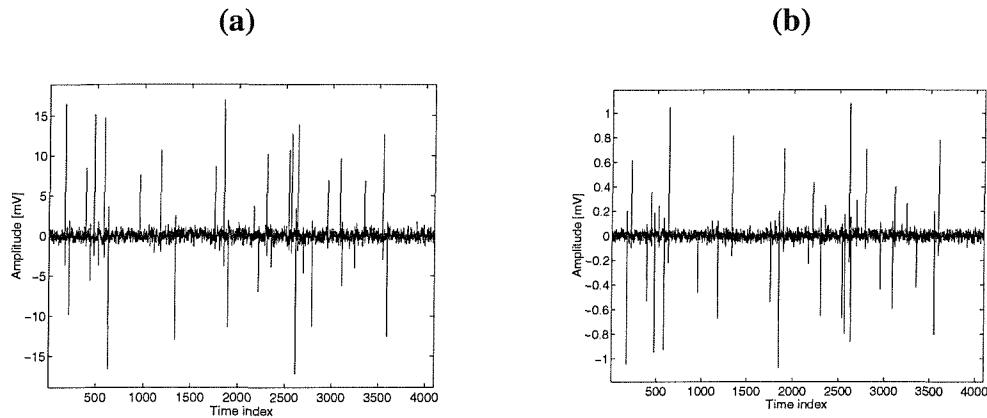


Figure 6.3.9 Restoration of signal from fourth-order Wiener solution and DE method using ARMA(4,2) inverse filter. (a): restored signal by using Wiener optimisation, (b) : restored signal by using DE optimisation method.

As with the MA(4) inverse filter, the restored signals using both the Wiener approach and the DE method are very similar (with only a different magnitude and sign reversal). The two results are compared in numerical form in the following table using the parameters defined in Chapter 3 and Chapter 5.

Table 6.3.1 Comparison of restored signals from Wiener optimisation and DE process

Parameters	Methods	Wiener optimisation	DE process	Remarks (Input/output)
MA(4) Inverse filter	α -coefficient	0.32	0.32	(0.29/0.35)
	Entropy	1.40	1.29	(0.07/1.23)
	Skewness	1.70	1.77	(3.05/0.76)
	Kurtosis	123.15	128.50	(206.98/92.83)
	CPU time	1	125.5	(normalised)
ARMA (4,2) Inverse filter	α -coefficient	0.32	0.32	-
	Entropy	1.42	1.32	-
	Skewness	2.09	-2.17	-
	Kurtosis	120.30	127.99	-
	CPU time	0.8	155.2	(normalised)

6.3.3 Computational efficiency of DE optimisation method

The term ‘computational efficiency’ used in this blind signal reconstruction problem relates to the ability of reaching the global maximum (in terms of the performance index and now also includes the convergence speed i.e., CPU time). We now compare the computational efficiency of blind signal reconstruction between the Wiener optimisation and the DE method.

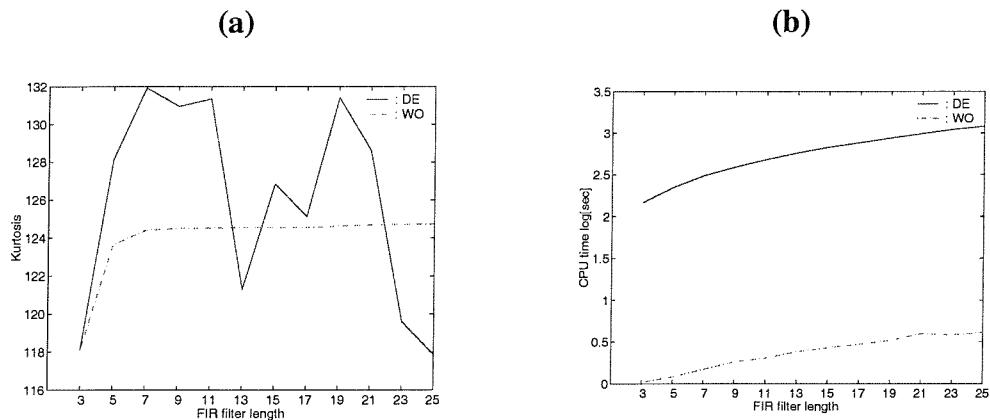


Figure 6.3.10 Comparison of signal restoration from fourth-order Wiener solution, and DE method. (a): kurtosis of restored signal, (b): computational time.

As can be seen in Figure 6.3.10 (a), the kurtosis of restored signal using Wiener optimisation (dotted line) and from DE optimisation method (solid line) for each FIR inverse filter length is similar. However, as shown in (b), the computational time required to yield similar kurtosis of the restored signal using the DE method is significantly greater than the Wiener optimisation.

The nature of the blind deconvolution involved in this work demands major computing power because of the following;

- (i) There exist many local maxima positioned closely.
- (ii) Unlike the Wiener approach which iterates a single parameter vector of inverse filter coefficients, DE iterates a set of parameter vectors leading to greatly increased computational loads. If the signal restoration using the Wiener approach fails then DE is also liable to produce a poor reconstruction despite having incurred the computational load.

We now reconsider the previous example using different measurement noise. Again, the objective is to compare the computation efficiency between the Wiener optimisation and the DE method. For the Wiener optimisation the maximum number of iterations is restricted to 100 and for the DE method the iteration is stopped.

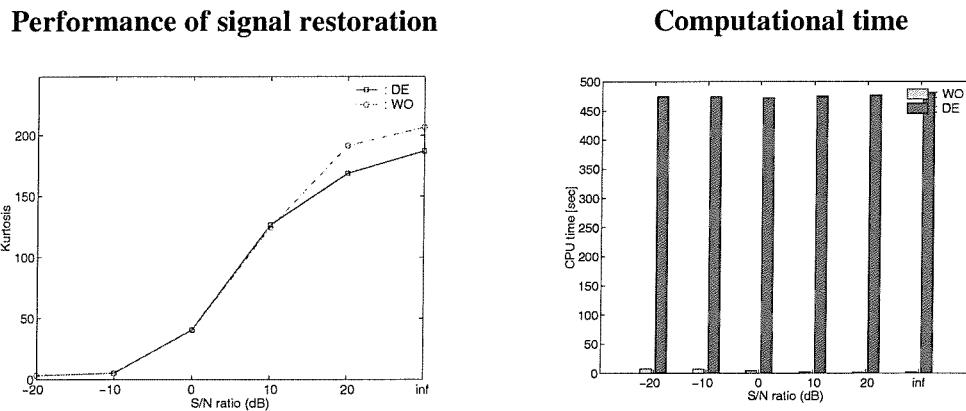


Figure 6.3.11 Comparison of signal restoration from fourth-order Wiener solution, and DE method (FIR inverse filter, $L=11$).

As can be seen in Figure 6.3.11, again we see the computational load is heavy for the DE method. Also, in the severe noise corruption cases (S/N ratio $20 \sim 0$ dB), we cannot expect ‘better’ reconstruction from DE method but also incur dramatically increased computational time.

From our extensive simulations and other work [Storn and Price, 1997], it is emphasised that three parameters govern the effectiveness of DE method in locating the global minimum. These are the population size (Figure 6.2.1), the scaling factor F (Figure 6.2.2), and the cross-over ratio CR (Figure 6.2.3). Guidance in selecting these parameters is given below;

The population size should be kept as small as possible so keeping the number of trial runs as low as possible. Use a population size P between five and ten times longer than the size of the parameter D as;

$$P \geq 5D \sim 10D.$$

We employ the FIR inverse filter because of its robustness. Confining the coefficients of the inverse filter to have unit norm is sufficient to explore the cost function (objective function) range as;

$$\|f_k\| \leq 1, \quad k = 0, 1, \dots, L-1.$$

The scaling factor F is determined by the constraint on the inverse filter coefficients;

$$0 < F < \|f_k\|.$$

The cross-over ratio CR should be small enough to increase the likelihood of 'birth' of a new vector with novel characteristics coming from the mutated vector. On the other hand, it should be large enough to maintain any existing inheritance. In this way, DE ensures the newly created trial vector provides many possibilities in the parameter space leading to a promising optimum. The CR is taken as;

$$0.1 \leq CR < F.$$

6.4 Summary and conclusion

In this chapter, we investigated the performance of just one of the numerous global optimisation methods in blind restoration of non-Gaussian impacting signals. This optimisation is the Differential Evolution (DE) method, which utilises an evolutionary process similar to the Genetic Algorithms (GA). The performance of signal reconstruction and computational efficiency are considered and compared to the Wiener optimisation method discussed in Chapter 3. The following conclusions arise;

1. The DE method can restore the input signal and is thus applicable to the Blind Deconvolution process.
2. Since the Wiener approach is believed to yield only one of the local maxima, the restoration of an impacting signal using a global optimisation method has been carried out and compared to the Wiener optimisation method. A major conclusion is that even through the Wiener optimisation results generally in a local maximum, the reconstruction seems very comparable to the global optimisation. We note here that in the Wiener approach, the choice of initial inverse filter type affects the performance of signal restoration. That is to say, by selecting an impulsive type initial inverse filter, the starting point of optimisation is initiated near an ‘acceptable’ local maximum.
3. Given the equivalence of results, DE scores relative low as a method because of its computational overhead.

To make DE a serious candidate for use in blind inversion, a further study relating to choice of filter length and computational efficiency is needed.

Chapter 7

Multiple Channel Signal Processing

7.1 Introduction

Multiple Input Multiple Output (MIMO) models arise frequently in signal processing (e.g., seismic exploration, digital communications, antenna array processing, biomedical signal processing, and multi-channel machine conditioning systems). The sensor signals from a p element array are a p -dimensional vector process representing the mixture of m different independent source signals. The propagation characteristics between m source signals and p sensor signals is often modeled as a linear (instantaneous or convolutive) MIMO system.

The Blind Source Separation (BSS) problem is a basic and difficult problem (e.g., [Bar-Ness et al, 1981]). Since then many solutions have been proposed, most of which are based on independence criteria which involve higher-order moments generated by non-linear functions [Jutten and Herault, 1991], cumulants [Lacoume and Ruiz, 1988; Cardoso, 1989; Comon, 1989] or contrast functions [Moreau and Macchi, 1993; Comon, 1994; Delfosse and Loubaton, 1995]. Most of these works are related to the separation of instantaneous linear mixtures of sources: the observation at any time t are a linear superposition (with real coefficients) of the sources at time t . For the linear instantaneous mixture model, source separation can be achieved by utilising the higher-order based joint diagonalisation of the observed signal matrix [Cardoso and Comon, 1990; Comon, 1994; Yellin and Weinstein, 1994; Cardoso, 1998a, 1998b; Lathauwer et al, 1999; Zhu et al, 1999]. A solution for this problem uses a time-based model in which the source separation is achieved through higher-order statistics and independency criteria.

This approach utilises the joint-cumulant of output signals.

Multichannel blind deconvolution (MBD) is an extension of the single channel blind deconvolution which utilises an objective function maximisation based on the central limit theorem [Donoho, 1981] and already discussed in Chapter 3. For this, a variable norm deconvolution method is introduced [Gray, 1979]. This is a multichannel technique which iteratively estimates an inverse to the unknown system which when convolved with the measured signals yields the unknown impacting signals. This inverse is estimated such that the resulting output signals are as spiky and non-Gaussian as possible. The proposed method is derived from statistical theory and utilises the non-linear optimisation technique discussed in Chapter 3 (relevant references are found in that chapter).

This chapter introduces multichannel blind source separation and multichannel blind deconvolution based on higher order statistics of signals from convolutive mixtures. In particular, we are concerned with the case that the number of inputs is the same as the number of outputs. Simulations for two input two output cases are carried out and their performances are assessed.

7.2 Model selection and problem formulation

7.2.1 General

The model of the convolutive mixture for multi-channel case is shown in Figure 7.2.1.

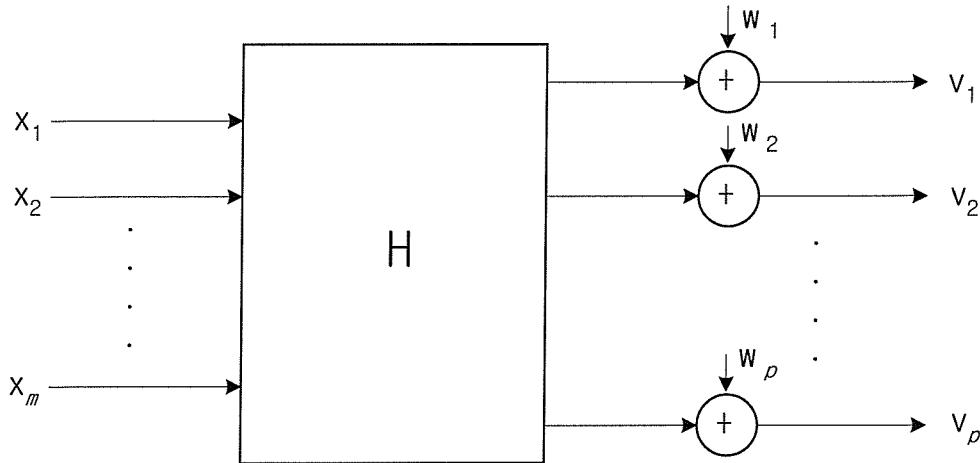


Figure 7.2.1 Multi-input multi-output system with unknown system \mathbf{H} and independent input signals \mathbf{x} , unknown white Gaussian noise \mathbf{w} , and observed signal \mathbf{v} .

In general situations, the observed signal $\mathbf{v}(k)$ at the sensor output is a p vector with components that result from the unknown m sources $\mathbf{x}(k)$ i.e.,

$$\begin{aligned}\mathbf{v}(k) &= [v_1(k) \cdots v_p(k)]^T \\ \mathbf{x}(k) &= [x_1(k) \cdots x_m(k)]^T\end{aligned}\tag{7.2.1}$$

and the input signals are assumed to be independent each other and at least one of the input signals is non-Gaussian.

The system \mathbf{H} is a $(p \times m)$ matrix of impulse responses (in the time domain) or equivalently a $(p \times m)$ matrix of transfer functions in the z-domain.

7.2.2 Two input two output systems and input-output relationship

The aim of this study focuses on the restoration of impacting signals from observed multichannel signals using both source separation and deconvolution.

We concentrate on a two input two output system based on models of two forms of cross coupling referred to as singly coupled (feed forward) and doubly coupled (feed back).

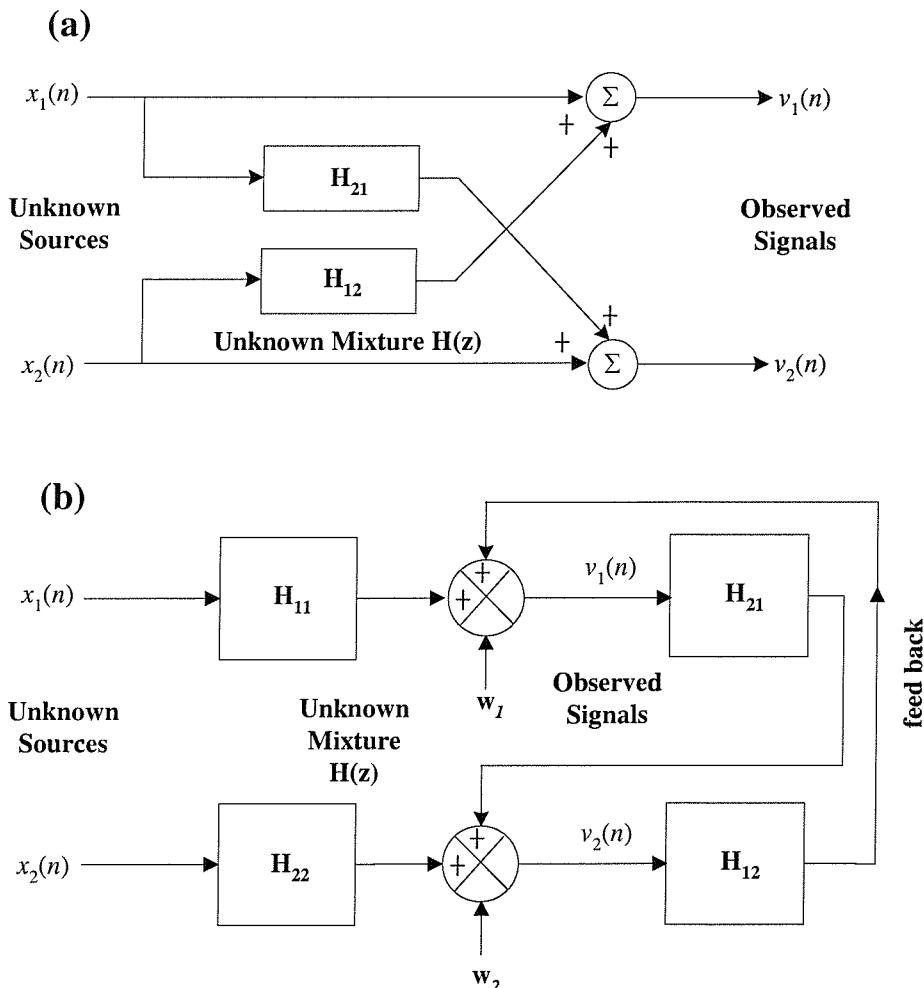


Figure 7.2.2 Two-input two-output system, (a): Single-coupled, (b): Double coupled model

Case (a);

For Figure 7.2.2, the observed signals are the results of single coupling of two input signals modelled as

$$v_i(n) = x_i(n) + \sum_{k=0}^{L-1} h_{ij}(k) x_j(n-k) \quad (7.2.2)$$

where $h_{ij}(k)$ is the k th coefficient of filter H_{ij} with $i \neq j$, $\forall i, j \in \{1, 2\}$.

The model described in Figure 7.2.2 is noise free ($w=0$) and $H_{11}=H_{22}=1$. These simplifications can be thought as representing measurement points near source positions with precise instrumentation. The input-output relation is

$$\begin{aligned} V_1(z) &= X_1(z) + H_{12}(z)X_2(z) \\ V_2(z) &= H_{21}(z)X_1(z) + X_2(z) \end{aligned} \quad (7.2.3)$$

where H_{12} and H_{21} are FIR filters satisfying $|h_{ij}(k)| < 1$.

Case (b);

When the observed signals are the result of ‘feed back’ as in Figure 7.2.2, they may be written as

$$v_i(n) = \sum_{k=0}^{L_i-1} h_{ii}(k) x_i(n-k) + \sum_{k=0}^{L_j-1} h_{ij}(k) x_j(n-k) + w_i(n) \quad (7.2.4)$$

where $h_{ij}(k)$ is the k th coefficient of filter H_{ij} with $\forall i, j \in \{1, 2\}$.

In contrast to the case of single coupling, this models the outputs of remote measurement points as affecting other outputs.

In the z-domain

$$\begin{aligned} V_1(z) &= H_{11}(z)X_1(z) + H_{12}(z)V_2(z) \\ V_2(z) &= H_{21}(z)V_1(z) + H_{22}(z)X_2(z) \end{aligned} \quad (7.2.5)$$

7.2.3 Blind Source Separation via joint-cumulant cancellation (nulling)

The structure of the source separation system is chosen as being of the form in Figure 7.2.3 [Yellin and Weistein, 1994]. Detailed descriptions for this structure are listed in Appendix F.

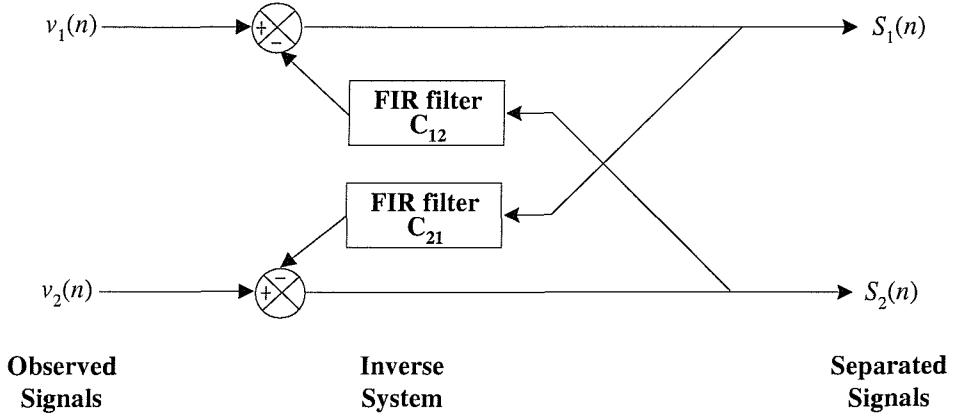


Figure 7.2.3 Recursive source separation structure

This is a recursive form [Jutten and Herault, 1991]. Using this recursive separation system based on the observed signals $v_1(n)$ and $v_2(n)$, the outputs $s_1(n)$ and $s_2(n)$ are calculated as

$$s_i(n) = v_i(n) - \sum_{k=0}^{L_{ij}} c_{ij}(k) s_j(n-k) \quad (7.2.6)$$

in which the updating of the filter coefficients (of the filters c_{ij}) are iterated by using the fourth-order joint cumulants cancellation of above two outputs which takes the form of the steepest descent algorithm as (for example),

$$c_{ij}(q+1, k) = c_{ij}(q, k) + \mu_{ij} E \{ s_i(n)^3 \cdot s_j(n-k) \} \quad , \quad i \neq j, \forall i, j \in \{1, 2\} \quad (7.2.7)$$

where q represents each iteration step.

The magnitude of adaptation gain should be in a range that can assure stable convergence and computational efficiency. In this study, this value commonly has

been used as 0.01 arising from the adaptation gain μ_{ij} in the range [Thi and Jutten, 1995],

$$0.005 \leq \mu_{ij} \leq 0.05$$

and the threshold t_r for the iteration to stop is set to [Cardoso, 1998]

$$t_r = \frac{1}{100 \cdot \sqrt{N}}$$

where N is the total number of data points.

7.2.4 Blind deconvolution via multichannel objective function maximisation

The structure of the multichannel deconvolution operator is similar to the single channel blind deconvolution operator discussed in Chapter 3. However, unlike the source separation structure, this deconvolution operator uses a single inverse filter which is shown below.

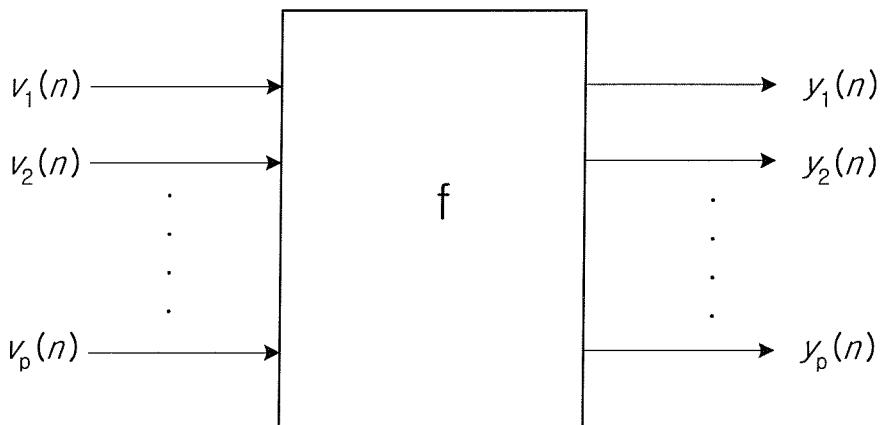


Figure 7.2.4 Multichannel blind deconvolution operator

As shown in Figure 7.2.4, since the deconvolution operator is commonly applied to the input signal to produce the output signal, the number of input signals (denoted by p) to the inverse system, which is the same as the number of output signals. The output signal coming from each channel ($j=1 \sim p$) is expressed as

$$y_j(n) = \sum_{k=0}^{L-1} v_j(n-k) \cdot f(k) \quad (7.2.8)$$

For the restoration of an impacting signal, we introduce a multichannel objective function as

$$O(\mathbf{y}, r) = \sum_{j=1}^p \frac{\frac{1}{N} \sum_{n=0}^{N-1} |y_j(n)|^r}{\left[\frac{1}{N} \sum_{n=0}^{N-1} |y_j(n)|^2 \right]^{r/2}} \quad (7.2.9)$$

where \mathbf{y} is the output p -vector of the inverse filter and the parameter r governs the relative weightings of the nature of an impulsive signal [Claerbout, 1973], which is similar to the single channel case. This parameter r is adjusted depending on the characteristics of input impacting signals concerned, defining ‘variable norm deconvolution’ [Gray, 1979].

The impacting signal reconstruction is achieved by maximising the objective function (the summation of normalised variable higher order moments or cumulants of each output signal) with respect to the filter coefficients.

The final equation of the multichannel inverse filter (deconvolution operator) is expressed as

$$\sum_{l=0}^{L-1} \left[\sum_{j=1}^p \frac{1}{u_{2j}} \sum_{n=0}^{N-1} v_j(n-l) \cdot v_j(n-k) \right] \cdot f(l) = \sum_{j=1}^p \left[\frac{1}{u_{1j}} \sum_{n=0}^{N-1} |y_j(n)|^{r-1} \cdot \text{sgn}(y_j(n)) \cdot v_j(n-l) \cdot v_j(n-k) \right] \\ k = 0, 1, \dots, L-1 \quad (7.2.10)$$

where

$$u_{1j} = \frac{1}{N} \sum_{n=0}^{N-1} |v_j(n)|^r$$

$$u_{2j} = \frac{1}{N} \sum_{n=0}^{N-1} y_j(n)^2$$

$v_j(n)$: n -th sample of the signal recorded on channel j (N samples per channel and p channels)

$f(l)$: l -th sample of an inverse filter having L length

$y_j(n)$: n -th sample of the output of the inverse filter coming from channel j (N samples per channel and p channels)

The inverse filter equation (7.2.10) can be expressed in matrix form

$$\mathbf{R} \cdot \mathbf{f} = \mathbf{g} \quad (7.2.11)$$

where \mathbf{R} is the autocorrelation matrix of observed signals, \mathbf{f} is the inverse filter coefficient vector, and \mathbf{g} denotes the cross-correlation of the observed signals and the output of the inverse filter.

7.3 Simulations of source separation and deconvolution for impacting signal reconstruction

The objective of this sub-section is to restore an impacting signal from the measured (observed) signals through source separation and blind deconvolution.

7.3.1 Signals and systems for simulations

The two input signals consist of an impacting signal and a Gaussian signal (normal operating excitation) are shown in Figure 7.3.1. For each input signal, N represents the number of samples, γ_3 and γ_4 are the skewness and kurtosis of the signal, respectively.

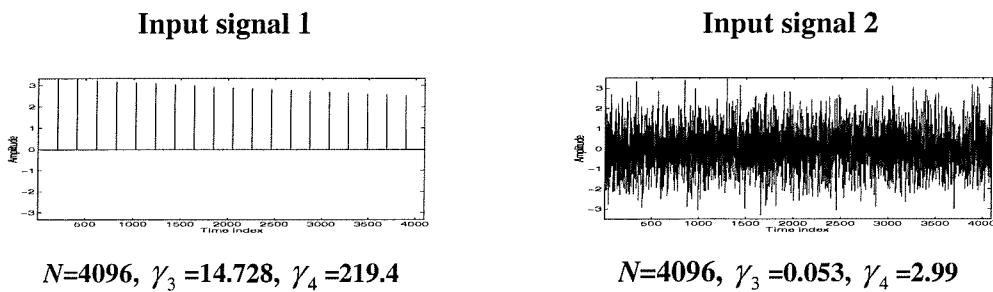


Figure 7.3.1 Input signals

Using these input signals, examples of the blind source separation from observed signals and restoration of the impacting signal procedures are demonstrated in two cases as follows;

Case (a): Two-input two-output without feedback (single coupling)

Since with $H_{ii} = H_{jj} = 1$, the observed signals have been generated as in Figure 7.2.2 (a), which can be redrawn in Figure 7.3.2.

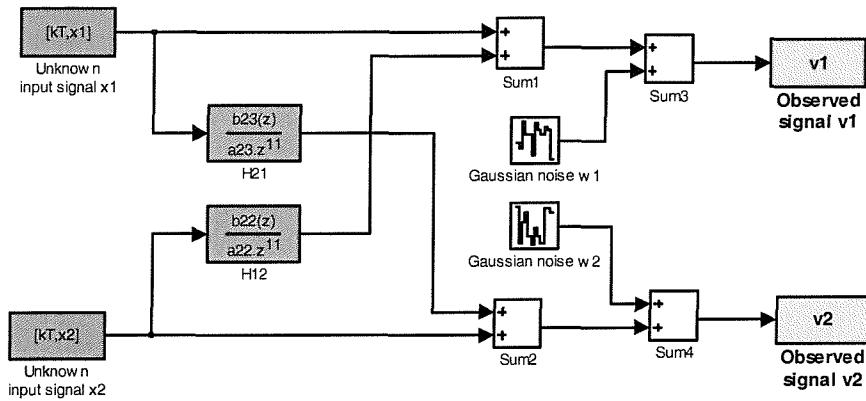


Figure 7.3.2 Simulation model for case (a)

In this simulation, the unknown system is selected to have a FIR filter length of 10 ($L=10$) and one Gaussian and one non-Gaussian impacting signal is filtered under noise free conditions ($w_i(n) = 0$).

Considering a 2-input 2-output MA(10) system model and $m=2$ and $p=2$ in equation.(7.2.3). Its 2×2 transfer function $H(z)$ is chosen as

$$H(z) = \begin{bmatrix} 1 & H_{12}(z) \\ H_{21}(z) & 1 \end{bmatrix} \quad (7.3.1)$$

where $H_{12}(z) = [0.03 \ -0.15 \ 0.35 \ -0.42 \ 0.2 \ 0.23 \ 0.5 \ 0.48 \ -0.26 \ 0.08 \ -0.01]z^{-l}$ and

$$H_{21}(z) = [0.1 \ -0.34 \ 0.56 \ -0.57 \ 0.33 \ -0.02 \ -0.19 \ 0.23 \ -0.15 \ 0.06 \ -0.01]z^{-l}$$

The unknown systems are shown in the following figure;

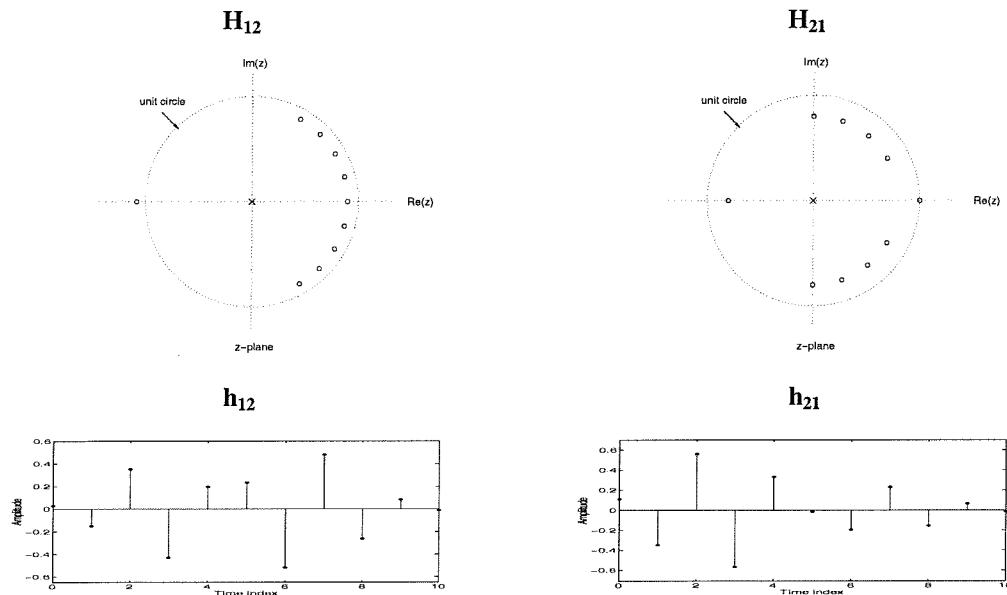


Figure 7.3.3 Unknown systems' pole-zero map and impulse response function shape used in simulation case (a)

The outputs of this system are shown below;

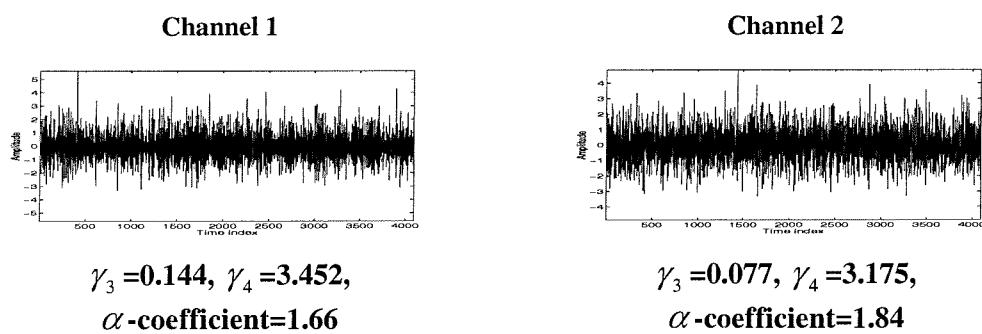


Figure 7.3.4 Two channel observed signal from the simulation case (a)

Case (b): Two-input two-output with feed back (double coupling)

For a more practical situation, the observed signal includes feed back of system outputs as in Figure 7.2.2 (b) and illustrated as a simulation model in the following figure;.

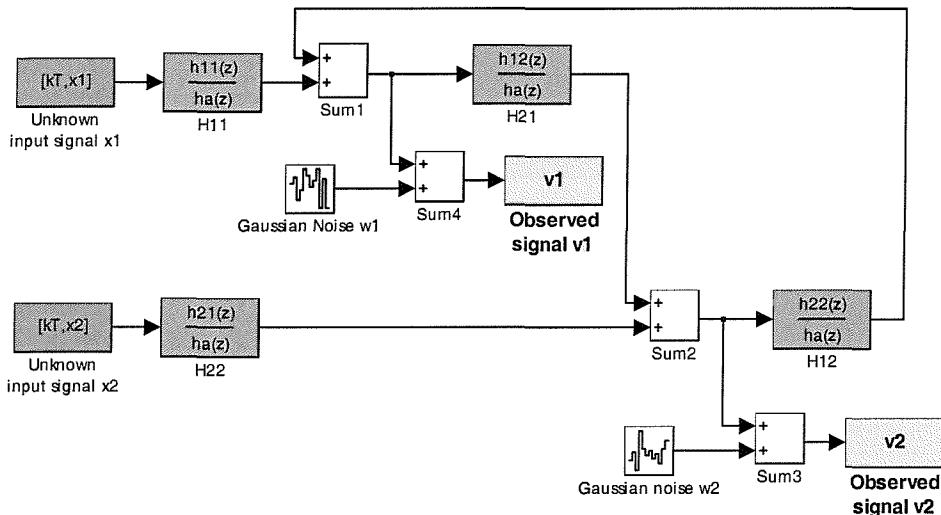


Figure 7.3.5 Simulation model for case (b)

We assume that $H_{11}(z) \neq H_{22}(z) \neq 0$, $\tilde{H}_{12}(z)$ and $\tilde{H}_{21}(z)$ have finite impulse responses and that we assume noises are zero. The unknown system MA(6) is modelled as

$$H(z) = \begin{bmatrix} 0.2 + 0.8z^{-1} + 0.4z^{-2} & 0.5 - 0.3z^{-1} \\ 0.3z^{-1} - 0.6z^{-2} & -0.21z^{-1} - 0.5z^{-2} + 0.72z^{-3} + 0.36z^{-4} + 0.12z^{-6} \end{bmatrix} \quad (7.3.2)$$

The input signals used in this simulation case are the same signals as those used in case (a) (Figure 7.3.1). The system's pole-zero maps and impulse response function shapes with their output signals are shown in the following figure;

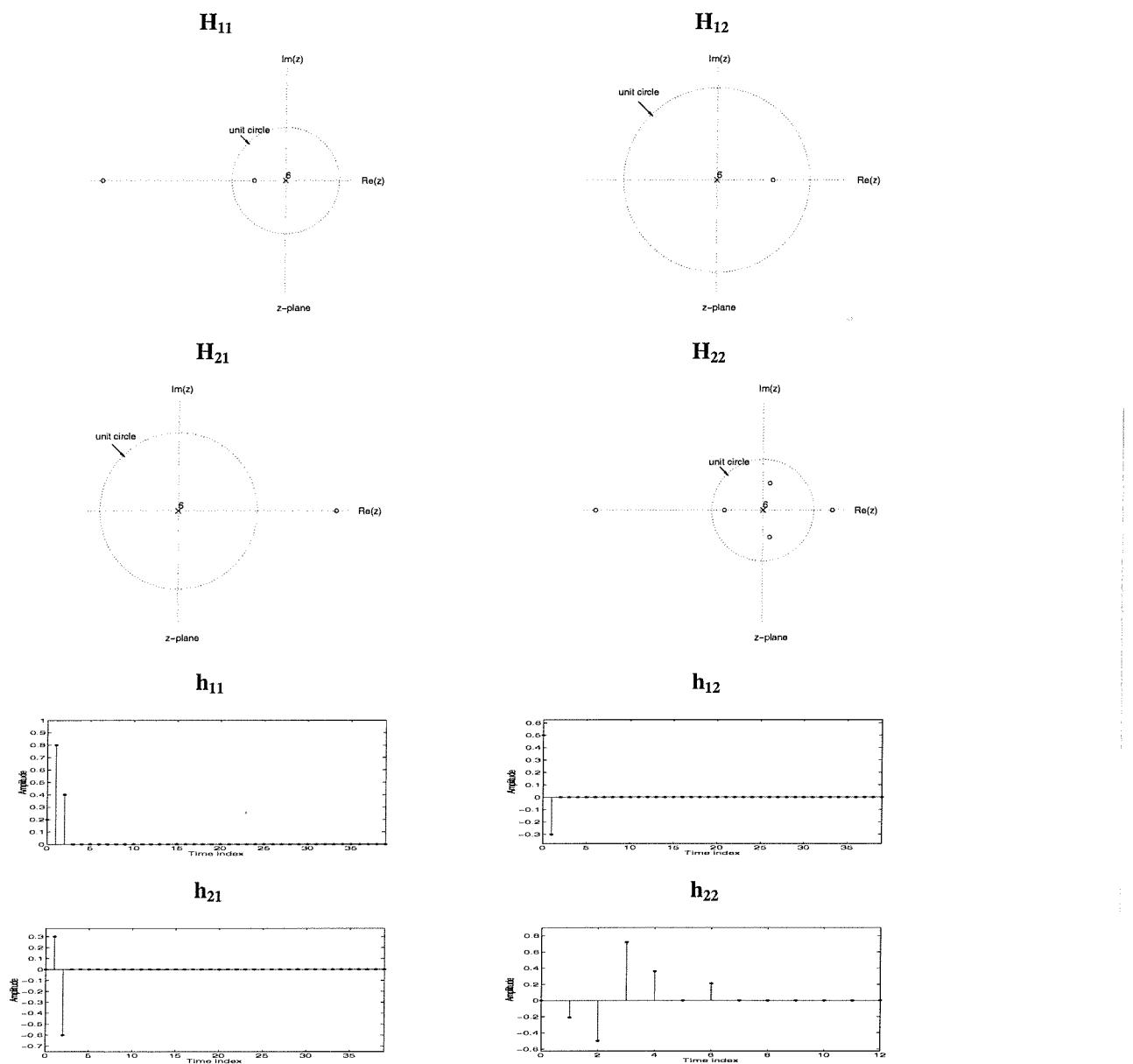


Figure 7.3.6 Unknown systems used in case (b)

The outputs of this system are shown below;

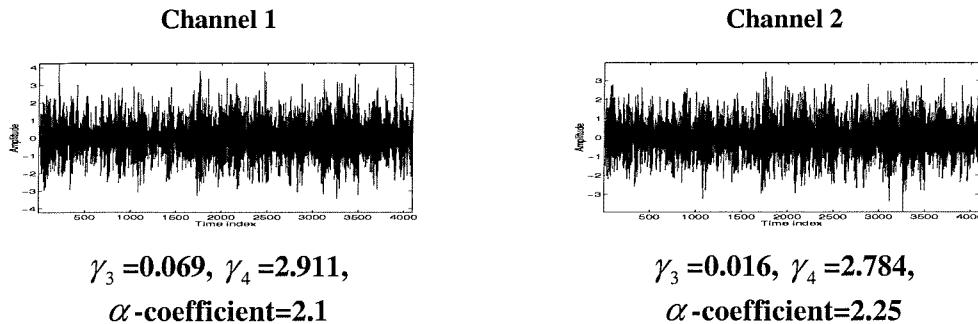


Figure 7.3.7 Two channel observed signal from simulation case (b)

Using two observed signals in each simulation case, we tried to identify the impacting signal through Blind Source Separation (BSS) and Multichannel Blind Deconvolution (MBD).

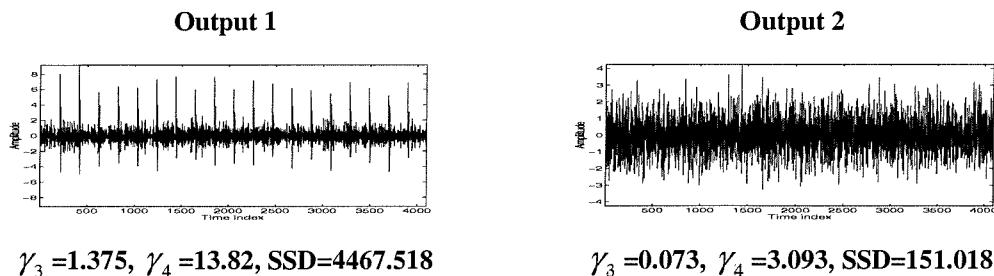
The results are shown in subsequent section with the shapes of the separated/restored signals and their statistical parameters summarised in table.

7.3.2 Results and discussion

The impacting signal is restored from both the BSS and MBD process for two different simulation cases. The performance of each method is compared.

Results of signal separation and deconvolution

Separated signal from BSS ($\mu = 0.01$)



Restored signals from multichannel blind deconvolution

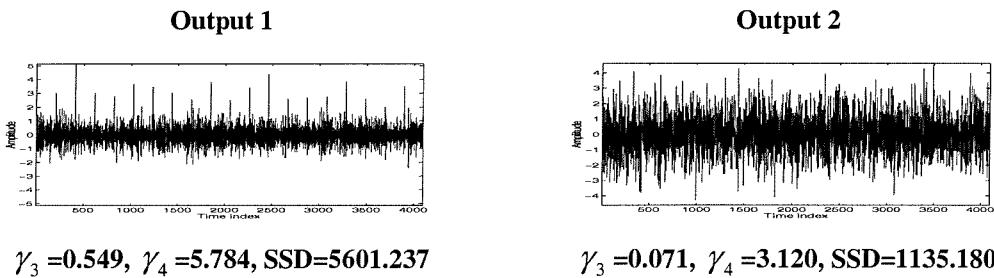
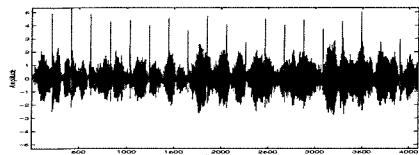
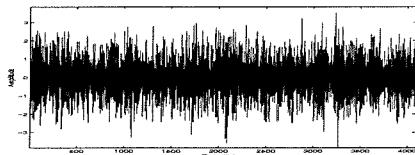


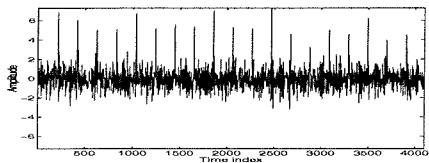
Figure 7.3.8 Separated and restored signals of simulation case (a)

Separated signal from BSS ($\mu = 0.01$)**Output 1**

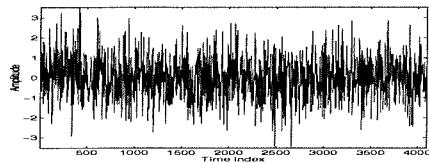
$$\gamma_3 = 0.345, \gamma_4 = 3.808, \text{ SSD} = 5898.843$$

Output 2

$$\gamma_3 = -0.023, \gamma_4 = 2.874, \text{ SSD} = 13948.460$$

Restored signals from multichannel blind deconvolution**Output 1**

$$\gamma_3 = 1.532, \gamma_4 = 9.548, \text{ SSD} = 5377.893$$

Output 2

$$\gamma_3 = 0.092, \gamma_4 = 3.243, \text{ SSD} = 13027.086$$

Figure 7.3.9 Separated and restored signals of simulation case (b)

Overall comparison for impacting signal restoration from both BSS and MBD process is given by the following table.

Table 7.3.1 summarises the statistical performances and features of each signal to compare the impacting signal restoration through the BSS and MBD process.

Table 7.3.1 Comparison of impacting signal restoration results from BSS and MBD process

Simulation case 1; $L=11$		
	BSS ($\mu=0.01, t_r=1.4e-3$)	MBD
skewness	1.375	0.549
kurtosis	13.82	5.784
SSD	4467.518	5601.237
Restored impact signal		
Simulation case 2; $L=6$		
	BSS ($\mu=0.01, t_r=2.6e-3$)	MBD
skewness	0.345	1.532
kurtosis	3.808	9.548
SSD	5898.843	5377.893
Restored impact signal		

Discussions

1) Discussion for source separation process (iteration stopping criteria)

Restoration of the impacting signal from observed signals through BSS is based on the statistical independency of the input signal. Hence, the independency criteria represented by the higher order sense are inspected and used as stopping criteria for iterative calculation of filter coefficients in equation (7.2.7). The various independency measures (higher order joint cumulants) are applied and compared in the following table.

Table 7.3.2 Comparison of impacting signal separation results for both case (a) and case (b) simulation

Simulation case (a); $L=11$, $\mu=0.01$, $t_r=1.4e-3$							
	J_1	J_2	J_3	J_4	J_5	J_6	J_7
Iteration	32	-*	218	129	13	168	225
Skewness	1.375	-	0.021	0.946	0.419	0.366	0.096
Kurtosis	13.82	-	3.011	9.983	4.863	5.258	3.153
SSD	4.4e+3	-	9.2e+3	4.8e+3	5.73e+3	5.71e+3	6.7e+3

Simulation case (b); $L=6$, $\mu=0.01$, $t_r=0.0026$

	J_1	J_2	J_3	J_4	J_5	J_6	J_7
Iteration	183	126	36	209	62	58	458
Skewness	0.345	0.352	0.230	0.342	0.322	0.398	0.101
Kurtosis	3.808	3.814	3.48	3.835	3.731	3.647	3.003
SSD	5.89e+3	5.88e+3	6.2e+3	5.89e+3	5.93e+3	5.97e+3	6.6e+3

‘*’: not converges to the given threshold up to 500 iterations

The various higher order joint cumulants expressed in Table 7.3.2 are defined below (Cum_{ij} , $i + j = 4$ denotes cross cumulants);

$$\mathbf{J}_1 : |Cum_{22}\{s_1(n), s_2(n)\}|$$

$$\mathbf{J}_2 : |Cum_{31}\{s_1(n), s_2(n)\}|$$

$$\mathbf{J}_3 : |Cum_{31}\{s_2(n), s_1(n)\}|$$

$$\mathbf{J}_4 : E\{Cum_{22}\{s_1(n), s_2(n-k)\}\} \text{ for } k = 0, 1, \dots, L-1 \text{ and takes minimum value}$$

$$\mathbf{J}_5 : E\{Cum_{22}\{s_2(n), s_1(n-k)\}\} \text{ for } k = 0, 1, \dots, L-1 \text{ and takes minimum value}$$

$$\mathbf{J}_6 : E\{Cum_{31}\{s_1(n), s_2(n-k)\}\} \text{ for } k = 0, 1, \dots, L-1 \text{ and takes absolute minimum value}$$

$$\mathbf{J}_7 : E\{Cum_{31}\{s_2(n), s_1(n-k)\}\} \text{ for } k = 0, 1, \dots, L-1 \text{ and takes absolute minimum value}$$

As a result, with the given adaptation gain μ commonly selected as 0.01, and each threshold values t_r , it has been turned out that the criterion \mathbf{J}_1 ($|Cum_{22}\{s_1(n), s_2(n)\}|$) can yield consistent identification for both case (a) and case (b) simulation.

2) The variable norm parameter selection for the multichannel blind deconvolution process

As expressed in equation (7.2.9), the multichannel objective function contains the variable norm value which can be any real value r ($r > 2$). From two different simulations, we observed the kurtosis of the restored impacting signal for each r value selection.

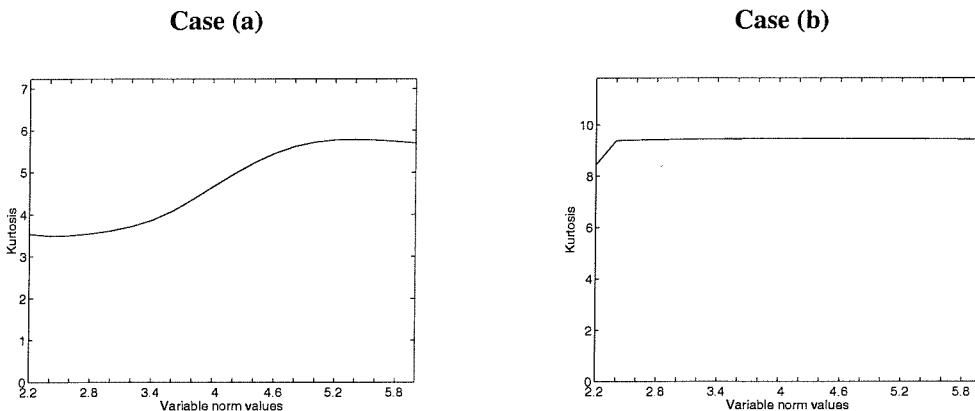


Figure 7.3.10 The effect of variable norm value on the impacting signal restoration through MBD process

For example, when r takes the value 4, the multichannel objective function corresponds *kurtosis* and the algorithm becomes the multichannel ‘Minimum Entropy Deconvolution (MED)’ [Wiggins, 1977].

For simulation case (a), the kurtosis of the restored impacting signal varies as the selection of r changes, whereas there are consistent results for simulation case (b) especially for $r > 2.4$.

The effect of these r values become less significant when the observed signals are close to Gaussian and possess an α -coefficient over 2 (see Figure 7.3.7). This means that for the problem of impacting signal reconstruction with severe noise corrupted observed signals, r can be selected to be greater than 4 (for more detailed information about variable r deconvolution, refer to Gray’s thesis [Gray, 1979]). In this study, we selected the r as 5.3 which gives the maximum kurtosis value for simulation case (a).

7.4 Summary and conclusion

In this chapter, we have addressed the problem of the blind source separation and deconvolution of sources for convolutive mixtures for the reconstruction of an impacting signal among a set of observed signals. A theoretical derivations and numerical simulations have been carried out assuming that one of the sources are non-Gaussian (impacting signal) and the other signals are Gaussian and are independent of each other.

We showed that the cancellation of fourth-order output joint-cumulants leads to a satisfactory condition for the source separation and non-Gaussian impacting signal reconstruction. Three different types of joint-cumulants have been

$Cum_{22}\{s_i(n), s_j(n-k)\}$, $Cum_{31}\{s_i(n), s_j(n-k)\}$, and $Cum_{13}\{s_i(n), s_j(n-k)\}$. In fact, the selection of these output joint-cumulants possibly depends on the nature of the unknown signals and is not known a priori. Nevertheless, focusing on the highly impulsive signal dealt in this study, the $Cum_{31}\{s_i(n), s_j(n-k)\}$ is considered to be the most effective cost function (updating function), and is thus employed here. To check the degree of independency of the separated signal, the criterion has been selected as $|Cum_{22}\{s_1(n), s_2(n)\}|$. This factor has shown consistency of convergence from numerous simulation results (refer Table 7.3.2).

The multichannel blind deconvolution process which utilises variable norm values as a multichannel objective function is introduced.

For a complicated system having multi-coupling or heavy noise interference, the multichannel blind deconvolution approach turns out to be a more effective method (see Table 7.3.1).

The selection of the variable norm values (r) depending on the status of each input and output signal of the inverse system remains to be studied further.

CHAPTER 8

Application to a mechanical impacting problem

8.1 Introduction

In this chapter, the behaviour of a randomly excited vertical cantilever beam with an endstop is investigated experimentally. The aim of this chapter is the practical verification of blind signal separation and recovery of the impacting signal. In the context of ‘condition monitoring’ of a mechanical system, this might be a key element related to fault detection. The aim would be to obtain a diagnostic tool that avoids false alarms. In a practical system, there may be (unexpected) nonlinearities and other hidden effects among the signals and structures. Accordingly, in our study, the experimental conditions are controlled closely to ensure that the mathematical models used in previous chapters are relevant.

Figure 8.1.1 illustrates a simple schematic view of a mechanical system in which an impacting signal arises due to an end stop. In a practical case, we assume the detailed form of the system and input sources are unavailable. The impulse signal due to the impacting between the moving body and surrounding structure is to be identified from the measured data (in this case the acceleration signal). The measured signal consists of the response of the mass-spring-damper or beam structure when the excitation and the impact signal acting simultaneously on the structure.

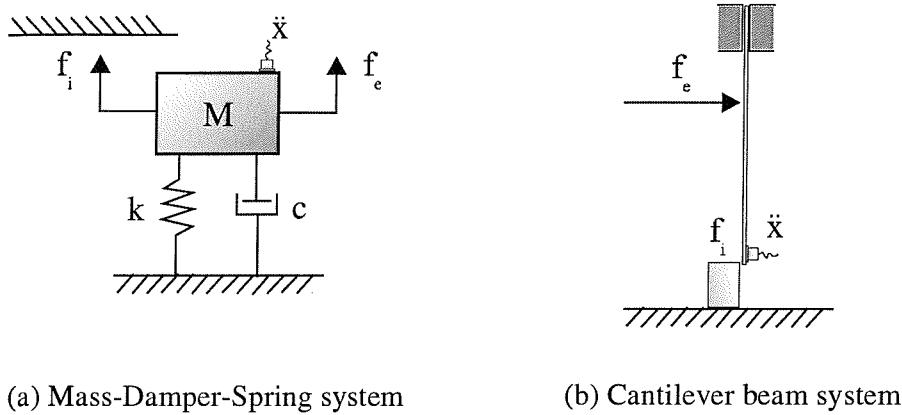


Figure 8.1.1 Mechanical model for blind deconvolution

8.2 Description of the mechanical impacting system

This section provides i) detailed descriptions of the experimental equipment used, ii) graphical illustration for data acquisition, iii) definitions of signals obtained, iv) relationship between the signals and v) pre-examination of the measured signals in each impacting or non-impacting test.

8.2.1 Instrumentation and experimental layout

Table 8.2.1 has summarised the details of equipment used in this experiment.

Table 8.2.1 Equipment list and settings

No.	Equipment name	Model	Serial No.	Status and settings
1	Accelerometer	B&K 4375	1238990	Sensitivity : 0.318 pC/ms ⁻²
2	Charge Amplifier 1	B&K 2635	1575831	Freq. range : 2Hz-1kHz Connect to force transducer on exciter Output : 100mV/Unit
3	Charge Amplifier 2	B&K 2635	1278234	Freq. Range : 2Hz-1kHz Connect to force transducer on impact Output : 100mV/Unit
4	Charge Amplifier 3	B&K 2635	1690266	Freq. Range : 2Hz-1kHz Connect to accelerometer Output : 10mV/Unit
5	DAT recorder.	TEAC RD-135T	7230701	Volt range : 2Volt for Ch1 : Signal generator Ch2 : Force transducer on impact point Ch3 : Accelerometer Ch4 : Force transducer on exciter Tape speed : x1 Mode : PGM
6	Exciter	LDS M101	9444	3V random excitation from R9211C
7	Force transducer 1	B&K 8200	1321344	Sensitivity : 4.13 pC/N
8	Force transducer 2	B&K 8200	1288285	Sensitivity : 3.82 pC/N
9	Low pass filter	Kemo 2ch. VBF 8 MK4	810806	Freq. range : 400 Hz Cut-Off to force transducer
10	Oscilloscope	Thander T0315	000158	Scale : 0.5V/div, dc Sweep time : 5ms/div, impact signal monitoring
11	Power Amplifier	790 series	139	Normal output : x1 Out
12	Signal analyser	Advantest R9211C	22020161	Random signal output : 3V, offset 0.0V Time signal sampling rate : 1.95msec (781μsec) Frame time : 2048 samples Ch A : Force transducer with sensitivity – 10dBV Ch B : Accelerometer with sensitivity 3dBV

The experimental layout is shown in Figure 8.2.1 in which a cantilever beam is driven from a random source through an exciter. The primary (only) input to the system is this exciter and impacting is induced as a consequence. However, as we

shall see in Section 8.2.2, we will model this as a two input linear system (the inputs being the force and the impacts due the end stop).

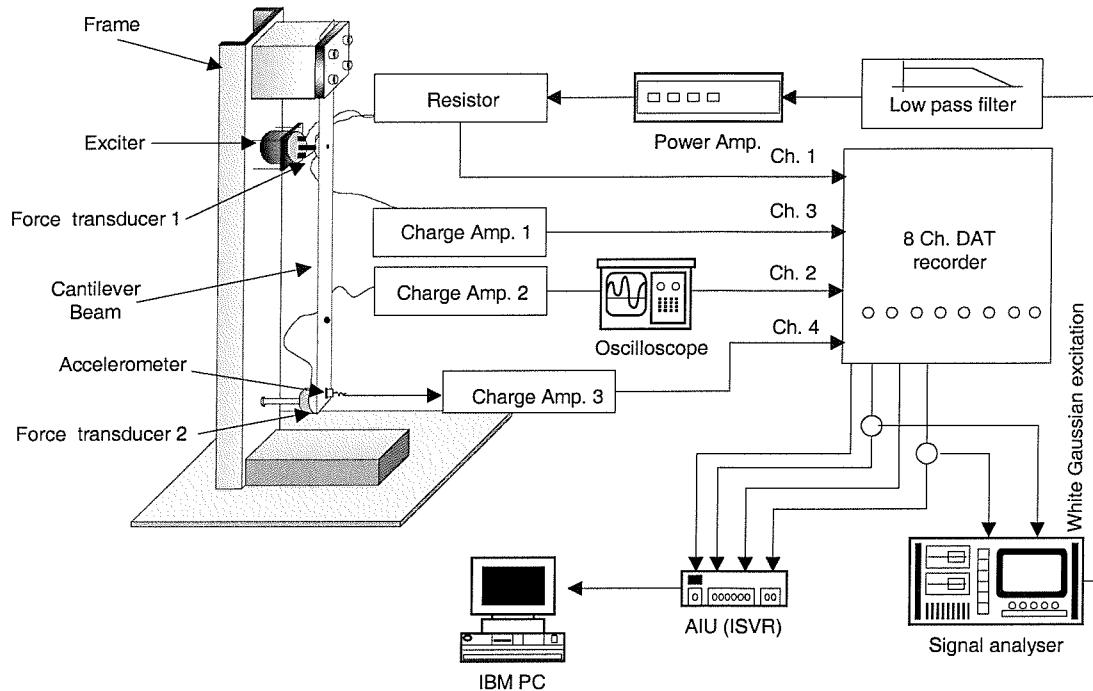


Figure 8.2.1 Experimental set-up for blind signal separation

With reference to Figure 8.2.1, a broad band Gaussian signal is the excitation. Impacting is induced by placing an end stop restricting the motion of the beam (at the beam tip). The Gaussian excitation signal is produced from a signal analyser (Ch.1) and fed to the exciter. The signal from the exciter (Ch.3) is responsible for the movement of the beam. The impacting signal (Ch. 2) is generated by the end stop incorporating a force transducer which is placed 1.5 mm behind the beam's steady state position. The accelerometer is attached at the free end of the cantilever beam to collect the signal (Ch. 4) mixed with the vibration signal of the beam and impacting signal caused by the end stop. Note the impacting signal (Ch. 2) is captured so as to assess the performance of the inversion process – normally of course this would be unavailable.

8.2.2 Definitions and relationship of experimental data

Model

The whole structure of Figure 8.2.1 can be thought as a system shown below;

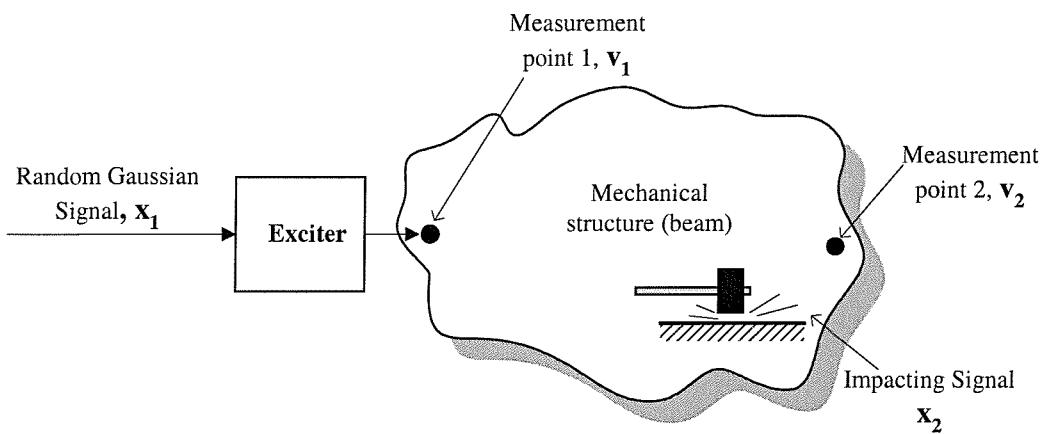


Figure 8.2.2 Structural components of beam excitation

This structure is a complicated system as the impacting signal x_2 is **nonlinearly** related to the excitation signal x_1 , and so the analysis of this situation thus becomes a very complex problem. In order to restore the hidden impacting signal from measured signals, we intend to reduce this complexity to a simplified linear equivalent system modelled as a two input two output system. A simplified experimental layout is shown in Figure 8.2.3.

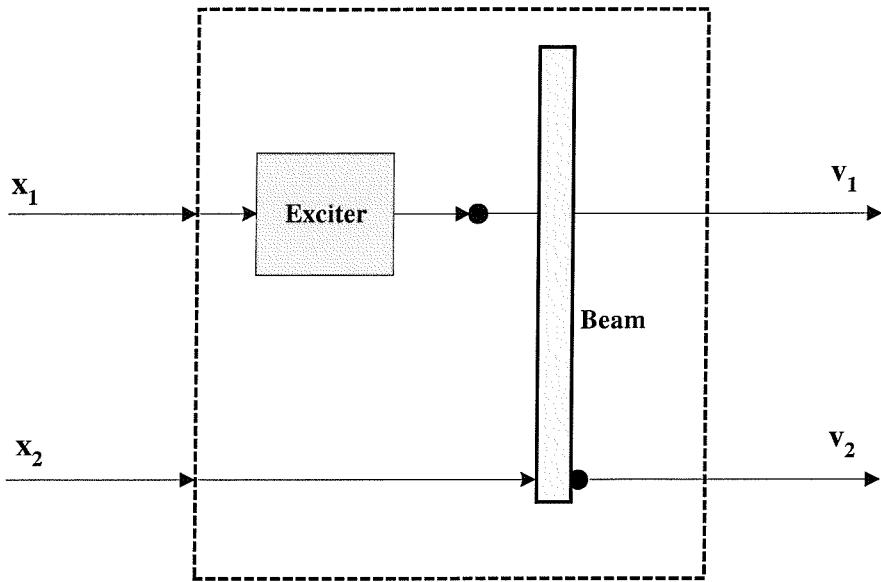


Figure 8.2.3 Simplified input-output structure for beam excitation experiment

The definitions of signals given in the simplified system (Figure 8.2.3) are as follows;

- x_1 : unknown input signal 1 (Gaussian driving signal, recorded in Ch. 1)
- x_2 : unknown input signal 2 (non-Gaussian impacting signal, recorded in Ch. 2)
- v_1 : measured signal 1 (force transducer signal, recorded in Ch. 3)
- v_2 : measured signal 2 (accelerometer signal, recorded in Ch. 4)

Note that the force transducer signal v_1 is the measured at the point where the force transducer is in direct contact with the structure (the beam).

Relationship of signals

Figure 8.2.3 illustrates the situation where the white Gaussian excitation force (normal operating signal) acts simultaneously with the impacting signal on the beam. The accelerometer signal v_2 is the sum of the beam motion driven by the forcing term v_1 and the impacting term x_2 . Due to the motion of the beam, the force transducer signal v_1 is affected by the feed back effect. Hence, the force transducer signal contains both the exciter output and beam motion. Suppose we

measure both the force transducer signal v_1 and accelerometer signal v_2 in Figure 8.2.3, the two input and two output system can be arranged as follows

$$\begin{aligned} V_1 &= H_1 X_1 + H_4 V_2 \\ V_2 &= H_2 V_1 + H_3 X_2 \end{aligned} \quad (8.2.1)$$

Inverting this

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{H_1}{1-H_2H_4} & \frac{H_3H_4}{1-H_2H_4} \\ \frac{H_1H_2}{1-H_2H_4} & \frac{H_3}{1-H_2H_4} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad (8.2.2)$$

Denoting

$$\begin{aligned} G_1 &= \frac{H_1}{1-H_2H_4}, & G_2 &= \frac{H_3H_4}{1-H_2H_4} \\ G_3 &= \frac{H_1H_2}{1-H_2H_4}, & G_4 &= \frac{H_3}{1-H_2H_4} \end{aligned} \quad (8.2.3)$$

Then, (8.2.2) can be expressed

$$\begin{aligned} V_1 &= G_1 X_1 + G_2 X_2 \\ V_2 &= G_3 X_1 + G_4 X_2 \end{aligned} \quad (8.2.4)$$

Assuming the accessible signals are V_1 (Force transducer signal) and V_2 (Accelerometer signal), and X_1 is Gaussian, using the accelerometer signal (V_2) can provide the reconstruction of X_2 and G_4 , whereas the use of the force transducer signal (V_1) can provide X_2 and G_2 .

8.2.3 Survey of the status of the measured signals

Two experiments are carried out in this study, (i) an impacting case, (ii) a non-impacting case. Each of these will be discussed below.

Impacting case: the end stop is positioned to ensure that impacting occurs.

From Figure 8.2.3, we can model the structural components and signals as follows

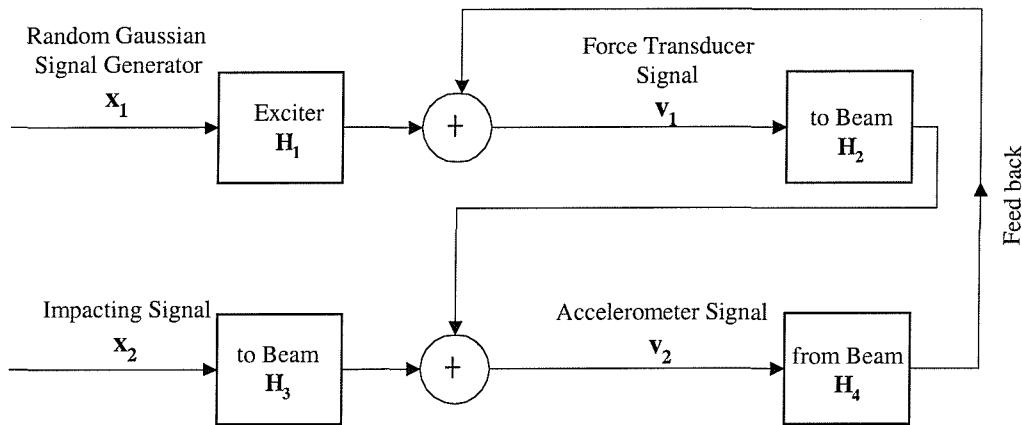


Figure 8.2.4 Two input two output model

The measured signals are shown below.

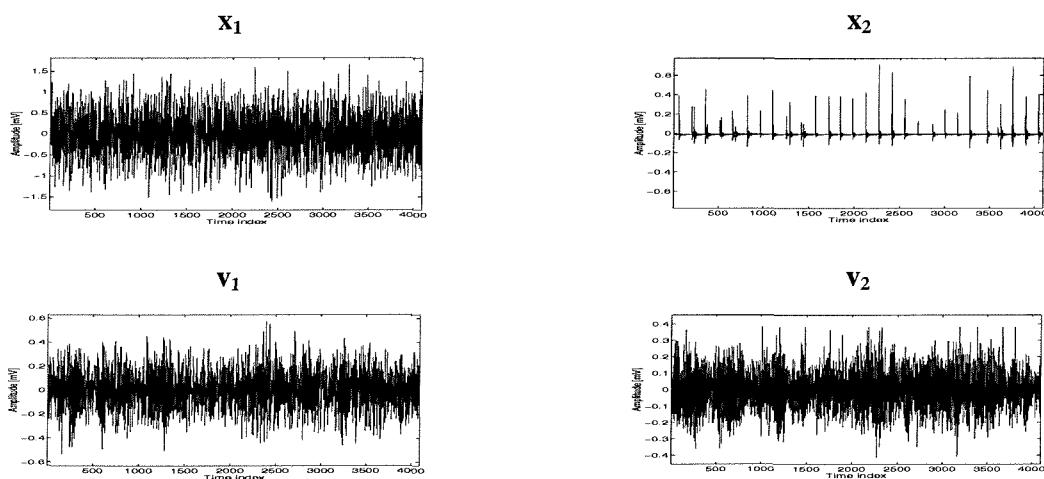


Figure 8.2.5 Signals from impacting experiment; x_1 : unknown Gaussian excitation signal, x_2 : impacting signal, v_1 : force transducer signal, v_2 : accelerometer signal.

Non-impacting case: the end stop is removed to allow the beam move freely.

When the Gaussian driving signal alone drives the beam. The model becomes,

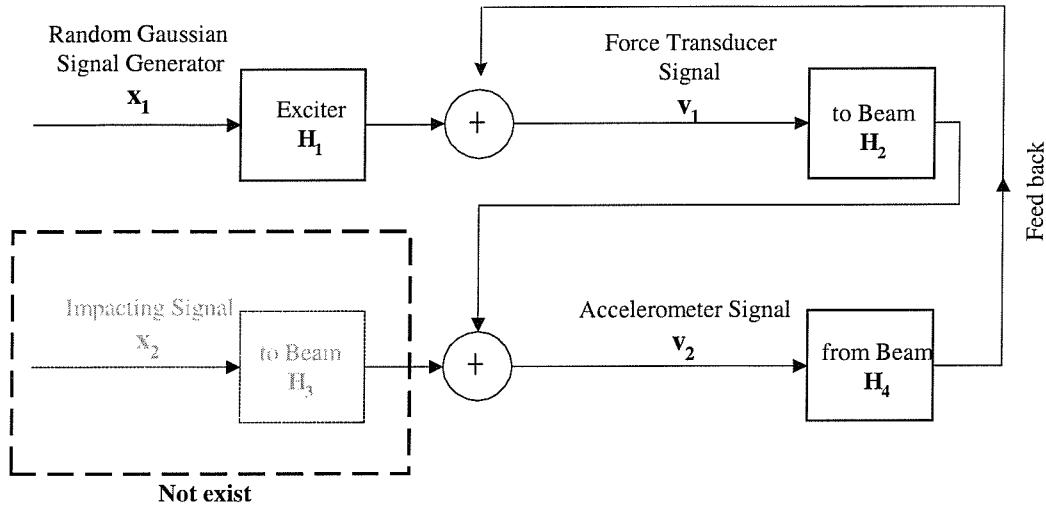


Figure 8.2.6 Single input two outputs model and deconvolution

The signals are shown below.

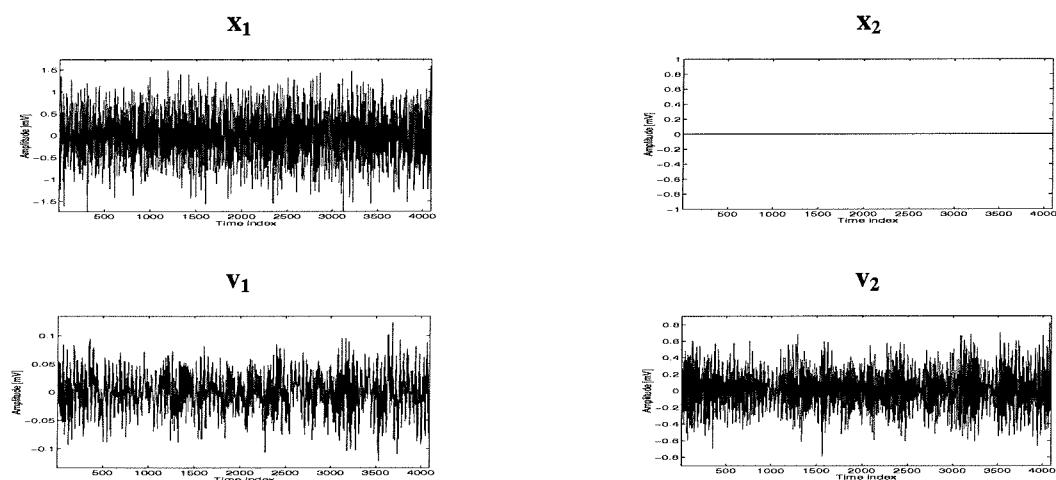


Figure 8.2.7 Signals from the non-impacting experiment; x_1 : unknown Gaussian excitation signal, x_2 : impacting signal (zero), v_1 : force transducer signal, v_2 : accelerometer signal.

From these signals, we assume we can measure (in practice) only two signals (i.e., force transducer signal, v_1 and accelerometer signal, v_2). The statistical properties of the signals are summarised in the table below;

Table 8.2.2 Statistical properties of each signal

Impacting case				
Signals Statistical parameters	Input signals (unknown)		Observed signals	
	x_1	x_2	v_1	v_2
Number of data points (N)	4096	4096	4096	4096
Skewness	-0.0186	9.454	-0.006	0.066
Kurtosis	2.9001	128.533	2.994	2.983
Non-impacting case				
Signals Statistical parameters	Input signals (unknown)		Observed signals	
	x_1	x_2	v_1	v_2
Number of data points (N)	4096	4096	4096	4096
Skewness	-0.049	-	-0.005	-0.049
Kurtosis	2.904	-	2.994	2.928

Impacting signal detection

Using the Higher Order Singular Value Decomposition (HOSVD) described in Chapter 4, Section 4.3, detection of the impacting signal is carried out.

We consider three different cases: (i) using only the force transducer signal (v_1), (ii) using only the accelerometer signal (v_2), (iii) using two signals (multichannel approach).

The results are shown below.

(i) using only the force transducer signal (v_1 signals from Figure 8.2.5 and Figure 8.2.7)

As a reminder of the terms used in Figure 8.2.8, D_3 represents the detection parameter originated from the third order cumulant tensor and D_4 from the fourth order cumulant tensor. The dotted line in each figure is the threshold of detection. Thus, when the bar graph exceeds the dotted line, it means that an impacting signal exists in the measured signals.

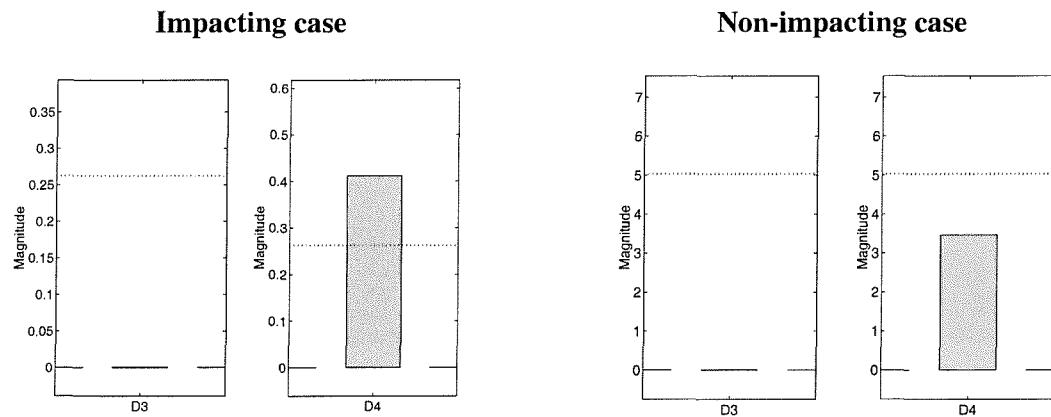


Figure 8.2.8 Detection results from HOSVD.

From the left D_4 graph in Figure 8.2.8, we can verify the existence of an impacting signal, which corresponds to the situation when the beam is excited with the end stop in position. In contrast, the right D_3 or D_4 graph tells us that there is no impacting signal.

(ii) using only the accelerometer signal (v_2 signals from Figure 8.2.5 and Figure 8.2.7)

We see the results in Figure 8.2.9.

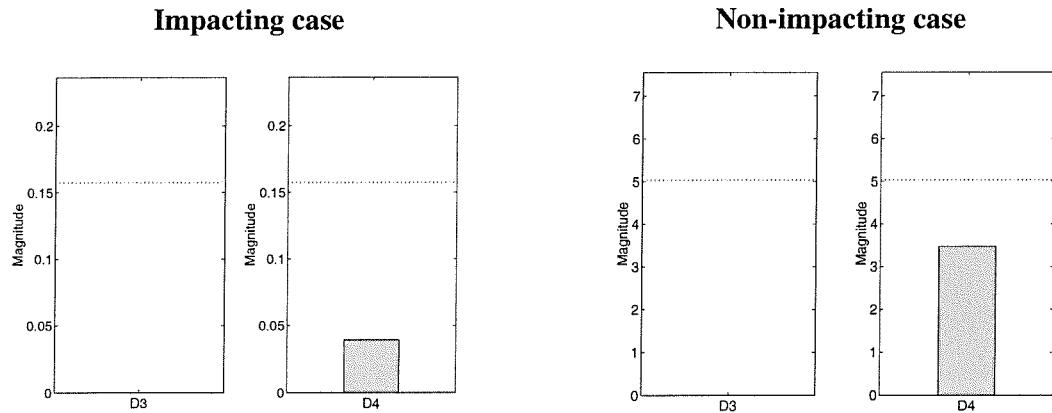


Figure 8.2.9 Detection results from HOSVD.

Unlike the detection results from the force transducer signal, the bar graphs do not indicate the existence of the impacting signal. One possible reason for this is that the construction of the higher order cumulant tensors from a single signal is liable to give inconsistent results. This inconsistency may be corrected when we have more than one signal, which is demonstrated next.

(iii) using two measurements (multichannel approach) (\mathbf{v}_1 and \mathbf{v}_2 signals from Figure 8.2.5 and Figure 8.2.7)

Following similar procedure given in Chapter 4, we create a multichannel based impacting signal detection algorithm and the result is shown in Figure 8.2.10.

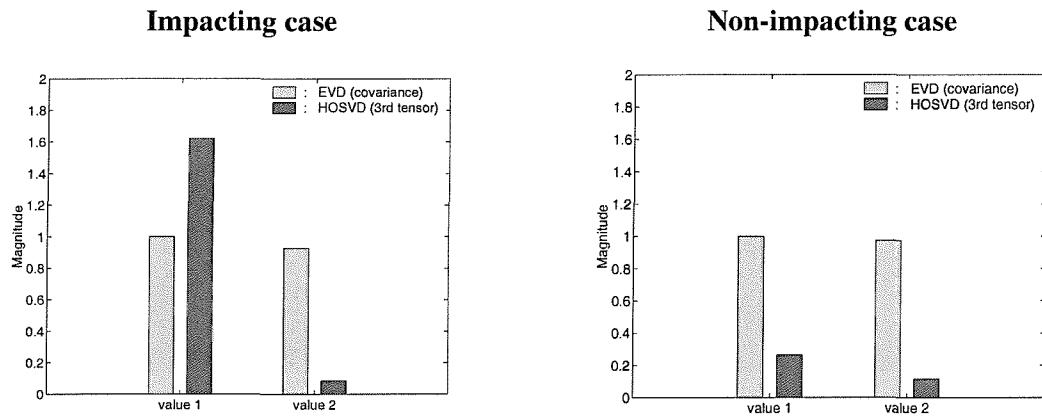


Figure 8.2.10 Detection results from HOSVD.

Using two channels, the comparison of values from eigenvalue decomposition (EVD) and higher order singular value decomposition (HOSVD) can serve as a further detection tool. As a result, we can detect the impacting signal properly through this multichannel impacting signal detection procedure.

We now move to the **restoration** of the impacting signal using various methods (single channel BS, DE, BSS, and MBD) in the next section.

8.3 Reconstruction of the input signal from measured mechanical structure's response

From the measured motion of the beam at two different points, the impacting signal (x_2) is to be reconstructed.

8.3.1 Single observation case

For the ordinary blind deconvolution process, we assume that either the force transducer signal (v_1) or the accelerometer signal (v_2) is available. Two methods (Wiener optimisation and DE method) are applied to recover the impacting signal.



Figure 8.3.1 Single Input Single Output (SISO) inverse filtering

Reconstruction of impacting signal from force transducer signal (v_1)

As a first step, the length of inverse filter (FIR) determination is required (using the methods of Chapter 5).

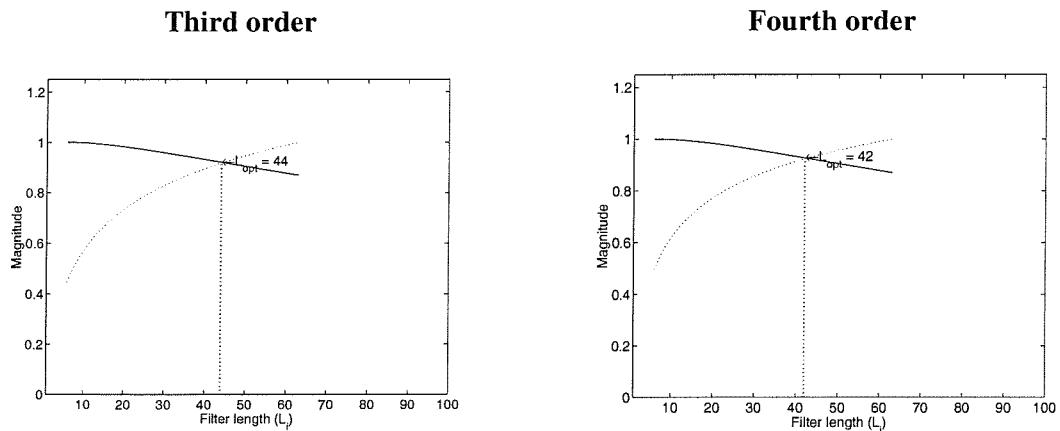


Figure 8.3.2 Result of FIR inverse filter determination

Wiener optimisation approach

The initial FIR inverse filter is selected as an impulse (initial) type (see Chapter 5) with 44 coefficients (for third order case as shown in Figure 8.3.2) as;

$$\mathbf{f}_{b_ini} = \underbrace{[1 \ 0 \ 0 \ \dots \ 0]}_{44}, \quad \mathbf{f}_a = 1$$

The shapes of the restored signals using third and fourth order Wiener optimisation are shown below.

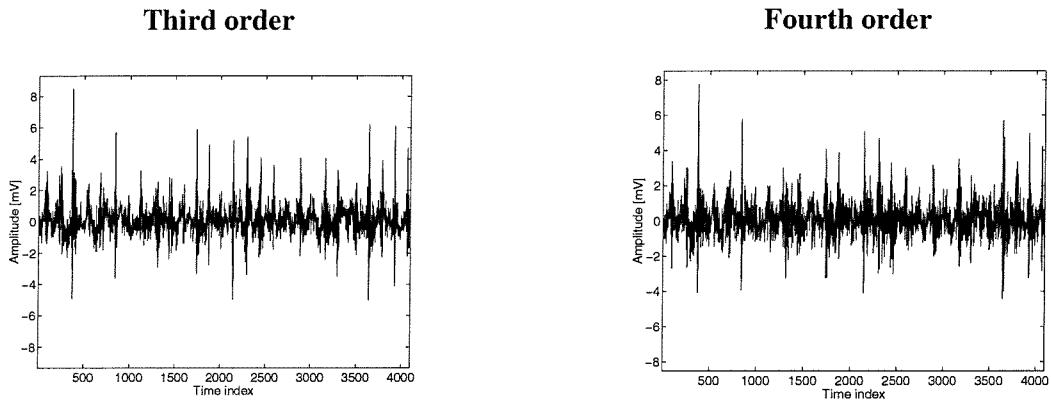


Figure 8.3.3 Restored impacting signals from force transducer signal (Wiener optimisation)

From the above figure, we can see the impulsive nature of signal. The statistical performances of each restored signal are summarised in Table 8.3.1.

DE method (see Chapter 6)

The initial FIR inverse filter is composed of 100 candidates (population) of 44 (for third order case) random numbers (having uniform distribution) as;

$$\mathbf{f}_{b_ini} = \begin{bmatrix} 0.2 & -1.2 & \dots & 0.5 & -0.1 \\ \vdots & \vdots & & & \vdots \\ & & \vdots & & \\ 1.4 & \dots & -1.7 & \dots & 0.5 \end{bmatrix}^{44}, \quad \mathbf{f}_a = 1$$

The shapes of the restored signals from the third and fourth order DE method are shown below.

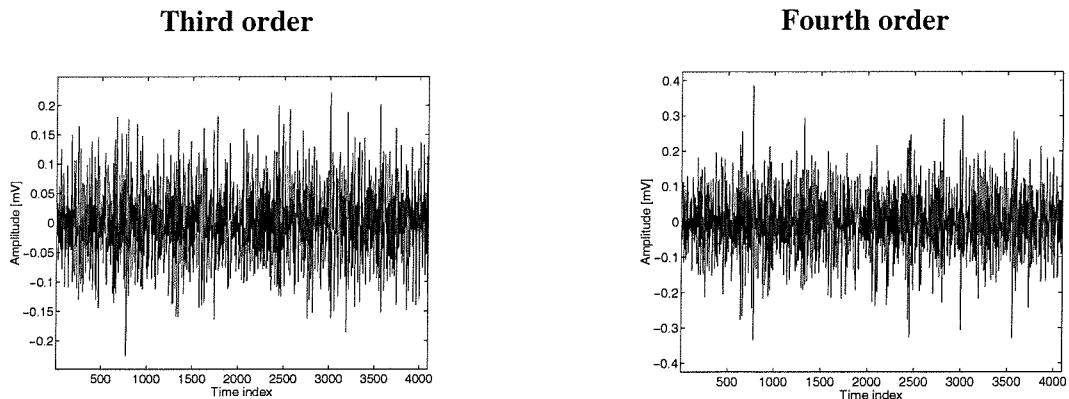


Figure 8.3.4 Restored impacting signals from force transducer signal (DE method)

Unlike the results of Wiener optimisation, the output of the inverse filter calculated from DE does not reveal the impulse components. The observed force transducer signal (\mathbf{v}_1) is severely affected by the Gaussian excitation signal and the optimisation iteration of DE is probably terminated (500 iterations) before the

global maximum is found (a considerably longer iteration allowance may yield better results but make this less attractive than Wiener optimisation).

The statistical performances of each restored signal from DE method are summarised in Table 8.3.1 and compared with Wiener optimisation.

Table 8.3.1 Results of impacting signal reconstruction from the force transducer signal (v_1)

Methods Statistical parameters	Wiener approach		DE	
	Third order	Fourth order	Third order	Fourth order
Skewness	0.894	0.554	0.18	-0.035
Kurtosis	9.303	7.477	3.391	3.719
SSD	7206.422	7190.730	7831.570	9019.917
CPU time (sec)	8.56	29.38	5181.78	4782.96

Referring to the results of Table 8.3.1 and considering the computational time of Wiener optimisation and DE, it is clear that Wiener approach is a more efficient approach to impacting signal reconstruction.

For another single channel blind deconvolution, we will next use the accelerometer signal (v_2) to restore the impacting signal through both the Wiener optimisation and DE method.

Reconstruction of impacting signal from accelerometer signal (v_2)

The length of inverse filter (FIR) determination is carried out and the result is shown below.

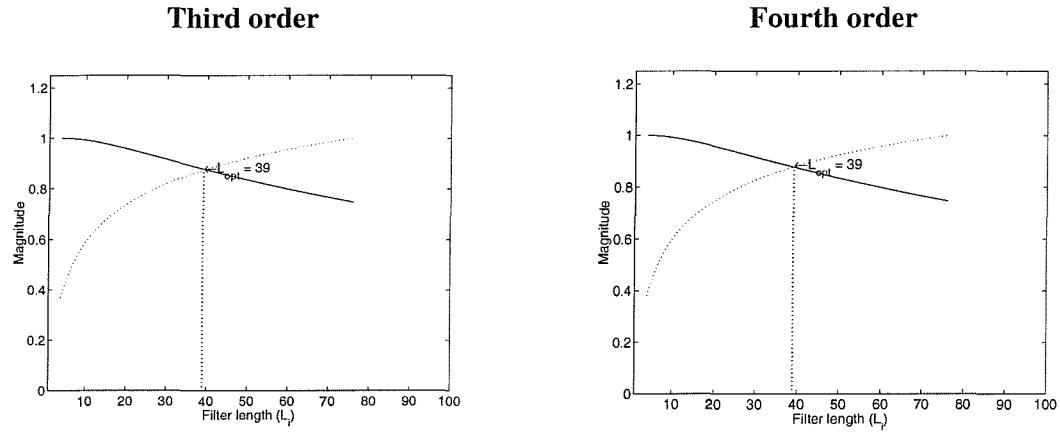


Figure 8.3.5 Result of FIR inverse filter determination

which suggests an FIR inverse filter length of 39 coefficients. Thus, from the result of optimal FIR inverse filter length determination given in Figure 8.3.5, the impacting signal restoration procedures are carried out.

Wiener optimisation approach

The initial FIR inverse filter is selected as an impulse (initial) type with 39 coefficients for both the third and fourth order case as;

$$\mathbf{f}_{b_ini} = \underbrace{[1 \ 0 \ 0 \ \dots \ 0]}_{39}, \quad \mathbf{f}_a = 1$$

The shapes of restored signals from the third and fourth order Wiener optimisation are shown below.

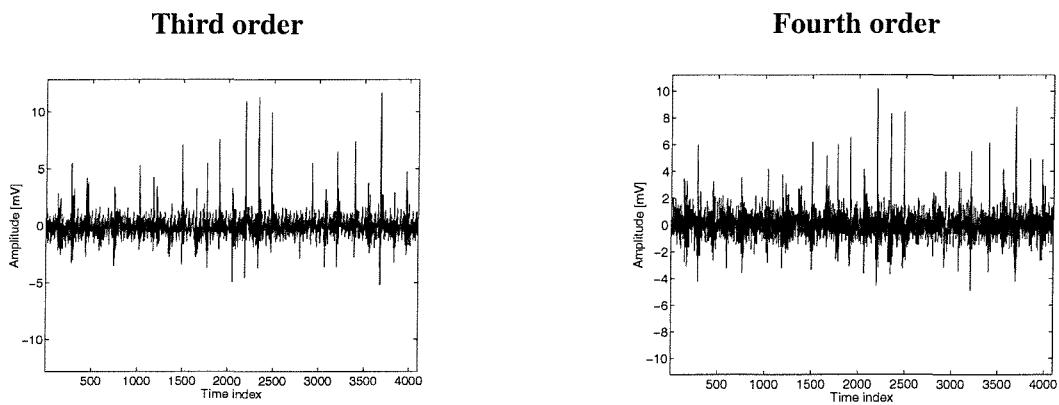


Figure 8.3.6 Restored impacting signals from accelerometer signal (Wiener optimisation)

As the measurement position of the accelerometer (channel 2, signal v_2) is quite close to the impacting point of the beam, the impulsive nature of each signal in Figure 8.3.6 shows a clear contrast to the restored signals from the force transducer (channel 1, signal v_1). Note that the statistical properties of both v_1 and v_2 are similar (refer to Table 8.2.2)

DE method

The initial FIR inverse filter is composed with 100 candidates (population) of 39 random numbers (having uniform distribution) as;

$$\mathbf{f}_{b_ini} = \begin{bmatrix} 0.2 & -1.2 & \dots & 0.5 & -0.1 \\ \vdots & \vdots & & & \vdots \\ & & \vdots & & \\ 1.4 & \dots & -1.7 & \dots & 0.5 \end{bmatrix}^{39}, \quad \mathbf{f}_a = 1$$

The restored signals from the third and fourth order DE method are shown below.

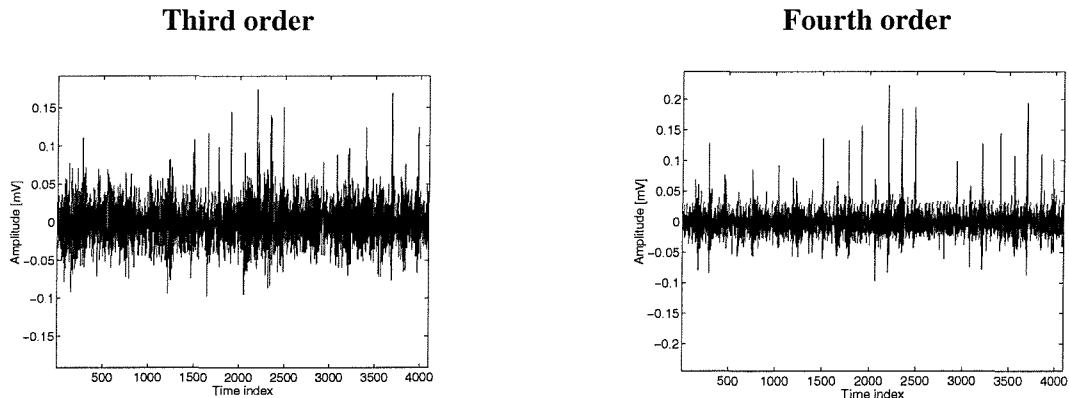


Figure 8.3.7 Restored impacting signals from accelerometer signal (DE method)

By using the accelerometer signal, the result of impacting signal restoration by DE shows a strong contrast to the result of the force transducer signal (compared to the signal of Figure 8.3.4). Also this is now competitive with the result of the Wiener optimisation. These results are summarised in Table 8.3.2.

Table 8.3.2 Results of impacting signal reconstruction from the accelerometer signal (v_2)

Methods Statistical parameters	Wiener approach		DE	
	Third order	Fourth order	Third order	Fourth order
Skewness	2.283	1.066	0.482	1.640
Kurtosis	25.355	13.185*	4.930	17.275**
SSD	4711.445	5459.333	6574.665	4949.491
CPU time (sec)	5.98	22.30	4310.77	12555.580

Table 8.3.2 illustrates that the DE method may be an acceptable approach to impacting signal reconstruction. We can see an improvement in the result of the fourth order DE (marked by '**') over those of Wiener approach (marked by '*'). However, it should be noted that the computational time of DE is more than 500 times greater than the Wiener approach.

8.3.2 Multiple observation case

For multi-channel source separation and blind deconvolution process, we assume that both the force transducer signal (v_1) and the accelerometer signal (v_2) are available.

We utilise the multichannel blind signal separation (BSS) and the multichannel blind deconvolution (MBD) developed in Chapter 7. The results of both methods for impacting signal reconstruction are given below.

Signal separation by BSS

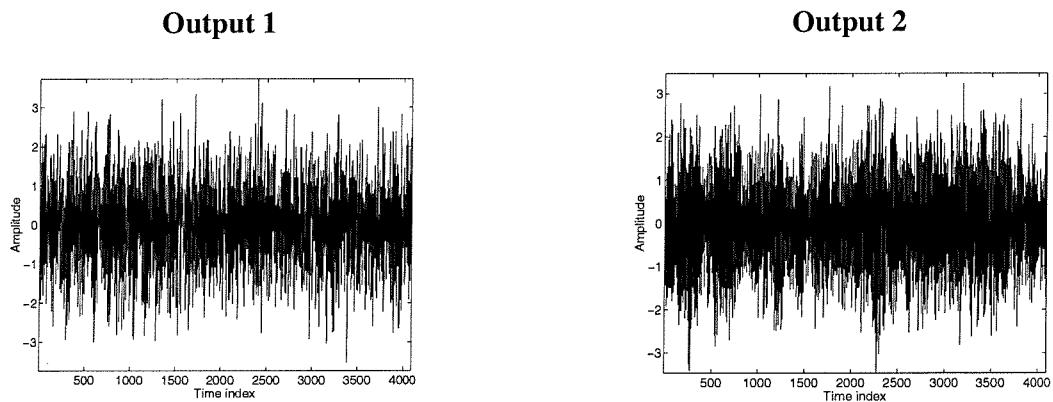


Figure 8.3.8 Separated signals from blind source separation method

Signal reconstruction from MBD

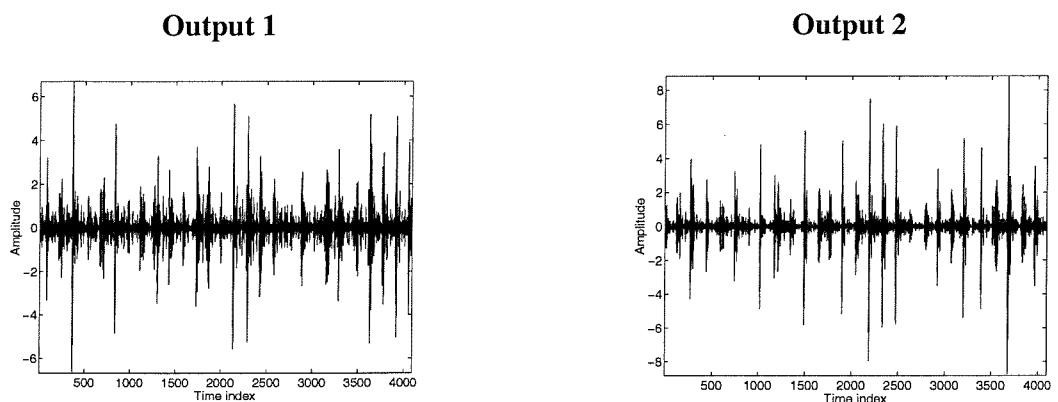


Figure 8.3.9 Separated signals from multichannel blind deconvolution method

The statistical properties are compared in following table.

Table 8.3.3 Comparison of multichannel blind source separation and blind deconvolution for impacting signal reconstruction

Methods Statistical parameters	BSS		MBD	
	y_1	y_2	y_1	y_2
Skewness	-0.022	-0.049	0.028	0.006
Kurtosis	2.909	2.830	11.208	19.874
SSD	-	-	6865.046	6139.029

As can be seen in Figure 8.3.8, Figure 8.3.9 and Table 8.3.3, it is noticeable that the BSS cannot recover this impacting signal, whereas the MBD can.

A necessary condition for BSS is that the source signals should be independent and from the arguments of Section 8.2.2, this is not satisfied in this experiment (refer to the simplification of the system from Figure 8.2.2 to Figure 8.2.3 in Section 8.2.2). However, in spite of this simplification, the MBD process can successfully recover the impacting signal and shows impulsive components in both outputs.

8.4 Summary and conclusion

This chapter has demonstrated the practical application of the blind deconvolution technique through a simple beam excitation experiment. The signals from single channel measurements or two channel measurements have been used to identify a fault signal that can hardly be recognised by a simple examination (direct observation or for example, crest factor or kurtosis exploration of the signal from the pick-up sensor).

Using any single BD or DE method, the impacting signal has been successfully identified in a statistical sense. Among these, the efficiency of BD from the point of view of computational time is preferable to DE.

For a more general approach, two different multichannel signal reconstruction procedures are carried out namely; (i) the Blind Source Separation (BSS) and (ii) the Multichannel Blind Deconvolution (MBD).

A comparison for impacting signal reconstruction is summarised below.

Table 8.4.1 Comparison of the best impacting signal reconstruction results for single channel BD, DE and multichannel blind deconvolution

	Single channel BD	Single channel DE	Multichannel BD
Skewness	2.283	1.640	0.006
Kurtosis	25.355	17.275	19.874
SSD	4711.445	4949.491	6139.029
Remarks	From the third order deconvolution method using the accelerometer signal	From the fourth order DE method using the accelerometer signal	-

From this table, we can see that the single channel BD process is an effective method for restoring the impacting signal.

The clear impulsive nature of restored signal can help identify the cause of the impacts within the mechanical systems. This encourages us to propose single channel BD as an effective tool to reveal impacting signals in condition monitoring.

Due to the computational inefficiency, DE can only be used as a reference for further comparative works.

Future work could include the use of cross signal manipulation using the reconstructed multichannel signals.

Chapter 9

Conclusions

9.1 General conclusion

Based on the study of this thesis, conclusions are presented relating to seven stages of the work.

Stage 1 (Chapter 2) The thesis begins with the explorations of the characteristics of higher order cumulants of signals and their properties through convolution. Based upon theory and simulation, this stage provides evidence of the validity of the application of higher order statistics to the source signal reconstruction problem.

Selecting the value of the normalised cumulant which incorporates the second order and higher (third or fourth) order cumulant, the inequality condition of this value is established. This condition now provides the key motivation in the blind deconvolution problems and hence is *employed for the blind reconstruction* of source signal.

Stage 2 (Chapter 3) This stage has been devoted to the fundamental consideration of the utilisation of the Higher Order Statistics in order to reconstruct an unknown impacting signal from only a measured signal. Starting from the basic Wiener optimisation approach for FIR systems, the blind deconvolution procedure has been justified utilising the objective function and its correspondence to ‘partial order’. In restoring the impacting signal, two different objective functions (constrained and normalised higher order cumulant) have been justified from which FIR (non-recursive, MA inverse system) filter coefficients are calculated. For completeness, the inverse system has been selected as having a recursive nature (i.e., AR or ARMA system) with the expectation of improved performance over that of the non-recursive system. These three different inverse systems are compared with the shapes of each objective function with respect to the filter coefficients. The result of signal restoration from each system has been also compared. Even though all the systems possess more than one maximum, the MA inverse system turned out to be more robust than the recursive systems.

Stage 3 (Chapter 4) The aim of this contribution is to put HOSVD/tensors into a practical context. To do this, applications of higher order statistics through the construction of higher order tensors and their singular value decomposition is introduced. From this, we have devised the utilisation of HOSVD for detection, classification and reconstructability of non-Gaussian signals. From the constructed higher order tensors of measured signal, the higher order singular values are estimated. The essence of non-Gaussian signal detection is based on the comparison of the second order singular values and higher order singular values. From the simulations with various systems, we conclude that the detection, classification and reconstructability assessment using HOSVD can be a useful tool for blind processing of impacting processes.

Stage 4 (Chapter 5) This stage provides guides to open questions in blind deconvolution namely (i) initial filter coefficient vector selection and (ii) choice of length of the inverse filter. From this, firstly, it is demonstrated that it becomes natural to take the *initial inverse filter as impulsive* in form. Secondly, criteria on the selection of optimal inverse filter length have been proposed resulting a *predictive determinator* which utilises the objective function (skewness or kurtosis), α -coefficient and/or entropy of the observed signal that can avoid very time consuming computation.

Stage 5 (Chapter 6) We investigated the performance of just one of numerous global optimisation methods for the blind restoration of non-Gaussian impacting signals. This optimisation is the Differential Evolution (DE) method, which utilises an evolutionary process similar to the Genetic Algorithms (GA). The performance of signal reconstruction and computational efficiency are considered and compared to the Wiener optimisation method discussed in Chapter 3. Since the Wiener approach is believed to yield only one of the local maxima, the restoration of an impacting signal using a global optimisation method has been carried out. The restoration of impacting signal from DE has resulted in a successful method which is comparable to the Wiener optimisation method. However, given the equivalence of the results, DE scores relatively poor method because of its computational overhead.

Stage 6 (Chapter 7) In this stage, we have addressed the problem of the blind source separation and deconvolution of sources for convolutive mixtures for the reconstruction of an impacting signal. The theoretical derivations and numerical simulations have been carried out and an approach suggested for the multichannel blind non-Gaussian (impacting signal) reconstruction from multipoint measurement of a mechanical structure.

Stage 7 (Chapter 8) In this stage, we used an experimental impacting cantilever beam to validate the blind deconvolution process from computer simulations. Firstly, we justified a multichannel model of this experiment and established an approach to the practical application of BD process by simplifying the experimental model. Secondly, we tried to reconstruct the impacting signal through single or multichannel BD process. The experimental results for restoring the faulty impacting signal gave acceptable information to identify the cause of mechanical system.

9.2 Further research

In this final section, aspects of further research are proposed which relate to the results of Chapter 6, Chapter 7 and Chapter 8.

From the results of Chapter 6, possible future research could include;

- (1) Study of stability of the IIR (AR and ARMA) inverse filters in DE process
- (2) The search for computational efficiency when the filter length increases

From the results of Chapter 7 and Chapter 8, enhancement on the restored signal using multichannel restored signals through;

- (1) Signal vector splitting based on the SVD
- (2) A cumulant-based adaptive enhancer

On a more general note relating to applications, the results of this thesis are sufficiently encouraging to go forward to a more unified approach to condition monitoring. For example, for detection and diagnosis of a faulty impacting component of a mechanical system where the signal cannot be measured directly, one must identify the signal and characterise the internal state of the system. A diagrammatical illustration of this monitoring system is depicted in Figure 9.2.1.

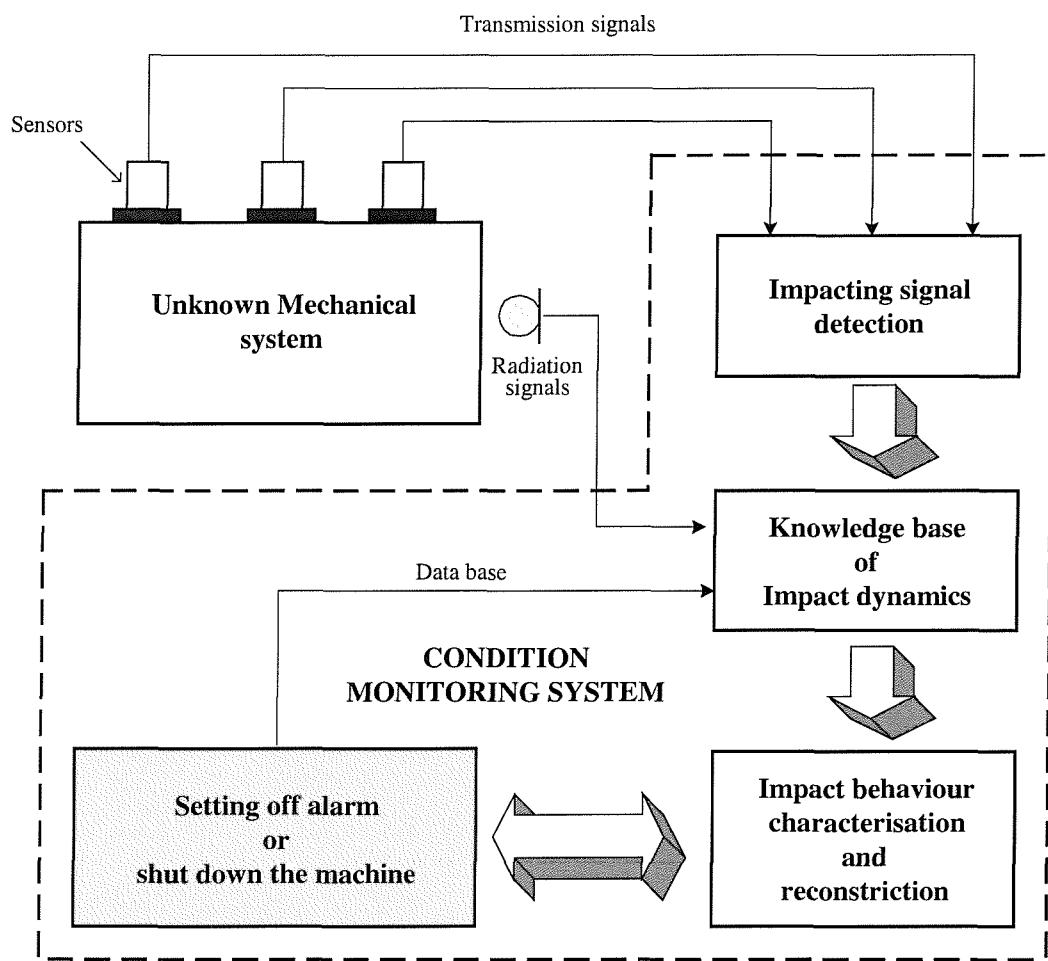


Figure 9.2.1 Condition monitoring system based on this thesis

Using the results of this thesis, we could have applied the impacting signal detection tool based on the HOSVD and signal identification using BD process. However, a pattern recognition of impacting signal may be required in order to make a precise diagnosis. To do this, a **frequency based higher order signal processing** (bi-spectrum and tri-spectrum) becomes an essential element as well as the time based higher order statistics.

References

- [1] Abed-Meriam, K., Qui, W. and Hua, Y., "Blind System Identification", *Proceedings of the IEEE*, Vol. 85, No. 8, pp 1310-1322, Aug. 1997.
- [2] Akaike, H., "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, Vol. **19**, No. 6, pp. 716-723, Dec. 1974.
- [3] Bar-Ness, Y. and Rokach, J., "Cross coupled boot-strapped interference canceller," Proc. International Conference on Antennas and Propagation, USA, June 1981, pp 292-295, 1981.
- [4] Barrett, J. F., "The use of characteristic functionals and cumulant-generating functionals to discuss the effect of noise in linear systems," *Journal of Sound and Vibration* Vol **I**, No.3, pp229-238, 1964.
- [5] Bendat, J. S. and Piersol, A. G., *Random Data - Analysis and Measurement Procedures*, John Wiley & Sons, Inc., 1986.
- [6] Bradaric, I. and Petropulu, A. P., "Design of low rank estimators for higher order statistics based on the second order statistics", *Proceedings, IEEE Signal Processing Workshop on Higher-Order Statistics*, June 14-16, Caesarea, Israel, 1999.
- [7] Braun, S. and Hammond, J. K., *Mechanical Signature Analysis -Theory and Applications*- Chapter 5, Academic Press, 1986.
- [8] Brillinger, D. R, "An Introduction to Polyspectra", *Annual Mathematics and Statistics*, Vol. **36**, pp 1351-1374, 1965.
- [9] Cadzow, J. A., "Blind Deconvolution via Cumulant Extrema", *IEEE Signal Processing Magazine*, pp 24-42, May. 1996.
- [10] Cardoso, J. F. and Comon, P., "Tensor-based independent component analysis, in Signal Processing V: Theories and Applications", *Proceedings EUSIPCO-90*, L. Torres, ed., Elsevier, Amsterdam, pp. 673-676, 1990.

[11] Cardoso, J. F., "Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem", *Proc. ICASSP-90*, Albuquerque, NM, pp. 2655-2658, 1990.

[12] Cardoso, J. F., "Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors", *Proceedings ICASSP-91*, Toronto, Canada, pp. 3109-3112, 1991.

[13] Cardoso, J. F., "Fourth-order cumulant structure forcing. Application to blind array processing", *Proceedings SSAP-92*, pp. 136-139, 1992.

[14] Cardoso, J. F., "Multidimensional Independent Component Analysis", *Proceedings, ICASSP-98*, Seattle, USA, 1998 (a).

[15] Cardoso, J. F., "Blind signal separation: Statistical principles", *Proceedings of the IEEE*, Vol. **9**, No. 10, pp 2009-2025, Oct. 1998 (b).

[16] Cardoso, J. F., "Blind Identification of independent signals," *Workshop on Higher-Order Spectral Analysis*, Vail, CO, June 1989.

[17] Chan, Y.T. and Wood, J. C., "A New Order Determination Technique for ARMA Processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. **32**, No. 3, pp 517-521, Jun. 1984.

[18] Chow, J. C., "On Estimating the Orders of an ARMA Process with Uncertain Observations", *IEEE Transactions on Automatic Control*, Vol. **17**, pp 707-709, Oct. 1972.

[19] Claerbout, J. F. and Muir, F., "Robust Modeling with Erratic Data", *Geophysics*, Vol. **38**, pp 826-847, 1973.

[20] Claerbout, J. F., "Parsimonious deconvolution", Stand. Exploration Project, Vol. 13, pp 1-9, 1977.

[21] Comon, P., "Independent component analysis, a new concept?", *Signal Processing, Special Issue on Higher Order Statistics*, Vol. **36**, pp. 287-314, 1994.

[22] Comon, P., "Separation of sources using higher-order cumulants," in: *SPIE Vol. 1152, Advanced Algorithms and Architectures for Signal Processing*, Vol. **IV**, San Diego, CA, 8-10 August 1989.

[23] Delfosse, N. and Loubaton, P., "Adaptive blind separation of independent sources: A deflation approach", *Signal Processing*, Vol. **45**, No. 1, pp 59-83, July 1995.

[24] Donoho, D., *On minimum entropy deconvolution*, In *Applied Time series Analysis II* ed., D. Findley, Academic Press, New York, pp 556-608, 1981.

[25] Duntzman, G. H., *Principal Component Analysis*, SAGE Publications Inc., 1989.

[26] Dyer, D. and Stewart, R. M., "Detection of rolling element bearing damages by statistical vibration analysis", *Trans. ASME, J. Mech. Design*, Vol. **100**, No. 2, pp 229-235, 1978.

[27] Foschini, G. J., "Equalising without altering or detecting data", *AT&T Technical Journal*, Vol. **64**, pp 1885-1911, 1985.

[28] Giannakis, G. B. and Mendel, J. M., "Identification of Nonminimum Phase Systems Using Higher Order Statistics", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. **37**, No. 3, pp 360-377, Mar. 1989.

[29] Giannakis, G. B. and Mendel, J. M., "Cumulant-Based Order Determination of Non-Gaussian ARMA Models", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. **38**, No. 8, pp 1411-1423, Aug. 1990.

[30] Giannakis, G. B., "Signal processing via higher order statistics", Ph.D. dissertation, Univ. Southern California, Los Angeles, CA, 1986.

[31] Godard, D. N., "Self-recovering equalisation and carrier tracking in two-dimensional data communication systems", *IEEE Transactions on Communications*, Vol. COM-**28**, pp 1867-1875, 1980.

[32] Goldberg, D. E., *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, 1989.

[33] Gray, W., "Variable norm deconvolution", PhD. Dissertation, Stanford University, Stanford, CA, 1979.

[34] Hammond, J. K. and Clarkson, P. M., "Lecture notes on Signal Processing", ISVR, University of Southampton, UK, 1989.

[35] Horn, R. A. and Johnson, C. R., *Topics in Matrix Analysis*, Cambridge University Press, N.Y., 1991.

[36] Ingber, L. and Rosen, B., "Genetic Algorithms and Very Fast Simulated Reannealing: A Comparison," *Journal of Mathematical and Computer Modelling* **16** (11), pp 87-110, 1992.

[37] Juang, B., Perdue, R. J. Jr., and Thomson, D.L., "Deployable automatic speech recognition systems: Advances and challenges," *AT&T Technical Journal*, pp 45-55, Mar./Apr. 1995.

[38] Jutten, C. and Herault, J., "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing* Vol. **24**, No. 1, pp 1-10, 1991, July 1991.

[39] Kendal, M. G. and Stuart, A., *The advanced theory of statistics*, Volume 1 Distribution theory, Ch. 2, Charles Griffin & Company Limited, London, 1958.

[40] Kim, Donghae, "Identification of Nonstationary Parametric Models Using Higher-Order Statistics", PhD Thesis, Univ. Southampton, ISVR, UK, 1998.

[41] Kirkeby, O., Nelson, P A, H. Hamada and F. Orduna-Bustamante (1996) Fast Deconvolution of Multi-Channel Systems using Regularisation. *ISVR Technical Memorandum, No. 255* April 1996.

[42] Kundur, D. and Hatzinakos, D., "Blind image deconvolution", *IEEE Signal processing Mag.*, Vol. 13, pp 43-64, May 1996.

[43] Lacoume, J. L. and Ruiz, P., "Source identificatiion: A solution based on cumulants" *IEEE Acoust. Speech Signal Process. Workshop V*, Minneapolis, USA, August 1988.

[44] Lamm P K (1993) Inverse problems and Ill-Posedness', in 'Inverse Problem in Engineering : Theory and Practice. *The American Society of Mechanical Engineers*, 1993.

[45] Lathauwer, De L., De Moor, B. and Vandewalle, J., "Blind source separation by higher-order singular value decomposition, in Signal Processing VII: Theories and Applications", Proceedings EQSIPCO-94, Edinburgh, U.K., pp. 175-178, 1994.

[46] Lathauwer, De L., Comon, P., De Moor, B. and Vandewalle, J., "Higher-order power method – application in independent component analysis", Proceedings NOLTA'95, Las Vegas, UT, 1995, pp,91-96, 1995.

[47] Lathauwer, De L., De Moor, B. and Vandewalle, J., "Blind Source separation by simultaneous third-order tensor diagonalisation, in Signal Processing VIII: Theories and Applications", Proceedings EUSIPCO-96, Trieste, Italy, 1996 (a).

[48] Lathauwer, De L., De Moor, B. and Vandewalle, J., "Independent Component Analysis based on higher-order statistics only", Proc. SSAP-96, Corfu, Greece, 1996 (b).

[39] Lathauwer, De L., "Signal Processing based on Multilinear Algebra", PhD. Dissertation, Dept. of Electrical Engineering (ESAT), Katholieke Universiteit Leuven, Belgium, 1997.

[50] Lathauwer, De L., De Moor, B., "Orthogonal Super-Symmetric Tensor Decompositions and Independent Component Analysis", *Journal of Chemometrics Special Issue on Multi-Way Analysis*, May 1999 (a).

[51] Lathauwer, De L., Comon, P., De Moor, B. and Vandewalle, J., "ICA algorithms for 3 sources and 2 sensors", *Proceedings, IEEE Signal Processing Workshop on Higher-Order Statistics*, pp 116-120, Caesarea, Israel, June 14-16, 1999 (b).

[52] McCullagh, P., "Tensor Methods in Statistics," Chapman and Hall, London, 1987.

[53] Mendel, J. M., "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory : Theoretical Results and Some Applications," *Proceedings of IEEE*, Vol **79**, No. 3, pp 278-305, March 1991.

[54] Miller, J. H. and Thomas, J. B., "Detector for Discrete-Time Signals in non-Gaussian Noise", *IEEE Trans. On Information Theory*, IT-**18**, 2, pp 241-250, March 1972.

[55] Moreau, E. and Macchi, O., "New self-adaptive algorithms for source separation based on contrast functions", *IEEE Signal Processing Workshop on Higher-Order Statistics*, South Lac Tahoe, CA, USA, pp 215-219, June 1993.

[56] Nandi, A.K., Mampel, D. and Roscher, B., "Blind Deconvolution of Ultrasonic Signals in Nondestructive Testing Applications", *IEEE Transactions on Signal processing*, Vol **45**, No. 5, May, 1997.

[57] Nelder, J. A. and Mead, R., "A Simplex method for function optimisation", *Computer J.* **7** pp 308-313, 1965.

[58] Nikias, C. L. and Mendel, J. M., "Signal Processing with Higher-Order Spectra", *IEEE Signal Processing Magazine*, Vol. **10**, No. 3, pp 10-37, 1993.

[59] Nikias, C. L. and Petropulu, A. P., *Higher-Order Spectra Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[60] Nikias, C. L. and Raghubeer, M. R., "Bispectrum estimation : A digital signal processing frame work," *Proceedings of IEEE* Vol. **75**, pp 869-891, July 1987.

[61] Ooe, M. and Ulyrych, T. J., "Minimum entropy deconvolution with an exponential transformation", Annual Meeting of European Association of Exploration Geophysicists, Dublin, Scotland, 1978.

[62] Oppenheim, A. V., Schafer, R.W., *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, N.J., 1989.

[63] Otte, D., Fyfe, K., Sas, P. and Leuridan, J., "Use of Principal Component Analysis for dominant noise source Identification", *ImechE*, C21/88, pp 123-132, 1988.

[64] Pachaud, C., Salvatet, R. and Fray, C., "Crest factor and Kurtosis contributions to identify defects inducing periodical impulsive forces", *Mechanical Systems and Signal Processing (MSSP)*, vol.11, No. 6, pp 903-916, Jun. 1997

[65] Papoulis, A., *Probability, Random Variables, and Stochastic Process*, Ch. 15, McGraw-Hill International Editions 1991.

[66] Peacock, K. L. and Treitel, S., "Predictive Deconvolution – Theory and Practice", *Geophysics*, Vol. 34, pp. 155-169, June 1969.

[67] Pozidis, H. and Petropulu, A. P., "Cross-Spectrum Based Blind Channel Identification", *IEEE Transactions on Signal processing*, Vol. 45, No. 12, pp 2977-2992, Dec. 1997.

[68] Proakis, J. G., "Digital Communications, 3rd ed.", McGraw-Hill New York, 1995.

[69] Rice, S. O., "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, Vol 23, pp 282-332, 1944.

[70] Robinson, E., "Predictive Decomposition of Time Series with Application to Seismic Exploration", *Geophysics*, Vol. 32, no. 3, pp. 418-484, June 1967.

[71] Robinson, E. A. and Treitel, S., "Digital Signal Processing in Geophysics", in *Application of Digital Signal Processing*, A.V. Oppenheim, ed., Prentice Hall, Englewood Cliffs, N.J., 1978.

[72] Robinson, E. A. and Treitel, S., *Geophysical Signal Analysis*, Prentice Hall, Englewood Cliffs, N.J., 1980.

[73] Shin K., "Characterisation and identification of chaotic dynamical systems", ISVR 1996 PhD Thesis, 1996.

[74] Stockham, T., Cannon, T. and Ingerbretsen, R., "Blind deconvolution through digital signal processing", *Proceedings of the IEEE*, Vol. 63, pp 678-692, 1975.

[75] Storn, R. and Price, K., "Differential Evolution – A Simple and Efficient Heuristic for Global Optimisation over Continuos Spaces," *Journal of Global Optimization* 11 pp 341-359, 1997.

[76] Takens, F., "Detecting strange attractors in turbulence", *Lecture notes in mathematics*, 898, pp 365-381, 1981.

[77] Thi, Hoang-Lan Nguyen and Jutten, C., "Blind source separation for convolutive mixtures," *Signal Processing* Vol. **45**, pp 209-229, 1995.

[78] Tikhonov A N and Arsenin V Y (1977) Solutions of ill-posed problems. *John Wiley & Sons.*

[79] Tsatsanis, M. K. and Giannakis, G. B., "Subspace Methods for Blind Estimation of Time-Varying FIR Channels", *IEEE Transactions on Signal processing*, Vol. **45**, No. 12, pp 3084-3093, Dec. 1997.

[80] Wiggins, R. A., "Minimum entropy deconvolution", Thirty-ninth Meeting of the European Association of the Exploration Geophysicists, 1977.

[81] Wiggins, R. A., *Minimum entropy deconvolution*, Elsevier Scientific publishing Geoexploration, 16, pp 21-35, 1978.

[82] Woods, L. C. and Treitel, S., "Seismic Signal Processing", *Proc. IEEE*, Vol. **63**, pp. 649-661, April, 1975.

[83] Yellin, D. and Weinstein, E., "Criteria for multichannel signal processing", *IEEE Transactions on Signal processing*, Vol. **42**, No. 8, pp 2158-2168, Aug. 1994.

[84] Yoon, S H and Nelson, P A (1995) Some techniques to improve stability in identifying acoustic source strength spectra. *ISVR Technical Memorandum*, **No. 779** Nov. 1995.

[85] Yoon S H and Nelson P A (1997) On the condition number of the matrix to be inverted in an acoustic inverse problem. *ISVR Technical Memorandum*, **No. 817** April 1997.

[86] Zhu, Jie, Cao, Xi-Ren and Ding, Zhi, "An Algebraic Principle for Blind Separation of White non-Gaussian Sources", *Signal processing*, Vol. **76**, pp 105-115, 1999.

Appendix A

Central limit theorem and partial order

This Appendix summarises results relating to the effect of filters on random signals and Donoho's concept of partial order. It includes a brief explanation of the *central limit theorem* and simulations to confirm the validity for our class of signals.

A.1 Central Limit Theorem

Let X_1, X_2, \dots, X_N be mutually independent random variables whose individual distributions are not specified. Denote m_{X_i} and σ_{X_i} as the mean and standard deviation of each X_i , $i = 1, 2, \dots, N$. Then the sum random variable

$$X = \sum_{i=1}^N X_i \quad (\text{A.1.1})$$

will have a mean m_X and standard deviation σ_X , where

$$m_X = \sum_{i=1}^N m_{X_i}, \quad \sigma_X^2 = \sum_{i=1}^N \sigma_{X_i}^2 \quad (\text{A.1.2})$$

The *central limit theorem* states that under certain conditions the sum random variable X will be normally distributed as $N \rightarrow \infty$ with the above mean and variance.

$$p_X(x) \simeq \frac{1}{\sigma_X \sqrt{2\pi}} e^{-(x-m_X)^2/2\sigma_X^2} \quad (\text{A.1.3})$$

Proof

Equation (A.1.3) can be proved as follows [Bendat, 1958].

For simplicity, we shall assume that the respective mean values are zero.

$$m_{X_i} = E[X_i] = 0, \quad \sigma_{X_i}^2 = E[X_i^2] \quad (\text{A.1.4})$$

Denote the higher (central) moments of X_i as

$$\gamma_{X_i} = E[X_i^3], \quad \tau_{X_i} = E[X_i^4] \quad (\text{A.1.5})$$

For the sum random variable $X = \sum_{i=1}^N X_i$ since the X_i are mutually independent, it

follows that

$$m_X = 0, \quad \sigma_X^2 = N\sigma_{X_i}^2, \quad \gamma_X = N\gamma_{X_i}, \quad \tau_X = N\tau_{X_i}, \quad \text{etc.} \quad (\text{A.1.6})$$

We next normalise X to a new random variable Y whose variance is unity by

$$Y = \frac{X}{\sigma_X} = \frac{\sum_{i=1}^N X_i}{\sigma_X} \quad (\text{A.1.7})$$

Using the *characteristic function* definition in Chapter 2 (section 2.2), the relationship between the probability density function of the random variable Y and its characteristic function is given by

$$\begin{aligned} \phi_Y(\omega) &= E[e^{j\omega Y}] \\ &= E\left[e^{j\omega \sum_{i=1}^N X_i / \sigma_X}\right] \\ &= \prod_{i=1}^N E\left[e^{j(\omega / \sigma_X) X_i}\right] \\ &= \prod_{i=1}^N \phi_{Y_i}(\omega / \sigma_X) \end{aligned} \quad (\text{A.1.8})$$

where $\phi_{Y_i}(\omega / \sigma_X) = E[e^{j(\omega / \sigma_X) X_i}]$.

Taking the logarithm of $\phi_Y(\omega)$ changes the product term in (A.1.8) into a sum,

$$\log \phi_Y(\omega) = \sum_{i=1}^N \log \phi_{Y_i}(\omega / \sigma_X) \quad (\text{A.1.9})$$

By expanding the exponential term into a power series, and using (A.1.5) and (A.1.6),

$$\begin{aligned}
 \phi_{Y_i}(\omega/\sigma_x) &= E\left[1 + j(\omega/\sigma_x)X_i - \frac{(\omega/\sigma_x)^2}{2!}X_i^2 - j\frac{(\omega/\sigma_x)^3}{3!}X_i^3 + \frac{(\omega/\sigma_x)^4}{4!}X_i^4 + \dots\right] \\
 &= 1 - \frac{(\omega/\sigma_x)^2}{2!}\sigma_{xi}^2 - j\frac{(\omega/\sigma_x)^3}{3!}\gamma_{xi} + \frac{(\omega/\sigma_x)^4}{4!}\tau_{xi} + \dots \\
 &= 1 - \frac{\omega^2}{2N} - j\frac{\omega^3}{6N\sqrt{N}}\left(\frac{\gamma_{xi}}{\sigma_{xi}^3}\right) + \frac{\omega^4}{24N^2}\left(\frac{\tau_{xi}}{\sigma_{xi}^4}\right) + \dots
 \end{aligned} \tag{A.1.10}$$

For large values N , and taking the logarithm of (A.1.10),

$$\log \phi_{Y_i}(\omega/\sigma_x) \approx \log\left(1 - \frac{\omega^2}{2N}\right) \approx -\frac{\omega^2}{2N} \tag{A.1.11}$$

Consequently, as N approaches infinity, substituting (A.1.11) into (A.1.9) yields

$$\log \phi_Y(\omega) = \sum_{i=1}^N \log \phi_{Y_i}(\omega/\sigma_x) \rightarrow -\frac{\omega^2}{2} \tag{A.1.12}$$

Thus, the characteristic function of Y becomes,

$$\phi_Y(\omega) = e^{-\omega^2/2} \tag{A.1.13}$$

The relationship between the probability density function $P_Y(y)$ of the random variable Y and the characteristic function is

$$\phi_Y(\omega) = E[e^{j\omega Y}] = \int_{-\infty}^{\infty} e^{j\omega y} p_Y(y) dy \tag{A.1.14}$$

which is a Fourier transform. The inverse Fourier transform of $\phi_Y(\omega)$ yields

$$\begin{aligned}
 p_Y(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_Y(\omega) e^{-j\omega y} d\omega \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\omega^2/2} e^{-j\omega y} d\omega \\
 &= \frac{1}{\sqrt{2\pi}} e^{-y^2/2}
 \end{aligned} \tag{A.1.15}$$

which is a normal probability density function with zero mean and unity variance.

Detraction from the central limit theorem

Expressing the probability density function of the total sum X as $p_X(x)$ and the individual random variable X_i 's densities as $p_{X_i}(x)$, then $p_X(x)$ is the convolution of the densities $p_{X_i}(x)$ of X_i

$$p_X(x) = p_{X_1}(x) * p_{X_2}(x) * \dots * p_{X_N}(x) \quad (\text{A.1.16})$$

where $*$ means 'convolution'. Thus, the central limit theorem can be viewed as property of convolution of positive functions. Because of this, the central limit theorem does not hold if only a small number m of the given densities are dominant. That is to say, if the densities of the other random variables are relatively narrow (leptokurtic) in the sense that, in the evaluation of the total convolution, they can be approximated by impulses. In this case, $p_X(x)$ is effectively the convolution of only m dominant densities and it need not be close to a normal curve.

Simulation for central limit theorem

Let μ_i and σ_i^2 be the mean value and variance of each random variable $x_i(k)$, $i = 1, 2, \dots, N$. Consider the sum random variable as illustrated in Figure A.1.1

$$y(k) = \sum_{i=1}^N a_i x_i(k) \quad (\text{A.1.17})$$

where a_i are arbitrary fixed constants. Now, the mean value m_y and variance σ_y^2 become

$$\begin{aligned} m_y &= E[y(k)] = E\left[\sum_{i=1}^N a_i x_i(k)\right] = \sum_{i=1}^N a_i E[x_i(k)] = \sum_{i=1}^N a_i m_i \\ \sigma_y^2 &= E[(y(k) - m_y)^2] = E\left[\sum_{i=1}^N a_i (x_i(k) - m_y)\right]^2 = \sum_{i=1}^N a_i^2 \sigma_i^2 \end{aligned} \quad (\text{A.1.18})$$

Following figure illustrates the sum of impulsive random variables (signals).

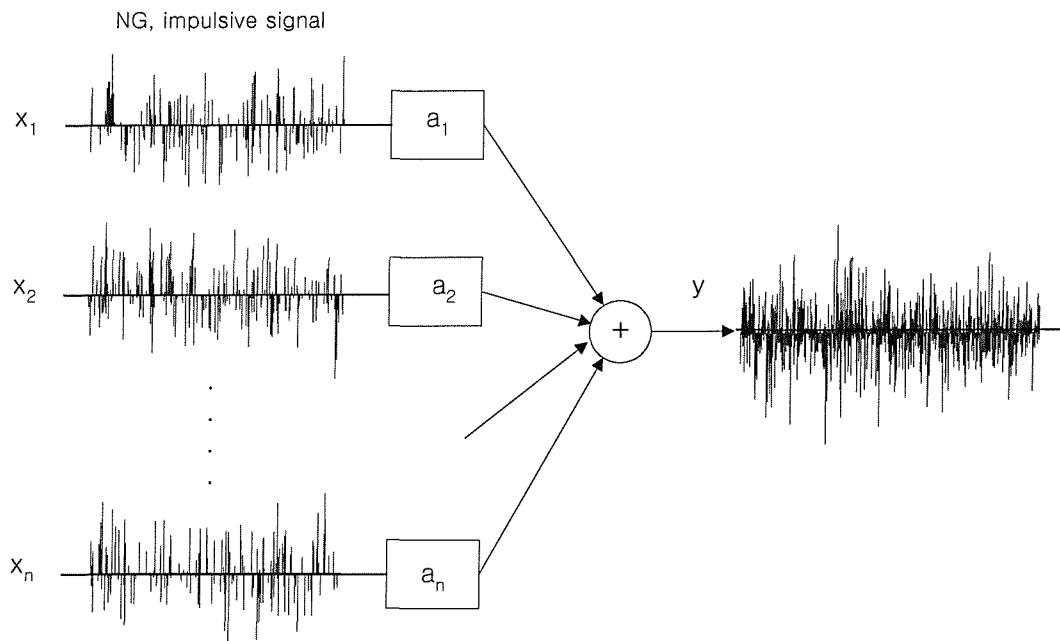


Figure A.1.1 Process of the linear combination of the non-Gaussian impulsive signals

As expressed in equation (A.1.17), the combination (from $N=1$ to 160) of random signals with 2000 sample points ($k=2000$) are simulated. The statistical changes of each output case (denoted by 'Combined signal' in the figure) are then calculated and are plotted by a solid line in Figure A.1.2. Also, the Gaussian distribution in each case (denoted by 'Gaussian signal' possessing the same mean and variance as expressed in equation (A.1.18)) are plotted as a dotted line for comparison.

Appendix A, Central limit theorem and partial order

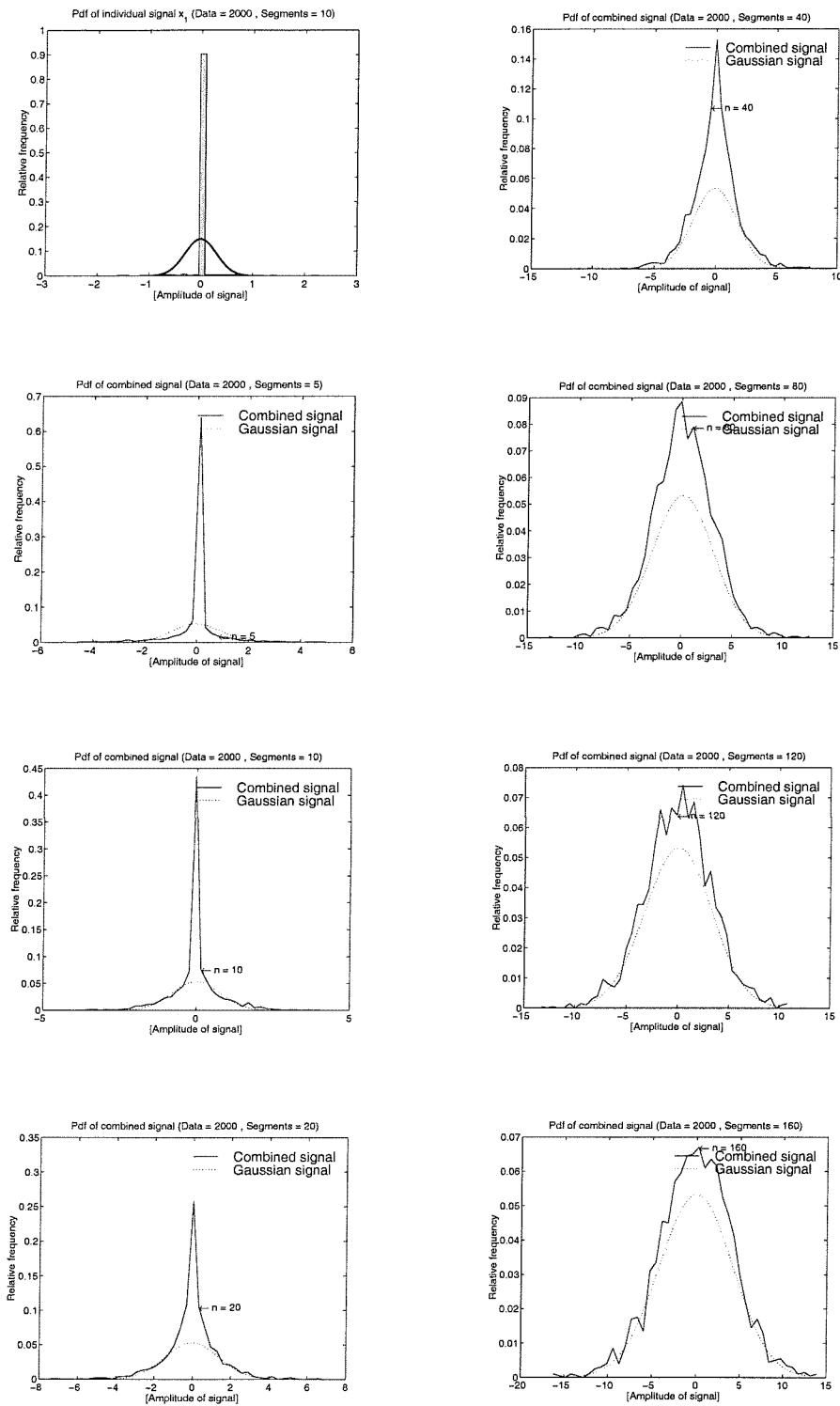


Figure A.1.2 The changes of the statistical property of the linear combination of the impulsive random signals

A.2 Partial order and BD process

Given a time series $v(n)$ which is a filtered version of an input sequence $x(n)$ as shown in Figure A.2.1

$$\mathbf{v} = \mathbf{h} * \mathbf{x} \quad (\text{A.2.1})$$

where $*$ means ‘convolution’. The deconvolution problem is to find a filter $f(n)$ which recovers from the observed series $v(n)$;

$$\hat{\mathbf{x}} = \mathbf{f} * \mathbf{v} \quad (\text{A.2.2})$$

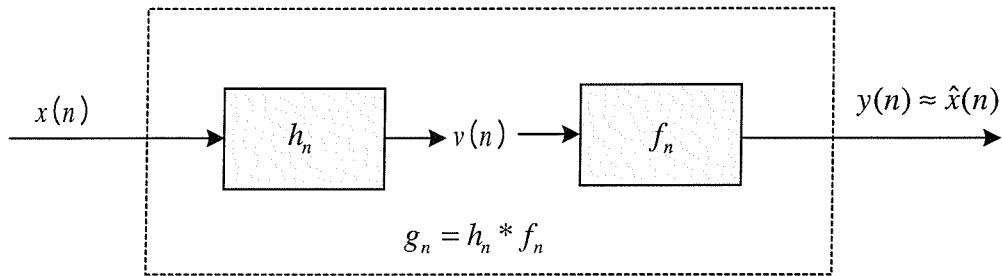


Figure A.2.1 Deconvolution model

To solve this, we introduce (in Chapter 3) the degree of Gaussianity of a random variable (i.e., how close a signal is to the Gaussian distribution in a probabilistic sense). The key idea of recovering $v(n)$ is maximise a non-Gaussian parameter (expressed in higher order terms) of the output $\hat{x}(n)$ of an inverse filter $f(n)$ with respect to the coefficients of the filter provided that $x(n)$ is non-Gaussian. If $x(n)$ is Gaussian then $v(n)$ is also Gaussian, thus the recovery procedure given in Figure A.2.1 is not applicable.

The state of Gaussianity of signals in each filtering process is now defined in terms of ‘partial order’.

A.2.1 Partial order in linear combination

Suppose two random variables X and Y (note that their lower case letters, x and y are realisations of each random variable) are related by $Y = aX + c$ (a and c are non zero constants), these random variables are regarded as equivalent from a probabilistic point of view and have the same probability distribution and denoted as

$$X \doteq Y \quad (A.2.3)$$

which means the probabilistic characteristics of any random variable is not affected by a constant scaling and addition (i.e., the shapes of the probability distributions of X and Y are identical).

On the other hand, for appropriate constants a_i , ($i=1,2,\dots$ with $\sum a_i^2 < \infty$), we write

$$Y \doteq \sum a_i X_i \quad (A.2.4)$$

to mean the linear combination (linear convolution or linear filtering if a_i are assumed to be filter coefficients) of independent random variable X . Then, output Y has a relationship with the X and defined below;

$$X \cdot \geq Y \quad (A.2.5)$$

The notation $\cdot \geq$ is now called the *partial order* of random variables (i.e., Y is more Gaussian than X). This is supported by the following two properties:

- a) Transitivity : if $X \cdot \geq Y$ and $Y \cdot \geq Z$ then $X \cdot \geq Z$, Z is more Gaussian than X
- b) Asymmetry : Let X and Y have finite variances. If $X \cdot \geq Y$ and $Y \cdot \geq X$ then $X \doteq Y$. X and Y are equivalent in a probabilistic sense.

If Z is Gaussian,

$$Z \cdot \geq X, \quad \text{for any } X \quad (A.2.6)$$

meaning that there are no other random variables which are more Gaussian than Z .

On the other hands, for X having a certain distribution but not exactly Gaussian,

$$X \cdot \geq \sum a_i X_i \cdot \geq Z \quad (A.2.7)$$

Equation (A.2.7) is strict unless

- a) X is Gaussian (then $X \doteq \sum a_i X_i \doteq Z$), i.e., all the random variables are equivalent in statistical sense.
- b) X is not Gaussian, but the linear combination is trivial (i.e., no two a_i is non zero) (then $X \doteq \sum a_i X_i \cdot \geq Z$). Simple scaling of a random variable doesn't

change any statistical properties of itself.

The case a) implies that the efficiency of the deconvolution process is significantly degraded when the input signal to be restored has the form of the Gaussian distribution. In a word, relation (A.2.7) says that linear combinations of independent random variables are “more nearly” Gaussian than the individual components of the combination. Note that $X \geq Y$ means Y is more Gaussian than X .

A.2.2 Partial order and blind deconvolution (inverse filtering)

When a non-Gaussian, i.i.d. signal excites a linear system- the output of the linear system tends to be closer to Gaussianity due to the linear filtering. In other words, the probability density function of the output from the linear combination (convolution) is closer to Gaussianity than that of the individual signal. This has been already explained through the relationship of Gaussianity to linear filtering – i.e., the central limit theorem.

The blind deconvolution scheme aims to find the inverse filter coefficients from which the output of the inverse filter should be *less* Gaussian. The equations used in FIR inverse filter coefficient calculation have to take a certain form of coefficients with a certain length to solve the equation.



Figure A.2.2 Gaussianity and filtering

The above figure notes that if an *i.i.d.*, non-Gaussian signal is filtered by a linear system, its output tends to be more Gaussian (Donoho, 1981).

Starting from this idea, a relationship between the partial order and blind deconvolution problem will be discussed. To do this, Figure A.2.1 is developed in Figure A.2.3 in detailed form

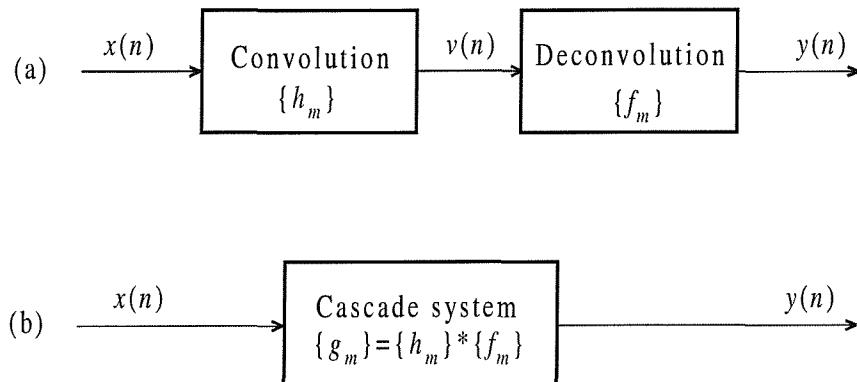


Figure A.2.3 Deconvolution: (a) convolution-deconvolution operation; (b) equivalent system (for a noise free case)

A certain statistical value calculated from the output of an arbitrarily given inverse filter is defined to represent the degree of the Gaussianity (e.g., skewness or kurtosis of y). In such a manner that the statistical value can discriminate between the distribution of v and that of $\sum a_i v(n-i)$ (an output of linear filter). In other words, this requires the selection of an *Objective* function O which can be equivalent to the partial order $\cdot \geq$ described in the previous section. Thus, O agrees with order $\cdot \geq$ on v if $v \cdot \geq y$ implies

$$O(v) > O(y) \quad (\text{A.2.8})$$

for every v, y which are filtered white noise sampled from random variable in v , i.e., if y is more Gaussian than v , then the Objective function on v is larger than on y . Hence, making the role of the objective function cope with the partial order becomes the basis of the BD process. This objective function can be formed from using the characteristics of cumulants, which is defined by the normalised cumulant of order (r,s) of v as

$$O_v(r,s) = c_r^v / (c_s^v)^{r/s}, \quad r > s \quad (\text{A.2.9})$$

where c_r^v represents the r -th order cumulant of signal v and the order of denominator's cumulant s is even integer greater than or equal 2 less than the integer r . If $y(n) \doteq \sum a_i v(n-i)$, where $v(n-i)$ are independent samples of random variable v ,

$$O_y(r, s) = O_v(r, s) \frac{\sum (a_i)^r}{(\sum a_i^s)^{r/s}} \quad (\text{A.2.10})$$

So, for $r > s$ and even $s \geq 2$, $|O_y(r, s)| \leq |O_v(r, s)|$. Therefore, $\mathbf{v} \geq \mathbf{y}$ implies

$$|O_v(r, s)| \geq |O_y(r, s)| \quad \text{even } s \geq 2, \quad \text{and} \quad r > s \quad (\text{A.2.11})$$

Thus, equation (A.2.11) says that normalised cumulants of order (r, s) for \mathbf{v} and \mathbf{y} are consistent with $\cdot \geq$. When \mathbf{v} has zero mean, $O_v(4, 2)$ thus corresponds to the value of ‘kurtosis’ of signal \mathbf{v} and has been used in Wiggins’s Minimum Entropy Deconvolution (MED). This shows that the original MED procedure is consistent with equation (A.2.8) when \mathbf{v} has a kurtosis greater than 3 (leptokurtic). If \mathbf{v} has a kurtosis smaller than 3 (platokurtic), *minimising*, not maximising $O_y(4, 2)$ is the appropriate strategy.

Also, choosing the order r equals 3 and s as 2, the normalised cumulant takes the form of ‘skewness’ of a signal.

The input-output relationship in a linear system is usually expressed in a linear form

$$y(n) = \sum_m f_m v(n-m) \quad (\text{A.2.12})$$

where $y(n)$ is the output of the inverse system which has an impulse response $\{f_m\}$ and $v(n)$ indicates the input of the inverse system. Relating (A.2.12) to (A.2.11) and denoting the objective function of output $O_y(r, s)$ then

$$O_y(r, s) = \frac{\sum_m (f_m)^r}{\left| \sum_m (f_m)^s \right|^{r/s}} \cdot O_v(r, s) \quad (\text{A.2.13})$$

By taking the absolute value of (A.2.13),

$$|O_y(r, s)| = \frac{\left| \sum_m (f_m)^r \right|}{\left\{ \left| \sum_m (f_m)^s \right|^{1/s} \right\}^r} \cdot |O_v(r, s)| \quad (\text{A.2.14})$$

in which the term $\left| \sum_m (f_m)^s \right|^{1/s}$ designates the standard l_s norm $\|\underline{f}\|_s$ and the numerator is $\{\|\underline{f}\|_r\}^r$. From the inequality of the rational term for s an even integer and $r > s$ case,

$$\left\{ \frac{\|\underline{f}\|_r}{\|\underline{f}\|_s} \right\}^r \leq 1 \quad (\text{A.2.15})$$

Using (A.2.15) for (A.2.14) yields a bound for a positive even integer s ,

$$|O_y(r, s)| \leq |O_v(r, s)| \quad \text{for } r > s \quad (\text{A.2.16})$$

The upper bound of $O_y(r, s)$ thus can be interpreted as follows; Any time-invariant linear operation (non-Gaussian i.i.d. time series input) results in a time series whose magnitude of the normalised cumulants is less than or equal to that of the excitation for all even $s < r$.

Thus, the required deconvolving operator must generate a response whose normalised cumulants have magnitudes that are the largest over the class of all linear operators for all even $s < r$.

The study in this thesis exclusively considers the case in which $s < r$ with s even, whereby it is desired to select \mathbf{g} (the impulse response sequence of the convolution-deconvolution operator in Figure A.2.3) so as to maximise an estimate of $|O_y(r, s)|$. That is, the ideal deconvolution operation is performed by maximising the magnitude of the normalised response cumulant $O_y(r, s; \mathbf{g})$, where s is any positive even integer less than r for which cumulant of order s , c_s^y is nonzero (e.g., use the order s equals 2 meaning the variance of signal). This maximisation is to be made with respect to the unit-impulse response $\{g_m\}$ of the combined convolution-deconvolution operator of the Figure A.2.3 (b). Since the unit-impulse response of the unknown linear convolution operator $\{h_m\}$ is implicitly contained within the output data $y(n)$, this maximisation must be made with respect to the deconvolution operator's unit-impulse response $\{f_m\}$. The required maximisation therefore takes the form

$$\max_{\mathbf{f}} |O_y(r, s)| = \max_{\mathbf{f}} [O_y(r, s) \operatorname{sgn}[O_y(r, s)]] \quad (\text{A.2.17})$$

where \mathbf{f} is an appropriately dimensioned vector whose components are the elements of the unit-impulse response of the deconvolving operator (inverse filter). In this way, the equation (A.2.17) is devised to find a global maximum of $|O_y(r, s)|$. Normally, the necessary condition which give *a local maximum is done by differentiating and equating to zero with respect to the filter coefficients*.

To support this, the restoration of the impacting signals that consist of a few impulsive variables can be achieved through the inversion of the system h_m in Figure A.2.3 unless there is noise interference. However, the observed signals are often contaminated by noise signal (assumed Gaussian) making the observation ‘more Gaussian’. Thus, a simple inversion of the system cannot provide a correct restoration of the impulse signal. Focusing on the problem of restoring the highly non-Gaussian spiky signals, it would be more general to say that the BD process can be achieved by making the inverse system produce an output that is strongly non-Gaussian. This is the main objective of the Blind deconvolution process.

Appendix B

Numerical expansion of the higher order deconvolution operator (part of Chapter 3)

This Appendix provides numerical solutions for the nonrecursive (MA) and recursive (AR or ARMA) inverse filter coefficient calculation which are used in Chapter 3.

B.1 Linear Nonrecursive Deconvolution operator (MA inverse filter)

The input-system-output relationship can be expressed as the linear constant-coefficient difference equation

$$y(n) = \sum_{k=0}^{L-1} f_k v(n-k) \quad (\text{B.1.1})$$

where f_k is inverse filter coefficient ($k=0,1,\dots,L-1$). The normalised cumulant of order (r,s) is used in which $r>s$ and integers (for example, $r=4$ and $s=2$)

$$O_y(r,s) = \sum_{k=0}^{N-1} y^r(k) / \left[\sum_{k=0}^{N-1} y^s(k) \right]^{r/s} \quad (\text{B.1.2})$$

Optimising this objective function with respect to the filter coefficient f_i ($i=0,1,\dots,L-1$), the necessary condition is

$$\partial O_y(r,s) / \partial f_i = 0 \quad (\text{B.1.3})$$

with the expansion form

Appendix B, Numerical expansion of the Higher order deconvolution operator

$$\begin{aligned} \frac{\partial O_y(r,s)}{\partial f_i} &= \frac{r \sum_{k=0}^{N-1} y^{r-1}(k) \frac{\partial y(k)}{\partial f_i} \left(\sum_{k=0}^{N-1} y^s(k) \right)^{r/s} - r \left(\sum_{k=0}^{N-1} y^s(k) \right)^{r/s-1} \cdot \sum_{k=0}^{N-1} y^{s-1}(k) \cdot \frac{\partial y(k)}{\partial f_i} \cdot \sum_{k=0}^{N-1} y^r(k)}{\left(\sum_{k=0}^{N-1} y^s(k) \right)^{2r/s}} \quad (\text{B.1.4}) \\ &= 0 \end{aligned}$$

Rearranging the numerator term of equation (B.1.4) and setting to zero becomes

$$\sum_{k=0}^{N-1} y^{s-1}(k) \cdot \frac{\partial y(k)}{\partial f_i} = \left(\sum_{k=0}^{N-1} y^s(k) / \sum_{k=0}^{N-1} y^r(k) \right) \sum_{k=0}^{N-1} y^{r-1}(k) \frac{\partial y(k)}{\partial f_i} \quad (\text{B.1.5})$$

for $s=2$ case,

$$\sum_{k=0}^{N-1} y^{s-1}(k) = \sum_{k=0}^{N-1} y(k) = \sum_{k=0}^{N-1} \left(\sum_{i=0}^{L-1} f_i v(k-i) \right) \quad (\text{B.1.6})$$

$$\text{Since, } \frac{\partial y(n)}{\partial f_0} = \frac{\sum_{i=0}^{L-1} f_i v(n-i)}{\partial f_0} = f_0 \frac{\partial v(n-0)}{\partial f_0} + f_1 \frac{\partial v(n-1)}{\partial f_0} + f_2 \frac{\partial v(n-2)}{\partial f_0} + \dots = v(n-0) \quad ,$$

the term $\frac{\partial y(k)}{\partial f_i}$ can be written as

For $m=0,1,2,\dots,L-1$

$$\frac{\partial y(k)}{\partial f_m} \equiv v(k-m) \quad (\text{B.1.7})$$

Thus, substituting (B.1.6) and (B.1.7) into (B.1.5) yields

$$\sum_{i=0}^{L-1} f_i \sum_{k=0}^{N-1} v(k-m) v(k-i) = \left(\sum_{k=0}^{N-1} y^s(k) / \sum_{k=0}^{N-1} y^r(k) \right) \sum_{k=0}^{N-1} y^{r-1}(k) v(k-m) \quad (\text{B.1.8})$$

for $m=0,1,\dots,L-1$.

Expressing this in matrix form

$$\begin{bmatrix} \sum_{k=0}^{N-1} v(k) v(k) & \sum_{k=0}^{N-1} v(k) v(k-1) & \dots & \sum_{k=0}^{N-1} v(k) v(k-q) \\ \sum_{k=0}^{N-1} v(k-1) v(k) & \sum_{k=0}^{N-1} v(k-1) v(k-1) & \dots & \sum_{k=0}^{N-1} v(k-1) v(k-q) \\ \vdots & \vdots & & \vdots \\ \sum_{k=0}^{N-1} v(k-m) v(k) & \sum_{k=0}^{N-1} v(k-m) v(k-1) & \dots & \sum_{k=0}^{N-1} v(k-m) v(k-q) \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_{L-1} \end{bmatrix} = \begin{bmatrix} A \cdot \sum_{k=0}^{N-1} y^{r-1}(k) v(k) \\ A \cdot \sum_{k=0}^{N-1} y^{r-1}(k) v(k-1) \\ \vdots \\ A \cdot \sum_{k=0}^{N-1} y^{r-1}(k) v(k-m) \end{bmatrix} \quad (\text{B.1.9})$$

where $A = \sum_{k=0}^{N-1} y^s(k) / \sum_{k=0}^{N-1} y^r(k)$. equation (B.1.9) can be written equivalently,

$$\mathbf{R}_{vv} \cdot \mathbf{b} = \mathbf{g} \quad (\text{B.1.10})$$

where \mathbf{R}_{vv} denotes the symmetry $L \times L$ autocorrelation matrix of the observed signal, \mathbf{b} is $L \times 1$ inverse filter coefficient vector, and \mathbf{g} is $L \times 1$ cross-correlation vector between the observed signal and the output of the inverse filter. This is the same equation used in Chapter 3 (equation (3.3.17)).

B.2 Linear Recursive Deconvolution operator (ARMA inverse filter)

B.2.1 The MA part

The optimisation condition can be solved for the MA part of the inverse filter coefficients

$$\begin{aligned} \frac{\partial O_y(r, s)}{\partial b_i} &= \frac{r \sum_{k=0}^{N-1} y^{r-1}(k) \frac{\partial y(k)}{\partial b_i} \left(\sum_{k=0}^{N-1} y^s(k) \right)^{r/s} - r \left(\sum_{k=0}^{N-1} y^s(k) \right)^{r/s-1} \cdot \sum_{k=0}^{N-1} y^{s-1}(k) \cdot \frac{\partial y(k)}{\partial b_i} \cdot \sum_{k=0}^{N-1} y^r(k)}{\left(\sum_{k=0}^{N-1} y^s(k) \right)^{2r/s}} \quad (\text{B.2.1}) \\ &= 0 \end{aligned}$$

Rearranging the numerator term of equation (B.2.1) and setting to zero becomes

$$\sum_{k=0}^{N-1} y^{s-1}(k) \cdot \frac{\partial y(k)}{\partial b_i} = \left(\sum_{k=0}^{N-1} y^s(k) / \sum_{k=0}^{N-1} y^r(k) \right) \sum_{k=0}^{N-1} y^{r-1}(k) \frac{\partial y(k)}{\partial b_i} \quad (\text{B.2.2})$$

For mathematical brevity, let's denote

$$\frac{\partial y(n)}{\partial b_m} \equiv s_b(n - m) \quad (\text{B.2.3})$$

for $m=0,1,2,\dots,q-1$ and $n=0,1,\dots,N-1$. The term s_b stands for the derivation of the output of the inverse filter with respect to the MA part coefficient and is called as the elements of the response sensitivity sequence for the MA part.

Since the output of the AR, MA part can be expressed

Appendix B, Numerical expansion of the Higher order deconvolution operator

$$\sum_{k=0}^{N-1} y^{s-1}(k) = \sum_{k=0}^{N-1} \left(-\sum_{j=0}^{p-1} a_j y(k-j) + \sum_{i=0}^{q-1} b_i v(k-i) \right)^{s-1} \quad (B.2.4)$$

Setting $s=2$ and combining (B.2.2), (B.2.3), and (B.2.4) yields (for $m=0,1,\dots,q-1$)

$$\sum_{i=0}^{q-1} b_i \sum_{k=0}^{N-1} s_b(k-m) v(k-i) = \sum_{j=0}^{p-1} a_j \sum_{k=0}^{N-1} s_b(k-m) y(k-j) + \left(\sum_{k=0}^{N-1} y^r(k) / \sum_{k=0}^{N-1} y^r(k) \right) \sum_{k=0}^{N-1} y^{r-1}(k) s_b(k-m) \quad (B.2.5)$$

which is the similar formulation to that of FIR estimation, e.g., the filter coefficient can be calculated by the between the correlation of x , s_b and correlation of y , s_b . To solve the term $s_b(k-m)$ in equation (B.2.3) and (B.2.5), we first calculate the term $\frac{\partial y(k)}{\partial b_i}$ as following ($a_0 = 1$)

$$\begin{aligned} \text{For } i = 0; \quad a_0 \frac{\partial y(n)}{\partial b_0} + a_1 \frac{\partial y(n-1)}{\partial b_0} + a_2 \frac{\partial y(n-2)}{\partial b_0} + \dots &= \frac{\sum_{i=0}^{q-1} b_i v(n-i)}{\partial b_0} = v(n-0) \\ \text{For } i = 1; \quad a_0 \frac{\partial y(n)}{\partial b_1} + a_1 \frac{\partial y(n-1)}{\partial b_1} + a_2 \frac{\partial y(n-2)}{\partial b_1} + \dots &= \frac{\sum_{i=0}^{q-1} b_i v(n-i)}{\partial b_1} = v(n-1) \\ \text{For } i = q-1; \quad a_0 \frac{\partial y(n)}{\partial b_{q-1}} + a_1 \frac{\partial y(n-1)}{\partial b_{q-1}} + a_2 \frac{\partial y(n-2)}{\partial b_{q-1}} + \dots &= \frac{\sum_{i=0}^{q-1} b_i v(n-i)}{\partial b_{q-1}} = v(n-q+1) \end{aligned} \quad (B.2.6)$$

Expressing (B.2.6) in compact form with AR part coefficient and the observed signal

$$\sum_{j=0}^{p-1} a_j s_b(n-m) = v(n-m) \quad (B.2.7)$$

for $m=0,1,2,\dots,q-1$ and $n=0,1,\dots,N-1$.

Note that in (B.2.7), when the AR part coefficient is $a_0=1$ and $a_i=0$ for $i=1,2,\dots,p-1$ the sensitivity factor s_b becomes identical to the observed signal $v(n)$. When we express (B.2.6) in matrix form for every coefficient of the MA part,

For $m = 0$

$$\begin{bmatrix} \frac{\partial y(0-0)}{\partial b_0} & \frac{\partial y(0-1)}{\partial b_0} & \dots & \frac{\partial y(0-p+1)}{\partial b_0} \\ \frac{\partial y(1-0)}{\partial b_0} & \frac{\partial y(1-1)}{\partial b_0} & \dots & \frac{\partial y(1-p+1)}{\partial b_0} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y(N-1-0)}{\partial b_0} & \frac{\partial y(N-1-1)}{\partial b_0} & \dots & \frac{\partial y(N-1-p+1)}{\partial b_0} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{bmatrix} = \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(N-1) \end{bmatrix}$$

For $m = 1$

$$\begin{bmatrix} \frac{\partial y(0-0)}{\partial b_1} & \frac{\partial y(0-1)}{\partial b_1} & \dots & \frac{\partial y(0-p+1)}{\partial b_1} \\ \frac{\partial y(1-0)}{\partial b_1} & \frac{\partial y(1-1)}{\partial b_1} & \dots & \frac{\partial y(1-p+1)}{\partial b_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y(N-1-0)}{\partial b_1} & \frac{\partial y(N-1-1)}{\partial b_1} & \dots & \frac{\partial y(N-1-p+1)}{\partial b_1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{bmatrix} = \begin{bmatrix} v(-1) \\ v(0) \\ \vdots \\ v(N-2) \end{bmatrix}$$

For $m = q-1$

$$\begin{bmatrix} \frac{\partial y(0-0)}{\partial b_{q-1}} & \frac{\partial y(0-1)}{\partial b_{q-1}} & \dots & \frac{\partial y(0-p+1)}{\partial b_{q-1}} \\ \frac{\partial y(1-0)}{\partial b_{q-1}} & \frac{\partial y(1-1)}{\partial b_{q-1}} & \dots & \frac{\partial y(1-p+1)}{\partial b_{q-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y(N-1-0)}{\partial b_{q-1}} & \frac{\partial y(N-1-1)}{\partial b_{q-1}} & \dots & \frac{\partial y(N-1-p+1)}{\partial b_{q-1}} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{bmatrix} = \begin{bmatrix} v(0-q+1) \\ v(1-q+1) \\ \vdots \\ v(N-1-q+1) \end{bmatrix} \quad (B.2.8)$$

Appendix B, Numerical expansion of the Higher order deconvolution operator

The s_b terms in equation (B.2.7) are calculated via the matrix form,

$$\left[\begin{array}{c|cccc} & m=0 & m=1 & \dots & m=p-1 \\ \hline n=0 & s_b(0-0) & s_b(0-1) & \dots & s_b(0-p+1) \\ n=1 & s_b(1-0) & s_b(1-1) & \dots & s_b(1-p+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n=N-1 & s_b(N-1-0) & s_b(N-1-1) & \dots & s_b(N-1-p+1) \end{array} \right] \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{bmatrix} = \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(N-1) \end{bmatrix} \quad (\text{B.2.9})$$

Using the obtained sensitivity sequence matrix from (B.2.9), equation (B.2.5) can be expressed in matrix form

$$\begin{aligned} & \left[\begin{array}{cccc} \sum_{k=0}^{N-1} s_b(k)v(k) & \sum_{k=0}^{N-1} s_b(k)v(k-1) & \dots & \sum_{k=0}^{N-1} s_b(k)v(k-q+1) \\ \sum_{k=0}^{N-1} s_b(k-1)v(k) & \sum_{k=0}^{N-1} s_b(k-1)v(k-1) & \dots & \sum_{k=0}^{N-1} s_b(k-1)v(k-q+1) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=0}^{N-1} s_b(k-m)v(k) & \sum_{k=0}^{N-1} s_b(k-m)v(k-1) & \dots & \sum_{k=0}^{N-1} s_b(k-m)v(k-q+1) \end{array} \right] \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{q-1} \end{bmatrix} = \\ & \left[\begin{array}{cccc} \sum_{k=0}^{N-1} s_b(k)y(k-1) & \sum_{k=0}^{N-1} s_b(k)y(k-2) & \dots & \sum_{k=0}^{N-1} s_b(k)y(k-p+1) \\ \sum_{k=0}^{N-1} s_b(k-1)y(k-1) & \sum_{k=0}^{N-1} s_b(k-1)y(k-2) & \dots & \sum_{k=0}^{N-1} s_b(k-1)y(k-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=0}^{N-1} s_b(k-m)y(k-1) & \sum_{k=0}^{N-1} s_b(k-m)y(k-2) & \dots & \sum_{k=0}^{N-1} s_b(k-m)y(k-p+1) \end{array} \right] \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{bmatrix} + \begin{bmatrix} A \cdot \sum_{k=0}^{N-1} y^{r-1}(k)s_b(k) \\ A \cdot \sum_{k=0}^{N-1} y^{r-1}(k)s_b(k-1) \\ \vdots \\ A \cdot \sum_{k=0}^{N-1} y^{r-1}(k)s_b(k-m) \end{bmatrix} \quad (\text{B.2.10}) \end{aligned}$$

where $m=q-1$ and $A = \sum_{k=0}^{N-1} y^s(k) / \sum_{k=0}^{N-1} y^r(k)$.

Or, in a compact form

$$\mathbf{R}_{s_b v} \cdot \mathbf{b} = \mathbf{R}_{s_b y} \cdot \mathbf{a} + \mathbf{r}_{y s_b} \quad (\text{B.2.11})$$

B.2.2 The AR part

The optimisation condition can be solved for the AR part inverse filter coefficient

$$\frac{\partial O_y(r,s)}{\partial a_j} = \frac{r \sum_{k=0}^{N-1} y^{r-1}(k) \frac{\partial y(k)}{\partial a_j} \left(\sum_{k=0}^{N-1} y^s(k) \right)^{r/s} - r \left(\sum_{k=0}^{N-1} y^s(k) \right)^{r/s-1} \cdot \sum_{k=0}^{N-1} y^{s-1}(k) \cdot \frac{\partial y(k)}{\partial a_j} \cdot \sum_{k=0}^{N-1} y^r(k)}{\left(\sum_{k=0}^{N-1} y^s(k) \right)^{2r/s}} = 0 \quad (\text{B.2.12})$$

Rearranging the numerator term of equation (B.2.12) and setting to zero becomes

$$\sum_{k=0}^{N-1} y^{s-1}(k) \cdot \frac{\partial y(k)}{\partial a_j} = \left(\sum_{k=0}^{N-1} y^s(k) / \sum_{k=0}^{N-1} y^r(k) \right) \sum_{k=0}^{N-1} y^{r-1}(k) \frac{\partial y(k)}{\partial a_j} \quad (\text{B.2.13})$$

Let assume $a_0=1$ and denote

$$\frac{\partial y(n)}{\partial a_m} \equiv s_a(n-m) \quad (\text{B.2.14})$$

for $m=0,1,\dots,p-1$ and $n=0,2,\dots,N-1$. The term s_a is called as sensitivity factor for AR part coefficient.

And for $s=2$ case, the left part of equation (B.2.12) can be written

$$\sum_{k=0}^{N-1} y^{s-1}(k) = \sum_{k=0}^{N-1} \left(- \sum_{j=0}^{p-1} a_j y(k-j) + \sum_{i=0}^{q-1} b_i x(k-i) \right) \quad (\text{B.2.15})$$

Thus, combining (B.2.13), (B.2.14), and (B.2.15) yields (for $m=0,1,\dots,p-1$)

$$\sum_{j=0}^{p-1} a_j \sum_{k=0}^{N-1} s_a(k-m) y(k-j) = \sum_{i=0}^{q-1} b_i \sum_{k=0}^{N-1} s_a(k-m) v(k-i) - \left(\sum_{k=0}^{N-1} y^s(k) / \sum_{k=0}^{N-1} y^r(k) \right) \sum_{k=0}^{N-1} y^{r-1}(k) s_a(k-m) \quad (\text{B.2.16})$$

As the same manner to the MA part case, the sensitivity factor for AR part can be calculated as following:

$$\text{Since, } a_0 \frac{\partial y(n)}{\partial a_1} + a_1 \frac{\partial y(n-1)}{\partial a_1} + a_2 \frac{\partial y(n-2)}{\partial a_1} + \dots = y(n-1),$$

$$\sum_{j=0}^{p-1} a_j s_a(n-j) = y(n-1) \quad (\text{B.2.17})$$

The s_a terms in equation (B.2.17) are calculated via the matrix form imposing zero initial condition,

Appendix B, Numerical expansion of the Higher order deconvolution operator

$$\left[\begin{array}{c|cccc} & m=0 & m=1 & \dots & m=p-1 \\ \hline n=0 & s_a(0-0) & s_a(0-1) & \dots & s_a(0-p+1) \\ n=1 & s_a(1-0) & s_a(1-1) & \dots & s_a(1-p+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n=N-1 & s_a(N-1-0) & s_a(N-1-1) & \dots & s_a(N-1-p+1) \end{array} \right] \cdot \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{bmatrix} = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix} \quad (\text{B.2.18})$$

Hence, similarly to the MA part calculation, equation (B.2.16) can be expressed in matrix form for AR part coefficient

$$\left[\begin{array}{cccc} \sum_{k=0}^{N-1} s_a(k-0)y(k-0) & \sum_{k=0}^{N-1} s_a(k-0)y(k-1) & \dots & \sum_{k=0}^{N-1} s_a(k-0)y(k-p+1) \\ \sum_{k=0}^{N-1} s_a(k-1)y(k-0) & \sum_{k=0}^{N-1} s_a(k-1)y(k-1) & \dots & \sum_{k=0}^{N-1} s_a(k-1)y(k-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=0}^{N-1} s_a(k-m)y(k-0) & \sum_{k=0}^{N-1} s_a(k-m)y(k-1) & \dots & \sum_{k=0}^{N-1} s_a(k-m)y(k-p+1) \end{array} \right] \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} =$$

$$\left[\begin{array}{cccc} \sum_{k=0}^{N-1} s_a(k-0)v(k) & \sum_{k=0}^{N-1} s_a(k-0)v(k-1) & \dots & \sum_{k=0}^{N-1} s_a(k-0)v(k-q+1) \\ \sum_{k=0}^{N-1} s_a(k-1)v(k) & \sum_{k=0}^{N-1} s_a(k-1)v(k-1) & \dots & \sum_{k=0}^{N-1} s_a(k-1)v(k-q+1) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=0}^{N-1} s_a(k-m)v(k) & \sum_{k=0}^{N-1} s_a(k-m)v(k-1) & \dots & \sum_{k=0}^{N-1} s_a(k-m)v(k-q+1) \end{array} \right] \cdot \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{q-1} \end{bmatrix} + \begin{bmatrix} A \cdot \sum_{k=0}^{N-1} y^{r-1}(k)s_a(k-0) \\ A \cdot \sum_{k=0}^{N-1} y^{r-1}(k)s_a(k-1) \\ \vdots \\ A \cdot \sum_{k=0}^{N-1} y^{r-1}(k)s_a(k-m) \end{bmatrix} \quad (\text{B.2.19})$$

where $m=p-1$ and $A = \sum_{k=0}^{N-1} y^s(k) / \sum_{k=0}^{N-1} y^r(k)$.

Or, in a compact form

$$\mathbf{R}_{s_a y} \cdot \mathbf{a} = \mathbf{R}_{s_a v} \cdot \mathbf{b} + \mathbf{r}_{y s_a} \quad (\text{B.2.20})$$

Appendix C

Scheme of Independent Component Analysis (Blind Source Separation, BSS)

In this section we tackle a fundamental signal processing problem with tools from Higher Order Statistics (HOS) and multilinear algebra, namely the problem of *Independent Component Analysis* (ICA), which is also known as Blind Source Separation (BSS).

The goal of ICA is the decomposition of a multichannel data set in an a priori unknown linear mixture of a priori unknown source signals, relying on the assumption that the source signals are mutually statistically independent. This concept is in fact a fine-tuning of the more well-known *Principal Component Analysis* (PCA), where one aims at the decomposition in a linear mixture of uncorrelated components. PCA involves the Eigenvalue Decomposition (EVD) of the covariance matrix and ICA relies on tensorial generalisation of the EVD applied to a higher-order cumulant.

The ICA is based on the tensorial diagonalisation of the higher-order cumulant tensor (matrix) of the observation signal vector, which is similar to the diagonalisation of the covariance matrix in PCA. Hence, HOEVD of a cumulant tensor can be used to solve the ICA problem.

C.1 Basic equation

The multi-input multi-output linear system can be expressed by following equation,

$$\mathbf{v} = \mathbf{M}\mathbf{x} + \mathbf{w} \quad (\text{C.1.1})$$

where the observed signal $\mathbf{v} \in R^I$ (I observation channel), the unknown source signal $\mathbf{x} \in R^J$ (J independent source signal), the additive Gaussian noise signal $\mathbf{w} \in R^I$, and the unknown mixing matrix $\mathbf{M} \in R^{I \times J}$ whose columns $\{\mathbf{m}_i\}$ are the mixing vectors. From only the observed signal matrix \mathbf{y} (whose row consists of N samples of a time sequence and I observation channel, i.e., $\{v_{i,n}\}_{n=1,2,\dots,N \& i=1,2,\dots,I}$), the identification of the mixing matrix \mathbf{M} or the unknown source signal matrix \mathbf{x} is to be restored. Hence, this addresses the *blind identification/deconvolution* problem.

C.2 Assumptions

- The mixing vectors are linearly independent.
- The components of \mathbf{x} are mutually statistically independent.
- The noise components are Gaussian and uncorrelated with the source signal.

C.3 PCA and pre-whitening

The mixing matrix $\mathbf{M} \in R^{I \times J}$ can be decomposed and written as,

$$\mathbf{M} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T \quad (\text{C.3.1})$$

where \mathbf{U} is the left singular vector, \mathbf{V} is the right singular vector, \mathbf{S} is a matrix whose diagonal components are singular values, and T represents the matrix transpose [Golub and Van Loan, 1996].

Making zero mean of the observed signal of (C.1.1) gives

$$\mathbf{v} = \mathbf{v} - \mathbf{m}_v \quad (\text{C.3.2})$$

By considering that the noise is uncorrelated with the source, the output (observed) covariance can be written as

$$\mathbf{C}_v^{(2)} = \mathbf{M} \cdot \mathbf{C}_x^{(2)} \cdot \mathbf{M}^T + \mathbf{C}_w \quad (C.3.3)$$

The pre-whitened covariance matrix is estimated as following;

$$\tilde{\mathbf{C}}_v^{(2)} = [\mathbf{C}_v^{(2)} - \sigma_w \mathbf{I}] \quad (C.3.4)$$

For $I > J$ (more sensors than sources) case, the variance of the noise can be estimated by averaging $I-J$ eigenvalues of $\mathbf{C}_v^{(2)}$ [Vaseghi, 1996].

Also, by assuming the uncorrelated source signals have unit variance, the pre-whitened covariance matrix of the observed signal is reduced to

$$\begin{aligned} \tilde{\mathbf{C}}_v^{(2)} &= \mathbf{M} \cdot \mathbf{M}^T \\ &= \mathbf{U} \cdot \mathbf{S}^2 \cdot \mathbf{U}^T \end{aligned} \quad (C.3.5)$$

This relationship enables to estimate the left singular matrix and the squared singular value matrix through the Eigenvalue Decomposition (EVD) of the matrix of (C.3.4).

C.4 Independent Component Analysis (ICA)

In order to determine the mixing matrix \mathbf{M} we further need to know the right singular matrix \mathbf{V} in (C.3.1).

Step 1: Projection

The projection of the observed signal into source signal subspace is done by multiplying the left singular matrix and singular value as,

$$\mathbf{z} = \mathbf{S}^\dagger \cdot \mathbf{U}^T \cdot \mathbf{v} \quad (C.4.1)$$

where † represents the pseudo inverse.

Step 2 : Higher order cumulant matrix construction

Like the second order (covariance) matrix, the P -th order ($P > 2$) cumulant of the projected matrix \mathbf{z} can be written by using the P -th order tensor notation

$$C_{\mathbf{z}}^{(P)} = C_{\mathbf{v}}^{(P)} \times_1 (\mathbf{U} \cdot \mathbf{S})^+ \times_2 (\mathbf{U} \cdot \mathbf{S})^+ \times \cdots \times_p (\mathbf{U} \cdot \mathbf{S})^+ \quad (\text{C.4.2})$$

The P -th order cumulant expression of the observed signal becomes as,

$$C_{\mathbf{v}}^{(P)} = C_{\mathbf{x}}^{(P)} \times_1 (\mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T) \times_2 (\mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T) \times \cdots \times_p (\mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T) + C_{\mathbf{w}}^{(P)} \quad (\text{C.4.3})$$

in which the $C_{\mathbf{w}}^{(P)}$ term vanishes when assuming the Gaussian noise. Substituting equation (C.4.3) into (C.4.2) yields the relationship of the P -th order cumulant of projected signal, source signal and the right singular matrix of the mixing matrix \mathbf{M} as,

$$C_{\mathbf{z}}^{(P)} = C_{\mathbf{x}}^{(P)} \times_1 \mathbf{V}^T \times_2 \mathbf{V}^T \times \cdots \times_p \mathbf{V}^T \quad (\text{C.4.4})$$

Step 3 : Higher Order Eigen Value Decomposition (HOEVD)

The P -th order cumulant of the source signal can be expressed from the equation (C.4.4),

$$C_{\mathbf{x}}^{(P)} = C_{\mathbf{z}}^{(P)} \times_1 \mathbf{V} \times_2 \mathbf{V} \times \cdots \times_p \mathbf{V} \quad (\text{C.4.5})$$

The initial matrix \mathbf{V} can be obtained by the HOEVD of the P -th order cumulant of the projected signal $C_{\mathbf{z}}^{(P)}$.

For example, the fourth-order cumulant tensor can be calculated by

$$\mathbf{z} = [z_1 \quad z_2]^T \text{ and } z_1 \text{ and } z_2 = N \times 1 \text{ vector, } N : \text{number of sample points.}$$

Hence, \mathbf{z} is pre-whitened standardised random vector ($2 \times N$ matrix, 2 means number of channels).

For $p=4$ (Fourth-order cumulant matrix), according to the JADE program, $C_{\mathbf{z}}^{(4)}$ is calculated as,

$$C_{\mathbf{z}}^{(4)} = \begin{bmatrix} \sum_{n=1}^N \left[\frac{z_1(n)^4}{N} \right] - 3 & \sum_{n=1}^N \left[\frac{z_1(n)^3 z_2(n)}{N} \right] & \sum_{n=1}^N \left[\frac{z_1(n)^2 z_2(n)^2}{N} \right] - 1 & \sum_{n=1}^N \left[\frac{z_1(n) z_2(n)^3}{N} \right] \\ \sum_{n=1}^N \left[\frac{z_1(n)^3 z_2(n)}{N} \right] & \sum_{n=1}^N \left[\frac{z_1(n)^2 z_2(n)^2}{N} \right] - 1 & \sum_{n=1}^N \left[\frac{z_1(n) z_2(n)^3}{N} \right] & \sum_{n=1}^N \left[\frac{z_2(n)^4}{N} \right] - 3 \\ \sqrt{2} \sum_{n=1}^N \left[\frac{z_1(n)^3 z_2(n)}{N} \right] & \sqrt{2} \left[\sum_{n=1}^N \left\{ \frac{z_1(n)^2 z_2(n)^2}{N} \right\} - 1 \right] \\ \sqrt{2} \left[\sum_{n=1}^N \left\{ \frac{z_1(n)^2 z_2(n)^2}{N} \right\} - 1 \right] & \sqrt{2} \sum_{n=1}^N \left[\frac{z_1(n) z_2(n)^3}{N} \right] \end{bmatrix}$$

which is a 2×6 matrix.

Step 4 : Diagonalisation of the tensor product of the projected signal cumulant and eigen matrix

Since the source signal is independent, the P -th order cumulant of the source signal $C_{\mathbf{x}}^{(P)}$ must be diagonal. As a result, we can obtain the true left singular matrix \mathbf{V} when the above tensor algebra can yield a diagonal matrix of $C_{\mathbf{x}}^{(P)}$.

This diagonalisation of the right side of equation (C.4.5) can be accomplished through the Jacobi method (or Givens rotation) from which the matrix \mathbf{V} is updated.

Step 5 : Mixing matrix and source signal reconstruction

Finally, we can get a matrix \mathbf{F}

$$\mathbf{F} = \mathbf{V} \cdot \mathbf{S}^\dagger \cdot \mathbf{U}^T \quad (C.4.6)$$

which is the inverse of the mixing matrix \mathbf{M} .

$$\mathbf{F} = \mathbf{M}^{-1} \quad (C.4.7)$$

The source signal can be obtained by

$$\hat{\mathbf{x}} = \mathbf{F} \cdot \mathbf{v} \quad (C.4.8)$$

C.5 Application of ICA for ECG signal separation

Following figure gives a biomedical example taken from J. F. Carsodo's 'separation of foetal ECG' titled "Multidimensional Independent Component Analysis", listed in Proceedings of ICASSP, 1998., Seattle, USA. The left figure (reproduced by permission of the author) displays 3 channels of cutaneous potential signals of a pregnant woman. The large pulses correspond to the mother electrocardiogram (ECG). The middle figure displays the extracted mother's ECG and the right signals represent the baby's ECG separated by the ICA.

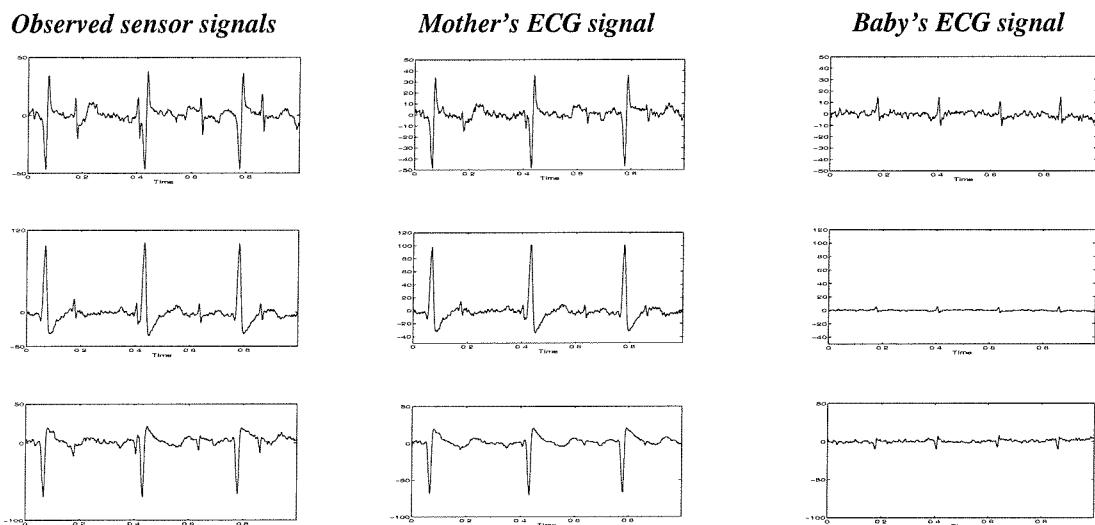


Figure C.5.1 Example of ICA application for blind biomedical source separation problem. Left : 3-channel set of cutaneous potential recordings, Middle : Estimated mother's ECG, Right : Estimated baby's ECG obtained via ICA [Cardoso, 1998].

Appendix D

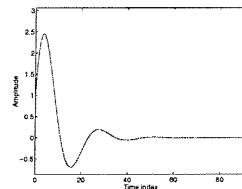
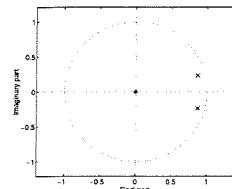
Detection, classification and reconstruction of input signals using HOSVD (summary of simulation results of Chapter 4)

This appendix summarises the extensive simulation results for detection of non-Gaussian impacting signal, classifications of signals and/or systems, and assessment of the impacting signal reconstructability using Higher Order Singular Value Decomposition (HOSVD). These are examples of the work of Chapter 4 and consider 10 different systems (only two of these are discussed in Chapter 4) and three inputs (one is a Gaussian signal and the others are non-Gaussian impacting signals). The validity of HOSVD application is supported through simulations. All terms used in the figure captions are defined in Chapter 4.

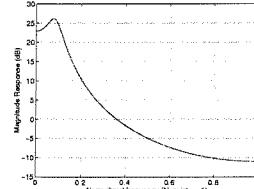
D.1 Systems used in the simulation

System 1 Pole-zeros

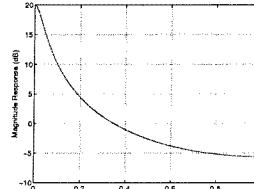
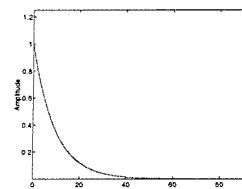
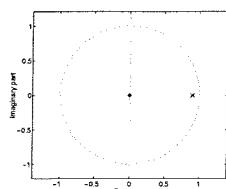
Impulse response



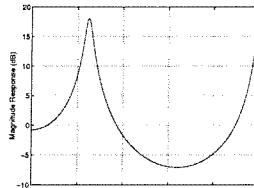
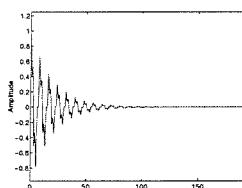
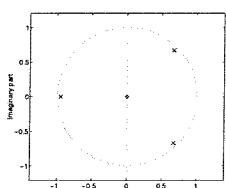
Frequency response



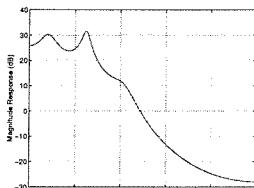
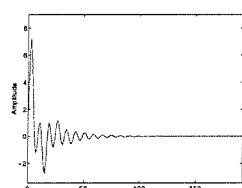
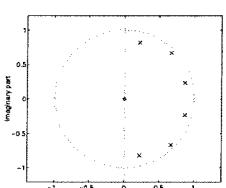
System 2



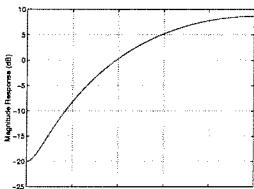
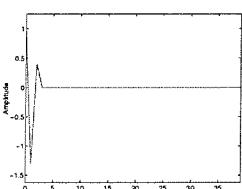
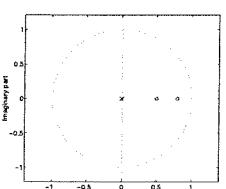
System 3



System 4

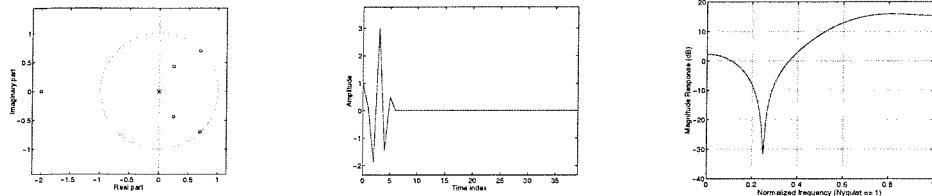


System 5

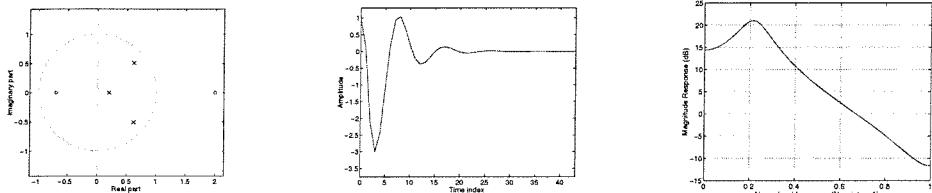


Appendix D, Detection, classification and reconstruction of input signals using HOSVD

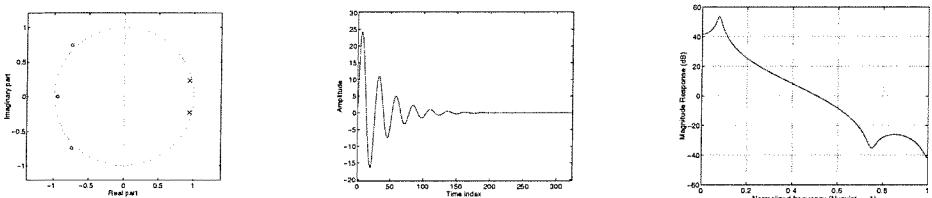
System 6



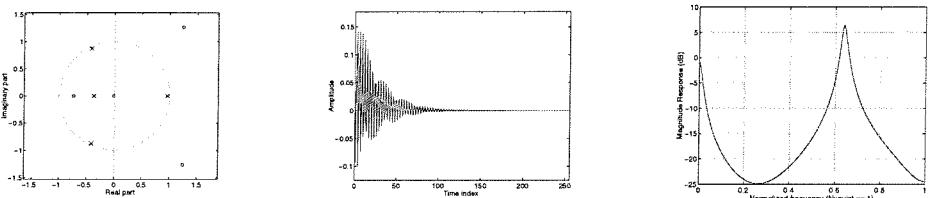
System 7



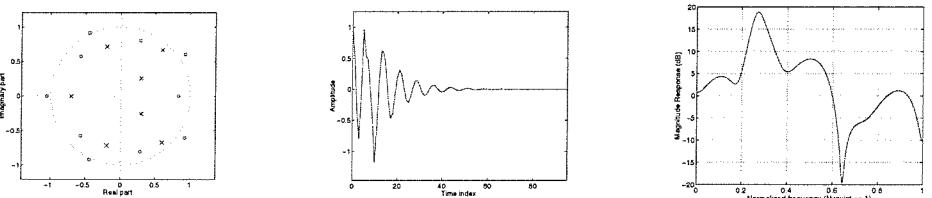
System 8



System 9

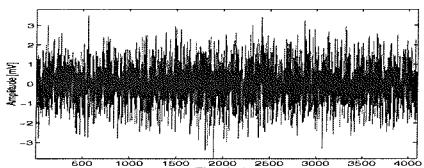


System 10



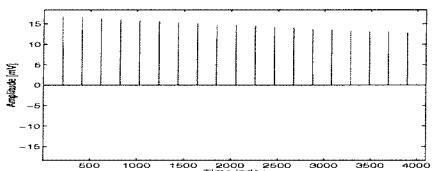
D.2 Signals used in the simulation

Case 1; Gaussian signal input



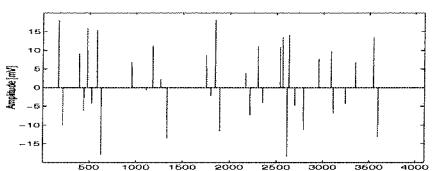
Mean : 0.017
Variance : 1.0
Skewness : 0.009
Kurtosis : 2.94
Crest factor : 3.86

Case 2; Uni-directional impacting signal input



Mean : 0.068
Variance : 0.999
Skewness : 14.72
Kurtosis : 219.98
Crest factor : 16.69

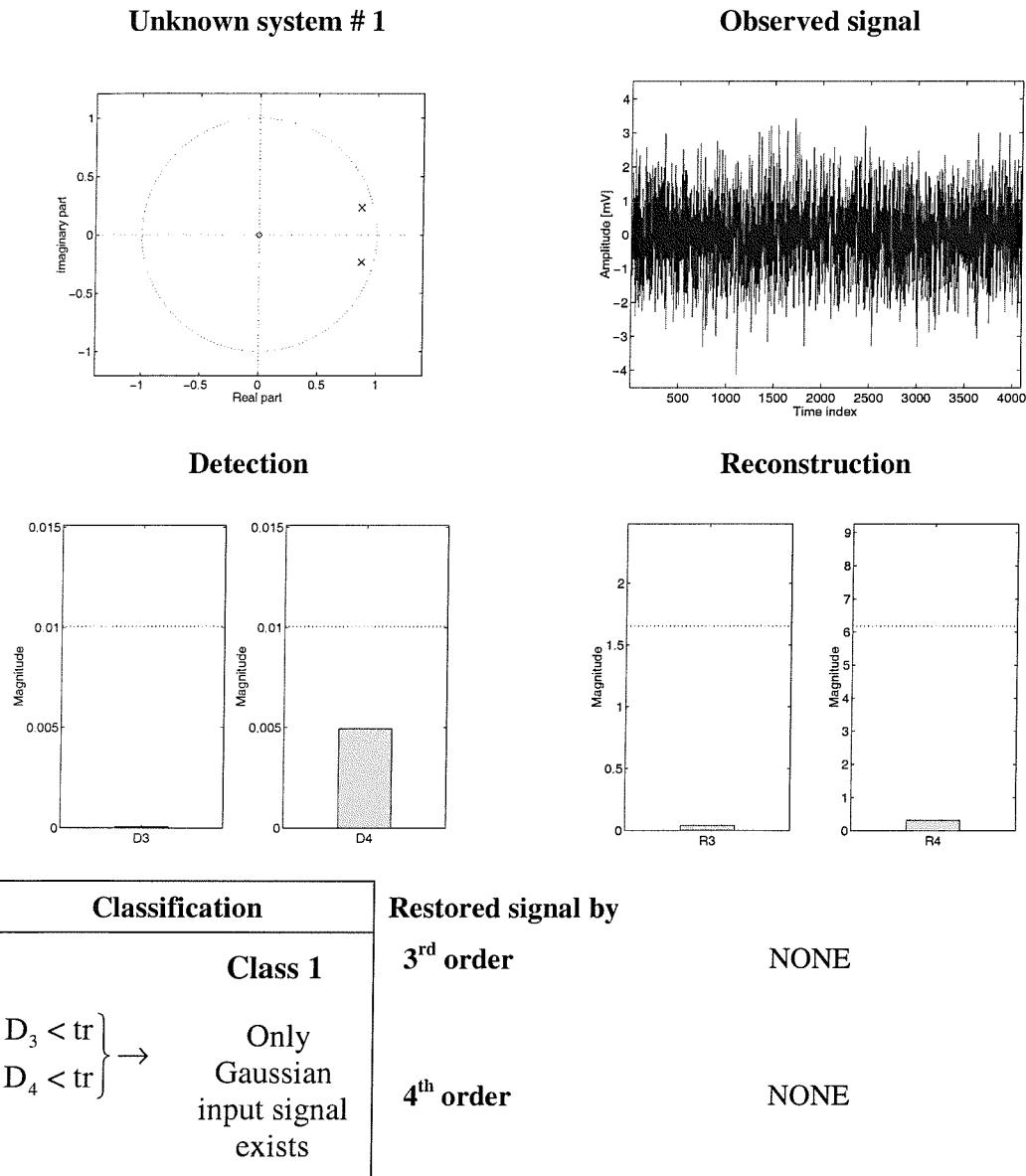
Case 3; Bi-directional impacting signal input



Mean : 1.4e-16
Variance : 0.999
Skewness : 3.05
Kurtosis : 206.98
Crest factor : 18.40

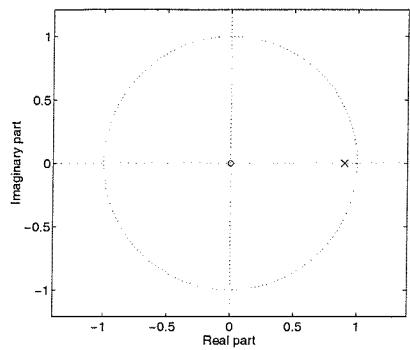
D.3 Simulation results of detection, classification and reconstruction of input signals (refer to Chapter 4 for notations)

D.3.1 Gaussian input case

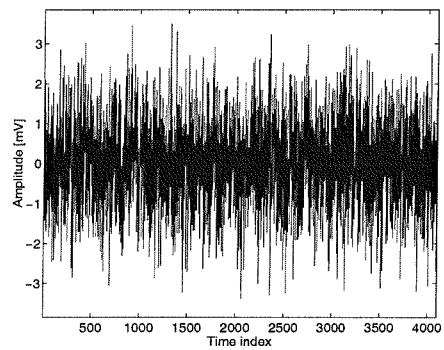


Appendix D, Detection, classification and reconstruction of input signals using HOSVD

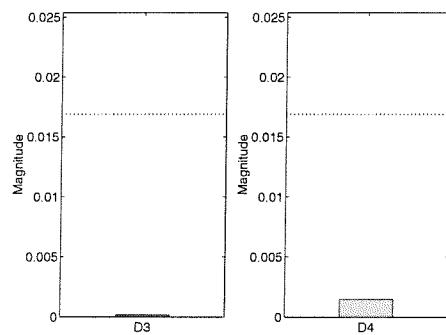
Unknown system # 2



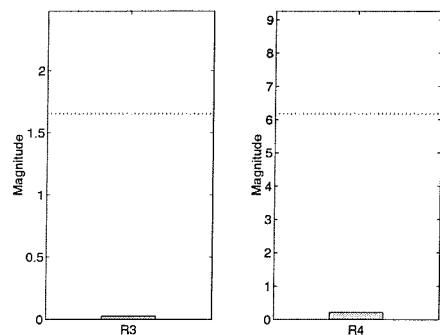
Observed signal



Detection



Reconstruction



Classification	
$D_3 < tr$ $D_4 < tr$	Only Gaussian input signal exists

Restored signal by

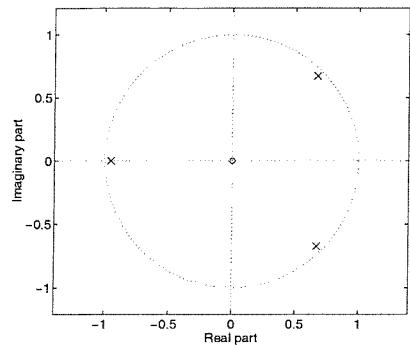
3rd order

NONE

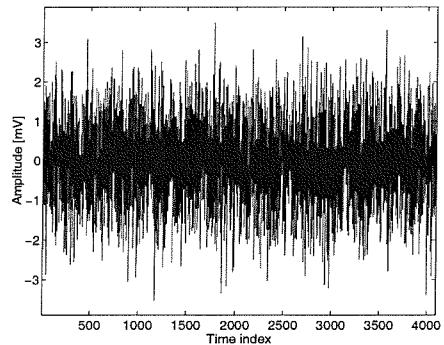
4th order

NONE

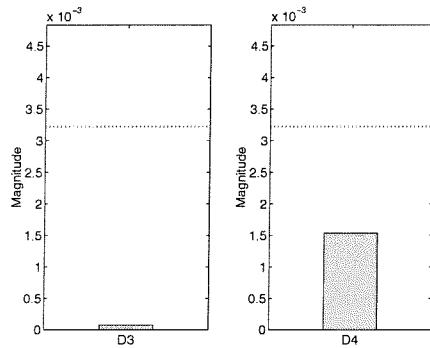
Unknown system # 3



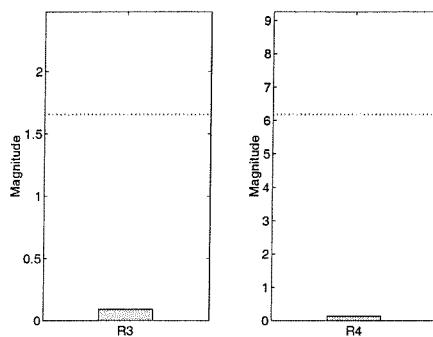
Observed signal



Detection



Reconstruction

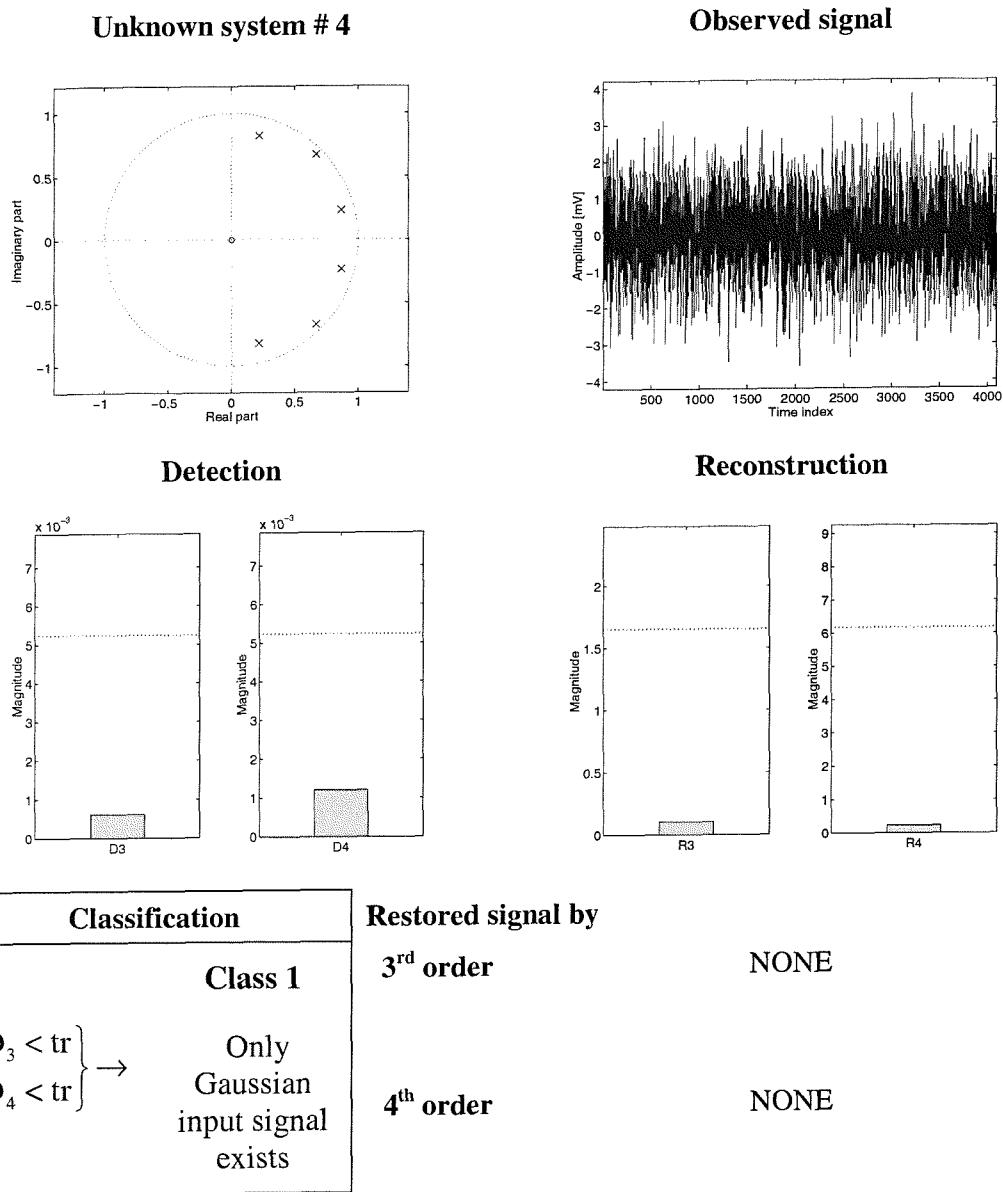


Classification	
	Class 1
$D_3 < \text{tr}$ $D_4 < \text{tr}$	Only Gaussian input signal exists

Restored signal by

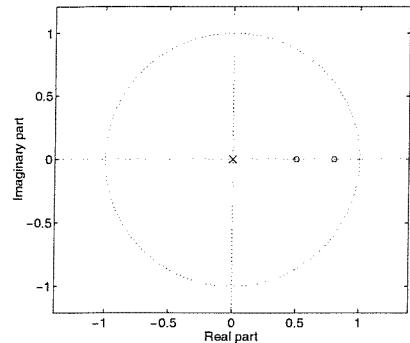
3rd order NONE

4th order NONE

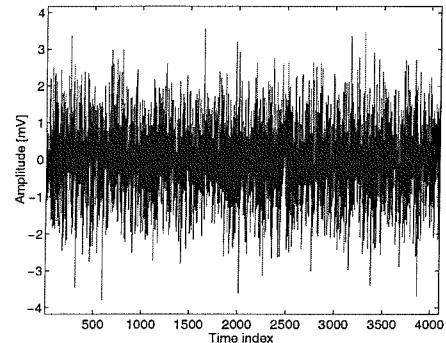


Appendix D, Detection, classification and reconstruction of input signals using HOSVD

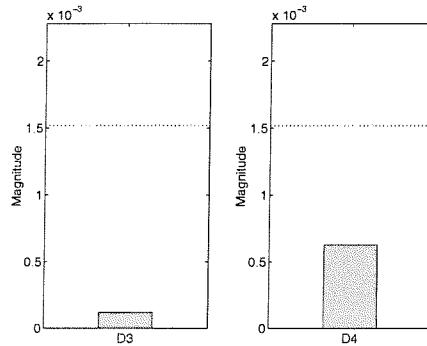
Unknown system # 5



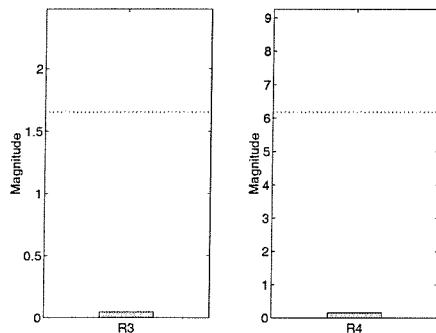
Observed signal



Detection



Reconstruction



Classification	
Class 1	
$D_3 < \text{tr}$ $D_4 < \text{tr}$	Only Gaussian input signal exists

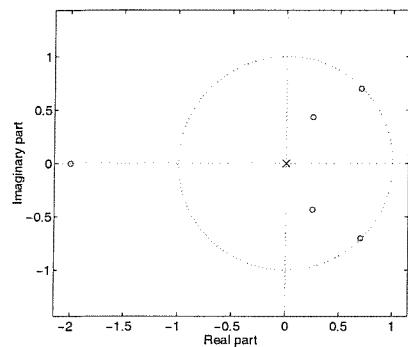
Restored signal by

3rd order **NONE**

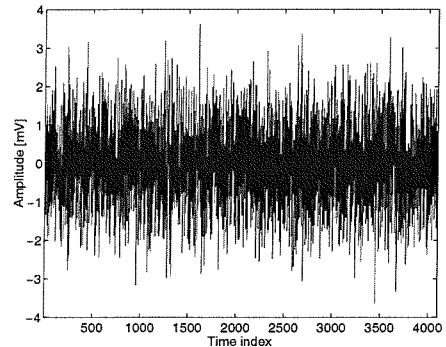
4th order **NONE**

Appendix D, Detection, classification and reconstruction of input signals using HOSVD

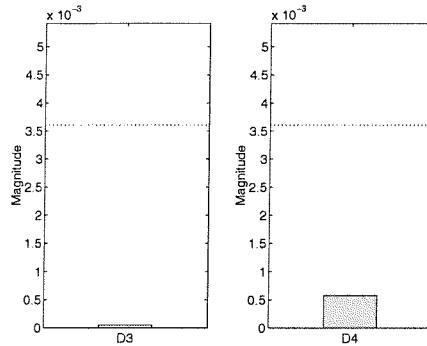
Unknown system # 6



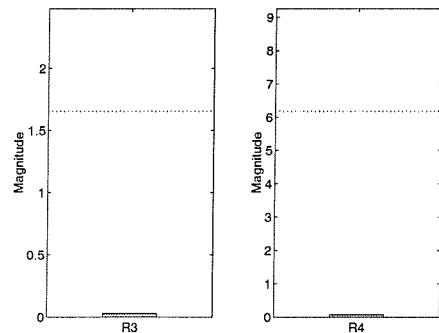
Observed signal



Detection



Reconstruction



Classification

Class 1

Restored signal by

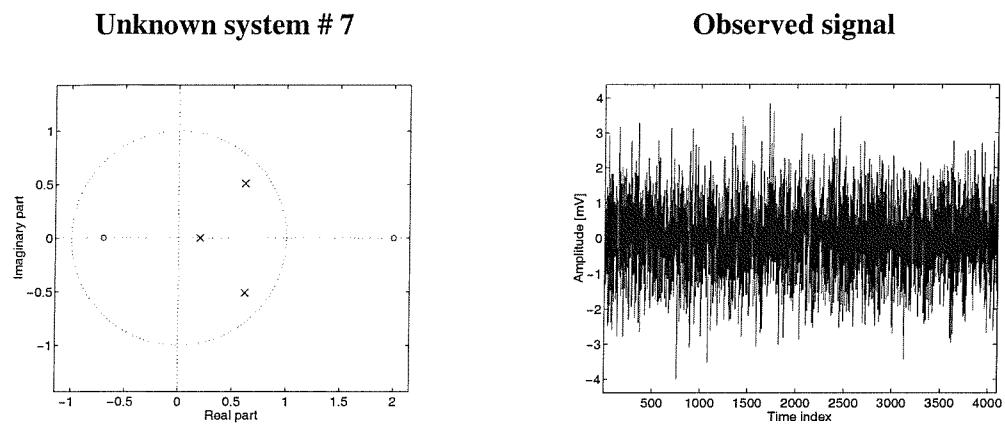
3rd order

NONE

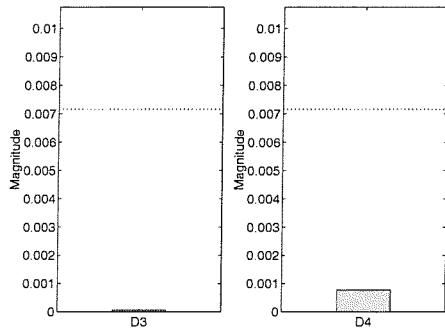
4th order

NONE

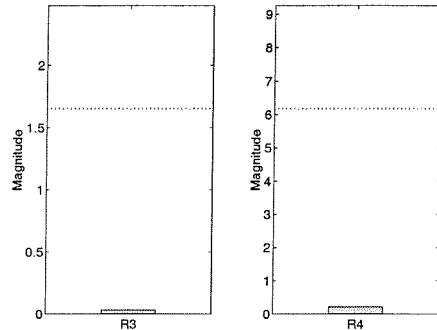
Appendix D, Detection, classification and reconstruction of input signals using HOSVD



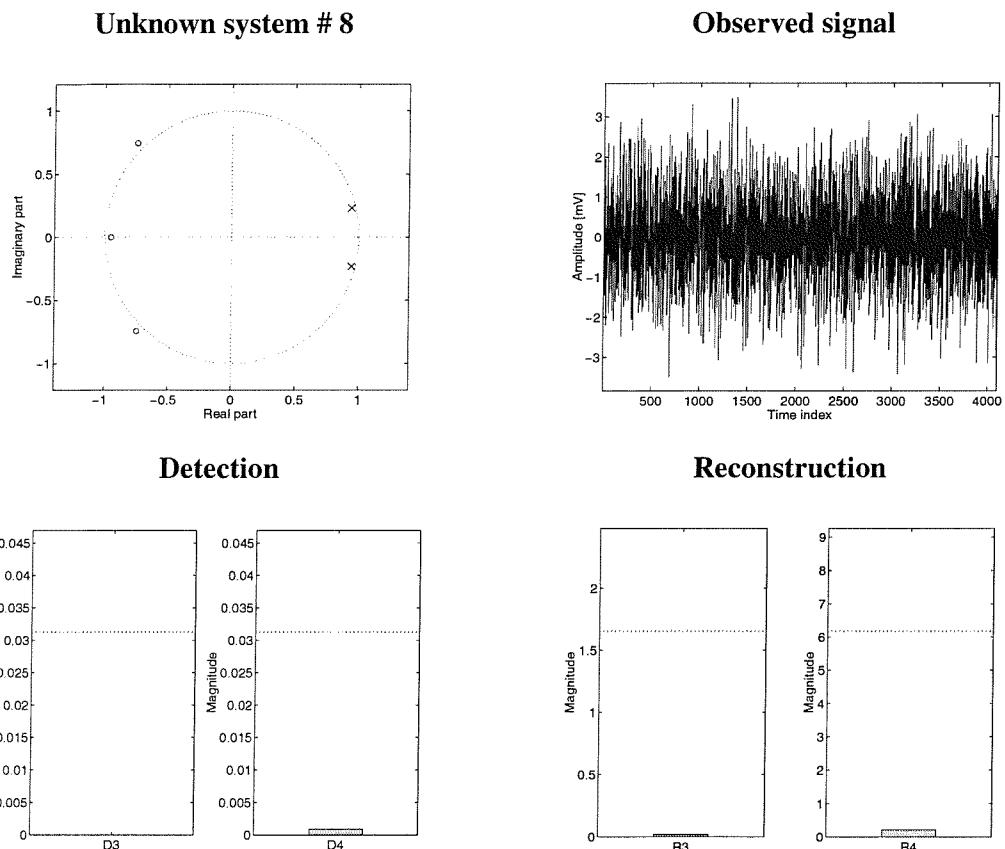
Detection



Reconstruction



Classification	
$D_3 < tr$ $D_4 < tr$	Class 1 Only Gaussian input signal exists



Classification

Class 1

Restored signal by

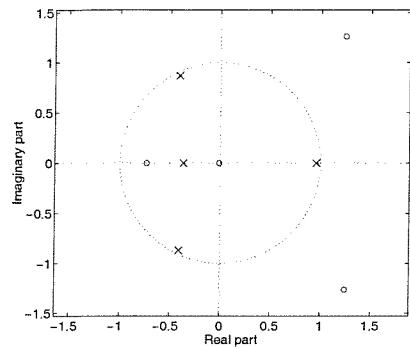
3rd order

NONE

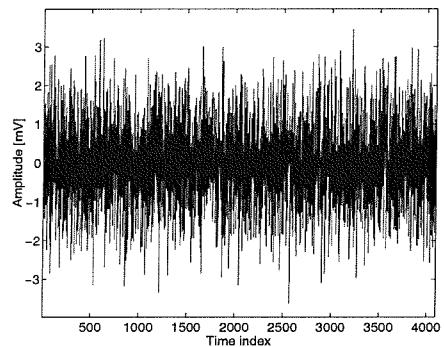
4th order

NONE

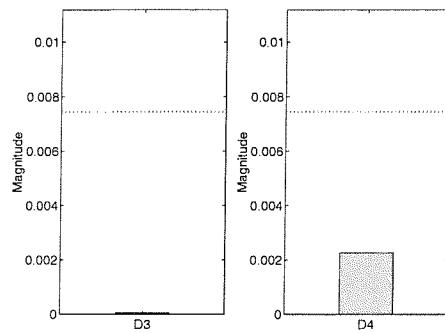
Unknown system # 9



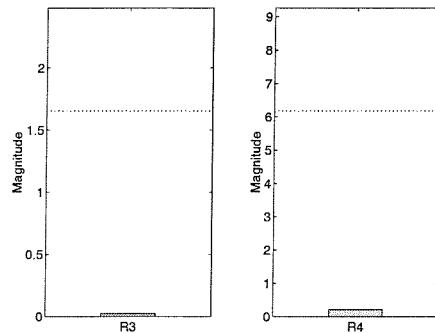
Observed signal



Detection



Reconstruction



Classification	
	Class 1
$D_3 < \text{tr}$ $D_4 < \text{tr}$	Only Gaussian input signal exists

Restored signal by

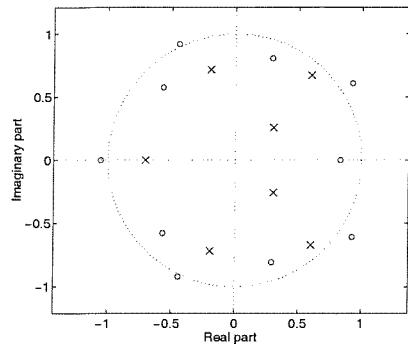
3rd order

NONE

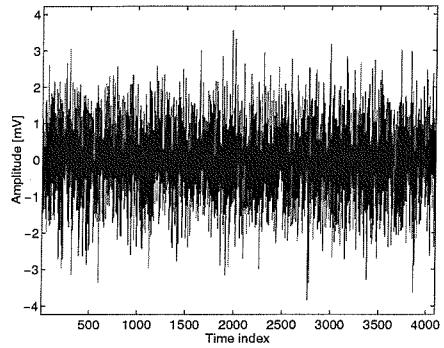
4th order

NONE

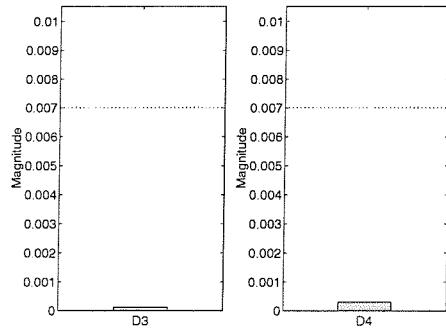
Unknown system # 10



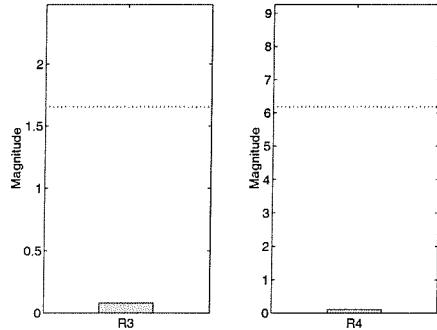
Observed signal



Detection



Reconstruction



Classification	
	Class 1
$D_3 < \text{tr}$ $D_4 < \text{tr}$	Only Gaussian input signal exists

Restored signal by

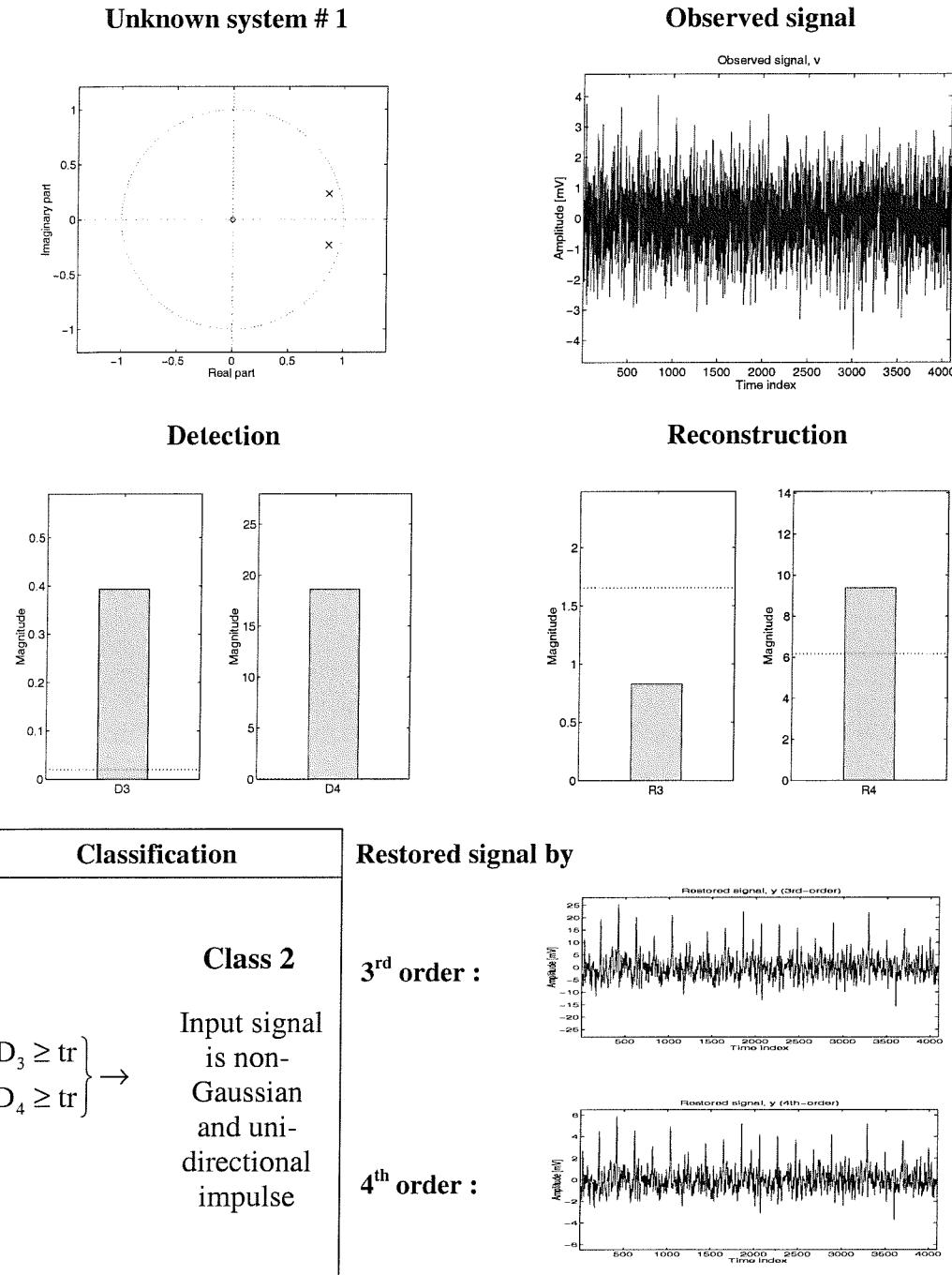
3rd order

NONE

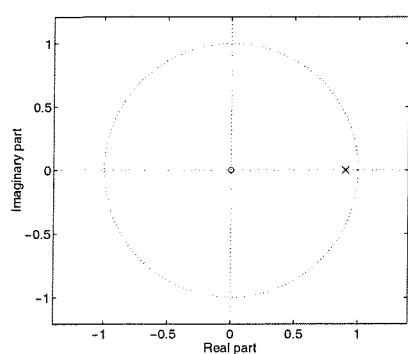
4th order

NONE

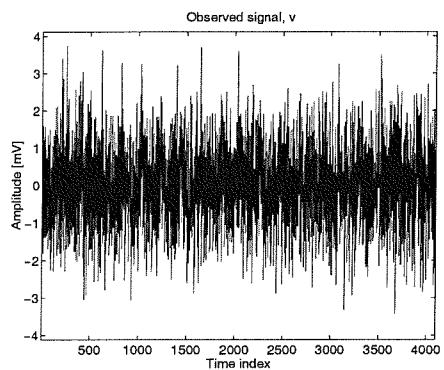
D.3.2 Non-Gaussian input (uni-direction impacting signal) case



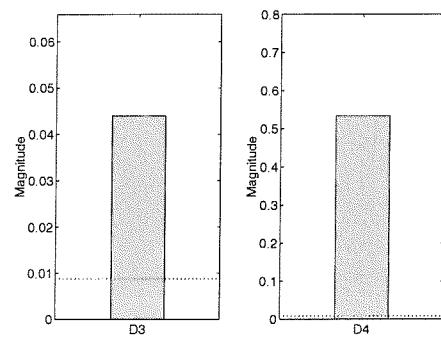
Unknown system # 2



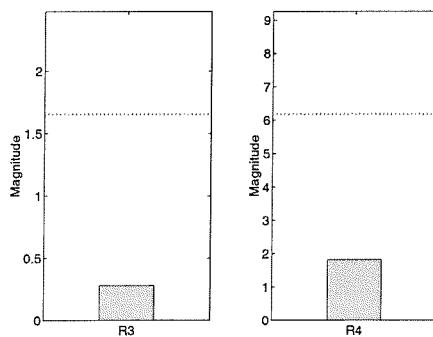
Observed signal



Detection



Reconstruction

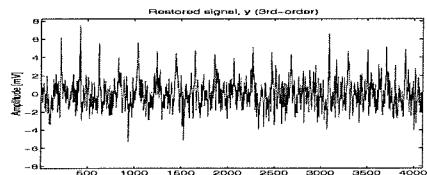


Classification

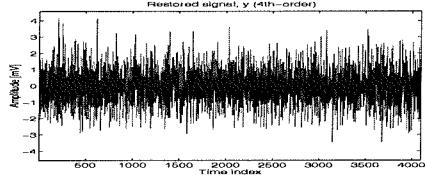
Class 2
 $D_3 \geq \text{tr}$ }
 $D_4 \geq \text{tr}$ }
 Input signal
 is non-
 Gaussian
 and uni-
 directional
 impulse

Restored signal by

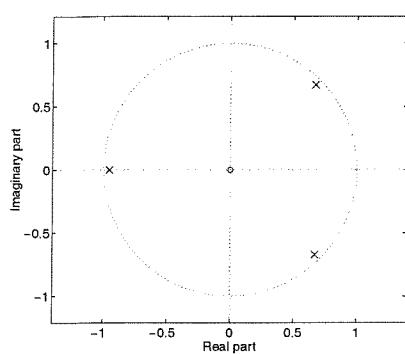
3rd order :



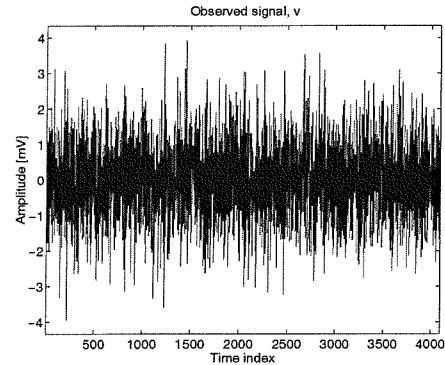
4th order :



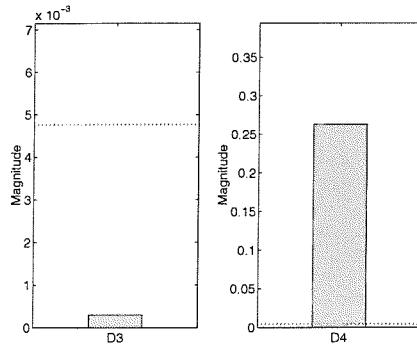
Unknown system # 3



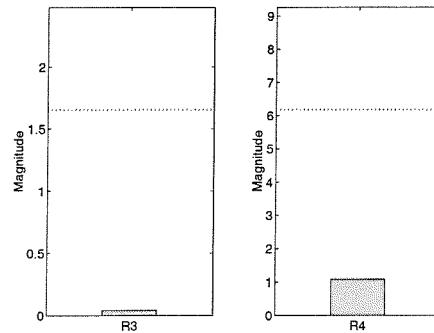
Observed signal



Detection



Reconstruction



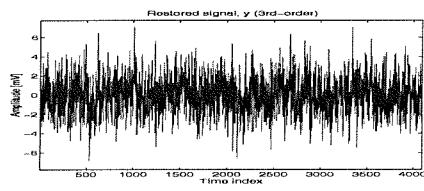
Classification

Class 3

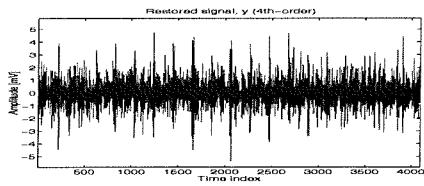
$D_3 < \text{tr}$ } $D_4 \geq \text{tr}$ } \rightarrow Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

Restored signal by

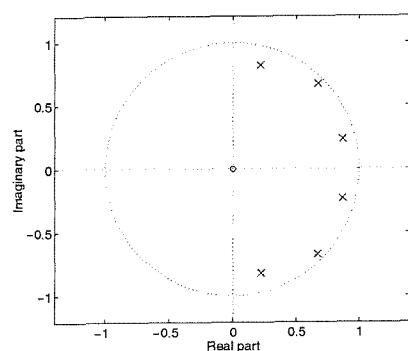
3rd order :



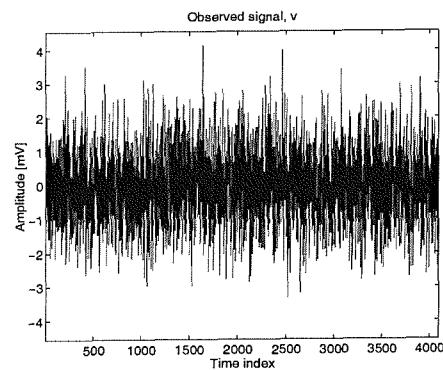
4th order :



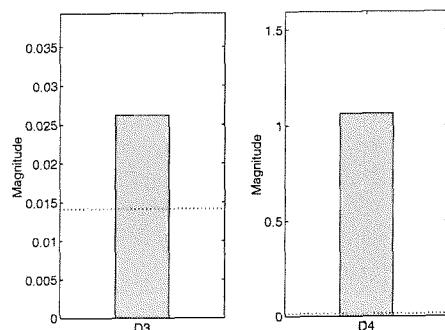
Unknown system # 4



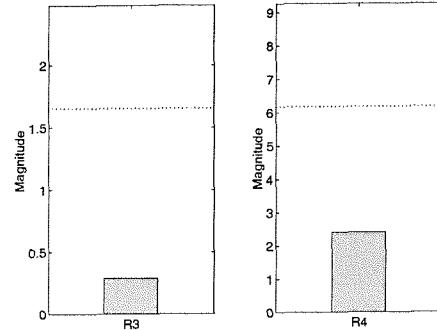
Observed signal



Detection



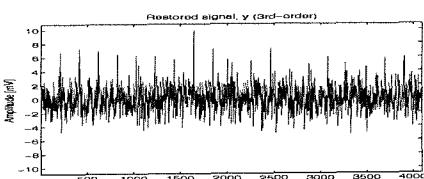
Reconstruction



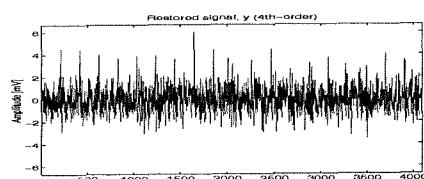
Classification	
Class 2	
$D_3 \geq \text{tr}$ $D_4 \geq \text{tr}$	\rightarrow Input signal is non- Gaussian and uni- directional impulse

Restored signal by

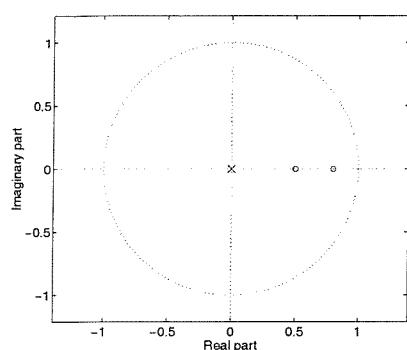
3rd order :



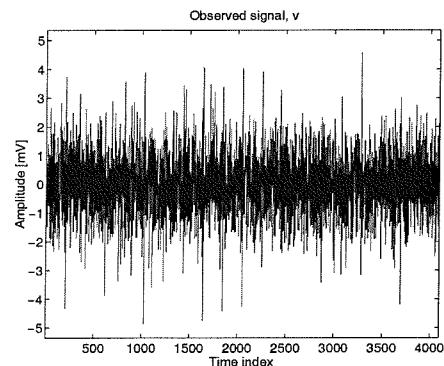
4th order :



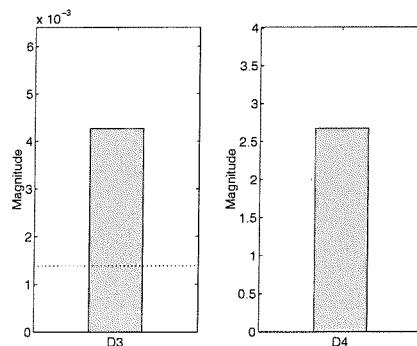
Unknown system # 5



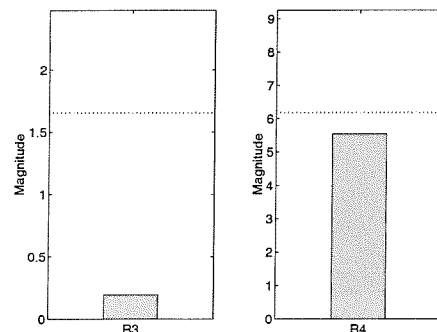
Observed signal



Detection



Reconstruction

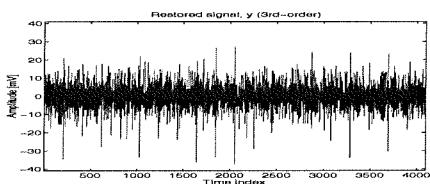


Classification

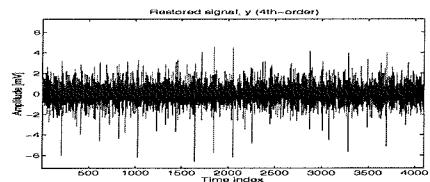
Class 2
 $D_3 \geq \text{tr}$ }
 $D_4 \geq \text{tr}$ }
 Input signal
 is non-
 Gaussian
 and uni-
 directional
 impulse

Restored signal by

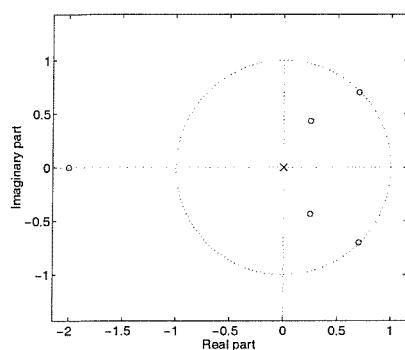
3rd order :



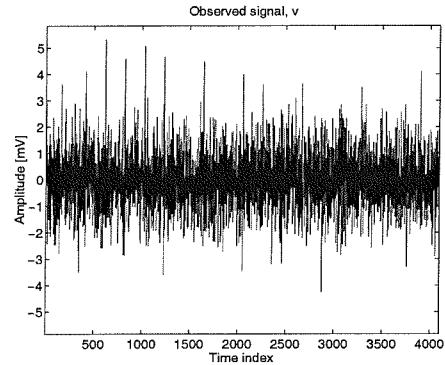
4th order :



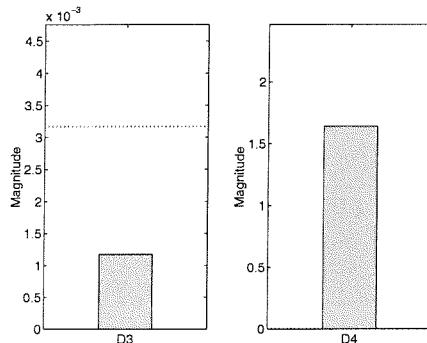
Unknown system # 6



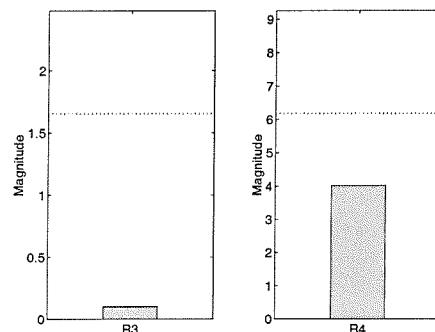
Observed signal



Detection



Reconstruction



Classification

Class 3

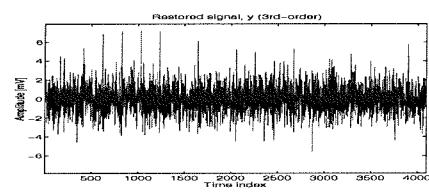
Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

$D_3 < \text{tr}$

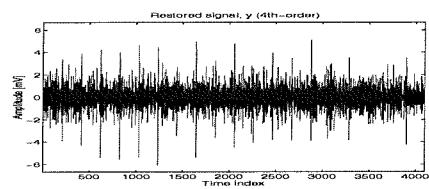
$D_4 \geq \text{tr}$

Restored signal by

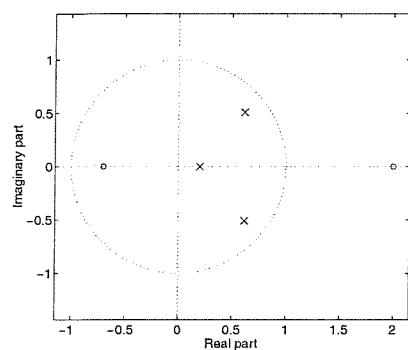
3rd order :



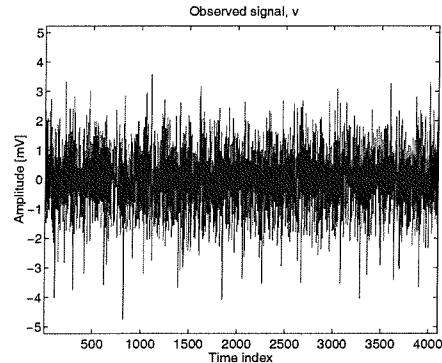
4th order :



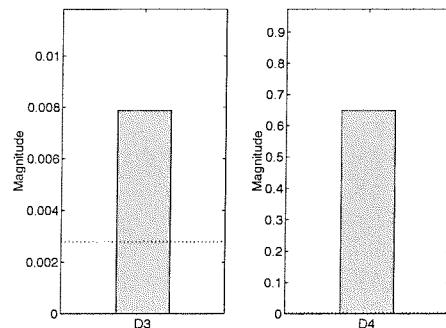
Unknown system # 7



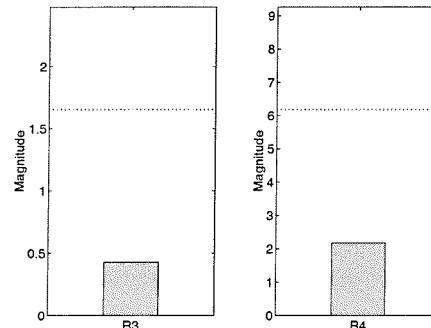
Observed signal



Detection



Reconstruction

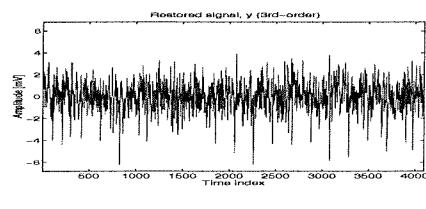


Classification

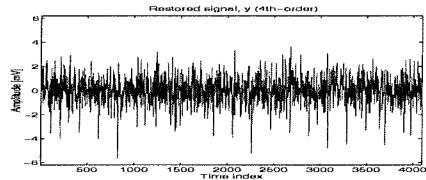
Class 2
 $D_3 \geq \text{tr}$
 $D_4 \geq \text{tr}$ } \rightarrow
 Input signal
 is non-
 Gaussian
 and uni-
 directional
 impulse

Restored signal by

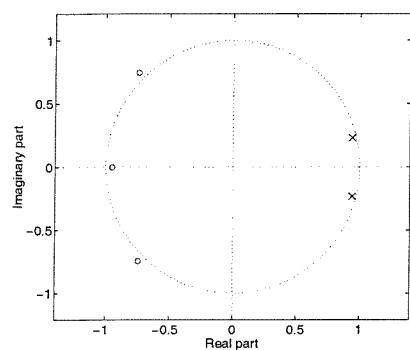
3rd order :



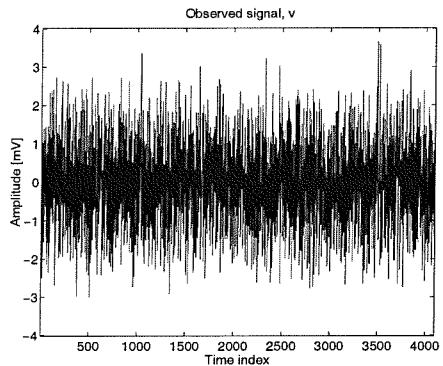
4th order :



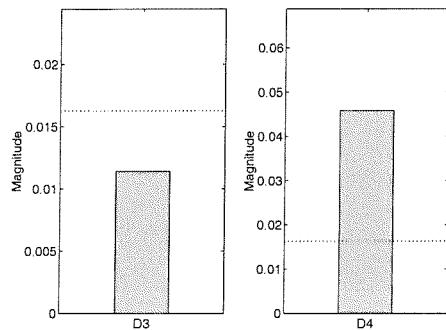
Unknown system # 8



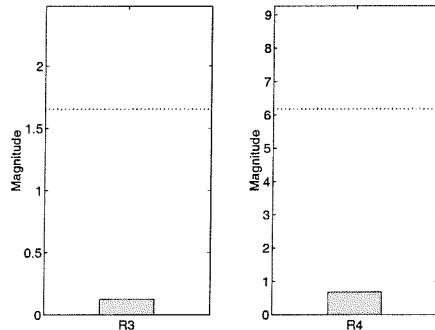
Observed signal



Detection



Reconstruction



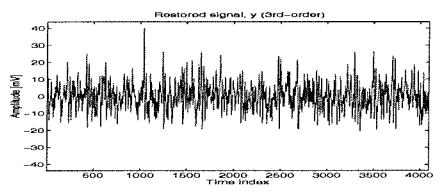
Classification

Class 3

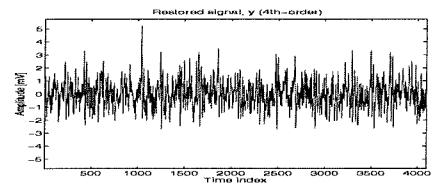
$D_3 < \text{tr}$ } $D_4 \geq \text{tr}$ } \rightarrow
 Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

Restored signal by

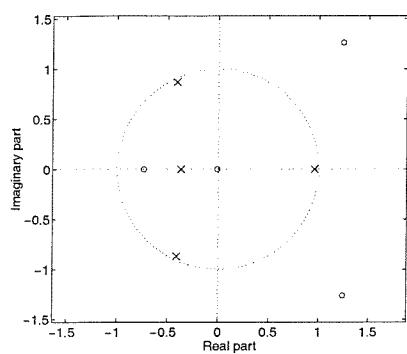
3rd order :



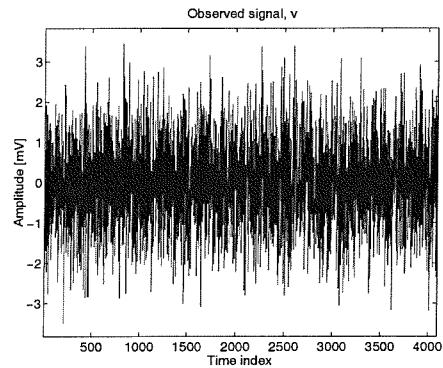
4th order :



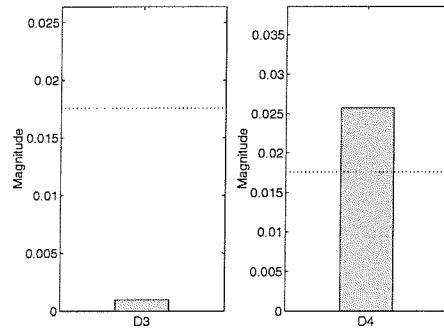
Unknown system # 9



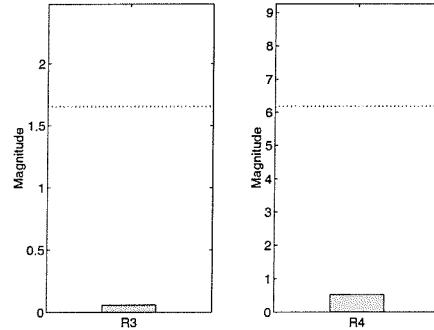
Observed signal



Detection



Reconstruction



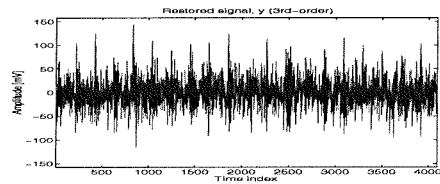
Classification

Class 3

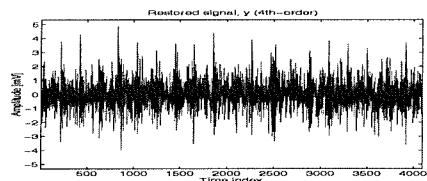
$D_3 < \text{tr}$ } $D_4 \geq \text{tr}$ } \rightarrow Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

Restored signal by

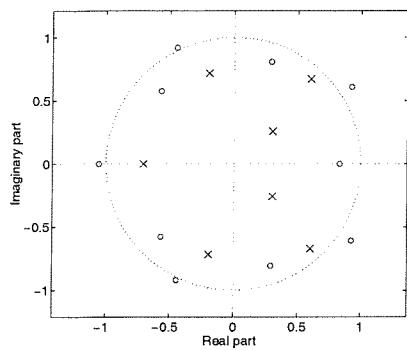
3rd order :



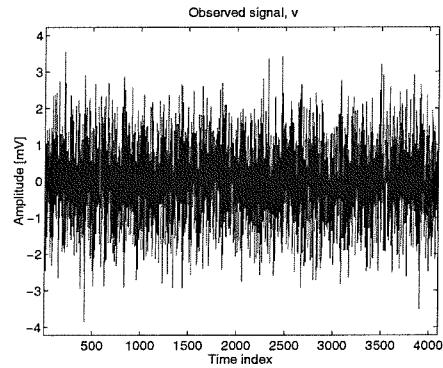
4th order :



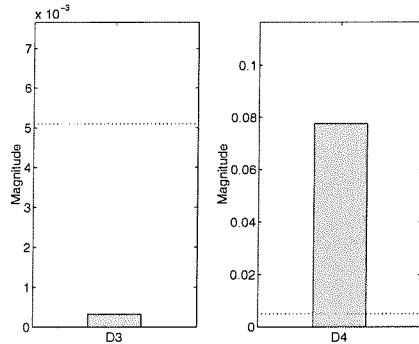
Unknown system # 10



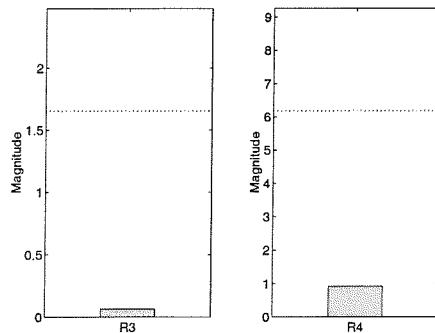
Observed signal



Detection



Reconstruction



Classification

Class 3

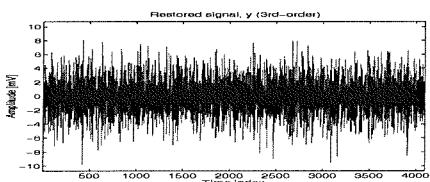
Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

$D_3 < \text{tr}$

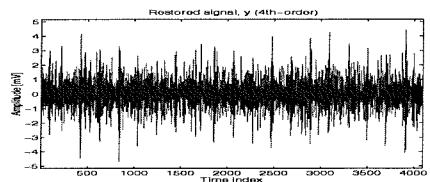
$D_4 \geq \text{tr}$

Restored signal by

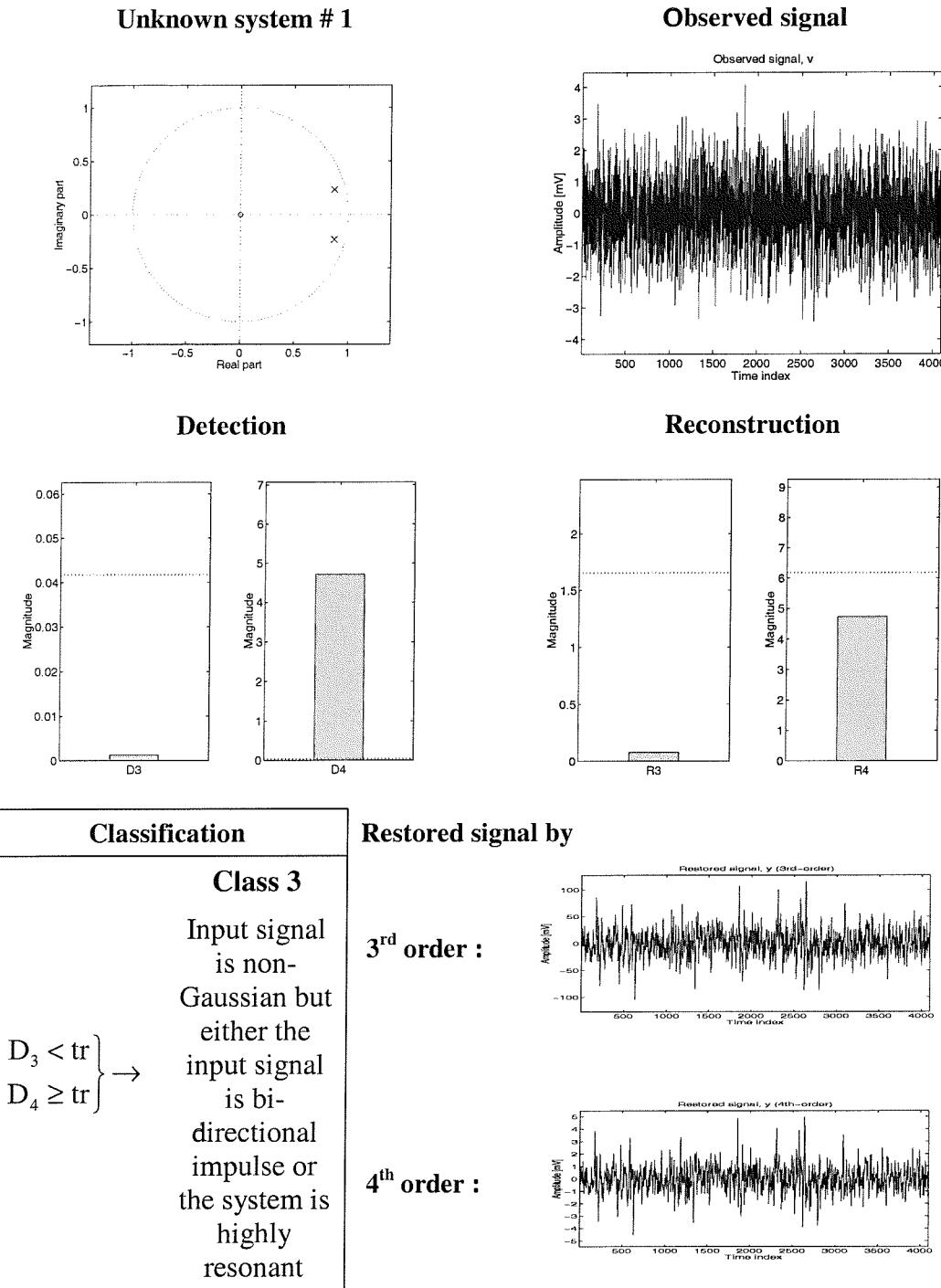
3rd order :



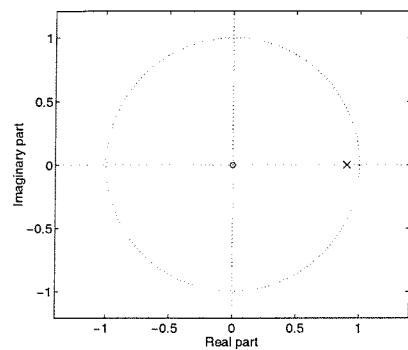
4th order :



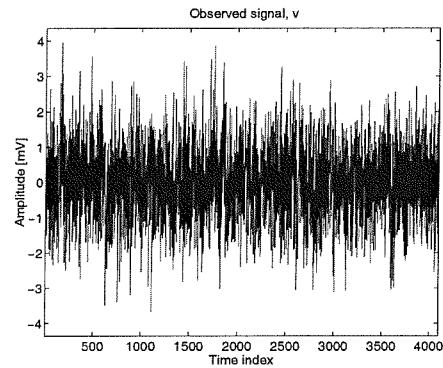
D.3.3 Non-Gaussian input (bi-direction impacting signal) case



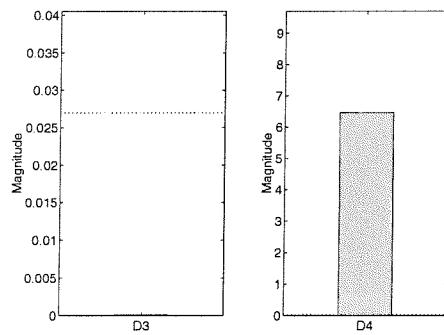
Unknown system # 2



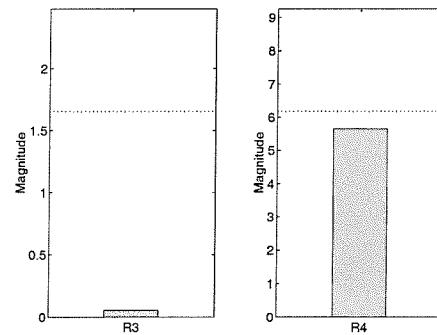
Observed signal



Detection



Reconstruction



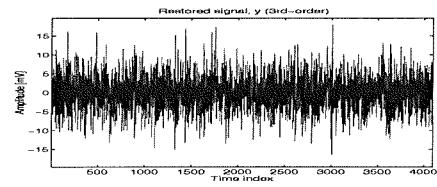
Classification

Class 3
 Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

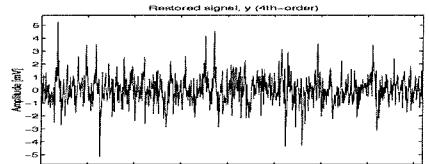
$D_3 < \text{tr}$ } $D_4 \geq \text{tr}$ } \rightarrow

Restored signal by

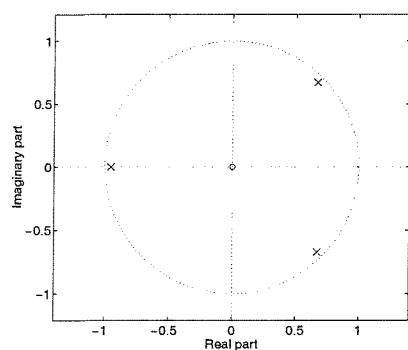
3rd order :



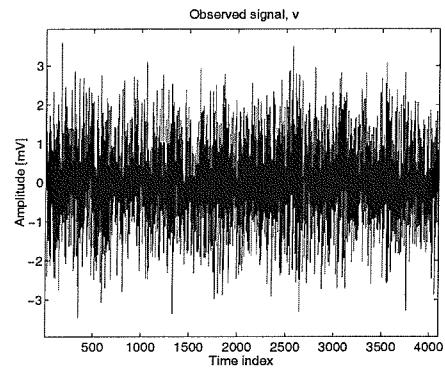
4th order :



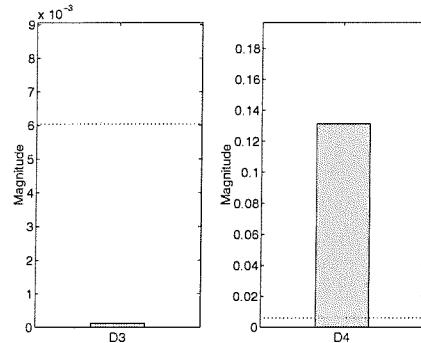
Unknown system # 3



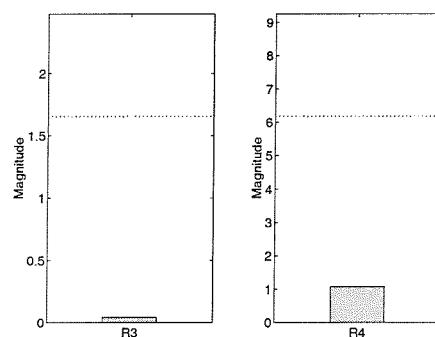
Observed signal



Detection



Reconstruction



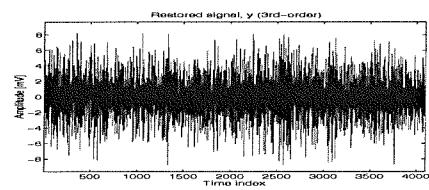
Classification

Class 3

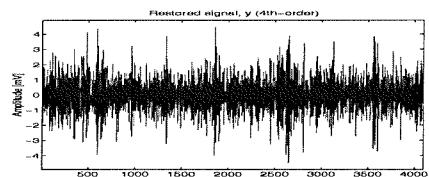
$D_3 < \text{tr}$ } \rightarrow Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

Restored signal by

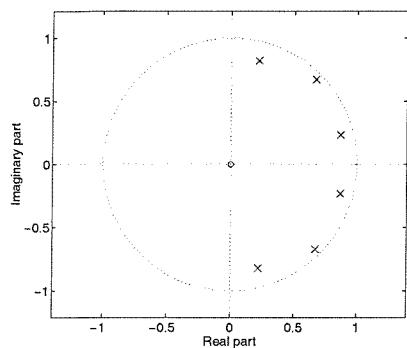
3rd order :



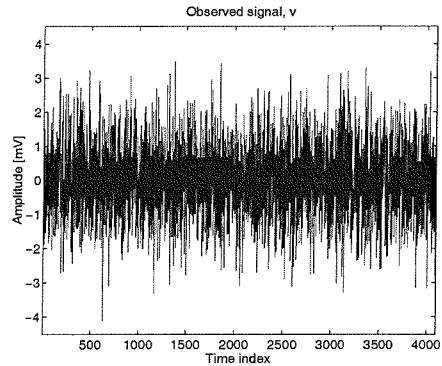
4th order :



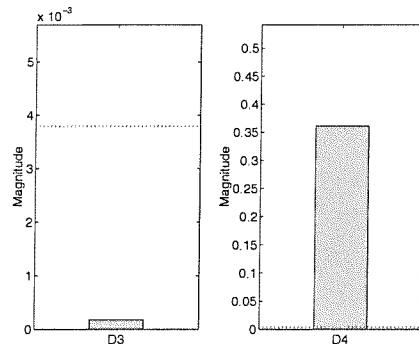
Unknown system # 4



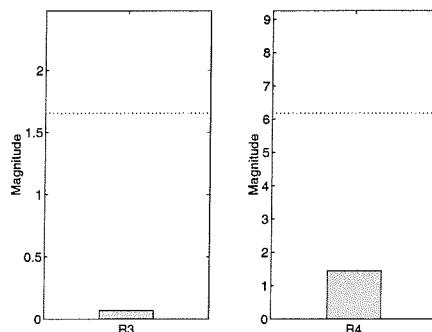
Observed signal



Detection



Reconstruction



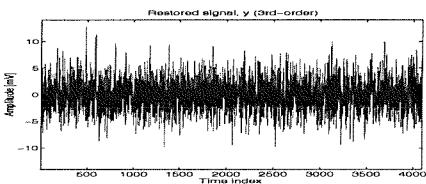
Classification

Class 3
 Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

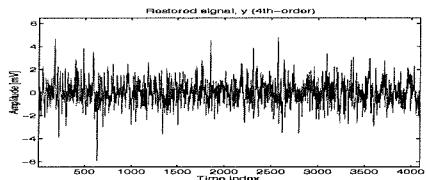
$D_3 < \text{tr}$ } $D_4 \geq \text{tr}$ } \rightarrow

Restored signal by

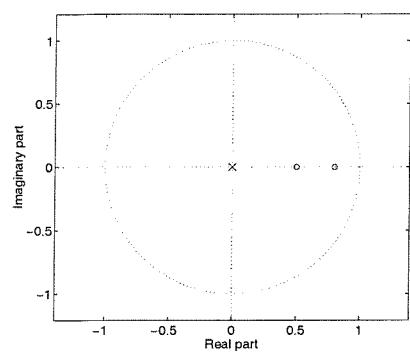
3rd order :



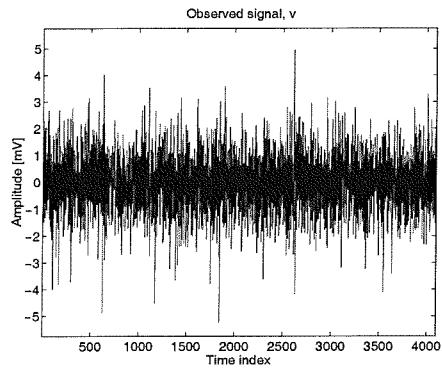
4th order :



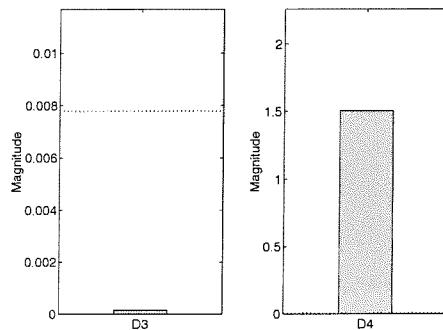
Unknown system # 5



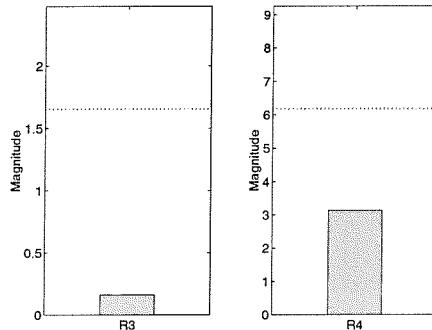
Observed signal



Detection



Reconstruction



Classification

Class 3

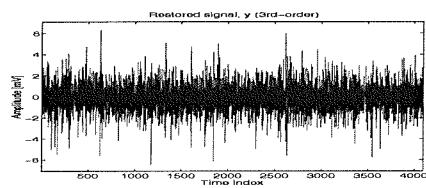
Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

$D_3 < \text{tr}$

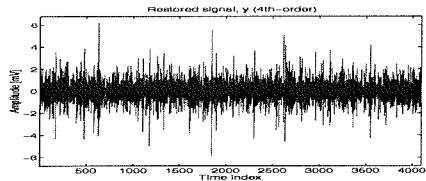
$D_4 \geq \text{tr}$

Restored signal by

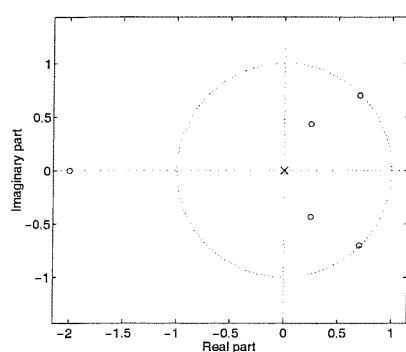
3rd order :



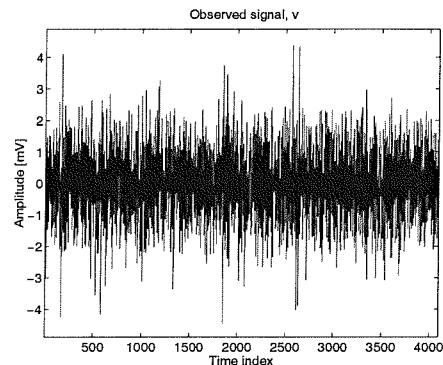
4th order :



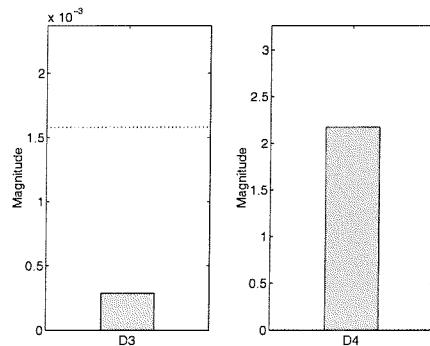
Unknown system # 6



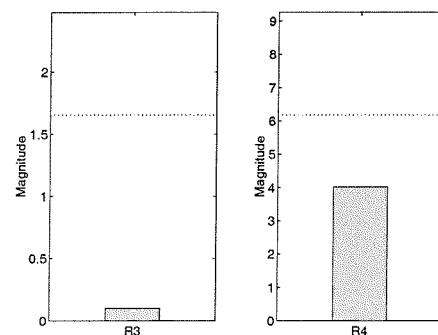
Observed signal



Detection



Reconstruction



Classification

Class 3

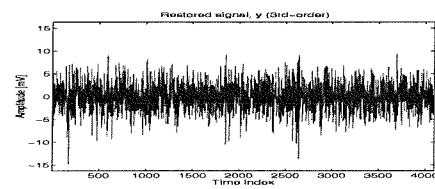
Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

$D_3 < \text{tr}$

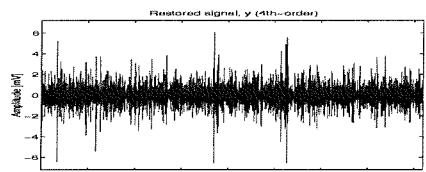
$D_4 \geq \text{tr}$

Restored signal by

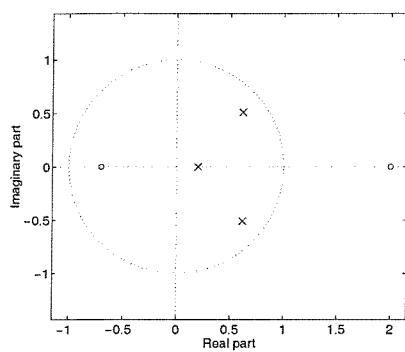
3rd order :



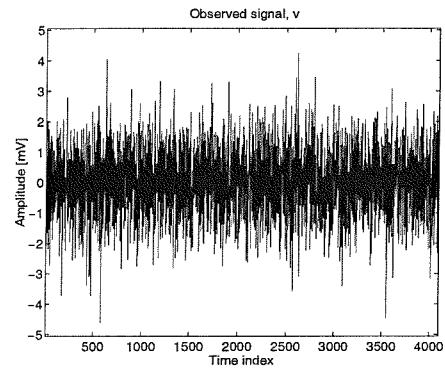
4th order :



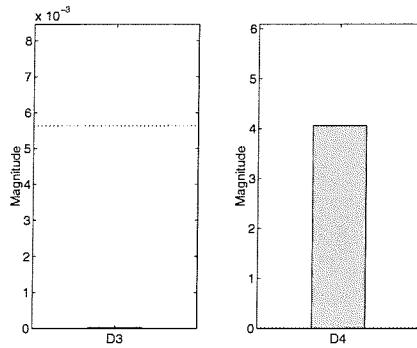
Unknown system # 7



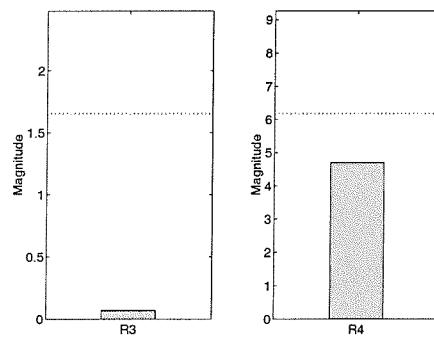
Observed signal



Detection



Reconstruction



Classification

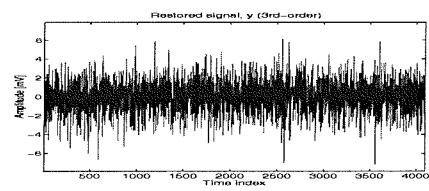
Class 3

Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

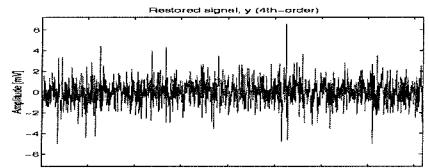
$$\left. \begin{array}{l} D_3 < \text{tr} \\ D_4 \geq \text{tr} \end{array} \right\} \rightarrow$$

Restored signal by

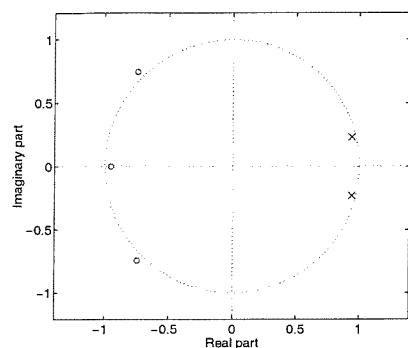
3rd order :



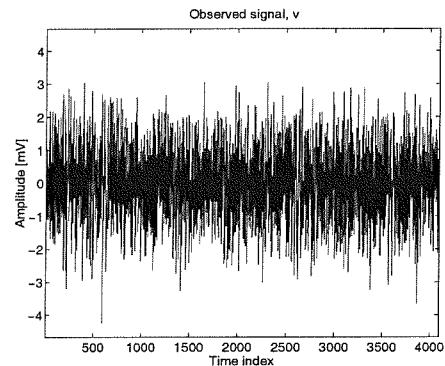
4th order :



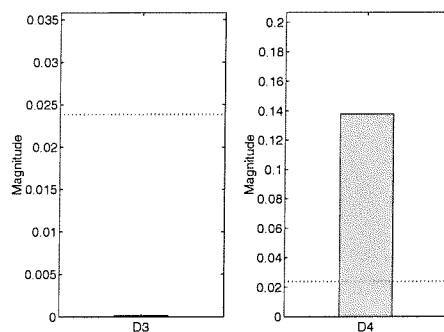
Unknown system # 8



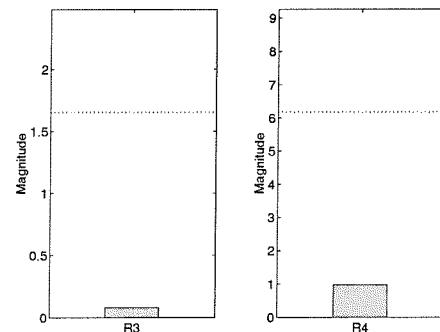
Observed signal



Detection



Reconstruction



Classification

Class 3

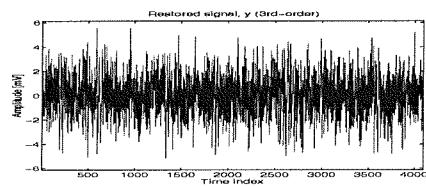
Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

$D_3 < \text{tr}$

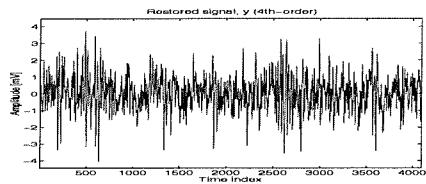
$D_4 \geq \text{tr}$

Restored signal by

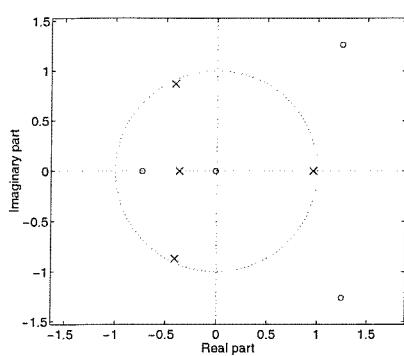
3rd order :



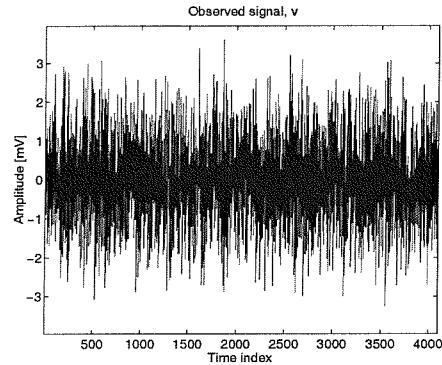
4th order :



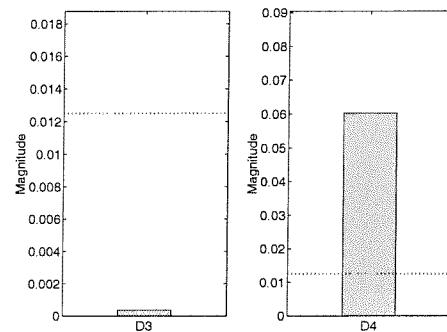
Unknown system # 9



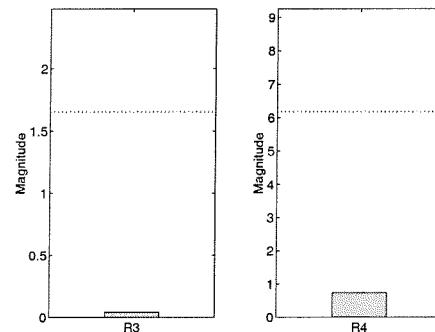
Observed signal



Detection



Reconstruction



Classification

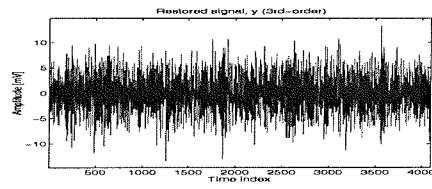
Class 3

Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

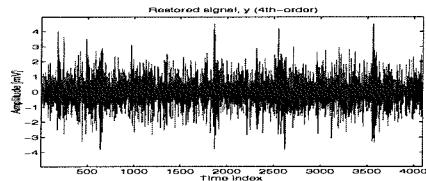
$$\left. \begin{array}{l} D_3 < \text{tr} \\ D_4 \geq \text{tr} \end{array} \right\} \rightarrow$$

Restored signal by

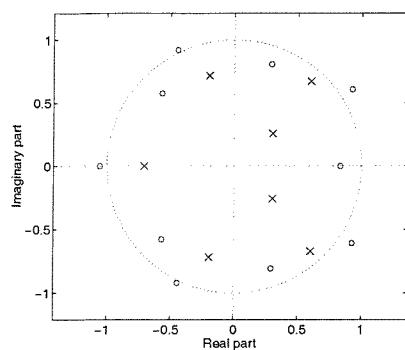
3rd order :



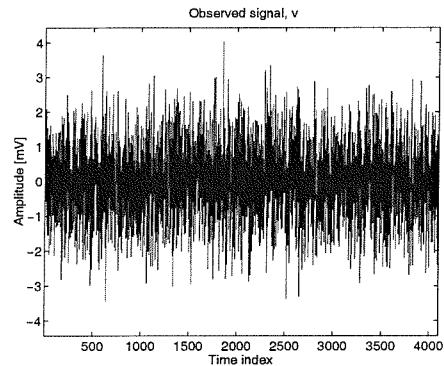
4th order :



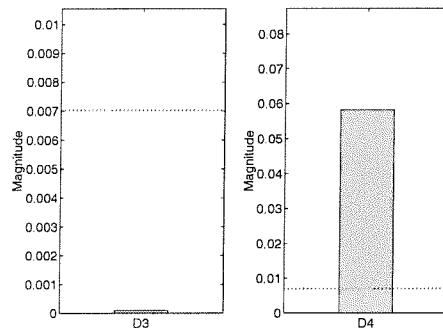
Unknown system # 10



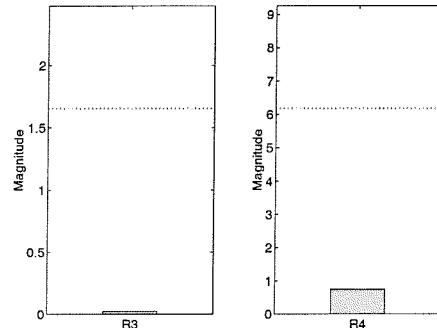
Observed signal



Detection



Reconstruction



Classification

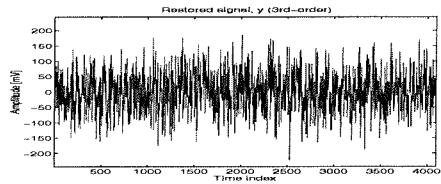
Class 3

Input signal is non-Gaussian but either the input signal is bi-directional impulse or the system is highly resonant

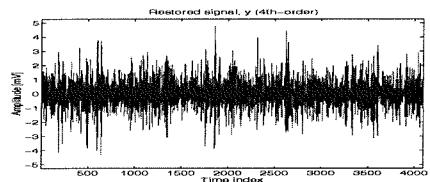
$D_3 < \text{tr}$ } $D_4 \geq \text{tr}$ } \rightarrow

Restored signal by

3rd order :



4th order :



Appendix E

Determination of the inverse filter length (simulation results for Chapter 5)

This Appendix provides details in connection with the task of inverse filter length determination carried out in Chapter 5, section 5.4. Using various length of impulse response and degree of noise interference simulations, the effect of an inverse filter length is considered by comparing signals reconstructed by different length filters.

E.1 Preliminaries

This subsection examines the effect of the length of unknown system's impulse response and noise interference on the performance of signal reconstruction when the length of the inverse filter is fixed. Thus, the parameters which play dominant roles in the determination of the inverse filter length are explored.

E.1.1 Impacting signals used in simulation

Two different impacting signal has been used in this simulation. One is an impacting signal with regular uni-directional impulse sequence and the other is an irregular bi-directional impulse signal both of which have large high order statistical values (e.g. skewness and kurtosis). The two different types of the impacting signals are shown in the following figure.

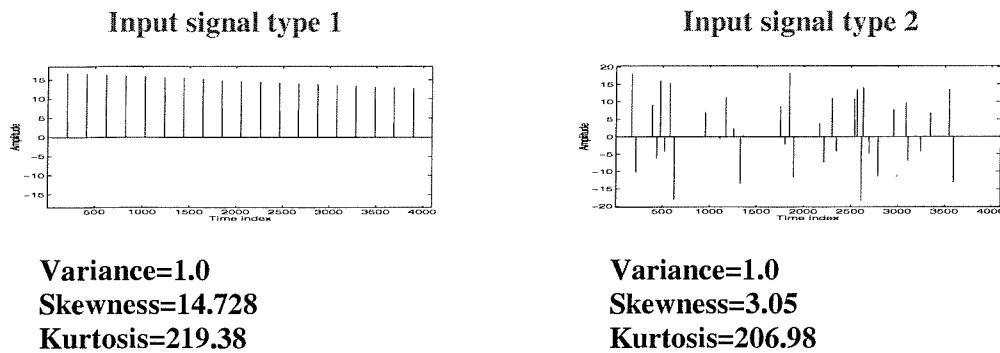


Figure E.1.1 Input signals used for simulation

E.1.2 Model of unknown system for simulation

To generate the variable length of the impulse response, an AR(2) system (an unknown system) is selected whose pole positions are at with the radius (r_i) and angle (θ_i).

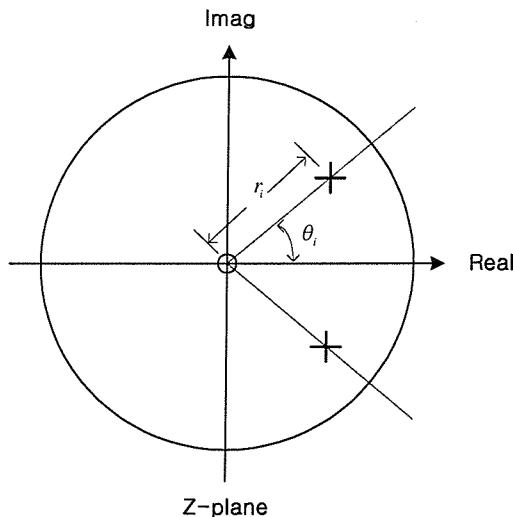


Figure E.1.2 The AR(2) system generating the variable length of impulse response.
 'o' : zero of the system and 'x': poles of the system.

For this AR(2) system, its impulse response denoted as $h(k)$ can have different values and length by changing the r_i and θ_i . The following figure shows one example of the impulse response with pole position at $r_i = 0.9$ and $\theta_i = \pi/4$.

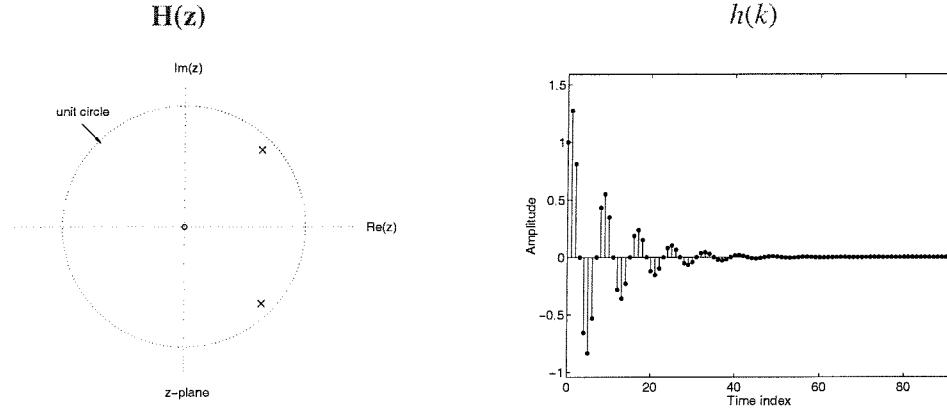


Figure E.1.3 Pole-zero maps and impulse responses of selected system

Even though we select the system as AR, the length of the impulse response (denoted by IR) of the system is assumed to be a finite length truncating the impulse response function at the point when its magnitude is less than 0.001 (this is done for only for simulation). Thus, for the system shown in Figure E.1.3, the length of IR becomes 74 (i.e., equivalent length of the system $L=74$).

Now, the output of this system can be expressed as convolution process

$$z(n) = \sum_{k=0}^{L-1} h(k)x(n-k) \quad (\text{E.1.1})$$

$z(n)$ is thought as ‘noise free’ output of the unknown system $H(z)$.

E.1.3 Noise signal generation

The Signal to Noise ratio (S/N ratio) of the (unknown) additive Gaussian noise $w(n)$ is defined as

$$\text{SNR} = 10 \log \frac{\sigma_z^2}{\sigma_w^2}, \text{ dB} \quad (\text{E.1.2})$$

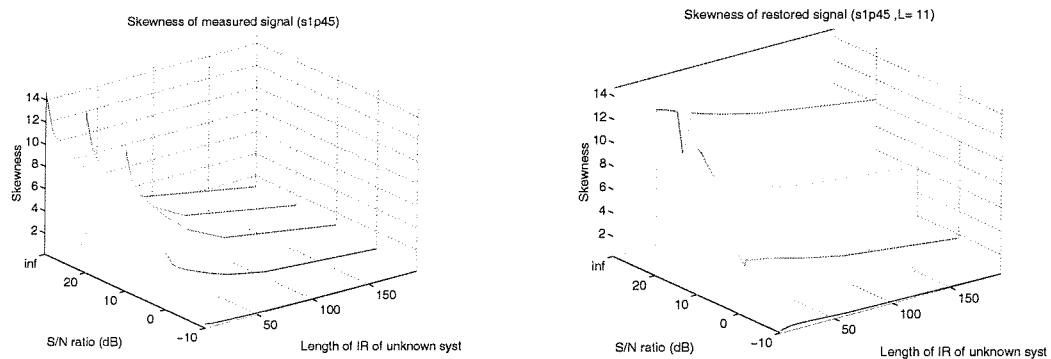
where σ_z^2 and σ_w^2 represent the variance of the noise free output of the system and additive Gaussian noise signal, respectively. The observed signal can be obtained by adding the noise free output signal and noise signal as

$$v(n) = \sum_{k=0}^{L-1} h(k)x(n-k) + w(n) \quad (\text{E.1.3})$$

E.2 Simulation 1: Effects of the length of unknown system's impulse response and noise interference on the inverse filter

The change of higher-order statistical properties for each length of unknown system's impulse response and the degree of Gaussian noise interference for the measured and the restored signal with fixed length of the inverse filter has been monitored.

Third-order case, $\theta_i = \pi / 4$



Fourth-order case, $\theta_i = \pi / 4$

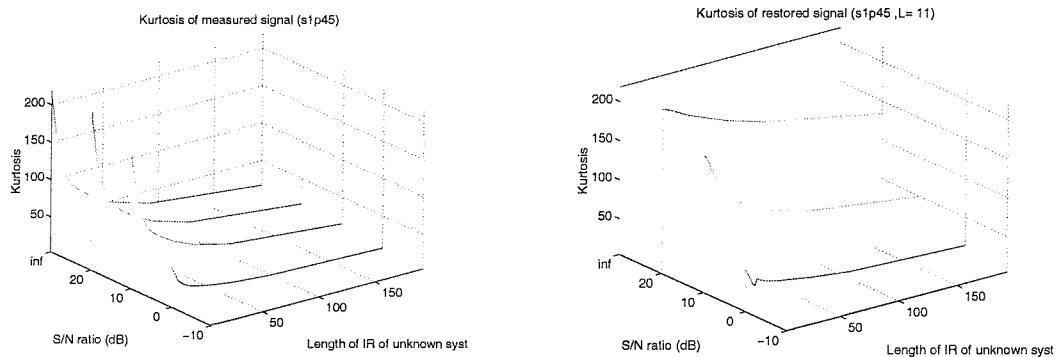
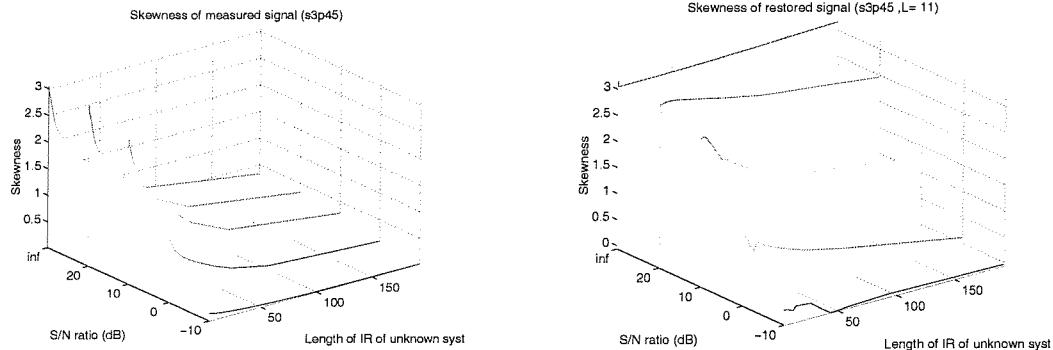


Figure E.2.1 Skewness and kurtosis changes in measured and restored signal (for type 1 input signal)

Third-order case, $\theta_i = \pi / 4$



Fourth-order case, $\theta_i = \pi / 4$

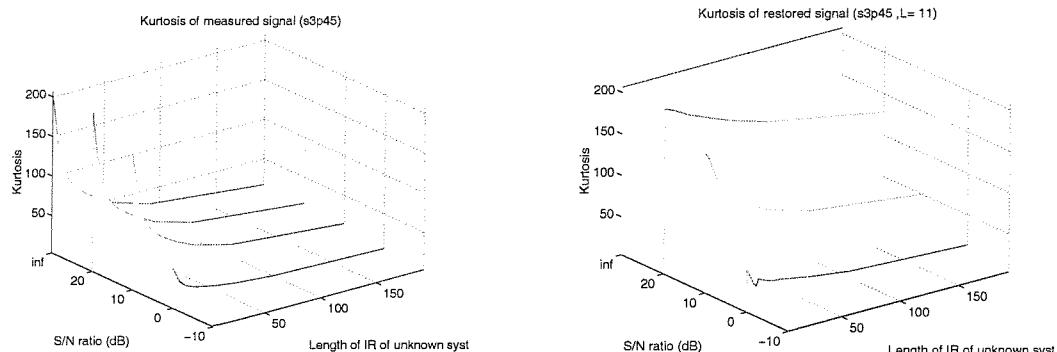


Figure E.2.2 Skewness and kurtosis changes in measured and restored signal (for type 2 input signal)

Figure E.2.1 (type 1 input signal case) and Figure E.2.2 (type 2 input signal case) demonstrate a three dimensional view of the third- and fourth-order statistical values (i.e. skewness and kurtosis) of the measured (left column of the figures) and restored signal (right column of the figures) with different length of the impulse responses (r_i varying from 0 to 0.95) and noise interference (S/N ratio varying from -10 to ∞ dB). Each row of figures is divided into the plots of higher order statistical values of measured and restored signals for unknown system's θ_i is varying from $\pi/12$ to $5\pi/12$. As shown in Figure E.1.3, by varying the pole position of AR(2) system, the different impulse response sequence is acquired and its corresponding length is marked in the right axis of each graph. Also, the different Signal to Noise ratio (SN ratio) interfering on the measured signal is plotted in the left axis of the graph. Thus, starting from those varying parameters,

the higher order statistical values of the measured and restored signal are plotted in the vertical axis of each graph.

The signal reconstruction has been achieved using MA(11) inverse system (i.e. FIR, $L=11$) with a normalised objective function in this simulation.

E.3 Simulation 2: Effects of the length of inverse filter and noise interference on the performance of impacting signal reconstruction for system with a long impulse response ($r_i = 0.9$ and $\theta_i = \pi/4$ as shown in the second row of Figure E.2.1)

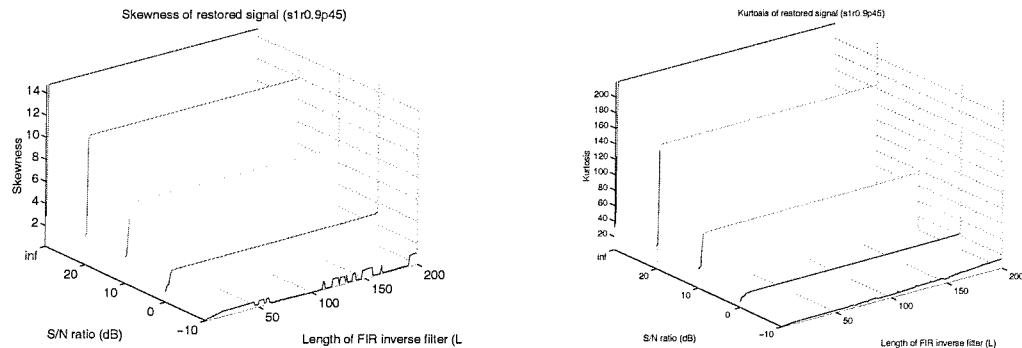
In previous section, we have observed the change of skewness and kurtosis of the restored signal from various measured signals. The signal restoration has been only achieved through a FIR inverse system in which L is fixed as 11.

In conjunction with the previous simulation, the length L of the FIR inverse filter is changed to trace the higher order statistical values in each reconstructed signal.

Input signal type 1 ($r_i = 0.9$, $\theta_i = \pi/4$)

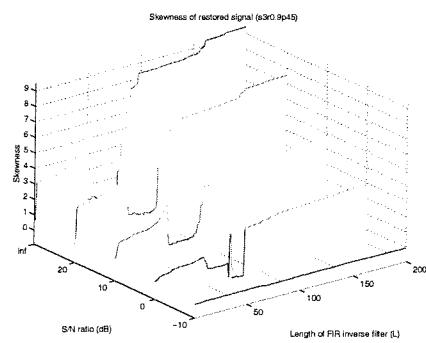
Third-order case,

Fourth-order case



Input signal type 2 ($r_i = 0.9$, $\theta_i = \pi/4$)

Third-order case



Fourth-order case

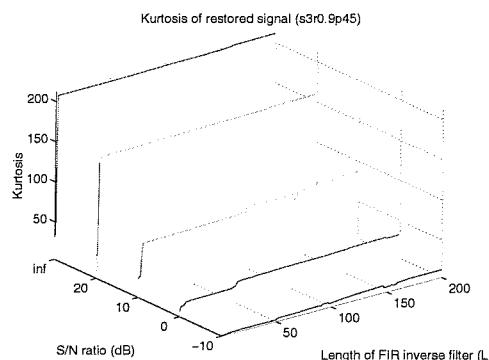


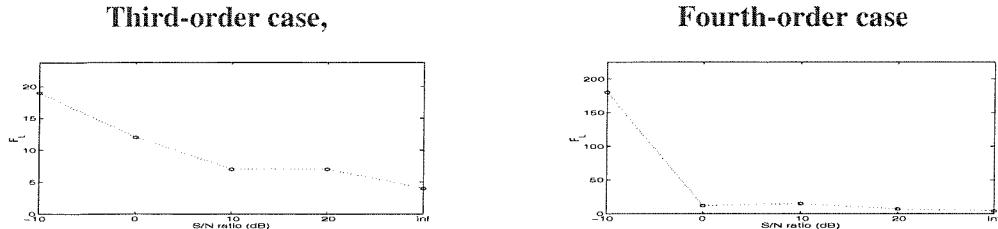
Figure E.3.1 Skewness and kurtosis changes in restored signals with different length of FIR inverse filter

In Figure E.3.1, another three dimensional view of the skewness (left column of the figures) and kurtosis (right column of the figures) of the signal restored from the inverse filter of different length (L varying from 1 to 200) is demonstrated. The selected AR(2) system is selected as r_i and θ_i to be 0.9 and $\pi/4$, respectively (shown in the second row of Figure E.2.1). Again, two different types of impacting signal used in the previous simulation (shown in Figure E.1.1) have been used as inputs to the system and plotted in separate rows.

The length of the inverse filter is marked in the right axis of each graph and the Signal to Noise ratio varying from -10 to ∞ dB interfering on the output of the system is plotted in the left axis of the graph. Thus, with those varying parameters (L and SN ratio), the higher order statistical values of the restored signals are plotted in the vertical axis of each graph.

From the results of restored signal's higher-order statistical values, the minimum upper bound of inverse filter length is calculated for each condition of the measured signal (different S/N ratio).

Input signal type 1 ($r_i = 0.9$, $\theta_i = \pi/4$)



Input signal type 2 ($r_i = 0.9$, $\theta_i = \pi/4$)

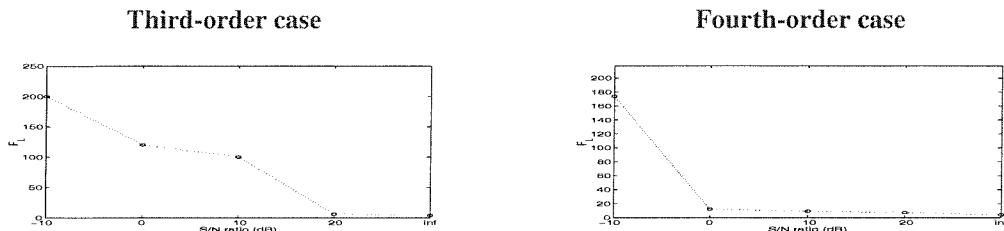


Figure E.3.2 Selection of minimal inverse filter length for different measured signal

The points marked as ‘o’ in each graph indicate the lower bound of the length of FIR inverse filter from which the reconstructed signal can be considered as to be satisfactory result (for the literal convenience, the terminology of ‘minimal’ length of inverse filter is reserved to this length).

For example, let us compare the result of restored signals with minimal length of FIR inverse filter L_{\min} and with arbitrarily selected length L_2 ($L_2 \gg L_{\min}$) using the third-order deconvolution filter, type 1 input and measured signal having 10 dB S/N ratio.

The shape of each restored signal and its statistical values are compared in the following figure.

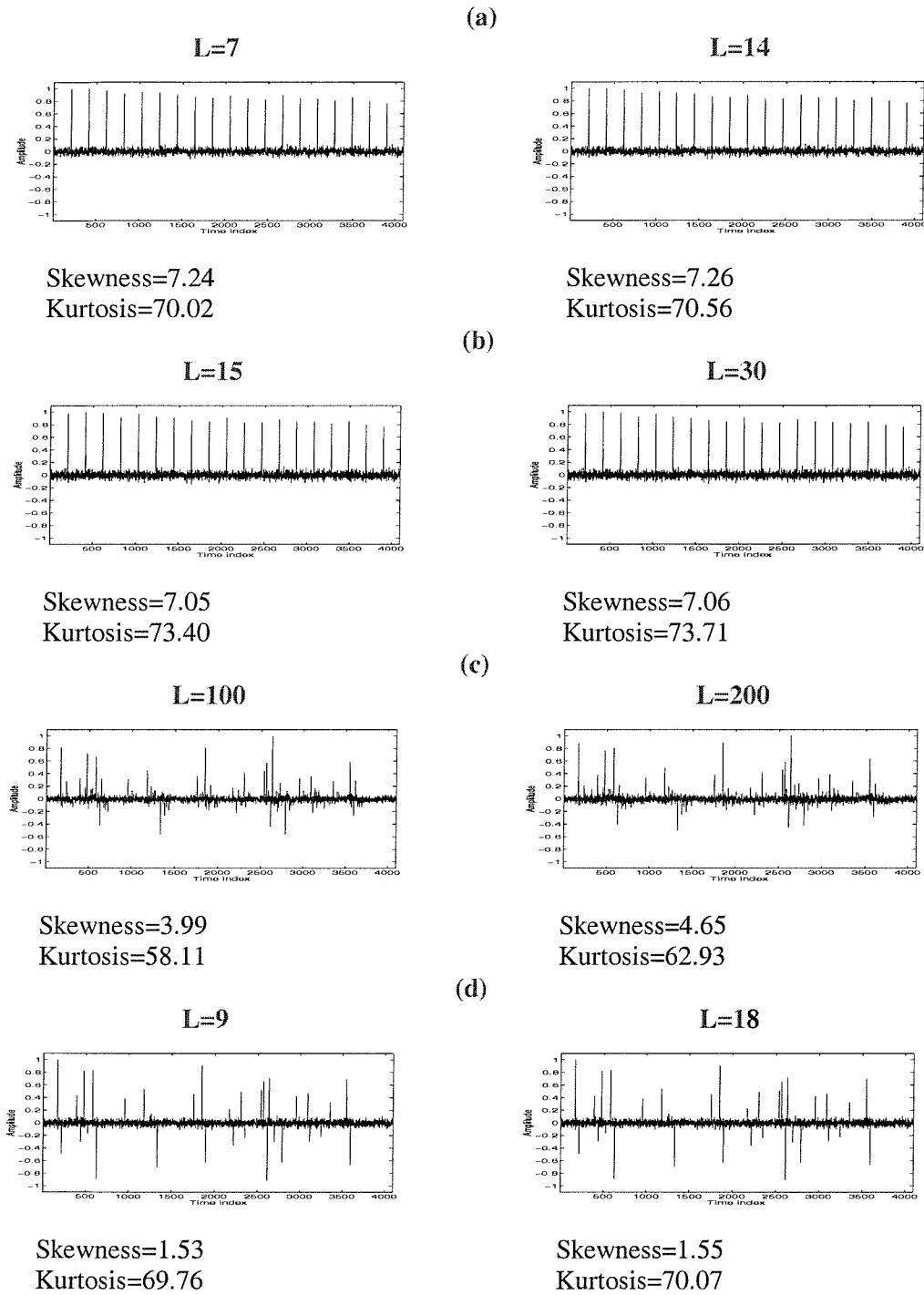


Figure E.3.3 Restored input signals with different inverse filter length when unknown system is simulated by AR(2) system ($r_i = 0.9$, $\theta_i = \pi/4$), and S/N ratio = 10 dB. (a): restored signal type 1 from third-order method, (b): restored signal type 1 from fourth-order method, (c): restored signal type 2 from third-order method, (b): restored signal type 2 from fourth-order method

From the comparison of restored signals in Figure E.3.3, we notice the higher order statistical values (skewness and kurtosis) of each signal are slightly increased when the length of the inverse filter is doubled. Due to the increased FIR filter coefficients, the components of noise signal takes the form of impacting components (signals possessing the amplitudes beyond the threshold) and thus becomes highly possible to give incorrect reconstruction of the impacting signal when the length of inverse filter is too excessive.

It has been observed from simulation 1 that both the skewness and kurtosis has similar trends, which are changing with the variations of impulse response length and noise power for measured and restored signal. In fact, the type of impacting signal gives little differences in this simulation.

For the measured signal, a significant decrease of higher order statistical values (skewness and kurtosis) have been observed as the length of the impulse response of the unknown system is increased for both noise free (S/N ratio is ∞ dB) and noise interference case. On the other hand, however, for the restored signal, it is noticeable for noise free case that the performance of reconstruction is unaffected by the length of the impulse response of unknown system. Thus, the measured signal coming from the filtering of impacting signal without noise interference provides the results of reconstruction identical regardless of the length of the impulse response of the system (i.e. system independent for noise free observed signal).

Throughout the simulation 2, we can see a certain trend on the length of the FIR inverse filter coping with the statistical status of the measured signal.

Above a certain range of inverse filter length in Figure E.3.1, the performance (higher order statistical values) of the restored signal becomes independent of the length of the inverse filter, which infers the existence of compact FIR inverse filter length. This is also supported from the comparison of restored signals from the minimal inverse filter which is coming from the results of Figure E.3.2 and from any long inverse filter illustrated in Figure E.3.3. The results of restored signals given in Figure E.3.3 demonstrate the fact even the longer inverse filters cannot guarantee the better restoration.

Appendix E, Determination of the inverse filter length

From these simulations, we conclude the following:

The performance of impacting signal reconstruction based on the higher order statistical method strongly depends on the degree of noise interference.

The results given in simulation 2 supports the fact that the blind reconstruction of an impacting signal from a measured signal can be acceptable for a certain range of length of inverse filter. Hence, a **criterion for the selection of optimal inverse filter length** is required, which is carried out in Chapter 5.

Appendix F

Independence and source separation structure (part of Chapter 7)

In this section we prove the fact that independence criteria based on the nulling of the fourth-order output joint-cumulant is a sufficient condition to separate the sources for statistically independent input signal.

Based on this, a blind source separation structure is introduced which can separate the observed signals into statistically independent ones from which the restoration of the input signals is achieved.

F.1 Independence using higher order statistics

The random signals (*i.i.d*) $x_i(n)$ and $x_j(n)$ are said to be mutually statistically independent if the joint probability density function is separable,

$$p_{x_i, x_j}(\alpha, \beta) = p_{x_i}(\alpha) \cdot p_{x_j}(\beta) \quad (\text{F.1.1})$$

The higher-order (order ≥ 3) joint-cumulants equal zero. Let $x_i(n)$ be the zero-mean signals, then the $(l+m)$ -th order joint-moments can be expressed as,

$$M_{lm}(x_i, x_j) = E[x_i^l x_j^m] \quad (\text{F.1.2})$$

where $l + m > 2$, $l \neq 0$, $m \neq 0$.

For the fourth-order, the joint-cumulants of two independent signals satisfy [Nikias and Petropulu, 1993],

$$\begin{aligned}
 Cum_{31}(x_i, x_j) &= M_{31}(x_i, x_j) - 3M_{20}(x_i, x_j)M_{11}(x_i, x_j) = 0 \\
 Cum_{22}(x_i, x_j) &= M_{22}(x_i, x_j) - M_{20}(x_i, x_j)M_{02}(x_i, x_j) - 2M_{11}^2(x_i, x_j) = 0 \quad (\text{F.1.3}) \\
 Cum_{13}(x_i, x_j) &= M_{13}(x_i, x_j) - 3M_{02}(x_i, x_j)M_{11}(x_i, x_j) = 0
 \end{aligned}$$

Also, $x_i(n-k)$ and $x_j(n-p)$ with any delays k, p are also independent. Therefore, to fourth-order,

$$Cum_{lm}(x_i(n-k), x_j(n-p)) = 0, \quad \text{with } l+m=4 \text{ and } l=m \neq 0 \quad (\text{F.1.4})$$

Moreover, if the signals \mathbf{x}_i and \mathbf{x}_j are statistically independent, then \mathbf{x}_i and $\mathfrak{I}(\mathbf{x}_j)$, where $\mathfrak{I}(\mathbf{x}_j)$ is the signal \mathbf{x}_j filtered by linear operator \mathfrak{I} , are also independent.

F.2 Source Separation structure

F.2.1 Blind Source Separation via joint-cumulant cancellation

In the context of multichannel blind deconvolution, the observation vector $\mathbf{v}(k)$ is assumed to be generated from an unknown source vector $\mathbf{x}(k)$ through the unknown multivariate filter $\mathbf{H}(z)$ i.e.,

$$V(z) = \mathbf{H}(z)X(z) + W(z) \quad (\text{F.2.1})$$

where $W(z)$ is a p dimensional additive white Gaussian noise that is assumed to be statistically independent of the source vector $\mathbf{x}(k)$ and the elements of the matrix $\mathbf{H}(z)$ are given by

$$H_{ij}(z) = \sum_{k=0}^{L_{ij}} h_{ij}(k)z^{-k} \quad (\text{F.2.2})$$

where L_{ij} represents the length of FIR filter length (we assume the unknown systems have finite impulse response functions).

Equation (F.2.1) may be expanded in the time domain, and the output of the i -th sensor is

$$v_i(n) = \sum_{j=1}^m \sum_{k=0}^{L_{ij}} h_{ij}(k) x_j(n-k) + w_i(n), \quad 1 \leq i \leq p \quad (\text{F.2.3})$$

The task of multichannel blind deconvolution is to restore the source vector $\mathbf{x}(k)$ from the observation vector $\mathbf{v}(k)$, up to possibly scaled, reordered, and delayed estimates, i.e., $\hat{\mathbf{x}}(k) = \mathbf{P} \Lambda \mathbf{D}(z) \mathbf{x}(k)$, where $\mathbf{P} \in \mathbb{R}^{m \times m}$ is a permutation matrix, $\Lambda \in \mathbb{R}^{m \times m}$ is a nonsingular diagonal matrix, and $\mathbf{D}(z)$ is a diagonal matrix given by

$$\mathbf{D}(z) = \text{diag} \left\{ z^{-d_1}, z^{-d_2}, \dots, z^{-d_m} \right\} \quad (\text{F.2.4})$$

which implies the existence of delays in each estimated signal $\hat{\mathbf{x}}(k)$.

In other words, the objective of multichannel blind deconvolution is to design a multichannel inverse system so that the global system $\mathbf{G}(z)$ (which combines the effect of unknown system and inverse system) has a decomposition of the following form:

$$\mathbf{G}(z) = \mathbf{P} \Lambda \mathbf{D}(z) \quad (\text{F.2.5})$$

The structure of source separation is a set of linear filters through which the observed signal is filtered to yield the source signals as

$$\mathbf{S}(z) = \mathbf{F}(z) \mathbf{V}(z) \quad (\text{F.2.6})$$

where $\mathbf{F}(z)$ is a matrix of filters.

Combining (F.2.1) with noise free condition, we obtain

$$\mathbf{S}(z) = \mathbf{F}(z) \mathbf{H}(z) \mathbf{X}(z) = \mathbf{G}(z) \mathbf{X}(z) \quad (\text{F.2.7})$$

where $\mathbf{G}(z)$ is a matrix form of system-inverse system combined filters, which is shown in Figure F.2.1.

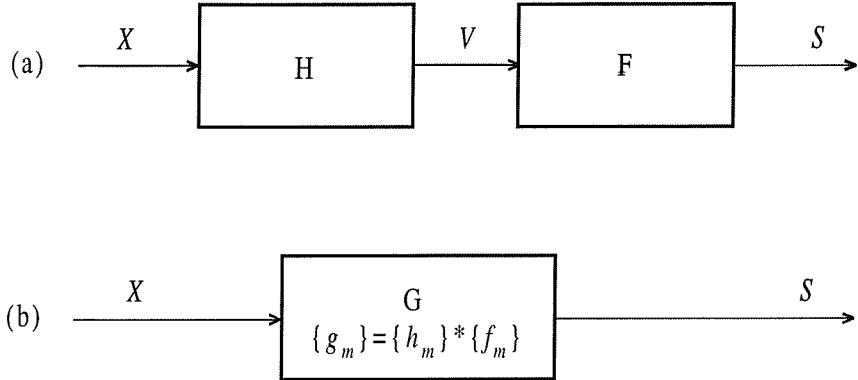


Figure F.2.1 Source separation structure. (a): convolution-deconvolution system, (b): cascade system

Using simplifications mentioned above, matrix $G(z)$ becomes

$$G(z) = F(z)H(z) = \begin{bmatrix} F_{11}(z) + F_{12}(z)H_{21}(z) & F_{11}(z)H_{12}(z) + F_{12}(z) \\ F_{22}(z)H_{21}(z) + F_{21}(z) & F_{22}(z) + F_{21}(z)H_{12}(z) \end{bmatrix} \quad (\text{F.2.8})$$

The separation will be achieved if matrix $F(z)$ is such that $G(z)$ becomes diagonal up to a permutation:

$$F(z)H(z) = \begin{bmatrix} G_{11}(z) & 0 \\ 0 & G_{22}(z) \end{bmatrix} \quad (\text{F.2.9})$$

or

$$F(z)H(z) = \begin{bmatrix} 0 & G_{12}(z) \\ G_{21}(z) & 0 \end{bmatrix} \quad (\text{F.2.10})$$

Combining (F.2.9) or (F.2.10) with (F.2.8) leads to the two separating conditions, respectively,

$$\begin{aligned} F_{12}(z) / F_{11}(z) &= -H_{12}(z) \\ F_{21}(z) / F_{22}(z) &= -H_{21}(z) \end{aligned} \quad (\text{F.2.11})$$

or

$$\begin{aligned} F_{11}(z) / F_{12}(z) &= -H_{21}(z) \\ F_{21}(z) / F_{22}(z) &= -H_{12}(z) \end{aligned} \quad (\text{F.2.12})$$

When we consider the system H_{ij} are assumed to be FIR filters, the inverse system F_{ij} may be considered so that solutions (F.2.11) and (F.2.12) leads to FIR filters.

F.2.2 Blind Source Separation via joint-cumulant cancellation

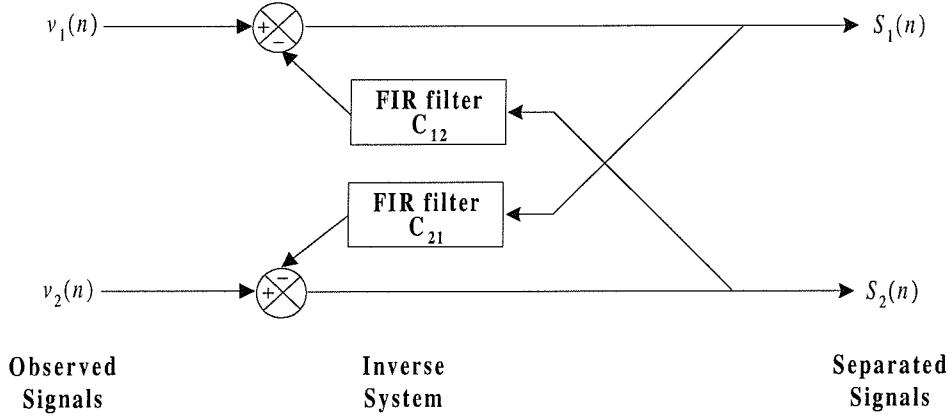


Figure F.2.2 Recursive source separation structure

In the z-domain, provided $(I - C(z))^{-1}$ exists, the matrix equation is

$$S(z) = (I + C(z))^{-1}V(z) \quad (\text{F.2.13})$$

where $C(z)$ is the matrix of inverse filters $C_{ij}(z)$.

Combining (F.2.10) and (F.2.13) leads to

$$\begin{aligned} F(z) &= (I + C(z))^{-1} \\ &= \frac{1}{1 - C_{12}(z)C_{21}(z)} \begin{bmatrix} 1 & -C_{12}(z) \\ -C_{21}(z) & 1 \end{bmatrix} \end{aligned} \quad (\text{F.2.14})$$

and matrix $G(z)$ in (F.2.8) becomes

$$G(z) = \frac{1}{1 - C_{12}(z)C_{21}(z)} \begin{bmatrix} 1 - C_{12}(z)H_{21}(z) & H_{12}(z) - C_{12}(z) \\ H_{21}(z) - C_{21}(z) & 1 - C_{21}(z)H_{12}(z) \end{bmatrix} \quad (\text{F.2.15})$$

Then, combining (F.2.9) and (F.2.10), we derive the two solutions,

$$C_{ij}(z) = H_{ij}(z) \quad (\text{F.2.16})$$

or

$$C_{ij}(z) = 1 / H_{ji}(z) \quad (\text{F.2.17})$$

with outputs, respectively

$$S_i(z) = X_i(z) \quad (\text{F.2.18})$$

or

$$S_i(z) = H_{ij}(z)X_j(z) \quad (\text{F.2.19})$$

If we constrain $C_{ij}(z)$ in Figure F.2.2 to be FIR filters, only solution (F.2.16) can be achieved under the condition such that the $H_{ij}(z)$ are FIR filters.

With the condition $F_{ii}(z)=1$, equation (F.2.8) becomes

$$G(z) = \begin{bmatrix} 1 + F_{12}(z)H_{21}(z) & H_{12}(z) + F_{12}(z) \\ H_{21}(z) + F_{21}(z) & 1 + F_{21}(z)H_{12}(z) \end{bmatrix} \quad (\text{F.2.20})$$

Let L be the maximum order of filters H_{ij} , then F_{ij} , and G_{ij} , must also be L th-order FIR filters and G_{ij} becomes $2L$ th-order FIR filter;

$$\begin{aligned} G_{ij}(z) &= \sum_{k=0}^{L-1} (h_{ij}(k) + f_{ij}(k))z^{-k}, \\ G_{ii}(z) &= 1 + \left(\sum_{k=0}^{L-1} f_{ij}(k)z^{-k} \right) \left(\sum_{k=0}^{L-1} h_{ji}(k)z^{-k} \right) \end{aligned} \quad (\text{F.2.21})$$

From the relation $S(z) = G(z)X(z)$, we utilise the fourth-order joint-cumulants between the outputs at different times: $s_1(n)$ and $s_2(n-k)$, and use the linearity properties of these cumulants [Nikias and Petropulu, 1993] with the assumption of source independency. Thus, as already described in sub-section 7.2.2, all the joint-cumulants involving \mathbf{x}_1 and \mathbf{x}_2 are zero, and the expression is expanded as (using the cascade structure of Figure F.2.1)

$$\begin{aligned} \text{Cum}_{lm}(s_i(n), s_j(n-k)) &= \\ &= \text{Cum}_{lm} \left(\sum_{p=0}^{2L-1} g_{ii}(p)x_i(n-p), \sum_{q=0}^{L-1} g_{ji}(q)x_i(n-k-q) \right) \\ &= \text{Cum}_{lm} \left(\sum_{q=0}^{L-1} g_{ij}(q)x_j(n-q), \sum_{p=0}^{2L-1} g_{jj}(p)x_j(n-k-p) \right) \\ l + m &= 4, \quad l \neq m \neq 0 \end{aligned} \quad (\text{F.2.22})$$

If signal $x(n)$ is a zero-mean independent, identically distributed (*i.i.d.*) and non-Gaussian process, then

$$\begin{aligned}
 & \text{Cum}_p^x(k_1, k_2, \dots, k_{p-1}) \\
 &= \text{Cum}\left(x(n), x(n-k_1), \dots, x(n-k_{p-1})\right) \\
 &= \begin{cases} \gamma_p^x & \text{if } k_1 = k_2 = \dots = k_{p-1} = 0 \\ 0, & \text{otherwise} \end{cases}
 \end{aligned} \tag{F.2.23}$$

where

$$\gamma_p^x = \text{Cum}_p^x(0, 0, \dots, 0) \tag{F.2.24}$$

If the signal $x(n)$ is not i.i.d., but relation (F.2.23) holds up to p th-order, $x(n)$ is said to be a to p th-order white signal [Nikias and Mendel, 1993]. Expanding the equation (F.2.22), the joint-cumulants between outputs at different delays $0 \leq k \leq L-1$ and for $i, j \in \{1, 2\}$, are expressed as

$$\begin{aligned}
 & \text{Cum}_{31}(s_i(n), s_j(n-k)) \\
 &= \sum_{l=0}^{L-1} g_{ii}^3(l+k) g_{ji}(l) \gamma_4^{x_i}(n-k-l) \\
 &+ \sum_{l=0}^{L-k-1} g_{jj}^3(l) g_{ij}^3(l+k) \gamma_4^{x_j}(n-k-l)
 \end{aligned} \tag{F.2.25}$$

$$\begin{aligned}
 & \text{Cum}_{13}(s_i(n), s_j(n-k)) \\
 &= \sum_{l=0}^{L-1} g_{ii}^3(l+k) g_{ji}^3(l) \gamma_4^{x_i}(n-k-l) \\
 &+ \sum_{l=0}^{L-k-1} g_{jj}^3(l) g_{ij}^3(l+k) \gamma_4^{x_j}(n-k-l)
 \end{aligned} \tag{F.2.26}$$

$$\begin{aligned}
 & \text{Cum}_{22}(s_i(n), s_j(n-k)) \\
 &= \sum_{l=0}^{L-1} g_{ii}^2(l+k) g_{ji}^2(l) \gamma_4^{x_i}(n-k-l) \\
 &+ \sum_{l=0}^{L-k-1} g_{jj}^2(l) g_{ij}^2(l+k) \gamma_4^{x_j}(n-k-l)
 \end{aligned} \tag{F.2.27}$$

Using the assumption that H_{ij} are FIR filters and the constraint on F_{ij} are FIR filters with $F_{ii} = F_{jj} = 1$, it can be deduced that $G_{jj} = 1 + F_{ji}H_{ij}$ cannot be zero. Also assuming $|h_{ij}(0)| < 1$ and at least one of the input signal is non-Gaussian, we

derive $g_{jj}(0) = 1 + f_{ji}(0)h_{ij}(0) \neq 0$. Thus, the terms $g_{ij}(k)$ in equations (F.2.25), (F.2.26), and (F.2.27) should be zero to satisfy the separation condition.

In this section, we utilise the cancellation of $Cum_{31}(s_i(n), s_j(n-k))$ to derive the c_{ij} s in an iterative manner, i.e.,

$$\begin{aligned} \text{For } k \in [0, L] \\ c_{ij}(q+1, k) = c_{ij}(q, k) - \mu Cum_{31}(s_i(n), s_j(n-k)) \end{aligned} \quad (\text{F.2.28})$$

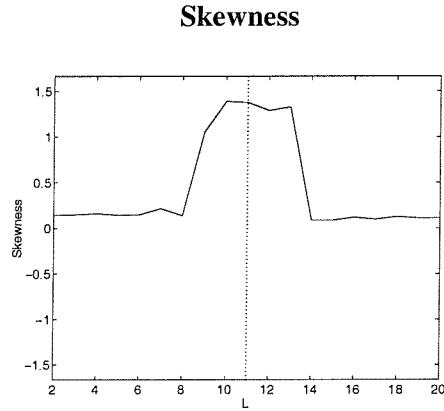
where q represents each iteration point and μ designates a positive adaptation step.

F.2.3 Other considerations on the blind source separation structure

The effect of length of the FIR separation system

The orders of the separation systems in these simulations have been selected as the same length of the unknown system's maximum MA order. However, from a practical point of view, this cannot be achieved. For simulation case (a) in Chapter 7 - which allows tight margin of FIR separation filter length only - requires an estimation of the unknown systems' MA order to yield a reasonable selection of the length of the FIR separation system. This task can be realised using various methods of MA order and parameter determination using higher order cumulant as mentioned in Chapter 5.

Simulation case (a); $\mu = 0.01$, $t_r = 1.4e-3$



Simulation case (b); $\mu = 0.01$, $t_r = 0.0026$

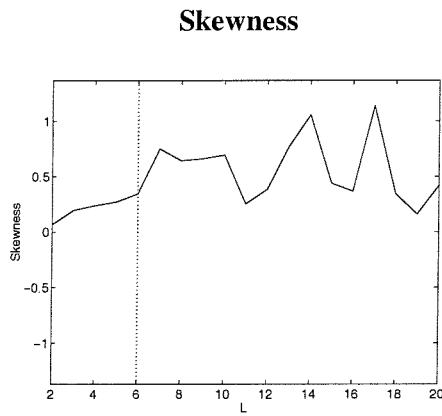


Figure F.2.3 The effect of FIR separation system length. Dotted line represents the maximum MA order of unknown system.

Note that the performance of impacting signal reconstruction for simulation case (b) in Chapter 7 is less severe in selection of the length of FIR separation system than the case (a). This fact is relieving aspect that the practical systems are mostly takes the form of the case (b) types. Thus, once the length of the FIR filter length can be above a maximum length of the unknown systems' impulse response function, the separation task will not be severely affected by the length of the separation system.

The effect of Gaussian noise interference on each channel output

The simulations performed here do not consider the effect of noise. However, in practice, there could be external disturbances such as noise interference or measurement error. To cope with this, Gaussian noise which is independent of any other signals are introduced to the output of each channel and the source separation is carried out using the same factors (L , μ , t_r) as the noise free case.

Simulation case (a); $L=11$, $\mu = 0.01$, $t_r=1.4e-3$

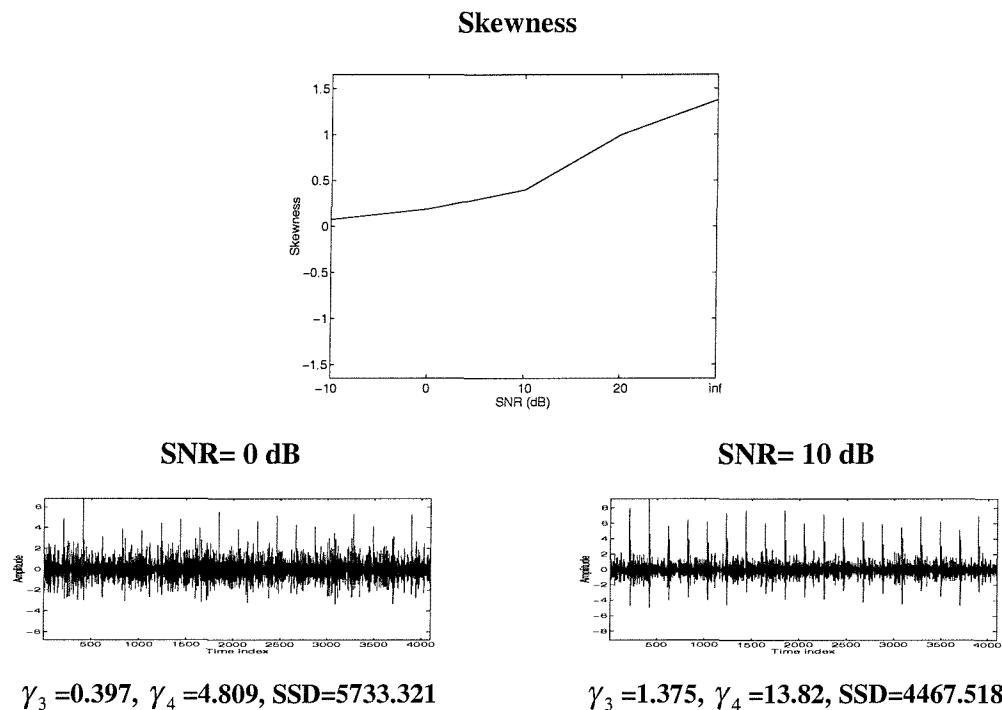


Figure F.2.4 The effect of Gaussian noise interference

Simulation case (b); L=6, $\mu = 0.01$, $t_r = 0.0026$

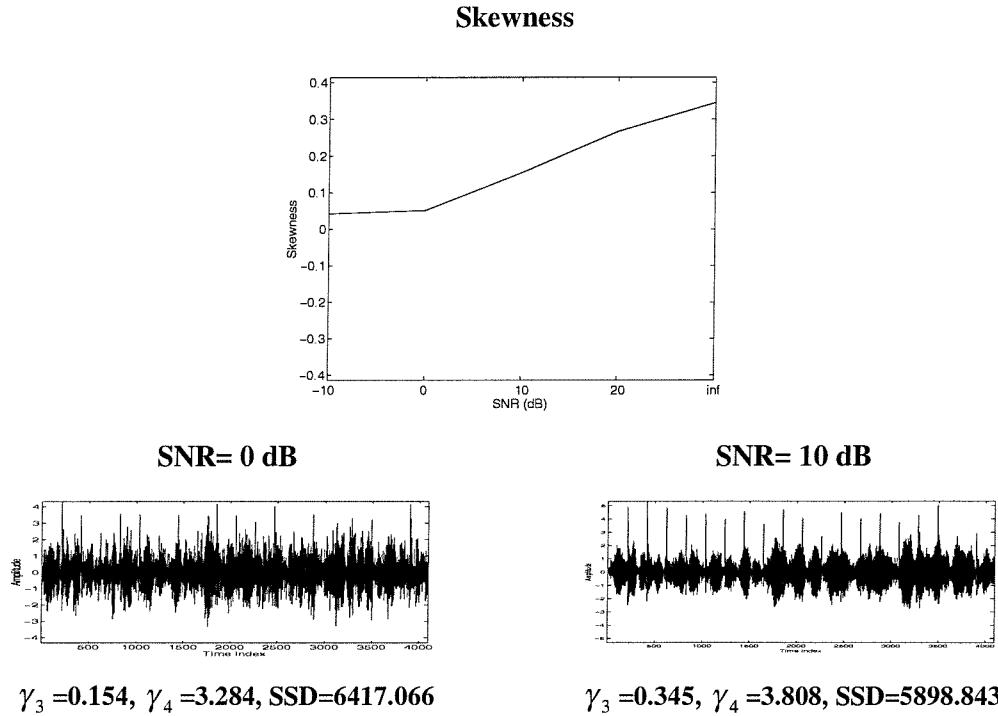


Figure F.2.5 The effect of Gaussian noise interference