BLIND DECONVOLUTION: TECHNIQUES AND APPLICATIONS

Fu-Chun Zheng BSc MSc AMIEE

A thesis submitted for the degree of *Doctor of Philosophy* to the Faculty of Science and Engineering at

University of Edinburgh

--- March 1992 ---

CONTENTS

ABSTRACT

This thesis is primarily concerned with developing new parameter based blind deconvolution algorithms and studying their applications. The blind deconvolution problem for minimum phase (MP) systems is well understood, and in this case the well known predictive schemes can be employed. When systems are nonminimum phase (NMP), however, the predictive deconvolution methods can only generate the spectrally equivalent MP solution. This is because the predictive schemes are based only on autocorrelations, which are completely blind to the phase properties of systems. In order to solve the blind deconvolution problem of NMP systems, higher order cumulant (HOC) analysis is adopted in this thesis. The reason for this is that HOC carry the phase information of systems only to a linear phase shift. the parametric approach is adopted due to its advantages in terms of variance and resolution over nonparametric methods. Both MA and AR based models are studied in this work.

A new robust blind deconvolution algorithm for MP systems: variance approximation and series decoupling (VASD), is presented first. It is shown that this algorithm possesses some advantages over the existing ones with the same purpose.

Then, based on a MA system model, we proposed a HOC-based two-step relay algorithm, in which the close-form formula for MA parameters are combined with an optimal fitting scheme, and the thorny problem of multimodality is overcome to a very great degree. Thus, the optimal identification of the MA parameters of NMP systems can be obtained.

In the study of the AR based model, six new families of HOC based linear equations with respect to the AR parameters are derived. Since the inverse filter coefficients are simply the solution of a set of linear equations, their uniqueness can normally be guaranteed. In comparison with the existing AR based methods, only diagonal slices of cumulants are used in our algorithms, in which simplicity and elegance are fully embodied. It has been shown that our algorithm can offer more accurate results than the existing ones.

Finally, the algorithms obtained above are made adaptive through the novel use of the successive over-relaxation (SOR) scheme, and a fast convergence rate is realised. As a result, the derived adaptive algorithms are capable of dealing with both linear time invariant (LTI) and linear time variant (LTV) systems. Equalisation of multilevel pulse amplitude modulation (PAM) series, which have been transmitted through NMP channels, is simulated, and the results are presented.

The research recorded in this thesis is of interest to both the data communications (NMP channel equalisation) and seismic data processing (NMP wavelet removal) community.

LIST OF ABBREVIATIONS

AGC	automatic gain control
AR	autoregressive
AST	asymmetric-to-symmetric transformation
ARMA	autoregressive moving average
BGR	Benveniste-Goursat-Ruget
DD	decision directed
DFT	discrete Fourier transform
EMG	electromyogram
EW	elementary waveform
FDM	frequency domain multiplexing
GM	(1) Giannakis-Mendel
	(2) Giannakis method
HOC	higher order cumulant
HOS	higher order spectrum
IDFT	inverse discrete Fourier transform
IID	independent and identically distributed
ISI	inter-symbol interference
LMS	least mean squares
LS	least squares
LTI	linear time-invariant
LTV	linear time-variant
MA	moving average
MC	Monte-Carlo
MED	minimum entropy deconvolution
ML	maximum likelihood
MLD	maximum likelihood deconvolution
MP	minimum phase
NMP	non-minimum phase
NMU	neuro-muscular-unit
OPT	optimal
PAM	pulse amplitude modulation
RLS	recursive least squares
sEMG	surface electromyogram
SAT	symmetric-to-asymmetric transformation
SLP	searching linear programming
SNR	signal-to-noise ratio
SOR	successive over-relaxation
TDM	time domain multiplexing
TDR	time domain reflector
VASD	variance approximation and series decoupling

LIST OF PRINCIPAL SYMBOLS

a_i	(1) parameter of AR part in an ARMA model
	(2) asymmetrically distributed data
b_i	parameter of MA part in a ARMA model
$c(k)$	restored series
$c_l()$	diagonal slice of lth-order cumulants
e	2.7182818 $\cdots$
$e(t)$	elementary waveform
$h(k)$, h_k	system unit-sample response
i	parameter number
j	(1) parameter number
	(2) $\sqrt{-1}$
k	sample number
l	parameter number
$max(\cdot)$	maximum of a set of real numbers
$min(\cdot)$	minimum of a set of real numbers
n	parameter number
$n(k)$	sampled additive (or measurement) noise
$n(t)$	additive (or measurement) noise
p	AR order of ARMA model
p_l	predictor coefficient
q	MA order of ARMA model
r_1	order of non-causal part of inverse filter
r_2	order of causal part of inverse filter
s_i	symmetrically distributed data
$s(t)$	surface electromyogram
s_c^2	estimated variance
t	time
u_i	coefficient of elementary waveform
$w(k)$	realisation of an IID process
$x(k)$	(1) system input series
	(2) system output series
$x_{nf}(k)$	noise-free system output
$y(k)$	system output series
$y_{nf}(k)$	noise-free system output series
$\hat{y}(k+1\|k)$	one-step prediction of $y(k)$ at time k
z	variable in the z-transform
z_i	zero of system transfer function
$A(\omega)$	amplitude frequency response of a system
$C_l(\tau)$	lth-order cumulant
$D[\cdot]$	characteristic system

$D(N)$	system enery delay
$E(N)$	system enery evolution
$\mathbf{E}[\cdot]$	mean of a random process
$H(z)$	system transfer function
$Imag(\cdot)$	imaginary part of a complex number
$M_l(\tau)$	l th-order moment
N	number of samples
$P(\cdot)$	probability
$\hat{R}(\tau)$	estimate of 3rd-order cumulant
$Real(\cdot)$	real part of a complex number
$\mathbf{S}[\cdot]$	system operator
V	varimax norm
$V(k)$	seismic wavelet
$W(k)$	IID process
$X(z)$	z-transform of $x(k)$
$Y(k)$	system output process
λ	parameter of Bernoulli distribution
ϵ^2	value of cost function
γ	scaling factor in BGR method
γ_2, σ^2	variance
γ_3	skewness of a random process
γ_4	kurtosis of a random process
μ	relaxation factor
$\mu(k)$	reflectivity of the earth
ω	angular frequency in radians
$\phi(\omega)$	phase frequency response of a system
δ	step size in VASD
$\delta(\tau)$	Dirac function
τ	(1) time delay (2) step size in steepest descent algorithm
θ_j	coefficient of inverse filter
$\Phi(n)$	(1) value of autocorrelation function at the n th delay (2) value of crosscorrelation function at the n th delay
∇_l	gradient in an optimisation procedure

ACKNOWLEDGEMENTS

During my study of the last three years here in Scotland, many people have provided me with their invaluable help.

I am indebted to my two supervisors, Dr. S. McLaughlin and Dr. B. Mulgrew, without whose supervision this thesis would never have been generated.

I am particularly grateful to Prof. P. Grant who has supplied much advice which was of enormous help in all aspects.

I would also like to thank **Sino-British Friendship Scholarship Scheme (SBFSS)** for the comprehensive financial support for my study here.

Last but not least, I must express my gratitude to my parents, my grandparents, and especially my wife, Wanjie. Without their consistent encouragement and support I might never have survived the slog of the last three years.

Chapter 1:

INTRODUCTION

The key to our understanding of natural phenomena is data acquisition (measurement) and analysis by some combination of appropriate techniques. However, it is not always possible to obtain data of the necessary accuracy directly by measurement in the majority of cases. What is acquired are only images of the true process. Moreover, the data is always blurred by measurement noise of some form, a typical example of this being seismic exploration. In this situation what is obtained is not the true reflectivity series of the earth but the convolution series, corrupted by noise, of the true reflectivity and wavelet. Similarly, received data in communication systems is often contaminated because of the presence of noise and the effect of non-ideal channel characteristics. Thus, restoring the original data from measurement data plays an extremely important role in many applications. This is the role of blind deconvolution, which will be described in this thesis.

1.1 Blind Deconvolution

The term, *blind deconvolution*, first appeared in T. G. Stockham *et al*'s paper: " Blind Deconvolution through Digital Signal Processing" in 1975 [1]. In this paper, results were presented of improving the voice quality of recordings produced before the mid-1920's by the homomorphic method first described by A. V. Oppenheim in [2]. However, the practice of the technique proceeded the adoption of the term "blind deconvolution". From 1950's onwards, geologists started using blind deconvolution methods in seismic exploration. Over the next two decades, several deconvolution algorithms were routinely used in commercial geological data processing packages [3]-[10]. These algorithms manifested their great power and efficacy in the discoveries in the North Sea, Alaska, Africa, the Far East, the Persian Gulf, and many continental shelf regions of the world [11].

Mathematically, the formulation of blind deconvolution can be generally described as follows. With reference to Fig(1.1), let the input signal of a system be $x(k)$, and the output signal $y(k)$. If we assume that,

$$y(k)=\mathrm{S}[x(k)]+n(k) \tag{1.1.1}$$

is satisfied, where S[·] is the system operator, and $n(k)$ the measurement noise associated with the observation $y(k)$, then the problem of determining both S[·] and $x(k)$ from only $y(k)$ and some assumed information regarding S[·] and $x(k)$ is referred to as *blind deconvolution*. As to S[·] (or the system), it may be linear or nonlinear, and time-invariant or time-variant. But in this thesis, it will always be assumed to be linear time-invariant (LTI), unless explicitly indicated otherwise. Consequently, according to linear system theory, Eq(1.1.1) can be reduced to

$$y(k)=x(k)*h(k)+n(k) \quad , \tag{1.1.2}$$

where the symbol $*$ denotes the convolution operation, and $\{h(k)|\ k=0,\ 1,\ \ldots\}$ represents the (unit) impulse response of the system, as in Fig(1.2).

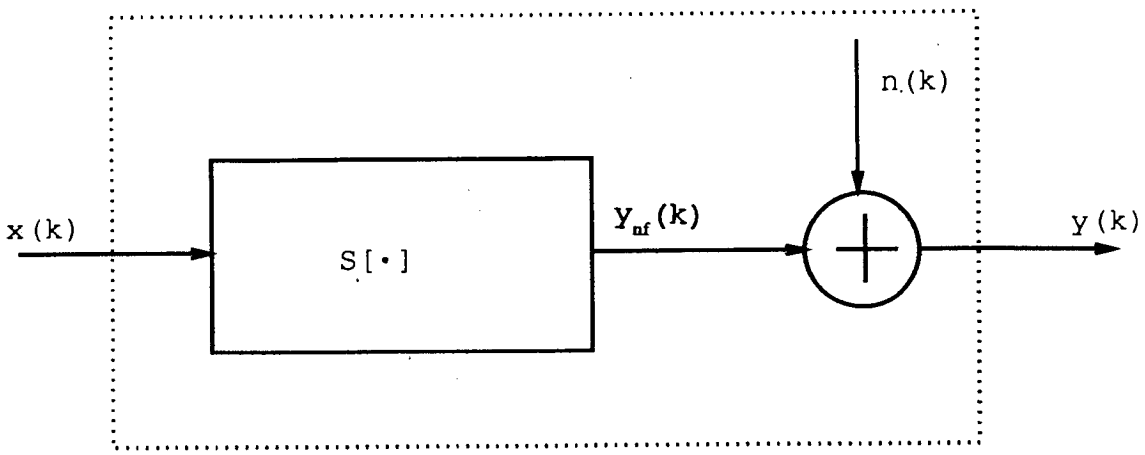
In fact, deconvolution involves two steps: system identification and signal restoration. System identification, which involves the estimation of the system model parameters, is the first step, and generally the more difficult one, since the identification is required to be correct both in amplitude and in phase. Restoration is the process by which the inverse filter is formed from the corresponding system model parameters obtained in the identification process and then the estimated input signal determined. However, the interface between these two steps may not be very explicit, since it depends on the adopted system model. Therefore, we can say that, deconvolution is the identification of system impulse response and the restoration of the input signal both in amplitude and in phase.

From Eq(1.1.2), it is clear that to determine $h(k)$ and $x(k)$ only from $y(k)$ is mathematically intractable. Consequently, blind deconvolution cannot be realised under the absolute condition that only $y(k)$ is known. Thus some assumptions regarding $h(k)$ and $x(k)$ are required to solve the problem of blind deconvolution. The most conventional assumption for $x(k)$ is that it is an independently and identically distributed (IID) stochastic signal (or series). This implies that $x(k)$ is stationary and ergodic. In this thesis, unless being pointed out specifically, the following conditions will be assumed to hold: *(a)* the system input $x(k)$ is zero-mean and IID; and *(b)* the diagonal slices of the ith order cumulants of $x(k)$ are limitedly large, where $i=2,\ 3,\ 4,$ and 6 [15].

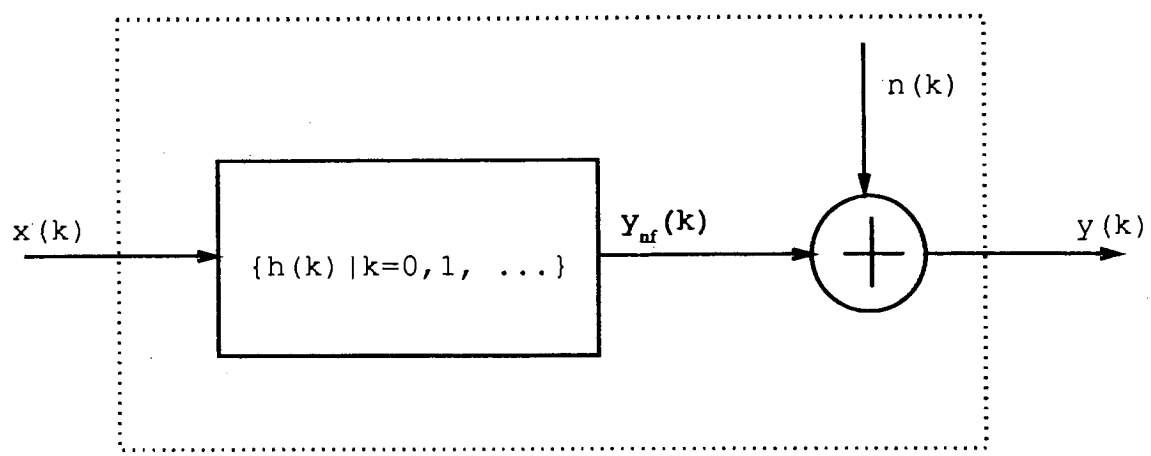
As to $h(k)$ (*i. e.*, the system), it has often been assumed to be minimum phase (MP). This assumption comprises the prime basis of prediction based deconvolution algorithms, which form an important class of deconvolution techniques [3]-[9]. In fact, predictive deconvolution has been a routine scheme in some fields such as seismic geoexploration [4]-[7][10]. When the considered system is non-minimum phase (NMP), however, predictive deconvolution must fail, since predictive schemes are based only on the autocorrelation functions of system output. The autocorrelation functions of a signal are totally insensitive (or blind) to the phase properties of the signal itself. Unfortunately, many (artificial or natural) systems are exactly NMP. The practical examples for NMP case include Vibroseis wavelets in geoexploration [9][10] and telephone and fading radio channels in telecommunications [19][20][36]. Thus, the MP assumption becomes a major obstacle, and must be relaxed or removed (for definitions of MP and NMP systems, see Section 2.2 of this thesis).

Over the past decade, higher order cumulants, (HOC, or equivalently, higher order spectra (HOS), which are Fourier transform of HOC), based techniques have been found to be a powerful tool when applied to the problem of blind deconvolution of NMP systems [12]-[15]. The HOC of a signal carry the phase information of the signal up to a linear phase shift, consequently, HOC based techniques can be applied to NMP systems.

In fact, HOC (or HOS) analysis technology is emerging as a very important technique in the field of signal processing. The potential of HOC analysis has been found to be useful in many other domains in addition to system identification and deconvolution [28][33]. However, this thesis in the main concentrates on the application of HOC analysis to blind identification and deconvolution of NMP systems.



Fig(1.1) A general system model.



Fig(1.2) A linear system model.

1.2 Applications of Blind Deconvolution

If Eq(1.1.2) is studied, it is clear that this model representation is applicable in many fields. In fact, blind deconvolution has in recent decades found an extremely wide range of applications. These include: speech processing [1][16][17], seismic signal processing [3]-[11], adaptive equalisation of communication channels [18]-[21], echo cancellation (dereverberation), image restoration [22][23], and time domain automatic measurement [24]. Due to space limitations, only four specific examples will be briefly discussed as follows.

1.2.1 Geoexploration

Fig(1.3) illustrates a typical geoexploration data acquisition configuration, consisting of a seismic source and a waveform receiver [5]. The most efficient and simplest seismic sources include dynamite and other high explosives. The waveform receiver normally comprises a geophone array. By using this set of equipment, the reflection data from the earth, termed a seismogram, can be acquired and recorded on magnetic tape for further processing.

In Fig(1.3), let $y(k)$ be the seismogram, $V(k)$ the wavelet and $\mu(k)$ the true reflectivity of the earth, then following equation holds:

$$y(k) = V(k)*\mu(k) + n(k) \quad , \tag{1.2.1}$$

i. e., the received signal $y(k)$ is always the convolution of wavelet $V(k)$ with reflectivity $\mu(k)$ (plus noise). Consequently, it is necessary to deconvolve $y(k)$ in order to obtain the true reflectivity $\mu(k)$. Obviously, this is a typical blind deconvolution problem.

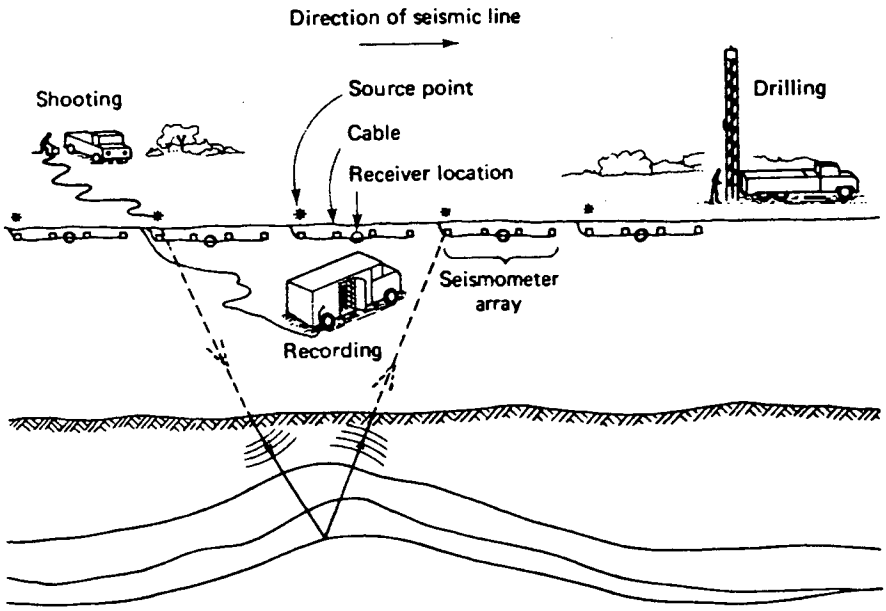
Comparing Eq(1.2.1) and Eq(1.1.2), we are led to the following important system interpretation for the seismic reflection model [11]: $y(k)$ can be thought of as the output of a system whose unit impulse response $h(k)=V(k)$ and whose input signal $x(k)=\mu(k)$. Although this interpretation is counterintuitive from a physical point of view, it is more useful and convenient for general description from a mathematical point of view [11]. As a result, in this thesis, we will directly use Eq(1.1.2), and indiscriminately employ terms "system" and "wavelet".

The deconvolution of a seismogram can greatly improve the resolution of the seismic traces, and greatly aid interpretation of the stratum structure of the earth.

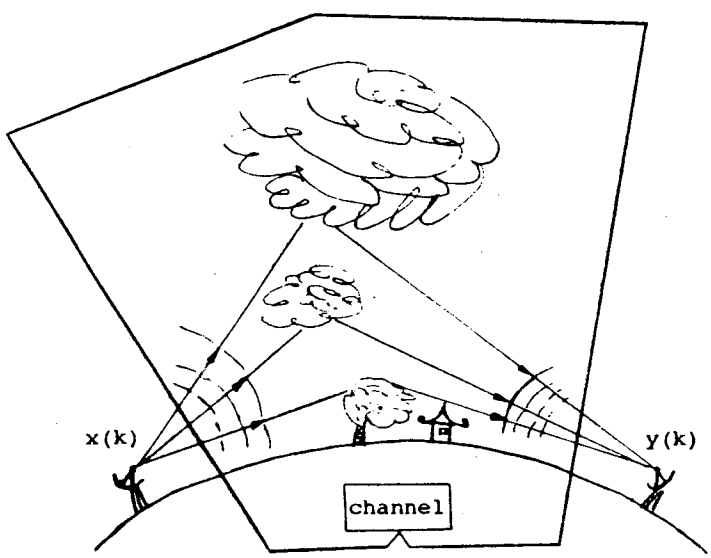
1.2.2 Equalisation of channels

Fig(1.4) illustrates a typical communication process, where $y(k)$ represents the received signal, and $x(k)$ the transmitted signal. There are many types of interference during the transmission of the data which can seriously distort the received signal $y(k)$, in comparison with the true signal $x(k)$. These interferences result in inter-symbol interference (ISI) [25][26], and consequently, it is necessary to remove the distortion from $y(k)$.

Theoretically, the above scenario can be described by a linear channel model, and then, Eq(1.1.2) holds. Thus, blind deconvolution, which is termed *blind equalisation* in communications, can greatly reduce the distortion caused by the channel and the additive noise.



Fig(1.3) A geoexploration data acquisition configuration [5].



Fig(1.4) A practical channel.

The main advantage of blind equalisation over other equalisation techniques is that it is possible to realise blind starting and remove the training stage which is essential in other approaches [21].

1.2.3 Biomedical signal processing

Many cases in this field may also be modelled as a blind deconvolution problem. Here, only the analysis of surface electromyogram (sEMG) is taken as an example [27].

A set of a moto-neuron and muscles governed by the neuron is referred to the neuro-muscular unit (NMU). Action potentials propagating along muscle fibers cause changes in potential at the skin surface. The temporal superposition of such potential changes is termed s-EMG in biomedical engineering. In fact, what is of most interest is the NMU activity signal rather than s-EMG. But it is only possible to measure the s-EMG in practice. Fortunately, sEMG is the superposition of a series of NMU signals, and the following expression holds:

$$s(t) = \sum_i u_i * e(t-i) + n(t) \quad , \tag{1.2.2}$$

where, $s(t)$ is the sEMG, $e(t)$ is the Elementary Waveform (EW): the NMU signal, and u_i is the coefficient of each EW. It can be seen that Eq(1.2.2) is identical in form to Eq(1.1.2). By means of blind deconvolution, the NMU signal can be extracted, and more information regarding possible muscular disorder obtained. This area is another promising application area for blind deconvolution.

1.2.4 The improvement of acoustic quality

This example involves the processing of voice signals [1]. Before the mid-1920's, the music records were made using the acoustic method. As a result of limitations in the recording setup, there is always distortion from the disks recorded in that era. The advantage of the techniques used then was an ability to avoid what is presently termed nonlinear distortion, and the major effect of the recording system reduces to be the same as that of a linear filter. In other words, only linear distortion remains. Therefore, the waveform $v(t)$ which is available to us today by playing an old recording can be modeled as

$$v(t) = s(t) * h(t) + n(t) \quad , \tag{1.2.3}$$

where, $s(t)$ denotes the sound being recorded, $h(t)$ the impulse response of the recording mechanism, and $n(t)$ the effects of surface noise. Clearly, Eq(1.2.3) is of the same form as Eq(1.1.2), and represents another typical blind deconvolution problem.

Today, by means of homomorphic deconvolution, one of the early blind deconvolution techniques, the effect of $h(t)$ has been successfully reduced or removed, and the voice quality is greatly improved.

1.3 Project Motivation

As discussed in the last section, blind deconvolution is being used in an increasingly wide range of fields. This fact itself forms the practical drive for the study of blind deconvolution. In the present section, we will show the theoretical motivation of this project. Although the blind deconvolution problem, for the case of MP systems, has been well

understood, and the well known predictive deconvolution technique can fulfill the task, further research on the minimum phase assumption based techniques is still required. This is reflected in the work recorded in Chapter 3 of this thesis.

On the other hand, when systems are NMP (or mixed phase), in order to obtain the correct solution via blind deconvolution, both in amplitude and in phase, the higher order cumulants (HOC) (or HOS) of the system output must be investigated and analysed. This results from the fact that only the HOC can reflect the phase properties of signals (or systems). However, HOC analysis technology is still in its infancy in comparison with autocorrelation domain. Therefore, there are still many problems to be solved both theoretically and practically. A glimpse of this point can be obtained by taking parameter based NMP system deconvolution for example.

<1> MA model based techniques: Although several simple close-form formulae have been proposed [15][28]-[30], their practical results are often inaccurate, because it is much more difficult to estimate the higher order (>2) cumulants accurately than autocorrelation functions. Secondly, as a result of the nonlinearity associated with HOC, the problem of multimodality arises when the fitting scheme is used to determine the optimal solution [31].

<2> AR model based techniques: This type of methods are preferred when restoration is the only objective. But the current algorithms employ both diagonal and non-diagonal slices of the cumulants [32][14]. This results in two problems:

i) The forms of the algorithms become over-complicated; and

ii) the algorithm accuracy may be degraded.

As in the autocorrelation domain, parametric approaches offer the possible advantages of lower variance and higher resolution over nonparametric techniques. Therefore, it becomes vitally necessary to overcome the above problems associated with the parametric methods so that the advantages can be brought into full play. Undoubtedly, the improvements in the theoretic algorithms will greatly benefit the practical applications of blind deconvolution. Based on the above points, HOC based parametric blind deconvolution was chosen as the focal point of this project.

1.4 Layout and Organisation of Thesis

This thesis deals primarily with parametric blind deconvolution algorithms and their applications. The basic concepts, application areas, and the motivation of this project have been concisely illuminated in the foregoing three sections. The following chapters mainly present the design of new parametric blind deconvolution algorithms and improvement of existing ones; and the application of these algorithms to data communications (channel equalisation).

Chapter 2 presents a review of the current blind deconvolution techniques: both off-line and on-line algorithms. The advantages and disadvantages of these algorithms are discussed. Following on from this, HOC (or HOS) analysis is introduced, a special emphasis being placed on the phase sensitivity of HOC.

Chapter 3 presents a new robust blind deconvolution algorithm for MP systems: variance approximation and series decoupling (VASD). It is shown that this algorithm possesses some advantages over existing predictive techniques.

Chapter 4 begins the study of HOC based blind deconvolution. In this chapter, a two-step relay approach, based on a MA system model, is proposed in which the closed-form formula for MA parameters are combined with the HOC fitting schemes, and the problem of multimodality is partially overcome. In the meantime, the optimal identification of the MA model parameters of systems is realised as a result of optimal HOC fitting.

Chapter 5 continues with HOC based algorithm development. In this chapter, however, a noncausal AR system model is adopted for the convenience of signal restoration. Six new families of linear equations with respect to the AR parameters are derived. In comparison with the existing algorithms, only the diagonal slices of cumulants are employed. Simplicity and elegance are fully embodied in the obtained sets of linear equations.

Chapter 6 is devoted to the application of the algorithms obtained in above chapters to the equalisation of multilevel pulse amplitude modulation (PAM) signals, which have been transmitted through NMP channels. In order to meet the requirements of communications, an adaptive version of the algorithms is developed by the novel use of the successive over-relaxation (SOR) iteration scheme. A fast convergence rate is realised. As a result, these adaptive algorithms are capable of dealing with not only linear time invariant (LTI) but also linear time variant (LTV) channels.

Chapter 7 summarises the research work recorded in this thesis, followed by an indication of further possible directions of the blind deconvolution research.

Chapter 2:

BLIND DECONVOLUTION (EQUALISATION) ALGORITHMS

2.1 Introduction

As pointed out in the previous chapter, blind deconvolution (or equalisation) cannot be realised under the absolutely blind condition that only noisy system output $y(k)$ is known. Thus, some assumptions for $x(k)$ and $h(k)$ (or even $n(k)$) are necessary. Over last three decades, a great number of blind deconvolution algorithms have been proposed in different fields. This proliferation of algorithms is a result of two key points. Firstly, the assumptions which can be accepted vary from application to application. Secondly, from the point of view of applicability, the more relaxing the assumptions to which an algorithm is subject, the wider the range of applications to which the algorithm applies. In order to illuminate the current situation and existing problems of blind deconvolution technology, we will in this chapter survey various existing blind deconvolution (equalisation) algorithms.

Since the phase property of a system plays a fundamental role in derivation of most algorithms, the concepts of minimum phase (MP), maximum phase, and mixed phase are clarified in section 2.2. Predictive deconvolution used to be the dominant scheme, and is still an important part of many commercial seismic data processing packages. Consequently, all of section 2.3 is devoted to this pioneering technique. To overcome the main constraint with predictive deconvolution--minimum phase (MP), many other methods have been suggested by the researchers in two main fields: geoexploration (seismic data processing) and telecommunications (channel equalisation). Seismic data oriented (off-line or block) methods are scanned and compared in section 2.4, and channel equalisation (on-line or adaptive) algorithms in section 2.5.

Although the foregoing techniques are indeed effective for some specific cases, the blind deconvolution problem of NMP systems still remained far from being completely solved. In the 1980's, however, a powerful tool appeared: higher-order cumulant (HOC) analysis. The reason for this is that HOC contain system (signal) phase information. In section 2.6, basic theory of higher-order cumulant analysis is introduced, and in section 2.7, the phase sensitivity of HOC is studied in detail.

2.2 Phase Properties of Systems

In terms of phase properties, all linear systems fall into three categories: *minimum phase (MP), maximum phase, and mixed phase systems*. The phase properties of systems comprise one of the bases of the derivation and discussion in this thesis, consequently the definitions of the above three categories of systems are briefly demonstrated in this section according to various criteria: zero distribution, phase-frequency response, and energy evolution [6][34][35].

<1> Zero distribution:

This is the most commonly used criterion for system phase properties, especially in parameter based system theory. Let us assume the transfer function (in z-domain) of a stable and causal system to be

$$H(z) = \frac{\sum_{i=0}^{q} b_i z^{-i}}{1+\sum_{i=1}^{p} a_i z^{-i}} = \frac{b_0 \cdot \prod_{i=1}^{q}(1-u_i z^{-i})}{\prod_{i=1}^{p}(1-v_i z^{-i})} , \qquad (2.2.1)$$

where u_i and v_i denote the zeros and poles of the system, respectively. Then we have the following definitions:

i) If and only if all of the zeros, u_i, are located inside the unit circle in the z-plane, the system is termed a *minimum phase* (MP) *system*.

ii) If and only if all of the zeros, u_i, are located outside the unit circle in the z-plane, the system is termed a *maximum phase system*.

iii) If and only if only some of the zeros, u_i, are located inside the unit circle, but the rest of them are not, the system is termed a *mixed phase system*.

In addition, maximum phase and mixed phase are jointly referred to as *non-minimum phase* (NMP).

As will be shown in the later sections, the zero distribution, *i. e.*, the phase property, of a system determines the causality of its inverse system (or the invertibility of the system itself). This fact becomes vitally important in system identification and deconvolution.

<2> Phase-frequency response:

From Eq(2.2.1), the frequency response of the system can be expressed as

$$H(e^{j\omega}) = H(z)\big|_{z=e^{j\omega}} = A(\omega)e^{-j\phi(\omega)} , \qquad (2.2.2)$$

where $A(\omega)$ and $-\phi(\omega)$ represent the amplitude and phase frequency response of system $H(z)$, and ω is the normalised angular frequency. Then, a system which satisfies the following condition is called MP system: *It has the least phase lag $\phi(\omega)$ among all causal systems with the same amplitude frequency response $A(\omega)$*. Otherwise, the system is NMP. Obviously, this description indicates the origin of such terms as minimum phase, maximum phase, and mixed phase.

<3> Energy evolution:

Let the unit impulse response of system $H(z)$ be $\{h(k)|\ k=0,\ 1,\ ...\}$. The system energy evolution [34] is defined as

$$E(N) = \sum_{k=0}^{N} h^2(k) \quad , \tag{2.2.3}$$

and the system energy delay [6] as

$$D(N) = \text{total system energy} - E(N) \quad . \tag{2.2.4}$$

Considering Eq(2.2.3) and

$$\text{total system energy} = \sum_{k=0}^{\infty} h^2(k) \quad , \tag{2.2.5}$$

system energy delay in Eq(2.2.4) can be written as

$$D(N) = \sum_{k=N+1}^{\infty} h^2(k) \quad , \tag{2.2.6}$$

Then, *among the systems* $\{h_j(k)|\ j \in J,\ k=0,\ 1,\ ...\}$ *with the same total energy, the system* $\{h_m(k)|\ k=0,\ 1,\ ...\}$ *which satisfies the condition below is MP:*

$$D_m(N) = \min_{j \in J}\{D_j(N)\} \quad . \tag{2.2.7}$$

Otherwise, *the system is NMP.* This definition bestows upon MP (or NMP) system another name: *minimum (or nonminimum) delay system*, which is mainly used in geoexploration.

It should be noted that, the energy evolution $E_m(N)$ of system $h_m(k)$ must be maximum among $\{h_m(k)|\ j \in J\}$ if $h_m(k)$ is MP. This fact indicates that the energy of a MP system is more concentrated at the "front end" than any other systems in the whole family of systems with the same total energy (or autocorrelation).

Above, we have described three criteria for system phase properties. In fact, any of the above three criteria can equally serve as the definition of MP (or NMP), and each of the remaining ones is then a necessary and sufficient condition for MP (or NMP) [6]. As a result, we will directly use them without explanation according to the need of discussion proceeding from next section.

2.3 Predictive Deconvolution

Predictive deconvolution, probably the earliest blind deconvolution approach, was invented and developed by some geophysicists in 1950's. It is based mainly on Wiener optimal filtering theory and Robinson "statistical" seismic model [3]-[7][37]. As mentioned at the beginning of this chapter, predictive deconvolution has played a particularly significant role in seismic geoexploration. In this section, we will present the main points of this pioneering technique.

First, let us study the design of a linear predictor, which has a straightforward relation with the design of the inverse filter (deconvolver) used in predictive deconvolution. Consider signal $y(k)$, $k=0,\ 1,\ ...$, which is the output of a linear system H. Let $\hat{y}(k+m|k)$ be the predictive estimate of $y(k+m)$ obtained by using data segment $\{y(k),\ y(k-1),\ ...,\ y(k-L)\}$ and a linear predictor, *i. e.*,

$$\hat{y}(k+m|k) = \sum_{l=0}^{L} p_l y(k-l) \quad , \tag{2.3.1}$$

where L is the length of the predictor $\{p_l | \, l=0, 1, \ldots, L\}$, and m is prediction distance. According to Wiener optimal filtering theory, the tap coefficients in Eq(2.3.1) can be determined by minimising the following cost function:

$$\epsilon^2 = \frac{1}{N}\sum_{k=0}^{N}[\hat{y}(k+m|k)-y(k+m)]^2 \quad , \tag{2.3.2}$$

where N is the data record length. Application of optimisation theory to Eq(2.3.2), using Eq(2.3.1), immediately leads to the following equation:

$$\sum_{l=0}^{L} p_l \Phi_{yy}(l-j)=\Phi_{yy}(j+m) \quad , \quad j=0, 1, \ldots, L \, , \tag{2.3.3}$$

where, Φ_{yy} is the autocorrelation function (ACF) of $y(k)$, and can be calculated by

$$\Phi_{yy}(n) = \frac{1}{N}\sum_{k=n}^{N} y(k)y(k-n) \quad . \tag{2.3.4}$$

It is easy to find that filter $\{p_l \, | \, l=0, 1, \ldots, L\}$ depends only on the autocorrelation, not on the phase property, of $y(k)$. Considering $\Phi(-n)=\Phi(n)$, Eq(2.3.4) can be rewritten as the more compact matrix form:

$$\mathbf{A}_{(L+1)\times(L+1)}\mathbf{P}_{L+1} = \mathbf{B}_{L+1} \quad , \tag{2.3.5}$$

where

$$\mathbf{A}_{(L+1)\times(L+1)} = \begin{bmatrix} \Phi_{yy}(0) & \Phi_{yy}(1) & \cdots & \Phi_{yy}(L) \\ \Phi_{yy}(1) & \Phi_{yy}(0) & \cdots & \Phi_{yy}(L-1) \\ \cdots & \cdots & \cdots & \cdots \\ \Phi_{yy}(L) & \Phi_{yy}(L-1) & \cdots & \Phi_{yy}(0) \end{bmatrix} \quad , \tag{2.3.6a}$$

$$\mathbf{P}_{L+1} = [p_0, p_1, \ldots, p_L]^T \quad , \tag{2.3.6b}$$

and

$$\mathbf{B}_{L+1} = [\Phi_{yy}(m), \Phi_{yy}(m+1), \ldots, \Phi_{yy}(m+L)]^T \quad . \tag{2.3.6c}$$

Let us notice that $\mathbf{A}_{(L+1)\times(L+1)}$ in Eq(2.3.6a) is a symmetric Toeplitz matrix, consequently Eq(2.3.5) can be efficiently solved by using the well known Levinson-Durbin recursive algorithm [5][37]-[40].

Now, let the prediction distance $m=1$, then Eq(2.3.3) becomes

$$\sum_{l=0}^{L} p_l \Phi_{yy}(l-j)=\Phi(j+1) \quad , \quad j=0, 1, \ldots, L. \tag{2.3.7}$$

Shifting the left hand side terms to the right hand side, we can also rewrite the above equation as

$$\Phi_{xx}(j+1) - \sum_{l=0}^{L} p_l \Phi_{yy}(l-j) = 0 \quad , \quad j=0, 1, \cdots, L. \tag{2.3.8}$$

Let

$$\Phi_{yy}(0) - \sum_{l=0}^{L+1} p_l \Phi_{yy}(l) = \beta \ , \tag{2.3.9}$$

by combining Eq(2.3.8) with Eq(2.3.9), we can obtain the augmented form of Eq(2.3.5):

$$\mathbf{A}_{(L+2)\times(L+2)}\mathbf{P}_{L+2} = \mathbf{B}_{L+2} \ , \tag{2.3.10}$$

where

$$\mathbf{A}_{(L+2)\times(L+2)} = \begin{bmatrix} \Phi_{yy}(0) & \Phi_{yy}(1) & \cdots & \Phi_{yy}(L+1) \\ \Phi_{yy}(1) & \Phi_{yy}(0) & \cdots & \Phi_{yy}(L) \\ \cdots & \cdots & \cdots & \cdots \\ \Phi_{yy}(L+1) & \Phi_{yy}(L) & \cdots & \Phi_{yy}(0) \end{bmatrix} \ , \tag{2.3.11a}$$

$$\mathbf{P}_{L+2} = [1, -p_0, -p_1, \ldots, -p_L]^T \ , \tag{2.3.11b}$$

and

$$\mathbf{B}_{L+2} = [\beta, 0, \ldots, 0]^T \ . \tag{2.3.11c}$$

After completing the above predictor design, let us reconsider the deconvolution problem: Assume the input signal of the system, H, is $x(k)$, which is a zero-mean white noise series with variance σ^2, and the output signal of system is $y(k)$. Here, H is assumed to be a minimum phase (MP) system with impulse response $h(k)$. Our objective is to design an inverse filter $\{\theta_i \,|\, i=0, 1, \ldots, L+1\}$ whose output $c(k)$ approximates $x(k)$ in the Wiener (or least square) sense, *i. e.*,

$$c(k) = \sum_{i=0}^{L+1} \theta_i y(k-i) \approx x(k) \ , \tag{2.3.12}$$

where the θ_i's are determined by minimising the cost function

$$\epsilon^2 = \frac{1}{N}\sum_{k=0}^{N}[c(k)-x(k)]^2 \ . \tag{2.3.13}$$

Similarly, by applying optimisation theory to the above expression , the following equation can be derived:

$$\mathbf{A}_{(L+2)\times(L+2)}\Theta = \mathbf{F} \ , \tag{2.3.14}$$

where

$$\Theta = [\,\theta_0, \theta_1, \ldots, \theta_{L+1}]^T \ , \tag{2.3.15a}$$

and

$$\mathbf{F} = [h(0)\sigma^2, 0, \ldots, 0]^T \ . \tag{2.3.15b}$$

Note that Eq(2.3.14) is identical to Eq(2.3.10) in form apart from the constant $h(0)\sigma^2$, but the latter is not essential since it only affects the amplitude of output signal of the inverse filter $\{\theta_i \,|\, i=0, ,1, \ldots, L+1\}$. Thus, we can conclude that, for a zero-mean white input signal, the inverse filter of a system is completely determined by the one-step-forward predictor of the output signal. In other words, as long as we obtain the prediction filter $\{p_i \,|\, i=0, 1, \ldots, L\}$, we can immediately determine the inverse filter:

$\{1, -p_0, -p_1, \ldots, -p_L\}$. In fact, it is not strictly necessary to let $y(k)$ pass through $\{1, -p_0, -p_1, \ldots, -p_L\}$ after $\{p_0, p_1, \ldots, p_L\}$ is determined, since we can obtain the deconvolution result directly from the predictive filter as shown below.

Notice that the deconvolution result

$$c(k) = y(k) - \sum_{l=0}^{L} p_l \, y(k-l-1) = y(k) - \hat{y}(k|k-1) \, . \tag{2.3.16}$$

From this equation, it can be easily seen that the one-step ahead prediction error is exactly the deconvolution result! Therefore, we can construct the inverse filter as shown in Fig(2.1).

A physical explanation of predictive deconvolution is as follows: Since $x(k)$ is white noise, it is completely unpredictable. But after passing through the linear system, it contains the predictable information of the system operator. Since we can only predict the predictable part of the output $y(k)$, the error must be the unpredictable part of $x(n)$, which is the system input signal (or the innovation series).

In order to confirm the predictive deconvolution technique, the following simple example is presented. Assume the transfer function of a system is

$$H(z) = \frac{(1-0.5z^{-1})(1-0.1z^{-1})}{1+0.2z^{-1}} \, , \tag{2.3.17}$$

and the input signal $x(k)$ is a uniformly distributed pseudo-white-noise series, as shown in Fig(2.2), and the noisy output signal $y'(k)$ is synthesised from

$$y(k) = -0.2y(k-1)+x(k)-0.6x(k-1)+0.05x(k-2) \, , \tag{2.3.18a}$$

and

$$y'(k) = y(k)+n(k) \, , \tag{2.3.18b}$$

where $n(k)$ is Gaussian white noise. The deconvolved signal $c(k)$ is then determined according to the predictive deconvolution method described above. Fig(2.3a) illustrates the deconvolution result in the presence of 30 dB additive noise, and Fig(2.3b) is the corresponding error waveform $e(k)=c(k)-x(k)$. In the simulation, the predictor length $L=10$ and the data record length $N=1500$, but only the first 150 samples are plotted in the figures and similarly thereafter.

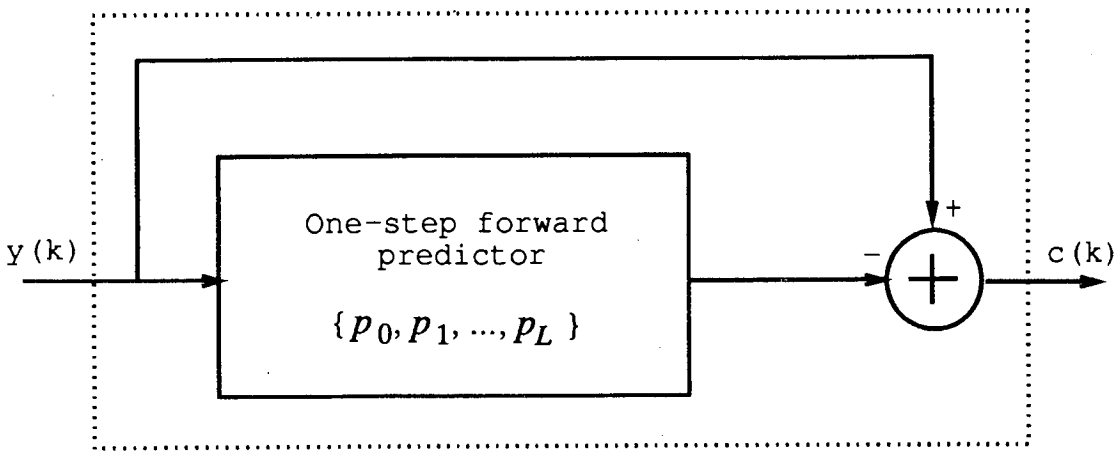
Obviously, the above results are reasonablely satisfactory, and in fact, predictive deconvolution was the main work horse of statistical wavelet deconvolution, and is still currently an important routine in geoexploration data processing [41][6][53]. This technique has the advantages of high stability and ease of implementation as long as the prerequisite assumptions are satisfied.

The main assumptions of predictive deconvolution are:

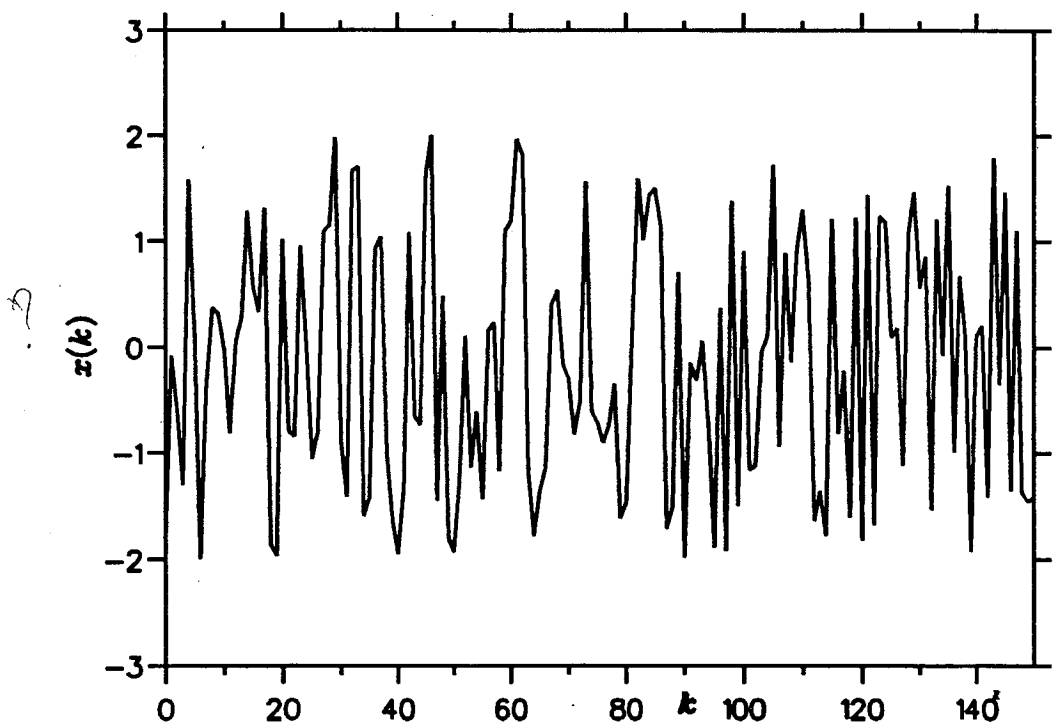
1) The input signal is zero-mean and white, and

2) the system transfer function (or wavelet) is minimum phase.

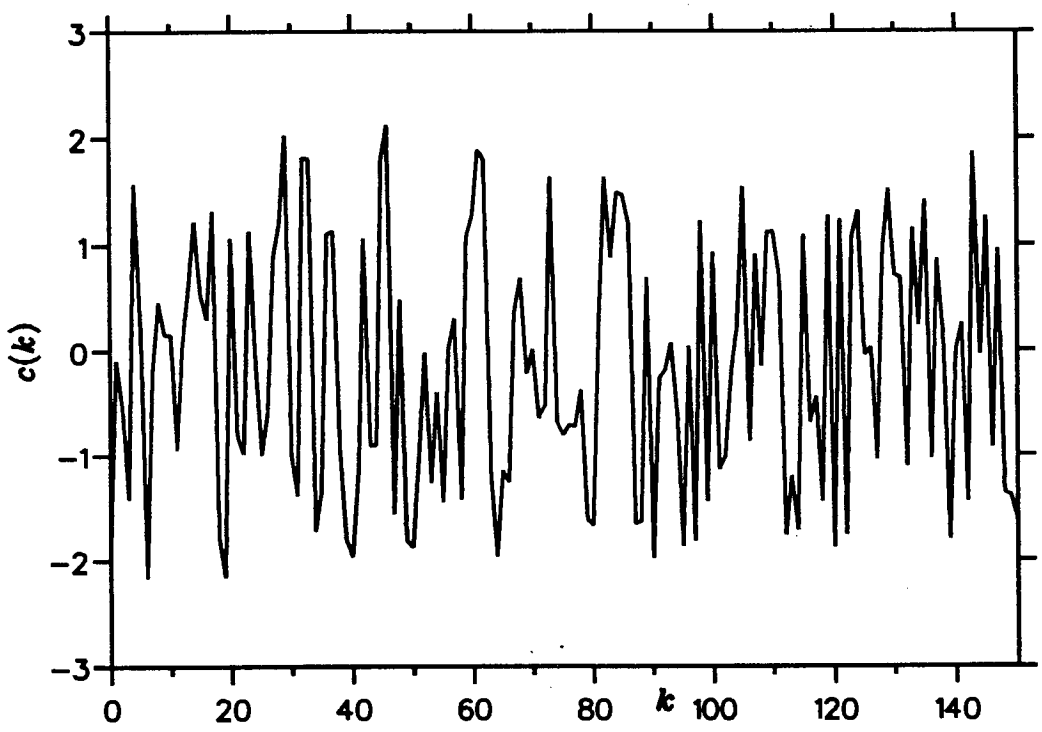
The former has proved to be normally reasonable. The latter, unfortunately, is not reasonable in many cases. This fact results in the failure of predictive deconvolution for NMP



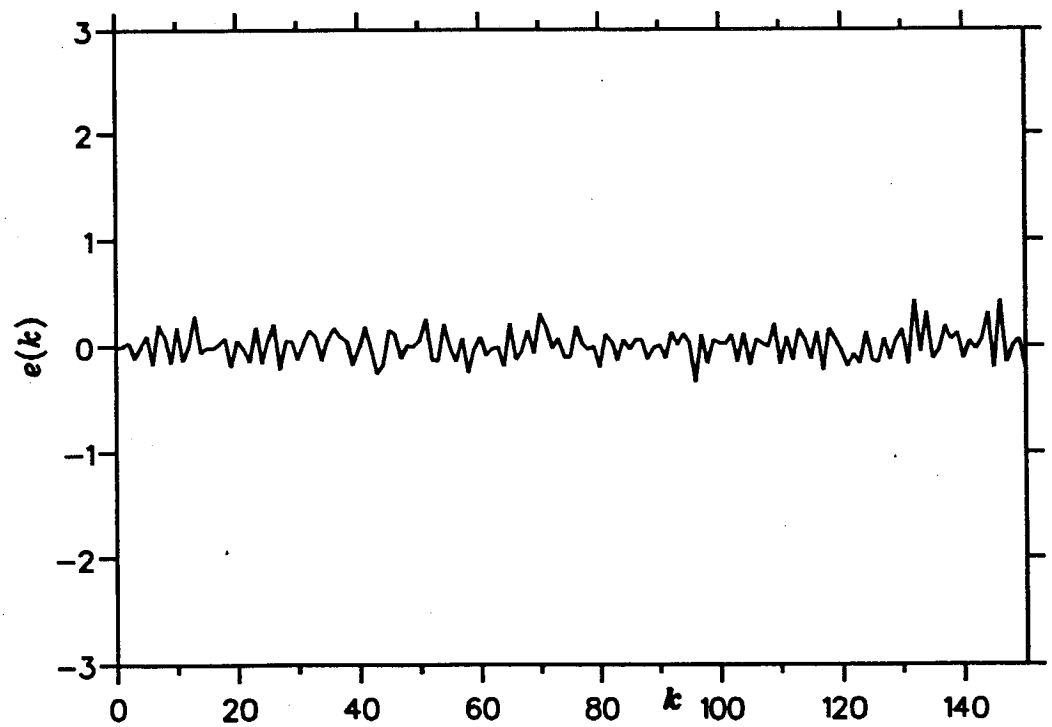
Fig(2.1) Predictive deconvolver.



Fig(2.2) A segment of the input signal x(k).



Fig(2.3a) Deconvolved signal c(k).



Fig(2.3b) Deconvolution error e(k).

systems. To obtain a clearer view about the effect of this assumption, we present another example below.

The input signal is the same as in previous example. But this time the transfer function of system is

$$H(z) = \frac{(1-2z^{-1})(1-0.1z^{-1})}{1+0.2z^{-1}} \quad . \tag{2.3.19}$$

Fig(2.4a) is the deconvolution result (noise-free case). Clearly, it is completely different from the input signal in Fig(2.2). The error waveform is shown in Fig(2.4b). Thus, the predictive deconvolution fails in this case.

Why can the predictive deconvolution method give correct results only in the MP case? Simply speaking, the reasons are: 1) Wiener's theory is based only on the second-order cumulants (autocorrelations) of the time series, and the latter is not able to reflect the phase characteristic of the series, and 2) when system is NMP, its causal inverse filter is unstable, but the predictive method always gives a stable MP solution which has the same second-order cumulants as the true input series. From the point of view of frequency domain, predictive deconvolution always generates the spectrally equivalent MP solution. The more detailed analysis about phase sensitivity will be presented in the last section of this chapter.

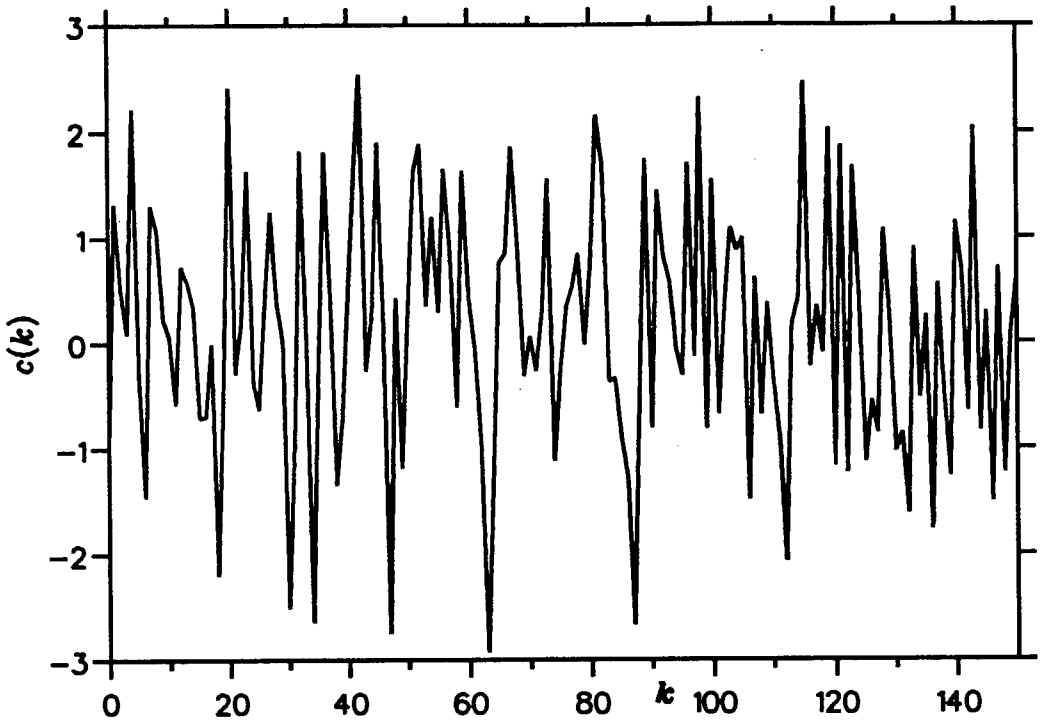
In order to relax the assumption of MP, *i.e.*, to obtain the correct deconvolution solution even in NMP case, many approaches have been proposed, which will now be reviewed.

2.4 Blind Deconvolution of Seismic Data with Nonminimum Phase Wavelets

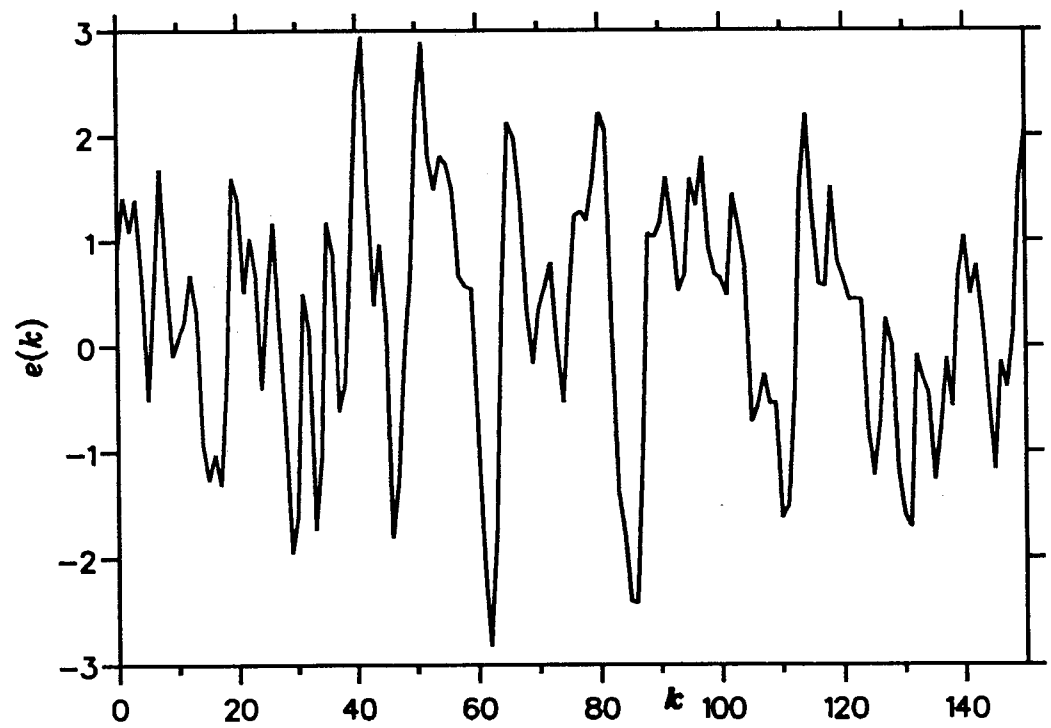
As discussed in the previous section, the predictive deconvolution model assumed MP systems (or wavelets) for which the corresponding causal inverses exist and consequently can be used to recover the system input series. Since the MP assumption may not always be valid, it is important to find alternative blind deconvolution (equalisation) techniques which will remove (or relax) the assumption. In fact, many such techniques have been published in the literature. In this section, some of the important seismic reflection data oriented (*i.e.*, off-line or block-type) approaches are reviewed, including homomorphic deconvolution, maximum likelihood deconvolution, and minimum entropy deconvolution. The next section will be devoted to the equalisation (*i.e.*, on-line or adaptive) methods.

2.4.1 Homomorphic deconvolution

Homomorphic deconvolution, a nonlinear technique, was originally proposed in 1965 by Oppenheim as an application of generalised superposition theory [2][42][43]. Later at the end of 1960's, Oppenheim and Schafer applied this nonlinear approach to the design of an echo canceller [2][44]. Also, as described in Section 1.1 of this thesis, T. G. Stockham successfully applied this method to the improvement of voice quality [1]. Riad and Nahman realised the separation of TDR signals occurring in overlapping time windows by using homomorphic technique [24]. In early 1970's, Ulrych and Stoffa *et al* introduced homomorphic technique into the area of seismology [45]-[47], and found this approach



Fig(2.4a) Deconvolved signal c(k).



Fig(2.4b) Deconvolution error e(k).

offers the considerable advantage that no prior assumption about the phase nature of seismic wavelet (system) need to be made, and it also does not require the random distribution assumption about the earth reflectivity series (system input). The main points of homomorphic deconvolution are now briefly introduced. In addition, since our current context is deconvolution, only the signals combined by means of convolution are considered here.

The key step in homomorphic deconvolution is to find such a system D, termed the *characteristic system* here, that D and its inverse D^{-1} can form the canonical representation shown in Fig(2.5). In our deconvolution case, it has been shown [42] that D should be some transformation from a convolutional space to an additive space. Mathematically, system D is defined by the relation [2][42][45]:

$$D[\,{}^{(a)}x_1(k) * {}^{(b)}x_2(k)] = aD[x_1(k)] + bD[x_2(k)] \quad , \tag{2.4.1}$$

where (a) and (b) denote scalar multiplication, $x_1(k)$ and $x_2(k)$ are the two convolution signals. The system L is a linear system, and the system D^{-1}, the inverse of D, is also a homomorphic system, and performs the inverse transformation (from an additive space back to the output convolutional space). Let us notice that the configuration in Fig(2.5) has great flexibility: once D has been determined, it remains fixed for all deconvolution problems.

Considering that the Fourier transform of two convolved signals is equal to the product of their Fourier transforms and a logarithmic operation can transform multiplication to addition, system D may be accomplished as illustrated in Fig(2.6a), *i.e.*,

$$\bar{x}(k) = IDFT[\log(DFT[x(k)])] \quad , \tag{2.4.2}$$

where $DFT[\cdot]$ represents the discrete Fourier transform, and $IDFT[\cdot]$ the inverse. In Eq(2.4.2) k has time unit, and has been termed *quefrency*, and $\bar{x}(k)$ has been termed the *complex cepstrum* [48][2]. Correspondingly, the inverse system D^{-1} in Fig(2.5) can be realised by

$$y(k) = IDFT[\exp(DFT[\bar{y}(k)])] \quad , \tag{2.4.3}$$

which can also be canonically represented as shown in Fig(2.6b). Fig(2.5) and Fig(2.6a,b) comprise the basis of homomorphic deconvolution.

Consequently, if we have

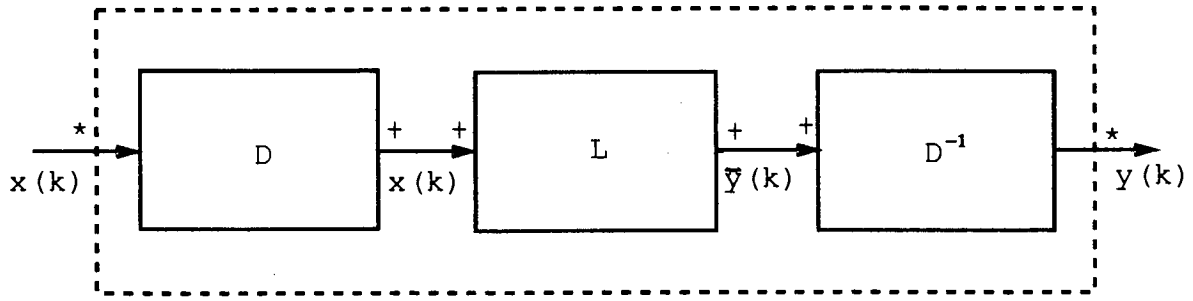
$$x_3(k) = x_1(k) * x_2(k) \tag{2.4.4}$$

in the time domain, we will have

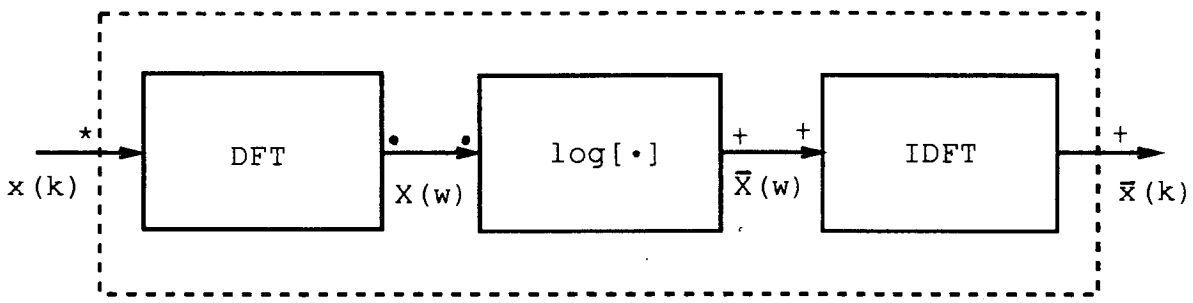
$$\bar{x}_3(k) = \bar{x}_1(k) + \bar{x}_2(k) \tag{2.4.5}$$

in the quefrency domain. Under ideal conditions, $\bar{x}_1(k)$ and $\bar{x}_2(k)$ can be separated by using a low-pass and a high-pass filter. The inherent assumption here is that wavelet--either $s_1(k)$ or $x_2(k)$--has such a smooth spectrum that its cepstrum can be forced into small values of quefrency (or a narrow quefrency band) [49].

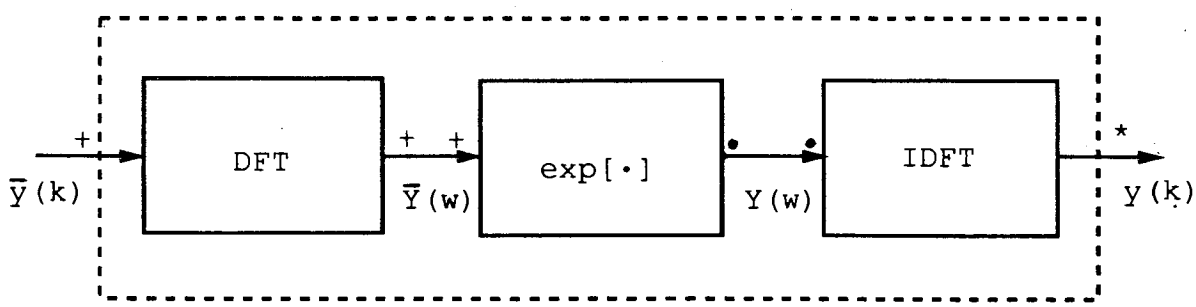
It is clear that, in principle, the above scheme does not depend upon any assumption about the system phase properties (MP or NMP) or the randomness of the input signals. It



Fig(2.5) Canonic representation for homomorphic deconvolution.



(a) System D.



(b) System D^{-1}.

Fig(2.6) Canonic representation for system D and D^{-1}.

only requires that the complex cepstra of two signals convolving together with each other occupy different quefrency bands. That is, the cepstra of the two signals should be well-separated enough so that a comb filter can be employed. In practice, the seismic data from certain areas do indeed satisfy the above condition. For example, Ulrych *et al* obtained very encouraging results in the application of homomorphic deconvolution to a series of teleseismic events recorded at Leduc, Alberta [45]. Also, Buhl and Stoffa *et al* successfully applied homomorphic deconvolution to the data which were recorded on the Argentine continental shelf near Bahia Blanca [47].

However, homomorphic deconvolution also has several limitations. In the first place, the logarithmic operation to complex numbers (spectrum of $x(k)$) involves the calculation of the arctangent. Due to the multivalued nature of the arctangent, a process normally referred to as phase unwrapping (or phase unfolding) has to be introduced. Unfortunately, the unique unwrapping of the phase curve has not been completely solved since all existing methods are complicated by a sensitivity to measurement noise in the system output (*e.g.*, seismogram) [49]. But we did not consider this point in Eq(2.4.4) to aid the simplicity of the discussion. Secondly, in the case of badly separated cepstrum, although the employment of exponential weighting procedure may improve the condition, the choice of the exponential factor is based only on experience (or trial and error), and must be carefully handled. Third, homomorphic deconvolution is usually not robust to measurement noise, and again some trial-and-error based methods have to be used to reduce the effect of noise as much as possible. In fact, there has rarely been any report on the application of homomorphic deconvolution to land data in seismology [49][54].

As a result, while the homomorphic deconvolution still has promise, it has not solved the problem of identifying the system (wavelet) quantitatively and reliably on real data (especially the real seismic data). Many questions remain to be explored.

In summary, although homomorphic deconvolution does not require the MP assumption about the system, it does require an assumption that the system input signal separates well from the impulse response of the system in the quefrency domain.

2.4.2 Maximum likelihood deconvolution

The principal inventors of the maximum likelihood deconvolution (MLD) are Mendel and his numerous students [11]. The basic concept in MLD is a time-domain state-variable model of the seismogram (or system output). Mathematically, this model can be described as follows [11][49].

Assume the seismogram is modelled as

$$y(k) = \sum_{i=1}^{k-1} x(i)h(k-i)+n(k) \quad , \tag{2.4.6}$$

which is equivalent to Eq(1.1.2) except that $h(0)=0$ and $i=1,2, \ldots, N$ here. For the wavelet $h(k)$, an ARMA model

$$H(z) = \frac{b_1 z^{-1}+b_2 z^{-2}+ \cdots +b_n z^{-n}}{1+a_1 z^{-1}+ \cdots +a_n z^{-n}} \tag{2.4.7}$$

is assumed, where $\mathbf{a}=[a_1, a_2, \ldots, a_n]^T$ and $\mathbf{b}=[b_1, b_2, \ldots, b_n]^T$ are the parameters of the

wavelet. Then the state-variable expression for $y(k)$ can be derived:

$$\mathbf{u}(k+1) = \mathbf{A}\mathbf{u}(k)+\mathbf{g}x(k) \ , \tag{2.4.8a}$$

$$y(k) = \mathbf{f}^T\mathbf{u}(k)+n(k) \ , \tag{2.4.8b}$$

$$\mathbf{u}(0) = \mathbf{0} \ , \tag{2.4.8c}$$

where

$$\mathbf{A} = \left[\begin{array}{c|c} \begin{matrix}0\\0\\\cdot\\\cdot\\\cdot\\0\end{matrix} & \mathbf{I}_{n-1} \\ \hline -a_n \quad -a_{n-1} \quad \cdots \quad -a_1 & \end{array}\right] \ , \tag{2.4.9a}$$

$$\mathbf{g} = [0, 0, \ldots, 1]^T \ , \tag{2.4.9b}$$

and

$$\mathbf{f} = [b_n, b_{n-1}, \ldots, b_1]^T \ . \tag{2.4.9c}$$

In Eq(2.4.9a), $\mathbf{I}_{n-1}$ represents the $(n-1)\times(n-1)$ unit matrix.

In the actual procedure, the MLD requires the following assumptions.

1) Measurement noise $n(k)$ is Gaussian and white, and

$$\mathrm{E}[n^2(k)] = \sigma_n^2 \ . \tag{2.4.10}$$

2) The system input sequence $x(k)$ (reflectivity) satisfies a Bernoulli-Gaussian (BG) distribution, and can be expressed in the product form:

$$x(k) = r(k)q(k) \ . \tag{2.4.11}$$

In this model, $q(k)$ is a Bernoulli sequence with parameter λ [50]:

$$Pr[q(k)] = \begin{cases} \lambda & q(k)=1 \\ 1-\lambda & q(k)=0 \end{cases} \ , \tag{2.4.12}$$

and $r(k)$ is a zero mean Gaussian white noise sequence with variance σ_r^2. $q(k)$ and $r(k)$ are statistically independent. Under this assumption, it can be easily proved that $x(k)$ is also a white sequence, and

$$\mathrm{E}[x^2(k)] = \lambda\sigma_r^2 \ . \tag{2.4.13}$$

According to Mendel [11], the physical interpretation of the product model in Eq(2.4.11) is that all of the reflection amplitude information is contained in the "amplitude" sequence $r(k)$ and all of the reflection location information is contained in the "event" sequence $q(k)$. As a result, we can estimate the reflection amplitudes and locations separately.

Let us assume $\hat{r}(k)$ and $\hat{q}(k)$ denote the maximum likelihood estimates for $r(k)$ and $q(k)$, respectively, then the *invariance property* of maximum likelihood estimation allows us

to compute the maximum likelihood estimates of $x(k)$, *i.e.* $\hat{x}(k)$, as

$$\hat{x}(k) = \hat{r}(k) * \hat{q}(k) \quad . \tag{2.4.14}$$

Now, in order to determine $\hat{x}(k)$ in the above equation, we must find the maximum likelihood estimates of the following unknowns:

1) The wavelet ARMA parameters b_i's and a_i's, and the statistical parameters σ_n^2, σ_r^2, and λ. These deterministic quantities can be expressed as a parameter vector θ:

$$\theta = [a_1, \cdots, a_n, b_1, \cdots, b_n, \sigma_{n_2}, \sigma_r^2, \lambda]^T \quad . \tag{2.4.15}$$

2) The sequence $r(k)$ and $q(k)$. They can also be written as the corresponding vector forms:

$$\mathbf{r} = [r(1), r(2), \ldots, r(N)]^T \tag{2.4.16}$$

and

$$\mathbf{q} = [q(1), q(2), \ldots, q(N)]^T \quad , \tag{2.4.17}$$

respectively. In addition, we denote

$$\mathbf{y} = [y(1), y(2), \ldots, y(N)]^T \quad . \tag{2.4.18}$$

To obtain the maximum likelihood estimates of $\mathbf{q}$ and θ, namely $\hat{\mathbf{q}}$ and $\hat{\theta}$, Kormylo and Mendel [51][11] proposed the following likelihood function:

$$L\{\hat{\mathbf{r}}, \mathbf{q}, \theta \,|\, \mathbf{y}\} \propto p(\hat{\mathbf{r}}, \mathbf{y} \,|\, \mathbf{q}, \theta) Pr(\mathbf{q} \,|\, \theta) =$$

$$(2\pi)^{-N} (\sigma_r^2 \sigma_n^2)^{-N/2} \exp(-\mathbf{y}^T \Omega_{\theta q}^{-1} \mathbf{y}/2) Pr(\mathbf{q} \,|\, \theta) \quad . \tag{2.4.19}$$

By maximising Eq(2.4.19), $\hat{\mathbf{q}}$ and $\hat{\theta}$ can be determined. As to $\hat{\mathbf{r}}$, it can be computed from

$$\hat{\mathbf{r}} = \hat{\sigma}_r^2 \mathbf{Q}_{\hat{q}} \mathbf{V}_{\hat{\theta}}^T \Omega_{\hat{\theta}\hat{q}}^{-1} \mathbf{y} \quad , \tag{2.4.20}$$

where $\mathbf{Q}_q$, $\mathbf{V}_\theta$ and $\Omega_{\theta q}$ are matrices related to $q(k)$, $h(k)$, σ_r^2, and σ_n^2 [11].

When the MLD is applied to real data, the reflectivity is further assumed to consist of a small number of large spikes superimposed on a background of smaller spikes (termed as backscatter noise) [11][52][49]. The large spikes sequence has been modelled as the product form as described above. For the more detailed discussion about MLD, see [11] or [52].

It can be concluded that MLD in essence possesses the following advantages: (*i*) it is based on a nonGaussian reflectivity model (BG model); (*ii*) it uses a parsimonious wavelet model, but does not require the MP assumption; (*iii*) it is a high-resolution technique not only for the broad band data but also for narrow band data; (*iv*) it employs nonlinear signal processing procedure which is necessary because of (*i*).

The main disadvantage of MLD is in the level of computational complexity.

Finally, it should be pointed out that, since MLD requires a very strict assumption, namely BG distribution, for reflectivity, and requires considerable computation, it has not developed as rapidly as originally expected over the last ten years. Another factor which results in the current situation of MLD is probably the appearance of higher order cumulant (HOC) based deconvolution techniques which offer far more advantages than MLD.

HOC based deconvolution is the primary topic of this thesis.

2.4.3 Minimum entropy deconvolution

As discussed above, predictive deconvolution requires the assumptions of a MP wavelet and IID series. The success of homomorphic deconvolution depends upon whether the wavelet cepstrum is separable from that of the reflectivity series. In addition, homomorphic deconvolution is rather sensitive to the presence of additive noise. The maximum likelihood deconvolution requires a Bernoulli-Gaussian distribution for the reflectivity series. Since all these requirements can be far too restrictive in many instances, a radically different alternative called minimum entroy deconvolution (MED) was presented by Wiggins in 1977 [55][56]. The reason why this scheme is so named is that it seeks to minimise the entropy or randomness of the data. In practical terms, minimising of entropy of the data is equivalent to characterising the reflectivity by only a relatively few large spikes, *i.e.*, equivalent to simplifying the data structure, since minimum entropy is an expression of simplicity or certainty [57][49]. As a result, Wiggins adopted a mathematical measure of simplicity (or parsimony) as the norm (cost function) instead of directly using the entropy itself. Wiggins' MED technique can be briefly described as follows [55]-[57][49]. The case of multitrace is considered here.

Let us denote the seismogram as $y(i, j)$, $i=1, 2, \cdots, M$, $j=0, 1, \cdots, N$, where M is the number of traces and N the number of data samples per trace. If the corresponding inverse filter for deconvolution is $\{\theta_l \,|\, l=-L, \cdots, 0, \cdots, L\}$, then

$$c(i, j) = \sum_{l=-L}^{L} \theta_l y(i, j-l) \tag{2.4.21}$$

is the output of the inverse filter. Considering the simple structure that $c(i, j)$ is desired to have, the following Varimax norm (cost function) can be employed:

$$V = \sum_{i=1}^{M} V_i \quad , \tag{2.4.22}$$

where

$$V_i = \frac{\sum_{j=0}^{N} c^4(i, j)}{[\sum_{j=0}^{N} c^2(i, j)]^2} \quad . \tag{2.4.23}$$

Notice that, if $c(i, j)$ were a single unit spike (impulse) for every i, V_i's would be unity. As the number of spikes increases, the V_i's decrease. So does V. Naturally, to exploit the simplicity of $c(i, j)$, it is only necessary to maximise V with respect to θ_l's. According to the least-square optimisation principle, we must have

$$\frac{\partial V}{\partial \theta_k} = 0 \quad , \quad k=-L, \cdots, L, \tag{2.4.24}$$

which leads to

$$\sum_{l=-L}^{L} \theta_l \sum_{i=1}^{M} V_i u_i^{-1} \sum_{j=0}^{N} y(i, j-l) y(i, j-k) = \sum_{i=1}^{M} u_i^{-2} \sum_{j=0}^{N} c^3(i, j) y(i, j-k) \; , \tag{2.4.25}$$

where $u_i = \sum_{j=0}^{N} c^2(i, j)$. Eq(2.4.25) can also be written in matrix form:

$$\mathbf{R}\theta = \mathbf{g} \quad , \tag{2.4.26}$$

where the meanings of $\mathbf{R}$, θ, and $\mathbf{g}$ are obvious. It can be seen that Eq(2.4.25) or Eq(2.4.26) is highly nonlinear with respect to θ_i's. Fortunately, $\mathbf{R}$ is a Toeplitz matrix. This fact suggests that an iterative solution of Eq(2.4.26) is very convenient. Consequently, the following algorithmic steps can be employed: *i*) assume an initial estimate for θ, and compute $\mathbf{R}$ and $\mathbf{g}$; *ii*) solve Eq(2.4.26) for θ using the well known Levinson recursive scheme; *iii*) recompute $\mathbf{R}$ and $\mathbf{g}$, return to step *ii*), and proceeding in this manner until convergence has been achieved.

Clearly, the strength of MED is that it does not require the assumption that the wavelet be MP, nor the IID feature for the reflectivity series, and it only assumes that the reflectivity consists of a relatively few large spikes (events). In practice, there are some geophysical environments which fit the above assumptions [49].

Like other techniques discussed so far, however, MED also has drawbacks [49][57][58]: 1) ambiguity of output signal polarity; 2) non-uniqueness of the solutions; 3) sensitivity to burst noise; 4) insensitivity to smaller spikes; and 5) heavy computational burden. In order to overcome or reduce the effects of these problems, various approaches have been adopted by researchers.

To improve MED's sensitivity to smaller spikes (events) and make it more generally applicable, Claerbout in 1977 proposed a modified version of Wiggins' MED: parsimonious deconvolution [59][58]. In his work, Claerbout developed an analytical relation between the adopted norm (simplicity criterion) and Shannon's entropy. Two years later, Ooe and Ulrych incorporated an exponential transformation into Wiggins' MED [57], and considerably enhanced the noise suppression characteristics of MED and improved the identification of smaller events (spikes) in the reflectivity series. In 1984, Cabrelli [60] geometrically analysed the Varimax norm adopted by Wiggins, and suggested a new simplicity criterion termed D norm. Based on this new norm, a noniterative algorithm was derived. It was shown that the algorithm produced output of greater simplicity than those generated by the "traditional" Wiggins' MED. Recently, Wang *et al* improved the efficiency of the D norm based scheme by introducing a fast algorithm [61]. Another big step forward in understanding MED-type techniques was achieved by Gray in 1979 [58]. In his approach called variable norm deconvolution, Gray linked the norm that he employed with the probability distribution of reflectivity series.

In addition, White opened another important avenue of study of the MED technique: use MED only as a phase shifter (corrector) after conventional deconvolution instead of a complete deconvolution operator [62]. That is, MED is intended as a sequel to standard deconvolution, not an alternative to it. White's work made MED more practically applicable.

Other contributors to MED-type techniques include Donoho, Godfrey, Deeming, and Walden [58]. Due to the limited space, we will not discuss further their work in this thesis. It should be pointed out that, despite all the above efforts, MED-type methods still have a

relatively long way to go in terms of the practical applicability.

Finally, let us notice an important fact: V_i in Eq(2.4.23) is an non-standard estimate of the higher-order statistical measure termed *kurtosis*. This is why MED is also named maximum kurtosis deconvolution in literature [62]. However, more important is that this fact suggested the adoption of higher-order statistical measures, termed higher-order cumulants (HOC) in statistics. These measures offered the possibility of deconvolution of non-minimum phase (NMP) series. This point has been reinforced by the development of higher-order cumulant (HOC) based deconvolution techniques, as will be discussed later in this thesis.

Above, three main block-type (off-line) deconvolution techniques, which can deal with nonminimum phase (NMP) wavelet, have been discussed: homomorphic deconvolution, maximum likelihood deconvolution, and minimum entropy deconvolution. A common feature among them that needs to be paid attention is that nonlinearity and higher-order statistics are employed.

2.5 Blind Equalisation of Nonminimum Phase Channels

As mentioned in Section 1.2, communications is another important application area of deconvolution, but a different nomenclature is used for deconvolution in communications: *equalisation*. In communications systems, there are two main factors which can cause distortion of the transmitted signals (or data), *i.e.*, intersymbol interference (ISI) and additive noise [25]. The former, which results from the channel departing (possibly in a time variant manner) from ideal characteristics, can be removed by means of adaptive equalisation. The latter, when very serious, can be suppressed by using matched filtering technology. In this thesis, however, we will only address the problem of adaptive equalisation, or more specifically, that of blind equalisation.

Typically, an adaptive equaliser is a hybrid system, *i.e.*, a combination of a preset mode and a decision-directed (DD) adaptive mode [25][63]. In the preset mode, equalisers need an initial training period in which a particular training data sequence (typically a pseudo-noise (PN) sequence) known and available in proper synchronism at the receiver, is transmitted. The purpose for this is to enable the system to acquire a correct initial set of equaliser parameters [21][25]. After the preset mode, the actual transmission can begin with unknown data (real information), and the equaliser enters the DD mode. The main procedure in DD mode can be briefly described as follows [63][64].

Let $x(k)$ be the transmitted (scalar) data at time k, $h(k)$ the impulse response of transmission channel, and $y(k)$ the received data at the channel output. The objective of equalisation is to determine an inverse filter $\{\theta_l\}$ such that

$$\sum_{l=-L}^{L} \theta_l y(k-l) = c(k) \approx x(k) \quad , \tag{2.5.1}$$

or

$$\sum_{l=-L}^{L} \theta_l h(k-l) \approx \delta(k) \tag{2.5.2}$$

holds, where $\delta(k)$ represents the Dirac delta function, and $L-(-L)+1=2L+1$ is the inverse filter length. In addition, we will always denote the inverse filter output as $c(k)$ in this section, as in Eq(2.5.1). Since we consider the NMP systems here, the inverse filter is noncausal. On the other hand, considering the specific feature of digital communications, it can be reasonablely assumed that $x(k)$ is IID, and equally distributed over a finite set D, *e.g.*, $D=\{\pm 1, \pm 3, \cdots, \pm(2M-1)\}$. In fact, this assumption corresponds exactly to the pulse amplitude modulation (PAM) scheme in digital communications. Then, according to adaptive filtering theory, in order to reconstruct $x(k)$ only from $y(k)$, the inverse filter $\{\theta_l\}$ can be adaptively formed in the following manner:

$$\theta_l^{(k+1)} = \theta_l^{(k)} - \tau \cdot e_C(k) \cdot y(k-l) \quad , \quad l=-L, \cdots, L, \tag{2.5.3a}$$

and

$$e_C(k) = c(k) - \hat{x}(k) \quad , \tag{2.5.3b}$$

where

$$\hat{x}(k) = \mathrm{dec}[c(k)] \tag{2.5.3c}$$

is the decision concerning $c(k)$, that is made by "slicing" $c(k)$ (called multilevel decision). As to the slicer, it can be, for example, an ordinary nonlinear memoryless threshold device. In our problem, $\hat{x}(k)$ is the nearest value to $c(k)$ in the finite set D. The equaliser adapted through Eq(2.5.3a-c) is termed a *decision directed* (DD) equaliser.

The above hybrid equaliser is normally able to offer satisfactory performance in practice, and can also be justified in theory [63][64]. Unfortunately, it is not practical in certain modern data communication systems. For example, in multipoint communication networks where a control station communicates with many tributary stations, it is inconceivable to interrupt the whole transmission (because lines are shared) so that the control modem can initiate a new preset mode causing all tributaries to retrain for the sake of a single particular tributary that would enter the network in the interim or that need reinitialisation simplely because of drastic changes in channel characteristics [21]. In addition, in transmission of time-division multiplexing (TDM) signals over existing microwave or radio coaxial facilities which employ a frequency-division multiplexing (FDM) system, the hybrid mode based equaliser is also not suitable due to the route reconnections during transmission [18]. As a result, blind equalisation, *viz.*, equalisation without training startup period, is necessary to such systems.

In the following part of this section, several important blind equalisation techniques will be reviewed. All these approaches, compared to those described in last section, are adaptive (on-line), due to the natural requirement of communications.

2.5.1 Sato algorithm

This algorithm is the first blind equalisation algorithm, and is of pioneering significance. In the case where the known training sequence is unavailable, a natural suggestion is to replace the known sequence with a sequence of data symbols estimated (*e.g.*, by multilevel decision) from the inverse filter output [64]. This is equivalent to using the DD mode at the very beginning of transmission. Unfortunately, this scheme fails to work, as

Sato found [18]. The theoretical analysis for the reasons of failure was later carried out by Macchi and Mazo *et al* [63][64]. In the Sato algorithm, a simple binary decision mechanism (slicer) is adopted, no matter whether the transmitted data are binary or not. Mathematically, this can be expressed as

$$\hat{x}(k) = \text{dec}[x(k)] = \gamma \cdot \text{sign}[c(k)] \ , \tag{2.5.4}$$

where sign[·] is the sign function, and γ the scaling factor given by

$$\gamma = \frac{\mathbf{E}[x^2(k)]}{\mathbf{E}[|x(k)|]} \ . \tag{2.5.5}$$

Substitution of Eq(2.5.4) into Eq(2.5.3) leads to the following algorithm:

$$\theta_l^{(k+1)} = \theta_l^{(k)} - \tau \cdot e_s(k) \cdot y(k-l) \ , \quad l = -L, \ \cdots, L, \tag{2.5.6a}$$

where

$$e_s(k) = c(k) - \gamma \cdot \text{sign}[c(k)] \ . \tag{2.5.6b}$$

Eq(2.5.6) is the so called *Sato algorithm*. Fig(2.7) illustrates the basic structure of the Sato equaliser.

The idea behind Sato's algorithm is to presume that the multilevel signal is decomposed into its polarity signal and the remaining signal, and then treat the remaining signal as random source noise. Consequently, the correct parameter adjustment will be performed after eliminating the remaining signal according to the average effect in the parameter adjusting mechanism [18].

The advantages of Sato algorithm include: 1) it is very easy to implement; 2) it is very robust (polarity signal directed); 3) it is independent of the probability density, except for the scaling factor γ [65]; and 4) it does not require any assumption for the channel phase properties.

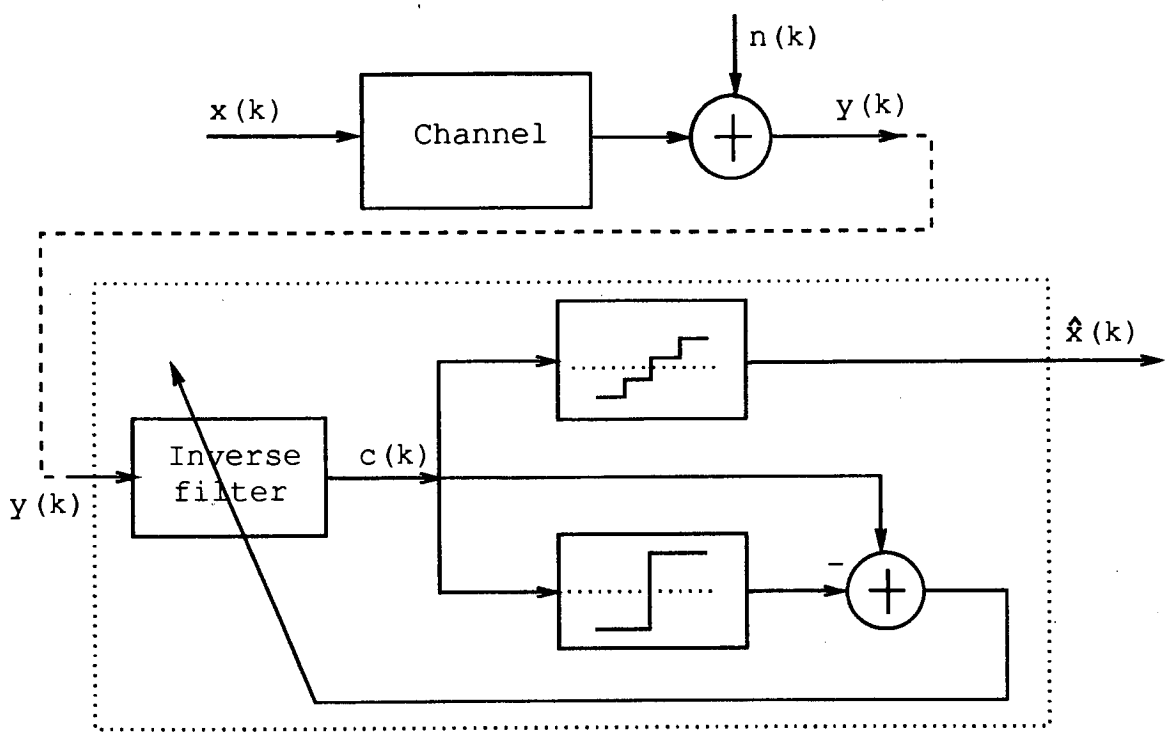
On the other hand, however, Sato could not prove whether his algorithm could converge in any case, and if not, under what conditions it could converge. Additionally, since $e_s(k) \neq 0$ even after the algorithm converges, the parameter adjustment is very noisy in Sato algorithm, except that the transmitted data are binary (2-level PAM). The next algorithm will improve these aspects.

2.5.2 Benveniste-Goursat algorithm

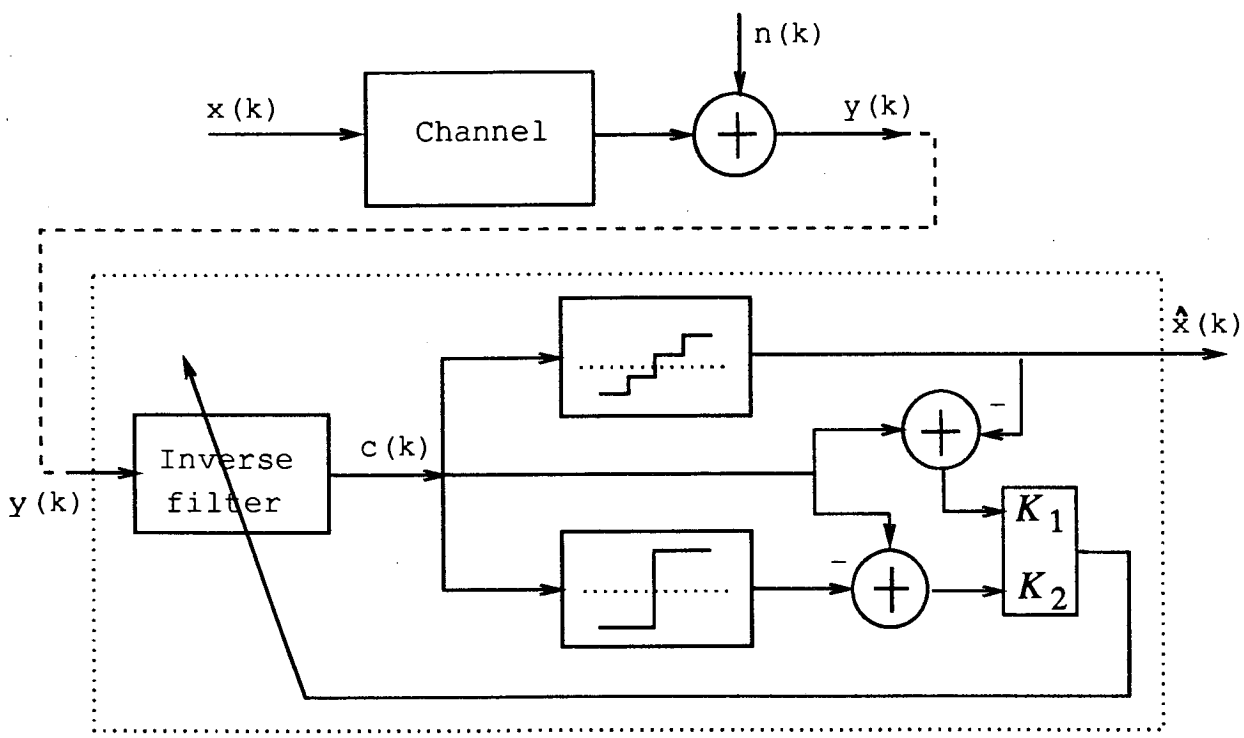
Although Sato's algorithm works well practically, the related theoretical analysis remained to be established. In 1980, A. Benveniste, M. Goursat, and G. Ruget (BGR) theoretically justified and extended Sato's result. In [19], BGR studied the general blind deconvolution problem, and pointed out the following two points which comprise the theoretical basis of blind deconvolution:

i) the blind deconvolution problem cannot be solved with the second-order statistics in the case of nonminimum phase (NMP) systems.

ii) the blind deconvolution problem has no solution when the system is NMP but the input signal has a Gaussian distribution.



Fig(2.7) Sato equaliser.



Fig(2.8) Benveniste-Goursat equaliser.

In other words, in the case of NMP systems, only when the input is non-Gaussian, can blind deconvolution problem possibly be solved, and the only alternative for the analysis tool is the non-second (higher or lower) order statistics.

BGR suggested the following algorithm for blind deconvolution:

$$\theta_l^{(k+1)} = \theta_l^{(k)} - \tau \cdot \psi[c(k)] \cdot y(k-l) \quad , \quad l = -L, \cdots, L, \tag{2.5.7}$$

where $\psi[\cdot]$ is a function whose form should be chosen according to the distribution of system input signals in order to enable the algorithm to converge to the global solution. BGR investigated two classes of distributions: super-Gaussian and sub-Gaussian, and proposed the corresponding choice for $\psi[\cdot]$.

The uniform distribution of PAM signals in communications can be considered to be close to the sub-Gaussian distribution. In this case, the simplest choice for $\psi[\cdot]$ is

$$\psi[v] = v - \gamma \cdot \text{sign}(v) \quad . \tag{2.5.8}$$

Substitution of Eq(2.5.8) into Eq(2.5.7) immediately gives Eq(2.5.6), *i.e.*, the Sato algorithm, thus the theoretical basis for Sato algorithm was found.

As discussed earlier, one of Sato algorithm's foibles is that its convergence is very noisy. On the other hand, the conventional procedure in Eq(2.5.3) obviously enjoys the desirable property of $e_C(k)=0$ after convergence. Benveniste and Goursat thus took advantage of the combination of these two schemes and suggested the following algorithm [20]:

$$\theta_l^{(k+1)} = \theta_l^{(k)} - \tau \cdot e_G(k) \cdot y(k-l) \quad , \quad l = -L, \cdots, L, \tag{2.5.9a}$$

where

$$e_G = K_1 e_C(k) + K_2 |e_C(k)| e_S(k) \quad , \tag{2.5.9b}$$

and, K_1 and K_2 are constants. The above algorithm, called the *Benveniste-Goursat algorithm* in some literature, has the advantage that it is both robust and "quietly" convergent, thus it provides us an automatic switch between the blind startup period and the conventional DD mode. It also possesses the advantages of the Sato algorithm. The basic structure of Benveniste-Goursat scheme is illustrated in Fig(2.8).

Benveniste *et al*'s results revealed that, the non-second order statistics play a vital role in blind deconvolution of NMP system output signals. A similar fact was also found by Scargle [66]. Although Benveniste *at al*'s techniques only employed the first-order statistics, and there are still some limitations (in terms of the probability distribution of signals) with them, they opened the prelude for development of higher-order cumulant (statistics) based algorithms, which finally removed all the requirements for data distribution (except that the distribution be non-Gaussian) and the system phase properties.

2.5.3 Other schemes

In addition to the aforementioned Sato algorithm and Benveniste-Goursat algorithm, there are several other reported blind deconvolution techniques. But we will only briefly describe two of them here, due to the limited space.

<1> Godard's algorithms: In [21], Godard introduced a different class of cost functions and algorithms for blind equalisation in data receivers employing two-dimensional

modulation. As in the Sato and Benveniste-Goursat approaches, the cost function adopted by Godard is also nonconvex, but the multimodal problem is solved by using a strategy of "small step-size parameters", and the corresponding algorithms are extremely robust with respect to channel distortions. Additionally, Godard's algorithms have a similar computational complexity in comparison with the conventional gradient algorithm for minimisation of the mean square error, thus can be easily implemented even in the microprocessor-based data receivers.

There are two drawbacks with Godard's algorithms: *i*) speed of convergence is normally very slow; and *ii*) a severe degradation with respect to the classical systems arises owing to decoupling the phase estimation and the removal of the ISI [20].

<2> **Stop-and-go algorithm:** In 1987, G. Picchi and G. Prati [67] re-explored the possibility of application of the DD scheme to blind startup since the DD scheme possesses the attractive advantages of simplicity and smoothness. As a consequence, they suggested a "stop-and-go" algorithm based on the following idea: the adaptation is stopped when the reliability (in a probabilistic sense) of the self-decided output error is not sufficiently high. In the algorithm, an easy-to-generate binary-valued flag is employed to tell the equaliser whether the output error on the current decision may reliablely be used in the DD algorithm. Unlike Godard's algorithm, the "stop-and-go" algorithm can realise the blind joint equalisation and carrier recovery. In essence, however, "stop-and-go" approach still adopted the Sato-type error in the operation of its flags.

The four main blind equalisation techniques have been reviewed above. As can be seen, all of them are feasible and even efficient in many cases, but generally they suffer from different limitations. An emerging solution to this situation is higher-order cumulant (HOC) based blind deconvolution (equalisation) technology.

2.6 Higher-Order Cumulant Analysis

It has been seen that, predictive deconvolution is limited by a strict assumption, *i.e.*, the system (or wavelet) be minimum phase (MP), and the other techniques suffer from other different limitations with respect to system and signal characteristics. Fortunately, higher-order cumulant (HOC) analysis technology provides a powerful tool for the solution of the blind deconvolution problem. Since HOC based blind deconvolution (equalisation) techniques comprise the main part of this thesis, some theoretical preliminaries are concentrated and presented in this and next section.

By "higher order", we normally mean the integer order higher than two, *viz.*, the 3rd order, the 4th order, and the like. The higher-order statistics, which we have mentioned in the earlier sections, actually include two classes: higher-order *cumulants* and higher-order *moments*, which can be mathematically defined as follows [13][68]-[70].

Assume a set of random variables is $\{x_1, x_2, \cdots, x_n\}$ with the joint characteristic function

$$\Phi(\omega_1, \omega_2, \cdots, \omega_n) = E\{\exp[j(\omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n)]\} , \qquad (2.6.1)$$

then their joint cumulants of order $r=k_1+k_2+\cdots+k_n$ are defined as

$$C_{k_1,\cdots,k_n} = (-j)^r \frac{\partial^r \ln\Phi(\omega_1, \omega_2, \cdots, \omega_n)}{\partial\omega_1^{k_1}\partial\omega_2^{k_2}\cdots\partial\omega_n^{k_n}}\Big|_{\omega_1=\omega_2=\cdots=\omega_n=0} , \tag{2.6.2}$$

and their joint moments of the same order r can be defined as

$$M_{k_1,\cdots,k_n} = (-j)^r \frac{\partial^r \Phi(\omega_1, \omega_2, \cdots, \omega_n)}{\partial\omega_1^{k_1}\partial\omega_2^{k_2}\cdots\partial\omega_n^{k_n}}\Big|_{\omega_1=\omega_2=\cdots=\omega_n=0} . \tag{2.6.3}$$

Obviously, the joint cumulants can be expressed by the corresponding joint moments.

Let us consider a real stationary stochastic process $\{x(k)\}$ (we will not distinguish in denotation between stochastic process and its realisations in this thesis). Then the following relations between cumulants and moments of $x(k)$ can be derived*:

$$\mathrm{E}[x(k)x(k+\tau_1)] = M_2(\tau_1) = C_2(\tau_1) , \tag{2.6.4}$$

$$\mathrm{E}[x(k)x(k+\tau_1)x(k+\tau_2)] = M_3(\tau_1, \tau_2) = C_3(\tau_1, \tau_2) , \tag{2.6.5}$$

$$\mathrm{E}[x(k)x(k+\tau_1)x(k+\tau_2)x(k+\tau_3)] = M_4(\tau_1, \tau_2, \tau_3)$$

$$= C_4(\tau_1, \tau_2, \tau_3)+C_2(\tau_1)\cdot C_2(\tau_3-\tau_2)+C_2(\tau_2)\cdot C_2(\tau_3-\tau_1)+C_2(\tau_3)\cdot C_2(\tau_2-\tau_1) , \tag{2.6.6}$$

etc., which are respectively corresponding to the 2nd-, 3rd-, and 4th-order statistics. Clearly, the 2nd-order statistics here is exactly the autocorrelations in conventional spectral analysis. In particular, when $\tau_i=0$, the following important statistical measures can be obtained:

$$\gamma_2 = C_2(0) = \mathrm{E}[x^2(k)] \tag{2.6.7a}$$

which is termed *variance*,

$$\gamma_3 = C_3(0, 0) = \mathrm{E}[x^3(k)] \tag{2.6.7b}$$

skewness, and

$$\gamma_4 = C_4(0, 0, 0) = \mathrm{E}[x^4(k)]-3\gamma_2^2 \tag{2.6.7c}$$

kurtosis. Notice that, although the two classes of statistics, cumulants and moments, are identical in value in the case of the 3rd (or 2nd) order, they are no longer equal to each other in the case of the 4th (or still higher) order.

Corresponding to the above two classes of statistics are two classes of spectra: *cumulant spectra* and *moment spectra*. These are generally termed *higher-order spectra (HOS)*. Cumulant spectra are defined as the (multi-dimensional) Fourier transform of the cumulants, and moment spectra as the (multi-dimensional) Fourier transform of the moments [68]-[71][13]. The former are suitable for the analysis of stochastic signals (stationary random processes), and the latter more suitable for the analysis of deterministic signals (transient or periodic processes) [71]. But in this thesis, we will deal only with the parameter based approaches to stochastic signal analysis, thus we will not consider further the frequency domain and moments, *i.e.*, we will only study cumulant based blind deconvolution approaches.

The reasons why cumulants, as opposed to moments, are employed to analyse the stochastic signals, can be explained as follows [13][71].

* $x(k)$ is assumed to be zero mean here.

(1) Consider a random process

$$z(k)=x(k)+y(k) \quad , \tag{2.6.8}$$

where $x(k)$ and $y(k)$ are two independent random processes, then

$$C_n^z(\tau) = C_n^x(\tau)+C_n^y(\tau) \quad , \tag{2.6.9}$$

but

$$M_n^z(\tau) \neq M_n^x(\tau)+M_n^y(\tau) \quad , \tag{2.6.10}$$

where $\tau=\{\tau_1, \cdots, \tau_{n-1}\}$. The above two equations indicate that cumulants satisfy the superposition principle, but moments do not. This property of cumulants is very important in signal processing, as will be discussed next.

(2) Higher-order cumulants of Gaussian processes are theoretically identical to zero, but higher-order moments do not have this property. As a result, Cumulants provide a more explicit measure of how non-Gaussian signals are (compared with moments).

(3) Higher-order ergodicity requirements are met more easily with cumulants than with moments.

Although higher-order cumulants (HOC) analysis (or HOS analysis) has found a wide range of applications, the motivation behind these applications can be simply summarised by the following three aspects [13][71].

<1> Suppress coloured additive noise: Recall Eq(2.6.8). If $z(k)$ is the noisy version of $x(k)$ as a result of the coloured noise $y(k)=n(k)$, then the influence of $n(k)$ may be completely suppressed in cumulant based algorithms no matter whether $n(k)$ is coloured or not. For example, if

1) $x(k)$ is non-Gaussian, and $n(k)$ is Gaussian; or

2) $x(k)$ is non-Gaussian and skewed ($\gamma_3 \neq 0$), and $n(k)$ is non-Gaussian but unskewed ($\gamma_3=0$),

then $n(k)$ theoretically has no interference to the calculation of $C_n^z(\tau)$ since in these two cases,

$$C_n^z(\tau) = C_n^x(\tau)+C_n^n(\tau) = C_n^x(\tau) \quad . \tag{2.6.11}$$

Notice that, in case 2, Eq(2.6.11) holds only if n is an odd number (*viz.*, odd order).

<2> Extract the phase information of nonminimum phase (NMP) signals: Unlike autocorrelation functions, HOC retain the phase information of signals. This comprises the theoretical basis of cumulant based blind deconvolution techniques, and will be demonstrated in detail in the next section.

<3> Detect and characterise the nonlinearities of systems: This is based on the following fact: the 3rd-order cumulants are sensitive to quadratic phase coupling, and the 4th-order cumulants to cubic phase coupling. As a consequence, for harmonic signals, the existence of higher-order cumulants (or moments) implies the "presence" of nonlinearities. The

detailed discussion of this aspect is beyond the scope of this thesis. For details, see [13] and [71].

Finally, it should be pointed out that the support domain for n-th order cumulants (moments) does not need to be the whole $(n-1)$ dimensional space. This is due to the symmetry properties that exist with cumulants. For instance, in the case of 3rd-order cumulants $C_3(\tau_1, \tau_2)$, the infinite wedge shaped sector (*i.e.*, sector 1 in Fig(2.9)) bounded by lines $\tau_2=0$ and $\tau_1=\tau_2$, where $\tau_1, \tau_2 \geq 0$, is sufficient as the support domain. In Fig(2.9), the 3rd-order cumulant values in any other sector (sector 2 to 6) can be determined by the ones in sector 1, and *vice versa*. Research about the number of support domains for the 4th (or even higher) order cumulants is still under way.

2.7 Higher-Order Cumulants and System Phase Properties

It has been noted that the higher-order cumulants of Gaussian processes theoretically vanish, *i.e.*, the Gaussian processes are completely determined by their autocorrelation functions. Consequently, in the main we will consider non-Gaussian processes in this thesis. In fact, the requirement for non-Gaussian characteristics is realistic in many practical cases. For example, PAM data in communications and the earth reflectivity series in geoexploration [72][58] are non-Gaussian.

As is well known, the autocorrelation function of a series is completely blind to its phase properties, *i.e.*, autocorrelation functions do not carry any phase information. However, higher-order cumulants (HOC) of a series retain the phase information, and are very sensitive to the signal phase properties. As a result, HOC can be employed to extract the phase characteristics of a signal (or system). This comprises the main theoretical basis of HOC based blind deconvolution and equalisation techniques, as mentioned earlier. To further demonstrate the insensitivity of autocorrelations and the sensitivity of HOC to signal (system) phase properties, let us investigate the concrete example below.

Consider the following three systems [13]:

$$H_1(z) = (1-az^{-1})(1-bz^{-1}) \quad , \tag{2.7.1}$$

$$H_2(z) = (z^{-1}-a)(z^{-1}-b) \quad , \tag{2.7.2}$$

and

$$H_3(z) = (z^{-1}-a)(1-bz^{-1}) \quad , \tag{2.7.3}$$

where $|a|, |b|<1$. Clearly, $H_1(z)$ is minimum phase, $H_2(z)$ is maximum phase, and $H_3(z)$ is mixed phase (refer to Section 2.2). It can be shown that the above three systems possess the same amplitude frequency characteristics:

$$\begin{aligned} A(\omega) &= |H_1(e^{-j\omega})| = |H_2(e^{-j\omega})| = |H_3(e^{-j\omega})| \\ &= [(1-2a\cos\omega+a^2)(1-2b\cos\omega+b^2)]^{\frac{1}{2}} \quad , \end{aligned} \tag{2.7.4}$$

but different phase frequency characteristics:

$$\phi_1(\omega) = \arg[H_1(e^{-j\omega})] = \operatorname{arctg}\frac{a\sin\omega}{1-a\cos\omega} + \operatorname{arctg}\frac{b\sin\omega}{1-b\cos\omega} \quad , \tag{2.7.5}$$

$$\phi_2(\omega) = \arg[H_2(e^{-j\omega})] = \text{arctg}\frac{\sin\omega}{a-\cos\omega} + \text{arctg}\frac{\sin\omega}{b-\cos\omega} \quad , \tag{2.7.6}$$

and

$$\phi_3(\omega) = \arg[H_3(e^{-j\omega})] = \text{arctg}\frac{\sin\omega}{a-\cos\omega} + \text{arctg}\frac{b\sin\omega}{1-b\cos\omega} \quad . \tag{2.7.7}$$

Assume these three systems are driven by the same non-Gaussian series $w(k)$ with $\mathbf{E}[w(k)]=0$, $\mathbf{E}[x(k)x(k+\tau_1)]=\gamma_2\delta(\tau_1)$, and $\mathbf{E}[w(k)w(k+\tau_1)w(k+\tau_2)]=\gamma_3\delta(\tau_1,\tau_2)$, as shown in Fig(2.10). Then the corresponding output signals can be obtained:

$$y_1(k) = w(k)-(a+b)w(k-1)+abw(k-2) \quad , \tag{2.7.8}$$

$$y_2(k) = abw(k)-(a+b)w(k-1)+w(k-2) \quad , \tag{2.7.9}$$

and

$$y_3(k) = -aw(k)+(ab+1)w(k-1)-bw(k-2) \quad . \tag{2.7.10}$$

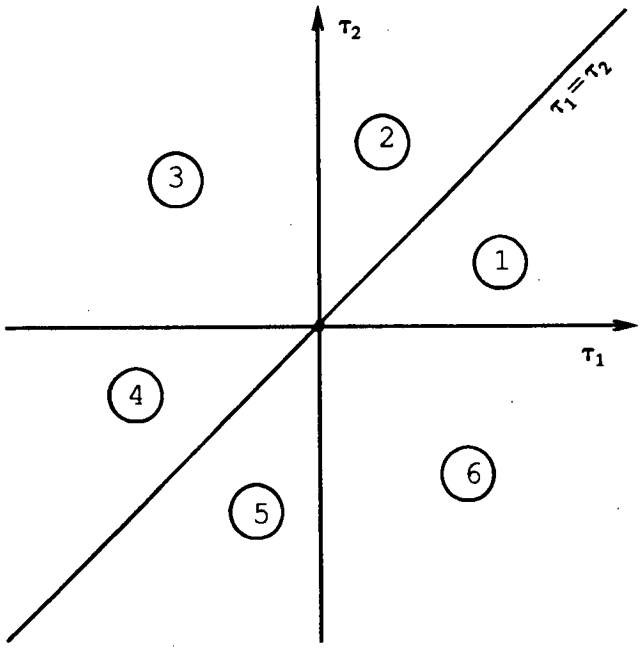
According to Eq(2.6.5) and Eq(2.6.4) in last section, the 3rd-order cumulants and the autocorrelation functions of the above three signals can be determined, and the results are listed in Table (2.1).

Table (2.1) Cumulants and Autocorrelations: a Comparison.

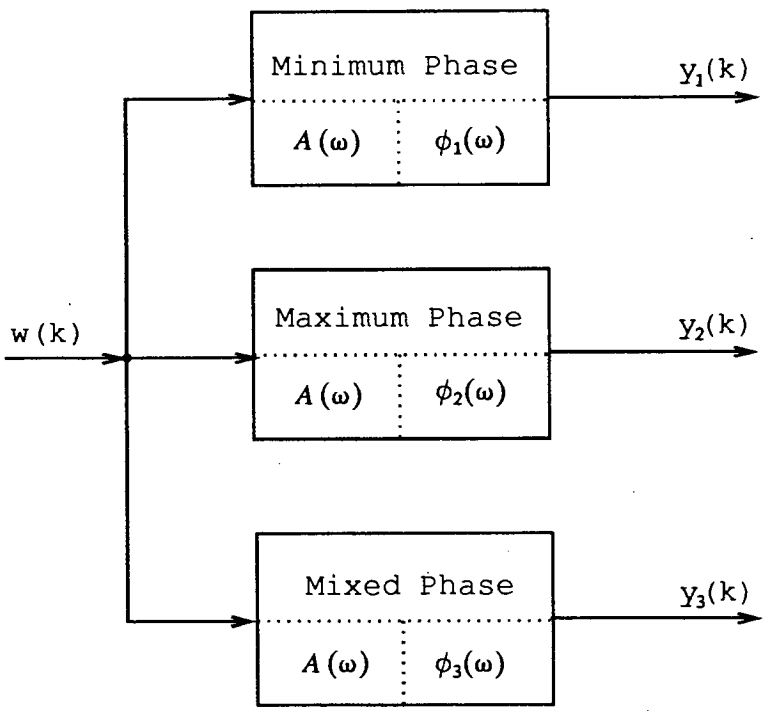
		Minimum Phase	*Maximum Phase*	*Mixed Phase*
$y(k)$		$y_1(k)$	$y_2(k)$	$y_3(k)$
Third Order Cumulants	$C_3(0,0)$	$1-(a+b)^3+a^3b^3$	$1-(a+b)^3+a^3b^3$	$(1+ab)^3-a^3-b^3$
	$C_3(1,1)$	$(a+b)^2-(a+b)a^2b^2$	$-(a+b)+ab(a+b)^2$	$-a(1+ab)^2+(1+ab)b^2$
	$C_3(2,2)$	a^2b^2	ab	$-ab^2$
	$C_3(1,0)$	$-(a+b)+ab(a+b)^2$	$(a+b)^2-(a+b)a^2b^2$	$a^2(1+ab)-(1+ab)^2b$
	$C_3(2,0)$	ab	a^2b^2	$-a^2b$
	$C_3(2,1)$	$-(a+b)ab$	$-(a+b)ab$	$ab(1+ab)$
Auto-correlation	$C_2(0)$	$1+a^2b^2+(a+b)^2$	$1+a^2b^2+(a+b)^2$	$1+a^2b^2+(a+b)^2$
	$C_2(1)$	$-(a+b)(1+ab)$	$-(a+b)(1+ab)$	$-(a+b)(1+ab)$
	$C_2(2)$	ab	ab	ab

Consequently, it has been confirmed that the HOC are really sensitive to signal (system) properties but autocorrelations totally lose the phase information of signals (systems). In fact, the latter has been implied earlier by Eq(2.7.4) through the well known Wiener-Kinchine theorem.

On the other hand, it should be noticed that HOC are blind to the linear phase shift of signals [15][13]. Given two systems: $S=\{h(k)\}$ and $S'=\{h'(k)=h(k-m)\}$, where m is an integer delay, it can be easily proved that the cumulant values of S output and those of S' output are identical. Thus, there exists a linear ambiguity with HOC. However, since this ambiguity only causes discrepancy in group delay of signals, it does not form any



Fig(2.9) Support domains for 3rd-order cumulants.



Fig(2.10) Three signals with different phase properties.

essential problem to blind deconvolution and identification.

2.8 Summary

The major blind deconvolution and equalisation techniques have been reviewed in this chapter. The advantages and drawbacks of every approach have been pointed out. Our discussions have mainly followed the line of introduction of nonlinearities or non-second order statistics. In addition, the general historical development of blind deconvolution (equalisation) technology has been discussed. It has also been indicated that HOC analysis has the potential to solve the blind deconvolution problem of NMP systems.

The study of blind deconvolution or equalisation problem can also be considered in another way: Bussgang techniques [65][67][73][36]. But this author believes that the key reason why the MP phase assumption has been relaxed in the techniques presented in Section 2.4 and 2.5 is the adoption of nonlinear mechanisms or non-second order statistics, instead of the employment of Bussgang type iteration. The essence of an algorithm is obviously far more important than its form. As a consequence, our discussions in this chapter have been made without reference to the Bussgang frame.

The theoretical preliminaries of HOC based blind deconvolution techniques, *i.e.*, the main principles and results of HOC analysis, have been introduced in Section 2.6 and 2.7. In the later chapters, we will cite the results from these two sections. But before looking further at HOC techniques, a novel autocorrelation based blind deconvolution algorithm is introduced in the next chapter.

Chapter 3:

A ROBUST BLIND DECONVOLUTION ALGORITHM: VARIANCE APPROXIMATION AND SERIES DECOUPLING

3.1 Introduction

In Section 2.3, the method of predictive deconvolution was introduced enabling us to efficiently obtain the estimates of both deterministic system impulse responses (or wavelet) as well as the random input series (or innovation) in the case of minimum phase (MP). Since the assumption of MP can be satisfied in practical models such as seismic ghost generation and water reverberation generation [5], it is necessary that MP assumption based techniques be further pursued.

Notice that, in predictive deconvolution, all of the autocorrelation function values: $\Phi(\tau)$, $\tau=0, 1, \ldots, p$, are used, p being the order of the corresponding inverse filter. Aiming at this feature, this chapter proposes a novel robust blind deconvolution algorithm which employs cost function based on a different idea: variance approximation and series decoupling (VASD). It has been found that not all autocorrelation function values $\{\Phi(\tau)\}$ are necessary to implement this algorithm.

Section 3.2 derives the algorithm, which is then analysed in Section 3.3. In Section 3.4, several computer simulation examples are presented, and the performance comparison with the predictive approach is also illustrated. The concerned concluding remarks are drawn in Section 3.5.

3.2 Algorithm Description

It is assumed in the following derivation that the system is minimum phase (MP), and the system input is an independent and identically distributed (IID) random series with zero mean and a known variance. The additive noise is Gaussian and white. Next, we adopt a rather different approach than those used in the existing techniques (*e. g.*, the predictive technique) to derive our algorithm.

Fig(3.1) illustrates the configuration of the proposed system. When the switch is at position 1, the deconvolver operates in the on-line state, and when at position 2 and 3, it is in the off-line state. As is shown in Fig(3.1), the input series is denoted as $x(k)$, the true

output series as $y_{nf}(k)$, the measured series $y(k)$, the measurement noise as $n(k)$, and the restored series as $c(k)$. Let the system model be assumed to be $ARMA(m, n)$ with transfer function

$$H(z) = \frac{1+\sum_{i=1}^{n} b_i z^{-i}}{\sum_{i=0}^{m} a_i z^{-i}} \quad . \tag{3.2.1}$$

Consequently, the parameter equations

$$\sum_{i=0}^{m} a_i y_{nf}(k-i) = x(k)+\sum_{i=1}^{n} x(k-i) \quad , \tag{3.2.2a}$$

and

$$y(k) = y_{nf}(k)+n(k) \quad , \quad k=0, 1, \cdots, N-1 \, , \tag{3.2.2b}$$

holds, where N is the number of samples. Since it is assumed that the system is minimum-phase, *i. e.*, all zeros of $H(z)$ are within the unit circle on the Z-plane (see Section 2.2), the corresponding inverse filter $H^{-1}(z)=\frac{1}{H(z)}$ can be expanded as

$$H^{-1}(z) = \sum_{l=0}^{\infty} \theta_l z^{-l} \quad . \tag{3.2.3}$$

Furthermore, considering that θ_l tends to zero with l increasing under the MP assumption, Eq(3.2.3) can be truncated as

$$H^{-1}(z) = \sum_{l=0}^{L} \theta_l z^{-l} \quad , \tag{3.2.4}$$

where L is a sufficiently large integer, and $\{\theta_l \,|\, l=0, 1, \cdots, L\}$ can be taken as the coefficients (or taps) of the inverse filter (MA model). Thus, $c(k)$, the restored signal, can be written as

$$c(k) = \sum_{l=0}^{L} \theta_l y(k-l) \quad , \quad k=0, 1, \cdots, N-1 \, . \tag{3.2.5}$$

As a result, restoration of $x(k)$ now reduces to determination of coefficients θ_l's.

Since $x(k)$ is an IID series, its autocorrelation function must satisfy

$$\Phi_{xx}(\tau) = 0 \quad , \quad \tau=1, 2, \ldots, q \, , \tag{3.2.6}$$

where q is a positive integer. With respect to the variance the following cost function is chosen:

$$\epsilon^2 = (s_c^2-\sigma_x^2)^2 + \sum_{\tau=1}^{q} \Phi_{cc}^2(\tau) \quad , \tag{3.2.7}$$

where,

$$s_c^2 = \frac{1}{N} \sum_{k=0}^{N-1} c^2(k) \quad , \tag{3.2.8}$$

$$\Phi_{cc}(\tau) = \frac{1}{N-\tau} \sum_{k=0}^{N-\tau-1} c(k) \cdot c(k+\tau) \quad , \tag{3.2.9}$$

and σ_x^2 is the known variance of input series $x(k)$. Note that, when

$$\epsilon^2 \to 0 \ , \tag{3.2.10}$$

it is necessary to have

$$s_c^2 \to \sigma_x^2 \ , \tag{3.2.11}$$

and

$$\Phi_{cc}(\tau) \to 0 \ , \quad \tau=1, 2, \ldots, q \ , \tag{3.2.12}$$

thereby $c(k)$ becomes a white series, which is an approximated version of the input series. Thus, as long as the coefficients θ_l's are adjusted to make ϵ^2 tend to the global minimum, the output series $c(k)$ of the filter $\{\theta_l \,|\, l=0, 1, \cdots, L\}$ will approximate the deconvolved series. Thus, deconvolution here is actually equivalent to such a process: the standard variance s_c^2 of $c(k)$ approximates the variance σ_x^2 of the true input series and $y(k)$ is gradually decoupled. This is the reason why the algorithm is termed *variance approximation* and *series decoupling* (VASD).

Under the assumption of a minimum phase system, ϵ^2 is a paraboloid with respect to the θ_l's ($l \geq 1$) when θ_0 is definite. But when θ_0 is a variable, the performance surface ϵ^2 takes the shape of symmetric double peaks, because both $\{\theta_l \,|\, l=0, 1, \ldots, L\}$ and $\{-\theta_l \,|\, l=0, 1, \ldots, L\}$ can produce the series $c(k)$ that allow ϵ^2 to reach the minimum. In this chapter, to obtain the unique solution of $\{\theta_l\}$, $\theta_0>0$ (or $\theta_0<0$) is assumed. The only ambiguity that can arise from this is that the deconvolution result is possibly the negative of the true one. This is not important, because the aim of deconvolution is to obtain the shape of $c(k)$. Thus, the deconvolution problem now reduces further to the global minimisation of ϵ^2 under the constraint of $\theta_0>0$ (or $\theta_0<0$). The steepest descent minimisation algorithm can be adopted to achieve this.

According to the steepest descent algorithm, the following recursion can be obtained:

$$\theta_l^{(i+1)} = \theta_l^{(i)} - \delta \cdot \nabla_l \ , \tag{3.2.13a}$$

$$l=0, 1, \cdots, L \ ,$$

$$\nabla_l = \frac{\partial \epsilon^2}{\partial \theta_l} \ , \tag{3.2.13b}$$

$$\theta_0>0 \ , \tag{3.2.13c}$$

$$\frac{\partial \epsilon^2}{\partial \theta_l} = 2(s_c^2 - \sigma_x^2) \cdot \frac{\partial s_c^2}{\partial \theta_l} + 2\sum_{\tau=1}^{q} \Phi_{cc}(\tau) \cdot \frac{\partial \Phi_{cc}(\tau)}{\partial \theta_l} \ , \tag{3.2.13d}$$

$$\frac{\partial s_c^2}{\partial \theta_l} = \frac{2}{N}\sum_{k=l}^{N-1} c(k) y(k-l) \tag{3.2.13e}$$

$$\frac{\partial \Phi_{cc}(\tau)}{\partial \theta_l} = \frac{1}{N-\tau}\sum_{k=l}^{N-\tau-1} [c(k) y(k-l+\tau) + c(k+\tau) y(k-l)] \ , \tag{3.2.13f}$$

$$c(k) = \sum_{j=0}^{l-1} \theta_j^{(i+1)} y(k-j) + \sum_{j=l}^{L} \theta_j^{(i)} y(k-j) \ , \tag{3.2.13g}$$

where δ is the step-size which controls the rate-of-convergence. The initial estimate of $\{\theta_l\}$

can be taken as $\{\theta_0(>0), 0, \cdots, 0\}$. In the iteration process, δ can be adjusted to make $\theta_0>0$.

One of the important features of VASD relates to the selection of q. It has been found that, unlike predictive deconvolution, it is not necessary to make $q=L$. In fact, when q is taken as 1 or 2, the results are quite acceptable. The general principle is that, the larger q is, the more accurate the result, and when $q=L$, the same number of autocorrelation function values as the predictive algorithm would be used. However, the larger q value will result in an increased computation cost, and this can be considered as the penalty for a more accurate result.

In order to enlarge the possible range of τ and to improve convergence of the algorithm, the gradient term is taken to be of the following form:

$$\nabla_l = \begin{cases} \text{sign}(\frac{\partial \epsilon^2}{\partial \theta_l})\cdot\ln(1+|\frac{\partial \epsilon^2}{\partial \theta_l}|) & \text{for} |\frac{\partial \epsilon^2}{\partial \theta_l}|\geq 1 \\ 2\cdot\text{sign}(\frac{\partial \epsilon^2}{\partial \theta_l})\cdot\ln(1+|\frac{\partial \epsilon^2}{\partial \theta_l}|) & \text{for} |\frac{\partial \epsilon^2}{\partial \theta_l}|<1 \end{cases} , \tag{3.2.14}$$

where sign$(\cdot)$ is the sign function. Eq(3.2.14) is actually based on the following two inequalities:

$$\ln(1+A) < A \quad \text{for } A>0 , \tag{3.2.15a}$$

and

$$2\cdot\ln(1+A) \geq A \quad \text{for } 0\leq A<1 , \tag{3.2.15b}$$

where A is a real number. Introduction of Eq(3.2.14) greatly reduces the risk of divergence of the algorithm when $|\frac{\partial \epsilon^2}{\partial \theta_l}|\geq 1$, and increases the rate-of-convergence, when $|\frac{\partial \epsilon^2}{\partial \theta_l}|<1$.

The proofs of the above two inequalities are straightforward and can be shown as follows.

Let

$$f(A) = \ln(1+A)-A , \tag{3.2.16}$$

then we must have

$$\frac{\partial f(A)}{\partial A} = -\frac{A}{1+A} . \tag{3.2.17}$$

Clearly,

$$\frac{\partial f(A)}{\partial A} < 0 , \quad \text{for } A>0 . \tag{3.2.18}$$

i.e., when $A>0$, $f(A)$ is strictly decreasing. Since

$$f(A)|_{A=0} = 0 , \tag{3.2.19}$$

it is necessary to have

$$f(A) < 0 \quad , \quad \text{for } A > 0 \quad , \tag{3.2.20}$$

which directly leads to Eq(3.2.15a).

By following a similar approach, Eq(3.2.15b) can be proved.

Finally, let us notice that it is not necessary to take ϵ^2 in Eq(3.2.7) as the criterion of stopping iteration in practical process, and the following standard can be employed:

$$\sum_{l=0}^{L} |\nabla_l| < \zeta \quad , \tag{3.2.21}$$

where ζ is a very small preset positive number. The reason for this is that Eq(3.2.21) and Eq(3.2.7) are consistent, and using the former can save some computation.

3.3 Algorithm Analysis

Following on from the above description of VASD algorithm, several comments are in order.

Firstly, since for $q=1$ or 2 an acceptable deconvolution result is achievable, it is not necessary to estimate the autocorrelation function $\Phi_{cc}(\tau)$ at all lag points, thus both the effect of estimation accuracy of $\Phi_{cc}(\tau)$ $(\tau>q)$ on the inverse filter coefficients and the computational burden are significantly reduced.

Secondly, although the measurement noise contaminates the output signal $y_{nf}(k)$ in terms of standard variance:

$$s_y^2 = s_{nf}^2 + \sigma^2 \quad , \tag{3.3.1}$$

VASD enforces s_c^2 to tend gradually to σ_x^2, which is equivalent to suppressing the noise. Thus, the VASD technique is robust to additive noise.

Thirdly, modification of the steepest descent minimisation algorithm ensures VASD possesses a fast and safe convergence.

Fourthly, VASD is effective for both Gaussian and non-Gaussian process, and consequently it is not necessary to know (or assume) the distribution of input series as in the MLM [11].

Finally, VASD can easily be implemented in real time. We have only derived the block version of VASD in last section, but the adaptive version can also be obtained through modification of Eq(3.2.13e) and Eq(3.2.13f). In this chapter, however, we will not further discuss this problem.

3.4 Simulation Results

In order to confirm the feasibility of VASD, the following computer simulations were implemented. The system transfer function is assumed to be

$$H(z) = \frac{1+0.467z^{-1}-0.133z^{-2}}{1-0.55z^{-1}-0.325z^{-2}+0.1z^{-3}} \quad .$$

The input $x(k)$ is a 500 point series generated by an uniformly-distributed pseudo-white-

noise generator, and Fig(3.2) shows its waveform (but only the first 150 points are drawn, similarly hereinafter). The noise-contaminated output series $y(k)$ is synthesised by using the following equations:

$$y_{nf}(k) = x(k)+0.467x(k-1)-0.133x(k-2)$$
$$+0.55y_{nf}(k-1)+0.325y_{nf}(k-2)-0.1y_{nf}(k-3) \quad , \tag{3.4.1a}$$

and

$$y(k)=y_{nf}(k)+n(k) \quad , \tag{3.4.1b}$$

where $n(k)$ is Gaussian white noise with the signal-to-noise ratio SNR=30dB. Other parameters include: $\delta=0.1$, $L=7$, $\zeta=0.3\times10^{-3}$, and $\{\theta_i^{(0)}\}=\{3, 0, \cdots, 0\}$. The following results are obtained using VASD:

1) The case of $q=1$ The deconvolved series $c(k)$ is shown in Fig(3.3a), and Fig(3.3b) illustrates the error waveform between $x(k)$ and $c(k)$: $e(k)=c(k)-x(k)$. This simulation takes only 10 iterations.

2) The case of $q=2$ The deconvolved series $c(k)$ is plotted in Fig(3.4a), and the error waveform $e(k)$ in Fig(3.4b). This implementation takes 16 iterations.

It is clear that the results and convergence rates in both above cases are satisfactory, and the case of $q=2$ is relatively better.

To compare VASD with the predictive deconvolution, we have applied the latter to the above synthesised series $y(k)$, and the result is shown in Fig(3.5a). The corresponding error waveform $e(k)$ is drawn in Fig(3.5b).

It can be seen that the results of VASD have a similar accuracy in comparison with those of the predictive deconvolution, but VASD is much simpler since only $\Phi(q)|_{q=0,1,2}$ are employed.

3.5 Conclusions

A new blind deconvolution method, VASD, has been presented in this chapter. Computer simulations indicate that this technique is effective and successful, and it has the following advantages over some existing approaches:

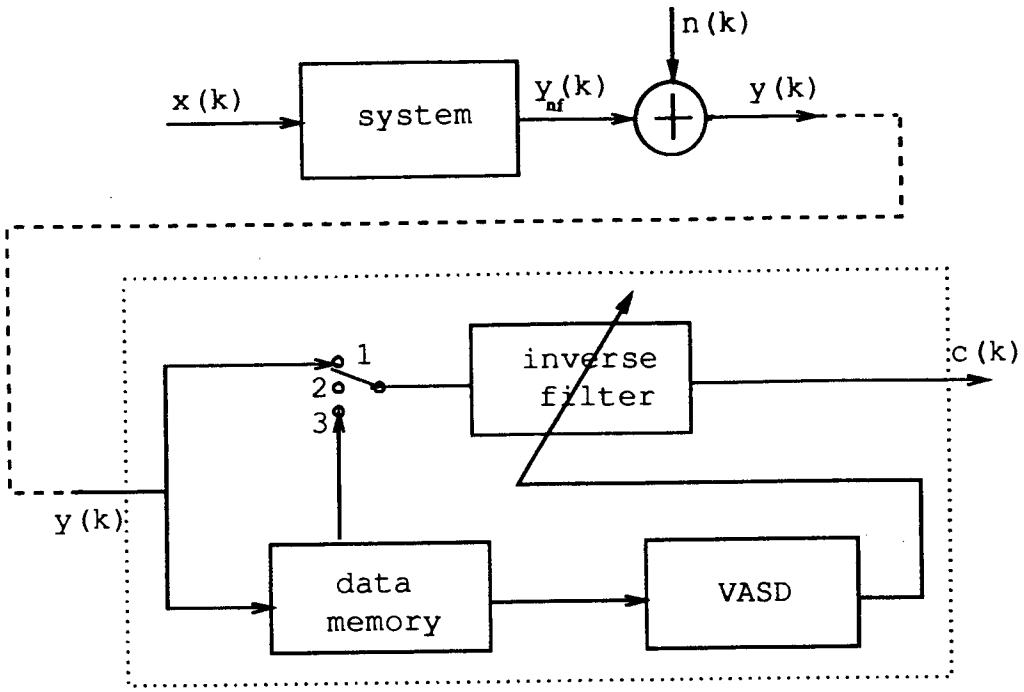
1) No need to estimate the autocorrelation function of output series at all lag points;

2) Fast convergence rate;

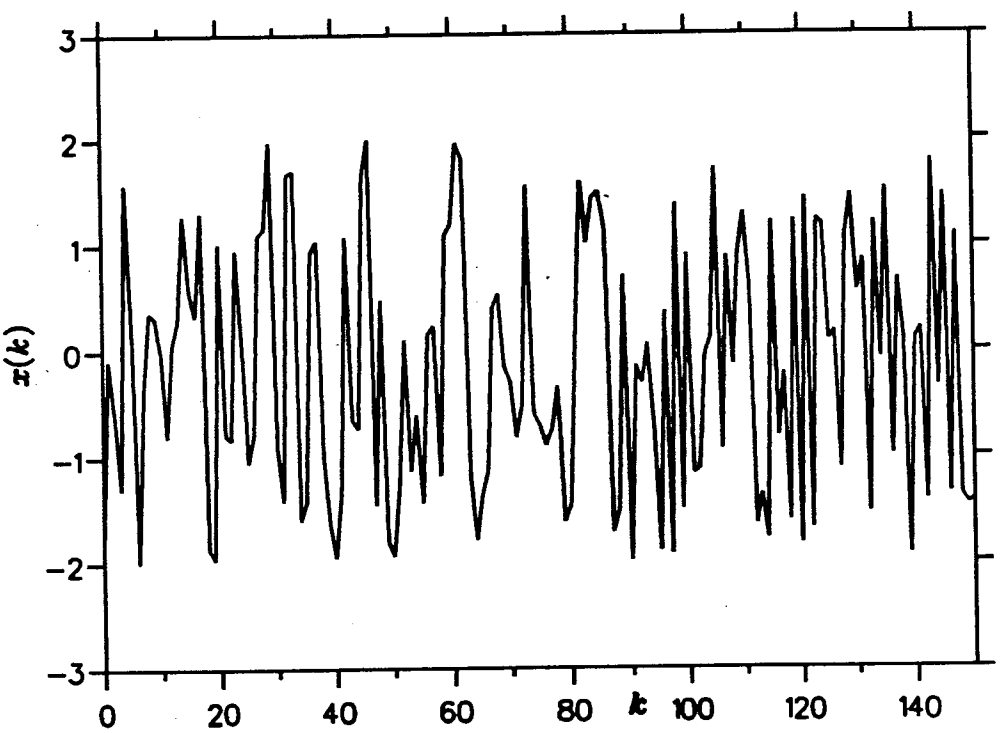
3) Strong robustness to additive noise; and

4) Being suitable to both Gaussian and non-Gaussian processes.

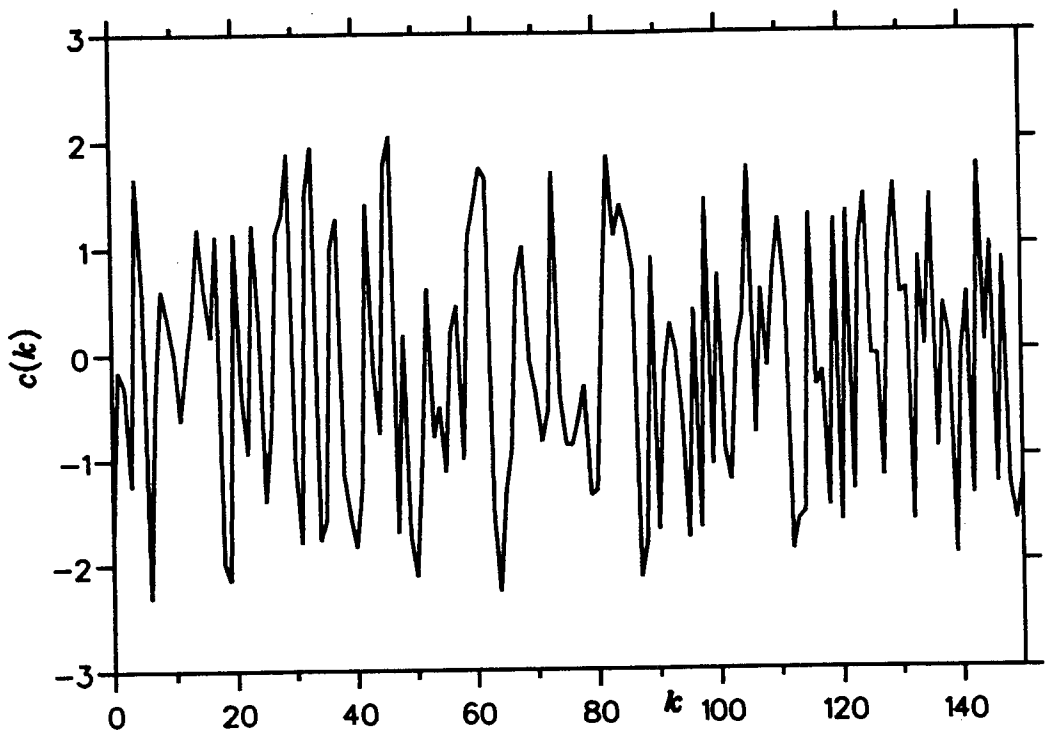
On the other hand, like the predictive deconvolution technique, VASD depends upon the assumption of minimum phase (MP). From next chapter, however, we will start the main theme of this thesis , higher-order cumulant based blind deconvolution techniques, which do not require the MP assumption.



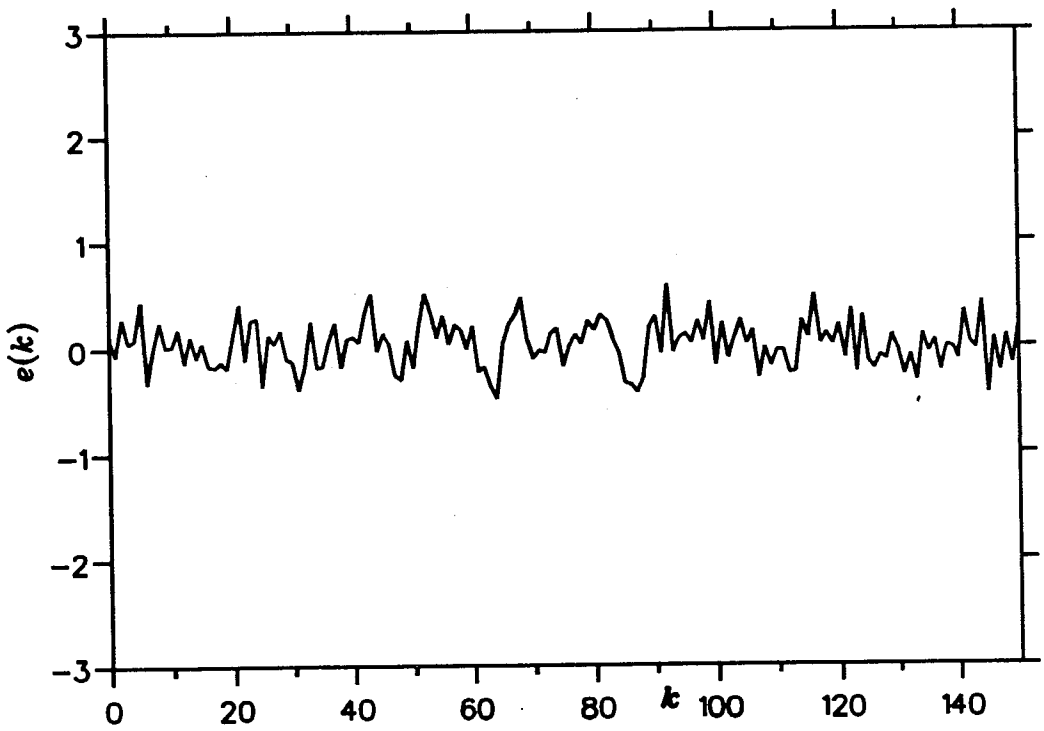
Fig(3.1) The proposed deconvolver system.



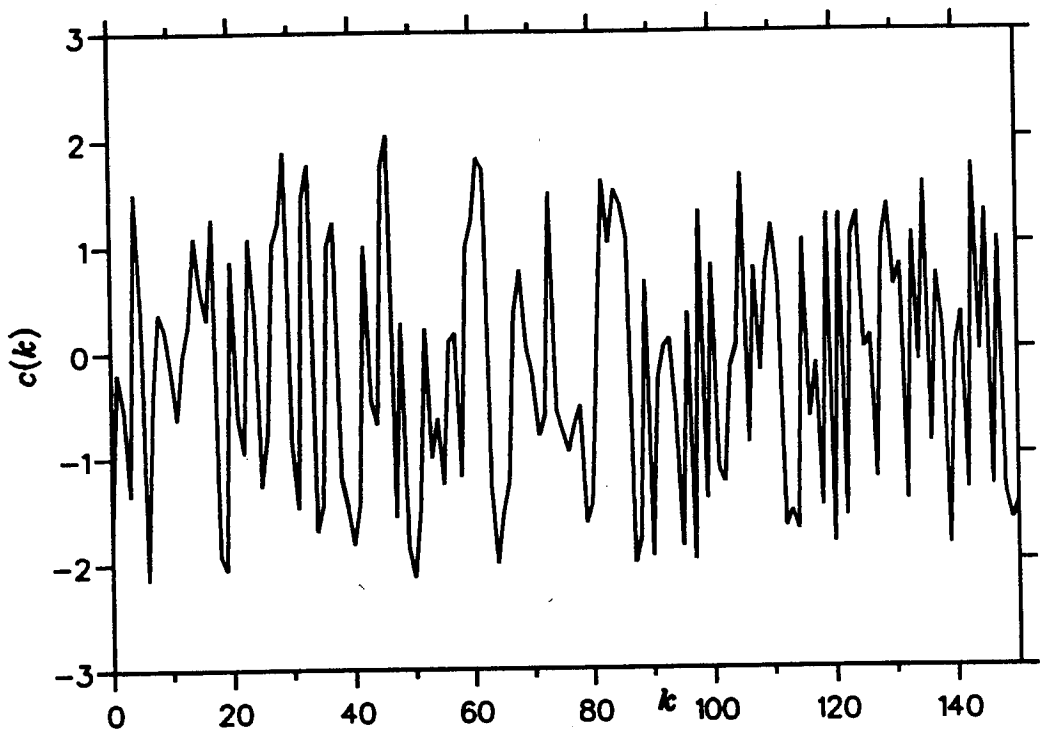
Fig(3.2) A segment of the input signal x(k).



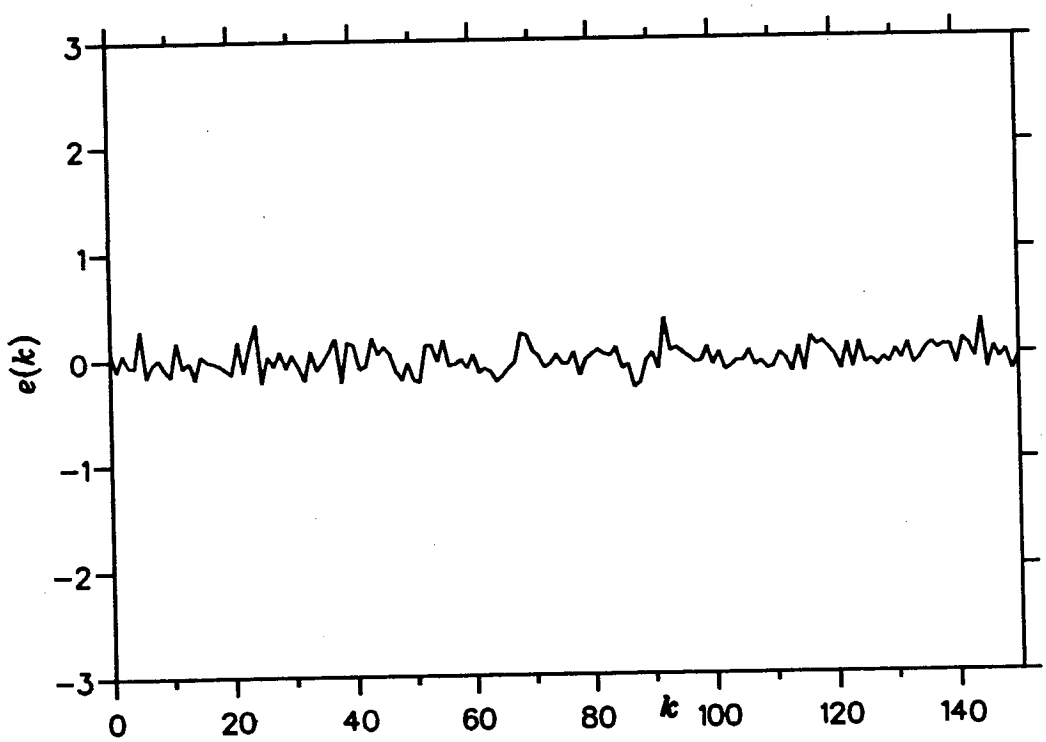
Fig(3.3a) Deconvolved signal c(k): q=1.



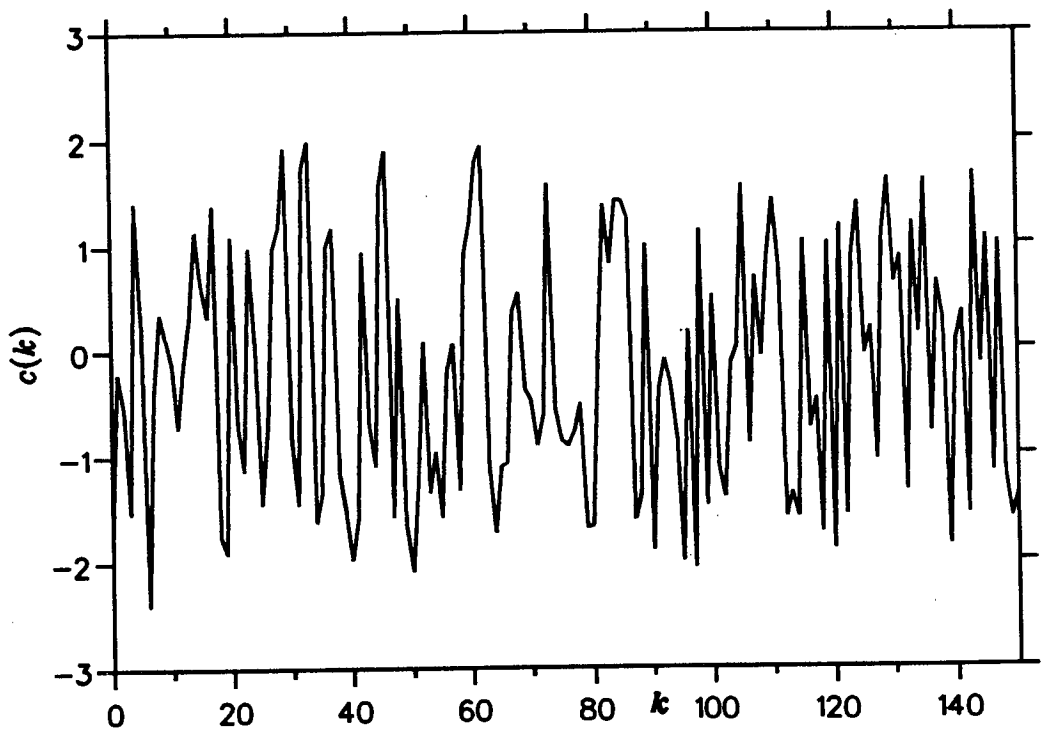
Fig(3.3b) Deconvolution error e(k): q=1.



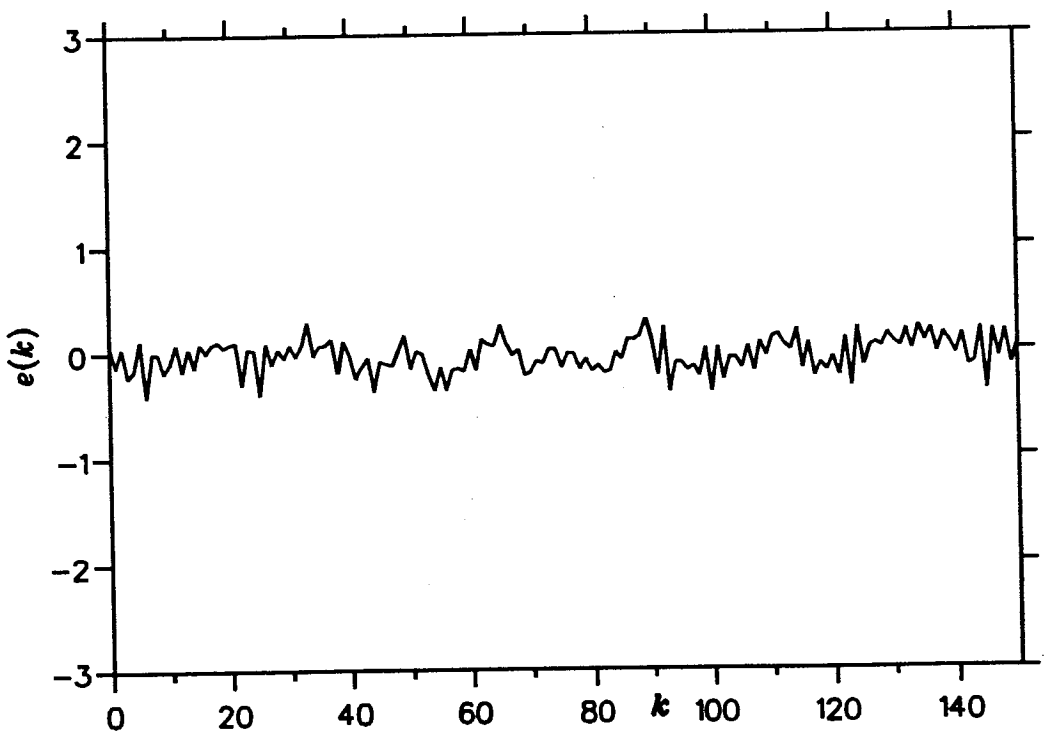
Fig(3.4a) Deconvolved signal c(k): q=2.



Fig(3.4b) Deconvolution error e(k): q=2.



Fig(3.5a) Deconvolved signal c(k): Predictive method.



Fig(3.5b) Deconvolution error e(k): Predictive method.

Chapter 4:

HIGHER-ORDER CUMULANT BASED BLIND DECONVOLUTION: MA MODEL

4.1 Introduction

The previous three chapters have shown that blind deconvolution is a problem of great significance in many practical fields including data transmission and seismic data processing, but its solution varies with different assumptions for input series and system models. In this chapter, the MA based blind deconvolution techniques will be studied. As is well known, for any stable system, its transfer function can be approximated by a MA model with finite number of coefficients (or parameters). From the general formulation of blind deconvolution problem, which has been presented in Chapter 1, we can redefine blind deconvolution problem in the context of MA model as follows.

Assuming that $x(k)$ is the observed output process of a linear time-invariant system driven by process $w(k)$, then the following equation holds:

$$x(k) = x_{nf}(k)+n(k) = \sum_{i=0}^{q} b_i w(k-i)+n(k)\ ,\ k=0,\ 1,\ \cdots,N-1, \qquad (4.1.1)$$

where b_i's are the MA model parameters (or the truncated unit impulse response) of the system (*e.g.*, the wavelet in seismic exploration), q is the order of MA model, $n(k)$ the additive noise, and N the number of samples. The basic objective of deconvolution involves two steps: 1) determine the coefficients b_i, *i.e.*, *system identification* (this step is termed channel identification in data communications, and wavelet extraction in seismic data processing); and 2) form an inverse filter to regenerate (or recover) the driving process $w(k)$, *i.e.*, *restoration.*

As noted in Section 1.1, it is impossible to determine the coefficients, b_i, and recover $w(k)$ in Eq(4.1.1) when only $x(k)$ can be observed. Fortunately, $w(k)$ can be reasonably assumed as a zero-mean, independent, and identically distributed (IID) process in many practical fields. With this assumption, when the system (or wavelet) is minimum phase (MP), $w(k)$ can be easily recovered by using the predictive deconvolution technique or the VASD method described in last chapter. When the system is nonminimum phase (NMP), however, the above two approaches cannot generate the correct results at all. The reason for this is that both predictive and VASD methods are based only on the second-order cumulants, *i.e.*, autocorrelation functions, which, as shown in Section 2.7, are completely blind to the phase property of the system. As a matter of fact, the results of predictive or

VASD technique are only the spectrally equivalent counterparts of the true solutions. Since the phase characteristics of a process is of vital importance in such fields as data communications and seismic exploration, it is necessary to employ the phase-sensitive higher-order cumulants (HOC) or their Fourier transform, *i.e.*, higher-order spectra (HOS).

In 1960's, a group of statisticians began further research on higher-order cumulants [68][69][82][13]. Over the following decade, however, the development of this field was not as fast as expected because of the lack of a sufficiently general theory and and the difficulties in computation. Not until the mid-1970's, when the rapid development of VLSI technology provided people with high speed and low price computers, was the potential of HOC analysis fully recognised. Since then, HOC (or HOS) analysis has been introduced into a wide range of fields: oceanography [74], geophysics [75], biomedicine [76], telecommunications [77], speech processing [78], and the diagnosis of mechanical systems [79].

The definition and properties of higher-order cumulants (HOC) have been demonstrated in Section 2.6 and 2.7. As pointed out earlier, this thesis will in the main be dedicated to the application of HOC to blind deconvolution of NMP signals.

In 1982, K. S. Lii and M. Rosenblatt successfully applied for the first time the higher order spectrum method to the identification and deconvolution of NMP nonGaussian processes (MA model) for the noise free case [12]. Lii and Rosenblatt's work has pioneering significance in the development of the HOC based blind deconvolution technology. Later in 1984, T. Matsuoka and T. J. Ulrych obtained satisfactory results in the case of noisy seismic data by using HOS based algorithms [80]. These successes greatly stimulated the interest of signal processing community in the application of HOC (or HOS) analysis to blind deconvolution of NMP signals. So far, considerable progress has been achieved.

In the beginning, conventional (Fourier type) techniques were adopted to estimate higher order spectra from which the phase characteristics of a signal (or system) can be determined. Like conventional spectral analysis techniques, however, conventional HOS estimation methods also suffer from similar limitations: lower resolution and higher variance. This can cause severe errors in phase extraction, thus, the corresponding deconvolution results normally become very rough. To overcome this problem, M. R. Raghuveer and C. L. Nikias first suggested a parametric approach of HOS estimation (AR model based) in 1985 [81]. Then in 1986, C. L. Nikias proposed a parametric modeling identification method of NMP ARMA channels [77]. Compared with conventional methods, parametric approaches offer the possibility of lower variance and higher resolution (at the cost of system order selection), and remove the step of phase extraction. Consequently, parametric approaches have now become increasingly preferable. In this thesis, we will only investigate parametric techniques.

In system identification by means of higher order-cumulants, G. B. Giannakis in 1987 published a simple and elegant closed-form formula [28]. In 1989, G. B. Giannakis and J. M. Mendel derived another class of closed-form identification formula, which, however, used both HOC and second-order cumulants (autocorrelation functions) and thus has a more complicated form [15]. In the same year, Swami and Mendel [30] also proposed a closed-form recursive estimation method, which has a similar complexity to the above method of Giannakis and Mendel's.

Although the formula given by Giannakis [28] is very simple, its results are often not satisfactory because of the errors in estimation of higher-order cumulants. In this chapter, a two-step correction algorithm for Giannakis' formula is suggested in Section 4.2. By using the approximate results of Giannakis' formula as the initial estimates of the corresponding parameters in an iterative optimization process, both the inaccuracy of Giannakis' formula and the multimodality of the nonlinear optimization problem are overcome. Additionally, a system order determination approach is introduced. Then in Section 4.3, an inverse filter is formed from the obtained model parameters by using a least squares (LS) approach. The order of the inverse filter can also be automatically selected. A similarity criterion based identification scheme is demonstrated in Section 4.4. This scheme results in one less parameter to be identified. Simulation results are presented in each section in order to confirm the feasibility of the corresponding algorithms. Related discussions are presented in Section 4.5. Finally, some conclusions are drawn in Section 4.6.

4.2 Parametrically Optimal Blind Identification: A Two-Step Approach

This section first derives a two-step parametrically optimal blind identification approach, then presents a new system order selection algorithm. Some analysis and simulations are presented in the last two sub-sections.

4.2.1 Algorithm derivation

As stated in Section 2.7 and 4.1, we assume that $w(k)$ is a zero-mean nonGaussian IID process. In this chapter, we will also assume $w(k)$ is an asymmetrically distributed process so that its 3rd-order cumulants are non-zero. For the case of a symmetric distribution, a symmetric-to-asymmetric transformation (SAT) can be introduced in order to make the techniques in this chapter applicable. The detailed discussion of this aspect will be demonstrated in Chapter 6 (Section 6.1). In addition, the AR model based methods, which will be studied in Chapter 5 (and Chapter 6), can also be employed to deal with the symmetrically distributed series. Although we will only investigate the third-order cumulants in this chapter, the relevant formulae can, in principle, be extended to the case of the fourth-order cumulants.

According to the definition of cumulants (see Section 2.6), the expression for the n-th order cumulant $C_x(m_1, m_2, \cdots, m_{n-1})$ of the system output $x(k)$ in Eq(4.1.1) can be derived as follows [15][82]:

$$C_x(m_1, m_2, \cdots, m_{n-1}) = \gamma_n \sum_{i=L}^{U} b_i \prod_{j=1}^{n-1} b_{i+m_j} + C_n(m_1, m_2, \cdots, m_{n-1}) , \quad (4.2.1)$$

where the m_i's represent the time lags (from this chapter, we will use this notation to denote time lag),

$$L = \max(0, -m_1, \cdots, -m_{n-1}) , \quad (4.2.2)$$

$$U = \min(q, q-m_1, \cdots, q-m_{n-1}) , \quad (4.2.3)$$

and

$$\gamma_n = C_w(m_1, m_2, \ldots, m_{n-1})\big|_{m_1=m_2=\ldots=m_{n-1}=0} . \quad (4.2.4)$$

In the above derivation, independence between $n(k)$ and $x_{wf}(k)$ has been assumed. In the 3rd-order case, Eq(4.2.1) reduces to

$$C(m_1, m_2)=\mathrm{E}[X(k)X(k+m_1)X(k+m_2)]$$

$$=\gamma_3\cdot\sum_{i=\max(0,\,-m_1,\,-m_2)}^{\min(q,\,q-m_1,\,q-m_2)} b_i b_{i+m_1} b_{i+m_2}\ ,\quad m_1, m_2=-q,\ \cdots,0,\ \cdots,q\ , \tag{4.2.5}$$

where the subscript for $C_x(\cdot)$ has been dropped, *i.e.*, $C(\cdot)=C_x(\cdot)$, and similarly hereinafter. In addition,

$$\gamma_3 = \mathrm{E}[w^3(k)] \tag{4.2.6}$$

is referred as *skewness*. A series is skewed if $\gamma_3\neq 0$ and unskewed if $\gamma_3=0$. Normally, an asymmetrically distributed series is skewed and a symmetrically distributed series is unskewed.

Notice that, Eq(4.2.5) comprises the basis of all MA model based deconvolution techniques. As the key step of deconvolution process, system identification now reduces to the estimation of the coefficient vector $\vec{b}=\{b_0,\ \cdots,\ b_q\}$ from Eq(4.2.5). Although several closed-form methods have emerged [28][15][30], as mentioned in last section, optimisation algorithms appear to be more attractive. The reasoning for this is as follows: the estimation of cumulants from a finite-length data sequence (especially from a "short" sequence) is usually of a relatively high error; and closed-form formula are sensitive to errors in cumulant estimation [15]. The combination of these two points may make the closed-form methods produce unacceptably poor results. Additionally, the autocorrelation function is also used in some closed-form methods, which can further weaken the robustness of these methods to the presence of additive noise.

To obtain the optimal estimate of $\vec{b}$, the nonlinear least squares cumulant fitting scheme can be adopted [12][83][84]. In the case of Eq(4.2.5), the cost function can be constructed as

$$\epsilon^2=\sum_{m_1=-q}^{q}\sum_{m_2=-q}^{q}\left[\hat{R}(m_1, m_2)-\gamma_3\cdot\sum_{i=\max(0,\,-m_1,\,-m_2)}^{\min(q,\,q-m_1,\,q-m_2)} b_i b_{i+m_1} b_{i+m_2}\right]^2\ , \tag{4.2.7}$$

where, $\hat{R}(m_1, m_2)$ denotes the estimate of $C(m_1, m_2)$, which, in principle, can be calculated as follows:

$$\hat{R}(m_1, m_2)=\frac{1}{k_u-k_l}\sum_{k=k_l}^{k_u} x(k)x(k+m_1)x(k+m_2)\ , \tag{4.2.8}$$

where $k_l=\max(0, -m_1, -m_2)$ and $k_u=\min(N-1, N-m_1-1, N-m_2-1)$. In a practical procedure, however, some other operations such as segmentation and averaging are usually needed to reduce the estimation error in $\hat{R}(m_1, m_2)$. For convenience, we briefly describe a well known segment-average scheme as follows.

First, segment N measured samples into P records of Q samples each, *i.e.*, $N=P\times Q$. Then estimate the cumulant $R^{(i)}(m_1, m_2)$ for the i-th record by

$$R^{(i)}(m_1, m_2)=\frac{1}{Q}\sum_{k=k_l}^{k_u} x^{(i)}(k)x^{(i)}(k+m_1)x^{(i)}(k+m_2)\ , \tag{4.2.9}$$

where $x^{(i)}(k)$ denotes the samples in the i-th record. In addition, $i=1, 2, ..., P$, $k_l=\max(0, -m_1, -m_2)$ and $k_u=\min(Q-1, Q-m_1-1, Q-m_2-1)$. Finally, average $R^{(i)}(m_1, m_2)$ over all records:

$$\hat{R}(m_1, m_2) = \frac{1}{P}\sum_{i=1}^{P} R^{(i)}(m_1, m_2) \quad . \tag{4.2.10}$$

The main purpose of the above averaging process is to reduce the estimation variance of $\hat{R}(m_1, m_2)$.

In practice, only the diagonal slice (where $m_1=m_2=m$) of $R_y(m_1, m_2)$ is used as this results in much reduced computational complexity. Then, Eq(4.2.7) can be simplified as

$$\epsilon^2 = \sum_{m=-q}^{q} [\hat{R}(m, m) - \gamma_3 \cdot \sum_{i=\max(0, -m)}^{\min(q, q-m)} b_i b_{i+m}^2]^2 \quad . \tag{4.2.11}$$

In fact, Eq(4.2.11) was also suggested in [12] and [13]. Theoretically no justification exists for the above simplification, but in the simulation results, it was observed that no obvious detrimental effect on the estimation accuracy of $\vec{b}$ resulted, *i. e.*, the empirical results indicate no problem. Obviously, the optimal estimate of $\vec{b}$ should be the global minimum point of Eq(4.2.11) or Eq(4.2.7). In addition, since only the output series is known, there is always a scale ambiguity for the waveform amplitude. For convenience of comparison, γ_3 is assumed to be known in this chapter.

Notice that, the minimisation of ϵ^2 with respect to $\vec{b}$ in Eq(4.2.11) or Eq(4.2.7) is a multimodal optimisation problem. For this, a gradient or Newton-Raphson type of technique is easy to implement, but it may converge to one of many local minima instead of the global minimum since no *a priori* knowledge about $\vec{b}$ is available. A similar problem arises in [83]. An alternative approach is the searching linear programming (SLP) method described in [12] and [83]. In the SLP method, autocorrelation functions and spectral factorisation are first used to estimate the q zeros of the spectrally equivalent minimum phase system, then linear programming is employed over 2^q sets of zeros to search the set of zeros with which Eq(4.2.11) reaches the global minimum. Clearly, SLP requires the estimation of autocorrelation functions and can involve great amount of computation (especially when q is relatively large) during the spectral factorisation and searching procedure. Hence, in order to obtain the optimal estimate of $\vec{b}$ from only the 3rd-order cumulants, the multimodal nature of the problem must be overcome. The following "two step relay algorithm" is proposed.

STEP 1: Use some method to obtain a relatively rough estimate of $\vec{b}$, which is sufficiently close to the global minimum point of ϵ^2;

STEP 2: Implement a gradient (or Newton-Raphson) optimisation algorithm, in which the initial estimate of $\vec{b}$ is taken to be the result of step 1.

In fact, for step 1, it is possible to use any one of the existing closed-form formulae proposed in [28][15][30]. As explained earlier, the results generated by these formulae may be very poor. But if they are used only as the initial estimates of the optimization processes, they are, intuitively, still much more advisable than arbitrarily chosen initial estimates. Since both the Giannakis-Mendel formula suggested in [15] and the Swami-Mendel

formula in [30] are relatively more complicated in comparison to Giannakis formula proposed in [28], the latter is naturally more preferable. Consequently, the following modified version of Giannakis' formula [28] is adopted here. The later simulation will justify this choice.

From Eq(4.2.5), it is necessary to have

$$C(q,l)=\gamma_3 b_0 b_q b_l \quad , \tag{4.2.12}$$

and

$$C(q,0)=\gamma_3 b_0^2 b_q \quad , \tag{4.2.13}$$

and then, it is easy to have

$$b_l = b_0 \cdot \frac{C(q,l)}{C(q,0)} \quad , \quad l=1, \ldots, q \, , \tag{4.2.14}$$

where b_0 can be determined as

$$b_0=\left[\frac{C^2(q,0)}{\gamma_3 \cdot C(q,q)}\right]^{\frac{1}{3}} \quad . \tag{4.2.15}$$

To derive Eq(4.2.15), we only need to consider Eq(4.2.13) and

$$C(q,q)=\gamma_3 b_0 b_q^2 \, . \tag{4.2.16}$$

Certainly, $C(\cdot)$ should be replaced with $\hat{R}(\cdot)$ in practice.

In step 2, the well known steepest descent method can be employed. From Eq(4.2.11), the gradient can be written as

$$\begin{aligned} \nabla_l &= \frac{\partial \epsilon^2}{\partial b_l} \\ &= 2 \cdot \sum_{m=-q}^{q} \left[\gamma_3 \cdot \sum_{i=max(0,\,-m)}^{min(q,\,q-m)} b_i b_{i+m}^2 - \hat{R}(m,m)\right](b_{l+m}^2 + 2 b_l b_{l-m}) \, , \\ & \qquad l=0, \ldots, q \, , \end{aligned} \tag{4.2.17}$$

where the summations include all non-zero terms. Although Eq(4.2.17) is a system of 5th-order simultaneous equations and has many sets of solutions, local minima, including a trivial one: $\{0, 0, \cdots, 0\}$, the global minimum of ϵ^2 can normally still be reached since the initial estimate is generally close enough to it.

In summary, the above two-step relay technique can be written as the following algorithmic form:

1): Initialize $\vec{b}$ by using Eq(4.2.15) and Eq(4.2.14);

2): Calculate ∇_l by using Eq(4.2.17), $l=0, \cdots, q$;

3): $\max(|\nabla_l|)<\zeta$ (ζ is a very small preset positive real number)?

If yes, stop.

If no, $b_l - \nabla_l \tau \rightarrow b_l$, $l=0, \cdots, q$, and go to *2).*

Here, τ is the step-size factor which controls the convergence rate.

4.2.2 System order selection

In the above discussion, it was assumed that the MA model order q was known. If q is not known, some HOC based methods [15][85][86] must be developed to determine it, since the previous autocorrelation function and Gaussian distribution based criteria such as the Akaike Information Criterion (AIC) do not apply here. A simple MA based method is proposed here, and it is as follows.

Assuming that the MA model order is J, from Eq(4.2.5), a new criterion function can be formed:

$$S(J) = \sum_{l=0}^{J} C^2(J, l) = b_J^2 \cdot (\gamma_3 b_0)^2 \cdot \sum_{l=0}^{J} b_l^2 \quad . \tag{4.2.18}$$

Considering that the cumulants are blind to the linear phase shift, we can always assume $b_0 \neq 0$. Then,

$$(\gamma_3 b_0)^2 \cdot \sum_{l=0}^{J} b_l^2 \neq 0 \tag{4.2.19}$$

must hold. Hence, we obtain the following theorem.

[THEOREM] If there exists a J_0, which makes

$$S(J) = 0 \tag{4.2.20}$$

for any $J \geq J_0$, then the order of the MA model is

$$q = J_0 - 1 \qquad \blacksquare \tag{4.2.21}$$

In fact, however, this theorem is only of theoretical value. The reason for this is that there always exists an error in the estimation of $C(m_1, m_2)$, which ensures $S(J)$ never vanishes even when $J > q$. Fortunately, we have observed an interesting phenomenon in our study: with variation of J ($J = 1, 2, \cdots$),

$$\hat{S}(J) = \sum_{l=0}^{J} \hat{R}^2(J, l) \tag{4.2.22}$$

reaches the minimum when $J = q + 1$. Hence, the following empirical method can be employed to determine q.

STEP 1: According to experience and computing limitation, preset J as a positive integer J_u, which should be "higher" than the anticipated order q;

STEP 2: Calculate $\hat{S}(J)$, $J = 1, \ldots, J_u$;

STEP 3: Search the minimum: $\hat{S}(J_m) = \min_J \{\hat{S}(1), \ldots, \hat{S}(J), \ldots, \hat{S}(J_u)\}$;

STEP 4: $q = J_m - 1$.

Our work shows the above method is normally reliable in the case of true MA channels as long as the estimation error of cumulants is not too high*. In fact, it is this case that requires the accurate order estimation, because the incorrect q value can cause such a serious error to the initial estimate in Eq(4.2.15) and Eq(4.2.14) that the identification algorithm fails to converge to the global minimum point. In the case of ARMA channels, the order of the equivalent MA model produced by the above method sometimes is not

* Our work indicated that the average relative error of cumulant estimations should approximately be less than 25%.

appropriate. But in this case, fortunately, the identification algorithm is not so sensitive to different MA order estimates, and a trial-and-error based scheme can be used. It has recently been brought to our attention that a similar formula was developed independently by Giannakis is his thesis [87].

4.2.3 Algorithm analysis

For the two-step relay identification algorithm suggested above, we have the following comments.

1) System phase: The algorithm can deal with both MP and NMP systems.

2) Input signals: The input signals must be non-Gaussian. It is well known that a NMP system is not identifiable if it is driven by a Gaussian random signal.

3) Assumptions for the additive noise $n(k)$: As pointed out in Section 2.6, the higher (≥ 3) order cumulants of a Gaussian process vanish. As a result, purely HOC based algorithms possess a very strong robustness to the interference of $n(k)$ if it is Gaussian. Furthermore, in our algorithm where only the 3rd-order cumulants are involved, the requirement for $n(k)$ can be relaxed to be "symmetrically distributed". The reason for this is that the 3rd-order (or any odd number order) cumulants of symmetrically distributed processes are zero.

4) Ambiguity of linear phase shift: Since higher-order cumulants are completely blind to the linear phase shift of system impulse responses, the algorithm actually assumes $b_0 \neq 0$, and the only effect with this is making the identification result present a linear phase left-shift in comparison to the true system impulse response.

5) Further study: The further detailed simulations will be implemented in the scenario of channel equalisation in Section 6.2.

4.2.4 Simulation results

To verify the above algorithm, three computer simulation examples are demonstrated in this subsection. In these simulations, $w(k)$ is taken to be an exponentially distributed IID series with zero mean, $\sigma^2=1$, and $\gamma^3=2$. $n(k)$ is Gaussian white noise. The total number of samples is 9000. In order to improve the accuracy of $\hat{R}(m_1, m_2)$ estimation, we segment the system output into 30 subrecords, and then estimate $\hat{R}(m_1, m_2)$ according to the segment-average scheme (Eq(4.2.9) and Eq(4.2.10)). Here, $N = P \times Q = 30 \times 300$.

In addition, the signal-to-noise ratio (SNR) in this thesis is defined as

$$SNR \text{ (dB)} = 10\log_{10}\left(\frac{\mathrm{E}[x^2(k)]}{\mathrm{E}[n^2(k)]}\right) . \tag{4.2.23}$$

In parameter updating, the Gauss-Seidel iteration technique is employed.

EXAMPLE (4.2.a): 2nd-order MA model:

$$x(k) = 1 - 2.3w(k-1) + 0.6w(k-2) + n(k) , \tag{4.2.24}$$

which is a NMP process with two zeros: $z_1=2$ and $z_2=0.3$.

In the case of 60 dB noise, the result of Giannakis' method (GM) is obtained and shown in Table (4.1). Then we take this result as the initial estimate of the iteration

optimization, and obtain the optimal (OPT) result, which is also listed in Table (4.1).

In the case of 10 dB and 5 dB noise, the above steps are repeated, and the corresponding results are illustrated in Table (4.1) as well. For completeness, the result for 5dB noise is also plotted in Fig(4.1).

Table (4.1) Results of Example (4.2.a).

SNR	algorithm	b_0	b_1	b_2	ϵ^2
	true $\vec{b}$	1.00000	-2.30000	0.60000	0.00000
60dB	GM	0.94509	-2.37819	0.59213	1.31109×10^{1}
60dB	OPT	0.95480	-2.27218	0.59973	2.99523×10^{-3}
10dB	GM	0.90601	-2.48740	0.58722	3.03432×10^{1}
10dB	OPT	0.94877	-2.27376	0.58907	1.58776×10^{-2}
5dB	GM	0.87310	-2.48369	0.58029	5.78282×10^{1}
5dB	OPT	0.94394	-2.27589	0.57932	3.53607×10^{-2}

EXAMPLE (4.2.b): 4th-order MA model:

$$x(k) = w(k)+2w(k-1)+3w(k-2)+5w(k-3)+w(k-4)+n(k) \quad , \qquad (4.2.25)$$

which is also a NMP process with 4 zeros: $z_{1,2}=-0.01743\pm1.59282$, $z_3=-1.73843$, and $z_4=-0.22670$. The same simulation as in Example (4.2.a) is implemented in the cases of 60 dB, 10 dB, and 5 dB noise, respectively, and the corresponding results are given in Table (4.2). The result for 5 dB noise is also drawn in Fig(4.2).

Table (4.2) Results of Example (4.2.b).

SNR	algorithm	b_0	b_1	b_2	b_3	b_4	ϵ^2
	true $\vec{b}$	1.00000	2.00000	3.00000	5.00000	1.00000	0.00000
60dB	GM	1.18893	1.98561	2.74182	3.83477	1.14676	4.17401×10^{4}
60dB	OPT	0.93859	2.01711	3.15318	5.04987	1.05421	4.22229×10^{0}
10dB	GM	1.13634	2.06415	2.80303	3.93140	1.16967	3.40373×10^{4}
10dB	OPT	0.94408	2.01659	3.16018	5.04327	1.04564	2.85831×10^{0}
5dB	GM	1.08203	2.13654	2.85664	4.01431	1.20086	2.77191×10^{4}
5dB	OPT	0.94898	2.01485	3.16502	5.03899	1.03932	2.34056×10^{0}

From the above two examples, It can easily be seen that the $\vec{b}$ determined by GM is located far from the bottom of ϵ^2 surface (because the correspondent ϵ^2 value is very large),

and is of a relatively larger error. In comparison, the result of OPT is much more accurate and nearly at the bottom point of the ϵ^2 surface (the corresponding ϵ^2 value is also much smaller than in GM). In addition, these two examples also fully demonstrate the robustness of our algorithm to the presence of additive noise.

EXAMPLE (4.2.c): ARMA model:

$$x_{nf}(k)-0.5x_{nf}(k-1) = w(k)+2w(k-1) \ , \tag{4.2.26a}$$

and

$$x(k) = x_{nf}(k)+n(k) \ , \tag{4.2.26b}$$

which, obviously, is also a NMP process. In simulation, we set the order of the equivalent MA model to be 8. For simplicity, here we only give the result in the case of 5 dB noise, which is shown in Table (4.3) and also in Fig(4.3).

Table (4.3) Result of Example (4.2.c).

SNR	algorithm	h_0	h_1	h_2	h_3	h_4
	true $\vec{h}$	1.00000	2.50000	1.25000	0.62500	0.31250
5dB	GM	0.17342	-0.26878	-0.31595	0.36988	0.82022
5dB	OPT	1.08360	2.54306	1.26516	0.62854	0.29067

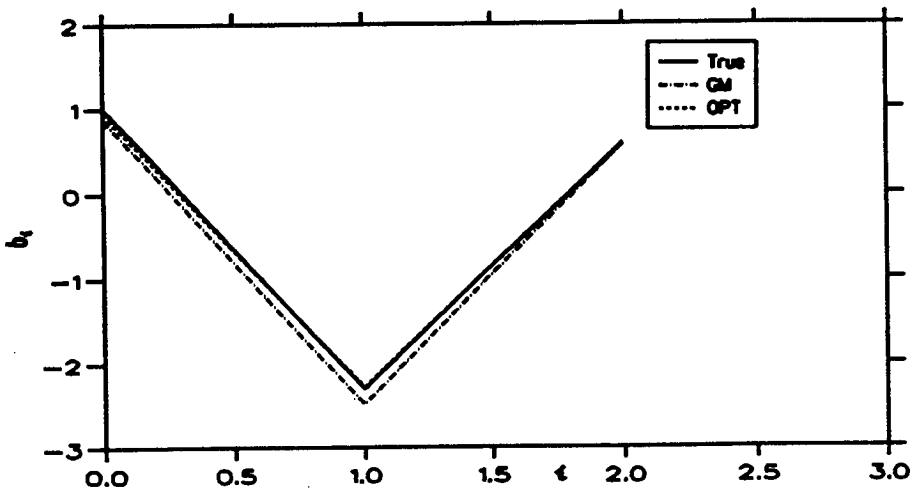
Table (4.3) (continued) Result of Example (4.2.c).

SNR	algorithm	h_5	h_6	h_7	h_8	ϵ^2
	true $\vec{h}$	0.15625	0.07812	0.03906	0.01953	0.00000
5dB	GM	1.11476	0.30661	-0.67120	-1.29485	3.06785×10^3
5dB	OPT	0.11997	0.13774	0.09978	0.04259	2.42479×10^0

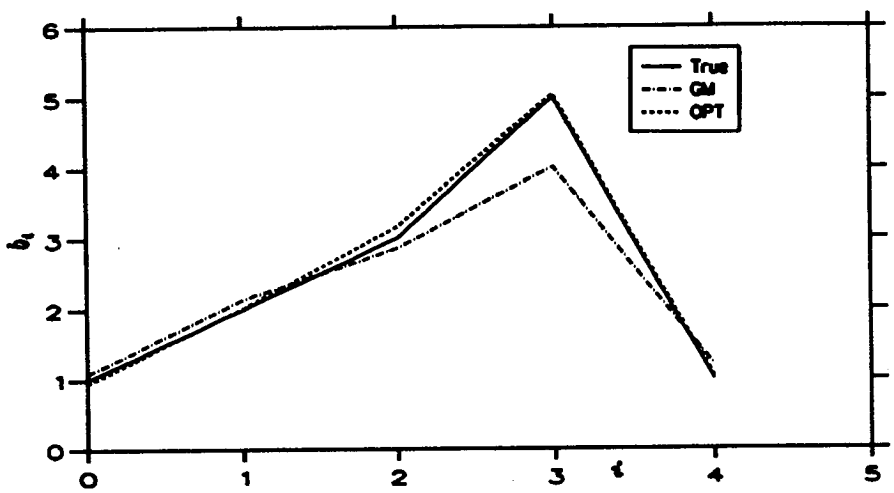
($\vec{h}$ denotes system impulse response)

This example again shows the strong robustness of our algorithm to the presence of additive noise. Although the result of Giannakis' method (GM) is very far from the true optimal solution, a very satisfactory result is still obtained by using our two-step relay approach.

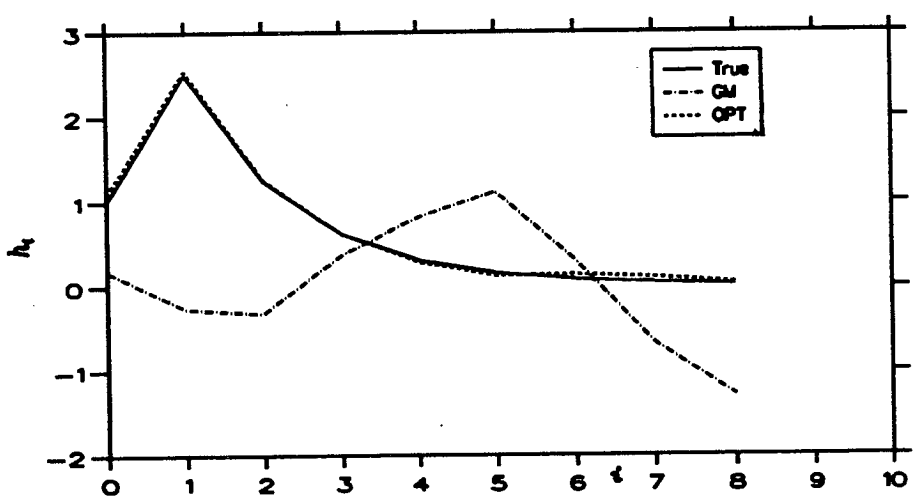
From the above description, it can be concluded that, the proposed algorithm, a combination of nonlinear least squares fitting and a simple closed-form formulae, is very effective and accurate to the problem of blind identification of NMP systems, and it is not only optimal but also very robust to the additive noise.



Fig(4.1) Results of Example (4.2.a): 5 dB noise.



Fig(4.2) Results of Example (4.2.b): 5 dB noise.



Fig(4.3) Results of Example (4.2.c): 5 dB noise.

4.3 Signal Restoration

The final objective of deconvolution is to restore the system input signals. After system identification, the kernel of signal restoration is to construct an appropriate inverse filter. This section first exhibits an inverse filter design algorithm, then presents several simulations.

4.3.1 Algorithm

Since it is assumed that the channel can be nonminimum phase, the general form of the inverse filter is noncausal, *i.e.*, the transfer function of the inverse filter should be

$$H^{-1}(z)=\frac{1}{H(z)}=\frac{1}{\sum_{i=0}^{q} b_i z^{-q}}$$

$$=\sum_{i=r_1}^{r_2}\theta_i z^{-i} \quad , \tag{4.3.1}$$

where r_1 is the order of the noncausal part of the inverse filter, and r_2 is the order of the causal part. The determination of θ_i's can be accomplished by the following procedures.

(a) Selection of r_1 and r_2

Let

$$H(z)=b_0\prod_{i=1}^{q}(1-z_i z^{-1}) \quad , \tag{4.3.2}$$

where the meaning of z_i is obvious, then

$$H^{-1}(z)=\sum_{i=1}^{q}\frac{A_i}{1-z_i z^{-1}} \quad . \tag{4.3.3}$$

Let

$$|z_{m_1}|,\ |z_{m_2}|,\ \ldots,\ |z_{m_s}|<1 \quad , \tag{4.3.4a}$$

and

$$|z_{n_1}|,\ |z_{n_2}|,\ \ldots,\ |z_{n_t}|>1 \quad , \tag{4.3.4b}$$

where $s+t=q$. Then, for any z_{m_i}, $i=1,\ \ldots,\ s$, clearly,

$$\frac{A_{m_i}}{1-z_{m_i}z^{-1}}=A_{m_i}\sum_{l=0}^{+\infty}(z_{m_i}z^{-1})^l \quad . \tag{4.3.5}$$

To determine r_2, an accuracy control factor ζ can be preset, which is a small positive number (certainly, $\zeta<1$), then ignoring the terms which satisfy

$$|z_{m_i}|^l<\zeta \quad ,\ i=1,\ \ldots,\ s, \tag{4.3.6}$$

and assuming that

$$|z_{m_M}|=\max_{i=1,\ \ldots,\ s}\{|z_{m_i}|\} \quad , \tag{4.3.7}$$

it is only necessary to choose

$$r_2=[\ln(\zeta)/\ln(|z_{m_M}|)]+1 \quad , \tag{4.3.8}$$

where $[\cdot]$ denotes taking the integer part of a number. If there is no z_{m_l} with $|z_{m_l}|<1$, *i. e.*, in the case of maximum phase, then r_2 can be selected as

$$r_2 = -1 \quad . \tag{4.3.9}$$

Similarly, considering z_{n_i}, $i=1, \ldots, t$, leads to

$$r_1=[\ln(\zeta)/\ln(|z_{n_N}|)]-1 \quad , \tag{4.3.10}$$

where

$$|z_{n_N}| = \max_{i=1, \ldots, t} \{|z_{n_i}|\} \quad . \tag{4.3.11}$$

If there is no z_{n_i} with $|z_{n_i}|>1$, i.e., in the case of minimum phase, then r_1 is selected to be:

$$r_1 = 0 \quad . \tag{4.3.12}$$

Therefore, the automatic selection of the inverse filter orders can be realized as long as the accuracy control factor ζ is input.

(b) Determination of θ_i's

In order to determine the θ_i's, we can first arbitrarily take L $(\geq r_2-r_1+1)$ different points on the unit circle:

$$g_l = e^{j\alpha_l} \quad , \quad l=0, \ldots, L, \tag{4.3.13}$$

where α_l is a uniformly random number over the interval $[0, 2\pi]$. Then, θ_i's can be obtained by minimising

$$\epsilon^2 = \sum_{l=0}^{L} |\sum_{i=r_1}^{r_2} \theta_i g_l^{-i} - H^{-1}(g_l)|^2 \quad , \tag{4.3.14}$$

considering Eq(4.3.13), Eq(4.3.14) can be written as,

$$\epsilon^2 = \sum_{l=0}^{L}\{[\sum_{i=r_1}^{r_2} \theta_i \cos(i\alpha_l) - H_R^{-1}(\alpha_l)]^2 + [\sum_{i=r_1}^{r_2} \theta_i \sin(i\alpha_l) + H_I^{-1}(\alpha_l)]^2\} \ , \tag{4.3.15}$$

where,

$$H_R^{-1}(\alpha_l) = Real(H^{-1}(g_l)) \quad , \tag{4.3.16a}$$

and

$$H_I^{-1}(\alpha_l) = Imag(H^{-1}(g_l)) \quad . \tag{4.3.16b}$$

Obviously, Eq(4.3.14) or Eq(4.3.15) is a unimodal Least Squares (LS) problem. Considering that θ_i's are real numbers, the normal equation of the above LS problem can be written as

$$\sum_{i=r_1}^{r_2} \theta_i \sum_{l=0}^{L} \cos((i-n)\alpha_l) = \sum_{l=0}^{L} [H_R^{-1}(\alpha_l)\cos(n\alpha_l) - H_I^{-1}(\alpha_l)\sin(n\alpha_l)] \quad , \tag{4.3.17}$$

$$n = r_1, \ldots, r_2 \ .$$

The parameters θ_i's can be easily obtained by solving this set of simultaneous equations. Clearly, the determined θ_i's are optimal in the least square sense. The system output signal $x(k)$ forms the input to the inverse filter designed above, and the restored signal can be obtained thus::

$$\hat{w}(k) = \sum_{i=r_1}^{r_2} \theta_i x(k-i) \quad . \tag{4.3.18}$$

4.3.2 Simulation results

This subsection will present three simulations of the two-step identification algorithm combined with the above restoration procedure. The same data series as in Section 4.2 is used, *i.e.*, $w(k)$ is an exponentially distributed process with $\mu=0$, $\sigma^2=1$, and skewness $\gamma_3=2$. The same segmentation-averaging procedure is also adopted: $N=P\times Q=300\times 30$. For convenience of comparison later, the first 150 points of $w(k)$ is plotted in Fig(4.4).

EXAMPLE (4.3.a): NMP MA model:

$$x(k) = 1-2.3w(k-1)+0.6w(k-2)+n(k) \quad . \tag{4.3.19}$$

The identification of this model for different additive noise levels has been investigated in Example (4.2.a). In this example, however, SNR is assumed to be 25 dB. First, the result of the Giannakis method is directly input into the restoration procedure described above in this section, and the reconstructed series $\hat{w}_{GM}(k)$ is displayed in Fig(4.5a) and the error $e_{GM}(k)=\hat{w}_{GM}(k)-w(k)$ in Fig(4.5b). The obtained inverse filter is illustrated in Fig(4.5c). Then, the result of our two-step algorithm is used to form the inverse filter in the restoration procedure, and the restored series $\hat{w}_{OPT}(k)$ and the error $e_{OPT}(k)=\hat{w}_{OPT}(k)-w(k)$ are drawn in Fig(4.6a) and Fig(4.6b), respectively. The corresponding inverse filter is illustrated in Fig(4.6c).

Clearly, the inverse filter resulting from our two-step algorithm is much more accurate than that directly from Giannakis' formula. Consequently, as has been seen, the deconvolution error of our algorithm is much smaller than that of Giannakis' method.

EXAMPLE (4.3.b): NMP ARMA model:

$$x_{nf}(k)+0.2x_{nf}(k-1) = 1+2w(k-1) \quad , \tag{4.3.20a}$$

and

$$x(k) = x_{nf}(k)+n(k) \quad . \tag{4.3.20b}$$

The SNR is 30 dB in this simulation. Similarly, by employing only Giannakis' formula, the following equivalent MA model is acquired:

$$\begin{aligned} x(k) = {} & 0.54172w(k)-0.17177w(k-1)-0.21535w(k-2) \\ & -0.02929w(k-3)+0.34758w(k-4)+0.43210w(k-5)+0.04367w(k-6) \\ & -0.00570w(k-7)-0.43109w(k-8) \quad . \end{aligned} \tag{4.3.21}$$

The impulse response of the above MA model is shown in Fig(4.7a). Obviously, it is very different from the true impulse response of the ARMA model in Eq(4.3.20a). Also, the zeros of the above MA model can be determined and listed as follows:

	$z_{1,2}$	$z_{3,4}$	$z_{5,6}$	z_7	z_8
Real	0.95902	−0.64306	−0.07951	0.83013	−0.98593
Imag.	±0.67159	±0.71241	±0.87395	0.00000	0.00000

It can be seen that trouble arises here because the zero z_8 is nearly on the unit circle! As a result, the inverse filter for this MA model does not exist (or, more exactly speaking, the inverse filter will become unacceptably long), and the restoration procedure cannot be implemented at all.

Next, as in last example, our two-step algorithm is adopted, and the following equivalent MA model is achieved:

$$\begin{aligned} x(k) &= 1.06076w(k)+1.78021w(k-1)-0.36216w(k-2) \\ &+0.09590w(k-3)-0.05943w(k-4)-0.02936w(k-5)+0.09420w(k-6) \\ &-0.00428w(k-7)+0.01555w(k-8) \quad . \end{aligned} \tag{4.3.22}$$

For comparison, we also demonstrate the impulse response of this model in Fig(4.7a), which is obviously extremely approximate to the true. On the other hand, the zeros of the above model are as follows:

	$z_{1,2}$	$z_{3,4}$	$z_{5,6}$	z_7	z_8
Real	0.49862	−0.13015	0.02943	−1.88678	−0.58726
Imag.	±0.25845	±0.48975	±0.40312	0.00000	0.00000

The inverse filter designed from this new model (*i.e.*, Eq(4.3.22)) is shown in Fig(4.7b). The corresponding deconvolved series $\hat{w}_{OPT}(k)$ and error $e_{OPT}(k)=\hat{w}_{OPT}(k)-w(k)$ are plotted in Fig(4.7c) and Fig(4.7d), respectively. Clearly, the result of our algorithm is very accurate.

EXAMPLE (4.3.c): NMP ARMA model:

$$x_{nf}(k)-0.6x_{nf}(k-1) = w(k)-3.3w(k-1)+0.9w(k-2) \tag{4.3.23a}$$

$$x(k) = x_{nf}(k)+n(k) \quad . \tag{4.3.23b}$$

The SNR is taken to be 15 dB in this example. Following the similar steps to above, we can first obtain the model from only Giannakis' method:

$$\begin{aligned} x(k) &= 0.26246w(k)-0.63380w(k-1)+0.22920w(k-2) \\ &-0.23163w(k-3)-0.24697w(k-4)-0.41744w(k-5)+0.33244w(k-6) \\ &0.12921w(k-7)+0.37675w(k-8)-1.53789w(k-9) \quad . \end{aligned} \tag{4.3.24}$$

Again, it can be easily found that the inverse filter for the above model does not exist. On the other hand, our algorithm generates the model:

$$x(k) = 0.94710w(k)-2.69776w(k-1)-0.73327w(k-2)$$

$$-0.46767w(k-3)-0.24215w(k-4)-0.07760w(k-5)-0.10044w(k-6)$$
$$-0.08557w(k-7)-0.06210w(k-8)-0.02026w(k-9) \quad . \qquad (4.3.25)$$

The impulse responses of both model Eq(4.3.24) and Eq(4.3.25) are shown in Fig(4.8a). The inverse filter obtained from model Eq(4.3.25) is illustrated in Fig(4.8b). Fig(4.8c) and Fig(4.8d) respectively show the corresponding deconvolution result $\hat{w}_{OPT}(k)$ and error $e_{OPT}(k)=\hat{w}_{OPT}-w(k)$. The relatively large error amplitude in Fig(4.8c) is due to the high additive noise level adopted in the simulation.

The last two examples indicate that Giannakis' method may cause not only the serious errors but also the non-existence of inverse filters. In fact, this phenomenon has been observed many times in our work. For example, it can also be seen with system model

$$x_{nf}(k)-0.5x_{nf}(k-1) = w(k)+3w(k-1) \quad , \qquad (4.3.26)$$

and so forth. It is our two-step algorithm that gets rid of this serious difficulty by optimising the results of Giannakis' method.

The results of the examples exhibited above have confirmed the feasibility and efficiency of our technique. The application of this technique to NMP channel equalisation will be studied in Chapter 6. But before ending the current chapter, let us next derive a variant of our two-step algorithm.

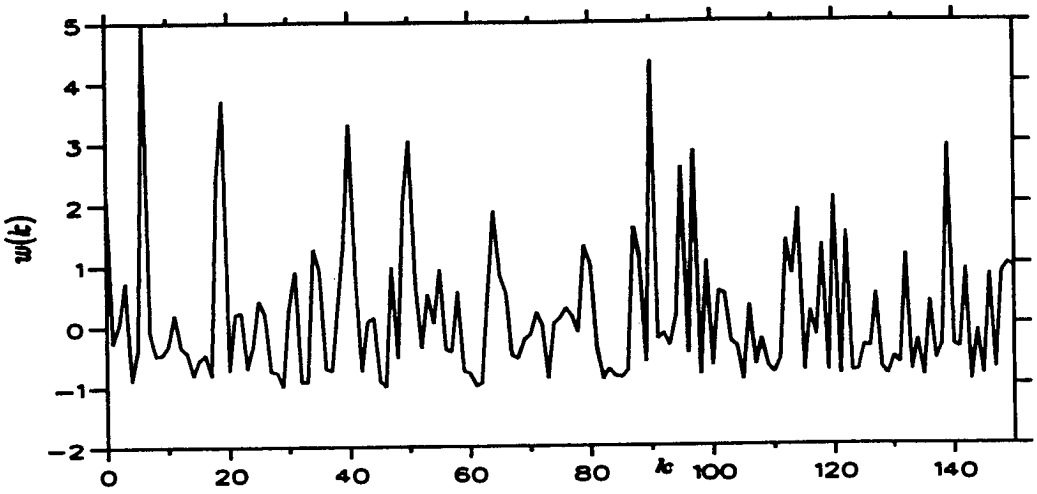
4.4 Blind Deconvolution via Similarity Criterion

Recall Eq(4.2.11) in Section 4.2. To minimise ϵ^2 in Eq(4.2.11), we actually have the following two alternatives. 1) Let γ_3 be a constant and then minimise ϵ^2 with respect to b_i, $i=0, 1, \ldots q$. This is exactly as before. 2) Minimise ϵ^2 with respect to $\beta=b_0^3\gamma_3$ and $b_i'=b_i/b_0$, $i=1, \ldots, q$. (apparently, $b_0'=1$). It is easy to find that there are $q+1$ parameters needing to be estimated in both of these two schemes. For the convenience, we will denote b_i' as b_i hereinafter wherever there is no ambiguity. Now, the question is, are all these b_i's necessary to our identification or deconvolution process? This section will demonstrate that the answer is: No. In fact, only q parameters are essential. For example, b_0 in the first scheme and β in the second scheme are redundant. Taking this point into consideration, a new cost function is suggested in this section, which is based on the concept of similarity, instead of the identity actually used in Eq(4.2.11).

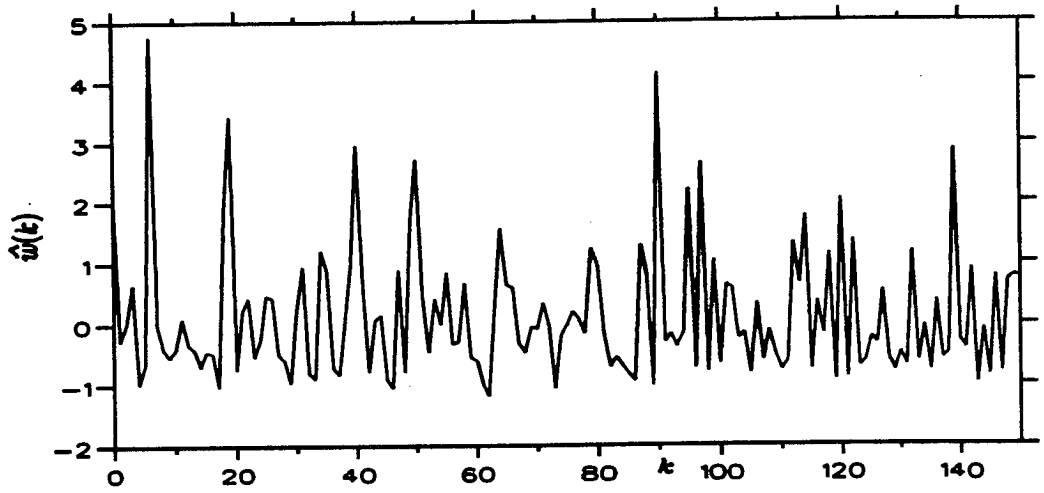
After the derivation of the new similarity based cost function for blind identification, a simulation example is also presented in this section to prove the feasibility of the proposed approach.

4.4.1 Preliminary discussion on similarity

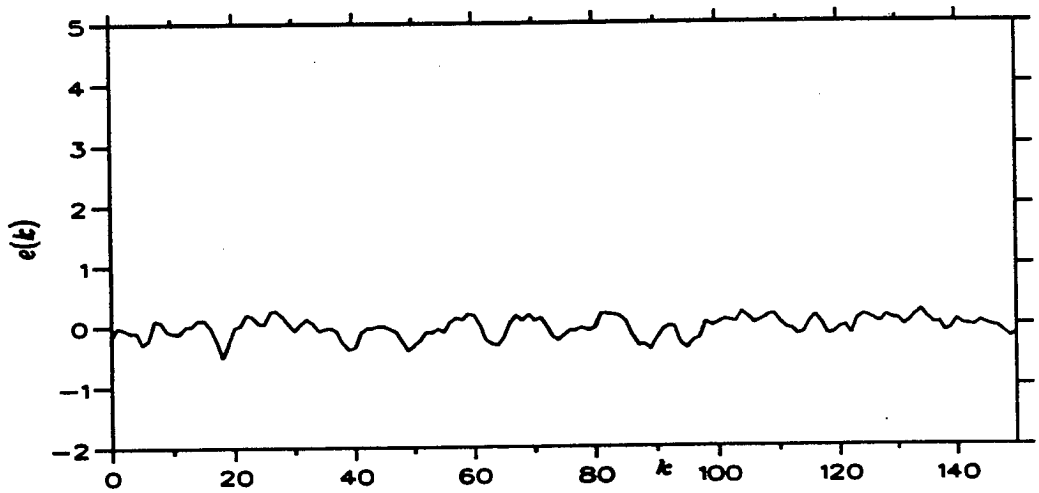
The parameter identification problem can be more generally formulated as follows. Let $\psi=\{\psi_i \,|\, i=0, \ldots, q\}$ be the parameter set needing to be identified, $\mathbf{g}=\{g_i \,|\, i=0, \ldots, N\}$ the measured data set, and $f(\psi, i)$ the assumed model. To estimate ψ, the cost function is constructed as



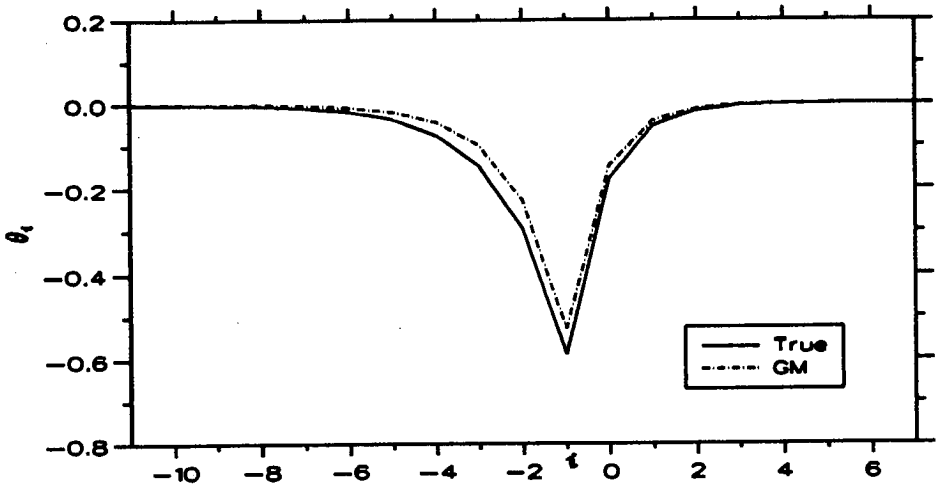
Fig(4.4) The true input series w(k).



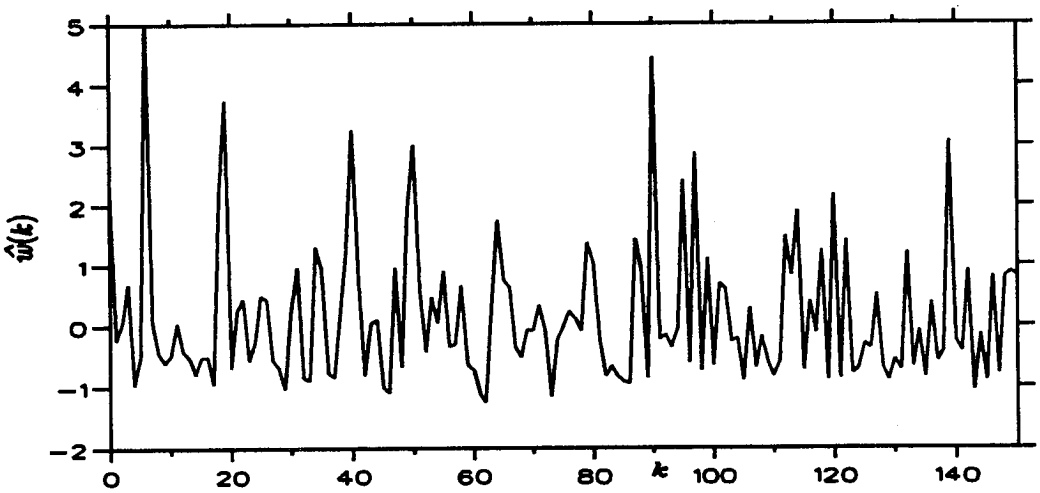
Fig(4.5a) Deconvolved series $\hat{w}(k)$ in Example (4.3.a): Giannakis method, 25 dB noise.



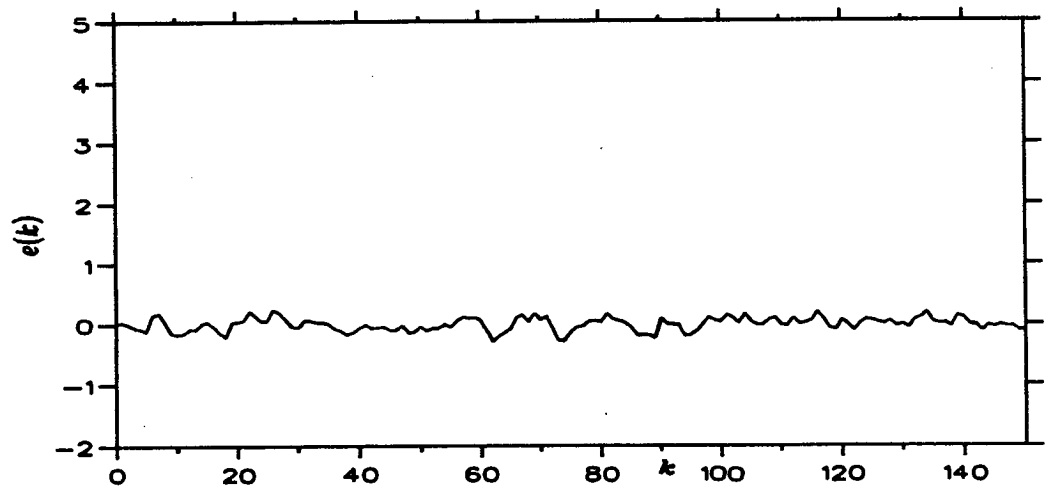
Fig(4.5b) Deconvolution error e(k) in Example (4.3.a): Giannakis method, 25 dB noise.



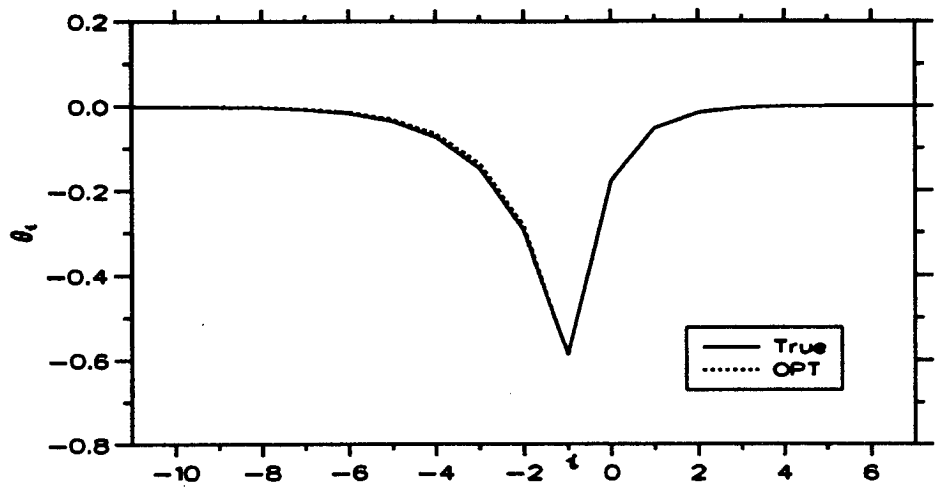
Fig(4.5c) Determined inverse filter in Example (4.3.a): Giannakis method, 25 dB noise.



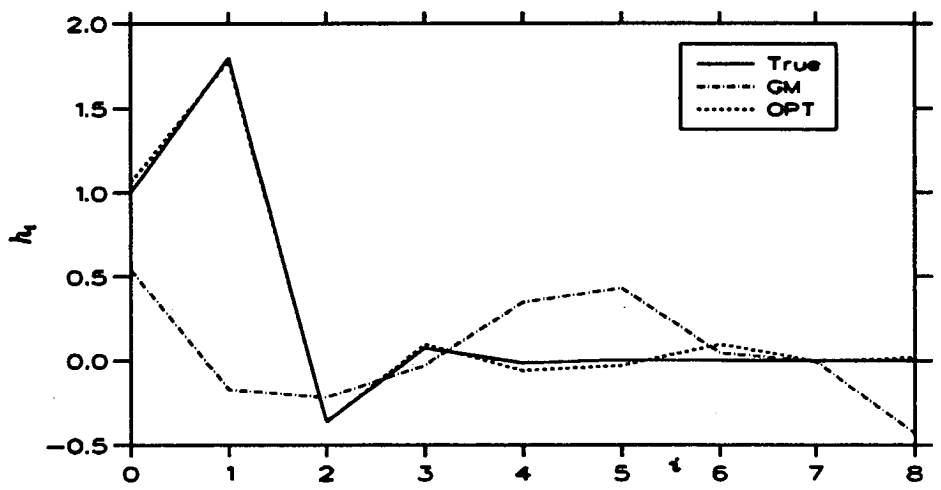
Fig(4.6a) Deconvolved series $\hat{w}(k)$ in Example (4.3.a): proposed algorithm, 25 dB noise.



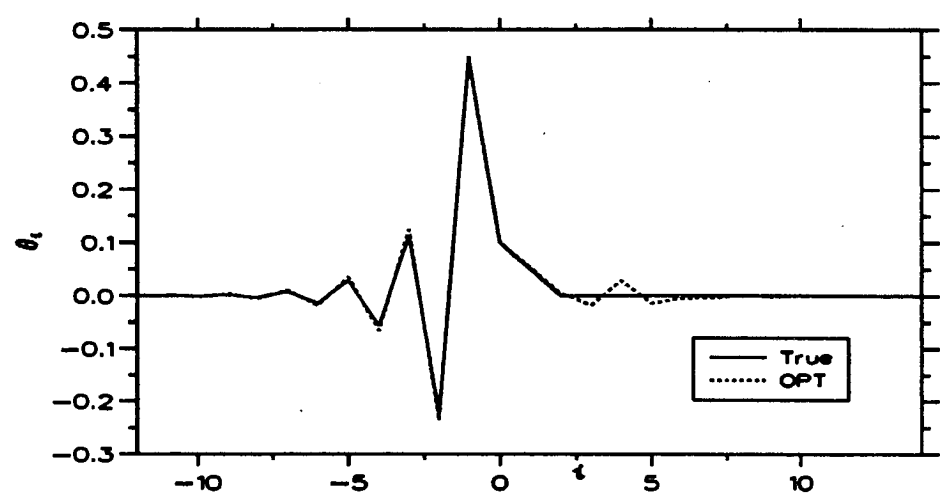
Fig(4.6b) Deconvolution error e(k) in Example (4.3.a): proposed algorithm, 25 dB noise.



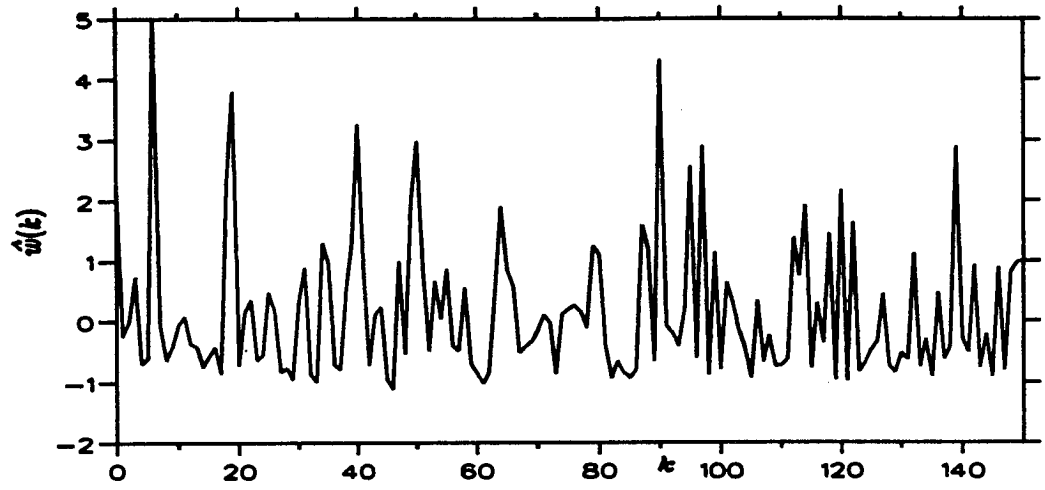
Fig(4.6c) Determined inverse filter in Example (4.3.a): proposed algorithm, 25 dB noise.



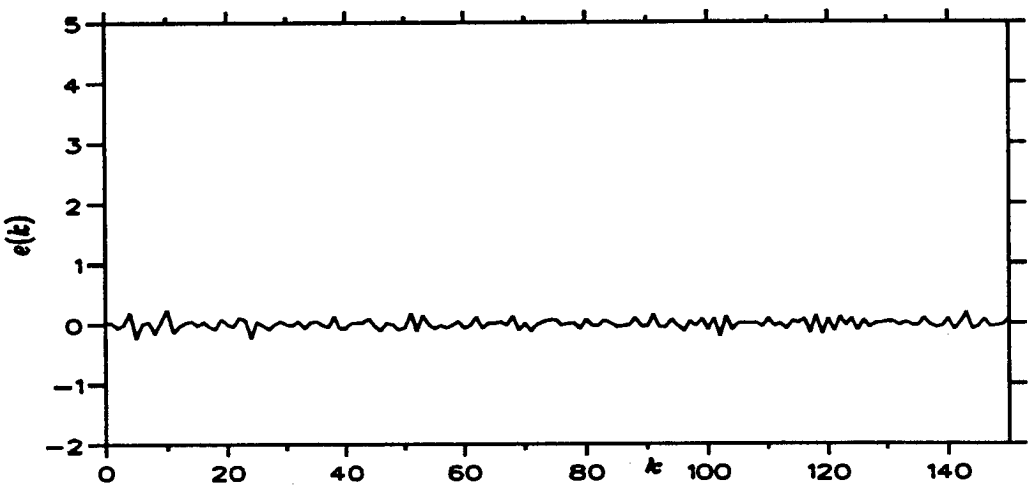
Fig(4.7a) Identification results in Example (4.3.b): 30 dB noise.



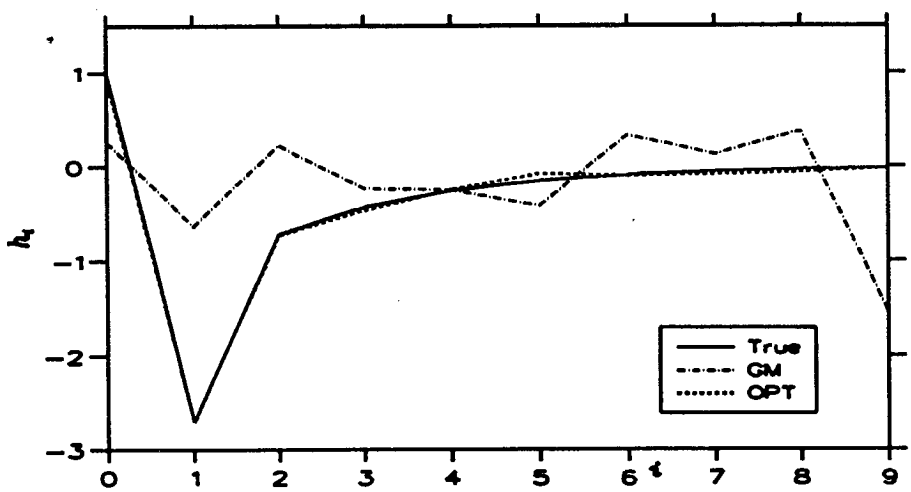
Fig(4.7b) Determined inverse filter in Example (4.3.b): proposed algorithm, 30 dB noise.



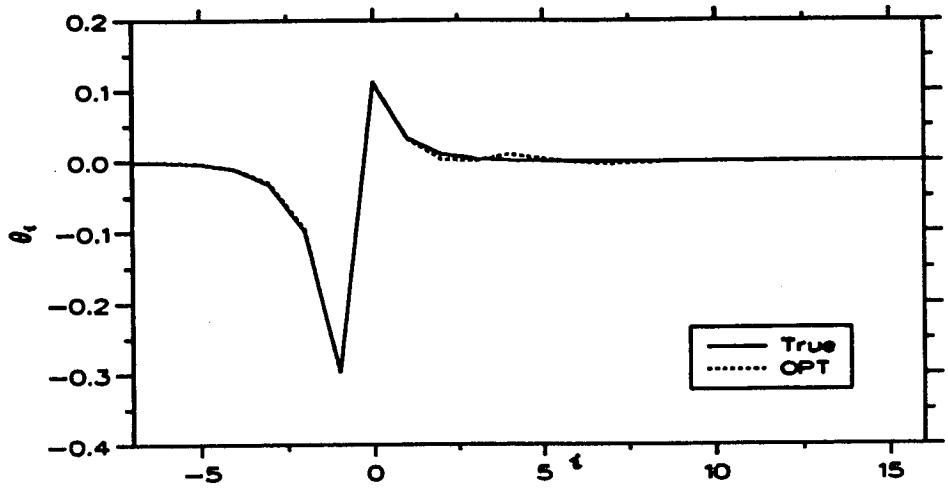
Fig(4.7c) Deconvolved series $\hat{w}(k)$ in Example (4.3.b): proposed algorithm, 30 dB noise.



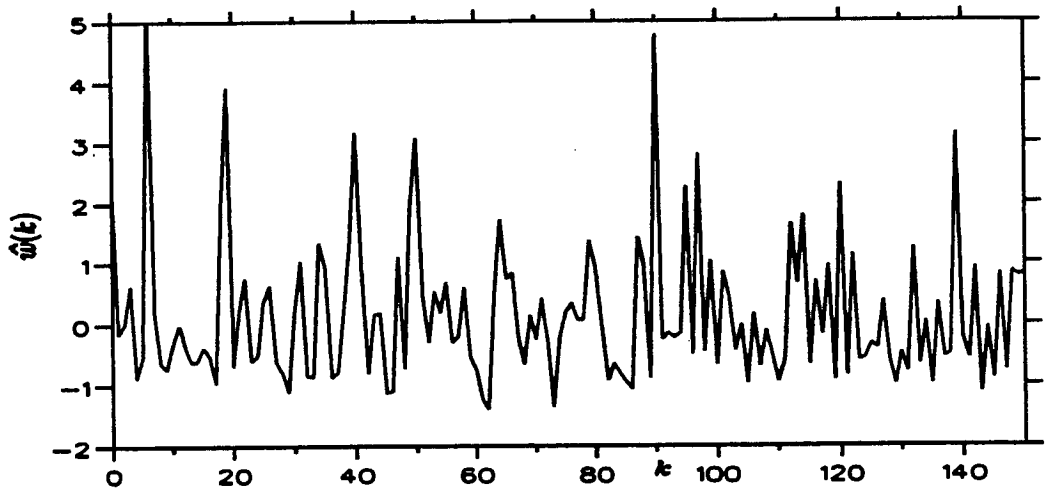
Fig(4.7d) Deconvolution error e(k) in Example (4.3.b): proposed algorithm, 30 dB noise.



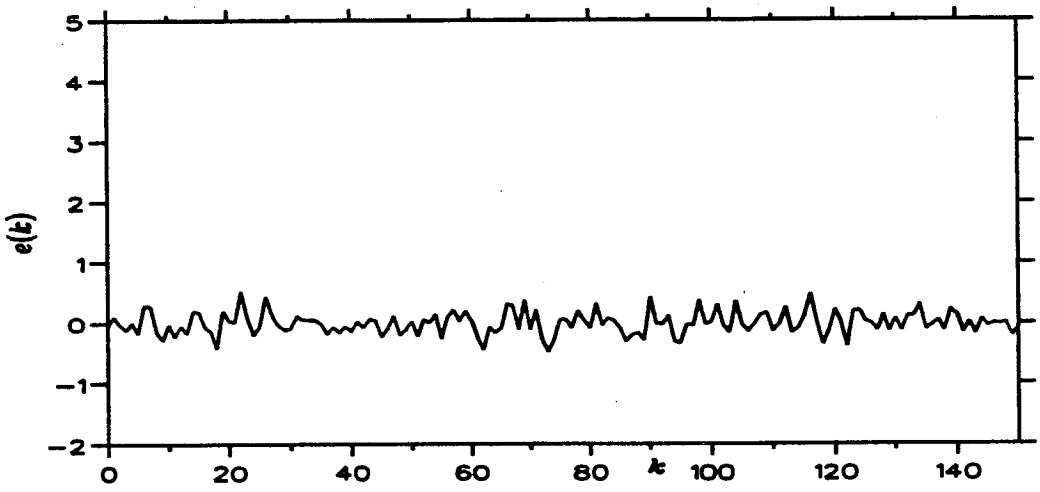
Fig(4.8a) Identification results in Example (4.3.c): 15 dB noise.



Fig(4.8b) Determined inverse filter in Example (4.3.c): proposed algorithm, 15 dB noise.



Fig(4.8c) Deconvolved series $\hat{w}(k)$ in Example (4.3.c): proposed algorithm, 15 dB noise.



Fig(4.8d) Deconvolution error e(k) in Example (4.3.c): proposed algorithm, 15 dB noise.

$$\epsilon^2 = \| g_i - f(\psi, i) \| , \tag{4.4.1}$$

where $\| \cdot \|$ represents the L_p-norm. The most commonly used norms include

$$\epsilon^2 = \sum_{i=0}^{N} | g_i - f(\psi, i) | , \tag{4.4.2a}$$

$$\epsilon^2 = \sum_{i=0}^{N} [g_i - f(\psi, i)]^2 , \tag{4.4.2b}$$

and

$$\epsilon^2 = \max_{0 \le i \le N} | g_i - f(\psi, i) | . \tag{4.4.2c}$$

ψ can then be determined by minimising ϵ^2. Notice that, no matter which type of norm is adopted, the cost function in Eq(4.4.1) actually always adopts a concept of identity: the identity between the curve represented by $\{g_i\}$ (g –curve) and the curve represented by $\{f(\psi, i)\}$ (f –curve). However, what is essentially necessary in blind identification is not the identity but the similarity: the similarity between the g –curve and the f –curve. This is because the difference of a constant amplitude factor does not affect the system characteristics. Fig(4.9) depicts the basic concept of similarity between g –curve and f –curve. The sufficient and necessary condition for this similarity can be expressed as

$$g_i = Q \cdot f(\psi, i) , \quad i = 0, 1, \ldots, N , \tag{4.4.3}$$

where Q is a constant. Since Eq(4.4.3) can be rewritten as

$$f(\psi, i+1) = \frac{1}{Q} g_{i+1} , \quad i = 0, 1, \ldots, N-1, \tag{4.4.4}$$

the following expression holds:

$$g_i \cdot f(\psi, i+1) = g_{i+1} \cdot f(\psi, i) , \quad i = 0, \ldots, N-1. \tag{4.4.5}$$

Obviously, Eq(4.4.5) is completely equivalent to Eq(4.4.3). In fact, Eq(4.4.5) can also be obtained directly from geometry. As can be seen from Fig(4.10), Eq(4.4.5) is exactly the condition under which the two polygons, which respectively take $\{g_i\}$ and $\{f(\psi, i)\}$ as the sides, are geometrically similar to each other. Now, corresponding to Eq(4.4.1), the following cost function can then be naturally constructed:

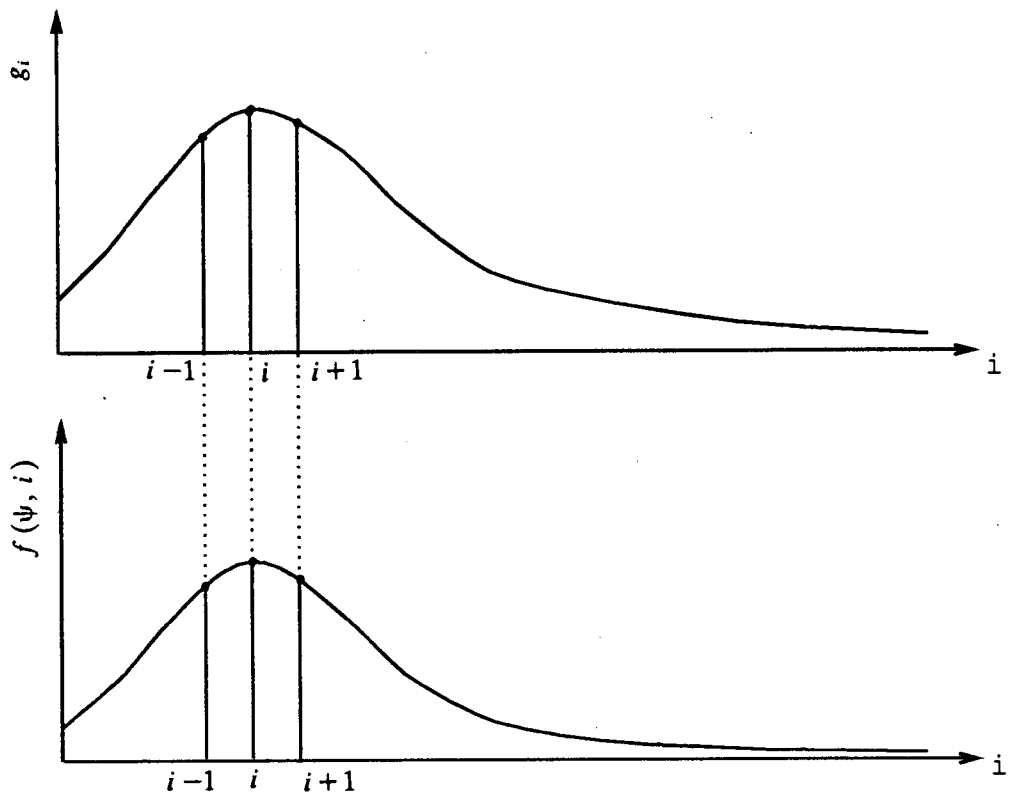
$$\epsilon^2 = \| g_i f(\psi, i+1) - g_{i+1} f(\psi, i) \| . \tag{4.4.6}$$

Specifically, in the case of L_2-norm, Eq(4.4.6) becomes

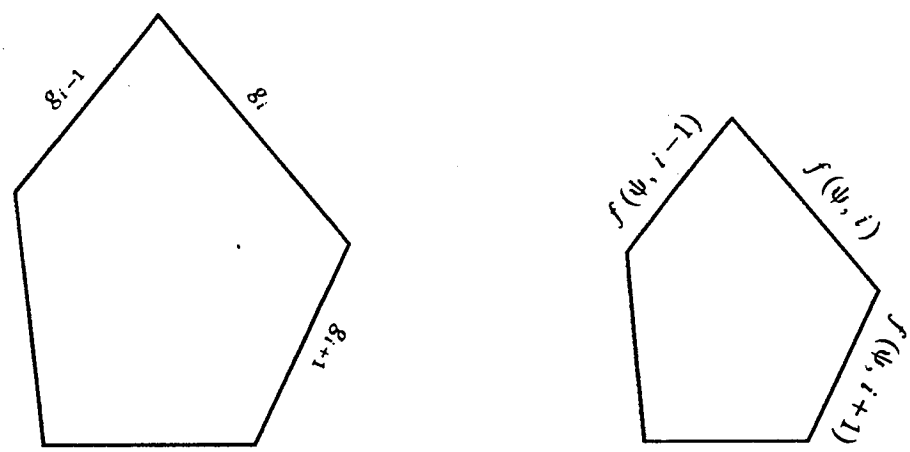
$$\epsilon^2 = \sum_{i=0}^{N-1} [g_i f(\psi, i+1) - g_{i+1} f(\psi, i)]^2 . \tag{4.4.7}$$

4.4.2 Application of similarity criterion to blind identification

As discussed at the beginning of this section, $\beta = \gamma^3 b_0$ in the second scheme (or b_0 in the first scheme) is a parameter which has to be estimated along with b_i, $i \neq 0$, if the Eq(4.2.11) based algorithm is employed. On the other hand, however, β (or b_0), is only an amplitude factor, and has no essential effect to the blind identification result, *i.e.*, the "shape" of the system impulse response. In this situation, the concept of similarity introduced earlier becomes a useful tool.



Fig(4.9) Similarity between two curves.



Fig(4.10) Similarity between two ploygons.

Let

$$f(\psi, m)=\beta \sum_{i=max(0,\ -m)}^{min(q,\ q-m)} b_i b_{i+m}^2 \ , \tag{4.4.8}$$

where $\psi=\{b_i \mid i=1, \ldots, q\}$, and

$$g_m=\hat{R}(m, m) \ , \tag{4.4.9}$$

a new similarity based cost function can thus be constructed from Eq(4.4.7) as

$$\epsilon^2=\sum_{m=-q}^{q-1}[\hat{R}(m, m)\sum_{i=max(0,\ -m-1)}^{min(q,\ q-m-1)} b_i b_{i+m+1}^2-\hat{R}(m+1, m+1)\sum_{i=max(0,\ -m)}^{min(q,\ q-m)} b_i b_{i+m}^2]^2 \ . \tag{4.4.10}$$

By minimising ϵ^2 in this equation, b_i, $i\neq 0$, can be estimated. Regarding the optimisation methods, both steepest descent method and Newton-Raphson method can be adopted. The gradient of ϵ^2 can be derived directly from Eq(4.4.10):

$$\nabla_l=\frac{\partial \epsilon^2}{\partial b_l}$$

$$=\sum_{m=-q}^{q-1}[\hat{R}(m, m)\sum_{i=max(0,\ -m-1)}^{min(q,\ q-m-1)} b_i b_{i+m+1}^2-\hat{R}(m+1, m+1)\sum_{i=max(0,\ -m)}^{min(q,\ q-m)} b_i b_{i+m}^2]\cdot$$

$$[\hat{R}(m, m)(b_{l+m+1}^2+2b_l b_{l-m-1})-\hat{R}(m+1, m+1)(b_{l+m}^2+2b_l b_{l-m})] \tag{4.4.11}$$

$$l=1, \ldots, q \ ,$$

where the calculations include all non-zero terms.

In order to achieve the global optimal solution, the idea of our two-step 'relay' approach can be borrowed here. From Eq(4.2.14), the initial estimate for the iteration process can be chosen as:

$$b_l=\frac{\hat{R}(q, l)}{\hat{R}(q, 0)} \ , \ l=1, \cdots, q \ . \tag{4.4.12}$$

4.4.3 Simulation results

In this subsection, the same system model as before is used to indicate the feasibility of the above similarity based approach. Namely, the model is

$$x(k)=w(k)-2.3w(k-1)+0.6w(k-2)+n(k) \ . \tag{4.4.13}$$

All other simulation conditions are also the same as in last section except the SNR=40 dB here. The identification result from the similarity based algorithm is listed in Table (4.4). For comparison, the results from the identity based algorithm and Giannakis' method are also given respectively in Table (4.4) (normalisation has been done here with respect to b_0). Clearly, the result of the similarity based identification algorithm proposed in this section is very encouraging. The signal restoration was also similarly realised , and the corresponding results are very satisfactory, but not shown here.

Table (4.4) Identification Results in Section 4.4.			
	b_0	b_1	b_2
true	1.00000	−2.30000	0.60000
similarity	1.00000	−2.36966	0.63083
identity	1.00000	−2.39198	0.63096
Giannakis	1.00000	−2.52928	0.62887

A similarity based identification approach, a variety of our basic two-step relay algorithm, was suggested in this section. This technique causes one less parameter which needs to be identified. The simulation confirmed the feasibility of this new approach.

4.5 Conclusions

A higher-order cumulant based two-step relay technique has been presented in this chapter. The technique has the following features. Firstly, the algorithm employs the result of Giannakis' method as the initial estimate of an optimisation procedure. As a result, the multimodality of the optimisation procedure are, to a great extent, overcome.

Secondly, the algorithm itself is not only parametric but also parametrically optimal, and it greatly improves the results of Giannakis' method in terms of accuracy. In fact, in some cases, the restoration process is unimplementable with only Giannakis' results (as has been seen in Section 4.3). In these cases, the application of our algorithm not only enables the restoration to become implementable but also generate very precise results. This, in some sense, is equivalent to the relaxation of the requirement for cumulant estimation.

Finally, the algorithm possesses a strong robustness to the additive noise. This is directly attributable to the exploitation of higher-order cumulants.

It has recently been brought to our attention that a very similar identification algorithm has been studied by Mehlan and Nandi [115] with similar results.

In addition, a similarity based variety of the algorithm has been introduced in the last section and it reduces by one parameter needing to be estimated.

It should also be pointed out that the main problem that remains is that the global solution of optimization can not be achieved when the result of Giannakis' method is poor or when the selected model order mismatches especially in the case of true MA systems. To solve this problem, the following three directions for the future work can be tried: *1)* improve the cumulant estimation accuracy as much as possible; *2)* develop more reliable order selection algorithm; and *3)* search for a globally convergent optimisation procedure. However, this problem cannot declare the failure of our algorithm. On the contrary, it is our algorithm that significantly overcome the multimodal difficulty in the optimisation process. This point will be further recognised in Chapter 6.

Chapter 5:

HIGHER-ORDER CUMULANT BASED BLIND DECONVOLUTION: AR MODEL

5.1 Introduction

To date, two types of parametric identification techniques for nonminimum phase systems have been suggested: MA model based methods and noncausal AR model based methods. As has been discussed in last chapter, one of the advantages of MA based approaches is that the MA coefficients can be determined directly from closed-form formula in most cases [15][28][30]. Unfortunately, the practical results of such closed-form formula are normally too inaccurate for the following two reasons: Firstly, these formula are over-sensitive to the estimation error of the cumulants, and secondly, estimates of higher-order cumulants generally suffer from higher variance than that of autocorrelation functions [65]. An alternative to this is to reduce the estimation of MA parameters to a nonlinear minimisation problem as in [15][12][88][92], however, this is a non-trivial problem. There are several existing schemes, such as good initial estimate finding methods [12][88] and overparametrisation methods [15][92], but they do not always work well. Simulated annealing algorithms [93][94] can achieve the global minimum, and have recently interested many researchers in the optimisation community. However, they do not show much potential in the deconvolution and identification domain, as a result of their slow rate of convergence.

In this chapter, an alternative set of parametric deconvolution and identification algorithms, based on non-causal AR system model, is presented. All these algorithms are derived directly from three theorems relating the inverse coefficients to the diagonal slices of cumulants of different orders. The algorithms requirements for the input are only that it be an independent and identically distributed (IID) non-Gaussian random series. The system can be nonminimum or minimum phase, but must have no zeros on the unit circle; 3rd- and 4th-order cumulants are employed respectively in our algorithms. Additionally, although our algorithms are in the main oriented to FIR (MA) systems, IIR (ARMA) systems can also be considered.

A major advantage for the noncausal AR model based algorithms (the type considered here) [13][32][14], is that the problem is no longer multimodal since a set of overdetermined linear equations results. In addition, since the identification results of this type of algorithms are simply the coefficients of the corresponding inverse filters, it is very convenient to implement deconvolution. In [32] and [14], both diagonal and nondiagonal

cumulant slices are used. The latter however has two main drawbacks: firstly, the estimation of nondiagonal cumulant slices is generally more noisy than that of diagonal slices [65], thus using nondiagonal slices may cause degradation to the performance of the algorithms, and secondly, the forms of the resulting matrices are very complicated, especially when 4th-order cumulants are used. In order to overcome these two points, in our algorithms, only diagonal cumulant slices and autocorrelation functions (2nd-order cumulants) are employed. As a result, both the algorithms involved with the 4th-order cumulants and their counterparts using the 3rd-order cumulants take an identical simple form. Thus, deconvolution of an unskewed series for the case of NMP systems, which has not received as much attention as the case of skewed series, is realised in this chapter as easily as that of a skewed series.

The general structure of this chapter is as follows. In Section 5.2, we first present three theorems regarding the inverse filter coefficients, and then develop the new deconvolution algorithms, each of them comprising a set of linear equations with respect to the inverse filter coefficients. The cases of 2nd-, 3rd- and 4th-order cumulants are studied respectively. Some remarks on the proposed technique are shown in Section 5.3. Then in Section 5.4, some computer simulation examples are presented. Finally, some conclusions are drawn in Section 5. In addition, the mathematical details of derivation are contained within Appendix 5.1.

5.2 Algorithm Kernel: Several Families of Linear Equations

The formation of blind deconvolution problem has been addressed in the preceding chapters. But for the convenience of description and reference in this chapter, we redefine it as follows. Let $x(k)$ be an ergodic series, which is the output of a stable linear time-invariant (LTI) mixed phase MA system $S=\{b_i \mid i=0, 1, \ldots q\}$. Consequently, we must have

$$x(k) = \sum_{i=0}^{q} b_i w(k-i) \quad , \tag{5.2.1}$$

where q denotes the order of system S, and $w(k)$ is the independent and identically distributed (IID) driving series. Then, the following procedure is termed *blind deconvolution*: Find an inverse filter $S^{-1}=\{\theta_i \mid i=r_1, \ldots, r_2\}$ which makes

$$w(k) \approx \sum_{i=r_1}^{r_2} \theta_i x(k-i) \quad , \tag{5.2.2}$$

given the output series $x(k)$ only, where r_1 and r_2 is the properly selected order of the non-causal part and causal part, respectively. But we will write "$\approx$" as "$=$" in the following derivation since the distinction is clear.

Let us notice in Eq(5.2.2) that $r_1=0$ when S is minimum phase and $r_2=-1$ when S maximum phase. Thus minimum phase and maximum phase are two special cases of the above model. It is also assumed that system S has no zeros on the unit circle. In addition, it is well known that $x(k)$ is not deconvolvable if S is nonminimum phase and $w(k)$ is a Gaussian IID series [19][13]. Hence, we always assume in this chapter that $w(k)$ is a non-Gaussian IID series with

$$\mathrm{E}[w(k)] = \gamma_1 = 0 \ , \tag{5.2.3a}$$

$$\mathrm{E}[w^2(k)] = \gamma_2 \ , \tag{5.2.3b}$$

$$\mathrm{E}[w^3(k)] = \gamma_3 \ , \tag{5.2.3c}$$

and

$$\mathrm{E}[w^4(k)] - 3\mathrm{E}^2[w^2(k)] = \gamma_4 \ , \tag{5.2.3d}$$

where $\gamma_4 \neq 0$ as a result of the assumption that $w(k)$ is non-Gaussian. In Eq(5.2.3c), γ_3 can be zero for non-skewed input signal, and in this case we may use the fourth-order cumulants.

It should be noted that the above assumption that the time series is non-Gaussian is often realistic in practical situations, as indicated in [72] and [96].

For the output series $x(k)$ assumed above, its nth-order cumulants can be denoted as $C_n(m_1, m_2, \ldots, m_{n-1})$. From the definition of cumulants (see [13] or Chapter 2), we have

$$C_2(m_1) = \mathrm{E}[x(k)x(k+m_1)] \ , \tag{5.2.4a}$$

$$C_3(m_1, m_2) = \mathrm{E}[x(k)\prod_{i=1}^{2} x(k+m_i)] \ , \tag{5.2.4b}$$

$$C_4(m_1, m_2, m_3) = \mathrm{E}[x(k)\prod_{i=1}^{3} x(k+m_i)]$$

$$-C_2(m_1)C_2(m_3-m_2) - C_2(m_2)C_2(m_3-m_1) - C_2(m_3)C_2(m_2-m_1) \ . \tag{5.2.4c}$$

Specially, we denote the diagonal slice of $C(m_1, m_2, \ldots, m_{n-1})$ as

$$c_n(m) = C_n(m_1, m_2, \ldots, m_{n-1})\big|_{m_1=m_2=\ldots=m_{n-1}=m} \ . \tag{5.2.5}$$

The algorithms proposed are based in the main on the following theorems.

Theorem 1: If system S is LTI, causal, and stable, and has no zeros on the unit circle, and its input $w(k)$ is zero-mean and IID, then the relation between S and its inverse filter S^{-1} can, in terms of 2nd-order cumulants, be expressed as

$$\frac{1}{\gamma_2}\cdot\sum_{j=r_1}^{r_2}\theta_j c_2(i+j) = \begin{cases} b_i & \text{if } i \in [0, q] \\ 0 & \text{other} \end{cases} \ . \tag{5.2.6}$$

Proof: See Appendix 5.1.

Theorem 2: If system S is LTI, causal, and stable, and has no zeros on the unit circle, and its input $w(k)$ is zero-mean, skewed and IID, then the relation between S and its inverse filter S^{-1} can, in terms of 3rd-order cumulants, be expressed as

$$\frac{1}{\gamma_3}\cdot\sum_{j=r_1}^{r_2}\theta_j c_3(i+j) = \begin{cases} b_i^2 & \text{if } i \in [0, q] \\ 0 & \text{other} \end{cases} \ . \tag{5.2.7}$$

Proof: See Appendix 5.1.

Notice that, only a limited number of equations are nontrivial in Eq(5.2.6) and Eq(5.2.7). Specifically, for a MA(q) system,

$$c_n(m) = 0 \quad \text{for } m \in (-\infty, -q-1] \bigcup [q+1, +\infty) \ , \tag{5.2.8}$$

where $n=2, 3, 4$, thus it can be seen that the equations in Eq(5.2.6) and Eq(5.2.7) are nontrivial only when $i+r_1 \le q$ and $i+r_2 \ge -q$, *viz.*, $-q-r_2 \le i \le q-r_1$ [82][12][13]. Thus, there are r_2-r_1+2q+1 nontrivial equations in Eq(5.2.6) and Eq(5.2.7), respectively. Since there are only r_2-r_1+1 coefficients θ_i's to be determined, many sets of equations can be constructed with respect to the θ_i's. As a result of the b_i's being unknown now, we can adopt either of the following two schemes to eliminate them.

<1> *Only adopting the zero equations--i.e, the ones with right hand sides (r.h.s.) equal to zero in Eq(5.2.6) and Eq(5.2.7).*

The number of such equations $2(r_2-r_1+q)$ is still much greater than the number of unknowns r_2-r_1+1. Hence, we need only take the following $2(r_2-r_1-q)$ equations in order to reduce unnecessary computation:

$$\begin{cases} \sum_{j=r_1}^{r_2} \theta_j c_2(i+j) = 0 \\ \sum_{j=r_1}^{r_2} \theta_j c_3(i+j) = 0 \end{cases} \qquad i = -r_2, \ldots, -1, q+1, \ldots, -r_1 \ . \tag{5.2.9}$$

Without loss of generality, we can always let $\theta_0=1$, *viz.*, the inverse filter coefficients have been normalised by θ_0 here. Then, Eq(5.2.9) can be written as

$$\mathbf{A}\Theta = \mathbf{B} \ , \tag{5.2.10}$$

where

$$\mathbf{A} = \begin{bmatrix} c_2(-r_2+r_1) & \cdots & c_2(-r_2-1) & c_2(-r_2+1) & \cdots & c_2(0) \\ c_3(-r_2+r_1) & \cdots & c_3(-r_2-1) & c_3(-r_2+1) & \cdots & c_3(0) \\ \cdots & & \cdots & \cdots & & \cdots \\ c_2(-1+r_1) & \cdot & c_2(-2) & c_2(0) & \cdot & c_2(-1+r_2) \\ c_3(-1+r_1) & \cdot & c_3(-2) & c_3(0) & \cdot & c_3(-1+r_2) \\ c_2(q+1+r_1) & \cdot & c_2(q) & c_2(q+2) & \cdot & c_2(q+1+r_2) \\ c_3(q+1+r_1) & \cdot & c_3(q) & c_3(q+2) & \cdot & c_3(q+1+r_2) \\ \cdots & & \cdots & \cdots & & \cdots \\ c_2(0) & \cdots & c_2(-r_1-1) & c_2(-r_1+1) & \cdots & c_2(r_2-r_1) \\ c_3(0) & & c_3(-r_1-1) & c_3(-r_1+1) & & c_3(r_2-r_1) \end{bmatrix} , \tag{5.2.11a}$$

$$\Theta = [\theta_{r_1}, \theta_{r_1+1}, \ldots, \theta_{-1}, \theta_1, \ldots, \theta_{r_2}]^T \ , \tag{5.2.11b}$$

and

$$\mathbf{B} = -[c_2(-r_2), c_3(-r_2), \ldots, c_2(-1), c_3(-1), c_2(q+1), c_3(q+1), \ldots, c_2(-r_1), c_3(-r_1)]^T \ . \tag{5.2.11c}$$

Notice that, Eq(5.2.10) is still overdetermined since r_1 and r_2 can always be taken to be sufficiently large to make $2(r_2-r_1-q) \ge r_2-r_1$, *viz*, $r_2-r_1 \ge 2q$. Thus, taking the least-squares solution of Eq(5.2.10), the inverse filter coefficients can be expressed as

$$\Theta = (\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{B}) \ . \tag{5.2.12}$$

Since $\mathbf{A}^T\mathbf{A}$ is positive definite, we must have $det(\mathbf{A}^T\mathbf{A})>0$ [99]. Hence, Θ can be uniquely determined from Eq(5.2.12).

<2> Eliminating b_i's in Eq(5.2.6) and Eq(5.2.7).

In this scheme, both zero and non-zero equations in Eq(5.2.6) and Eq(5.2.7) are adopted, but b_i's are eliminated by using the formula proposed by Giannakis and Mendel in [15].

In [15], Giannakis and Mendel derived the following equations (*i.e.*, Eq(24) in [15], but it is slightly modified here):

$$\gamma_3\sum_{i=0}^{q}c_2(n-i)b_i^2 = \gamma_2\sum_{i=0}^{q}c_3(n-i)b_i \;;\quad n=-q,\ldots,2q. \tag{5.2.13}$$

Applying the results of Theorem 1 and Theorem 2 to Eq(5.2.13), we can obtain

$$\sum_{j=r_1}^{r_2}\theta_j\sum_{i=0}^{q}c_2(n-i)c_3(i+j) = \sum_{j=r_1}^{r_2}\theta_j\sum_{i=0}^{q}c_3(n-i)c_2(i+j) \quad n\in[-q,\,2q]\;. \tag{5.2.14}$$

As before, θ_0 can still be assumed to be 1. Then, Eq(5.2.14) can be rewritten as

$$\sum_{\substack{j=r_1\\ j\neq 0}}^{r_2} f_{nj}\theta_j = g_n \;;\quad n\in[-q,\,2q]\;, \tag{5.2.15}$$

where

$$f_{nj} = \sum_{i=0}^{q}[c_2(n-i)c_3(i+j)-c_3(n-i)c_2(i+j)]$$

and

$$g_n = \sum_{i=0}^{q}[c_3(n-i)c_2(i)-c_2(n-i)c_3(i)]\;.$$

Clearly, many sets of equations with respect to θ_j's can be formed from Eq(5.2.15) and the zero equations in Eq(5.2.6) and Eq(5.2.7), but following two are the simplest.

Firstly, combining Eq(5.2.6) with Eq(5.2.15), we have

$$\begin{cases} \displaystyle\sum_{\substack{j=r_1\\ j\neq 0}}^{r_2}\theta_j c_2(n+j) = -c_2(n)\;; \quad \text{for } n\in[-r_2,\,-1]\cup[q+1,\,-r_1] \\ \displaystyle\sum_{\substack{j=r_1\\ j\neq 0}}^{r_2}\theta_j f_{nj} = g_n\;; \quad \text{for } n\in[0,\,q]\;, \end{cases} \tag{5.2.16}$$

or more compactly,

$$\mathbf{P}\Theta = \mathbf{Q}\;, \tag{5.2.17}$$

where

$$\mathbf{P}=\begin{bmatrix} c_2(-r_2+r_1) & \cdots & c_2(-r_2-1) & c_2(-r_2+1) & \cdots & c_2(0) \\ \vdots & & \vdots & \vdots & & \vdots \\ c_2(-1+r_1) & \cdots & c_2(-2) & c_2(0) & \cdots & c_2(-1+r_2) \\ f_{0,r_1} & \cdots & f_{0,-1} & f_{0,1} & \cdots & f_{0,r_2} \\ \vdots & & \vdots & \vdots & & \vdots \\ f_{q,r_1} & \cdots & f_{q,-1} & f_{q,1} & \cdots & f_{q,r_2} \\ c_2(q+1+r_1) & \cdots & c_2(q) & c_2(q+2) & \cdots & c_2(q+1+r_2) \\ \vdots & & \vdots & \vdots & & \vdots \\ c_2(0) & \cdots & c_2(-r_1-1) & c_2(-r_1+1) & \cdots & c_2(r_2-r_1) \end{bmatrix} \tag{5.2.18a}$$

and

$$\mathbf{Q} = [-c_2(-r_2), \ldots, -c_2(-1), g_0, \ldots, g_q, -c_2(q+1), \ldots, -c_2(-r_1)]^T . \tag{5.2.18b}$$

Then we can obtain

$$\Theta = (\mathbf{P}^T\mathbf{P})^{-1}(\mathbf{P}^T\mathbf{Q}) \ . \tag{5.2.19}$$

Secondly, combining Eq(5.2.7) and Eq(5.2.15), we can obtain following equation in a similar way:

$$\mathbf{U}\Theta = \mathbf{V} \ , \tag{5.2.20}$$

where

$$\mathbf{U}=\begin{bmatrix} c_3(-r_2+r_1) & \cdots & c_3(-r_2-1) & c_3(-r_2+1) & \cdots & c_3(0) \\ \vdots & & \vdots & \vdots & & \vdots \\ c_3(-1+r_1) & \cdots & c_3(-2) & c_3(0) & \cdots & c_3(-1+r_2) \\ f_{0,r_1} & \cdots & f_{0,-1} & f_{0,1} & \cdots & f_{0,r_2} \\ \vdots & & \vdots & \vdots & & \vdots \\ f_{q,r_1} & \cdots & f_{q,-1} & f_{q,1} & \cdots & f_{q,r_2} \\ c_3(q+1+r_1) & \cdots & c_3(q) & c_3(q+2) & \cdots & c_3(q+1+r_2) \\ \vdots & & \vdots & \vdots & & \vdots \\ c_3(0) & \cdots & c_3(-r_1-1) & c_3(-r_1+1) & \cdots & c_3(r_2-r_1) \end{bmatrix} \tag{5.2.21a}$$

and

$$\mathbf{V} = [-c_3(-r_2), \ldots, -c_3(-1), g_0, \ldots, g_q, -c_3(q+1), \ldots, -c_3(-r_1)]^T . \tag{5.2.21b}$$

Then, we must have

$$\Theta = (\mathbf{U}^T\mathbf{U})^{-1}(\mathbf{U}^T\mathbf{V}) \ . \tag{5.2.22}$$

If the input series $w(k)$ is unskewed, *viz.*, $\gamma_3=0$, the 4th-order cumulants must be used in order to deconvolve and identify the nonminimum phase systems. In this case, we employ the following theorem:

Theorem 3: If system S is LTI, causal, and stable, and has no zeros on the unit circle, and its input $w(k)$ is zero-mean, non-Gaussian and IID, then the relation between S and its inverse filter S^{-1} can, in terms of 4th-order cumulants, be expressed as

$$\frac{1}{\gamma_4}\cdot\sum_{j=r_1}^{r_2}\theta_j c_4(i+j) = \begin{cases} b_i^3 & \text{if } i\in[0,\ q] \\ 0 & \text{other} \end{cases} \quad \blacksquare \tag{5.2.23}$$

Proof: See Appendix 5.1.

From this theorem and foregoing discussion, the following descriptions are straightforward.

Firstly, combining Theorem 3 with Theorem 1, and along the line of scheme 1 presented above, the linear equation systems with the same form as Eq(5.2.10) can be obtained:

$$\mathbf{A}\Theta = \mathbf{B} \ , \tag{5.2.24}$$

where, however,

$$\mathbf{A}= \begin{bmatrix} c_2(-r_2+r_1) & \cdots & c_2(-r_2-1) & c_2(-r_2+1) & \cdots & c_2(0) \\ c_4(-r_2+r_1) & \cdots & c_4(-r_2-1) & c_4(-r_2+1) & \cdots & c_4(0) \\ \cdots & & \cdots & \cdots & & \cdots \\ c_2(-1+r_1) & . & c_2(-2) & c_2(0) & . & c_2(-1+r_2) \\ c_4(-1+r_1) & . & c_4(-2) & c_4(0) & . & c_4(-1+r_2) \\ c_2(q+1+r_1) & \cdot & c_2(q) & c_2(q+2) & \cdot & c_2(q+1+r_2) \\ c_4(q+1+r_1) & \cdot & c_4(q) & c_4(q+2) & \cdot & c_4(q+1+r_2) \\ \cdots & & \cdots & \cdots & & \cdots \\ c_2(0) & \cdots & c_2(-r_1-1) & c_2(-r_1+1) & \cdots & c_2(r_2-r_1) \\ c_4(0) & & c_4(-r_1-1) & c_4(-r_1+1) & & c_4(r_2-r_1) \end{bmatrix} \tag{5.2.25a}$$

and

$$\mathbf{B} = -[c_2(-r_2),\ c_4(-r_2),\ \ldots,\ c_2(-1),\ c_4(-1),\ c_2(q+1),\ c_4(q+1),\ \ldots,\ c_2(-r_1),\ c_4(-r_1)]^T \ . \tag{5.2.25b}$$

Secondly, considering Theorem 3 and the Giannakis-Mendel formula (Eq(29a) in [15]) for the case of 4th-order cumulants (with a slight modification here):

$$\gamma_4 \sum_{i=0}^{q} c_2(n-i) b_i^3 = \gamma_2 \sum_{i=0}^{q} c_4(n-i) b_i \ ; \quad n = -q, \ldots, 2q \ , \tag{5.2.26}$$

the following two equations, with the same form as Eq(5.2.17) and Eq(5.2.20), respectively, can be derived. Corresponding to Eq(5.2.17), we have

$$\mathbf{P}\Theta = \mathbf{Q} \ , \tag{5.2.27}$$

where

$$\mathbf{P}= \begin{bmatrix} c_2(-r_2+r_1) & \cdots & c_2(-r_2-1) & c_2(-r_2+1) & \cdots & c_2(0) \\ \cdots & & \cdots & \cdots & & \cdots \\ c_2(-1+r_1) & \cdots & c_2(-2) & c_2(0) & \cdots & c_2(-1+r_2) \\ f_{0,r_1} & \cdots & f_{0,-1} & f_{0,1} & \cdots & f_{0,r_2} \\ \cdots & & \cdots & \cdots & & \cdots \\ f_{q,r_1} & \cdots & f_{q,-1} & f_{q,1} & \cdots & f_{q,r_2} \\ c_2(q+1+r_1) & \cdots & c_2(q) & c_2(q+2) & \cdots & c_2(q+1+r_2) \\ \cdots & & \cdots & \cdots & & \cdots \\ c_2(0) & \cdots & c_2(-r_1-1) & c_2(-r_1+1) & \cdots & c_2(r_2-r_1) \end{bmatrix} \tag{5.2.28a}$$

and

$$\mathbf{Q} = [-c_2(-r_2),\ \ldots,\ -c_2(-1),\ g_0,\ \ldots,\ g_q,\ -c_2(q+1),\ \ldots,\ -c_2(-r_1)]^T \ . \tag{5.2.28b}$$

Let us notice that, f_{nj} and g_n in Eq(5.2.28a) and Eq(5.2.28b) are given by

$$f_{nj} = \sum_{i=0}^{q} [c_2(n-i)c_4(i+j) - c_4(n-i)c_2(i+j)]$$

and

$$g_n = \sum_{i=0}^{q}[c_4(n-i)c_2(i)-c_2(n-i)c_4(i)] \ .$$

Then, corresponding to Eq(5.2.20), we can also obtain

$$\mathbf{U}\Theta = \mathbf{V} \ , \tag{5.2.29}$$

where

$$\mathbf{U}= \begin{bmatrix} c_4(-r_2+r_1) & \cdots & c_4(-r_2-1) & c_4(-r_2+1) & \cdots & c_4(0) \\ \vdots & & \vdots & \vdots & & \vdots \\ c_4(-1+r_1) & \cdots & c_4(-2) & c_4(0) & \cdots & c_4(-1+r_2) \\ f_{0,r_1} & \cdots & f_{0,-1} & f_{0,1} & \cdots & f_{0,r_2} \\ \vdots & & \vdots & \vdots & & \vdots \\ f_{q,r_1} & \cdots & f_{q,-1} & f_{q,1} & \cdots & f_{q,r_2} \\ c_4(q+1+r_1) & \cdots & c_4(q) & c_4(q+2) & \cdots & c_4(q+1+r_2) \\ \vdots & & \vdots & \vdots & & \vdots \\ c_4(0) & \cdots & c_4(-r_1-1) & c_4(-r_1+1) & \cdots & c_4(r_2-r_1) \end{bmatrix} \tag{5.2.30a}$$

and

$$\mathbf{V} = [-c_4(-r_2), \ldots, -c_4(-1), g_0, \ldots, g_q, -c_4(q+1), \ldots, -c_4(-r_1)]^T . \tag{5.2.30b}$$

Here, f_{nj} and g_n take the same values as in Eq(5.2.28a) and Eq(5.2.28b).

An interesting point to note is that the results of Theorem 1, 2, and 3 can actually be expressed as the following general form:

$$\frac{1}{\gamma_l}\cdot\sum_{j=r_1}^{r_2}\theta_j c_l(i+j) = \begin{cases} b_i^{l-1} & \text{if } i\in[0,\ q] \\ 0 & \text{other} \end{cases} \ . \tag{5.2.31}$$

Here we have assumed $\gamma_l \neq 0$; $l=2, 3, 4$. Notice that, the equations involved with the 4th-order cumulants are of exactly the same form with their counterparts using the 3rd-order cumulants. As a result, the 4th-order cumulant based algorithms can be implemented almost as easily as the 3rd-order cumulant based ones. Furthermore, it can be found that all of the families of linear equations we have obtained above are of a similar elegant form, which makes the programming of the algorithms very regular. In addition, it should be noticed that the MA identification can also be realised directly from Theorem 3. This becomes even better when the *SNR* is low.

5.3 Remarks on Technique

On the technique proposed above, the following remarks are in order.

1) Since all the above three families of equations (Eq(5.2.10), Eq(5.2.17), & Eq(5.2.10)) are linear, the inverse filter coefficients $\{\theta_j \,|\, j=r_1, \ldots, 0, \ldots, r_2\}$ can normally be determined uniquely. Notice that, $\mathbf{A}^T\mathbf{A}$, $\mathbf{P}^T\mathbf{P}$, and $\mathbf{U}^T\mathbf{U}$ are positive definite, and therefore nonsingular.

2) All the above algorithms are derived from the MA (or FIR) systems. In the case of ARMA (including AR, similarly hereinafter), there are in principle two schemes to apply above formulae:

i) Since any stable LTI system can be approximated by a truncated MA model, we can use the truncated impulse response h_i's of ARMA systems as the parameters of the approximating MA models, an idea used in [28]. The advantage of this method is that the order of the AR part of ARMA model is not needed. But on the other hand, this method requires more accurate cumulant estimates.

ii) As is well known, for an ARMA(p, q) process

$$x(k)+\sum_{i=1}^{p}a_i x(k-i) = \sum_{i=0}^{q}b_i w(k-i)\ ;\ p\geq q\ , \tag{5.3.1}$$

the AR coefficients a_i's can be determined from the equations below:

$$c_3(-n)+\sum_{i=1}^{p}a_i c_3(-n+i) = 0\ ;\ n>q\ , \tag{5.3.2}$$

where, the AR order p can be selected by using other algorithms (for example, Hankel matrix based methods), see [15], [95] or [81] for detail. Then, the residual process

$$y(k) = x(k)+\sum_{i=1}^{p}\hat{a}_i x(k-i)\ , \tag{5.3.3}$$

where $\hat{a}_i$ denotes the estimate of a_i, reduces to an MA process:

$$y(k) = \sum_{i=0}^{q}b_i w(k-i)\ , \tag{5.3.4}$$

to which, the MA based algorithms can be applied.

3) Although the algorithms proposed above are derived from the viewpoint of deconvolution, they may conveniently be employed for system identification. As soon as the θ_j's are determined, the MA parameters b_i's (for MA systems), or the truncated impulse response h_i's (for ARMA systems), can be directly calculated up to a gain constant according to Theorem 1 (Eq(5.2.6)). But when the level of additive noise is very high, the obtained b_i's or h_i's may be rather poor estimates since the estimation of the autocorrelation function can be degraded. In this case, we have to directly invert $S^{-1}=\{\theta_i\}$ to obtain the relatively accurate results.

Obviously, when our algorithms are used to identify the MA parameters of systems, the problem of multimodality no longer exists, and the uniqueness of identification is guaranteed. To this extent, this chapter provides a feasible solution to the nonlinearity problem which exists with the Giannakis-Mendel formula [15].

4) Since both the 3rd-order cumulants and autocorrelation functions are used, the above algorithms can deal with both non-minimum and minimum phase systems for deconvolution and system identification. The only constraint is that the system to be deconvolved may not have any zeros on the unit circle. This can be considered as the penalty paid for the linearity of the equations.

5) Selection of the inverse filter order r_1 and r_2: The algorithms proposed above are insensitive to the choice of r_1 and r_2, as long as they are taken to be large enough. However the above algorithms only involve the general case of mixed phase systems. For the special cases of minimum and maximum phase systems, several minor changes must be

made to ensure the above algorithms are numerically more reliable.

i) For minimum phase systems r_1 should be set to be zero. Although this is not necessary theoretically, in practice, $r_1<0$ does sometimes cause large errors, especially in the presence of a high level of additive noise.

ii) Similarly, for maximum phase systems, r_2 should be set to be -1. Then, θ_{-1}, instead of θ_0, should be assumed to be 1. The corresponding changes should also be implemented for Eq(5.2.10), Eq(5.2.17) and Eq(5.2.20).

6) Determination of MA order q: This problem depends on the phase property of the considered system.

i) For minimum phase systems, the above algorithms are robust to the choice of q value. The reason for this is that the matrix structures in Eq(5.2.10), Eq(5.2.17) and Eq(5.2.20) become irrelevant with q value when systems are minimum phase.

ii) For nonminimum phase systems, the above algorithms are generally sensitive to the selection of MA order q, since q value directly affects the contents of the matrices in Eq(5.2.10), Eq(5.2.17), and Eq(5.2.20). Thus, reliable order determination is required. In fact, several higher-order cumulant based methods for order selection have emerged in recent years [90][14][15][95][97][98]. But in this chapter, we do not discuss this problem, but assume that the system order q is known. We will also demonstrate these two points by using simulation results in Section 5.4.

7) A theoretical analysis of the performances of above three families of linear equations (Eq(5.2.10), Eq(5.2.17), and Eq(5.2.20)) is not available at this time. Heuristically, Eq(5.2.17) and Eq(5.2.20) are more stable than Eq(5.2.10), especially when method (*i*) mentioned previously in *REMARK* 2 is used for ARMA systems. However, since Eq(5.2.20) involves more 3rd-order cumulant estimation, which is robust in the presence of additive Gaussian noise, it is recommended that Eq(5.2.20) be used as long as a sufficiently long output series can be obtained to give an acceptable estimate of the 3rd-order cumulants; similarly, Eq(5.2.17) can also be tried if the *SNR* is not too low. In next section where we consider simulations, we will confirm these points empirically.

8) All the above algorithms can be easily made recursive. A recursive version of the technique will be derived and examined in Chapter 6.

5.4 Simulation Results

In this section, we apply the algorithms described above to MA (FIR) system deconvolution. Both nonminimum and minimum phase systems are considered. To study the case of 3rd-order cumulants, an exponentially distributed series is used as the skewed input series $w(k)$. For the case of 4th-order cumulants, a uniformly distributed continuous series and an equally distributed discrete series are used as the unskewed input series $w(k)$. Also, the output series $x(k)$ is generated by convolving $w(k)$ with the system impulse response. The additive noise $n(k)$ to $x(k)$ is assumed to be white Gaussian, and the signal-to-noise ratio (*SNR*) in this chapter is defined the same way as in Section 4.2 (*i.e.*, Eq(4.2.24)):

$$SNR = 10\cdot\log_{10}\left(\frac{\mathrm{E}[x^2(k)]}{\mathrm{E}[n^2(k)]}\right) . \qquad (5.4.1)$$

In the simulations below, the l-th order cumulants $c_l(m)$ in the foregoing equations is substituted with their corresponding estimates $\hat{c}_l(m)$. In order to obtain a more accurate cumulant estimation, the well known segment-average scheme [13][15] was adopted again. Corresponding to the problem in the current chapter, this scheme can be briefly described as follows.

First, segment N output samples into P records of Q samples each, *viz.*, $N=P\times Q$. Then, estimate the cumulants $\hat{c}_l^{(i)}(m)$ for the i-th record by

$$\hat{c}_l^{(i)}(m) = \frac{1}{Q}\sum_{k=k_l}^{k_u} x^{(i)}(k)[x^{(i)}(k+m)]^{l-1} \quad , \tag{5.4.2}$$

where $x^{(i)}(k)$ denotes the samples in the i-th record. In addition, here, $i=1, 2, , . . . , P$, $k_l=max(0, -m)$, and $k_u=min(Q-1, Q-m-1)$. Finally, average $\hat{c}_l^{(i)}(m)$ over all records:

$$\hat{c}_l(m) = \frac{1}{P}\sum_{i=1}^{P}\hat{c}_l^{(i)}(m) \quad . \tag{5.4.3}$$

In this section, we always take $N=P\times Q=30\times 300$. It should be noted that the above segmentation procedure is not essential for the ergodic series. A more detailed analysis regarding these points will be exhibited in next chapter (Section 6.3), where the recursive version of the algorithms is derived.

Also, in the following simulations, only MA systems with known order are considered. In the case of ARMA systems, we can first use the methods mentioned in *REMARK* 2 to transform ARMA problems to MA problems. Then, the application of the MA identification and deconvolution algorithms to the obtained MA models becomes straightforward.

5.4.1 Skewed $w(k)$: the case of the 3rd-order cumulants

Here, $w(k)$ is taken to be exponentially distributed ($\gamma_2=1$,$\gamma_3=2$).

<Example 5.4.1> Mixed phase MA system: The model is

$$x'(k) = w(k)-2.4w(k-1)+0.8w(k-2)+n(k) \quad , \tag{5.4.4}$$

with two zeros: $z_1=2$ and $z_2=0.4$.

Set $r_1=-15$ and $r_2=15$. For the noise levels which produce $SNR=50\ dB$ and $SNR=10\ dB$, Eq(5.2.10), Eq(5.2.17) and Eq(5.2.20) are solved, respectively. The obtained inverse filter coefficients θ_j's are shown in Fig(5.1a), Fig(5.1b) and Fig(5.1c), respectively. It can be seen that, for this MA system,

i) the results of three equations are of approximately identical accuracy at the lower noise level ($SNR=50\ dB$), and

ii) the result of Eq(5.2.17) is of a relatively higher error in comparison with Eq(5.2.10) and Eq(5.2.20) for the higher noise level ($SNR=10\ dB$). The second point is because Eq(5.2.17) depends upon the 2nd-order cumulants, which cannot completely remove the effect of the additive Gaussian noise, to a greater extent than Eq(5.2.10) and Eq(5.2.20), as mentioned in *REMARK* 7.

Also, the identification results (SNR=50 dB) from Theorem 1 are listed in the following Table (5.1), where, for the convenience of comparison, the gain adjustment has been done by dividing $\hat{b}_i$'s by $\hat{b}_0$ (similarly hereinafter). Clearly, these identification results are satisfactory for the case of higher *SNR*.

Table (5.1) Identification Results in Example 5.4.1.

	Eq(5.2.10)	Eq(5.2.17)	Eq(5.2.20)	True
$\hat{b}_0$	1.00000	1.00000	1.00000	1.00000
$\hat{b}_1$	−2.43943	−2.44185	−2.44120	−2.40000
$\hat{b}_2$	0.84840	0.84959	0.84905	0.80000

In addition, the sensitivity of the algorithms to the order mismatch is tested. Fig(5.1d) illustrates the results of Eq(5.2.10), Eq(5.2.17) and Eq(5.2.20) when the system order is set to be 3 and $SNR = 50\ dB$. It can be seen that the algorithms are sensitive to the selected system order in the case of NMP systems, as explained in *REMARK* 6.

<**Example 5.4.2**> Minimum phase MA system: The model is

$$x'(k) = w(k)+0.3w(k-1)-0.4w(k-2)+n(k)\ , \tag{5.4.5}$$

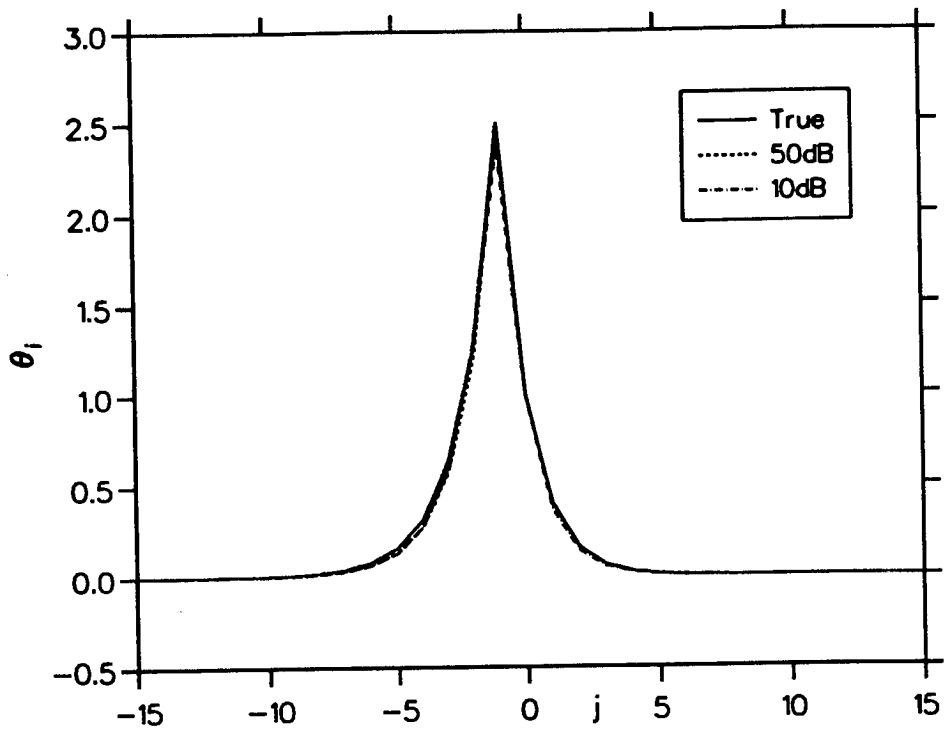
with two zeros: $z_1=-0.8$ and $z_2=0.5$.

First, set $r_1=0$ and $r_2=15$. Adding the noise which makes $SNR = 10\ dB$ and $SNR = 1\ dB$ this time, From Eq(5.2.10), Eq(5.2.17) and Eq(5.2.20), the corresponding inverse filters are determined, which are drawn in Fig(5.2a), Fig(5.2b), and Fig(5.2c), respectively. Again, it is clear that the result of Eq(5.2.20) is better than that of Eq(5.2.17) and Eq(5.2.10) when the additive noise is very strong. Similarly, the corresponding identification results from Theorem 1 are listed in the following Table (5.2) ($SNR = 10$ dB). They are comparable with the results listed in Table 1, where, however, the *SNR* is 50 dB. To this extent, the algorithms possess a stronger robustness to the additive noise for the MP systems than for the NMP ones.

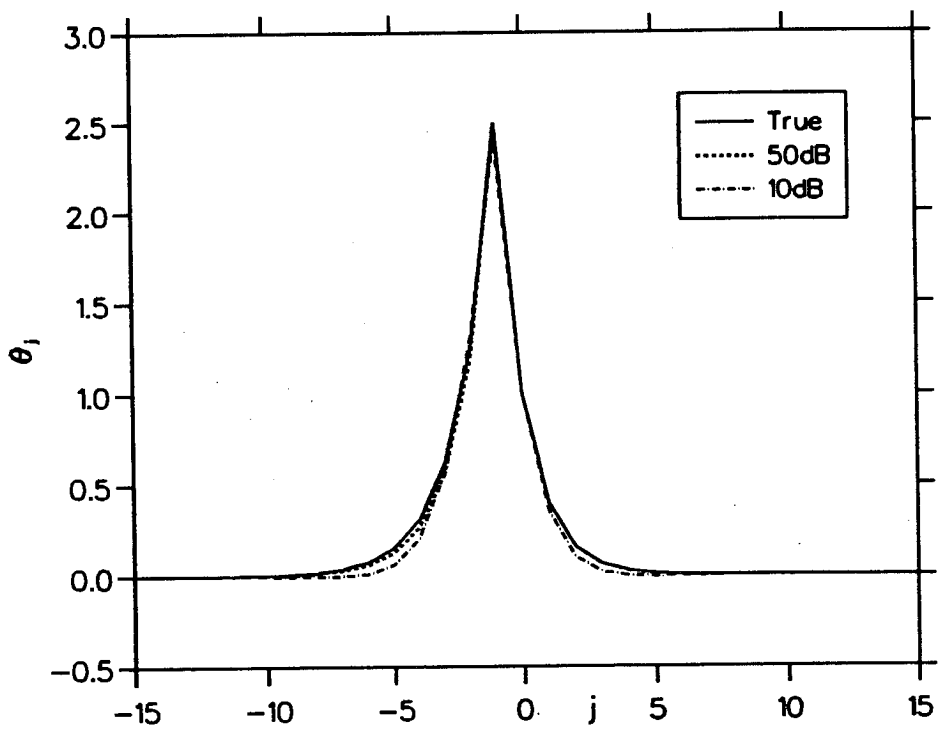
Table (5.2) Identification Results in Example 5.4.2.

	Eq(5.2.10)	Eq(5.2.17)	Eq(5.2.20)	True
$\hat{b}_0$	1.00000	1.00000	1.00000	1.00000
$\hat{b}_1$	0.27481	0.25478	0.28055	0.30000
$\hat{b}_2$	−0.35173	−0.34551	−0.35406	−0.40000

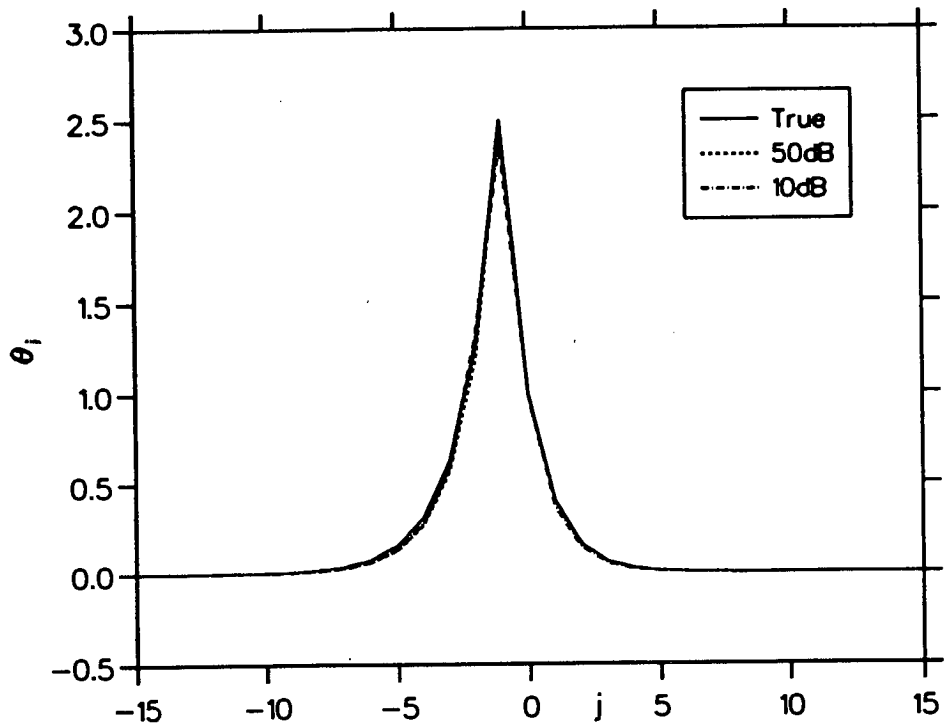
Then, set $r_1=-15$ and $r_2=15$. Fig(5.2d) demonstrates the obtained inverse filters ($SNR = 1\ dB$). As pointed out in *REMARK* 5, serious errors appear in the anti-causal parts of the inverse filters. But the result of Eq(5.2.20) is still more satisfactory than that of Eq(5.2.10) and Eq(5.2.17). In fact, during the simulation of other examples, not demonstrated here, we observed much more serious errors. These errors result from the



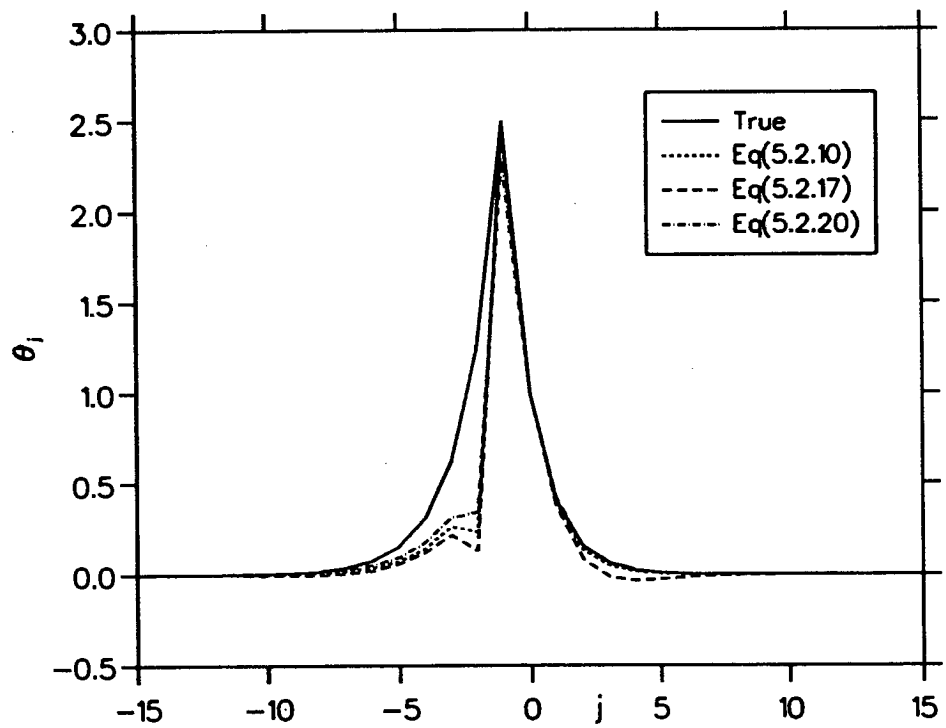
Fig(5.1a) Results of Eq(5.2.10) in Example 5.4.1: mixed phase MA(2) system.



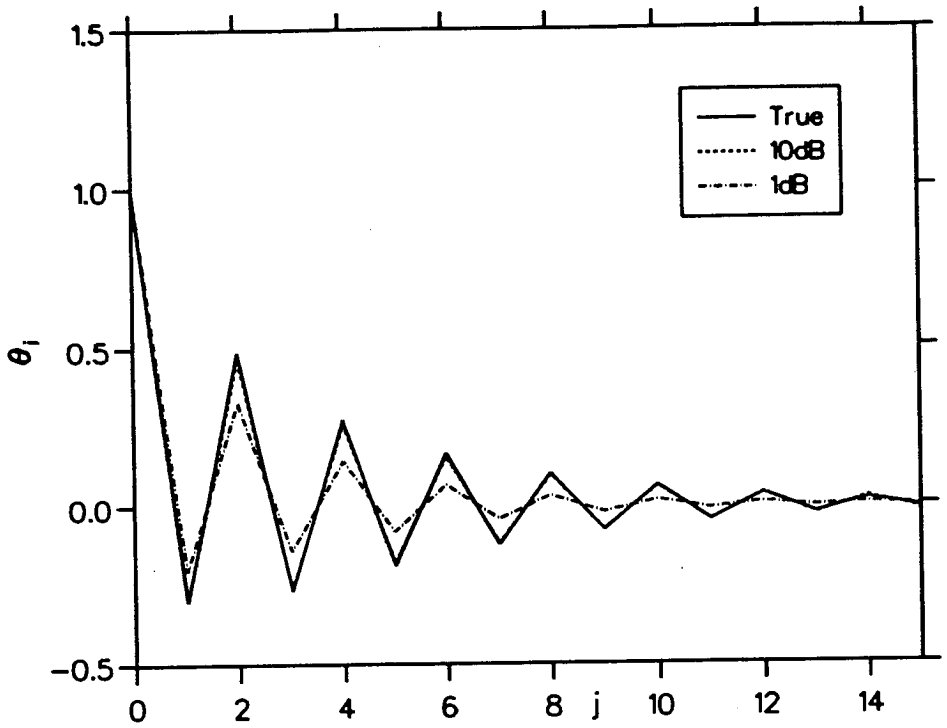
Fig(5.1b) Results of Eq(5.2.17) in Example 5.4.1: mixed phase MA(2) system.



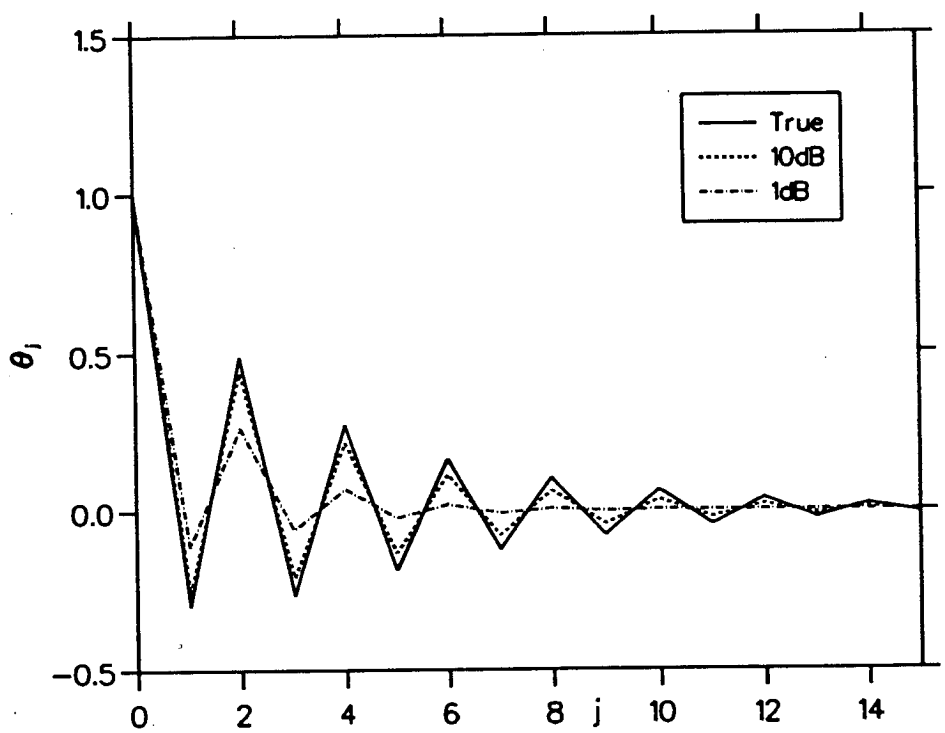
Fig(5.1c) Results of Eq(5.2.20) in Example 5.4.1: mixed phase MA(2) system.



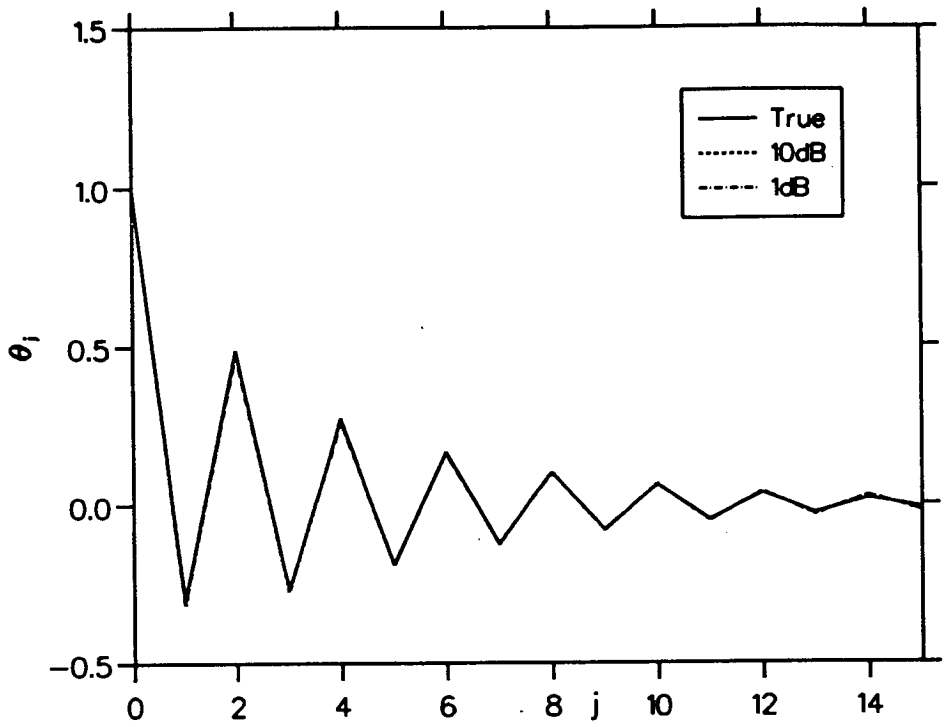
Fig(5.1d) Sensitivity of the algorithms to order mismatch: mixed phase MA(2) system, algorithm order=3.



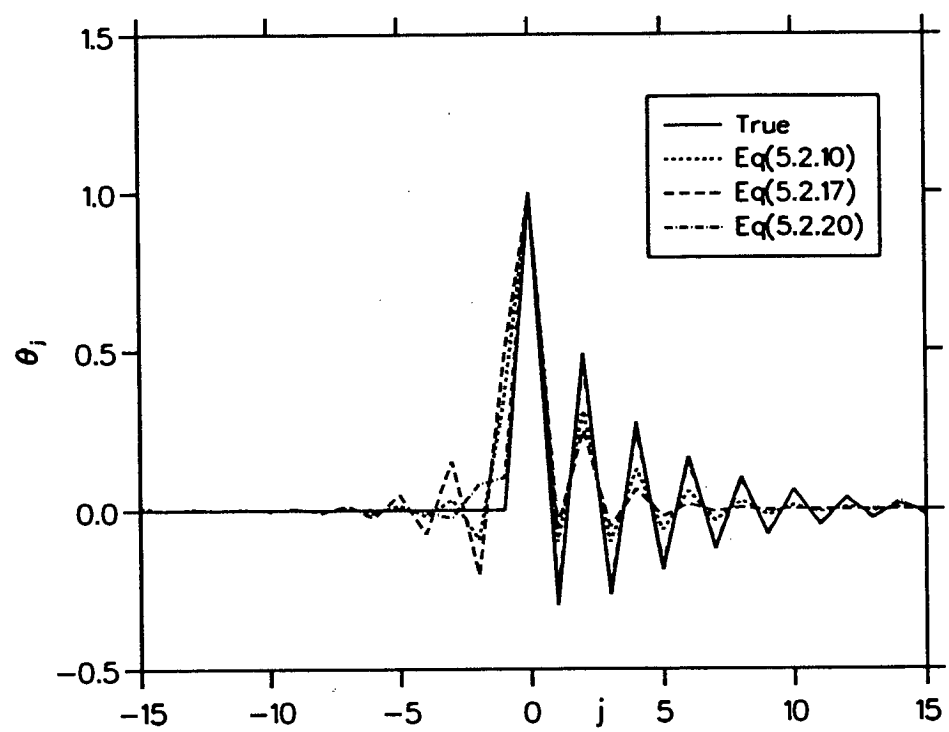
Fig(5.2a) Results of Eq(5.2.10) in Example 5.4.2: minimum phase MA(2) system.



Fig(5.2b) Results of Eq(5.2.17) in Example 5.4.2: minimum phase MA(2) system.



Fig(5.2c) Results of Eq(5.2.20) in Example 5.4.2: minimum phase MA(2) system.



Fig(5.2d) Algorithm comparison: r_1 is set to be −15, minimum phase MA(2) system, SNR=1dB.

estimation errors of the higher order cumulants in practical algorithm implementations. Thus, for minimum systems, it would be better to take $r_1=0$.

Finally, the system order is taken to be 3, and the obtained inverse filters from the three algorithms are illustrated in Fig(5.2e), where, $r_1=0$, $r_2=15$, and $SNR=10\ dB$. Clearly, the algorithms are robust to the order mismatch in the case of MP systems, as previously mentioned in *REMARK* 6.

From above two examples, it can be seen that all three equations, Eq(5.2.10), Eq(5.2.17), & Eq(5.2.20), are robust to the presence of additive noise, all producing satisfactory performances, though there are slight differences in accuracy.

5.4.2 Unskewed & continuous $w(k)$: the case of the 4th-order cumulants

The distribution of $w(k)$ is taken to be uniform over the interval $[-2, 2](\gamma_2=4/3\ ,\gamma_3=0,\text{and}\gamma_4=-32/15)$.

<Example 5.4.3> Mixed phased MA system: The model is

$$x'(k) = w(k)+3w(k-1)+2w(k-2)+w(k-3)+n(k) \tag{5.4.6}$$

with three zeros: $z_1=-2.32472$ and $z_{2,3}=-0.33764\pm j0.56228$.

Set $r_1=-15$ and $r_2=15$. Noise which results in a $SNR=50\ dB$ is added to the output series. First, solving Eq(5.2.24), we obtain the inverse filter shown in Fig(5.3a). Next, by solving Eq(5.2.27), the inverse filter shown in Fig(5.3b) is obtained. Finally, Eq(5.2.29) is solved, and the determined inverse filter is demonstrated in Fig(5.3c). The corresponding identification results are given in Table (5.3).

Table (5.3) Identification Results in Example 5.4.3.

	Eq(5.2.24)	Eq(5.2.27)	Eq(5.2.29)	True
$\hat{b}_0$	1.00000	1.00000	1.00000	1.00000
$\hat{b}_1$	2.97592	2.59161	2.73541	3.00000
$\hat{b}_2$	2.15296	1.82632	1.91538	2.00000
$\hat{b}_3$	1.01440	0.86431	0.91109	1.00000

In addition, the case for $SNR=15\ dB$ was also simulated. The inverse filters obtained from Eq(5.2.24), Eq(5.2.27), and Eq(5.2.29) are also demonstrated in Fig(5.3a), Fig(5.3b), and Fig(5.3c), respectively.

<Example 5.4.4> Minimum phase MA system: The model is

$$x'(k) = w(k)-0.5w(k-1)-0.14w(k-2)+n(k) \tag{5.4.7}$$

with two zeros: $z_1=0.7$ and $z_2=-0.2$.

Set $r_1=0$ and $r_2=15$. Noise which results in $SNR=50\ dB$ and $SNR=15\ dB$ is added, respectively. The determined inverse filters by Eq(5.2.24), Eq(5.2.27), and Eq(5.2.29) are shown in Fig(5.4a), Fig(5.4b), and Fig(5.4c), respectively. The corresponding identification results ($SNR=15$ dB) are listed in Table (5.4). Clearly, this table verifies our

observation about the identification function of the algorithms in Example 5.4.2, *i.e.*, the algorithms are are more robustness to the presence of additive noise for the MP systems than for the NMP ones.

Table (5.4) Identification Results in Example 5.4.4.

	Eq(5.2.24)	Eq(5.2.27)	Eq(5.2.29)	True
$\hat{b}_0$	1.00000	1.00000	1.00000	1.00000
$\hat{b}_1$	−0.47834	−0.47729	−0.47987	−0.50000
$\hat{b}_2$	−0.12946	−0.12929	−0.12973	−0.14000

It can be seen that all the results obtained by the algorithms in above two examples are satisfactory, but the results from Eq(5.2.29) are always better than the ones from the other two equations. As to the corresponding errors, their major sources include: additive noise, cumulant estimation errors, and numerical errors.

5.4.3 Unskewed & discrete $w(k)$: the case of the 4th-order cumulants

$w(k)$ is taken to be an equally distributed discrete series over set $\{\pm 1, \pm 3\}$ ($\gamma_2=5$, $\gamma_4=26$). Additive noise level is set to be 30 *dB*.

<Example 5.4.5> Mixed phased MA system: The model is

$$x'(k) = w(k)+5w(k-1)+4w(k-2)+2w(k-3)+n(k) \tag{5.4.8}$$

with three zeros: $z_1=-4.15276$ and $z_{2,3}=-0.42362\pm j0.54968$.

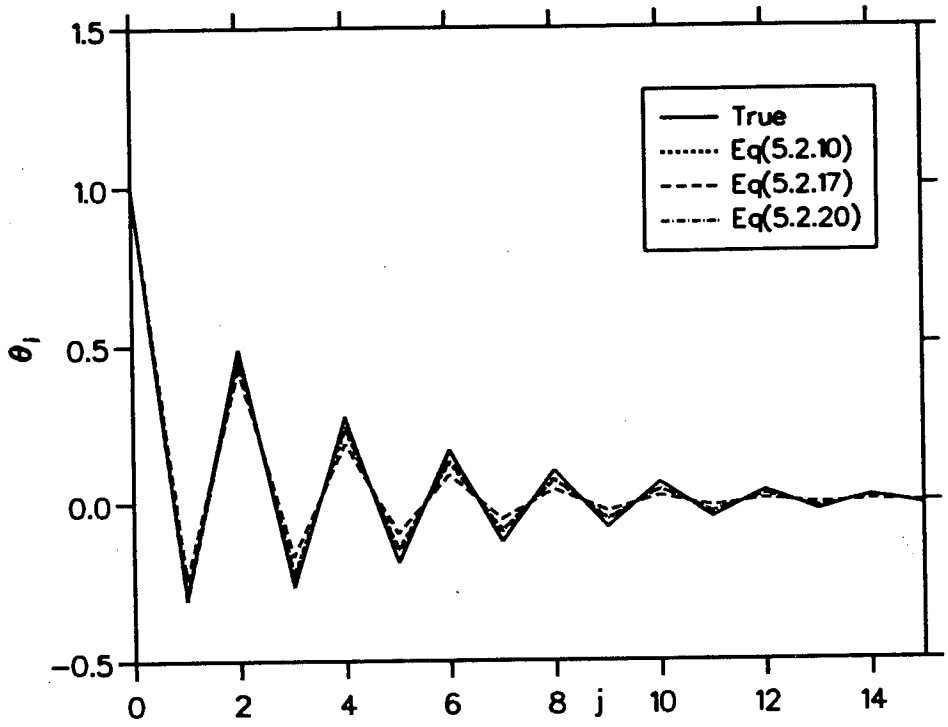
Obviously, this is a typical channel equalisation problem. Set $r_1=-15$ and $r_2=15$. In the presence of 30 dB additive noise, Eq(5.2.24), Eq(5.2.27), and Eq(5.2.29) are solved respectively. The results--inverse filters are as follows. Using Eq(5.2.24): results shown in Fig(5.5a). Using Eq(5.2.27): results shown in Fig(5.5b). Using Eq(5.2.29): results shown in Fig(5.5c). The corresponding identification results are listed in Table (5.5).

Table (5.5) Identification Results in Example 5.4.4.

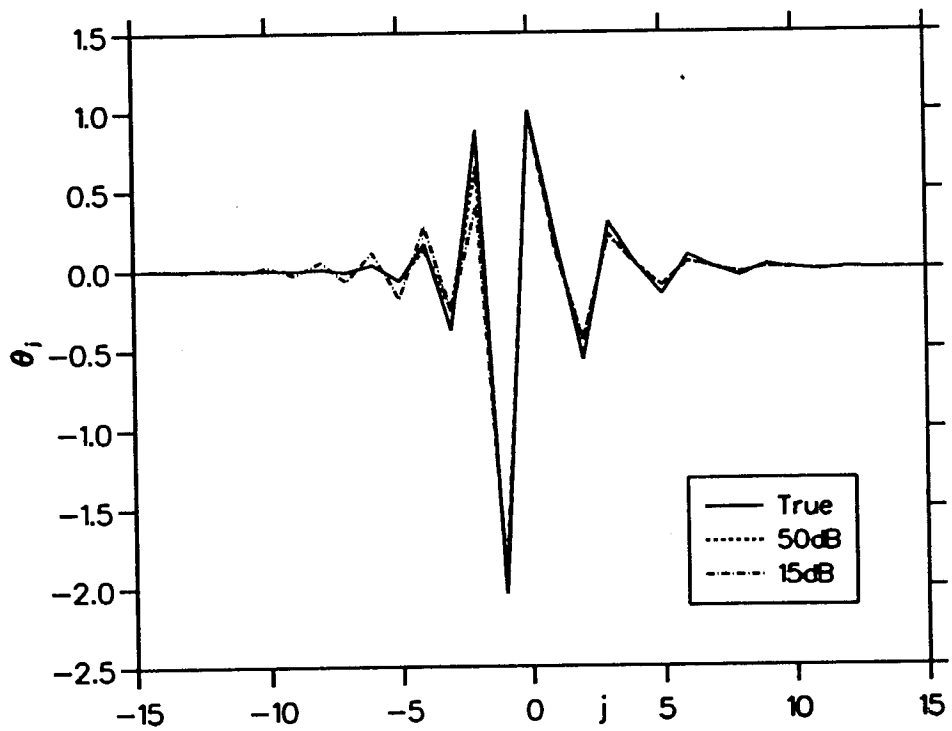
	Eq(5.2.24)	Eq(5.2.27)	Eq(5.2.29)	True
$\hat{b}_0$	1.00000	1.00000	1.00000	1.00000
$\hat{b}_1$	4.72570	5.23728	4.82235	5.00000
$\hat{b}_2$	3.53325	3.91082	3.59407	4.00000
$\hat{b}_3$	1.64270	1.84864	1.68567	2.00000

5.5 Conclusions

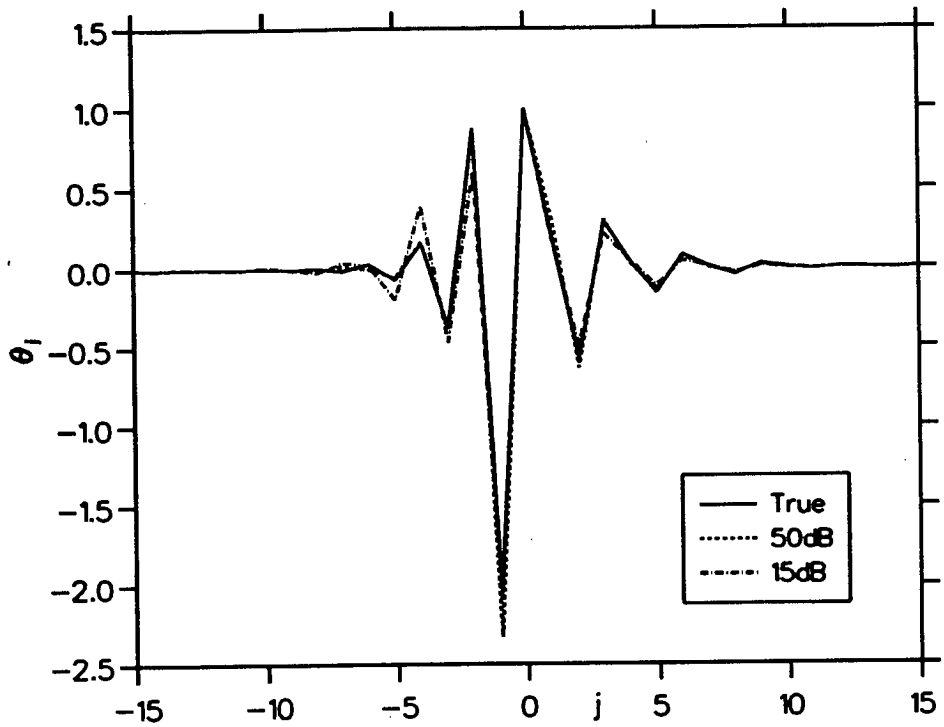
Through the above discussion, six families of linear equations have been derived according to three theorems, and they are all of a similar simple form. As a result of the linearity of these equations, the uniqueness of solutions can normally be guaranteed. Thus,



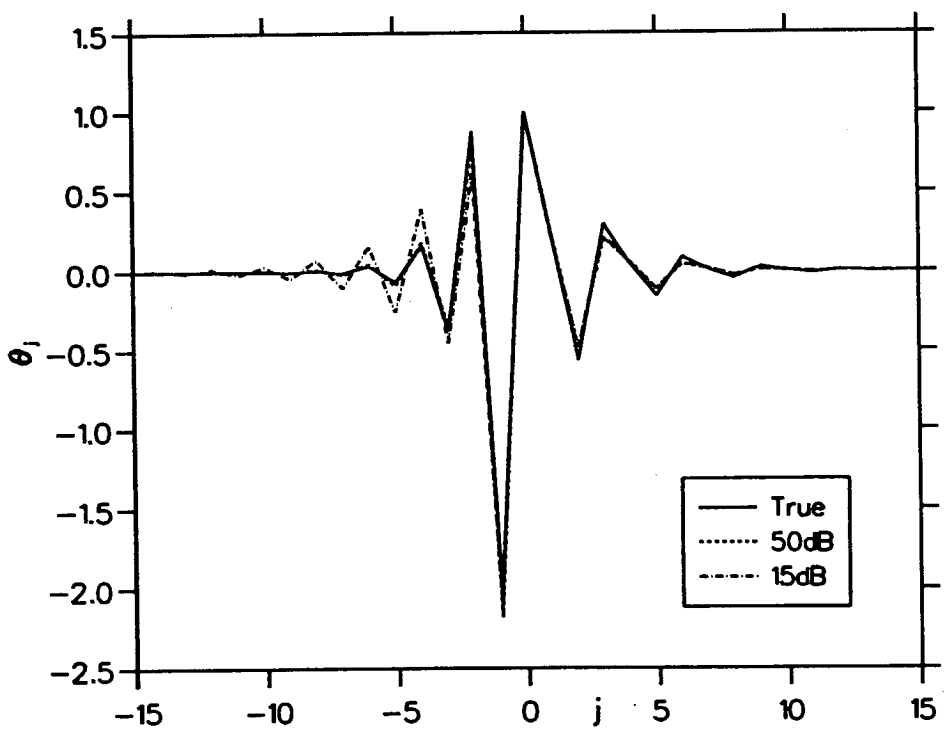
Fig(5.2e) Insensitivity of the algorithms to order mismatch: minimum phase MA(2) system, algorithm order=3.



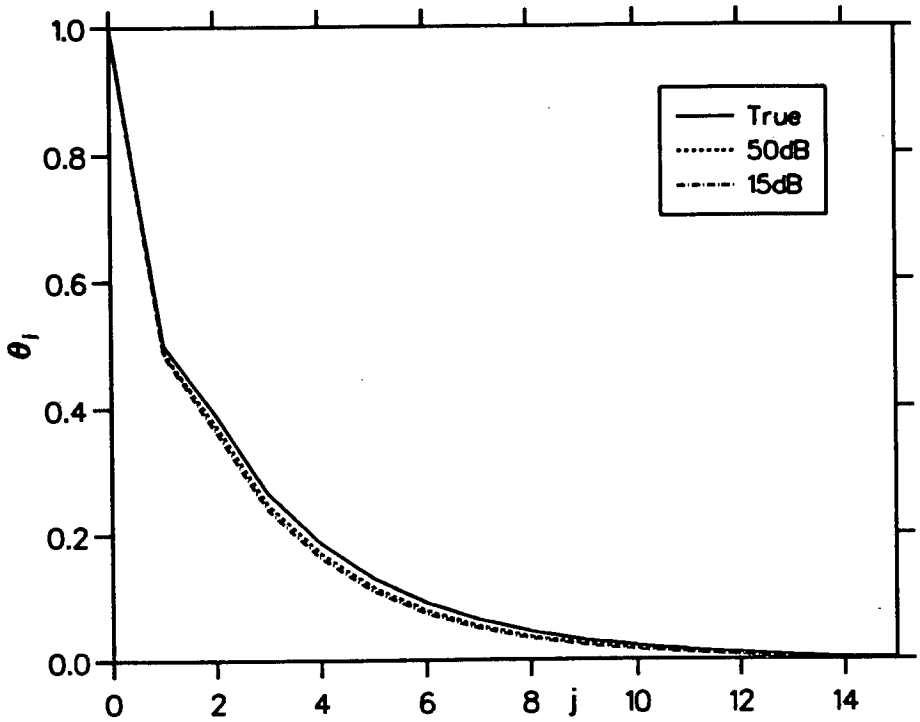
Fig(5.3a) Results of Eq(5.2.24) in Example 5.4.3: mixed phase MA(3) system.



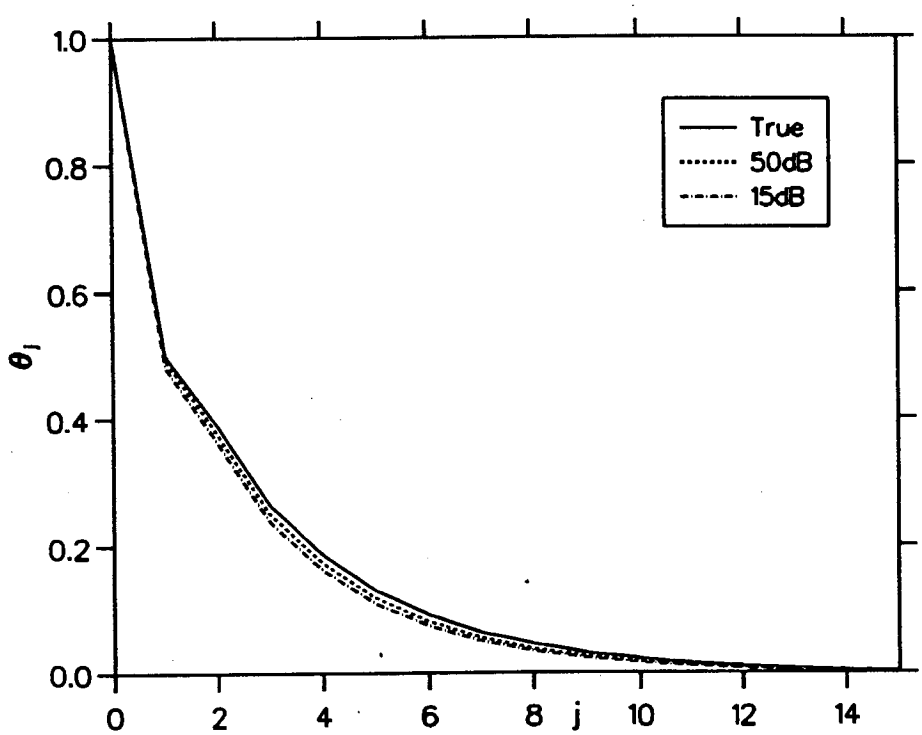
Fig(5.3b) Results of Eq(5.2.27) in Example 5.4.3: mixed phase MA(3) system.



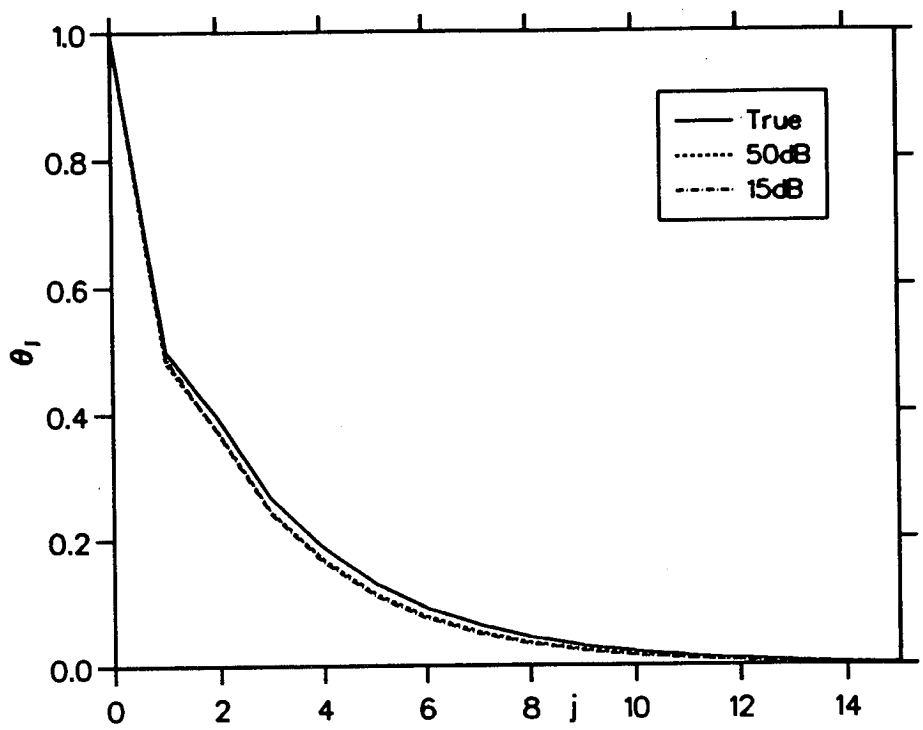
Fig(5.3c) Results of Eq(5.2.29) in Example 5.4.3: mixed phase MA(3) system.



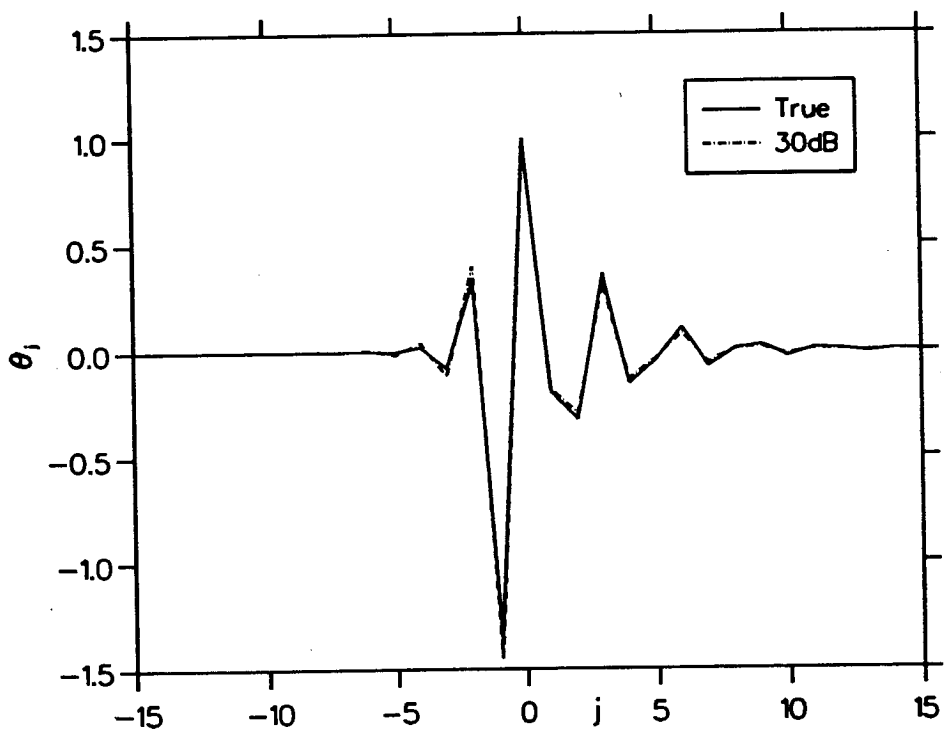
Fig(5.4a) Results of Eq(5.2.24) in Example 5.4.4: minimum phase MA(2) system.



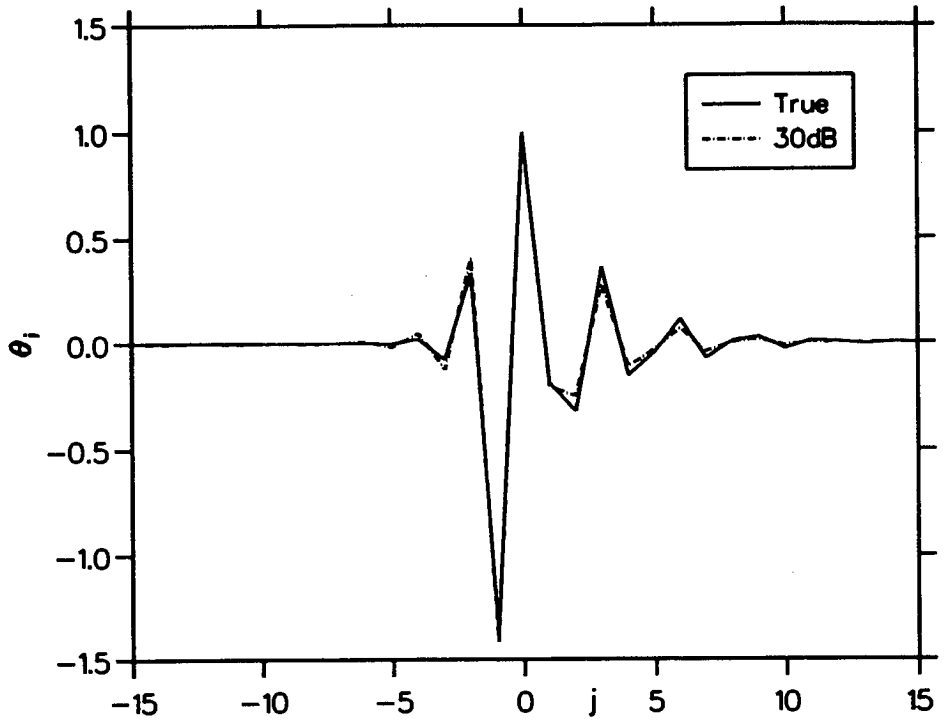
Fig(5.4b) Results of Eq(5.2.27) in Example 5.4.4: minimum phase MA(2) system.



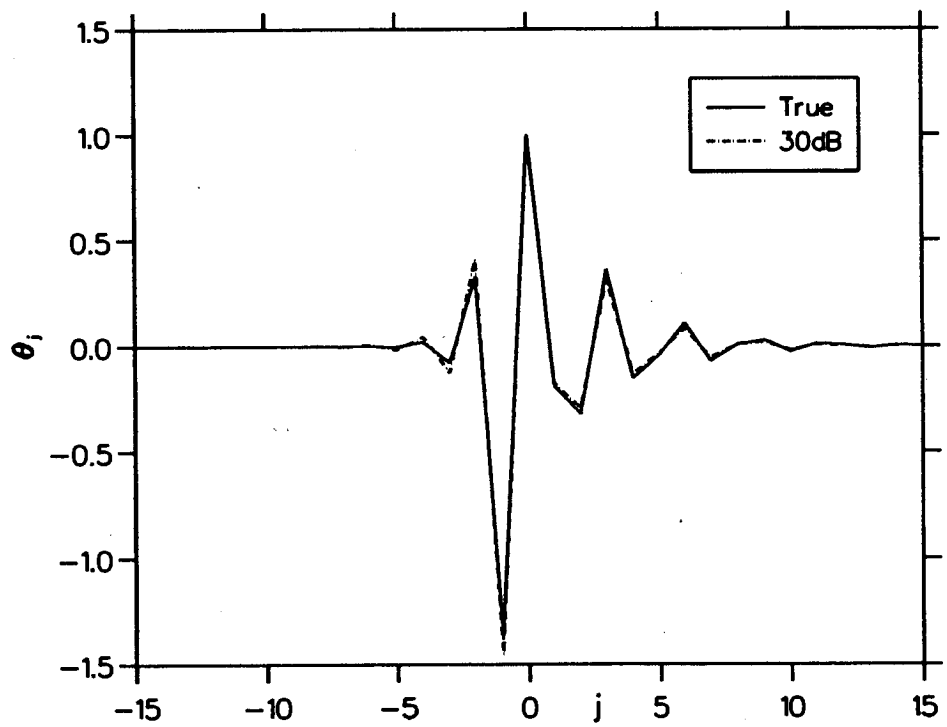
Fig(5.4c) Results of Eq(5.2.29) in Example 5.4.4: minimum phase MA(2) system.



Fig(5.5a) Results of Eq(5.2.24) in Example 5.4.5: mixed phase MA(3) system.



Fig(5.5b) Results of Eq(5.2.27) in Example 5.4.5: mixed phase MA(3) system.



Fig(5.5c) Results of Eq(5.2.29) in Example 5.4.5: mixed phase MA(3) system.

the risk of local minima or contradictory solutions, which exists with some other methods such as Giannakis-Mendel, is removed. To this extent, the results reported in this chapter can be seen as another important progress following Giannakis-Mendel method in blind system identification.

In comparison with existing linear (*i.e*, noncausal AR based) approaches, the algorithms proposed in this chapter possess the following two features:

a) Only the diagonal cumulant slices are used, which in general makes the algorithms more accurate, and also, makes the forms of the concerned matrices simple and elegant;

b) 4th-order cumulants are employed in the last three algorithms exactly as the substitute of 3rd-order cumulants in the first three algorithms, which enables the 4th-order cumulant based algorithms (for deconvolution of a unskewed series) to be implemented as easily as their 3rd-order cumulants based counterparts (for deconvolution of a skewed series).

Appendix 5.1

This appendix provides the mathematical proof for the three theorems involved in this chapter. The symbol system used in the text is followed.

1. Proof of Theorem 1.

From the conditions in Theorem 1 and Eq(5.2.1), we have

$$\Phi_{wx}(i) = \mathrm{E}[w(k)x(k+i)] \tag{A5.1.1}$$

$$= \mathrm{E}[w(k)\cdot\sum_{n=0}^{q} b_n w(k+i-n)] \tag{A5.1.2}$$

$$= \sum_{n=0}^{q} b_n \mathrm{E}[w(k)w(k+i-n)] \tag{A5.1.3}$$

$$= \sum_{n=0}^{q} b_n \gamma_2 \delta(n-i) \tag{A5.1.4}$$

$$= \begin{cases} \gamma_2 b_i & \text{if } i \in [0, q] \\ 0 & \text{other} \end{cases} . \tag{A5.1.5}$$

On the other hand, from Eq(5.2.2), we have

$$\Phi_{wx}(i) = \mathrm{E}[w(k)x(k+i)] \tag{A5.1.1}$$

$$= \mathrm{E}[\sum_{j=r_1}^{r_2} \theta_j x(k-j)x(k+i)] \tag{A5.1.6}$$

$$= \sum_{j=r_1}^{r_2} \theta_j \mathrm{E}[x(k-j)x(k+i)] \tag{A5.1.7}$$

$$= \sum_{j=r_1}^{r_2} \theta_j c_2(i+j) . \tag{A5.1.8}$$

Equating Eq(A5.1.5) and Eq(A5.1.8), Theorem 1 holds •

2. Proof of Theorem 2.

From the conditions in Theorem 2 and Eq(5.2.1), we have

$$\Phi_{wxx}(i) = \mathrm{E}[w(k)x^2(k+i)] \tag{A5.2.1}$$

$$= \mathrm{E}[w(k)\{\sum_{n=0}^{q} b_n w(k+i-n)\}^2] \tag{A5.2.2}$$

$$= \mathrm{E}[w(k)\cdot\sum_{m=0}^{q}\sum_{n=0}^{q} b_m b_n w(k+i-n)w(k+i-m)] \tag{A5.2.3}$$

$$= \sum_{m=0}^{q}\sum_{n=0}^{q} b_m b_n \mathrm{E}[w(k)w(k+i-m)w(k+i-n)] \tag{A5.2.4}$$

$$= \sum_{m=0}^{q}\sum_{n=0}^{q} b_m b_n \gamma_3 \delta(m-i, n-i) \tag{A5.2.5}$$

$$= \begin{cases} \gamma_3 b_i^2 & \text{if } i \in [0, q] \\ 0 & \text{other} \end{cases} . \tag{A5.2.6}$$

On the other hand, from Eq(5.2.2), we have

$$\Phi_{wxx}(i) = \mathrm{E}[w(k)x^2(k+i)] \tag{A5.2.1}$$

$$= \mathrm{E}[\sum_{j=r_1}^{r_2} \theta_j x(k-j)\cdot x^2(k+i)] \tag{A5.2.7}$$

$$= \sum_{j=r_1}^{r_2} \theta_j \mathrm{E}[x(k-j)x^2(k+i)] \tag{A5.2.8}$$

$$= \sum_{j=r_1}^{r_2} \theta_j c_3(i+j) \ . \tag{A5.2.9}$$

Equating Eq(A5.2.6) and Eq(A5.2.9), Theorem 2 holds •

3. Proof of Theorem 3.

From the conditions in Theorem 3 and Eq(5.2.1), we have

$$\Phi_{wxxx}(i) = \mathrm{E}[w(k)x^3(k+i)] \tag{A5.3.1}$$

$$= \mathrm{E}[w(k)\{\sum_{n=0}^{q} b_n w(k+i-n)\}^3] \tag{A5.3.2}$$

$$= \mathrm{E}[w(k)\cdot \sum_{m=0}^{q}\sum_{n=0}^{q}\sum_{l=0}^{q} b_m b_n b_l w(k+i-m)w(k+i-n)w(k+i-l)] \tag{A5.3.3}$$

$$= \sum_{m=0}^{q}\sum_{n=0}^{q}\sum_{l=0}^{q} b_m b_n b_l \mathrm{E}[w(k)w(k+i-m)w(k+i-n)w(k+i-l)] \ . \tag{A5.3.4}$$

For the input series $w(k)$, from Eq(5.2.4c), we have

$$\mathrm{E}[w(k)\prod_{i=1}^{3} w(k+m_i)] = C_4^w(m_1, m_2, m_3)$$

$$+ \gamma_2^2[\delta(m_1)\delta(m_3-m_2)+\delta(m_2)\delta(m_3-m_1)+\delta(m_3)\delta(m_2-m_1)] \ , \tag{A5.3.5}$$

where $C_4^w(m_1, m_2, m_3)$ is the 4th-order cumulants of $w(k)$. According to [82] and [15],

$$C_4^w(m_1, m_2, m_3) = \gamma_4\delta(m_1, m_2, m_3) \ . \tag{A5.3.6}$$

Then, from Eq(A5.3.4), we derive

$$\Phi_{wxxx}(i) = \sum_{m=0}^{q}\sum_{n=0}^{q}\sum_{l=0}^{q} b_m b_n b_l[\gamma_4\delta(i-m, i-n, i-l)$$

$$+\gamma_2^2\{\delta(i-m)\delta(n-l)+\delta(i-n)\delta(m-l)+\delta(i-l)\delta(m-n)\}] \tag{A5.3.7}$$

$$= \begin{cases} \gamma_4 b_i^3 + \gamma_2^2[\sum_{l=0}^{q} b_l^2 b_i + \sum_{m=0}^{q} b_m^2 b_i + \sum_{n=0}^{q} b_n^2 b_i] & \text{if } i \in [0, q] \\ 0 \qquad \text{other} \end{cases} \tag{A5.3.8}$$

$$= \begin{cases} \gamma_4 b_i^3 + 3\gamma_2^2 b_l \sum_{n=0}^{q} b_n^2 & \text{if } i \in [0, q] \\ 0 \qquad \text{other} \ . \end{cases} \tag{A5.3.9}$$

On the other hand, from Eq(5.2.2) and Eq(5.2.4c), we have

$$\Phi_{wxxx}(i) = \mathbf{E}[w(k)x^3(k+i)] \tag{A5.3.1}$$

$$= \mathbf{E}[\sum_{j=r_1}^{r_2} \theta_j x(k-j)\cdot x^3(k+i)] \tag{A5.3.10}$$

$$= \sum_{j=r_1}^{r_2} \theta_j \mathbf{E}[x(k-j)x^3(k+i)] \tag{A5.3.11}$$

$$= \sum_{j=r_1}^{r_2} \theta_j [c_4(i+j)+3c_2(i+j)c_2(0)] \tag{A5.3.12}$$

$$= \sum_{j=r_1}^{r_2} \theta_j c_4(i+j) + 3c_2(0) \sum_{j=r_1}^{r_2} \theta_j c_2(i+j) \ . \tag{A5.3.13}$$

Notice Theorem 1 (Eq(5.2.6)) and the equation

$$c_2(0) = \gamma_2 \sum_{n=0}^{q} b_n^2 \ , \tag{A5.3.14}$$

we can obtain

$$\Phi_{wxxx}(i) = \begin{cases} \sum_{j=r_1}^{r_2} \theta_j c_4(i+j) + 3\gamma_2^2 b_i \sum_{n=0}^{q} b_n^2 & \text{if } i \in [0, q] \\ \sum_{j=r_1}^{r_2} \theta_j c_4(i+j) & \text{other} \ . \end{cases} \tag{A5.3.15}$$

Equating Eq(A5.3.9) and Eq(A5.3.15), we must have

$$\sum_{j=r_1}^{r_2} \theta_j c_4(i+j) = \begin{cases} \gamma_4 b_i^3 & \text{if } i \in [0, q] \\ 0 & \text{other} \end{cases} , \tag{A5.3.16}$$

which is just Theorem 3 ●

Chapter 6:

BLIND EQUALISATION OF COMMUNICATION CHANNELS

6.1 Introduction

As indicated in Chapter 2, the development in modern telecommunications of time-division multiplexing (TDM) systems, multipoint networks, teletext broadcast systems, and so forth [18][21][100]-[102], have necessitated the need for blind equalisation techniques. The objective of blind equalization is to recover the transmitted message from the received one without any *a priori* knowledge of the unknown channel characters. The major advantage of blind equalization over conventional equalization is that it can realize automatic startup and remove the training stage, which is indispensable to conventional equalization. The problem of blind equalization has been well understood for the case of minimum phase (MP) channels, where it reduces to an innovation restoration problem and the well known predictive scheme [20] can be employed. Since not all channels satisfy the condition of MP in practice, the blind equalization of nonminimum phase (NMP) is of considerable significance and has already stimulated many research workers' interest over the past two decades.

In the pioneering work of [18], both the first and second order statistics were adopted although the concept of "order" was not explicitly mentioned. In [19] and [20], the general NMP channel equalization problem is mathematically addressed by assuming an exactly known non-Gaussian distribution: sub-Gaussian or sup-Gaussian distribution. On the other hand, it is well known that NMP channels cannot be equalized for the case of a Gaussian distribution [19]. Hence, the problem which remains to be solved is the equalization of NMP channels for the case of an unknown non-Gaussian distribution. Techniques based solely on autocorrelation functions have proved to be useless in solving this problem.

In the end, higher-order (>2) cumulant (HOC, Fourier transformation of which is termed *higher-order spectra (HOS)* or *polyspectra*) analysis technology provides a tool for solving the problem of blind equalisation of NMP channels. This is because communication channels are only a specific class of NMP systems, and many general results of NMP system identification can then be exploited to accomplish the estimation of NMP channels. Although HOC based blind equalisation techniques have only started to be studied by the signal processing and communications community in the past ten years, considerable progress has been achieved. As in blind identification or deconvolution, parametric methods are increasingly preferred to conventional approaches in blind equalisation as a consequence of their high resolution and low estimation variance. For example, among the suggested techniques are those of Chiang-Nikias' [32][14] and Hatzinakos-Nikias' [103]-[105].

However, it is because HOC based blind equalisation technology is still very young that there are some drawbacks with the existing approaches. As a result, the investigation in this field is still rapidly developing both in depth and in width. In this chapter, the two blind deconvolution techniques proposed respectively in Chapter 4 and 5 are modified and applied to the problem of blind equalisation of NMP channels. It turns out that the obtained equalisation schemes possess several advantages over the existing ones.

In this chapter, the transmitted data is assumed to be independently and uniformly distributed over some discrete finite set:

$$\begin{aligned} \mathbf{S} &= \{s_i = i \mid i = -(2M-1),\ -(2M-3),\ \ldots,\ 2M-1\} \\ &= \{\pm 1,\ \pm 3,\ \ldots,\ \pm(2M-1)\} \quad , \end{aligned} \tag{6.1.1}$$

which corresponds to a $2M$-level pulse amplitude modulation (PAM) scheme. Here, $M \geq 2$. If and only if M is sufficiently large, the above data is close to being sub-Gaussian [19][20]. In practice, however, M is usually not taken to be too large. This is a limitation of the algorithms in [19] and [20]. Fortunately, HOC based methods do not require any assumptions about the distribution except that it be non-Gaussian.

The general structure of this chapter is as follows. First in Section 6.2, the two-step algorithm suggested in Chapter 4 is applied to the discrete data: PAM series. An SAT-AST transformation pair is introduced in order that the 3rd-order cumulant based algorithm can be used. Meanwhile, the performance of our two-step algorithm is further tested in the context of blind equalisation. Then an adaptive version of the AR algorithm (described in Chapter 5) is systematically derived in Section 6.3, and a novel successive over-relaxation (SOR) is employed to accelerate the convergence rate. This SOR based algorithm can even track the change of linear time-variant (LTV) channels. In both of these two sections, several simulations are exhibited. Finally, some conclusions are summarised in Section 6.4.

6.2 MA Model Based Technique

The system studied in this section is shown in Fig(6.1). It is assumed that the channel is stable, linear, and time-invariant. Then, it can always be expressed as the following MA model with finite order q:

$$y_s(k) = y_{inf}(k) + n(k) = \sum_{i=0}^{q} b_i x_s(k-i) + n(k) \quad , \tag{6.2.1}$$

where, $n(k)$ is symmetrically-distributed additive noise, and $x_s(k)$ is the transmitted sequence in which the random data are independently and uniformly distributed over the set $\mathbf{S}$ in Eq(6.1.1).

The general expression for the nth-order cumulants $C_{y_s}(m_1, m_2, \cdots, m_{n-1})$ of $y_s(k)$ can then be obtained from Eq(4.2.1). But in this section, we will only use the 3rd-order cumulants ($n=3$). The reason for this is twofold:

1) The 3rd-order cumulants are relatively easier to estimate accurately than the cumulants of 4th or higher order.

2) The 3rd order cumulant of symmetrically-distributed noise $n(k)$ theoretically vanishes, which allows the developed channel identification algorithm to be robust in presence of additive noise with a symmetric distribution (*e.g.*, Gaussian noise).

In the following discussion of this section, will directly cite the related results derived in Section 4.2.

6.2.1 Symmetry-to-asymmetry transformation of data distribution

For the above data model, unfortunately, the skewness γ_3, and consequently the 3rd-order cumulants $C_{y_s}(m_1, m_2)$, theoretically vanishes, because $x_s(k)$ itself has a symmetric distribution! In order to overcome this point, we suggest the following scheme of transformation from a symmetric distribution to an asymmetric one.

The proposed system is illustrated in Fig(6.2). At the transmitter end, the following symmetric-to-asymmetric transformation (SAT) is implemented before transmitting:

$$\{s_i\} \rightarrow \{a_i = \ln[K \cdot (2M + s_i)] - \mu\} \quad , \tag{6.2.2}$$

i.e.,

$$x_s(k) \rightarrow x(k) = \ln[K \cdot (2M + x_s(k)] - \mu \quad , \tag{6.2.3}$$

where

$$\mu = \mathrm{E}[\ln(K \cdot (2M + x_s(k)))] = \boxed{\frac{1}{2M} \sum_{\substack{i=-(2M-1) \\ i \in \{\text{odd numbers}\}}}^{2M-1} \ln(K \cdot (2M + s_i))} \tag{6.2.4}$$

and K is a compression factor. The skewness of $x(k)$ can then be written as

$$\gamma_3 = \mathrm{E}[x^3(k)] = \boxed{\sum_{\substack{i=-(2M-1) \\ i \in \{\text{odd numbers}\}}}^{2M-1} a_i^3 P(a_i)} \quad , \tag{6.2.5}$$

where $P(a_i)$ denotes the probability at which the value a_i happens. Notice

$$P(a_i) = P(s_i) = \frac{1}{2M} \quad , \tag{6.2.6}$$

Eq(6.2.5) can be rewritten as

$$\gamma_3 = \boxed{\frac{1}{2M} \sum_{\substack{i=-(2M-1) \\ i \in \{\text{odd numbers}\}}}^{2M-1} [\ln(K \cdot (2M + i) - \mu]^3} \tag{6.2.7}$$

In this way, $x(k)$ and $y(k)$ become zero-mean skewed discrete sequences, which permit the application of the 3rd-order cumulant.

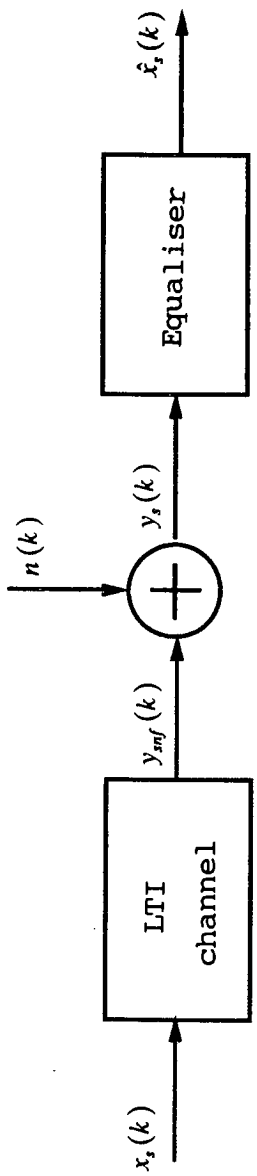
In order to restore the original sequence at the receiver end, it is necessary to perform the corresponding inverse transformation: asymmetric-to-symmetric (AST) one:

$$\{a_i\} \rightarrow \{s_i = \frac{1}{K} e^{\mu + a_i} - 2M\} \quad , \tag{6.2.8}$$

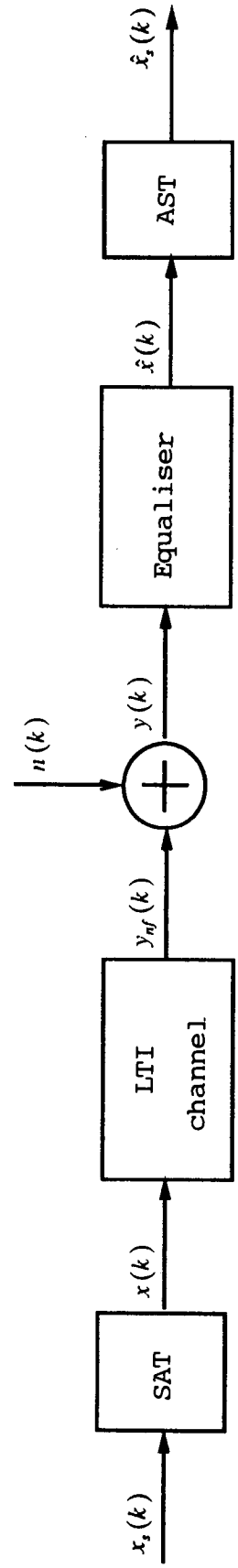
i.e.,

$$\hat{x}(k) \rightarrow \hat{x}_s(k) = \frac{1}{K} e^{\mu + \hat{x}(k)} - 2M \quad . \tag{6.2.9}$$

As to the value of factor K, considering transmission of the restoration error in Eq(6.2.9),



Fig(6.1) The channel-equaliser system.



Fig(6.2) The proposed channel-equaliser system.

it can be shown that $\frac{1}{2M} \le K \le \frac{3}{4M}$ is appropriate. Additionally, from Eq(6.2.4)and Eq(6.2.7), it can be proven that γ_3 does not change with variation of K.

Next the equalization of the received sequence $y(k)$ is considered. From Eq(6.2.1), the process model can be rewritten as

$$y(k) = y_{nf}(k) + n(k) = \sum_{i=0}^{q} b_i x(k-i) + n(k) \quad , \tag{6.2.10}$$

and from Eq(4.2.1), the corresponding 3rd-order cumulant expression here reduces to

$$C_y(m_1, m_2) = \gamma_3 \sum_{i=L}^{U} b_i b_{i+m_1} b_{i+m_2} + C_n(m_1, m_2) \quad , \tag{6.2.11}$$

where $L = \max(0, -m_1, -m_2)$, and $U = \min(q, q-m_1, q-m_2)$.

6.2.2 Application of two-step approach to channel equalisation

Channel identification is the key step in the equalization technique suggested here. From the foregoing model assumption, the problem of channel identification reduces to the estimation of the parameter vector $\vec{b} = \{b_0, b_1, \ldots, b_q\}$ in Eq(6.2.11) by means of the 3rd-order cumulants. Clearly, this problem is nearly the same as the one with which we dealt in Chapter 4. The only difference is that here the data consist of only discrete values: $\{\pm 1, \pm 3, \ldots\}$. As a consequence, the application of the two-step approach derived in Chapter 4 becomes very straightforward. The main points of the obtained equalisation algorithm are listed as follows.

<1> *Calculating* $R_y(m_1, m_2)$=estimate of $C_y(m_1, m_2)$ *by using the segment-average scheme:*

First, segment N received data samples into P records of Q samples each, *viz.*, $N = P \times Q$. Then estimate the cumulants $R_y^{(i)}(m_1, m_2)$ for the i-th record by

$$R_y^{(i)}(m_1, m_2) = \frac{1}{Q} \sum_{k=k_l}^{k_u} y^{(i)}(k) y^{(i)}(k+m_1) y^{(i)}(k+m_2) \quad , \tag{6.2.12}$$

where $y^{(i)}(k)$ denotes the samples in the i-th record. In addition, $i = 1, 2, \ldots, P$, $k_l = \max(0, -m_1, -m_2)$ and $k_u = \min(Q-1, Q-m_1-1, Q-m_2-1)$. Finally, average $R_y^{(i)}(m_1, m_2)$ over all records:

$$R_y(m_1, m_2) = \frac{1}{P} \sum_{i=1}^{P} R_y^{(i)}(m_1, m_2) \quad . \tag{6.2.13}$$

<2> *Order selection:*

If the channel order q is known, go to <3>; Otherwise, it needs to be determined. The empirical order selection method below has been employed in our work (for denotations, see subsection 4.2.2):

STEP 1: According to experience and computing limitation, preset J as a positive integer J_u, which should be "higher" than the anticipated order q;

STEP 2: Calculate $\hat{S}(J)$, $J = 1, \ldots, J_u$;

STEP 3: Search the minimum: $\hat{S}(J_m) = \min_J \{\hat{S}(1), \ldots, \hat{S}(J), \ldots, \hat{S}(J_u)\}$;

STEP 4: $q = J_m - 1$.

<3> *Initial estimate selection:*

Giannakis' formula can be adopted, *i.e.*,

$$b_0 = [\frac{C_y^2(q, 0)}{\gamma_3 \cdot C_y(q, q)}]^{\frac{1}{3}} \tag{6.2.14a}$$

and

$$b_l = b_0 \cdot \frac{C_y(q, l)}{C_y(q, 0)} \quad , \quad l = 1, \ldots, q. \tag{6.2.14b}$$

<4> *Parameter optimisation:*

$$\nabla_l = \frac{\partial \epsilon^2}{\partial b_l}$$

$$= 2 \cdot \sum_{m=-q}^{q} [\gamma_3 \cdot \sum_{i=\max(0, -m)}^{\min(q, q-m)} b_i b_{i+m}^2 - R_y(m, m)](b_{l+m}^2 + 2 b_l b_{l-m}) \quad , \tag{6.2.15a}$$

$$b_l \leftarrow b_l - \tau \cdot \nabla_l \quad , \tag{6.2.15b}$$

$$l = 0, \ldots, q \ .$$

<5> *Signal restoration:*

In order to determine the coefficients θ_i's of the equaliser, Eq(4.3.17) and the related equations in Section 4.3 can be directly employed here with no modification. The received sequence $y(k)$ then forms the input to the equaliser designed above, and the restored sequence can be obtained thus:

$$\hat{x}(k) = \sum_{i=r_1}^{r_2} \theta_i y(k-i) \quad , \tag{6.2.16}$$

and after implementing AST of Eq(6.2.9), the final equalization result can obtained.

6.2.3 Algorithm discussions

In order to apply the 3rd-order cumulant analysis to the equalization of channels driven by an unskewed discretely-distributed sequence, a transformation scheme (SAT) was proposed above. As to the form of the transformation function, it should be pointed out that, in addition to the logarithmic form, other forms such as reciprocal or square-root are also possible. However, it was observed that the cumulants of the received sequences can become more difficult to estimate when some forms of transformation functions are adopted. The study of this problem is under way.

On the other hand, if a "non-uniform modulation" at the transmitter was adopted according to the data distribution after the above SAT, and the corresponding change implemented for demodulation at the receiver , the "SAT" and "AST" stage in Fig(6.2) could be removed. In addition, for the transmitted sequence, the algorithm presented here requires $M \geq 2$, whereas Sato's scheme can be employed for the case of $M = 1$.

In the case of an unskewed driving sequence, 4th-order cumulant analysis can also be used, and in principle, the proposed algorithm can be extended to the 4th-order case. This, however, suffers from several limitations, as mentioned earlier. In addition, the danger of the algorithm failing to converge to the global minimum greatly increases. It is also worth noting that the assumption of "symmetrically-distributed noise" is often more reasonable than that of "Gaussian noise".

As to computational complexity, the burden of this algorithm is obviously higher than that of the simple Sato scheme. In the later simulation section, however, it can be seen that the proposed algorithm outperforms the corresponding Sato scheme in terms of convergence rate and steady state MSE. To this extent, the computational burden with our approach can be considered as the penalty incurred for the improved performance.

To summarise, the main features of the proposed algorithm are as follows. Firstly, it can correctly equalize received sequences for both MP channels and NMP channels, because HOC reflect the phase property of the channel. Secondly, the channel identification scheme is robust in the presence of additive symmetrically-distributed noise which corrupts the received data, since the 3rd-order cumulant of this type of noise theoretically vanishes. Thirdly, it can automatically determine the orders of inverse filters as long as an accuracy control factor is preset. Fourthly, since optimization schemes are adopted both in MA parameter estimation and in inverse filter design, the obtained equalization results are usually acceptable.

6.2.4 Computer simulations

In this section, the proposed algorithm is applied to the equalization of NMP channels. The transmitted data sequence $x_s(k)$ consists of random variables which are equally distributed over the 4-level set $S=\{\pm 1, \pm 3\}$ (*i.e.*, the case of $M=2$ in Eq(6.1.1). For each example, we perform twenty Monte-Carlo (MC) runs with $N=P\times Q=25\times 256$ data each. The $x_s(k)$ waveform for the first run is shown in Fig(6.3) (only the first 150 points are drawn, similarly hereinafter). Taking the compression factor $K=1/2M=0.25$, from Eq(6.2.4) and Eq(6.2.7), we can obtain: $\mu=-0.22280$ and $\gamma_3=-0.25191$, respectively. After the SAT of Eq(6.2.3), the obtained sequence $x(k)$ for the first run is given in Fig(6.4). The true received sequence $y_{nf}(k)$ for each run is generated, as indicated in Eq(6.2.10), by convolving $x(k)$ with the true channel model, and Gaussian noise $n(k)$ is then added to $y_{nf}(k)$ to produce $y(k)$ with the *SNR* being obtained from $10\log_{10}\dfrac{\mathrm{E}[y^2(k)]}{\mathrm{E}[n^2(k)]}$.

For comparison purposes, the corresponding Sato scheme is also implemented. A brief description of Sato scheme is given in Appendix 6.1.

[EXAMPLE 6.2.1] *The case of MA channel.*

The true channel transfer function is assumed to be

$$H(z) = 1+4.5z^{-1}+2z^{-2} \quad , \tag{6.2.17}$$

which is obviously mixed-phase. The first MC run is then carried out as follows. First, in the presence of 50 dB noise, the model order selection is implemented, and the curve of $\hat{S}(J)$ is illustrated in Fig(6.5), in which $\hat{S}(J)$ reaches the minimum at $J=3$, thus the order

$q=3-1=2$. The channel identification is then realized. The result is listed in Table (6.1), and the corresponding learning curve is drawn in Fig(6.6). In addition, channel identification is also implemented in the case of 20 dB noise. The corresponding $\hat{S}(J)$ curve and identification results are also illustrated in Fig(6.5) and Table (6.1), respectively. But the learning curve is given in Fig(6.7). To compare the identified channel with the true one more clearly, we plot the true and estimated channel frequency responses in Fig(6.8). All foregoing learning curves and frequency responses have been normalised, hereinafter. It can be seen that our algorithm is very robust to the additive noise in terms of the identification results.

Table (6.1) Channel identification results: 1st MC run.

		SNR = 50 *dB*		*SNR* = 20 *dB*	
$\vec{b}$	*true*	*initial*	*estimated*	*initial*	*estimated*
b_0	1.00000	0.14826	1.05982	0.46192	1.07033
b_1	4.50000	8.85113	4.53011	4.85908	4.53612
b_2	2.00000	5.65169	1.82962	3.48967	1.82223

The corresponding Sato scheme is also simulated in the presence of 50 dB and 20 dB noise. The order of noncausal part $r_1=-10$, the order of causal part $r_2=10$, and the step size $\tau=5.0\times10^{-4}$. The learning curves are also illustrated in Fig(6.6) and Fig(6.7), respectively. Clearly, the proposed approach has a much faster convergence rate than Sato scheme.

Nineteen other MC runs of our algorithm are similarly implemented, and the general identification results are listed in Table (6.2) (mean$\pm$variance).

Table (6.2) Channel identification results: 20 MC runs.

		SNR = 50 *dB*	*SNR* = 20 *dB*
$\vec{b}$	*true*	*estimated*	*estimated*
b_0	1.00000	0.93964$\pm$0.10736	0.95218$\pm$0.11144
b_1	4.50000	4.47019$\pm$0.06723	4.48838$\pm$0.06440
b_2	2.00000	1.95622$\pm$0.14801	1.97420$\pm$0.14654

[EXAMPLE 6.2.2] *The case of ARMA channel.*

The true channel transfer function is taken to be

$$H(z)=\frac{1+5.6z^{-1}+3z^{-2}}{1+0.3z^{-1}}\quad , \tag{6.2.18}$$

which is also mixed phase. The first run is then implemented. Let $SNR=50\ dB$. The curve of $\hat{S}(J)$ is plotted in Fig(6.9), from which the equivalent MA model order should be

selected as $q=4-1=3$. Similarly, the equivalent MA model parameters are estimated, and listed in Table (6.3). For comparison, the true impulse response (IR) $\vec{h}$ of the channel is also given in Table (6.3). The corresponding learning curve is shown in Fig(6.10). In order to study the robustness of our algorithm to order mismatching, we set $q=4$ and $q=5$, respectively, and make the same simulation as in the case of $q=3$. The corresponding results are also given in Table (6.3) and Fig(6.10). Fig(6.11) illustrates the corresponding channel frequency responses.

Table (6.3) Channel identification results under different q values: 1st MC run.

	$q=3$		$q=4$		$q=5$		*IR*
$\vec{b}$	*initial*	*estimated*	*initial*	*estimated*	*initial*	*estimated*	$\vec{h}$
b_0	1.86251	1.07464	4.81420	1.09246	2.29508	1.09297	1.00000
b_1	-1.16580	5.33401	13.49702	5.33343	1.00317	5.33343	5.30000
b_2	0.78238	1.21479	11.09953	1.20137	0.45498	1.20091	1.41000
b_3	1.54231	-0.48243	7.67586	-0.50251	0.50401	-0.50267	-0.42300
b_4	—	—	0.05581	0.22382	0.96133	0.21896	0.12690
b_5	—	—	—	—	1.68784	0.05016	-0.03807

Notice that, when $q=4$, the learning curve is smoothest, the minimum normalised MSE is lowest, and the frequency response is nearly equally close to the true one. Although the selection of $q=3$ does not seem to be as good as that of $q=4$ in this example, the final equalization results with $q=3$ and $q=5$ are still satisfactory, as can be seen.

Occasionally, the steepest descent process (SDP) may converge to a local minimum because the initial estimate from Eq(6.2.14) and Eq(6.2.15) is not accurate enough. To try and overcome this problem, the following scheme was adopted: when the SDP reaches a minimum point $\vec{b}_m$, reset the initial estimate of $\vec{b}$ as $-\vec{b}_m$, and restart the SDP, *etc*. Finally, take the $\vec{b}_m$ which makes ϵ^2 smallest as the global solution. This is why there exists a peak respectively when $q=3$ and $q=5$ in Fig(6.10): they are both restart points. However, the trouble encountered here by no means illustrates the total failure of the approach. On the contrary, as can be seen from the following simulations, it is the two-step scheme that, to a great extent, overcome the above problem.

Now, still let $q=4$ and $SNR=50\ dB$. We use the conventional (or randomly chosen!) initial estimate (CIE) in the SDP, *i. e.*, $\vec{b}_0=\{1, 0, 0, 0, 0\}$, instead of using the initial estimate generated by Giannakis formula (GIE) as before. It is observed that the SDP immediately converges to a local minimum: $\vec{b}_m=\{5.20882, 2.14187, -0.63981, 0.21779, 0.04560\}$. After resetting, the global minimum $\vec{b}_m=\{1.09246, 5.33343, 1.20137, -0.50251, 0.22382\}$ is reached. The learning curve is demonstrated in Fig(6.12). But under the same conditions, as we have seen earlier in this example, our algorithm directly converges to the

global minimum (listed in Table (6.3)). To compare even more clearly, we redraw the learning curve (see Fig(6.10)) of our algorithm in Fig(6.12).

Similar simulations are also implemented in the case of 35 *dB* noise. Again, when the CIE is used, the SDP goes to the same local minimum: $\vec{b}_m$ ={5.21076, 2.13764, -0.63702, 0.22706, 0.03933}, and arrives at the global minimum $\vec{b}_m$ ={1.09047, 5.33450, 1.20106, -0.50013, 0.23114} after resetting. But our algorithm directly catches the global: $\vec{b}_m$ ={1.09043, 5.33450, 1.20105, -0.50013, 0.23114} after resetting. The corresponding learning curves are also illustrated in Fig(6.12). Also, the identification results in this example show the robustness of our algorithm to the additive noise.

The corresponding learning curve of Sato scheme in the case of 50 dB noise is also shown in Fig(6.10). Similarly, $r_1 = -10$, $r_2 = 10$, and $\tau = 5.0 \times 10^{-4}$. Again, our algorithm outperforms the Sato scheme in convergence rate.

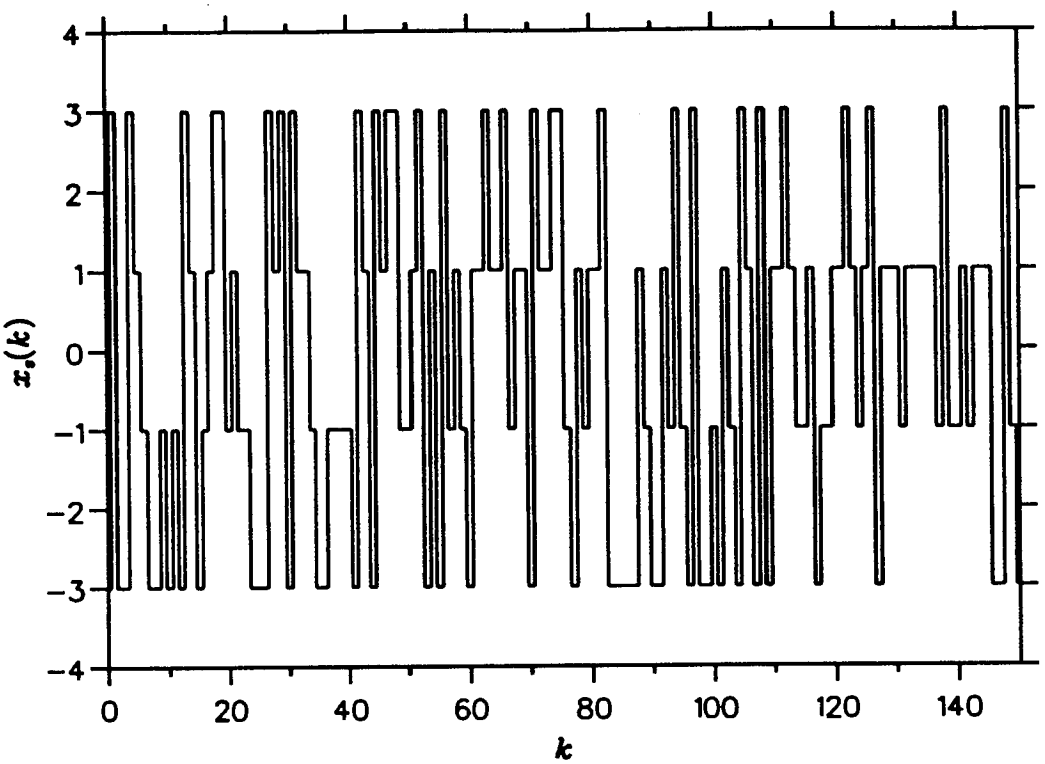
Similarly, nineteen MC runs of our algorithm are realised ($SNR = 50\ dB$), and the general identification results (mean±variance) are given in Table (6.4).

Table (6.4) Channel identification results under different q values: 20 MC runs.

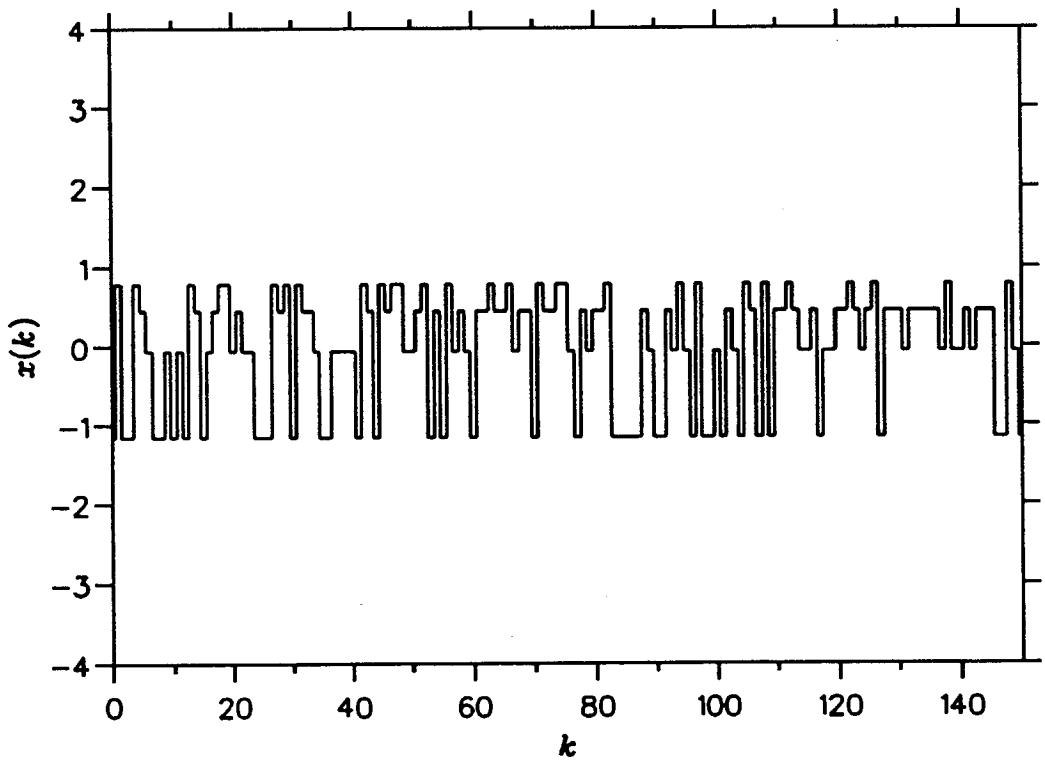
	$q=3$	$q=4$	$q=5$	*IR*
$\vec{b}$	*estimated*	*estimated*	*estimated*	$\vec{h}$
b_0	0.96917±0.09525	0.97381±0.09796	0.97457±0.10052	1.00000
b_1	5.27949±0.06382	5.28413±0.06919	5.27666±0.06457	5.30000
b_2	1.333050±0.13345	1.34793±0.15143	1.30858±0.13666	1.41000
b_3	-0.46967±0.09941	-0.48374±0.10846	-0.48984±0.10259	-0.42300
b_4	—	0.05305±0.13299	0.08223±0.13252	0.12690
b_5	—	—	-0.03703±0.12579	-0.03807

6.2.5 Conclusions

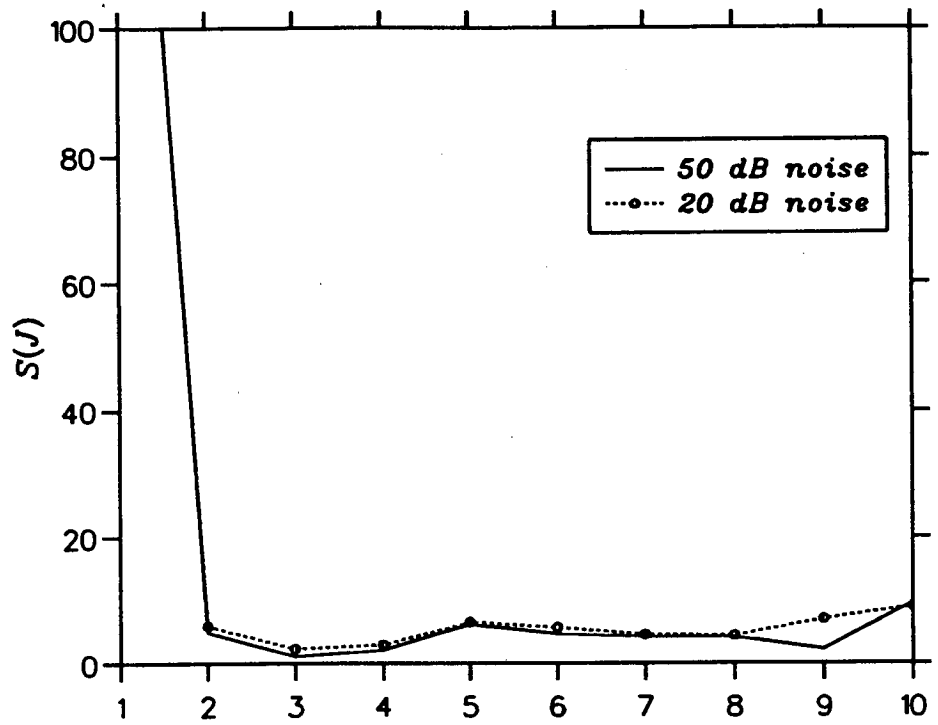
A novel third-order cumulant based equalization algorithm has been presented, in which a two-step relay channel identification method and a transformation scheme are proposed. For this algorithm, the channel can be nonminimum phase, and the transmitted data can be of any level number greater than 2. The multimodal problem is partially overcome. The algorithm has a faster convergence rate in comparison with such non-HOC based methods as Sato scheme. The computer simulations presented for 4-level PAM data verified these points. Also, in our work, we have successfully simulated, but not presented here, equalization of 8-level PAM data sequence using the above algorithm.



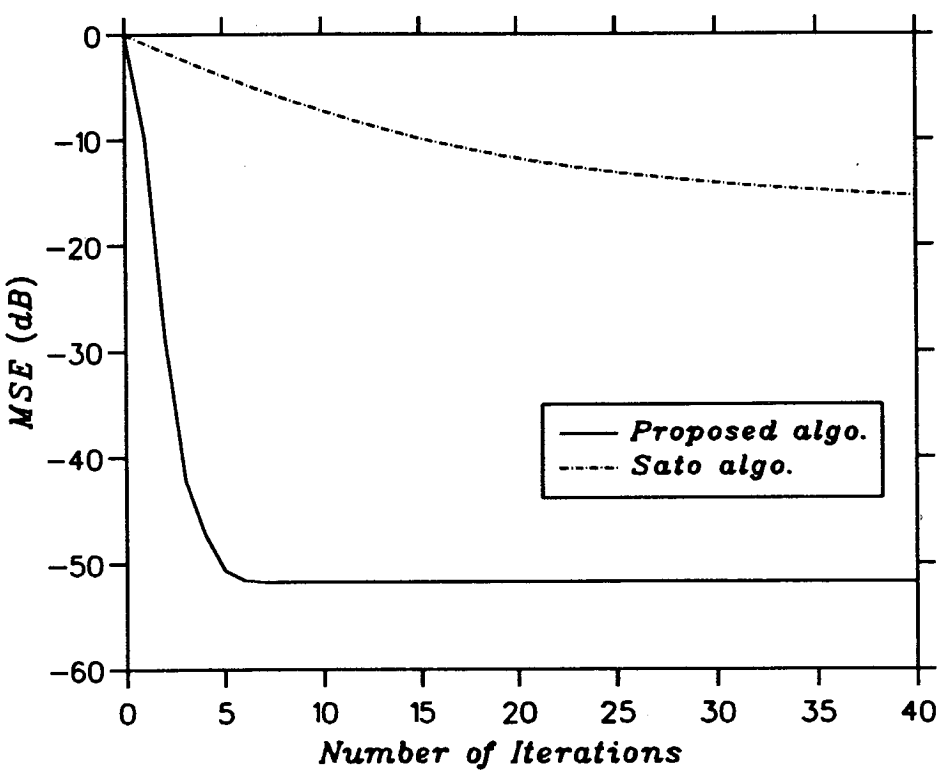
Fig(6.3) The transmitted sequence (before SAT).



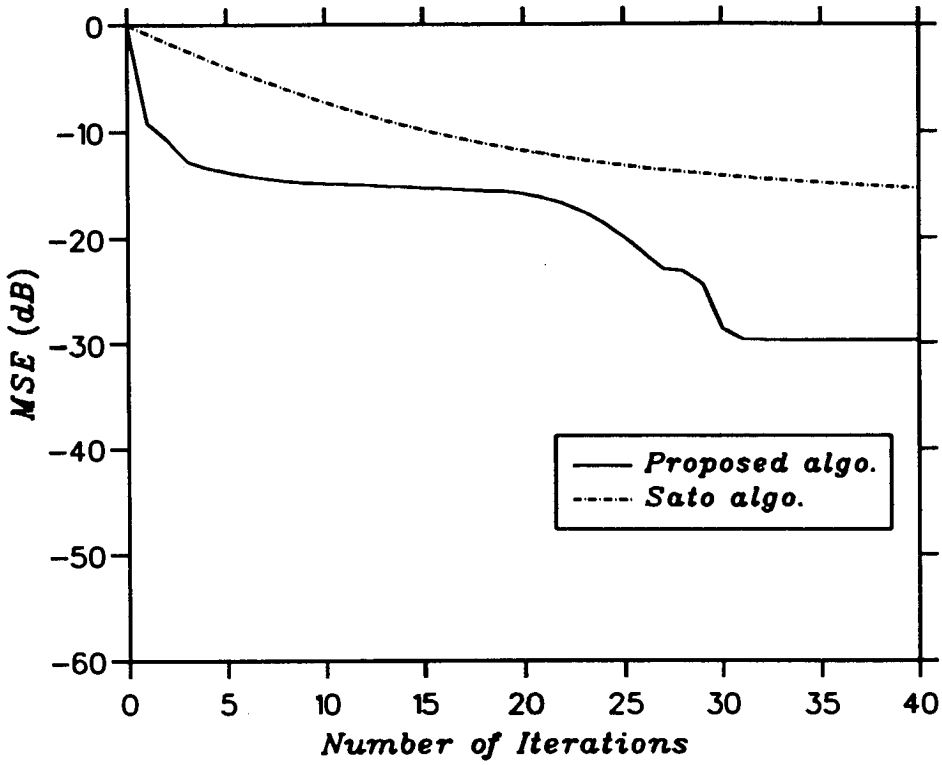
Fig(6.4) The transmitted sequence (after SAT).



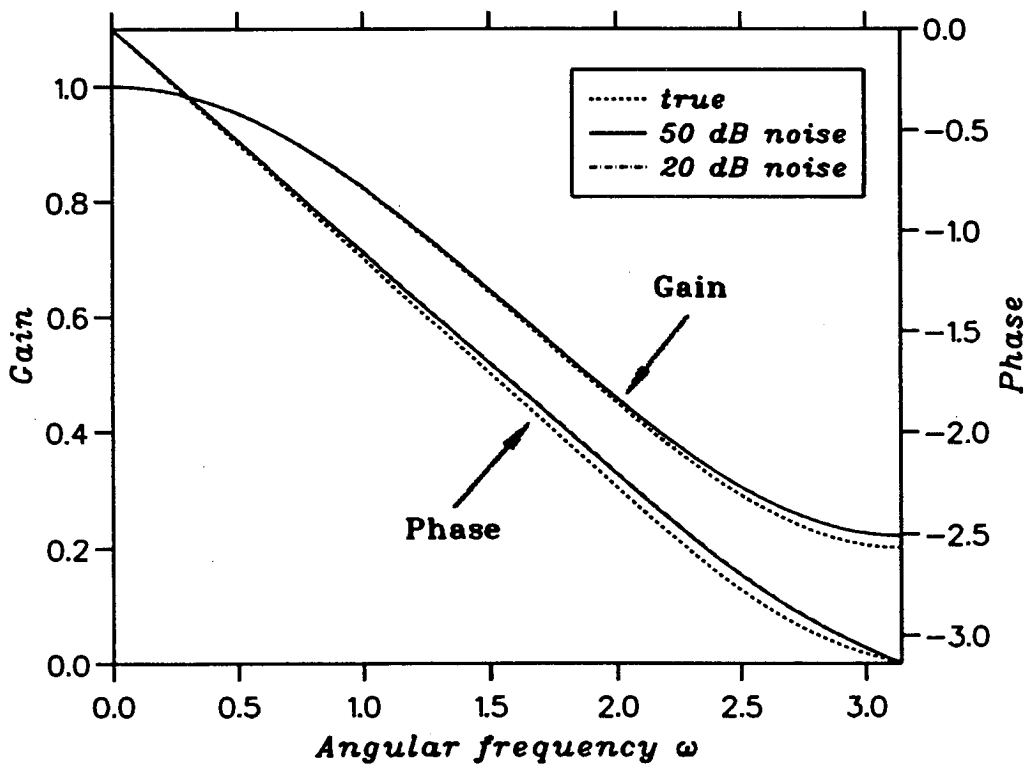
Fig(6.5) The curves of S(J) in Example 6.2.1.



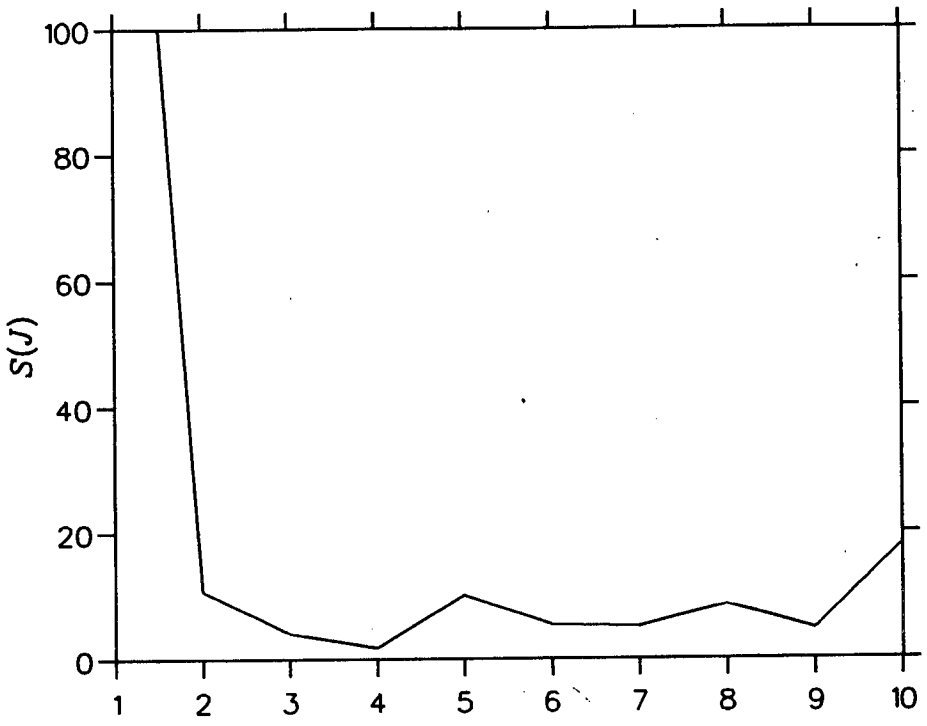
Fig(6.6) The learning curves in Example 6.2.1: 50 dB noise.



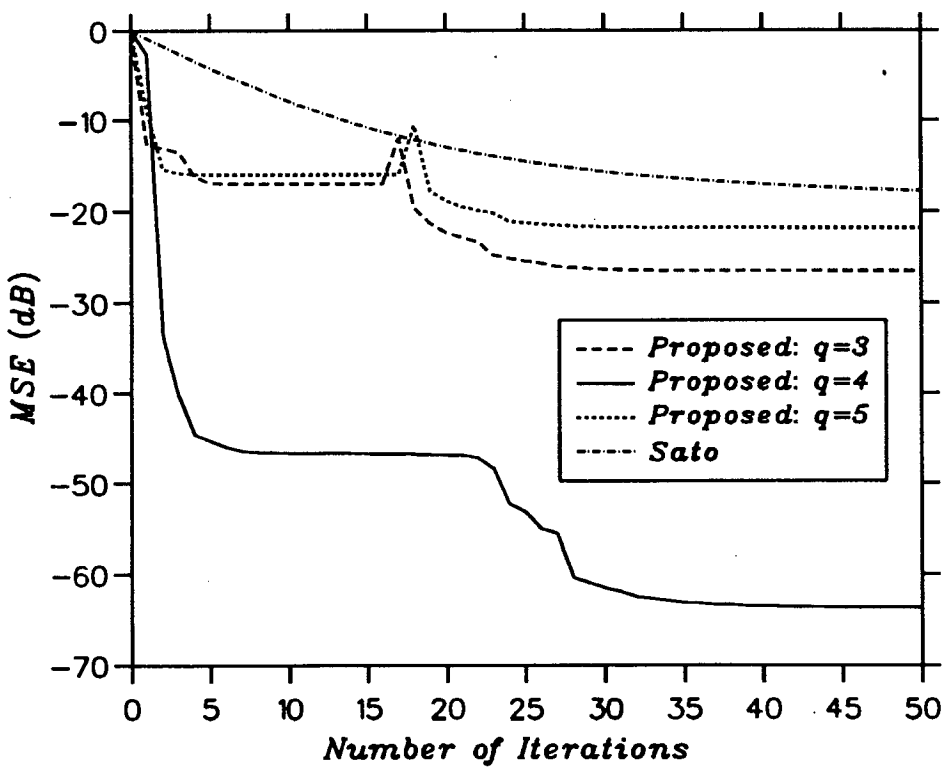
Fig(6.7) The learning curves in Example 6.2.1: 20 dB noise.



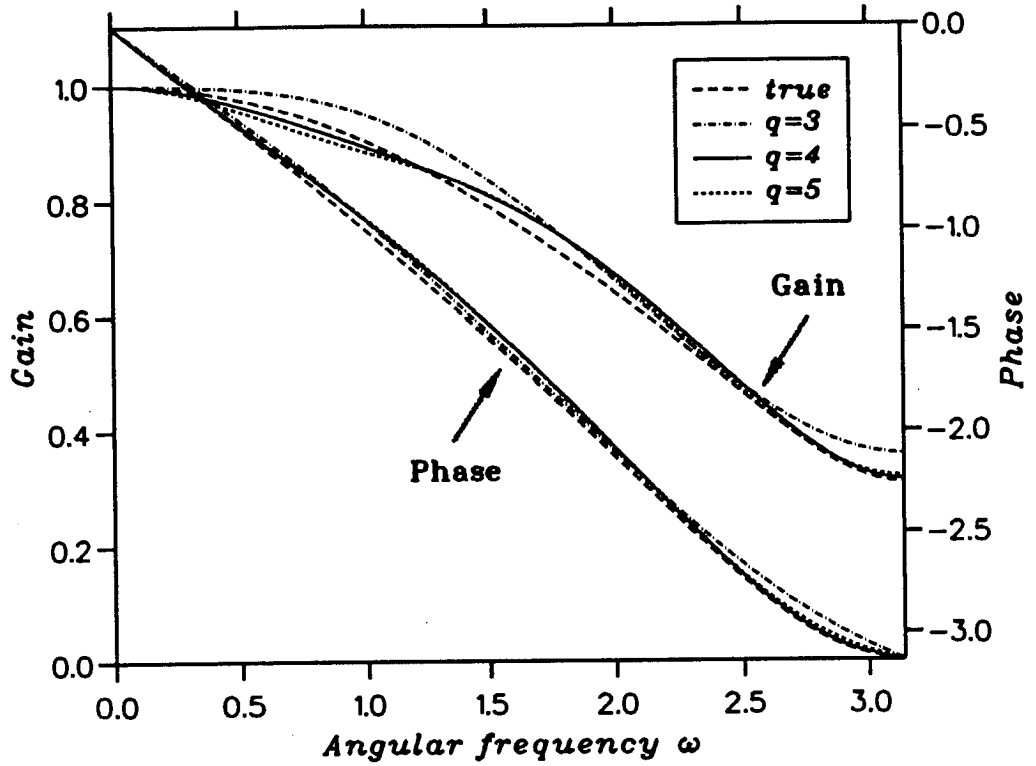
Fig(6.8) The frequency response in Example 6.2.1.



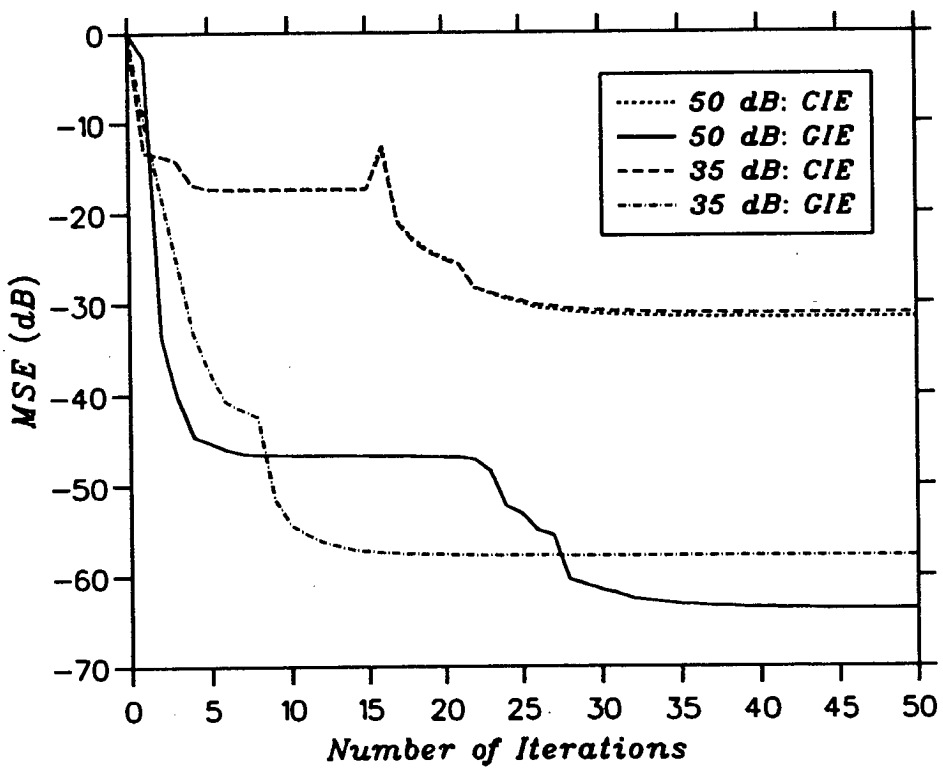
Fig(6.9) The curve of S(J) in Example 6.2.2.



Fig(6.10) The learning curves in Example 6.2.2.



Fig(6.11) The frequency response in Example 6.2.2.



Fig(6.12) The learning curves in Example 6.2.2: Different initial estimates.

6.3 AR Model Based Technique

In addition to MA model based methods, another basic class of parametric techniques are AR model based ones, which are often more preferable in the equalisation scenario. We derived in Chapter 5 six different families of linear equations with respect to the coefficients of inverse filters (or, equivalently, the AR parameters of channels), each family of linear equations comprising a single deconvolution algorithm. As addressed there, the proposed algorithms have two important advantages over the existing approaches: *i*) only the diagonal slice of cumulants is used; and *ii*) the algorithm form is much simpler. In this section, the adaptive performance of the algorithms will be examined in the context of blind equalisation of NMP channels.

The general structure of this section is as follows: In Subsection 6.3.1, the blind equalisation problem is re-formulated for convenience of the following description. Subsection 6.3.2 derives the new AR model based equalisation algorithm, where an iteration scheme termed successive over-relaxation (SOR) is introduced to accelerate the algorithm convergence rate. Some analysis and discussions on the algorithm are given in Subsection 6.3.3. Then, the comparison of SOR with the conventional LMS scheme is made in Subsection 6.3.4. Subsection 6.3.5 mathematically analyses the convergence behaviour of our algorithm. Several computer simulations are shown in Subsection 6.3.6. Finally, some conclusions are drawn in Subsection 6.3.7.

6.3.1 Preliminaries

The channel studied in this paper is the same as before, *i.e.*, it is assumed to be stable and linear, and can be expressed as the following MA model with finite order q:

$$y(k) = y_{nf}(k) + n(k) = \sum_{i=0}^{q} b_i x(k-i) + n(k) \quad , \tag{6.3.1}$$

where, $n(k)$ is white Gaussian noise, and $x(k)$ is the transmitted sequence in which the random data are independently and uniformly distributed over some discrete finite set: $S=\{\pm 1, \pm 3, \ldots, \pm(2M-1)\}$, which corresponds to a $2M$-level PAM scheme. Also, we assume that the transfer function of the above channel has no zeros on the unit circle. As defined before, the object of blind equalisation is to restore $x(k)$, given only $y(k)$, by determining an inverse filter $\{\theta_l\}$ which permits Eq(6.3.2) shown below hold.

$$\hat{x}(k) = \sum_{l=r_1}^{r_2} \theta_l y(k-l) \approx x(k) \quad . \tag{6.3.2}$$

Here, r_1 is the order of the noncausal part of the inverse filter, and r_2 that of the causal part. In fact, when $r_1=0$, Eq(6.3.2) represents the case of MP channels, where the well known prediction technique can be used to realise the blind equalisation [19][106]. On the other hand, when $r_2=-1$, Eq(6.3.2) models the maximum phase channels, for which the prediction method can also be employed [19]. However, for mixed phased channels (*i. e.*, $r_1<0$ and $r_2\geq 0$), no correct blind equalisation can be achieved by using the prediction scheme. In this case, as explained earlier, only higher-order cumulant (HOC) based techniques can offer a satisfactory solution without any other requirements for the data distribution except that it be non-Gaussian.

Similarly (to last section), as a result of symmetric distribution (or unskewedness) of the above PAM data, the 3rd-order cumulants of $x(k)$, hence that of $y(k)$, theoretically vanish. Thus, it is necessary (assume that no SAT is introduced here) to use the 4th-order cumulants of $y(k)$, which, mathematically, may be expressed as (see Section 2.6)

$$C_4(m_1, m_2, m_3) = \mathrm{E}[y(k)\prod_{i=1}^{3} y(k+m_i)]$$

$$-C_2(m_1)C_2(m_3-m_2)-C_2(m_2)C_2(m_3-m_1)-C_2(m_3)C_2(m_2-m_1) \quad , \tag{6.3.3}$$

where, $C_2(\cdot)$ denotes the autocorrelation function of $y(k)$, *viz.*,

$$C_2(m) = \mathrm{E}[y(k)y(k+m)] \quad . \tag{6.3.4}$$

In this section, we only use the diagonal slice of $C_4(m_1, m_2, m_3)$, where, $m_1=m_2=m_3$, and we denote it as

$$c_4(m) = C(m_1, m_2, m_3)|_{m_1=m_2=m_3=m} \quad . \tag{6.3.5}$$

For the autocorrelation function, the diagonal slice $c_2(\cdot)$ of $C_2(\cdot)$ reduces to $C_2(\cdot)$ itself, *i. e.*,

$$c_2(m)=C_2(m) \quad . \tag{6.3.6}$$

Thus, we have

$$c_4(m) = \mathrm{E}[y(k)y^3(k+m)]-3\mathrm{E}[y^2(k)]\mathrm{E}[y(k)y(k+m)] \quad . \tag{6.3.7}$$

On the other hand, it should be noticed that, determination of inverse filter $\{\theta_l\}$ is equivalent to modelling of channels by a noncausal AR structure, *viz.*,

$$H(z) = \sum_{i=0}^{q} b_i z^{-i} \approx \frac{1}{\sum_{l=r_1}^{r_2} \theta_l z^{-l}} \quad . \tag{6.3.8}$$

Consequently, the determination of the inverse filter $\{\theta_l\}$ can also be considered as a process of channel identification (AR model parameter identification).

In this section, a new 2nd- and 4th-order cumulant based SOR algorithm is used to adapt θ_l's and realise the corresponding blind equalisation procedure. SOR is a particular variety of the conventional least-mean-square (LMS) adaptive scheme, but possesses a faster convergence rate and a greater insensitivity to the different values of "step-size", termed "relaxation factor" here.

6.3.2 Algorithm derivation

In order to determine the inverse filter parameters θ_l's in Eq(6.3.2) or Eq(6.3.8), the following set of 2nd- and 4th-order cumulant based equations was obtained in Chapter 5 under the conditions defined above (regarding the other two related sets of equations in Chapter 5, they can be similarly processed):

$$\begin{cases} \sum_{\substack{j=r_1 \\ j\neq 0}}^{r_2} \theta_j c_4(n+j) = -c_4(n) \; ; \quad \text{for } n \in [-r_2, -1] \bigcup [q+1, -r_1] \\ \sum_{\substack{j=r_1 \\ j\neq 0}}^{r_2} \theta_j f_{nj} = g_n \; ; \quad \text{for } n \in [0, q] \quad , \end{cases} \tag{6.3.9}$$

where,

$$f_{nj} = \sum_{i=0}^{q}[c_2(n-i)c_4(i+j)-c_4(n-i)c_2(i+j)] \tag{6.3.10a}$$

and

$$g_n = \sum_{i=0}^{q}[c_4(n-i)c_2(i)-c_2(n-i)c_4(i)] \ . \tag{6.3.10b}$$

The difference between θ_j's in Eq(6.3.9) and θ_i's in Eq(6.3.2) is that normalisation has been done in Eq(6.3.9) so that $\theta_j|_{j=0}=1$. Eq(6.3.9) can be rewritten as following more compact form:

$$\mathbf{U}\Psi = \mathbf{V} \ , \tag{3.11}$$

where,

$$\Psi = (\psi_i)_{(r_2-r_1)}$$

$$= [\theta_{r_1}, \theta_{r_1+1}, \ldots, \theta_{-1}, \theta_1, \ldots, \theta_{r_2-1}, \theta_{r_2}]^T \ , \tag{6.3.12a}$$

$$\mathbf{U} = (u_{ij})_{(r_2-r_1+1)\times(r_2-r_1)}$$

$$= \begin{bmatrix} c_4(-r_2+r_1) & \cdots & c_4(-r_2-1) & c_4(-r_2+1) & \cdots & c_4(0) \\ \vdots & & \vdots & \vdots & & \vdots \\ c_4(-1+r_1) & \cdots & c_4(-2) & c_4(0) & \cdots & c_4(-1+r_2) \\ f_{0,r_1} & \cdots & f_{0,-1} & f_{0,1} & \cdots & f_{0,r_2} \\ \vdots & & \vdots & \vdots & & \vdots \\ f_{q,r_1} & \cdots & f_{q,-1} & f_{q,1} & \cdots & f_{q,r_2} \\ c_4(q+1+r_1) & \cdots & c_4(q) & c_4(q+2) & \cdots & c_4(q+1+r_2) \\ \vdots & & \vdots & \vdots & & \vdots \\ c_4(0) & \cdots & c_4(-r_1-1) & c_4(-r_1+1) & \cdots & c_4(r_2-r_1) \end{bmatrix} , \tag{6.3.12b}$$

and

$$\mathbf{V} = (v_i)_{(r_2-r_1+1)}$$

$$= [-c_4(-r_2), \ldots, -c_4(-1), g_0, \ldots, g_q, -c_4(q+1), \ldots, -c_4(-r_1)]^T \ . \tag{6.3.12c}$$

In practice, $c_2(\cdot)$ and $c_4(\cdot)$ will have to be replaced with their estimates $\hat{c}_2(\cdot)$ and $\hat{c}_4(\cdot)$, respectively. But we will not explicitly indicate the "hats" in the following parts of this section to aid the conciseness of description because it is easy to distinguish. From Eq(6.3.11), the following normal equation can be obtained:

$$\mathbf{U}^T\mathbf{U}\Psi = \mathbf{U}^T\mathbf{V} \ . \tag{6.3.13}$$

Since $\mathbf{U}^T\mathbf{U}$ is positive definite, we must have

$$det(\mathbf{U}^T\mathbf{U}) \neq 0 \ . \tag{6.3.14}$$

From Eq(6.3.14) and the relevant linear equation theory, the solution of Eq(6.3.13), then the least squares solution of Eq(6.3.11), exists uniquely, *viz.*,

$$\Psi = (\mathbf{U}^T\mathbf{U})^{-1}(\mathbf{U}^T\mathbf{V}) \ . \tag{6.3.15}$$

Eq(6.3.15) is the block form of our algorithm. The elements of **U** and **V** can be recursively estimated at each sample (or time step) by using the following expressions:

$$c_2^{(k+1)}(m) = \frac{1}{k+2}\sum_{i=0}^{k-m+1} x(i)x(i+m)$$

$$= \frac{1}{k+2}[(k+1)c_2^{(k)}(m)+x(k-m+1)x(k+1)]$$

$$= c_2^{(k)}(m)(1-\frac{1}{k+2})+\frac{x(k-m+1)x(k+1)}{k+2} \quad , \quad \text{for } 0\le m\le q; \qquad (6.3.16a)$$

$$c_2^{(k+1)}(m)=c_2^{(k+1)}(-m) \quad , \quad \text{for } -q\le m<0; \qquad (6.3.16b)$$

$$c_2^{(k+1)}(m) = 0 \quad , \quad \text{other}, \qquad (6.3.16c)$$

and

$$c_4^{(k+1)}(m) = M_4^{(k+1)}(m)-3c_2^{(k+1)}(0)c_2^{(k+1)}(m) \quad , \qquad (6.3.17)$$

where, $M_4^{(k+1)}(m)$ is the diagonal slice of the 4th-order moments at the $(k+1)$th time step, and it can be estimated as follows:

$$M_4^{(k+1)}(m) = \frac{1}{k+2}\sum_{i=0}^{k-m+1} x(i)x^3(i+m)$$

$$= \frac{1}{k+2}[(k+1)M_4^{(k)}(m) + x(k-m+1)x^3(k+1)]$$

$$= M_4^{(k)}(m)(1-\frac{1}{k+2}) + \frac{x(k-m+1)x^3(k+1)}{k+2} \quad , \quad \text{for } 0\le m\le q; \qquad (6.3.18a)$$

$$M_4^{(k+1)}(m) = \frac{1}{k+2}\sum_{i=-m}^{k+1} x(i)x^3(i+m)$$

$$= M_4^{(k)}(m)(1-\frac{1}{k+2}) + \frac{x^3(k+m+1)x(k+1)}{k+2} \quad , \quad \text{for } -q\le m<0; \qquad (6.3.18b)$$

$$M_4^{(k+1)}(m) = 0 \quad , \quad \text{other}. \qquad (6.3.18c)$$

Thus, Eq(6.3.15) also provides, in principle, a solution to the adaptive inverse filter parameters with the above recursive estimation of cumulants. The direct calculation of the r.h.s. of Eq(6.3.15) at each sample, however, as in autocorrelation function based adaptive filtering methods, involves a considerable amount of computation. In order to make full use of the information contained in the current set of parameters, the two most widely used schemes which have been proposed to solve such kind of problems in autocorrelation function domain are the recursive least-squares (RLS) and least-mean-squares (LMS) algorithms [109]. In our problem, however, the RLS type of scheme is not applicable. This is because the size and structure of both **U** and **V** are "time-invariant" here, *i. e.*, they do not change with the increase of time k. On the other hand, the ordinary LMS type of scheme can be borrowed and used in our problem. From Eq(6.3.11), we can define

$$\epsilon^2(k) = [\mathbf{U}(k)\Psi(k)-\mathbf{V}(k)]^T[\mathbf{U}(k)\Psi(k)-\mathbf{V}(k)] \qquad (6.3.19)$$

as the cost function at kth iteration, where $\mathbf{U}(k)$, $\mathbf{V}(k)$, and $\Psi(k)$ respectively denote the estimates of **U**, **V** and Ψ at the kth iteration. Consequently, we can obtain the following

update to $\Psi(k)$:

$$\Psi(k+1) = \Psi(k)+2\cdot\tau(k)\cdot\nabla(k) \quad , \tag{6.3.20a}$$

and

$$\nabla(k) = \mathbf{U}^T(k)[\mathbf{V}(k)-\mathbf{U}(k)\Psi(k)] \quad , \tag{6.3.20b}$$

where, $\tau(k)$ is generally taken from the range:

$$0 < \tau(k) < \frac{1}{tr[\mathbf{U}^T(k)\mathbf{U}(k)]} \quad . \tag{6.3.20c}$$

However, the convergence rate of the above scheme is generally slow, as in the autocorrelation function domain [14][109]. In addition, the convergence rate is greatly dependent on choosing of $\tau(k)$. In order to make best use of the information contained in the current set of parameters and obtain a faster convergence rate (without increasing the computation penalty too much), the successive overrelaxation (SOR) iterative scheme was adopted. The block digram of our equalisation procedure is illustrated in Fig(6.13).

The SOR scheme was introduced simultaneously by D. M. Young and S. P. Frankel in 1950 [110][111]. In fact, it is an accelerated scheme of Gauss-Seidel iterative method for solving the large scale sets of linear equations. Like the Gauss-Seidel method, determination of the inverse of matrix is not required in the SOR method. But the SOR method can achieve a faster convergence rate. Now, we apply SOR scheme to Eq(6.3.13), and the following iterative algorithm can be obtained.

Let

$$\mathbf{U}^T(k)\mathbf{U}(k) = \mathbf{A}(k) = (a_{ij}^{(k)})_{(r_2-r_1)\times(r_2-r_1)} \quad , \tag{6.3.21a}$$

and

$$\mathbf{U}^T(k)\mathbf{V}(k) = \mathbf{D}(k) = (d_i^{(k)})_{(r_2-r_1)} \quad . \tag{6.3.21b}$$

Then, $\Psi(k)$ can be updated by equation

$$\psi_i^{(k+1)} = \psi_i^{(k)} + \delta\psi_i^{(k)} \quad , \tag{6.3.22a}$$

where,

$$\delta\psi_i^{(k)} = \frac{\mu}{a_{ii}^{(k)}}\left(d_i^{(k)}-\sum_{j=1}^{i-1} a_{ij}^{(k)}\psi_j^{(k+1)}-\sum_{j=i}^{r_2-r_1} a_{ij}^{(k)}\psi_j^{(k)}\right) \quad , \tag{6.3.22b}$$

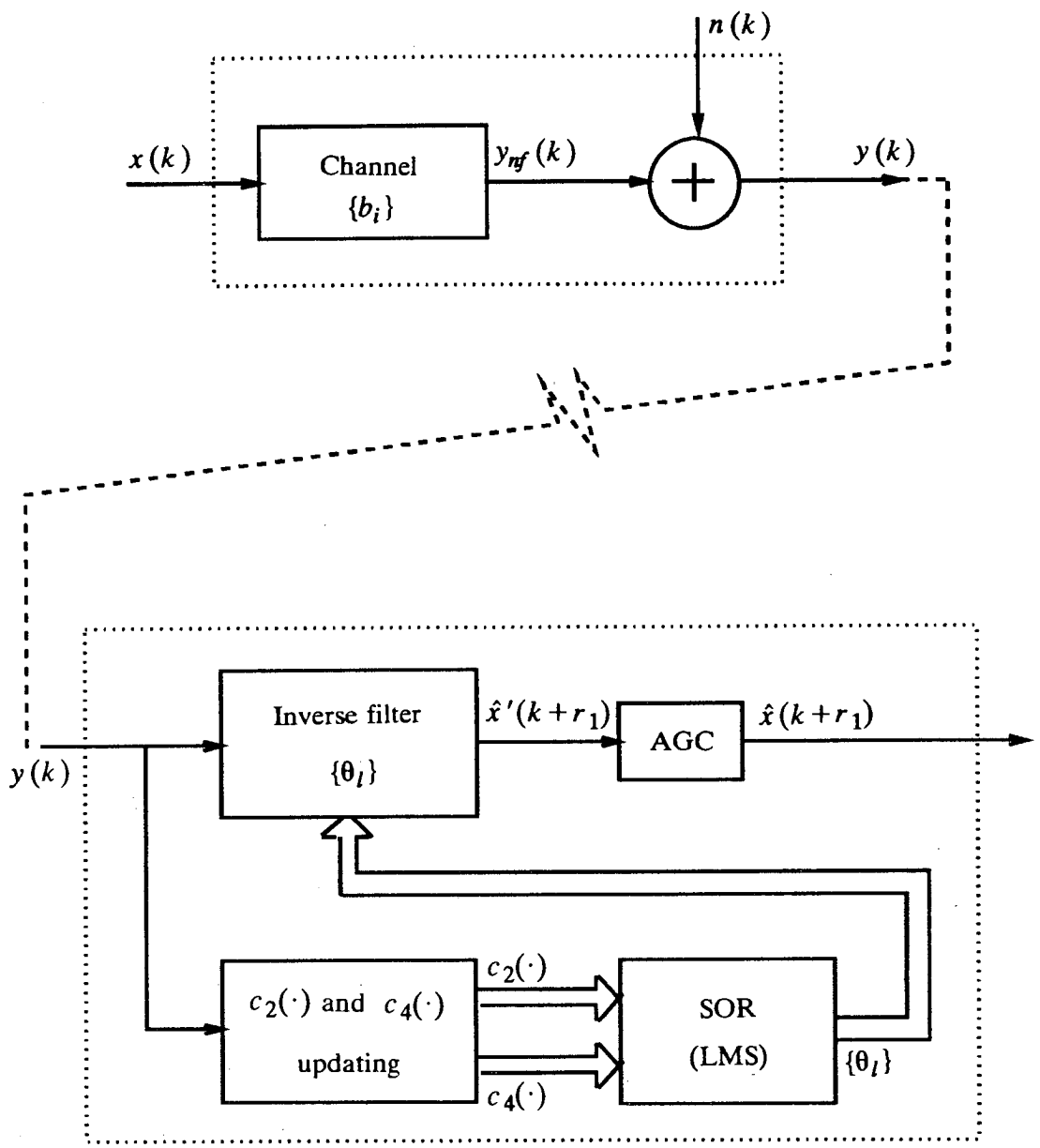
$$i=1, 2, \ldots, r_2-r_1,$$

and μ is termed *relaxation factor*, which, to guarantee the convergence of the above iterative process, must be chosen from the following range (a concise proof for this is given in Appendix 6.2. For more details, see [110] and [111]):

$$0 < \mu < 2 \quad . \tag{6.3.22c}$$

When $\mu=1$, SOR reduces to Gauss-Seidel method. Furthermore, Eq (6.3.22b) can be rewritten as

$$\delta\psi_i^{(k)} = \frac{\mu}{\sum_{l=1}^{r_2-r_1+1}(u_{li}^{(k)})^2}\cdot\sum_{l=1}^{r_2-r_1+1} u_{li}^{(k)}\cdot\left(v_l^{(k)}-\sum_{j=1}^{i-1} u_{lj}^{(k)}\psi_j^{(k+1)}-\sum_{j=i}^{r_2-r_1} u_{lj}^{(k)}\psi_j^{(k)}\right) \quad , \tag{6.3.22d}$$



Fig(6.13) The blind equaliser structure in Section 6.3.

where $u_{ij}^{(k)}$ and $v_i^{(k)}$ denote the values of u_{ij} and v_i after the kth iteration, respectively.

In practical programming, further arrangements can be made to reduce the computational load. For example, we can adopt the following algorithmic form:

For the $(k+1)$th iteration (or sample),

1) Filling $\mathbf{U}(k)$ and $\mathbf{V}(k)$ by using Eq(6.3.16a-c) and Eq(6.3.17-18c);

2) Calculate

$$a_{ii} = \sum_{l=1}^{r_2-r_1+1} (u_{il}^{(k)})^2 \quad , \quad i=1, 2, \ldots, r_2-r_1 \quad ; \tag{6.3.23a}$$

3) For $i=1, 2, \ldots, r_2-r_1$,

if $i=1$, **then** calculate

$$\Delta_{l1}^{(k)} = v_l^{(k)} - \sum_{j=1}^{r_2-r_1} u_{lj}^{(k)} \psi_j^{(k)} \quad , \tag{6.3.23b}$$

$$l = 1, 2, \ldots, r_2-r_1+1 \quad ,$$

$$\delta\psi_1^{(k)} = \frac{\mu}{a_{11}^{(k)}} \sum_{l=1}^{r_2-r_1+1} u_{l1}^{(k)} \Delta_{l1}^{(k)} \quad , \tag{6.3.23c}$$

$$\psi_1^{(k+1)} = \psi_1^{(k)} + \delta\psi_1^{(k)} \quad , \tag{6.2.23d}$$

otherwise, calculate

$$\Delta_{li}^{(k)} = \Delta_{l,i-1}^{(k)} - \delta\psi_{i-1}^{(k)} u_{l,i-1}^{(k)} \quad , \tag{6.2.23e}$$

$$l = 1, 2, \ldots, r_2-r_1+1 \quad ,$$

$$\delta\psi_i^{(k)} = \frac{\mu}{a_{ii}^{(k)}} \sum_{l=1}^{r_2-r_1+1} u_{li}^{(k)} \Delta_{li}^{(k)} \quad , \tag{6.2.23f}$$

$$\psi_i^{(k+1)} = \psi_i^{(k)} + \delta\psi_i^{(k)} \quad . \tag{6.2.23g}$$

Notice that, in above step 3, the information of $\Delta_{l,i-1}^{(k)}$ is fully utilised to calculate $\Delta_{li}^{(k)}$. As a result, the computation amount is greatly reduced than otherwise.

6.3.3 Algorithm discussions

Several discussions for the above algorithm are then in order:

<1> Initial estimate of $\Psi(k)$: In principle, $\Psi(0)$ can be chosen arbitrarily. But habitually, it is set to be **0**.

<2> Since the obtained inverse filter is noncausal, the restored data series has a delay of $-r_1$ samples (or time steps), as shown in Fig(6.13), *viz*.,

$$\hat{x}'(k+r_1) = \sum_{j=r_1}^{r_2} \theta_j y(k+r_1-j) \quad . \tag{6.3.24}$$

<3> As a result of normalisation of θ_j's (or ψ_i's), the output data of the inverse filter $\{\theta_j\}$ has an amplitude gain in comparison with the true transmitted data at each time step; thus, an automatic gain control (AGC) mechanism [103][104][106] is needed, as in Fig(6.13). From Eq(6.3.24), the power of $\hat{x}'(k)$ at the kth time step can be expressed as

$$\sigma_{\hat{x}'}^2(k) = \mathrm{E}[(\hat{x}'(k))^2]$$

$$= \sum_{i=r_1}^{r_2}\sum_{j=r_1}^{r_2}\theta_i\theta_j \mathrm{E}[y(k-i)y(k-j)]$$

$$= \sum_{i=r_1}^{r_2}\sum_{j=r_1}^{r_2}\theta_i\theta_j c_2(i-j) \quad . \tag{6.3.25}$$

Notice that $\theta_0=1$ in Eq(6.3.24-25). Let the power of the transmitted data be σ_x^2, then the AGC factor can be chosen as

$$f_{AGC} = \pm|\sigma_x^2/\sigma_{\hat{x}'}^2|^{1/2} \quad . \tag{6.3.26}$$

Since the earlier estimates of $c_2(\cdot)$ in adaptive procedure may cause $\sigma_{\hat{x}'}^2<0$, the operation of taking the absolute values is added in Eq(6.3.26) to force the blind equalisation to continue. The sign (+ or −) here is not identifiable as in many existing methods [103][104].

<4> The above algorithm is based on the linear time-invariant (LTI) channels. If the considered channel is linear time-variant (LTV), Eq(6.3.16a-c) and Eq(6.3.18a-c) must be modified so that $c_2^{(k)}(\cdot)$ and $c_4^{(k)}(\cdot)$ can track the change of channels as quickly as possible.

For slowly time-variant channels, the following "exponentially weighted" estimations of $c_2(\cdot)$ and $c_4(\cdot)$ should be adopted:

$$c_2^{(k+1)}(m) = \frac{1}{k+2}\sum_{i=0}^{k-m+1} f^{k-m+1-i}\cdot x(i)x(i+m)$$

$$= f\cdot c_2^{(k)}(m)(1-\frac{1}{k+2}) + \frac{x(k-m+1)x(k+1)}{k+2} \quad , \quad \text{for } 0\le m\le q; \tag{6.3.27a}$$

$$c_2^{(k+1)}(m) = c_2^{(k+1)}(-m) \quad , \quad \text{for } -q\le m<0; \tag{6.3.27b}$$

$$c_2^{(k+1)}(m) = 0 \quad , \quad \text{other}, \tag{6.3.27c}$$

where, f is termed "forgetting factor", which controls the tracking performance of the algorithm. Similarly,

$$M_4^{(k+1)}(m) = \frac{1}{k+2}\sum_{i=0}^{k-m+1} f^{k-m+1-i}\cdot x(i)x^3(i+m)$$

$$= f\cdot M_4^{(k)}(m)(1-\frac{1}{k+2}) + \frac{x(k-m+1)x^3(k+1)}{k+2} \quad , \quad \text{for } 0\le m\le q; \tag{6.3.28a}$$

$$M_4^{(k+1)}(m) = \frac{1}{k+2}\sum_{i=-m}^{k+1} f^{k+1-i}\cdot x(i)x^3(i+m)$$

$$= f\cdot M_4^{(k)}(m)(1-\frac{1}{k+2}) + \frac{x^3(k+m+1)x(k+1)}{k+2} \quad , \quad \text{for } -q\le m<0; \tag{6.3.28b}$$

$$M_4^{(k+1)}(m) = 0 \quad , \quad \text{other}, \tag{6.3.28c}$$

and then,

$$c_4^{(k+1)}(m) = M_4^{(k+1)}(m) - 3c_2^{(k+1)}(0)c_2^{(k+1)}(m) \quad , \tag{6.3.17}$$

On the other hand, if there are sudden changes in channels, the following "sliding window" based estimations of $c_2(\cdot)$ and $c_4(\cdot)$ are recommended.

Let the length of the sliding window be W. When $k<W$, Eq(6.3.16a-c) and Eq(6.3.18a-c) should still be used; but when $k>W$, the equations below should be adopted:

$$c_2^{(k+1)}(m) = \frac{1}{W}[c_2^{(k)}(m)\cdot W - x(k+1-W)x(k-m+1-W) + x(k+1)x(k+1-m)] \quad , \quad \text{for } 0\le m\le q; \tag{6.3.29a}$$

$$c_2^{(k+1)}(m)=c_2^{(k+1)}(-m) \quad , \quad \text{for } -q\le m<0; \tag{6.3.29b}$$

$$c_2^{(k+1)}(m) = 0 \quad , \quad \text{other.} \tag{6.3.29c}$$

Similarly,

$$M_4^{(k+1)}(m) = \frac{1}{W}[M_4^{(k)}(m)\cdot W - x(k-m+1-W)x^3(k+1-W) + x(k-m+1)x^3(k+1)] \quad , \quad \text{for } 0\le m\le q; \tag{6.3.30a}$$

$$M_4^{(k+1)}(m) = \frac{1}{W}[M_4^{(k)}(m)\cdot W - x^3(k+m+1-W)x(k+1-W) + x^3(k+m+1)x(k+1)] \quad , \quad \text{for } -q\le m<0; \tag{6.3.30b}$$

$$M_4^{(k+1)}(m) = 0 \quad , \quad \text{other,} \tag{6.3.30c}$$

then again,

$$c_4^{(k+1)}(m) = M_4^{(k+1)}(m)-3c_2^{(k+1)}(0)c_2^{(k+1)}(m) \quad . \tag{6.3.17}$$

Note that, the last $W+q$ data points must be stored to update the estimation of cumulants in the above sliding window based scheme. As a result, θ_j's will be automatically updated to the corresponding new AR model parameters. This will be demonstrated in the simulation section.

<5> Choosing of relaxation factor μ: D. M. Young obtained an expression for the optimum relaxation factor [110][111]:

$$\mu_{opt} = \frac{2}{1+\sqrt{1-\rho^2(J)}} \quad , \tag{6.3.31}$$

where $\rho(J)$ is the spectral radius of the Jacobi iterative matrix J associated with Eq(6.3.13). Since determination of $\rho(J)$ is generally very complicated, Eq(6.3.31) cannot be used in practical environments. Empirical values can usually be found by trial-and-error. Fortunately, the disparity in the convergence rates caused by different relaxation factor values in our SOR algorithm is considerablely less than resulted from different step-size values in the LMS scheme.

In addition, it is well known that Eq(6.3.20c) only represents an ad-hoc basis for selection of step-size τ, and this criterion does not always work well. In fact, other criteria have been suggested recently [112][113]. However, this is not such a major problem for the above SOR algorithm since the selection range for relaxation factor μ is clear: $0<\mu<2$. The mathematical justification for this selection range is shown in Appendix 6.2. More details can be found in [110] or [111].

<6> Computational complexity: This is another important index for adaptive algorithms. In comparison with the corresponding LMS scheme, there is only a small increase in the computational amount as the penalty of the faster converging speed. This will be further analysed in later Subsection 6.3.4.

<7> In our derivation above, the channel order q is assumed to be known. If q is unknown, several HOC based methods are available to select it [97][98][14]. But a systematic and full comparison about the performances of these order selection methods has still to be reported in the literature. However, this problem is beyond the scope of this paper.

6.3.4 Comparison of SOR with LMS

In Subsection 6.3.2, the adaptive SOR algorithm was derived. In this section, the relation between SOR and LMS is illustrated, and the comparison between them is also made in terms of computational complexity.

Now, we can rewrite Eq(6.3.20a-c) in their component forms:

$$\psi_i^{(k+1)} = \psi_i^{(k)} + 2\tau(k)\nabla_i^{(k)} \quad , \tag{6.3.32a}$$

$$\nabla_i^{(k)} = \sum_{l=1}^{r_2-r_1+1} u_{li}^{(k)} \cdot (v_l^{(k)} - \sum_{j=1}^{r_2-r_1} u_{lj}^{(k)} \psi_j^{(k)}) \quad , \tag{6.3.32b}$$

and

$$0 < \tau(k) < \frac{1}{\sum_{i=1}^{r_2-r_1} \cdot \sum_{l=1}^{r_2-r_1+1} (u_{li}^{(k)})^2} \quad . \tag{6.3.32c}$$

Notice that, when we calculate $\psi_i^{(k+1)}$, the values for $\psi_0^{(k+1)}$, $\psi_1^{(k+1)}$, ..., $\psi_{i-2}^{(k+1)}$, and $\psi_{i-1}^{(k+1)}$ are already available, and the latter are the most recently obtained values and generally better (in terms of their closeness to the true solution) than the values for $\psi_0^{(k)}$, $\psi_1^{(k)}$, ..., $\psi_{i-2}^{(k)}$, and $\psi_{i-1}^{(k)}$. Thus, from the viewpoint of taking advantage of the most current information, we can change the ordinary form of LMS in Eq(6.3.32a-b) into the following accelerated form (alternatively termed Gauss-Seidel form):

$$\psi_i^{(k+1)} = \psi_i^{(k)} + 2\tau_i(k)\nabla_i^{(k)} \quad , \tag{6.3.33a}$$

$$\nabla_i^{(k)} = \sum_{l=1}^{r_2-r_1+1} u_{li}^{(k)} \cdot (v_l^{(k)} - \sum_{j=1}^{i-1} u_{lj}^{(k)} \psi_j^{(k+1)} - \sum_{j=i}^{r_2-r_1} u_{lj}^{(k)} \psi_j^{(k)}) \quad . \tag{6.3.33b}$$

Comparison of Eq(6.3.33a-b) with Eq(6.3.22a) and Eq(6.3.22d) immediately leads to the conclusion that they are equivalent! Thus, the SOR scheme is actually an accelerated form of the ordinary LMS scheme. Considering the range (0, 2) of μ, the above comparison also gives the new range for $\tau(k)$ in Eq(6.3.33a):

$$0 < \tau_i(k) < \frac{1}{\sum_{l=1}^{r_2-r_1+1} (u_{li}^{(k)})^2} \quad . \tag{6.3.33c}$$

Notice that the range of $\tau_i(k)$'s in the new form of LMS scheme is greatly augmented.

On the other hand, it can be seen from the following discussion that the amount of computation for SOR becomes slightly larger than that for LMS. Since the computational amounts involved in cumulant estimation are the same for both SOR and LMS, we only need to compare the computational complexities of Eq(6.3.23a-g) and Eq(6.3.32a-c). In Eq(6.3.32a-c), the number of multiplications and divisions (MD) for one iteration is approximately $3(r_2-r_1)(r_2-r_1+1)+(r_2-r_1)$; In Eq(6.3.23a-g) (programming implementation of Eq(6.3.22a-b)), the corresponding number of MD becomes $2(r_2-r_1+1)(r_2-r_1)+2(r_2-r_1+1)(r_2-r_1-1)+(r_2-r_1+1)+(r_2-r_1)=4(r_2-r_1+1)(r_2-r_1)-1$. The numbers of additions and subtractions (AS) are very similar to MD. Hence, the SOR scheme approximately requires $(r_2-r_1+1)(r_2-r_1-1)=(r_2-r_1)^2-1$ more MD and the similar number more of AS. Table (6.5) shows this point further by several concrete examples. Apparently, the increase of computation load is normally not very large. In addition, This can be considered as the penalty to be paid for the faster convergence rate.

Table (6.5) Computational Complexity Comparison between SOR and LMS (MD & AS number).

r_1	r_2	SOR: Eq(6.3.23a-g)	LMS: Eq(6.3.32a-c)	Increase
-5	5	439	340	99
-10	10	1679	1280	399
-15	10	2599	1975	624
-15	15	3719	2820	899
-20	20	6559	4960	1599

6.3.5 Consistency and convergence

In this section, the consistency and convergence behaviour of our algorithm are analysed. It is shown that the algorithm is normally consistently convergent.

First, let us notice that the assumptions in Subsection 6.3.1 imply that:

1) For the transmitted data series $x(k)$, $E[x(k)]=0$, and the diagonal slices of the ith order cumulants of $x(k)$ are limitedly large, where $i=2, 3, 4, 5, 6$; and

2) For the channels: They are absolutely summable, *viz.*, $\sum_{i=0}^{q}|b_i|<\infty$. Then, according to the relevant theorem in [15], we can conclude that the estimators both in Eq(6.3.16) and in Eq(6.3.17) are unbiased and consistent. Also, it is assumed that the affect of additive noise is removed. Considering the nonsingularity of $\mathbf{U}^T(k)\mathbf{U}(k)$ and the results in [14] and [114], we have

$$\lim_{k\to\infty}\Psi(k) = \lim_{k\to\infty}(\mathbf{U}^T(k)\mathbf{U}(k))^{-1}(\mathbf{U}^T(k)\mathbf{V}(k)) = \Psi \qquad (6.3.34)$$

Thus, the estimator in Eq(6.3.15), then that in Eq(6.3.13), is unbiased and consistent.

Now, the remaining problem is: whether the SOR scheme converges? and if yes, under what conditions? To answer these two questions, we only need to cite the following

theorem:

[THEOREM] For a set of linear equations

$$\mathbf{A}\Psi = D \quad , \tag{6.3.35}$$

if **A** is symmetric and positive definite, and the relaxation factor is in the range: $0<\mu<2$, then SOR scheme converges •

The rigorous proof for the above theorem can be found in [111]. A concise and self-contained proof is also given in Appendix 6.2.

Obviously, our algorithm completely satisfies the conditions of the above theorem. Therefore, the convergence of our algorithm is guaranteed.

6.3.6 Simulation examples

In this subsection, several simulation examples are presented to demonstrate the performance of the algorithm proposed above. In the description below, the signal-to-noise ratio (SNR) is defined the same as before.

6.3.6.1 LTI channels

[Example 6.3.1] The mixed phase channel transfer function is assumed to be

$$H(z) = 1-2.3z^{-1}+0.6z^{-2} \quad . \tag{6.3.36}$$

The transmitted data is a 4-level PAM series: equally distributed over $\{\pm 1, \pm 3\}$, and the SNR is taken to be 40dB. The above SOR-based algorithm is applied to the received data series $y(k)$, where we set $r_1=-10$, $r_2=10$, and $\mu=1.3$. The mean-square-error (MSE) curve is drawn in Fig(6.14a), and the error rate (ER) curve in Fig(6.14b). For the convenience of comparison, the simulation of the LMS-based version of our algorithm (*i.e.*, Eq(6.3.20a-c)) is also presented, and the corresponding results are also illustrated in Fig(6.14a) and Fig(6.14b), where we similarly set $r_1=-10$ and $r_2=10$, but $\tau(k)=0.5/tr[\mathbf{U}^T(k)\mathbf{U}(k)]$. Notice that $\mu(k)$ has been chosen to make the LMS algorithm achieve the approximately fastest converging speed. Clearly, SOR-based algorithm is of a far faster convergence rate than the LMS-based one. But the MSE curve of the SOR algorithm is not as smooth as that of the LMS algorithm.

[Example 6.3.2] Let the channel transfer function be

$$H(z) = 1+4.5z^{-1}+2z^{-2} \quad . \tag{6.3.37}$$

Here, we set $\mu=1.5$ and $\tau(k)=0.9/tr[\mathbf{U}^T(k)\mathbf{U}(k)]$. The other assumptions are the same as Example 6.3.1. In the similar way to Example 6.3.1, the obtained results are illustrated in Fig(15a) and Fig(15b), respectively. This example demonstrates further the performance of the SOR algorithm much more.

[Example 6.3.3] The channel is of the transfer function (from [14])

$$H(z) = -0.7+1.63z^{-1}-0.9z^{-2} \quad . \tag{6.3.38}$$

As in [14], we set the following conditions: $r_1=-10$, $r_2=30$, $SNR=20dB$, and the transmitted data is a 2-level PAM series. Then, SOR algorithm is used exactly as in Example 6.3.1 and 6.3.2. When μ is let to be 1.2, the convergence traces of θ_j,

$j = \pm 1, \pm 2, \pm 3$, are drawn in Fig(6.16a), and the corresponding MSE curve in Fig(6.16b). In comparison, the results here are at least as good as that in [14], although the form of the algorithm here is very simple and regular. Another advantage of the algorithm here is that its convergence rate is much less sensitive to μ than the LMS scheme to τ. To see this point, we draw the MSE curves of $\mu=0.8$ and $\mu=1.8$ in Fig(6.16c). Clearly, there is no vital difference between them.

[Example 6.3.4] The channel transfer function is

$$H(z) = 1+3z^{-1}+2z^{-2}+z^{-3} \ . \tag{6.3.39}$$

Let $r_1=-10$, $r_2=10$ and $SNR=30$dB. The transmitted data is the same as in Example 6.3.3: 2-level PAM series. The SOR algorithm is implemented using different relaxation factor values: $\mu=0.5$, 1.0, 1.5, and 1.9. The corresponding MSE and ER curves are shown in Fig(17a) and Fig(17b), respectively. The robustness of our algorithm to μ can be seen again.

6.3.6.2 LTV Channels

[Example 6.3.5] The slowly variant channel with the transfer function [14]

$$H(z) = b_0+b_1z^{-1}+b_2z^{-2} \ , \tag{6.3.40}$$

where,

$$b_0 = -0.7+0.000025(k-1) \ ; \tag{6.3.41a}$$

$$b_2 = -0.3-0.000025(k-1) \ ; \tag{6.3.41b}$$

and

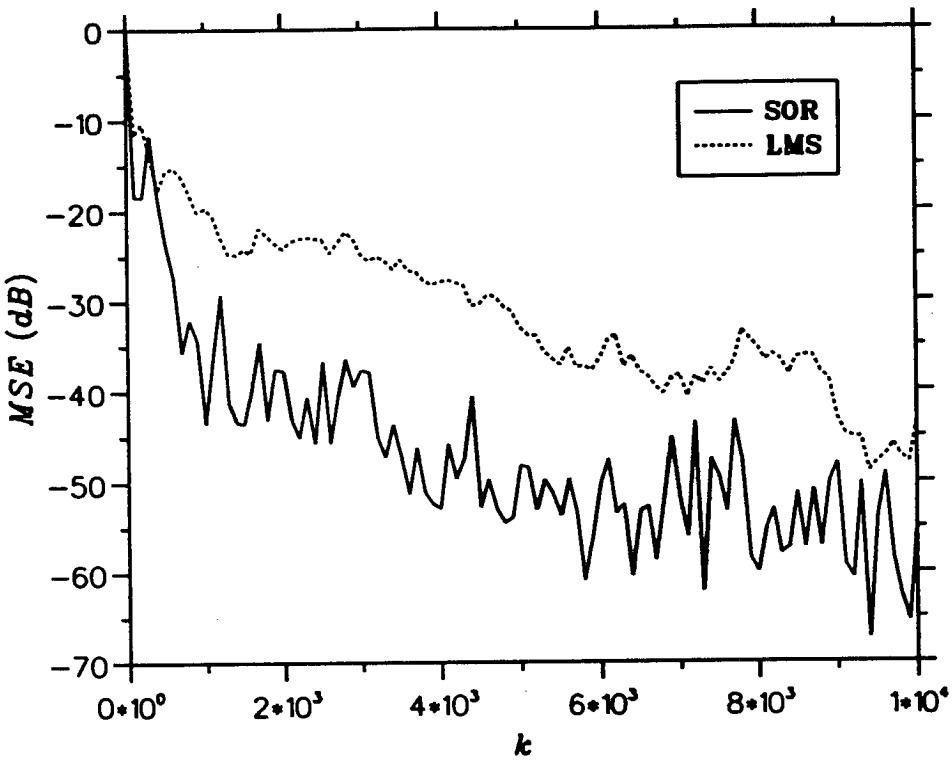
$$b_1 = 1+b_0b_2 \ . \tag{6.3.41c}$$

The transmitted data is a 2-level PAM series as in Example 6.3.3 and 6.3.4. First, set $SNR=30dB$, $r_1=-10$, $r_2=15$, and $\mu=1.4$. Then we apply the algorithm to the received data, where, Eq(6.3.27a-c) and Eq(6.3.28a-c) are used, and the forgetting factor f is taken to be 0.9999. The obtained converging traces of θ_j, $j=\pm 1, \pm 2, \pm 3$, are drawn in Fig(6.18a), and the MSE curve in Fig(6.18b). It is clear that the algorithm tracks the change of the channel very satisfactorily.

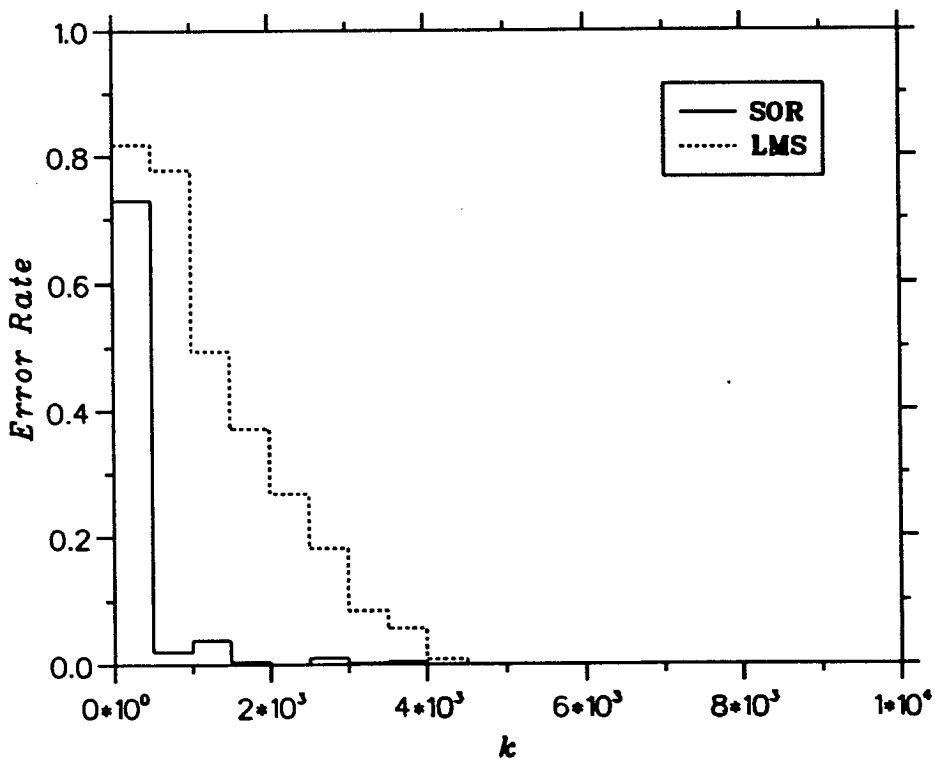
[Example 6.3.6] The sudden changed channel: The transfer function [14] is

$$H(z) = \begin{cases} -0.7+1.21z^{-1}-0.3z^{-2} \ , & 0 \le k < 4000; \\ -0.5+1.25z^{-1}-0.5z^{-2} \ , & k \ge 4000. \end{cases} \tag{6.3.42}$$

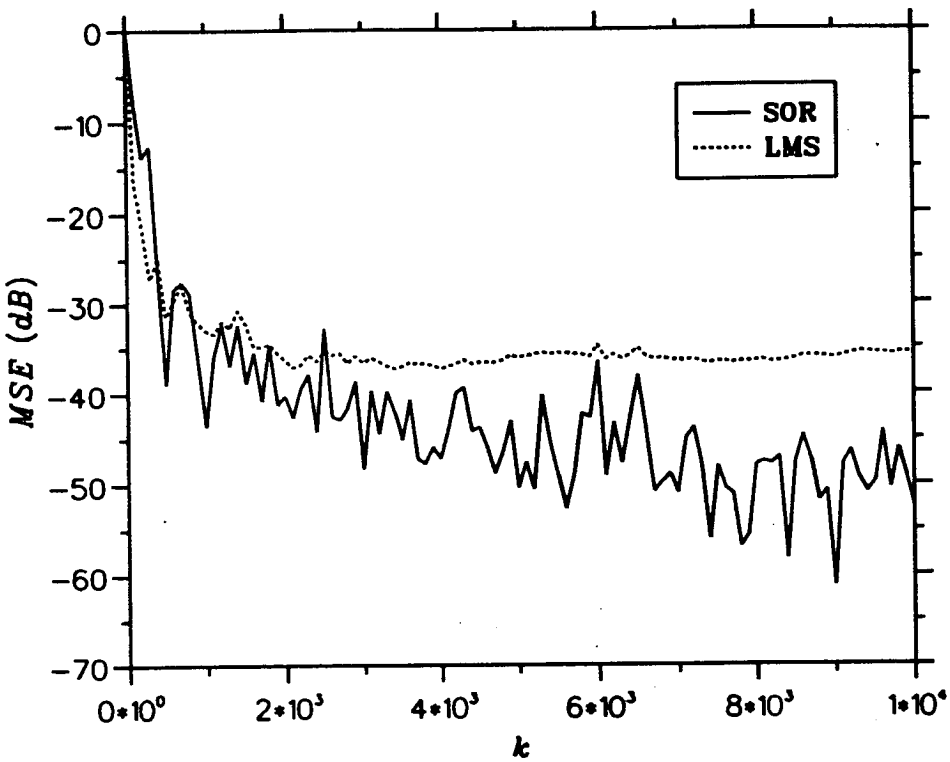
The transmitted data is also a 2-level PAM series. Again, set $SNR=30dB$, $r_1=-10$, $r_2=15$, and $\mu=1.1$. Eq(6.3.29a-c) and Eq(6.3.30a-c) are used to update the cumulants, where the window length is set to be 2000. The MSE curve is illustrated in Fig(6.3.19a), and the converging traces of θ_j, $j=\pm 1, \pm 2, \pm 3$, are plotted in Fig(6.3.19b). It can be seen that the change of the channel is closely tracked, and the deterioration caused by the sudden channel change is quickly eliminated.



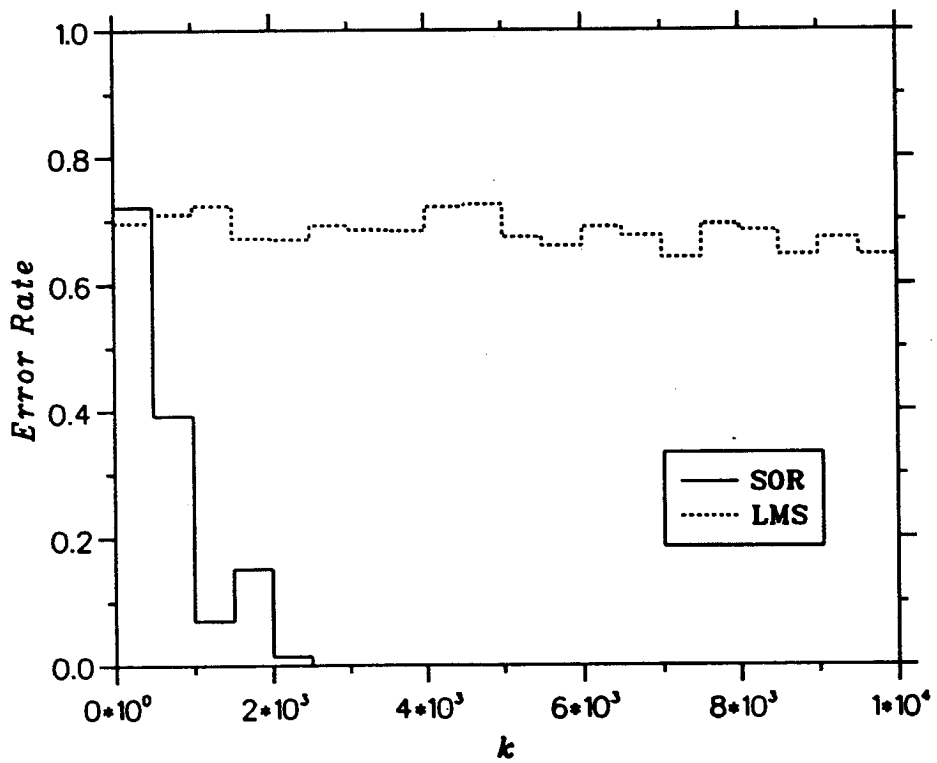
Fig(6.14a) The MSE curves of SOR and LMS algorithms in Example 6.3.1.



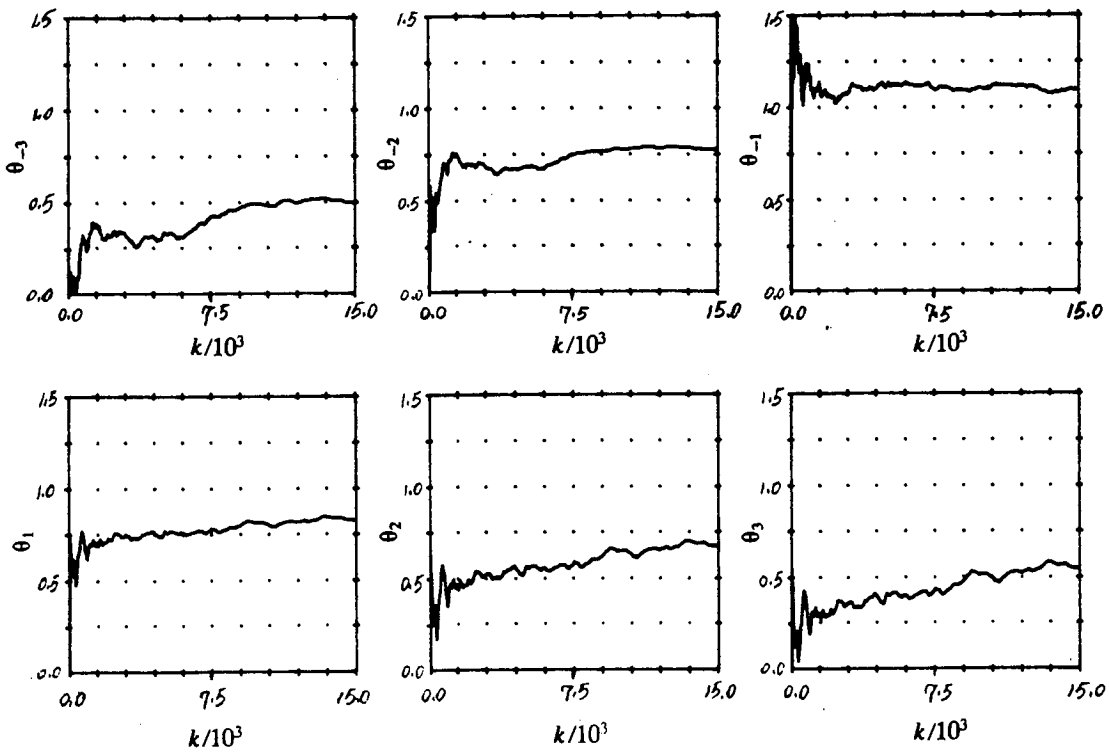
Fig(6.14b) The error rates (ER) of SOR and LMS algorithms in Example 6.3.1.



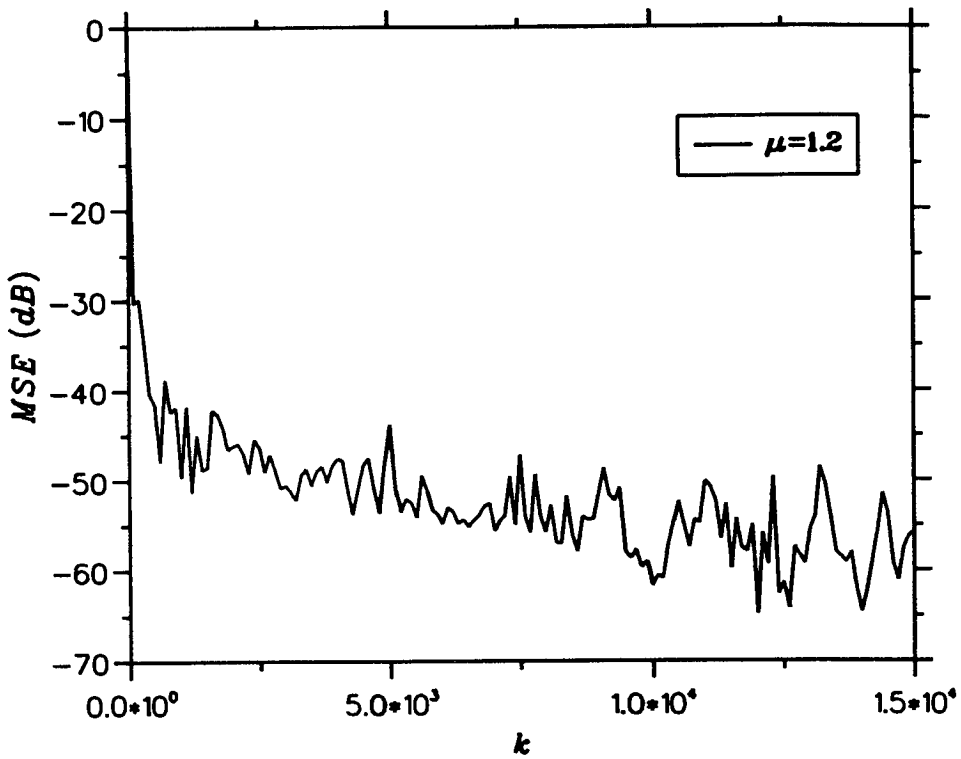
Fig(6.15a) The MSE curves of SOR and LMS algorithms in Example 6.3.2.



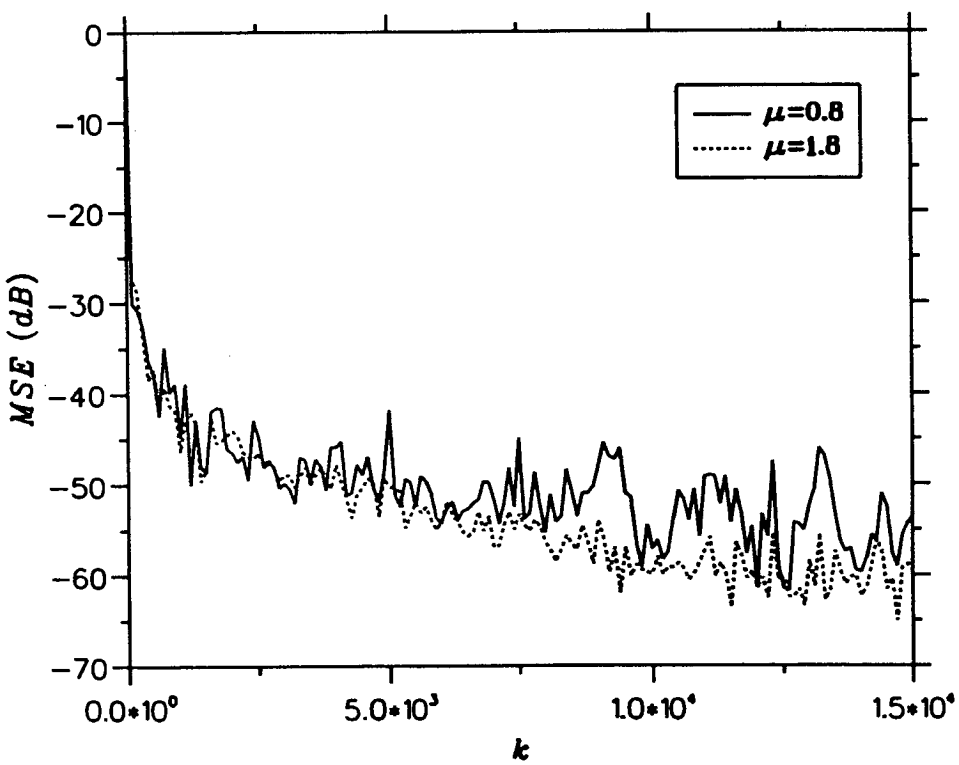
Fig(6.15b) The error rates (ER) of SOR and LMS algorithms in Example 6.3.2.



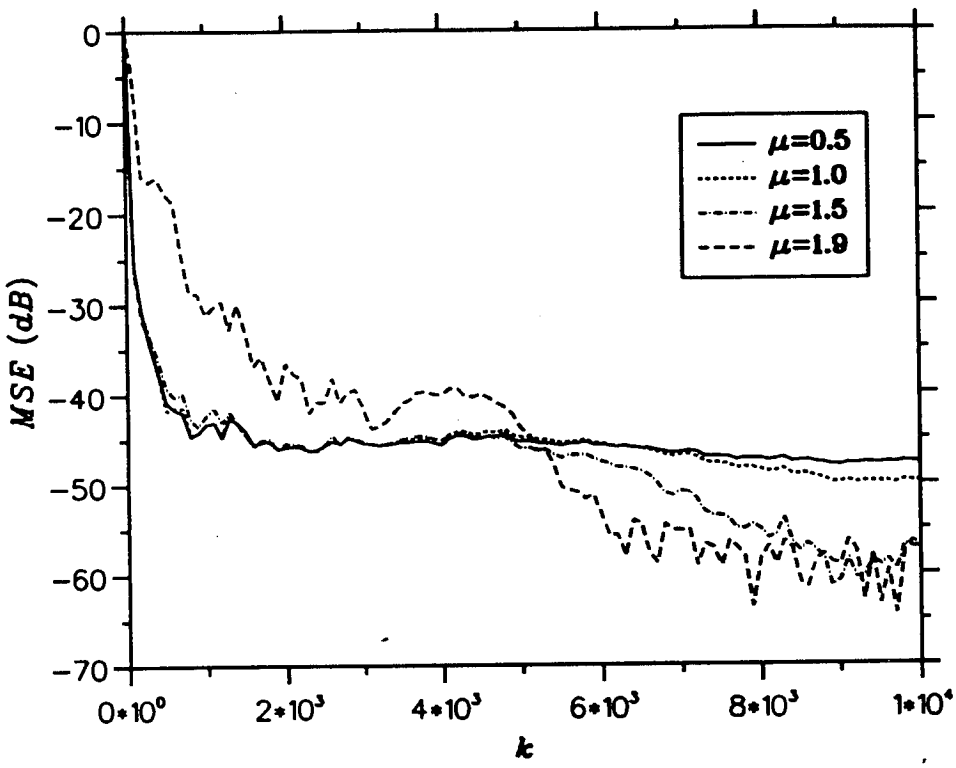
Fig(6.16a) The converging traces of the inverse filter coefficients in Example 6.3.3: SOR, μ=1.2 .



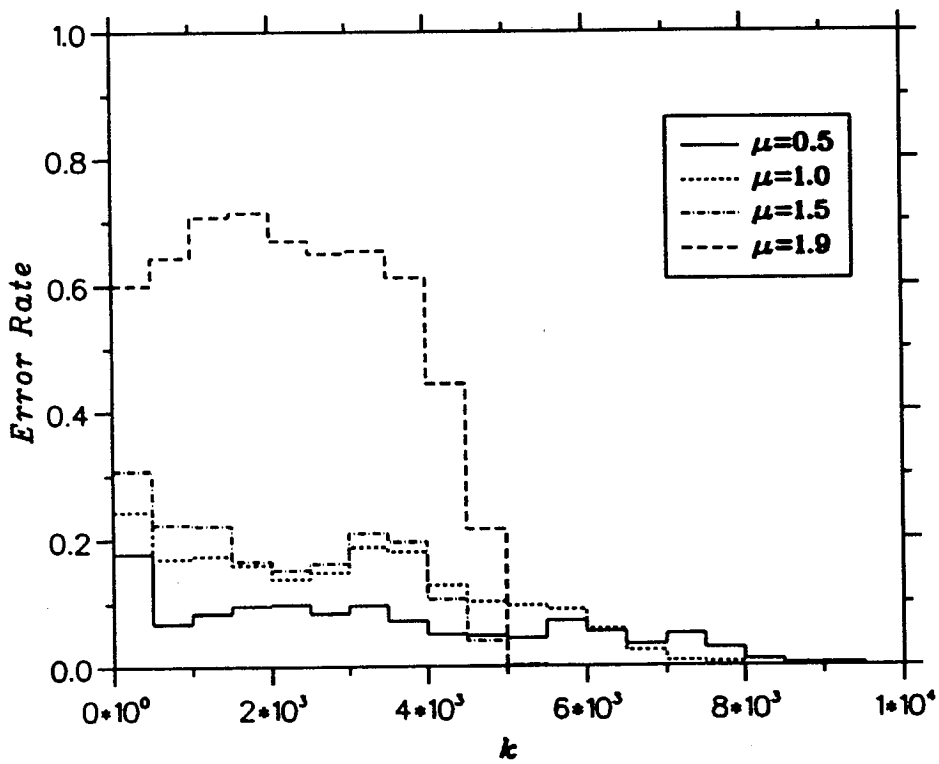
Fig(6.16b) The MSE curve of SOR algorithm in Example 6.3.3.



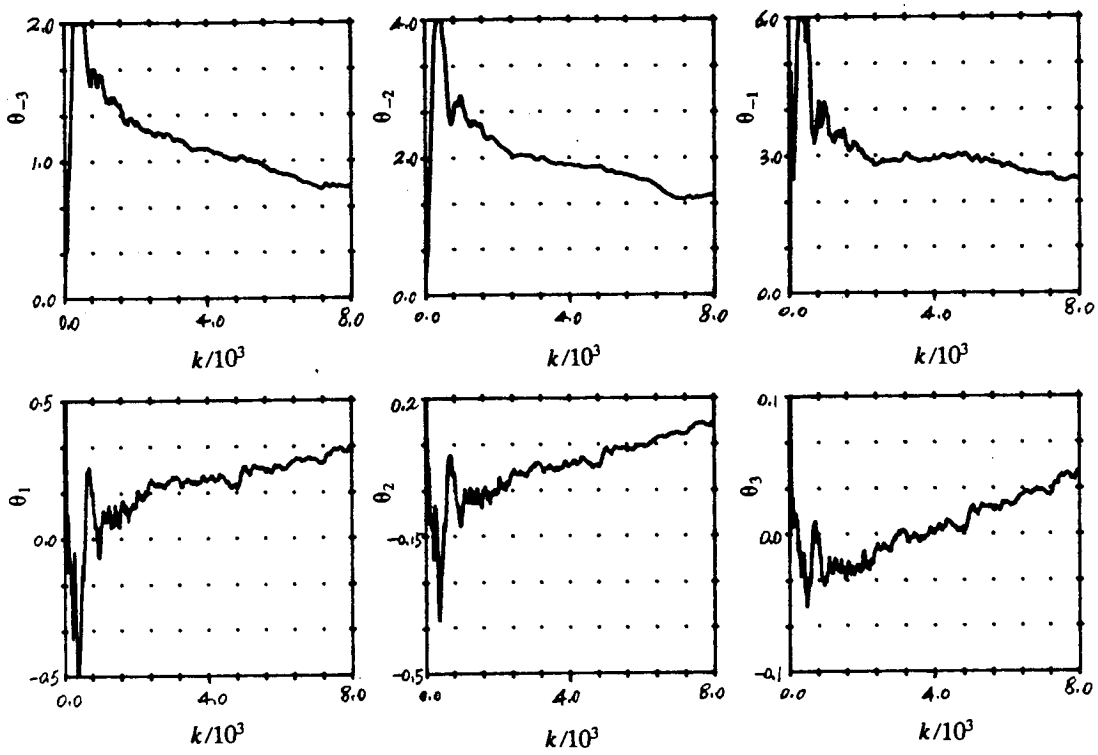
Fig(6.16c) The MSE curves of SOR algorithm in Example 6.3.3.



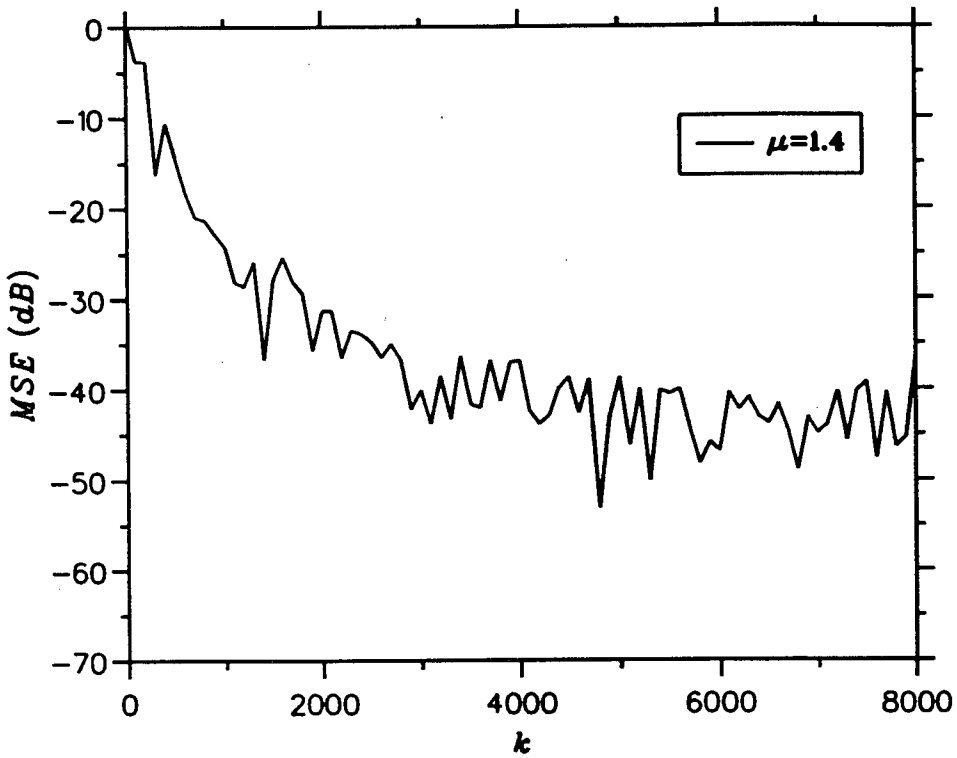
Fig(6.17a) The MSE curves of SOR algorithm in Example 6.3.4: different μ values.



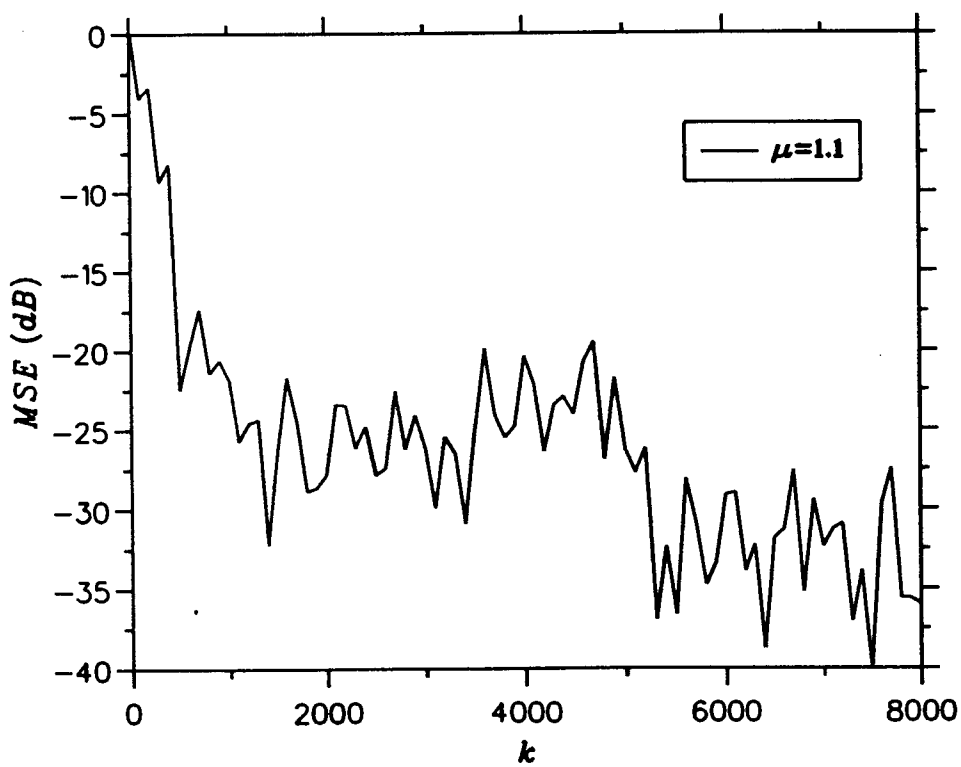
Fig(6.17b) The error rates (ER) of SOR algorithm in Example 6.3.4: different μ values.



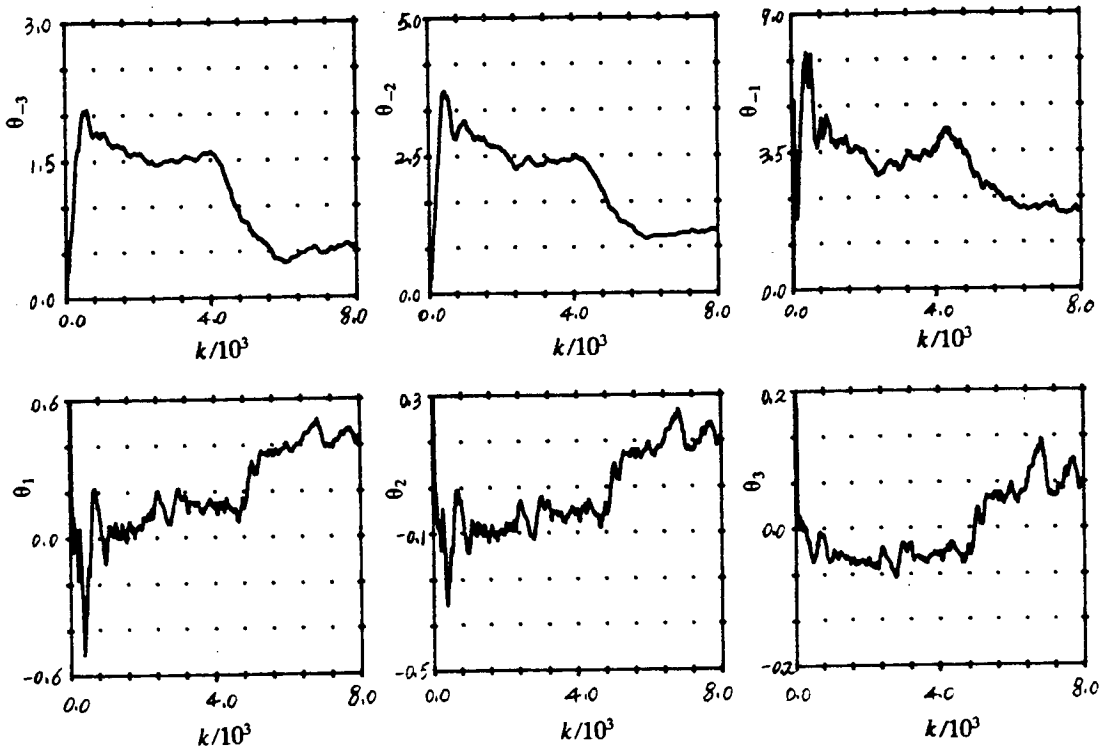
Fig(6.18a) The tracking traces of the inverse filter coefficients in Example 6.3.5: SOR, μ=1.4 .



Fig(6.18b) The MSE curve of SOR algorithm in Example 6.3.5.



Fig(6.19a) The MSE curve of SOR algorithm in Example 6.3.6.



Fig(6.19b) The tracking traces of the inverse filter coefficients in Example 6.3.6: SOR, μ=1.1 .

6.3.7 Conclusions

In this section, a new higher-order cumulant based equalisation algorithm has been described. In the algorithm, only the diagonal slices of cumulants are used. Several conclusions can be drawn:

(a) The algorithm is simple in form, and can be easily programmed and implemented.

(b) The introduction of the SOR scheme greatly accelerates the convergence rate and improves the adaptive performance of the algorithm. The algorithm also becomes insensitive to the selection of the step-size, *viz.*, the relaxation factor.

(c) The algorithm can deal with both LTI and LTV NMP channels.

The simulation results confirmed the feasibility and efficiency of the algorithm.

6.4 Summary

Two parametric HOC based blind equalisation techniques have been presented in this chapter: MA model based technique and AR model based technique. In both of these techniques, the channel can be not only MP but also NMP, and the data are multi-level PAM sequences, which are commonly used in modern communications.

In MA technique, a novel pair of data distribution transformations was introduced in order to enable the 3rd-order cumulant based two-step approach to become applicable to the (symmetrically distributed) PAM data. The thorny problem of multimodality is partially overcome. Also, the obtained algorithm has a much faster convergence rate than such non-HOC based methods as Sato approach.

In AR technique, we for the first time applied the SOR iteration scheme to an adaptive algorithm. As a result, the convergence rate of the algorithm is significantly accelerated. In the meantime, the algorithm becomes much less sensitive to the selection of the iteration step-size (relaxation factor). The algorithm is very simple in form and can be easily programmed and implemented. In addition, the algorithm can process not only LTI but also LTV channels.

All techniques are also justified by many computer simulations.

Appendix 6.1

In this appendix, the block version of Sato scheme is described as a reference of Section 6.2.

Here, we still use the denotations in Section 6.2, *i. e.*, $y(k)$ is the received signal, $\hat{x}_s(k)$ the restored signal, θ_i's the inverse filter coefficients, and W the number of received samples. In addition, the scaling factor is denoted as γ, and the step size as τ. According to [2], the corresponding cost function can be written as

$$\epsilon^2 = \frac{1}{W}\sum_{k=0}^{W-1}\epsilon_k^2 = \frac{1}{W}\sum_{k=0}^{W-1}[\hat{x}_s(k)-\gamma\cdot\mathrm{sign}(\hat{x}_s(k))]^2 \quad . \tag{A6.1.1}$$

From the above equation, we must have the following iterative algorithm (which we employed in the text):

$$\Phi_{yy}(i-l) = \frac{1}{W}\sum_{k=0}^{W-1}y(k-i)y(k-l) \quad , \tag{A6.1.2}$$

Iteration: $j=1, 2, \ldots.$

$$\hat{x}_s(k) = \sum_{i=r_1}^{r_2}\theta_i^{(j)}y(k-i) \quad . \tag{A6.1.3}$$

$$l = r_1, \ldots, r_2 \quad ,$$

$$\frac{\partial\epsilon^2}{\partial\theta_l} = \frac{2}{W}\sum_{k=0}^{W-1}[\hat{x}_s(k)-\gamma\cdot\mathrm{sign}(\hat{x}_s(k))]\cdot y(k-l)$$

$$= 2\cdot\sum_{i=r_1}^{r_2}\theta_i\Phi_{yy}(i-l) - \frac{2\gamma}{W}\sum_{k=0}^{W-1}[\mathrm{sign}(\hat{x}_s(k))\cdot y(k-l)] \quad , \tag{A6.1.4}$$

$$\theta_l^{(j+1)} = \theta_l^{(j)} - \tau\cdot\frac{\partial\epsilon^2}{\partial\theta_l} \quad . \tag{A6.1.5}$$

Appendix 6.2

In this appendix, we show a concise and self-contained proof for the theorem in Subsection 6.3.5.

Let $\mathbf{A}=(a_{ij})_{n\times n}$, $\Psi=(\psi_i)_n$, and $\mathbf{D}=(d_i)_n$. As in Eq(6.3.22a-b), the SOR iterative formula for Eq(6.3.35) can be expressed as

$$\psi_i^{(k+1)} = \psi_i^{(k)} + \delta\psi_i^{(k)} \quad , \tag{A6.2.1a}$$

where,

$$\delta\psi_i^{(k)} = \frac{\mu}{a_{ii}}(d_i - \sum_{j=1}^{i-1} a_{ij}\psi_j^{(k+1)} - \sum_{j=i}^{n} a_{ij}\psi_j^{(k)}) \quad . \tag{A6.2.1b}$$

Now, let $\mathbf{A}=\mathbf{A}_D+\mathbf{A}_L+\mathbf{A}_U$, where $\mathbf{A}_D=diag(a_{11}, \ldots, a_{nn})$, $\mathbf{A}_L$ and $\mathbf{A}_U$ are respectively the lower and upper triangular matrices corresponding to A. Then Eq(A6.2.1a-b) can be compactly written as one expression:

$$\mathbf{A}_D\Psi(k+1)=\mu(\mathbf{D}-\mathbf{A}_L\Psi(K+1)-\mathbf{A}_U\Psi(k))+(1-\mu)\mathbf{A}_D\Psi(k) \quad , \tag{A6.2.2}$$

where $\Psi(k)=(\psi^{(k)})_n$. Since $\mathbf{A}_D+\mu\mathbf{A}_L$ is nonsingular for any μ (from the fact $a_{ii}\neq 0$ in our situation), we can obtain the following matrix-form iterative formula of the SOR method:

$$\Psi(k+1) = \mathbf{L}_\mu\Psi(k)+\mathbf{f} \quad , \tag{A6.2.3}$$

where,

$$\mathbf{L}_\mu = (\mathbf{A}_D+\mu\mathbf{A}_L)^{-1}((1-\mu)\mathbf{A}_D-\mu\mathbf{A}_U) \quad , \tag{A6.2.4a}$$

and

$$\mathbf{f} = \mu(\mathbf{A}_D+\mu\mathbf{A}_L)^{-1}\mathbf{D} \quad . \tag{A6.2.4b}$$

Here, $\mathbf{L}_\mu$ is termed the *iterative matrix* of the SOR method.

According to the general iteration theory for linear equation systems [111], in order to prove the convergence of Eq(A6.2.3), we only need to show that the spectral radius of $\mathbf{L}_\mu$ is less than 1, or equivalently, any eigenvalue λ of $\mathbf{L}_\mu$ satisfies $|\lambda|<1$.

Let $\mathbf{Z}=(z_i)_n\neq 0$ be the eigenvector of $\mathbf{L}_\mu$ for λ, *i. e.*,

$$\mathbf{L}_\mu\mathbf{Z} = \lambda\mathbf{Z} \quad . \tag{A6.2.5}$$

Substituting Eq(A6.2.4a) into Eq(A6.2.5), we can obtain

$$((1-\mu)\mathbf{A}_D-\mu\mathbf{A}_U)\mathbf{Z} = \lambda(\mathbf{A}_D+\mu\mathbf{A}_L)\mathbf{Z} \quad . \tag{A6.2.6}$$

Thus, the following inner product expression holds:

$$(((1-\mu)\mathbf{A}_D-\mu\mathbf{A}_U)\mathbf{Z}, \mathbf{Z}) = (\lambda(\mathbf{A}_D+\mu\mathbf{A}_L)\mathbf{Z}, \mathbf{Z}) \quad , \tag{A6.2.7}$$

from which, the expression for λ can be derived:

$$\lambda = \frac{(\mathbf{A}_D\mathbf{Z}, \mathbf{Z})-\mu(\mathbf{A}_D\mathbf{Z}, \mathbf{Z})-\mu(\mathbf{A}_U\mathbf{Z}, \mathbf{Z})}{(\mathbf{A}_D\mathbf{Z}, \mathbf{Z})+\mu(\mathbf{A}_L\mathbf{Z}, \mathbf{Z})} \quad . \tag{A6.2.8}$$

Notice that

$$(\mathbf{A}_D\mathbf{Z}, \mathbf{Z}) = \sum_{i=1}^{n} a_{ii} |z_i|^2 = \sigma > 0 \ . \tag{A6.2.9}$$

Let

$$(\mathbf{A}_L\mathbf{Z}, \mathbf{Z}) = \alpha + i\beta \ \ (\text{here } i = \sqrt{-1}) \ . \tag{A6.2.10}$$

Since $\mathbf{A}=\mathbf{A}^T$ ($\mathbf{A}$ is symmetric), $\mathbf{A}_U = \mathbf{A}_L^T$ must hold. Thus we have

$$(\mathbf{A}_U\mathbf{Z}, \mathbf{Z}) = (\mathbf{Z}, \mathbf{A}_L\mathbf{Z}) = \overline{(\mathbf{A}_L\mathbf{Z}, \mathbf{Z})} = \alpha - i\beta \ . \tag{A6.2.11}$$

From Eq(A6.2.8), λ can be determined:

$$\lambda = \frac{(\sigma - \mu\sigma - \alpha\mu) + i\mu\beta}{(\sigma + \alpha\mu) + i\mu\beta} \ . \tag{A6.2.12}$$

Then,

$$|\lambda|^2 = \frac{(\sigma - \mu\sigma - \mu\alpha)^2 + (\mu\beta)^2}{(\sigma + \alpha\mu)^2 + (\mu\beta)^2} \ . \tag{A6.2.13}$$

On the other hand, since $\mathbf{A}$ is positive definite, we must have

$$0 < (\mathbf{AZ}, \mathbf{Z}) = ((\mathbf{A}_D + \mathbf{A}_L + \mathbf{A}_U)\mathbf{Z}, bZ) = \sigma + 2\alpha \ . \tag{A6.2.14}$$

Then, the inequality $0<\mu<2$ can guarantee

$$(\sigma - \mu\sigma - \alpha\mu)^2 - (\sigma + \mu\alpha)^2 = \mu\sigma(\sigma + 2\alpha)(\mu - 2) < 0 \ . \tag{A6.2.15}$$

Also, since $\sigma>0$ and $0<\mu<2$, we must have

$$(\sigma + \alpha\mu)^2 + (\mu\beta)^2 \neq 0 \ . \tag{A6.2.16}$$

Eq(A6.2.15) and Eq(A6.2.16) immediately lead to

$$|\lambda| < 1 \ . \tag{A6.2.17}$$

Therefore, the SOR method converges •

Chapter 7:

CONCLUSIONS AND FUTURE TRENDS

7.1 General Conclusions

Blind deconvolution and its applications have been studied in this thesis. The emphasis of our investigation has been primarily on the higher-order cumulant (HOC) based deconvolution techniques. Based on different parametric models, we have proposed several new blind deconvolution algorithms, which can be used to deal with not only minimum phase systems but also nonminimum phase (NMP) systems. The vital reason for this, as remarked in the preceding chapters, is that HOC carry the phase information of systems (up to a linear phase shift). In terms of the features of our algorithms (*i.e.*, the contributions of this thesis), the following conclusions are in order.

Firstly, based on a MA system model, a two-step relay identification approach was suggested (Chapter 4). By combining the existing simple close-form formulae with an optimal fitting scheme, we, to a great extent, overcame the problem of multimodality which is notorious in nonlinear optimisation. As a result, our algorithm greatly improves the results of the existing close-form methods.

Secondly, based on an AR system model, six families of linear equations with respect to the inverse filter coefficients were derived according to three new theorems (Chapter 5). Generally, they are all of a similar simple form. As a result of the linearity of these equations, the uniqueness of solutions can normally be guaranteed. Thus, the risk of local minima or contradictory solutions, which exists with some other methods such as Giannakis-Mendel's, is removed. It has been recognised that our result here represents another important progress following Giannakis-Mendel method in blind system identification. Particularly, compared with the existing linear (*i.e*, noncausal AR based) approaches, our algorithms here possess the following two important advantages:

a) Only the diagonal cumulant slices are used, which in general makes the algorithms more accurate, and also, makes the forms of the concerned matrices simple and elegant;

b) The 4th-order cumulants are employed exactly the same way as the 3rd-order cumulants, which enables the 4th-order cumulant based algorithms (for deconvolution of a unskewed series) to be implemented as easily as their 3rd-order cumulants based counterparts (for deconvolution of a skewed series).

Thirdly, the preceding two-step MA approach was successfully applied to the blind equalisation of multilevel pulse amplitude modulation (PAM) series received from NMP channels (Chapter 6). A novel data distribution transformation pair (SAT and AST) was

adopted to enable the 3rd-order cumulant based algorithms to become applicable to the symmetrically distributed data. The obtained equalisation scheme has been found to possess a faster convergence rate than the non-cumulant based methods such as the well known Sato method.

Fourthly, the adaptive version of the AR algorithm was achieved in the context of blind equalisation of NMP channels (Chapter 6). In this study, we for the first time introduced the successive over-relaxation (SOR) iteration scheme into an adaptive algorithm. As a consequence, the convergence rate of our algorithm is significantly accelerated, which enable our equaliser to deal with not only linear time-invariant (LTI) but also linear time-variant (LTV) NMP channels. In the meantime, the requirement for the selection of iteration step size is relaxed as well.

In addition, as an alternative to the predictive method, a new robust deconvolution algorithm (MP system oriented), named variance approximation and series decoupling (VASD), was obtained in Chapter 3. Unlike the predictive method, this algorithm does not need all the autocorrelation values of output series. No other approach with the similar feature has ever been reported do far.

The multifarious computer simulations given in every chapter provide the firm support for the above conclusions.

7.2 Future Trends

Originally, the solution of the blind deconvolution problem depended upon two assumptions: 1) minimum phase (MP) system transfer function, and 2) independent and identically distributed input series. As a result, the general trend of blind deconvolution research has been following, and will follow, the line of removal or relaxation of the above assumptions.

Regarding the first assumption, namely minimum phase, as demonstrated in this thesis, it has in theory been removed by the adoption of higher-order cumulant (HOC) analysis. Although HOC based techniques started to be investigated only about ten years ago, many encouraging results have been obtained. This is in no doubt a significant breakthrough in the development history of deconvolution technology. In the meantime, it has also been indicated that the future work in blind deconvolution should be based on HOC.

As an old saying goes, however, there is no pure gold in this world. Much work remains to be done in the domain of HOC based blind deconvolution. Below are several possible directions for the further research.

1) Developing more accurate methods to estimate the 3rd- and 4th-order cumulants from data records with finite length. This work is very important for further broadening the application range of HOC based techniques. As has been observed in this thesis and other relevant literatures, to reach the same accuracy, HOC estimation require much longer data record than autocorrelation estimation. The general law here is: the higher order one uses, the longer data record one needs.

2) Exploiting parallelism in HOC based approaches to develop fast algorithms. HOC based methods usually involve relatively larger computational amount. Parallelism appears to be one of the keys to deal with the problem.

3) Implementing HOC based algorithms in VLSI. In many application fields such as communications and real time control, most algorithms should finally be implemented in VLSI, so should HOC based deconvolution algorithms.

4) Comprehensively testing and comparing the many existing algorithms. Although this work is indispensable for the final maturisation of HOC based techniques, not much has been reported so far. The aspects of testing and comparison can include, for example, numerical stability, computational burden, convergence performance, and so forth.

5) Application of the existing algorithms to practical environments. This is obviously a more advanced, and the final, destination for HOC based deconvolution techniques.

On the other hand, in the second assumption---independence and identical distribution, the distribution problem has been well understood through the introduction of HOC: the only necessary constraint is non-Gaussianity. In fact, non-Gaussianity can be satisfied in the majority of cases. For the assumption of independence, however, the research is still in a very early stage. This work will sooner or later be carried out because independence is not alway the case in practice.

In summary, blind deconvolution is a wonderful domain which is full of potentials and challenges. Through people's consistent effort of several decades, many significant progresses, some of which are revolutionary, have been achieved. HOC based techniques is such an example.

REFERENCES

[1] T. G. Stockham, T. M. Cannon, and R. B. Ingebretsen, "Blind deconvolution through digital signal processing," *Proc. IEEE*, Vol. 63, No. 4, pp. 678-692, April 1975.

[2] A. V. Oppenheim, R. W. Schafer, and T. G. Stockham, "Non-linear filtering of multiplied and convolved signals," *Proc. IEEE*, Vol. 56, pp. 1264-1291, Aug. 1968.

[3] E. A. Robinson, *Predictive decomposition of time series with applications to seismic exploration*, PhD thesis, MIT, Cambridge, Mass., 1954.

[4] K. L. Peacock and S. Treitel, "Predictive deconvolution theory and practice," *Geophysics*, Vol. 34, No. 2, pp. 155-169, 1969.

[5] E. A. Robinson and S. Treitel, *Geophysical signal analysis*, Englewood Cliffs, NJ: Prentice-Hall, 1980.

[6] E. A. Robinson and T. S. Durrani, *Geophysical signal processing*, London: Prentice-Hall, 1986.

[7] A. Ziolkowski, *Deconvolution*, Boston: International Human Resources Development Corporation, 1984.

[8] L. C. Wood, and S. Treitel, "Seismic signal processing," *Proc. IEEE*, Vol. 63, No. 4, pp. 649-661, April 1975.

[9] O. Yilmaz, *Seismic data processing*, in series: *Investigations in geophysics*, Vol. 2, S. M. Doherty, Ed., SEG (USA), 1987.

[10] T. J. Ulrych and T. Matsuoka, "The output of predictive deconvolution," *Geophysics*, Vol. 56, No. 3, pp. 371-377, March 1991.

[11] J. M. Mendel, *Optimal seismic deconvolution: An estimation based approach*, New York: Academic Press, 1983.

[12] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for nonGaussian linear process," *Ann. Stat.*, Vol. 10, pp. 1195-1208, 1982.

[13] C. L. Nikias and M. R. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *IEEE Proc.*, Vol. 75, No. 7, pp. 869-891, July 1987.

[14] H.-H. Chiang and C. L. Nikias, "Adaptive deconvolution and identification of nonminimum phase FIR systems based on cumulants," *IEEE Trans. Automat. Contr.*, Vol. AC-35, pp. 36-47, Jan. 1990.

[15] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase system using higher order statistics," *IEEE Trans.*, Vol. ASSP-37, pp. 360-377, March 1989.

[16] R. W. Shafer and L. R. Rabiner, "Digital representations of speech signals," *Proc. IEEE*, Vol. 63, No. 4, pp. 662-677, April 1975.

[17] J. Makoul, "Linear prediction: A tutorial review," *Proc. IEEE*, Vol. 63, No. 4, pp. 561-581, April 1975.

[18] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Trans. Commu.*, pp. 679-682, June 1975.

[19] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a non-minimum phase system: Blind adjustment of a linear equaliser in data communications," *IEEE Trans. Automa. Contr.*, Vol. AC-25, No. 3, pp. 385-399, June 1980.

[20] A. Benveniste and M. Goursat, "Blind equalisers," *IEEE Trans. Commu.*, Vol. COM-32, No. 8, pp. 871-883, Aug. 1984.

[21] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commu.*, Vol. COM-28, No. 11, pp. 1867-1875, Nov. 1980.

[22] M. Cannon, "Blind deconvolution of spatially invariant image blurs with phase," *IEEE Trans.*, Vol. ASSP-24, No. 1, pp. 58-63, Feb. 1976,

[23] G. Thomas, C. Dussaussois, F. Buret, and P. Auriol, "Deconvolution without system model or new blind deconvolution," *IEEE ICASSP'85*, pp. 442-444, 1985.

[24] S. M. Riad and N. S. Nahman, "Application of the homomorphic deconvolution for the separation of TDR signals occurring in overlapping time windows," *IEEE Trans. Instru. Meas.*, Vol. IM-25, No. 4, pp. 388-391, Dec. 1976.

[25] F. G. Stremler, *Introduction to communication systems*, Chap. 9, Addison-Wesley Publishing Company, 2nd Edition, 1982.

[26] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, Vol. 73, No. 9, pp. 1349-1387, 1985.

[27] K. Yana, H. Marushima, H. Mino, and N. Takeuchi, "Bispectral analysis of filtered impulse process with applications to the analysis of bioelectric phenomena," *Proc. of WORKSHOP ON HIGHER ORDER SPECTRAL ANALYSIS*, pp. 140-145, Vail, Colorado, June 1989.

[28] G. B. Giannakis, "Cumulants: A Powerful tool in signal processing," *Proc. IEEE*, Vol. 75, No. 9, Sept. 1987, pp. 1333-1334.

[29] J. K. Tugnait, "Comments on "Cumulants: A powerful tool in signal processing"," *Proc. IEEE*, Vol. 77, No. 3, March, 1989, pp. 491-492.

[30] A. Swami and J. M. Mendel, "Closed-form recursive estimation of MA coefficients using autocorrelations and third-order cumulants," *IEEE Trans.*, Vol. ASSP-37, No. 11, pp. 1794-1795, Nov., 1989.

[31] J. K. Tugnait, "Order reduction of SISO nonminimum phase stochastic systems," *IEEE Trans. Automat. Contr.*, Vol. AC-31, No. 7, pp. 623-632, July 1986.

[32] C. L. Nikias and H.-H. Chiang, "Higher-order spectrum estimation via noncausal autoregressive modelling and deconvolution," *IEEE Trans.*, Vol. ASSP-36, pp. 1911-1913, Dec. 1988.

[33] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: The theoretical results and some applications," *IEEE Proc.*, Vol. 79, No. 3, pp. 278-305, March 1991.

[34] R. Kuk, *Introduction to digital signal processing*, New York: Mcgraw-Hill, 1990.

[35] R. A. Roberts and C. T. Mullis, *Digital signal processing*, Addison-Wesley Publishing Company, Inc., 1987.

[36] S. Haykin, *Adaptive filter theory*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 2nd edition, 1991.

[37] M. T. Silvia and E. A. Robinson, *Deconvolution of geophysical time series in the exploration for oil and natural gas*, Amsterdam: Elsevier Scientific Publishing Company, 1979.

[38] N. Levinson, "A heuristic exposition of Wiener's mathematical theory of prediction and filtering," *J. Math. Phys.*, Vol. 25, pp. 110-119, 1947.

[39] J. Durbin, "The fitting of time series models," *Rev. Inst. Stat.*, Vol. 28, pp. 233-243, 1960.

[40] R. A. Wiggins and E. A. Robinson, "Recursive solution to the multichannel filtering problem," *J. Geophys. Res.*, Vol. 70, pp. 1885-1891, 1965.

[41] W. A. Schneider, "Integral formulation for migration in two and three dimensions," *Geophysics*, Vol. 43, pp. 49-76, 1978.

[42] A. V. Oppenheim, *Superposition in a class of nonlinear systems*, Research Lab. of Electronics, MIT, Tech. Rep. 432, 1965.

[43] A. V. Oppenheim, *Optimum homomorphic filters*, Research Lab. of Electronics, MIT, Quart. Progr. Rep. 77, pp. 248-260, 1965.

[44] R. W. Schafer, *Echo removal by discrete generalized linear filtering*, Research Lab. of Electronics, MIT, Tech. Rep. 466, 1969.

[45] T. J. Ulrych, "Application of homomorphic deconvolution to seismology," *Geophysics*, Vol. 36, No. 4, pp. 650-660, 1971.

[46] P. L. Stoffa, P. Buhl, and G. M. Bryan, "The application of homomorphic deconvolution to shallow-water marine seismology--Part I: Models," *Geophysics*, Vol. 39, No. 4, pp. 401-416, 1974.

[47] P. Buhl, P. L. Stoffa, and G. M. Bryan, "The application of homomorphic deconvolution to shallow-water marine seismology--Part II: Real data," *Geophysics*, Vol. 39, No. 4, pp. 417-426, 1974.

[48] B. P. Bogert, M. J. Healey, and J. W. Tukey, "The quefrency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking," *Proc. of Symp. on Time Series Analysis*, M. Rosenblatt, Ed., NY: Wiley, pp. 209-243, 1963.

[49] D. G. Stone, "Wavelet estimation," *Proc. IEEE*, Vol. 72, No. 10, pp. 1394-1402, Oct. 1984.

[50] A. Papoulis, *Probability, random variables, and stochastic processes*, New York: Mcgraw-Hill, 1965.

[51] J. Kormylo and J. M. Mendel, "Applying maximum-likelihood deconvolution to well-log impedance data," presented at *the 49th Annual Meeting of the Society of*

Exploration Geophysicists, New Orleans, 1979.

[52] J. Goutsias and J. M. Mendel, "Maximum-likelihood deconvolution: An optimization theory perspective," *Geophysics*, Vol. 51, No. 6, pp. 1206-1220, June 1986.

[53] P. S. Schultz, "Seismic data processing: Current industry practice and new directions," *Geophysics*, Vol. 50, No. 12, pp. 2452-2457, Dec. 1985.

[54] K. H. Waters, *Reflection seismology: A tool for energy source exploration,*" New York: John Wiley and sons, Inc., 3rd Edition, 1987.

[55] R. A. Wiggins, "Minimum entropy deconvolution," *Proc. of Int. Symp. Computer Aided Seismic Analysis and Discrimination*, Falmouth, Mass., pp. 7-14, IEEE Computer Society, June 9-10, 1977.

[56] R. A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, Vol. 16, pp. 21-35, 1978.

[57] M. Ooe and T. J. Ulrych, "Minimum entropy deconvolution with an exponential transformation," *Geophys. Prosp.*, Vol. 27, No. 2, pp. 458-473, June 1979.

[58] A. T. Walden, "Non-Gaussian reflectivity, entropy, and deconvolution," *Geophysics*, Vol. 50, No. 12, pp. 2862-2888, Dec. 1985.

[59] J. F. Claerbout, "Parsimonious deconvolution," *Stanford Exploration Project*, Vol. 13, pp. 1-9, 1977.

[60] C. A. Cabrelli, "Minimum entropy deconvolution and simplicity: A noniterative algorithm," *Geophysics*, Vol. 50, No. 3, pp. 394-413, March 1984.

[61] C.-S. Wang, J. Tang, and B. Zhao, "An improvement on D norm deconvolution: A fast algorithm and the related procedure," *Geophysics*, Vol. 56, No. 5, pp. 675-680, May 1991.

[62] R. E. White, "Maximum kurtosis phase correction," *Geophysical J.*, Vol. 95, pp. 371-389, 1988.

[63] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inform. Theory*, Vol. IT-30, No. 2, pp. 161-176, March 1984.

[64] J. E. Mazo, "Analysis of decision-directed equalizer convergence," *Bell Syst. Tech. J.*, Vol. 59, No. 10, pp. 1857-1876, Dec. 1980.

[65] S. Bellini and F. Rocca, "Blind deconvolution: Polyspectra or Bussgang techniques?" in *Digital Communications*, E. Biglieri and G Prati, Ed., Elsevier Science Publishers B. V. (North-Holland), pp. 251-263, 1986.

[66] J. D. Scargle, "Absolute value optimization to estimate phase properties of stochastic time series," *IEEE Trans. Inform. Theory*, Vol. IT-23, pp. 140-143, Jan. 1977.

[67] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a "stop-and-go" decision-directed algorithm," *IEEE Trans. Commun.*, Vol. COM-35, No. 9, pp. 877-887, Sept. 1987.

[68] D. R. Brillinger, "An introduction to polyspectra," *Ann. Math. Statist.*, Vol. 36, pp. 1351-1374, 1965.

[69] M. Rosenblatt and J. W. Van Ness, "Estimation of the bispectrum," *Ann. Math. Statist.*, Vol. 36, pp. 1120-1136, 1965.

[70] M. Rosenblatt, "Cumulants and cumulant spectra," in *Time Series in Frequency Domain*, D. R. Brillinger and P. Krishnaiah, Ed., Amsterdam, the Netherlands: North-Holland, pp. 369-382, 1983.

[71] C. L. Nikias, "Signal processing with higher-order spectra," in *Tutorial Book for ICASSP'91 Tutorial 4: Higher-order Spectral Analysis*, Part I.

[72] A. T. Walden and J. W. J. Hosken, "The nature of the non-Gaussianity of primary reflection coefficients and its significance for deconvolution," *Geophys. Prosp.*, Vol. 34, pp. 1038-1066, 1986.

[73] R. Godfrey and F. Rocca, "Zero memory non-linear deconvolution," *Geophys. Prosp.*, Vol. 29, pp. 189-228, 1981.

[74] V. Kh. German *et al*, "Bispectral analysis of sea-level variations," *Sov. Meteorol. Hydrol.*, No. 11, pp. 50-57, 1980.

[75] J. W. J. Hosken, "A stochastic model of seismic reflections," presented at *the 50th Ann. Int. Meet. of SEG*, 1980.

[76] P. J. Huber *et al*, "Statistical methods for investigating phase relations in stationary stochastic processes," *IEEE Trans. Audio Electroacoust.*, Vol. AU-19, pp. 78-86, 1971.

[77] C. L. Nikias *et al*, "Identification of nonminimum phase communication channels via parametric modeling of third moments," *Proc. Int. Con. Communications (ICC)*, Toronto, Ont., Canada, pp. 667-670, June 1986.

[78] B. Wells, "Voiced/unvoiced decision based on the bispectrum," *Proc. ICASSP'85*, Tampa, FL, pp. 1589-1592, March 1985.

[79] T. Sato *et al*, "Real time bispectral analysis of gear noise and its application to contactless diagnosis," *J. Acoust. Soc. Am.*, Vol. 62, No. 2, pp. 382-387, Aug. 1977.

[80] T. Matsuoka and T. J. Ulrych, "Phase estimation using the bispectrum," *Proc. IEEE*, Vol. 72, No. 10, pp. 1403-1411, Oct. 1984.

[81] M. R. Raghuveer and C. L. Nikias, "Bispectrum estimation: a parametric approach," *IEEE Trans.*, Vol. ASSP-33, No. 4, pp. 1213-1230, Oct. 1985.

[82] D. R. Brillinger and Rosenblatt, "Computation and interpretation of kth-order spectrum," in *Spectral Analysis of Times Series*, B. Harris, Ed. New York: Wiley, pp. 189-232, 1967.

[83] J. K. Tugnait, "Order reduction of SISO nonminimum-phase stochastic systems," *IEEE Trans. Automat. Contr.*, Vol. AC-31, pp. 623-632, July 1986.

[84] J. K. Tugnait, "Identification of non-minimum phase linear stochastic systems," *Automatica*, Vol. 22, No. 4, pp. 4457-4464, 1986.

[85] G. B. Giannakis and J. M. Mendel, "Second or higher-order statistics for ARMA parameter estimation and order determination?" presented at *the 3rd ASSP Workshop on Spectrum Estimation and Modeling*, Boston, MA, 1986.

[86] Y. T. Chan and J. C. Wood, "A new order determination technique for ARMA process," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-32, pp. 517-521, June 1984.

[87] G. B. Giannakis, *Signal processing via higher-order statistics*, PhD thesis, Univ. of Southern Cali., Los Angeles, June 1986.

[88] J. K. Tugnait, "Identification of linear stochastic systems via second- and fourth-order cumulant matching," *IEEE Trans., Inform. Theory*, Vol. IT-33, No. 3, pp. 393-407, May 1987.

[89] B. Friedlander and B. Porat, "Adaptive IIR algorithms based on higher-order statistics," *IEEE Trans.*, Vol. ASSP-37, No. 4, pp. 485-495, April 1989.

[90] C. L. Nikias, "ARMA bispectrum approach to nonminimum phase system identification," *IEEE Trans.*, Vol. ASSP-36, pp. 513-524, Apr. 1988.

[91] A. M. Tekalp and A. T. Erdem, "Higher-order spectrum factorization in one and two dimensions with applications in signal processing and nonminimum phase system identification," *IEEE Trans.*, Vol. ASSP-37, pp. 1537-1549, Oct. 1989.

[92] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from higher-order moments," *IEEE Trans. Automat. Contr.*, Vol. AC-35, pp. 27-35, Jan. 1990.

[93] S. Kirkpatrick, C. D. Gelalt, and Jr. M. P. Vecchi, "Optimization by simulated annealing," *Science*, Vol. 220, No. 4598, pp. 671-680, 13 May 1983.

[94] W. H. Press *et al*, *Numerical recipes: The art of scientific computing*, Cambridge: Cambridge University Press, 1986, Ch. 10.

[95] G. Giannakis and J. M. Mendel, "ARMA order determination via higher-order statistics," in *Math. Theory of Networks and Systems*, C. L. Martin, Ed., Amsterdam, The Netherlands: North Holland.

[96] A. T. Walden and J. W. J. Hosken, "An investigation of the spectral properties of primary reflection coefficients," *Geophysical Prospecting*, Vol. 33, pp. 400-435, 1985.

[97] C. L. Nikias and R. Pan, "ARMA modeling of fourth-order cumulants and phase estimation," *Circuit Syst. Signal Processing*, Vol. 7, No. 3, pp. 291-325, 1988.

[98] G. B. Giannakis and J. M. Mendel, "Cumulant-based Order determination of non-Gaussian ARMA models," *IEEE Trans.*, Vol. ASSP-38, pp. 1411-1423, Aug. 1990.

[99] A. R. Collar and A. Simpson, *Matrices and Engineering Dynamics*, Chichester: Ellis Horwood Ltd, 1987, Ch. 1.

[100] J. Karaoguz and S. Ardalan, "Use of blind equalization for teletext broadcast systems," *IEEE Trans. Broadcast.*, Vol. BRO-37, No. 2, pp. 44-54, June 1991.

[101] F. Ross and D. P. Taylor, "An enhancement to blind equalization algorithms," *IEEE Trans. Commu.*, Vol. COM-39, No. 5, pp. 636-639, May 1991.

[102] V. Weerackody, S. A. Kassam, and K. R. Laker, "Convergence analysis of an algorithm for blind equalization," *IEEE Trans. Commu.*, Vol. COM-39, No. 6, pp. 856-865, June 1991.

[103] D. Hatzinakos and C. L. Nikias, "Polyspectral techniques for blind equalization of multilevel PAM schemes," *Proc. IEEE ICC'90*, pp. 1507-1511, 1990.

[104] D. Hatzinakos and C. L. Nikias, "Blind equalization based on second and fourth order statistics," *Proc. IEEE ICC'90*, pp. 1512-1516, 1990.

[105] D. Hatzinakos and C. L. Nikias, " Blind equalization using a trispectrum-based algorithm," *IEEE Trans. Commu.*, Vol. COM-39, No. 5, pp. 669-682, May 1991.

[106] O. Macchi and A. Hachicha, "Self-adaptive equalization based on a prediction principle," *Proc. of IEEE Globecom*, pp. 1641-1645, 1986.

[107] L. Ljung, "Consistency of the least squares identification method," *IEEE Trans. Automat. Contr.*, Vol. AC-22, pp. 539-551, Oct. 1977.

[108] C. L. Nikias and A. N. Venetsanopoulos, "Identification of nonminimum phase communication channels via parametric modelling of third moments," *Proc. IEEE ICC'86*, pp. 667-670, 1986.

[109] B. Friedlander, "Adaptive algorithms for finite impulse response filters," in *Adaptive Filters*, Chap. 3, C. F. N. Cowan and P. M. Grant, Ed., Prentice-Hall, Inc., New Jersey, 1985.

[110] R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall International, London, 1962.

[111] D. M. Young, *Iterative Solution of Large Linear Systems*, Academic Press, NY, 1971.

[112] W. Gardner, "Learning characteristics of stochastic gradient descent algorithms: A general study, analysis and critique," *Signal Processing*, Vol. 6, pp. 113-133, 1984.

[113] V. A. Gholkar, "Mean square convergence analysis of LMS algorithm," *Electron. Lett.*, Vol. 26, No. 20, pp. 1705-1706, Sept. 1990.

[114] K. Y. Wong and E. Polak, "Identification of linear discrete time systems using the instrumental variable method," *IEEE Trans. Automat. Contr.*, Vol. AC-12, pp. 707-718, Dec. 1967.

[115] R. Mehlan, *Investigation of Higher Order Statistics*, Diploma Thesis carried out at Imperial College, London, 1990.

PUBLICATION LIST

[1] F.-C. Zheng, S. McLaughlin, and B. Mulgrew, "Robust blind deconvolution algorithm: variance approximation and series decoupling," *Electron. Lett.*, Vol. 26, No. 13, pp. 921-923, 21st June 1990.

[2] F.-C. Zheng, S. McLaughlin, and B. Mulgrew, "Blind equalization of nonminimum phase channels: higher order cumulant based algorithm," to appear in *IEEE Trans. Signal Processing*.

[3] F.-C. Zheng, S. McLaughlin and B. Mulgrew, "Blind equalisation of PAM series via higher-order cumulant fitting," *Proc. of IEEE International Conference on Communications (ICC)*, Denver, Colorado, USA, pp. 1393-1397, June 23-26, 1991.

[4] F.-C. Zheng, S. McLaughlin and B. Mulgrew, "Blind deconvolution algorithms based on 3rd- and 4th-order cumulants," *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toronto, Ontario, Canada, pp. 1573-1576, May 14-17, 1991.

[5] F.-C. Zheng, S. McLaughlin, and B. Mulgrew, "Cumulant-based deconvolution and identification: several new families of linear equations," submitted to *Signal Processing*.

[6] F.-C. Zheng, S. McLaughlin, and B. Mulgrew, "Blind equalisation of multilevel PAM data for nonminimum phase channels via 2nd- and 4th-order cumulants," submitted to *Signal Processing*.

[7] F.-C. Zheng, S. McLaughlin, and B. Mulgrew, "A 2nd- and 4th-order cumulant based blind equalisation algorithm for nonminimum phase channels," *Proc. of IEEE & EURASIP International Signal Processing Workshop on Higher Order Statistics*, Chamrousse, France, pp. 79-82, July 10-12, 1991.