Independent Component Analysis Applications in CDMA Systems

By
Olcay Kalkan

A Dissertation Submitted to the
Graduate School in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

Department: Electrical and Electronics Engineering
Major: Electronics and Communication

İzmir Institute of Technology
İzmir, Turkey

July, 2004

We approve the thesis of **Olcay Kalkan**

Date of Signature

..... **23.07.2004**

Assist. Prof. Dr. Mustafa Aziz Altınkaya

Supervisor

Department of Electrical and Electronics Engineering

..... **23.07.2004**

Prof. Dr. F. Acar Savacı

Department of Electrical and Electronics Engineering

..... **23.07.2004**

Assist. Prof. Dr. Olcay Akay

Department of Electrical and Electronics Engineering

Dokuz Eylül University

..... **23.07.2004**

Prof. Dr. F. Acar Savacı

Head of Department

Department of Electrical and Electronics Engineering

ACKNOWLEDGEMENT

I would like to thank my supervisor Assist. Prof. Dr. Mustafa A. Altınkaya for his support and guidance throughout my research. Also I would like to thank Prof. Dr. F. Acar Savacı and Assist. Prof. Dr. Olcay Akay for serving on my thesis committee.

My friends in our department have helped me to make this thesis in success. I wish to express my thanks to Berna, Osman, Okan and Emre; they helped me greatly by taking care of many details and I would like to thank Erdal for his correction from "Guass" to "Gauss". Also I wish to give special thanks to Bora and Zuhal for their support during my thesis work, especially for their logistic support in my thesis defence.

Of course, I would like to thank my family. With their support and understanding there is nothing that I can't reach in my life.

ABSTRACT

Blind source separation (BSS) methods, independent component analysis (ICA) and independent factor analysis (IFA) are used for detecting the signal coming to a mobile user which is subject to multiple access interference in a CDMA downlink communication. When CDMA models are studied for different channel characteristics, it is seen that they are similar with BSS/ICA models. It is also showed that if ICA is applied to these CDMA models, desired user's signal can be estimated successfully without channel information and other users' code sequences. ICA detector is compared with matched filter detector and other conventional detectors using simulation results and it is seen that ICA has some advantages over the other methods.

The other BSS method, IFA is applied to basic CDMA downlink model. Since IFA has some convergence and speed problems when the number of sources is large, firstly basic CDMA model with ideal channel assumption is used in IFA application. With simulation of ideal CDMA channel, IFA is compared with ICA and matched filter.

Furthermore, Pearson System-based ICA (PS-ICA) method is used for estimating non-Gaussian multipath fading channel coefficients. Considering some fading channel measurements showing that the fading channel coefficients may have an impulsive nature, these coefficients are modeled with an α-stable distribution whose shape parameter α takes values close to 2 which makes the distributions slightly impulsive. Simulation results are obtained to compare PS-ICA with classical ICA. Also IFA is applied to the single path CDMA downlink model to estimate fading channel by using the advantage of IFA which is the capability to estimate sources with wide class of distributions.

ÖZ

Gözü kapalı kaynak ayrıştıma (BSS) yöntemlerinden bağımsız bileşen analizi (ICA) ve bağımsız etmen analizi (IFA), vericiden alıcıya kod bölüşümlü çoklu erişim (CDMA) iletişiminde kullanıcının sinyalinin diğer kullanıcıların girişim oluşturan sinyallerinden ayrıştırılması için kullanılmıştır. Farklı çokyollu sönümlemeli kanal karakteristikleri için CDMA modelleri incelendiğinde, oluşturulan modellerin BSS modelleri ile aynı olduğu görülmüştür. Bu CDMA modellerine ICA uygulandığında, kanal bilgisi ve kullanıcıların kod dizileri olmadan istenilen kullanıcının sinyallerinin kestirilebildiği görülmektedir. Benzetim çalışmaları yapılarak, BSS yöntemleri, uyumlu süzgeç alıcısı gibi klasik yöntemlerle karşılaştırılmış ve bu yöntemin avantajlarıyla öne çıktığı görülmüştür..

Bir diğer BSS yöntemi olan IFA da vericiden alıcıya basit CDMA modeline uygulanmıştır. Çok sayıda kaynak olan durumlarda IFA yavaş olduğu için ilk aşamada basit CDMA modeli kullanılmış ve benzetimler yoluyla ICA ve uyumlu süzgeç alıcısı ile karşılaştırılmıştır.

BSS için yeni bir yaklaşım olan Pearson Sistemi'ne dayalı ICA (PS-ICA) yöntemi, Gauss olmayan çokyollu sönümlemeli kanalın katsayılarının kestirilmesinde kullanılmıştır. Bu uygulamada, literatürde bulunan Gauss olmayan, dürtün özelliklere sahip kanallar üzerine yapılan çalışmalara dayanılarak, vericiden kullanıcıya sönümlemeli CDMA kanalı, şekil parametresi α dürtünce bir niteliği gösteren 2' ye yakın değerler alabilen α-kararlı bir dağılımla modellenmiştir. Benzetimler yoluyla PS-ICA yönteminin başarımı klasik ICA yönteminin kiyle karşılaştırılmıştır. Ayrıca tek yollu CDMA modeline IFA uygulanarak, bu yöntemde kaynakların Gauss karışımlarıyla modellendiğinden dolayı farklı dağılımlara sahip kaynakları ayrıştırabilmesindeki başarısından yararlanılmaya çalışılmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

BSS	Blind Source Separation
CDMA	Code Division Multiple Access
CMOE	Constrained Minimum Output Energy
DS-SS	Direct Sequence Spread Spectrum
EM	Expectation Maximization
FDMA	Frequency Division Multiple Access
ICA	Independent Component Analysis
IFA	Independent Factor Analysis
ISI	Inter Symbol Interference
KL	Kullback-Leibler
LMS	Least Mean Square
MAI	Multiple Access Interference
ML	Maximum Likelihood
MI	Mutual Information
MMSE	Minimum Mean Square Error
MOG	Mixture of Gaussians
PDF	Probability Density Function
POA	Probability of Acquisition
PS-ICA	Pearson System-based Independent Component Analysis
SNR	Signal to Noise Ratio
SUD	Single User Detector
SVD	Singular Value Decomposition
TDMA	Time Division Multiple Access

Chapter 1

INTRODUCTION

Code Division Multiple Access (CDMA) is a multiple access technique that uses spread spectrum modulation by each user with its own unique spreading code with all users sharing the same spectrum [1]. Spread spectrum modulation is done by pseudo-noise (PN) codes. By using PN codes, the narrowband information signal is modulated and spreaded to a substantially greater bandwidth prior to transmission. Digital cellular telephone system based on CDMA technology was proposed in 1990 by Qualcomm, Inc. In July 1993 it was adopted as a second U.S digital cellular standard, designated IS-95 [2]. The IS-95 system provided a very high capacity by using spread-spectrum signaling techniques. Third generation cellular systems are being designed to support wideband services like high speed internet access, video and high quality image transmission with the same quality as the fixed networks and one of the most promising approaches to 3G is to combine a Wideband CDMA (WCDMA) air interference with the fixed network of GSM. The 3G standard was created by the International Telecom Union (ITU) and is called IMT-2000 (International Mobile Telecommunications year 2000) based on CDMA (W-CDMA, CDMA 2000 and TD-SCDMA). 806-960 MHz, 1710-1885 MHz, 2110-2200MHz and 2500-2690 MHz frequency bands are separated by ITU for 3G mobile cellular communication systems.

In the CDMA downlink transmission in which information is transmitted from base stations to mobile units, symbol detection is an important problem in fading channels. In a downlink transmission the receiver is interested in only one of the information sources transmitted by the base stations [3]. In CDMA downlink problem, since desired user has not got any information about other users' spreading codes, separation of desired user's signals from other users' signals is a blind source separation (BSS) problem. Because of this, independent component analysis (ICA), one of the BSS methods, has been applied to CDMA models in last years.

In BSS problem, data measured by sensors arise from source signals that are mixed together by some linear transformation corrupted by noise. The task is to obtain

those source signals. However, the sources are not observable and there is no information about their properties beyond their statistical independence nor about the properties of the mixing process and the noise [4]. ICA is a statistical signal processing technique whose main application is BSS [5]. The sources can be found using ICA provided that all sources are independent and non-Gaussian [6]. Most of the work in the field of BSS since its emergence in the mid' 80s, in which ICA was used, aimed at idealized version of the problem where the mixing matrix is square and noiseless but in the mid' 90s a satisfactory solution was found (Bell and Sejnowski 1995; Cardoso and Laheld 1996; Pham 1996; Pearmuter and Parra 1997; Hyvarinen and Oja 1997). Algorithms for ICA assumes non-Gaussian models of the source densities and maximum likelihood (ML) estimate of mixing matrix is unique for chosen source densities.

One of the other solutions for BSS problem is independent factor analysis (IFA), which was proposed by Attias. In this method, each source model is described by a mixture of Gaussians and for learning of an associated probabilistic model of the mixing situation expectation-maximization (EM) algorithm is used.

In this thesis, the applications of ICA in CDMA downlink problem will be investigated. Simulation results will be shown to analyze the performance of these applications. Also, IFA is applied to CDMA. Simulation results will be given to compare ICA and IFA applications in CDMA systems.

Firstly, chapter 2 contains an overview about CDMA. Basic asynchronous model, asynchronous model in ideal channel and in multipath fading channels will be explained. Also in this chapter an overview about multipath fading channel will be given. In chapter 3, ICA will be explained. It contains objective functions ICA used which are classified as multi-unit and one-unit objective functions. A new technique, Pearson System-based ICA will be given in this chapter, too. Chapter 4 focuses on IFA. The IFA method and advantages of IFA over ICA will be given in this chapter. In Chapter 5, ICA is applied to detect desired user's symbols for basic CDMA downlink model and for CDMA transmission systems in multipath fading channel. Also in this chapter, FastICA is used for estimating multipath delays. Simulation results of these applications will be given too. Furthermore, IFA is applied and compared with ICA for basic CDMA downlink model. In Chapter 6, non-Gaussian fading coefficients of multipath channel were estimated by Pearson System-based ICA. The reason of

choosing Pearson System-based ICA and simulation results will be given in this chapter too.

Last chapter gives a discussion on ICA and IFA applications in CDMA downlink problem and also it contains some information about future work that can be worked on.

Chapter 2

CODE DIVISION MULTIPLE ACCESS

2.1 Multiaccess Communication

Multiaccess communication is a type of communication where several transmitters share a common channel like in the cases of a bus with multitaps, local area networks (LAN), cable television networks, mobile telephones transmitting to a base station, ground stations communicating with a satellite [3]. In these examples, the common point is that the receiver receives a superposition of signals sent by different transmitters and noise.

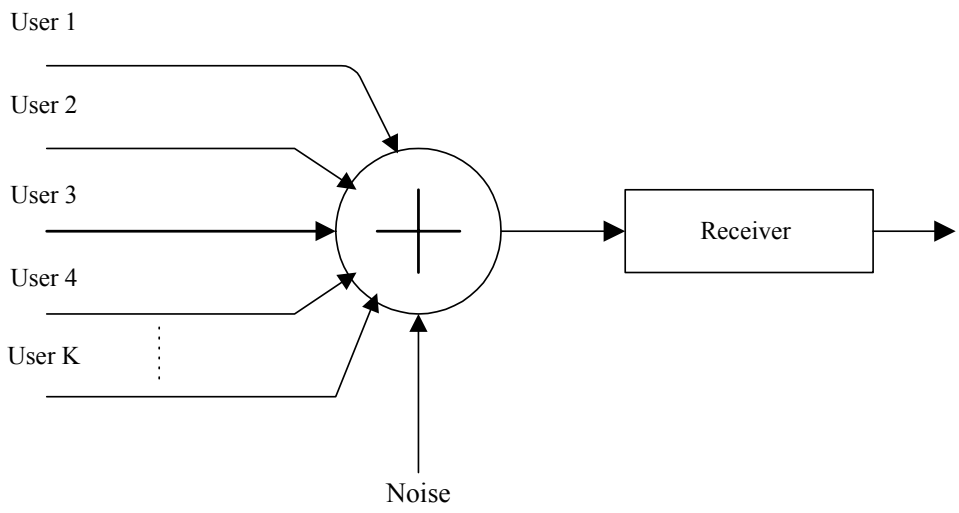

Figure 2.1 Multiaccess Communication

Message sources are not co-allocated and they can operate autonomously. For these reasons, the term “multiaccess” is used. In multiaccess communications, message sources are referred to as users.

Multiaccess communications is sometimes referred to as point to multipoint communication. Radio broadcasting and base station to mobile unit communication are the examples of this case. In the radio broadcasting, all receivers receive the same

information, but in the base station to mobile unit communication although all receivers receive the same information, one receiver or mobile unit is interested in only one of the information signals transmitted by the base station.

2.2 Comparison of CDMA with TDMA and FDMA

Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA) and Code Division Multiple Access (CDMA) are the three major multiaccess techniques used in communication systems. There are many extensions and hybrid techniques for these methods but an understanding of the three major methods is required firstly for understanding extensions of their methods [7]. Each multi-access technique has advantages and disadvantages related to the usage of the allowed bandwidth.

In TDMA, sending time is segmented into disjoint time slots as shown in Figure 2.2. All users are active for short periods of time on the same frequency [8]. Demodulation is carried out by simply switching on to the received signal at the appropriate epochs.

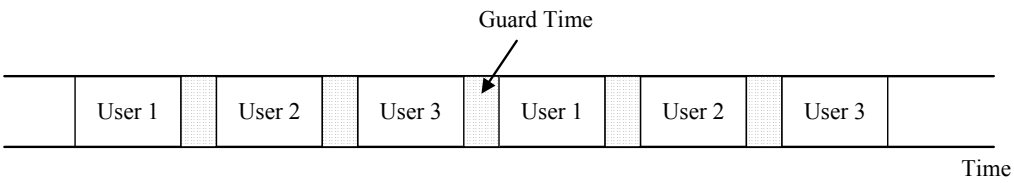

Figure 2.2 TDMA Communication

FDMA segments the frequency band into disjoint sub-bands and assigns a different carrier frequency to each user so that the resulting spectra do not overlap [3]. In FDMA each channel can be demodulated separately by using band-pass filtering in the frequency domain.

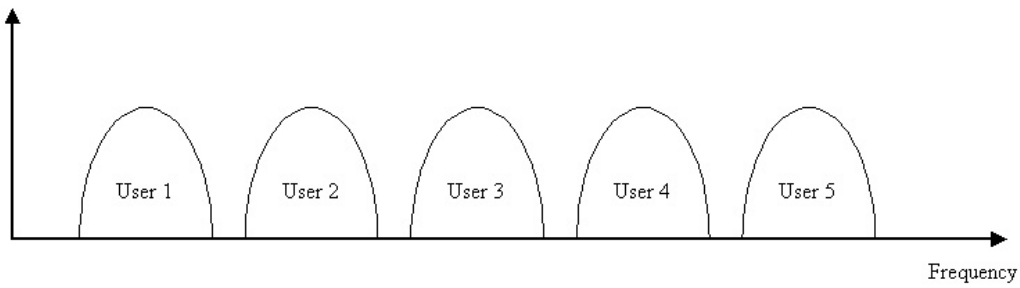

Figure 2.3 FDMA Communication

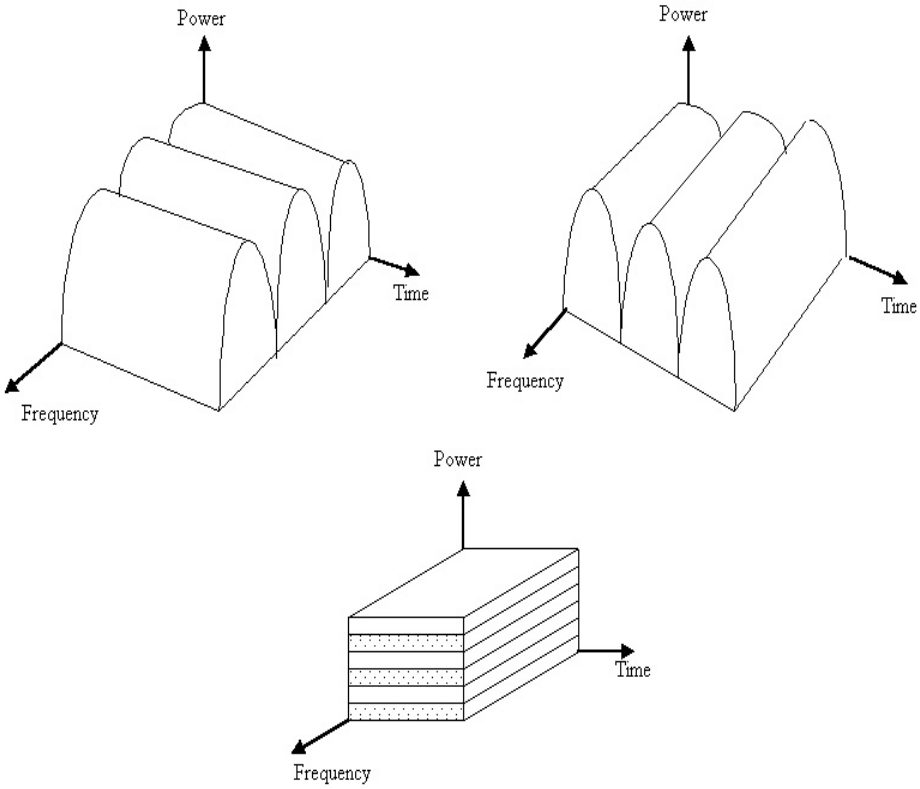

Figure 2.4 FDMA- TDMA- CDMA Communications

TDMA does not allow completely uncoordinated transmission and all transmitters and receivers must have access to a common clock [3]. Synchronization is very important in TDMA and it is difficult to synchronize all users. In FDMA, opposite to TDMA, no synchronization is required among the users. Because of nonideal effects of channel, guard time insertion in TDMA and spectral guard bands insertions in FDMA may be required as shown in Figures 2.2 and 2.3. Also

noninterfering multiaccess strategies may waste channel resources. Especially when the number of potential users is much greater than the number of simultaneous active users at any given time, wasting channel resources becomes an important problem. Because at these times, in TDMA most of the time slots are empty and in FDMA small part of spectrum is used.

In contrast to TDMA and FDMA, in CDMA, signals of users overlap both in time and frequency (Figure 2.7). All users send on the same frequency probably at the same time and they can use the whole bandwidth of the transmission channel. No planning and synchronization is required. However, it can be pointed as a disadvantage of CDMA that it requires more complex receiver structures. In the next sections of Chapter 2, CDMA will be given in more details.

2.3 CDMA Details

In CDMA, the narrowband message is multiplied by a large bandwidth signal, which is a pseudo random noise code (PN code) and these signals are called "code sequences", "user codes" or "signature sequences". When the users' codes are orthogonal, there is no interference between the users after despreading and the privacy of the communication of each user is protected. In practice, the codes cannot be orthogonal and cross-correlations between codes introduce performance degradation and this limits the number of active users. If signature waveforms' mutual interference is sufficiently low, the requirement that the waveforms be orthogonal can be dropped in CDMA. However selection of signature sequences is still very important since their cross-correlation must be very low. Non-orthogonal code sequences make CDMA a useful multiaccess technique for multiuser communication systems. As a result of this, users can be asynchronous and sharing of channel sources is dynamic since it depends on active users rather than on the number of potential users of the system.

The term "process gain", "spreading factor" or "bandwidth expansion factor" is one of the most important concepts in spread spectrum techniques. It points out to the signal to noise ratio (SNR) improvement as a result of spreading and despreading process. Process gain of a system is equal to the ratio of chip rate "R_C" to the data symbol rate "R_S".

$$N_C = \frac{BW_{SS}}{BW_{info}} = \frac{R_C}{R_S} = \frac{T_b}{T_C} \tag{2.1}$$

In Eq. (2.1), BW_{SS} is the bandwidth of the spreading data, BW_{info} is the bandwidth of the sent information data. T_C is the chip period of the spreading sequence, and T_b is the period of the transmitting signal.

Figure 2.5 shows a basic CDMA transmitter.

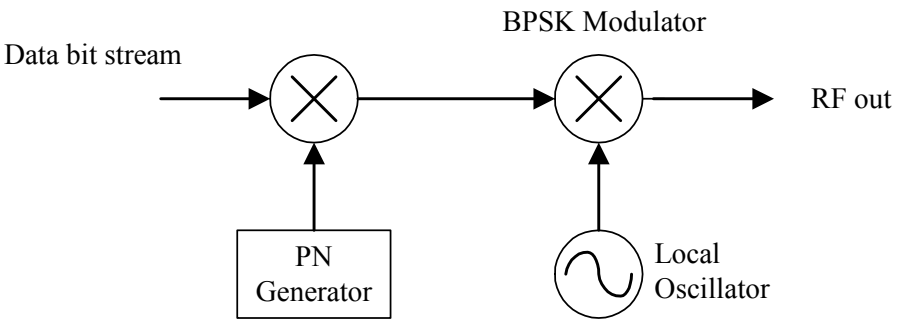

Figure 2.5 Basic CDMA Transmitter

As Figure 2.5 shows, data signal is modulated by the PN code which has a chip rate higher than the bit rate of the data. In Figure 2.6, the generation of a CDMA signal using a 10 chip length code is given.

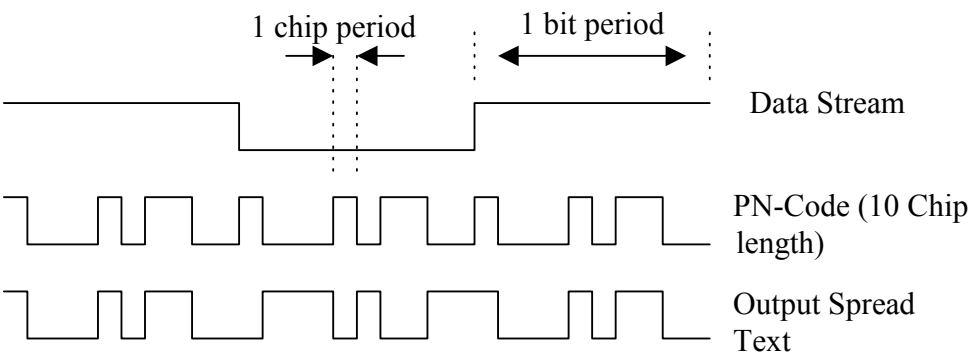

Figure 2.6 Generation of a CDMA signal using a 10 chip length code

The receiver side of the CDMA systems is more complex than other multiaccess system receivers. Because of non-orthogonal code sequences, simple correlator receivers cannot achieve good detection performances. As a consequence of

non-orthogonal signature waveforms, other users cause interference to the desired user called "Multiple Access Interference" (MAI). MAI affects the performance of the system and an interfering signal that is more powerful than the desired signal at a receiver which is the result of the near-far problem, causes high performance degradation. In addition to this problem, the characteristics of a wireless channel like multipath fading and propagation delay, affect the performance of CDMA systems. In subsection 2.3.3.2, the multipath channel will be given in more detail.

2.3.1 Signature Waveforms

In direct sequence spread spectrum (DS-SS) CDMA system, chip waveform and "spreading factor" or "processing gain" are common to all users. Also the selection of the code is important because the type and the length of the code set bounds on the system capability. In addition to this, as it was mentioned before, cross-correlations of codes must be sufficiently low. PN signature sequences, for given spreading factor, N, and number of users, K, achieve low cross-correlations for all possible offsets [3].

PN code sequence acts as a noise-like carrier used for bandwidth spreading of the signal energy. It is a pseudo-random sequence of 1's and 0's but not a real random sequence since it is periodic but it looks random for the user who doesn't know the code. The PN sequence code can be divided into two categories: short code and long code. In the short code, for each data symbol the same PN sequence is used but in the long code, the PN sequence period is much longer than the data symbol and so a different chip pattern is associated with each symbol [9].

PN sequence codes can be produced by many ways. Gold codes, Barker codes and m-sequence are examples of PN sequence codes. Since a large number of codes with the same length and low cross-correlations can be generated, Gold code sequences are often used in CDMA systems. So, in this thesis, Gold codes are used for spreading data in a CDMA downlink system.

2.3.2 Direct Sequence CDMA (DS-CDMA)

Direct sequence is a specific approach to construct spread spectrum waveforms.

If chip waveform is "P_{T_C}", number of chips per bit is "N" and binary sequence $p_t=\{c_1,c_2,...,c_N\}$ is used to modulate the chip waveform antipodally, the direct sequence spread spectrum waveform with duration NT_C is [3]:

$$s(t) = A\sum_{i=1}^{N}(-1)^{C_i} P_{T_C}(t-(i-1)T_C) \tag{2.2}$$

In the transmitter, the binary data d_t with symbol rate R_S is multiplied with the PN sequence p_t with chip rate R_C and transmitted baseband signal tx_b is produced:

$$tx_b = d_t \cdot p_t \tag{2.3}$$

Multiplication of d_t with p_t is spreading the baseband bandwidth R_S of d_t to a baseband bandwidth of R_C. In the receiver, the received baseband signal rx_b is multiplied with the PN sequence p_r. If $p_r = p_t$, tx_b is again produced and bandwidth of rx_b is despreaded to R_S. If $p_r \neq p_t$, the receiver cannot reproduce the transmitted data tx_b [9]. For BPSK modulation, the building blocks of DS-SS system are shown in Figure 2.7.

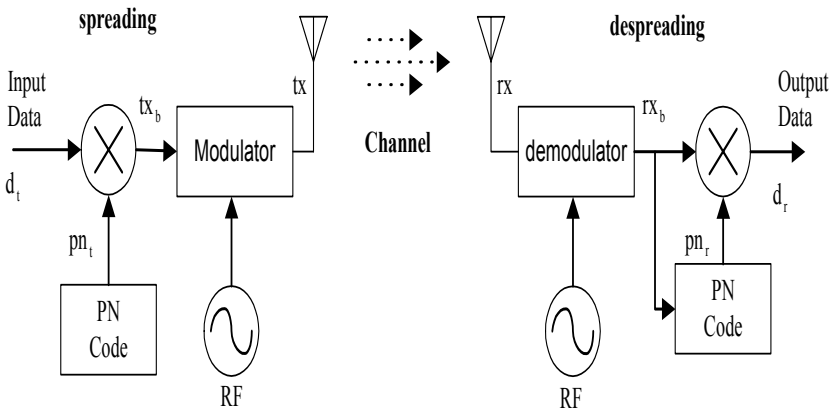

Figure 2.7 DS-SS System

2.3.2.1 Modulation

To modulate the input data d_t that has a bandwidth BW_{info}, it is multiplied with the PN sequence p_t and the transmitted data tx_b that has larger bandwidth BW_{SS} is produced [9].

$$BW_{\text{info}} \approx R_S << R_C \approx BW_{\text{SS}}$$

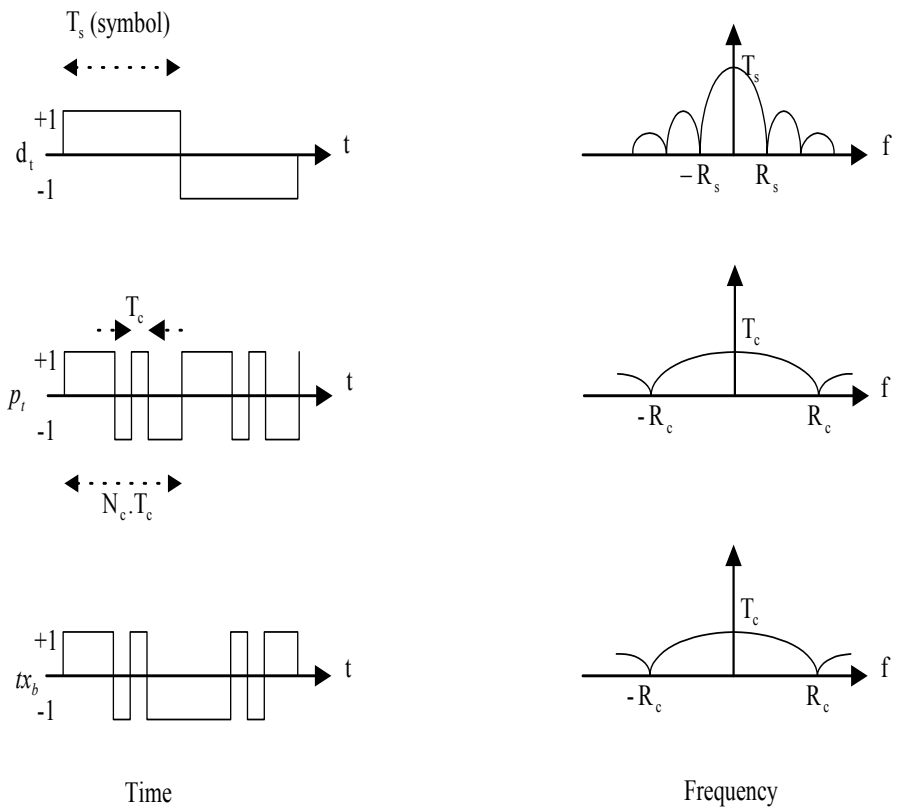

Figure 2.8 Modulation of data

2.3.2.2 Demodulation

To demodulate the received signal, rx_b is multiplied by the PN sequence p_r which is the same with p_t. This operation is called spectrum despreading. Figure 2.9 shows the demodulation of the received signal rx_b.

Then the receiver output, d_r, can be found as: (if $p_r = p_t$ and $p_r \cdot p_t = 1$ for all t)

$$d_r = rx_b \cdot p_r = (d_t \cdot p_t) \cdot p_t \tag{2.4}$$

$$d_r = d_t$$

If the received signal is multiplied by a PN sequence p_r which is different from p_t, the received output becomes:

$$d_r = rx_b \cdot p_r = (d_t \cdot p_t) \cdot p_r \tag{2.5}$$

For security in communication, the cross-correlation of p_t and p_r, which is given as

$$\text{Cross-correlation } R_C(\tau) = \left(p_t(t) p_r(t+\tau)\right) << 1 \text{ for all } \tau \tag{2.6}$$

As a consequence of Eq. (2.6), d_t cannot be reproduced by a user that does not know the PN sequence p_t. This orthogonality property of the allocated spreading codes means that the output of the correlator used in the receiver is approximately zero for all but the desired transmission [9].

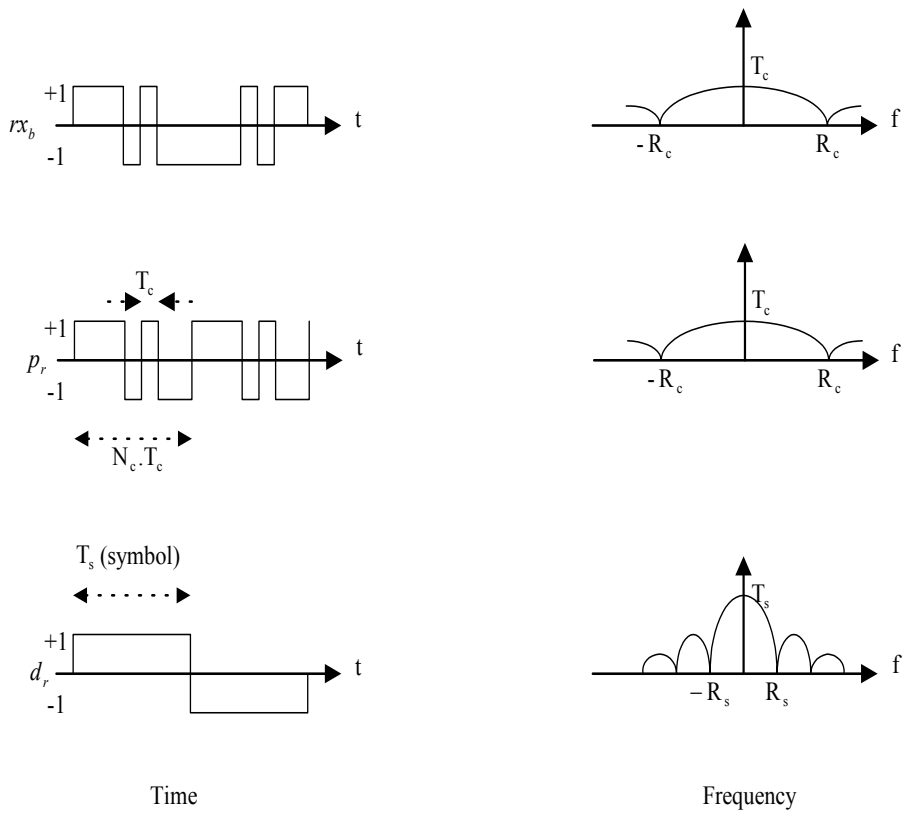

Figure 2.9 Demodulation of received signal

2.3.3 Basic CDMA Models

2.3.3.1 Basic Synchronous CDMA Model

A basic synchronous CDMA model for K users in an ideal channel consists of modulated synchronous signature waveforms embedded in additive white Gaussian noise (AWGN) [3]:

$$y(t) = \sum_{k=1}^{K} A_k b_k s_k(t) + \sigma n(t) \quad t \in [0,T] \tag{2.7}$$

In this model, $s_k(t)$ is the signature waveform of the k'th user and assumed to be zero outside the interval $[0,T]$, A_k is the amplitude of the k'th user's received signal, $b_k \in \{-1,+1\}$ is the transmitted bit of the k'th user and $n(t)$ is the AWGN with unity power spectral density. For special case of two users, Eq. (2.7) can be rewritten as:

$$y(t) = A_1 b_1 s_1(t) + A_2 b_2 s_2(t) + \sigma n(t). \tag{2.8}$$

2.3.3.2 Basic Asynchronous CDMA Model

In the asynchronous case the stream of bits must be taken into account opposite to one-shot synchronous CDMA model:

$$b_k[-M],\ldots\ldots,b_k[0],\ldots\ldots b_k[M]. \tag{2.9}$$

The CDMA model becomes:

$$y(t) = \sum_{k=1}^{K} \sum_{i=-M}^{M} A_k b_k[i] s_k(t - iT - \tau_k) + \sigma n(t). \tag{2.10}$$

In this equation, τ_k is the offset of the k'th user and the data rate, $1/T$, is equal for all users. In Figure 2.10, bit epochs are shown for three asynchronous users in a case

where $M=1$. If all offsets are identical ($\tau_1 = \tau_2 = \cdots = \tau_K$), the asynchronous CDMA model is reduced to a synchronous CDMA model.

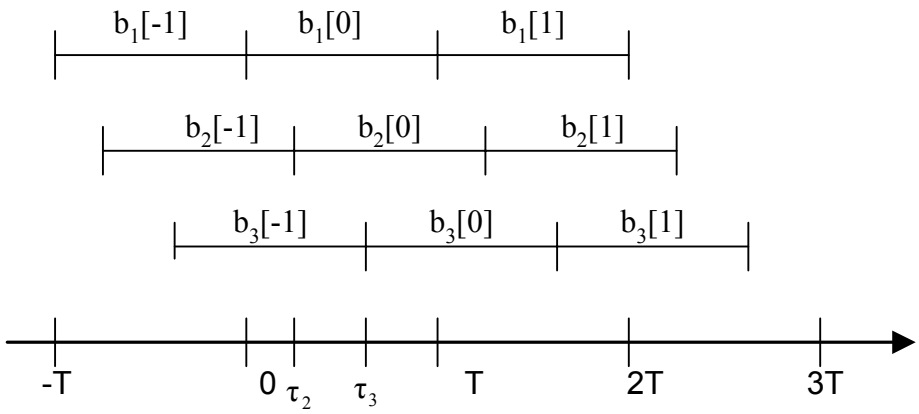

Figure 2.10 Bit epochs for K=3 and M=1 [3]

2.3.3.3 CDMA Downlink Model in a Multipath Channel

In transmission systems, the received signal consists of a combination of attenuated, reflected and refracted replicas of the transmitted signal and noise. If the receiver or the transmitter is moving, there can be a carrier frequency shift called "Doppler Effect". The amount of frequency change due to the Doppler Effect depends on the relative motion of the source and the receiver and on the speed of propagation of the wave.

In multipath channels, signals from transmitters may be reflected from objects such as hills, buildings or moving objects like vehicles. Figure 2.11 shows some of these possible reflection waves.

Because of the extra path length, the reflected signals arrive at a later time than the direct signal. This situation spreads the received energy. The time spread between the arrival of the first and last multipath seen by the receiver is called "delay spread" and it can lead to intersymbol interference (ISI). In Figure 2.12, multipath "delay spread" is shown.

Small scale fading or simply fading is used to describe the rapid fluctuations of the amplitude of a radio signal over a short period of time or travel distance [10]. If the channel possesses a constant gain and a phase response over a bandwidth that is smaller than the bandwidth of the transmitted signal, the fading depends on frequency

and it is called “frequency selective fading” else it is called “frequency flat fading” [10].

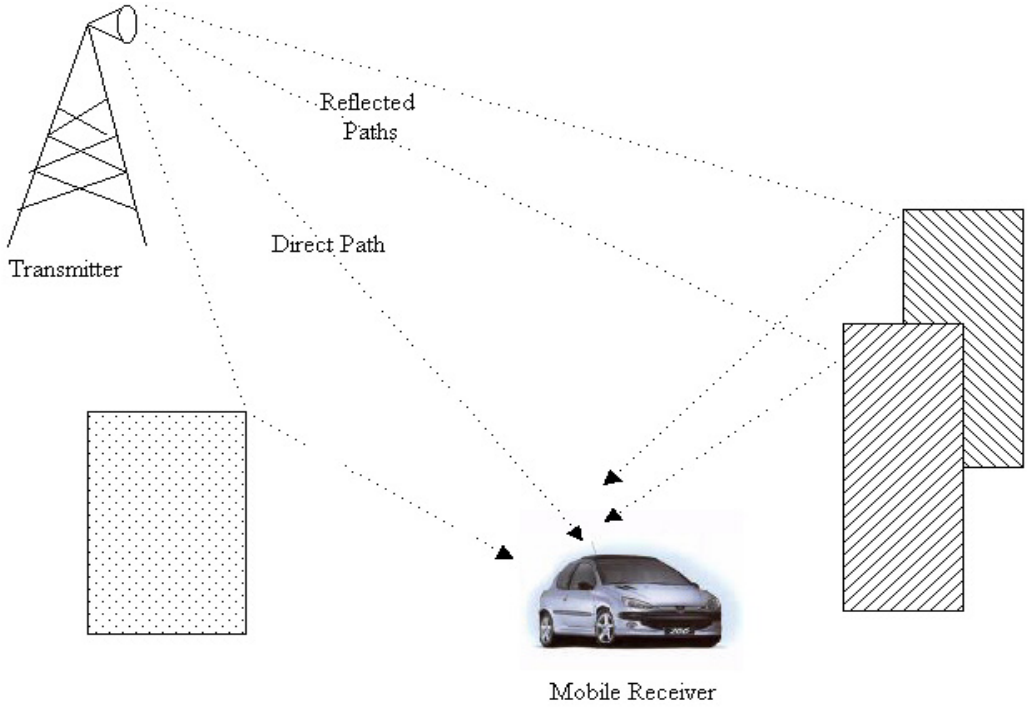

Figure 2.11 Reflection of the signal

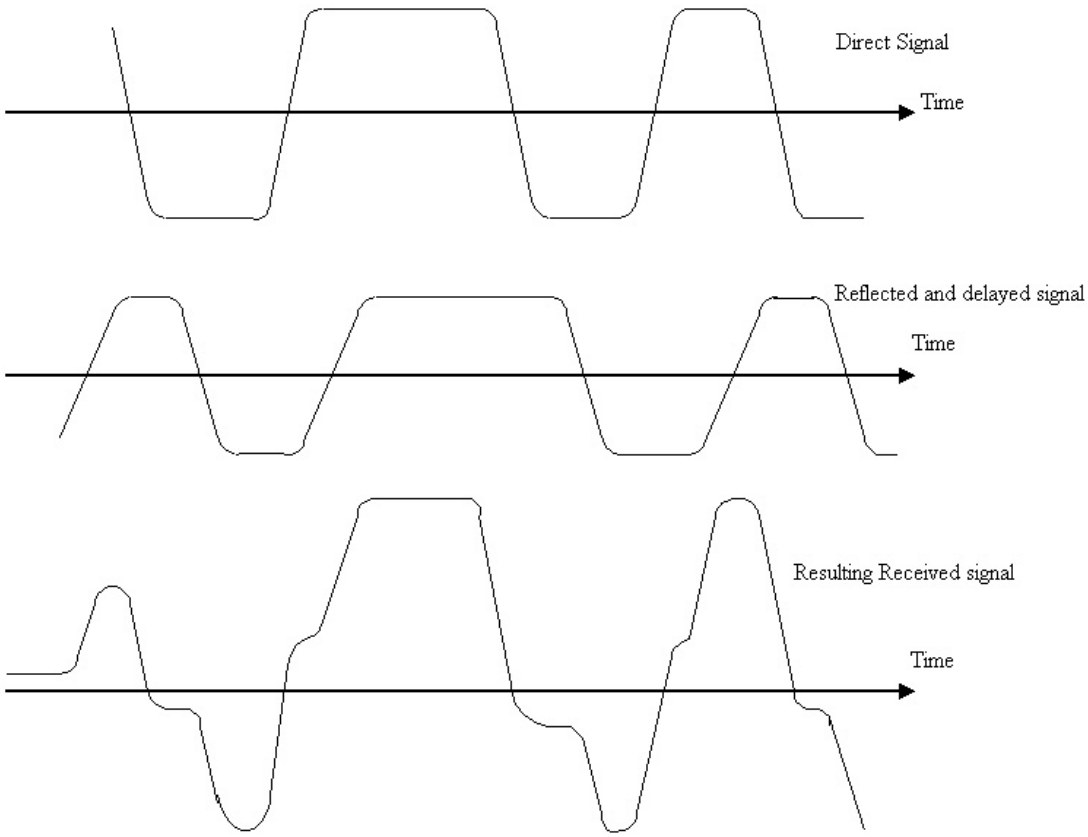

Figure 2.12 Multipath delay spread

When fading is studied in the time domain, channel may be classified as "fast fading" or "slowly fading" (Figure 2.13). In a fast fading channel, the channel impulse response changes within the symbol duration but in a slowly fading channel, channel impulse response changes at a rate much slower than that of the transmitted baseband signal.

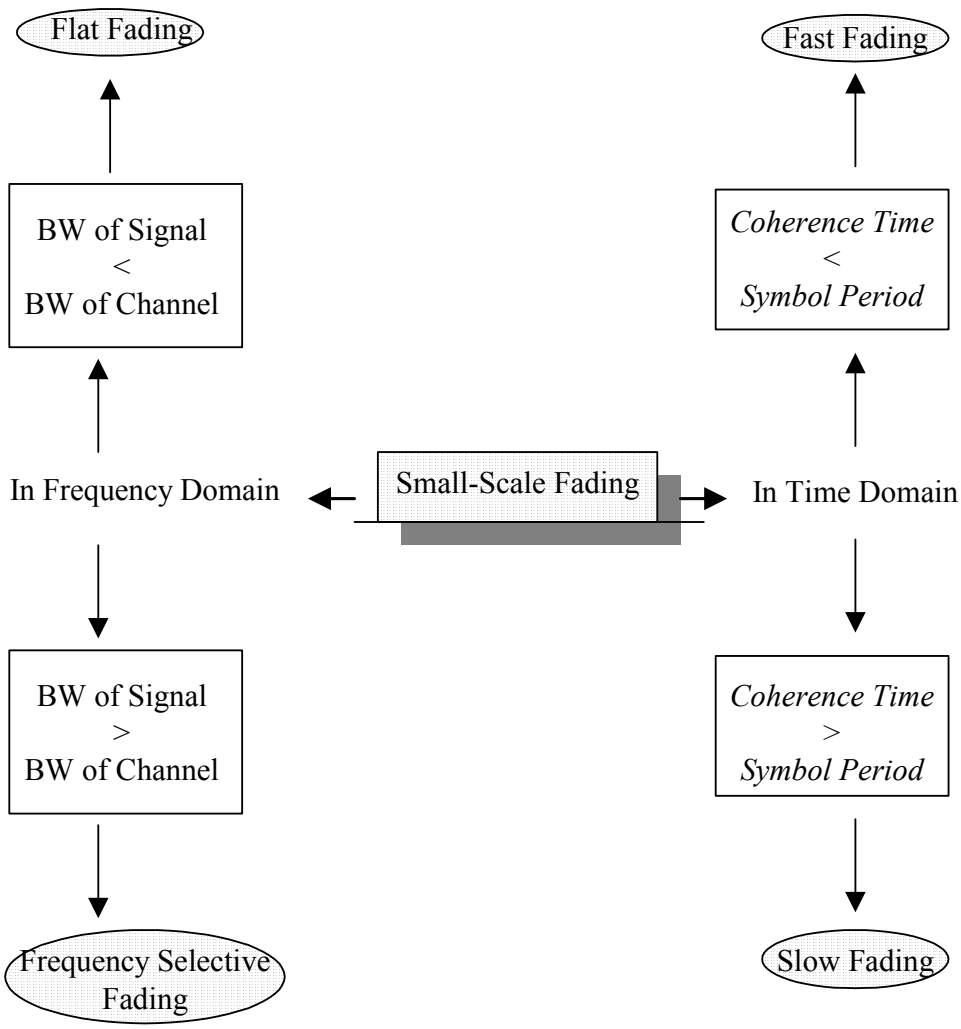

Figure 2.13 Fading in time and frequency domain

Rayleigh distribution and Rice distribution are commonly used in modeling fading channels. Rayleigh fading arises when the real and the imaginary parts of the channel fading coefficient are independent zero mean Gaussian random processes, in which case the phase of the coefficient is uniformly distributed on [0, 2π]. If the real and imaginary parts of coefficients are independent Gaussian random processes with non-zero means, Rician fading arises [11]. In the second case, there is a direct line of

sight between transmitter and receiver. But recent works show that some channel models can be non-Rayleigh and sometimes the real and imaginary parts of fading coefficients can deviate from a Gaussian process [12, 13 and 14]. In the following sections, these kinds of channel models will be explained.

If the multipath strength remains the same during the observation interval, which leads to a slow fading case, the downlink data in the observation interval have the form [3]:

$$r(t) = \sum_{m=1}^{N} \sum_{k=1}^{K} b_{km} \sum_{l=1}^{L} a_l s_k (t - mT - d_l) + n(t) . \tag{2.11}$$

In this model, K simultaneous users send N symbols via L independent paths. The strengths of each path is denoted by a_l. b_{km} is the k'th user's m'th symbol, $s_k(t)$ is the k'th user's binary chip spreading sequence in the interval $[0,T]$ where T is symbol duration and $s_k(t) \in \{-1,+1\}$. Each path's delay is denoted by d_l and it is assumed to change slowly for long time durations and be constant during the observation time. $n(t)$ denotes the AWGN.

If fading is fast and changes symbol by symbol, Eq. (2.11) can be rewritten as:

$$r(t) = \sum_{m=1}^{N} \sum_{k=1}^{K} b_{km} \sum_{l=1}^{L} a_{lm} s_k (t - mT - d_l) + n(t) \tag{2.12}$$

where a_{lm} is the fading factor of the l'th path corresponding to m'th symbol.

The two Equations (2.11) and (2.12) are the models for CDMA downlink communication in fast fading and slowly fading channels used in this thesis.

Chapter 3

INDEPENDENT COMPONENT ANALYSIS (ICA)

Independent component analysis is a statistical signal processing technique whose goal is to express a set of random variables as linear combinations of statistically independent component variables [15]. In order to distinguish the independent signals in a mixture, the separation method must take all higher order correlations into consideration. ICA does take these higher order correlations into account [6]. BSS, feature extraction, blind deconvolution are important applications of ICA. In BSS problem, data measured by sensors arise from source signals that are mixed together by some linear transformation corrupted by noise. The task is to obtain those source signals. The sources can be found using ICA provided that all sources are independent and non-Gaussian [6]. If the ICA model in Eq. (3.4) is considered, the goal is to find a matrix **W** from observations such that the output **y** is an estimate of source vectors. In Figure 3.1, this model is given.

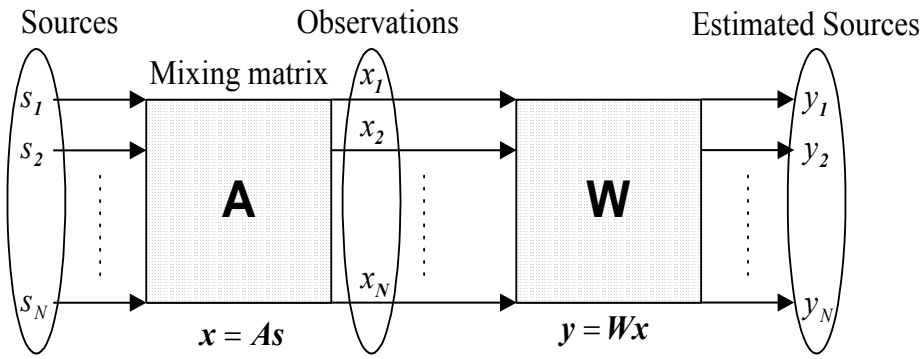

Figure 3.1 ICA Model

3.1 Statistical Independence

To understand ICA, firstly the term "statistical independence" must be known. Denote by $y_1, y_2, ..., y_m$ some random variables with zero mean and joint density $f(y_1, y_2, ..., y_m)$. If the density function can be factorized as:

$$f(y_1....y_m) = f_1(y_1) f_2(y_2).....f_m(y_m), \tag{3.1}$$

it is said that the variables y_i's are mutually independent. In Eq. (3.1), $f_i(y_i)$ denotes the marginal density of y_i.

Independence must be distinguished from uncorrelatedness. Independence is a much stronger requirement than uncorrelatedness [16].

If y_i's are uncorrelated,

$$E\{y_i y_j\} - E\{y_i\}E\{y_j\} = 0 \quad \text{for} \quad i \neq j. \tag{3.2}$$

If y_i's are independent,

$$E\{g_1(y_i) g_2(y_j)\} - E\{g_1(y_i)\}E\{g_2(y_j)\} = 0 \quad \text{for} \quad i \neq j$$

for any measurable functions g_1 and g_2. This is a stronger condition than the condition of uncorrelatedness. If two variables are independent, they are also uncorrelated but uncorrelatedness does not imply independency [17]. However there is one case in which uncorrelatedness and independence are equivalent. This is the case when $y_1, y_2, ..., y_m$ have a jointly Gaussian distribution [16]. Because of this ICA cannot separate source that has a Gaussian distribution.

3.2 Definition of ICA

In the literature there are at least three definitions of linear ICA. In the definitions given below, the observed m-dimensional random vector is denoted by $\mathbf{x}=(x_1, x_2 ... x_m)^T$ [16].

Definition 1: (General Definition) ICA of the random vector **x** consists of finding a linear transformation **s** = **Wx** so that the components *si*'s are as independent as possible in the sense of maximizing some function $F(s_1,s_2,.....,s_m)$ that measures independence.

Definition 2: (Noisy ICA Model) ICA of random vector **x** consists of estimating the following generative model for the data:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} \tag{3.3}$$

where the latent variables (components) $\{s_i,\ i=1,...,n\}$ in the vector $\mathbf{s} = (s_1,s_2,.....s_n)^T$ are assumed to be independent. The matrix **A** is a constant $m \times n$ "mixing matrix" and **n** is an m-dimensional random noise vector.

Definition 3: (Noise-free ICA Model) ICA of the random vector **x** consists of estimating the following generative model for the data.

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{3.4}$$

where **A** and **s** are the same as in Definition 2.

In Definition 3, the noise vector has been omitted. This definition is introduced by Jutten and Herault and it was probably the earliest explicit formulation of ICA [16]. In addition to the basic assumption of statistical independence, there are some fundamental restrictions. These are [16]:

1. All the independent components $\{s_i\}$ with the possible exception of one component must be non-Gaussian.
2. The number of observed linear mixtures, m, must be at least as large as the number of independent components n, i.e. $m \geq n$.
3. The mixing matrix **A** must have full column rank.

Non-Gaussianity of independent components is necessary for ICA. Because for Gaussian random variables, uncorrelatedness implies independence and any decorrelating representation would give independent components. The second restriction, $m \geq n$, is not completely necessary. Some recent works on the case $m<n$, called "ICA with overcomplete bases" can be found. Lastly third restriction about rank of mixing matrix **A** is also necessary for ICA.

In Eq. (3.4), by using ICA, independent components s_i and mixing matrix **A**'s columns can be estimated up to a multiplicative constant but it is an insignificant indeterminacy because any constant multiplying an independent component in Eq. (3.4) could be canceled by dividing the corresponding column of the mixing matrix **A** by the same constant. In addition to this, generally it is assumed that independent components $\{s_i,\ i=1,\ldots,n\}$ have unity variances. So, independent components are unique up to a multiplicative sign.

There is no ordering in estimating independent components in ICA but it is possible to introduce ordering in estimation by using norms of mixing matrix's columns or by using non-Gaussianity between the independent components.

The estimation of the data model of ICA is usually performed by formulating an objective function and then minimizing or maximizing it [16]. This objective function is called as "contrast function". The terms "loss function" or "cost function" are also used for objective functions.

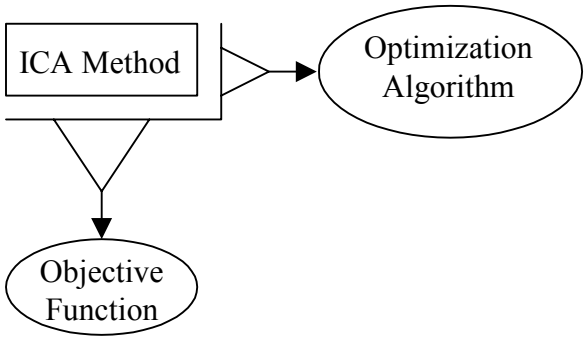

Figure 3.2 Summary of ICA Model

The choice of objective functions affects the statistical properties like consistency, asymptotic variance, robustness where the algorithmic properties like convergence speed, memory requirements, stability depend on the optimization algorithm.

Some ICA methods estimate the whole data model or all of the independent components at the same time. Others estimate one independent component at a time. The contrast functions of the first group are categorized as multi-unit contrast functions where the ones of the second group are one-unit contrast functions.

3.3 Objective Functions for ICA

3.3.1 Multi-unit Contrast Functions

In the literature, there are numerous multi-unit contrast functions but in this section major ones, which are likelihood and mutual information, will be depicted.

3.3.1.1 Likelihood and Network Entropy

Denoting by $\mathbf{W}=(w_1,w_2,.....,w_m)^T$ the matrix $\mathbf{A}^{-1}$, the log-likelihood of the noise-free ICA model in Eq. (3.4) takes the form [16]:

$$L = \sum_{k=1}^{K}\sum_{i=1}^{m} \log f_i\left(w_i^T x(k)\right) + K \ln\left|\det \mathbf{W}\right| \tag{3.5}$$

where f_i is the density function of the independent component s_i and it is assumed to be known and $x(k)$, $k = 1,.....,K$, is the realization of $\mathbf{x}$ [16].

Another contrast function is based on maximizing the output entropy of a neural network with non-linear outputs. If it is assumed that $\mathbf{x}$ is the input to the neural network whose output is of the form $g_i(\mathbf{w}^T\mathbf{x})$ where g_i's are non-linear functions and $\{\mathbf{w}_i,\ i=1,...,m\}$ are the weight vectors of the neurons, entropy of the output can be written as:

$$L_2 = H(g_1(\mathbf{w}_1^T\mathbf{x}),.....,g_m(\mathbf{w}_m^T\mathbf{x})) \tag{3.6}$$

where $H(\cdot) = -\int f(\cdot)\log f(\cdot)d(\cdot)$ denotes the differential entropy.

It is an interesting result that if g_i is chosen as the cumulative distribution function corresponding to f_i i.e. $g_i'(\cdot) = f_i(\cdot)$, network entropy maximization is equivalent to maximum likelihood (ML) estimation.

The advantage of the ML contrast function is its asymptotical efficiency but the requirement of the knowledge of the probability densities is the disadvantage of this approach. A second drawback is that the ML approach is sensitive to outliers.

3.3.1.2 Mutual Information

Mutual information is the most satisfactory contrast function in the multi-unit case. Mutual information "I" between m random variables $\{y_i, i=1,.....,m\}$ is found as follows [16]:

$$I(y_1,....,y_m) = \sum_i H(y_i) - H(\mathbf{y}). \tag{3.7}$$

The mutual information is a measure of dependence between two random variables. If random variables are independent, mutual information is equal to zero. Finding a transform that minimizes the mutual information between the components $\{s_i, i=1,...,m\}$ is a very natural way of estimating the ICA model. If $\mathbf{y} = \mathbf{W}\mathbf{x}$ where $\mathbf{x}$ is the observation matrix, Eq. (3.7) can be rewritten as:

$$I(y_1....y_m) = \sum_i H(y_i) - H(\mathbf{x}) - \log|\det \mathbf{W}|. \tag{3.8}$$

The use of mutual information can be motivated by using Kullback-Leibler (KL) divergence that can be considered as a kind of distance between the two probability densities:

$$\delta(f_1, f_2) = \int f_1(\mathbf{y}) \log \frac{f_1(\mathbf{y})}{f_2(\mathbf{y})} d\mathbf{y} \tag{3.9}$$

The independence of $\{y_i, i=1,...,m\}$ can be measured by finding KL divergence between $f(\mathbf{y})$ and the factorized density $\hat{f}(\mathbf{y}) = f_1(y_1).....f_m(y_m)$ where $\{f_i(\cdot), i=1,...m\}$ are marginal densities of the $\{y_i, i=1,...,m\}$. This quality is also equal to mutual information of y_i .

3.3.2 One-unit Contrast Functions

The expression "one-unit contrast function" is used to designate any function whose optimization enables estimation of a single independent component. In many

applications like the CDMA downlink problem, there is no need to estimate all the independent components. This reduces the computational complexity of the method if the input data has a high dimension. If one needs to estimate other independent components, they can be estimated one by one by using a simple decorrelator since independent components are by definition uncorrelated.

Negentropy and higher order cumulants are the major one-unit contrast functions used for estimating one independent component.

3.3.2.1 Negentropy (Negative Normalized Entropy)

The negentropy J is defined as:

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{Gauss}}) - H(\mathbf{y}) \tag{3.10}$$

where $\mathbf{y}_{\text{Gauss}}$ is a Gaussian random vector of the same covariance matrix as $\mathbf{y}$. Negentropy is always non-negative and it is zero if and only if $\mathbf{y}$ has a Gaussian distribution. When mutual information is expressed using negentropy:

$$I(y_1 .....y_m) = J(\mathbf{y}) - \sum_i J(y_i) + \frac{1}{2}\log\frac{\prod_{i=1}^{m}\mathbf{C}_{ii}^{y}}{\det\mathbf{C}^{y}} \tag{3.11}$$

where $\mathbf{C}^{\mathbf{y}}$ is the covariance matrix of $\mathbf{y}$ and $\mathbf{C}_{ii}^{y}$ is its i'th diagonal element. If the $\{y_i\}$ are uncorrelated the third term in Eq. (3.11) becomes zero and mutual information can be written as:

$$I(y_1 .....y_m) = J(\mathbf{y}) - \sum_i J(y_i)\,. \tag{3.12}$$

Finding maximum negentropy directions is equivalent to finding directions where the elements of the sum $J(y_i)$ are maximized and also it is equivalent to finding a representation in which mutual information is minimized.

Since the estimation of negentropy is difficult, it can be approximated by higher order cumulants like:

$$J(\mathbf{y}) \approx \frac{1}{12}\kappa_3(\mathbf{y})^2 + \frac{1}{48}\kappa_4(\mathbf{y})^2 \tag{3.13}$$

where $\kappa_i(\mathbf{y})$ is the *i*'th order cumulant of **y** but this approximation is too sensitive to outliers in many cases.

3.3.2.2 Higher Order Cumulants

Mathematically the simplest one-unit contrast functions are provided by higher order cumulants. Kurtosis is one of them and it is the mostly used function in applications (See Appendix A). The kurtosis of a Gaussian variable is equal to zero while the kurtosis of a non-Gaussian is nonzero. Therefore kurtosis can be used as a measure of the non-Gaussianity of a variable. For the ICA model given in Eq. (3.4), **x** is the observed data vector. To find one independent component, kurtosis of a linear combination of the observation $\{x_i,\ i=1,\ldots,N\}$, $\mathbf{w}^T\mathbf{x}$, is maximized or minimized. It is assumed that $E\{(\mathbf{w}^T\mathbf{x})^2\}=1$. Using the unknown mixing matrix **A**, **z** is defined such that $\mathbf{z}=\mathbf{A}^T\mathbf{w}$. Then using that data model $\mathbf{x}=\mathbf{A}\mathbf{s}$ one obtains $E\{(\mathbf{w}^T\mathbf{x})^2\} = \mathbf{w}^T\mathbf{A}\mathbf{A}^T\mathbf{w} = \|\mathbf{z}\| = 1$ ($E(\mathbf{s}\mathbf{s}^T)=\mathbf{I}$) and well known properties of kurtosis give:

$$\mathrm{kurt}(\mathbf{w}^T\mathbf{x}) = \mathrm{kurt}(\mathbf{w}^T\mathbf{A}\mathbf{s}) = \mathrm{kurt}(\mathbf{z}^T\mathbf{s}) = \sum_{i=1}^{m} z_i^4 \mathrm{kurt}(s_i) \tag{3.14}$$

Minimizing or maximizing the kurtosis in Eq. (3.14) under the constraint $\|\mathbf{z}\|=1$, one obtains one of the independent components as $\mathbf{w}^T\mathbf{x} = \pm s_j$. The drawback of kurtosis is that it is sensitive to outliers in the data.

3.3.2.3 General Contrast Functions

The generalized contrast functions, which can be considered as a generalization of kurtosis, have statistically appealing properties. They require no prior knowledge of

the densities of the independent components and allow a simple algorithmic implementation and they are simple to analyze.

One intuitive interpretation of contrast functions is that they are measures of non-Gaussianity. A family of such measures of non-normality could be constructed using practically any function G and considering the difference of the expectation of G for the actual data and the expectation of G for Gaussian data [16].

A contrast function J that measures the non-normality of a zero-mean random variable y using any even, non-quadratic, sufficiently smooth function G is given as follows:

$$J_G(y) = | E_y\{G(y)\} - E_v\{G(v)\} |^p \tag{3.15}$$

where v is a standardized Gaussian random variable, y is assumed to be normalized to unity variance and $p \in \{1,2\}$. For $G(y) = y^4$, J_G is simply the modulus of the kurtosis of y.

For suitable choice of G, the statistical properties of the estimator (asymptotic variance and robustness) are considerably better than the properties of the cumulant based estimators. The choices of G are:

$$\begin{aligned} G_1(\mathbf{u}) &= \log\cosh a_1 \mathbf{u} \\ G_2(\mathbf{u}) &= \exp(-\frac{a_2 \mathbf{u}^2}{2}) \end{aligned} \tag{3.16}$$

where a_1 , $a_2 \geq 1$ are suitable constants.

After choosing one of the contrast functions for ICA, a practical algorithm is needed to optimize the chosen contrast function. In ICA, before applying an optimization algorithm, some preprocessing should be done to make the ICA estimation simpler and that the convergence is faster.

3.4 Preprocessing of Data

In order to make a random vector, $\mathbf{x}$, zero-mean, the mean of $\mathbf{x}$, $\mathbf{m}=E\{\mathbf{x}\}$ is substracted from $\mathbf{x}$. In this way, there is no information loss in sources since after

estimating the mixing matrix **A** with ICA, the mean vector of **s** can be estimated by $\mathbf{A}^{-1}\mathbf{m}$. After making the data a zero-mean vector, second preprocessing step is whitening or sphering which means making the data uncorrelated and of unity variance [17]. By applying a linear transformation **V** on the data whitening can be done. If the original data is **x** with zero mean, then the whitened data $\mathbf{z} = \mathbf{V}\mathbf{x}$ has the property:

$$E\left[\mathbf{z}\mathbf{z}^T\right] = \mathbf{I}. \tag{3.17}$$

By applying eigenvalue decomposition to $\mathbf{R}_x = E\left[\mathbf{x}\mathbf{x}^T\right]$, **V** can be found as [17]:

$$\mathbf{R}_x = \mathbf{E}\mathbf{D}\mathbf{E}^T. \tag{3.18}$$

Then choose **V** as:

$$\mathbf{V} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T. \tag{3.19}$$

This results in:

$$\begin{aligned} E\left[\mathbf{z}\mathbf{z}^T\right] &= E\left[\mathbf{V}\mathbf{x}\mathbf{x}^T\mathbf{V}^T\right] \\ &= \mathbf{V}E\left[\mathbf{x}\mathbf{x}^T\right]\mathbf{V}^T \\ &= \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{E}\mathbf{D}\mathbf{E}^T\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T \\ &= \mathbf{I} \end{aligned} \tag{3.20}$$

where **E** is an orthonormal matrix.

This whitening operation also makes the mixing matrix orthonormal under the assumption of sources with unit variance ($\tilde{\mathbf{A}} = \mathbf{V}\mathbf{A}$), so:

$$E\left[\mathbf{z}\mathbf{z}^T\right] = \tilde{\mathbf{A}}\underbrace{E\left[\mathbf{s}\mathbf{s}^T\right]}_{\mathbf{I}}\tilde{\mathbf{A}}^T = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I} \tag{3.21}$$

After preprocessing operation of observed data, by using practical algorithms, the chosen contrast function can be optimized. The Jutten-Herault algorithm, non-linear decorrelating algorithms, non-linear PCA algorithms and least squares type algorithms are the most known algorithms for optimizing contrast functions but the fastest and simplest one is FastICA algorithm which is based on a fixed-point iteration [16,18].

3.5 FastICA Algorithm

FastICA is a general algorithm that can be used to optimize both one-unit and multi-unit contrast functions. Basic advantages of FastICA are [16]:

1. Algorithm is easy to use.
2. It does not need to estimate the pdfs first.
3. Its convergence is cubic.

Some disadvantages of FastICA are:

1. The choice of the non-linearity used in this algorithm, affects the stability of Fast ICA in estimating original signals.
2. It always has a certain lower bound limit which is the best possible estimate by FastICA.

For sphered data, the one-unit FastICA algorithm for maximization of non-Gaussianity by finding a direction, i.e. unit vector **w,** such that $\mathbf{w}^T\mathbf{x}$ maximizes non-Gaussianity, has the following form [16]:

$$\mathbf{w}(k) = E\{\mathbf{x}g(\mathbf{w}(k-1)^T\mathbf{x})\} - E\{g'(\mathbf{w}(k-1)^T\mathbf{x})\}\mathbf{w}(k-1) \tag{3.22}$$

where **w** is a weight vector and it is normalized to unit norm after every iteration. *g* is the derivative of the functions *G* in Eqs. (3.15) and (3.16) (Short derivation of the basic FastICA Algorithm is given in Appendix B).

In the FastICA algorithm, independent components can be estimated one by one by using a decorrelator or they can be estimated simultaneously by using a symmetric decorrelator. To estimate several independent components, FastICA algorithm is needed to be run several times with vector $\mathbf{w_1}$,..., $\mathbf{w_n}$ to prevent vectors from converging to the same maxima, so in every iteration each of them must be orthogonolized.

The simplest way to orthogonalize the vector $\mathbf{w_p}$ is to orthogonalize all previous vectors $\mathbf{w_1}$,....., $\mathbf{w_{p-1}}$ and apply orthogonalization at every iteration:

$$w_p \leftarrow w_p - \sum_{j=1}^{p-1}(w_p^T w_j)w_j \tag{3.23}$$

after estimating *p-1* independent components. Sometimes, it may be desirable to estimate all independent components in parallel. The symmetric orthogonalization of **W** is accomplished by the classical method involving square roots:

$$\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^T)^{-1/2}\mathbf{W}. \tag{3.24}$$

The steps of FastICA can be summarized as [16]:

1. Center the data to make it zero-mean.
2. Whiten the data to get **z**.
3. Choose initial vector **w** of unit norm.
4. Let $\mathbf{w} \leftarrow E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\} - E\{g'(\mathbf{w}^T\mathbf{z})\}\mathbf{w}$ where

   $g_1(\mathbf{y}) = \tanh(a_1\mathbf{y}) \quad g_2(\mathbf{y}) = \mathbf{y}(\exp(-\mathbf{y}^2/2)) \quad g_3(\mathbf{y}) = \mathbf{y}^3$ where $1 \leq a_1 \leq 2$
5. **w** = **w**/norm(**w**)
6. If not converged, go back to step 4.

3.6 Pearson System-Based ICA

By using ICA, independent sources are separated from their linear mixtures using only observed signals. In BSS, there is only limited information about sources. Distribution of sources should have a non-Gaussian distribution to be separated. Pearson System-based ICA is a BSS method that minimizes mutual information using a Pearson System-based parametric model. This model covers a wide class of source distributions those include skewed distributions with different values of kurtosis, e.g. Rayleigh distribution and log-normal distribution that are common in telecommunication applications [19].

For the ICA model in Eq. (3.4), the goal is to find a matrix **W** only using observations such that $\mathbf{y} = \mathbf{W}\mathbf{x}$ is a possibly scaled and permuted estimate of the source vector **s**. As it was mentioned before, a contrast function is maximized or minimized in order to separate sources. The derivative of the contrast function is called the estimating function [19].

KL divergence defined earlier is a measure employed in mutual information and likelihood contrasts. A ML type contrast function can be defined as

$\Phi_{ML}(\mathbf{y}) = K(\mathbf{y} // \mathbf{s})$ where $K(\mathbf{y} // \mathbf{s})$ is the KL divergence between the densities of $\mathbf{y}$ and $\mathbf{s}$. The mutual information contrast can be defined as $\Phi_{MI}(\mathbf{y}) = K(\mathbf{y} // \tilde{\mathbf{y}})$ where $\tilde{\mathbf{y}}$ denotes the vector with independent entries with each entry distributed as the corresponding marginal of $\mathbf{y}$. Likelihood is the sum of the mutual information and the term that gives the marginal mismatch between the output and the hypothesized sources leading to

$$K(\mathbf{y} // \mathbf{s}) = K(\mathbf{y} // \tilde{\mathbf{y}}) + K(\tilde{\mathbf{y}} // \mathbf{s}). \tag{3.25}$$

A well known result is that optimal choices for estimating functions in ML approach are the score functions of source distributions but this implies only when the source densities are positive everywhere [20].

3.6.1 Minimization of Mutual Information by Score Function

Relative gradient for minimization of mutual information leads to adaptive estimation of score functions of mixtures. Mutual information can be defined as the KL divergence between the joint density function and the product of marginal density functions found as [20]

$$\Phi_{MI}(\mathbf{w}) = K\left(f_{\mathbf{y}}(\mathbf{y}), \prod_{i=1}^{p} f_{y_i}(y_i) \right) \tag{3.26}$$

$$= \int f_{\mathbf{y}}(\mathbf{y}) log \frac{f_{\mathbf{y}}(\mathbf{y})}{\prod_{i=1}^{p} f_{y_i}(y_i)} d\mathbf{y}$$

$$= -\int f_{\mathbf{y}}(\mathbf{y}) log \frac{\prod_{i=1}^{p} f_{y_i}(y_i)}{f_{\mathbf{y}}(\mathbf{y})} d\mathbf{y}.$$

The relative gradient of the mutual information $\Phi_{MI}(\mathbf{w})$ is [21]

$$\Phi'_{MI}(\mathbf{w}) = \mathbf{w}^T \frac{\partial \Phi_{MI}(\mathbf{w})}{\partial \mathbf{w}} = \int \varphi_{\mathbf{y}}(\mathbf{y})\mathbf{y}^T f_{\mathbf{x}}(\mathbf{x})d\mathbf{x} - \mathbf{I} \tag{3.27}$$

where $\varphi_{\mathbf{y}}(\mathbf{y}) = -\frac{d\, log\, f_{\mathbf{y}}(\mathbf{y})}{d\mathbf{y}}$ is the score function of **y**. Eq. (3.27) can be rewritten by using the relation **y** =**Wx** as

$$\Phi'_{MI}(\mathbf{w}) = |\, det(\mathbf{W})\,| \int \varphi_{\mathbf{y}}(\mathbf{y})\mathbf{y}^T f_{\mathbf{y}}(\mathbf{y})d\mathbf{y} - \mathbf{I}\,. \tag{3.28}$$

If $y(t)$ is an ergodic random process, where individual samples are distributed according to $f_{\mathbf{y}}(\mathbf{y})$ and where |det(**W**)|=1 in the orthogonal case, the relative gradient can be estimated as

$$\hat{\Phi}'_{MI}(\mathbf{w}) = \frac{1}{T}\sum_{t=1}^{T} \hat{\varphi}_{\mathbf{y}}(\mathbf{y}(t))\mathbf{y}(t)^T \tag{3.29}$$

where $\hat{\varphi}_{\mathbf{y}}$ is an estimate of the score function of **y**. Since the output **y** changes at every iteration of the optimization algorithm, the optimal estimating functions also change at each iteration and in mutual information approach, the optimal estimating function is the score function of **y** [19].

3.6.2 The Pearson System

The Pearson System is a parametric family of distributions that may be used to model a wide class of source distributions [19]. Many widely used distributions, including normal, Student's t, gamma and beta distributions belong to the Pearson family. It is defined by the differential equation:

$$f'(x) = \frac{(x-a)f(x)}{b_o - b_1 x + b_2 x^2} \tag{3.30}$$

where a, b_0, b_1 and b_2 are the parameters of the distribution. The score function which is used as a contrast function is easily solved from Eq. (3.30) as

$$\varphi(x) = \frac{-f'(x)}{f(x)} = \frac{(x-a)}{b_o - b_1 x + b_2 x^2}, \tag{3.31}$$

and the derivative of the score function is

$$\varphi'(x) = -\frac{b_0 + ab_1 + 2axb_2 - x^2 b_2}{(b_0 + b_1 x + b_2 x^2)^2}. \tag{3.32}$$

The parameters of the Pearson System can be estimated by the method of moments when the mean is zero and the variance is one as

$$b_1 = a = -\frac{\mu_3(\mu_4 + 3\mu_2^2)}{C}, \tag{3.33}$$

$$b_0 = -\frac{\mu_2(4\mu_2\mu_4 - 3\mu_3^2)}{C}, \tag{3.34}$$

$$b_2 = -\frac{(2\mu_2\mu_4 - 3\mu_3^2 - 6\mu_2^3)}{C}, \tag{3.35}$$

where μ_2, μ_3, μ_4 are the second, the third and the fourth central moments, respectively and $C = 10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3$. For estimating theoretical moments to use in parameter estimation, sample moments are used as

$$\hat{\mu}_1 = \bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} \quad , \quad \hat{\mu}_2 = \hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}, \quad \hat{\mu}_3 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^3}{n\hat{\sigma}^3} \quad , \quad \hat{\mu}_4 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^4}{n\hat{\sigma}^4} \quad .$$

where $\hat{\sigma}$ is the estimated standard deviation of distribution. The source distributions are estimated through the marginal distributions by fitting them to the Pearson family. This is done iteratively until the optimization algorithm converges. Any suitable ICA algorithm can be used to optimize the derived criterion such as the natural gradient algorithm, the relative gradient algorithm and fixed-point algorithm [19]. In the relative gradient algorithm, the iteration procedure is done by using:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \eta(I - \varphi(\bar{y})\bar{y}^T)\mathbf{W}_k \tag{3.36}$$

where η is the learning rate. In fixed-point algorithm, iteration procedure is

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mathbf{D}\left(E\left\{\varphi(\bar{\mathbf{y}})\bar{\mathbf{y}}^T\right\} - \mathrm{diag}\left(E\left\{\varphi(y_i)y_i\right\}\right)\right)\mathbf{W}_k \tag{3.37}$$

where $\mathbf{D} = \mathrm{diag}\left(1/\left(E\{\varphi(y_i)y_i\} - E\{\varphi'(y_i)\}\right)\right)$.

When the kurtosis of the source distribution differs from the kurtosis of a normal distribution, in modeling distributions that are far from normal distribution there is no advantage of the Pearson System. For these situations fixed non-linearities can be used as a contrast function, so one speeds up the computation and avoids the estimation problems. When the source distribution is sub-Gaussian and super-Gaussian, the cubic contrast and the hyperbolic tangent contrast are used, respectively. When the distribution is close to Gaussian or has the same kurtosis as that of the normal distribution, Pearson System can be used for estimation. How to choose a function as a contrast can be found out from the third and the fourth moments according to Figure 3.3.

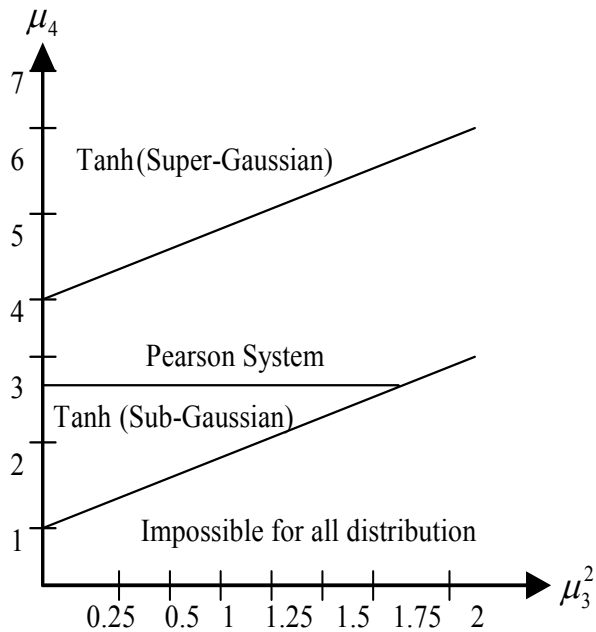

Figure 3.3 The choice of contrast function

The procedure to apply the Pearson System-based ICA may be given as:

1. Calculate the third and fourth order sample moments for current data $\mathbf{y}_k=\mathbf{W}_k\mathbf{x}$ and select the Pearson System or fixed contrasts according to Figure 3.3.
2. If the Pearson System is chosen, estimate parameters of the distributions by methods of moments.
3. Calculate scores for the Pearson System or for the fixed-point contrast.
4. Calculate demixing matrix $\mathbf{W}_{k+1}$ using Eqs (3.36) and (3.37).

3.6.3 A Simulation Example

In order to demonstrate the source separation performance of Pearson-System-based ICA, simulations are performed where four sources are mixed linearly. The sources are independent and they have Rayleigh distribution that is used to model the envelope of the received signals passed through a fading channel generally. Pearson System-based ICA (PS-ICA) is applied to noise-free observations and it can separate sources successfully [22]. Simulation results are given in the Figures 3.4-3.6. Figure 3.4 shows the sources, Figure 3.5 shows observations and Figure 3.6 shows the separated signals and source signals.

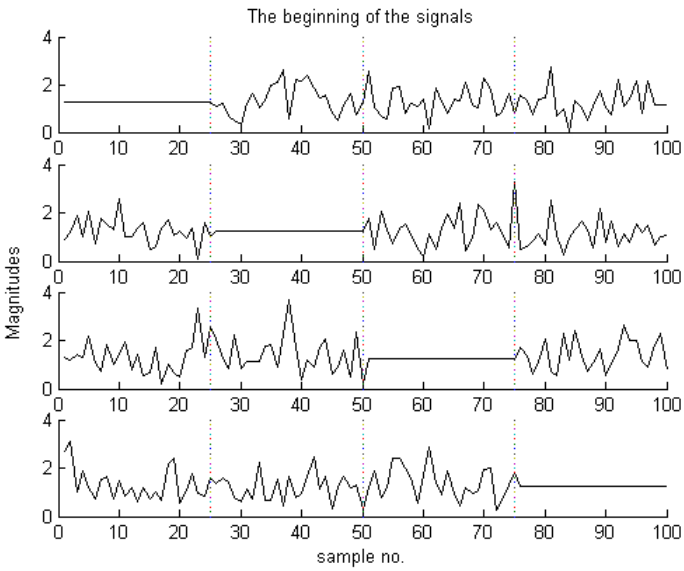

Figure 3.4 Source Signals

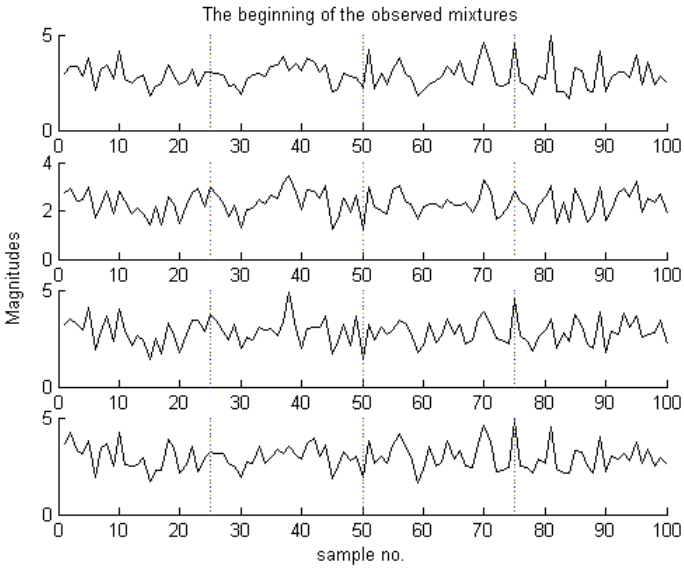

Figure 3.5 Observed Signals

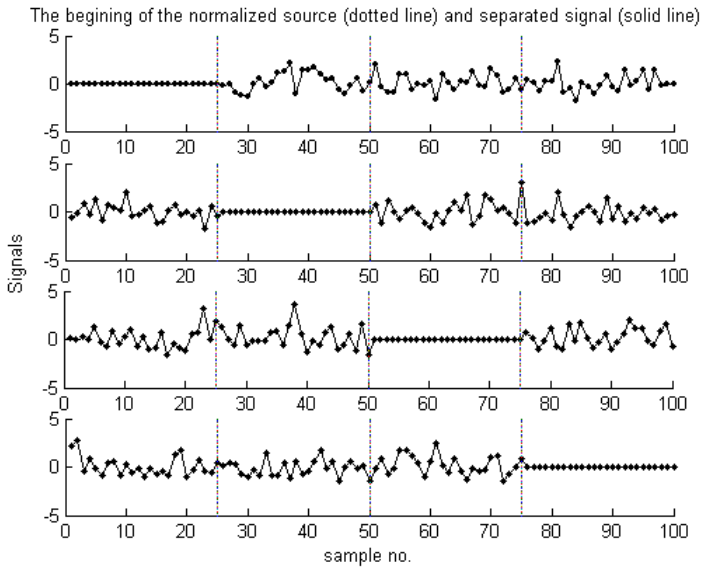

Figure 3.6 Separated Signals and Source Signals (points: original source signals, continuous lines: separated signals)

3.7 Applications of ICA

BSS is the classical application of ICA. By using observed signals, source signals can be estimated by ICA in BSS. Using EEG signals to separate brain signals, using ECG signals to separate heart signals, communication problems including

multipath propagation in mobile systems and separation of speech signals are among the popular applications of BSS/ICA. Another application of ICA is feature extraction. In such an application, columns of **A** in Eq. (3.4) represent features and s_i is the coefficient of the *i*'th feature of an observed data vector **x** and by using ICA, features can be extracted. In addition to these applications, in data analysis of economics, in psychology and in density estimation, ICA can be used [16].

Chapter 4

INDEPENDENT FACTOR ANALYSIS

Independent factor analysis (IFA) is a statistical method like ICA which is also used for recovering independent hidden sources from their observed mixtures [4]. The BSS problem considered in IFA allows also contamination of the mixture signal by additive noise. The task is to obtain those source signals without any information about the mixing process and the noise. The model is given by a probabilistic linear model [4]:

$$y_i = \sum_{j=1}^{L} H_{ij} x_j + u_i \quad i = 1.....L' \tag{4.1}$$

where y_i depends on linear combinations of the x_j's with constant coefficients H_{ij} and u_i are additive noise signals, L is the number of sources and L' is the number of sensors. In this model, the task is to estimate H_{ij} and x_j.

In the mid'80s, most of the work in the field of BSS aimed at a highly idealized version of the problem where mixing matrix is square, invertible and there is no noise. This model is called ICA which was the subject of the previous chapter. However in realistic situations, models have noise and number of sources and observations (sensors) can be different. As the noise level increases, separation quality of ICA decreases. Determining the source density model is another important problem in ICA since learning the densities of sources from the observed data is crucial.

IFA is a new algorithm for blind separation of noisy mixtures with a possibly non-square mixing matrix. It is performed in two steps:

1. Learning the IF model, mixing matrix, noise covariance, source density parameters from the data where each source density is modeled by a mixture of one-dimensional Gaussian density.
2. Recovering the sources from the sensor signals using the posterior density of the sources given the data.

4.1 The Independent Factor (IF) Generative Model

In the first step of IFA, a generative model named as the independent factor (IF) model is generated. The model in Eq. (4.1) can be rewritten in matrix form as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{u} \tag{4.2}$$

where $\mathbf{y}$ is $L' \times 1$ observed sensor signals vector, $\mathbf{x}$ is the unobserved source signals vector, $\mathbf{H}$ is the mixing matrix and $\mathbf{u}$ is the noise vector.

Before generating a model for probability density of the sensor signals, $p(\mathbf{y})$, density of the sources and noise must be specified. The sources x_i are modeled as L independent random variables with arbitrary distributions $p(x_i|\theta_i)$, where the individual i'th source density is parameterized by the parameter set θ_i.

The noise is assumed to be Gaussian with zero mean and covariance matrix $\mathbf{\Lambda}$. Even though sensor noise signals are independent, correlations may arise because of source noise or propagation noise. The density function of the noise is:

$$p(\mathbf{u}) = \zeta(\mathbf{u}, \mathbf{\Lambda}) \tag{4.3}$$

The generative IF model can be defined by Eqs. (4.2) and (4.3). It is parameterized by the source parameters $\boldsymbol{\theta}$, mixing matrix $\mathbf{H}$ and noise covariance matrix $\mathbf{\Lambda}$:

$$\mathbf{W} = (\mathbf{H}, \mathbf{\Lambda}, \boldsymbol{\theta}). \tag{4.4}$$

The resulting model sensor density is

$$\begin{aligned} p(\mathbf{y} \mid \mathbf{W}) &= \int p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})\, d\mathbf{x} \\ &= \int G(\mathbf{y} - \mathbf{H}\mathbf{x}, \mathbf{\Lambda}) \prod_{i=1}^{L} p(x_i|\theta_i)\, d\mathbf{x} \end{aligned} \tag{4.5}$$

where $d\mathbf{x} = \prod_i dx_i$. The IF parameters $\mathbf{W}$ should be adopted to minimize an error

function which measures the distance between the model and observed sensor densities.

4.1.1 Source Model: Factorial Mixture of Gaussians

To perform the integration in Eq. (4.5) analytically, it is important to choose a parametric form $p(x_i)$ which is sufficiently general to model arbitrary source densities. Mixture of Gaussians (MOG) model can satisfy these requirements. The density of source i is a mixture of n_i Gaussians given by [4]:

$$p(x_i|\theta_i) = \sum_{q_i=1}^{n_i} w_{i,q_i}\,\zeta\left(x_i - \mu_{i,q_i}, v_{i,q_i}\right) \quad \theta_i = \left\{w_{i,q_i}, \mu_{i,q_i}, v_{i,q_i}\right\} \tag{4.6}$$

with means μ_{i,q_i}, variances v_{i,q_i} and mixing proportions w_{i,q_i} where $q_i=1...n_i$.

For a normalized mixture, the mixing proportions for each source should sum up to unity $\sum_{q_i} w_{i,q_i} = 1$. In the generative description of the sources different Gaussians play the role of hidden states. To generate source signals x_i, first a state q_i is picked with probability $p(q_i) = w_{i,q_i}$, and then a number is found from the corresponding Gaussian density $p(x_i|q_i) = \zeta\left(x_i - \mu_{i,q_i}, v_{i,q_i}\right)$.

The joint source density $p(\mathbf{x})$ is formed by the product of the one-dimensional MOGs in Eq. (4.6) as

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{L} p(x_i|\theta_i) = \sum_{q} w_q\,\zeta\left(x - \mu_q, v_q\right) \tag{4.7}$$

where collective hidden states are

$$\mathbf{q} = \left(\mathbf{q}_1, ..., \mathbf{q}_L\right) \tag{4.8}$$

with mixing proportions $\mathbf{w}_q$, mean $\boldsymbol{\mu}_q$ and diagonal covariance matrix $\boldsymbol{\nu}_q$ which are defined as:

$$\mathbf{w_q} = \prod_{i=1}^{L} w_{i,q_i} = w_{1,q_1} \times \cdots \times w_{L,q_L}, \quad \boldsymbol{\mu}_\mathbf{q} = \left[\mu_{1,q_1} \cdots \mu_{L,q_L}\right]^T, \quad \mathbf{v_q} = \mathrm{diag}(\nu_{1,q_1} \cdots \nu_{L,q_L}) \quad (4.9)$$

In Eq. (4.7), Gaussians factorize as $\zeta(\mathbf{x} - \boldsymbol{\mu}_\mathbf{q}, \mathbf{v_q}) = \prod_i \zeta(x_i - \mu_{i,q_i}, v_{i,q_i})$ and the sum over collective states $\mathbf{q}$ represents summing over all the individual source states $\sum_\mathbf{q} = \sum_{q_1} \sum_{q_2} \cdots \sum_{q_L}$. For situations with many sources, this summation over all the individual source states causes problems and some other solutions must be considered. The joint source density $p(\mathbf{x})$ is itself a MOG as marginal source density in Eq. (4.6). Contrary to ordinary MOGs, the Gaussians in Eq. (4.7) is not free to adapt independently [4]. Changing the mean and variance of a single source state q_i leads to shifting the whole column of collective states $\mathbf{q}$. Then the source model in Eq. (4.7) is a mixture of "co-adaptive Gaussians" and called "Factorial MOG" [4].

4.1.2 Sensor Model

Combining Eq. (4.2), the noise model in Eq. (4.3) and source model in Eq. (4.7) give a two-step generative model of the observed sensor signals as shown in Figure 4.1.

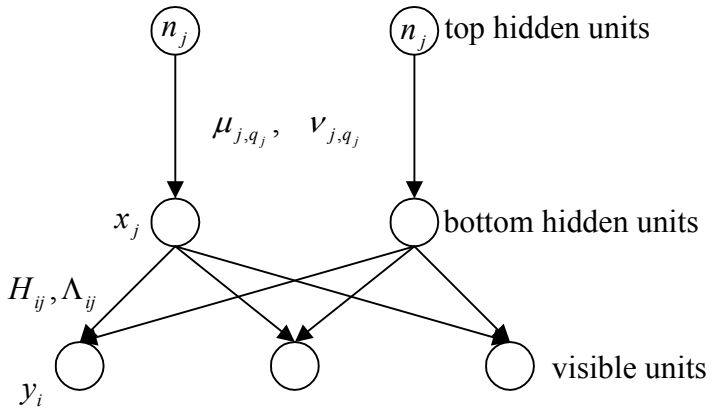

Figure 4.1 Network representation of the IF generative model [4].

The steps to generate a sensor signal $\mathbf{y}$ is given as follows [4]:

1. Pick a unit q_i for each source i with probability

$$p(\mathbf{q}) = \mathrm{w}_{\mathbf{q}} \tag{4.10}$$

from the top hidden layer of source states. This unit has a top-dawn generative connection with weight w_{j,q_j} to each of the units j in the bottom hidden layer. It causes the unit j to produce a sample from a Gaussian density centered at μ_{j,q_j} with a variance of v_{j,q_j} . The probability of generating a particular source vector **x** in the bottom hidden layer is:

$$p(\mathbf{x} \mid \mathbf{q}) = \zeta(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{q}}, \mathbf{v}_{\mathbf{q}}). \tag{4.11}$$

2. Each unit j in the bottom hidden layer has a connection with weight H_{ij} to each unit i in the visible layer. After the generation of **x**, unit i produces a sample y_i from a Gaussian density centered at $\sum_j H_{ij} x_j$. Although sensor noise is independent, generally the noise is correlated across sensors and probability for generating a particular sensor vector **y** in the visible layer is

$$p(\mathbf{y} \mid \mathbf{x}) = \zeta(\mathbf{y} - \mathbf{Hx}, \boldsymbol{\Lambda}). \tag{4.12}$$

This model is described by the joint density of the visible layer and the two hidden layers:

$$p(\mathbf{q}, \mathbf{x}, \mathbf{y} \mid \mathbf{W}) = p(\mathbf{q}) p(\mathbf{x} \mid \mathbf{q}) p(\mathbf{y} \mid \mathbf{x}). \tag{4.13}$$

Note that sensor signals depend on the sources but not on the source states so

$$p(\mathbf{y} \mid \mathbf{x}, \mathbf{q}) = p(\mathbf{y} \mid \mathbf{x}), \tag{4.14}$$

and IF network layers in Figure 4.1 form a top-down first order Markov chain. By using Eq. (4.13) and Eq. (4.14), $p(\mathbf{y})$ can be expressed as:

$$
\begin{aligned}
p(\mathbf{y} \mid \mathbf{W}) &= \sum_{\mathbf{q}} \int p(\mathbf{q}) p(\mathbf{x} \mid \mathbf{q}) p(\mathbf{y} \mid \mathbf{x})\, d\mathbf{x} \\
&= \sum_{\mathbf{q}} p(\mathbf{q}) p(\mathbf{y} \mid \mathbf{q})
\end{aligned}
\tag{4.15}
$$

where because of Gaussian forms in Eqs. (4.11) and (4.12)

$$
p(\mathbf{y} \mid \mathbf{q}) = \zeta\left(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}_{\mathbf{q}}, \mathbf{H}\boldsymbol{\nu}_{\mathbf{q}}\mathbf{H}^{\mathbf{T}} + \boldsymbol{\Lambda}\right). \tag{4.16}
$$

4.2 Learning the IF Model

4.2.1 Error Function and Maximum Likelihood

The IF model parameters can be found iteratively by minimizing error between model sensor density p(**y**|**W**) in Eq. (4.15) and observed one $p^o(\mathbf{y})$. KL distance function can be chosen as the error function [4]:

$$
\begin{aligned}
\xi(\mathbf{W}) &= \int p^o(\mathbf{y}) log \frac{p^o(\mathbf{y})}{p(\mathbf{y} \mid \mathbf{W})} d\mathbf{y} \\
&= -E[log\, p(\mathbf{y} \mid \mathbf{W})] - H_{p^o}
\end{aligned}
\tag{4.17}
$$

When $p^o(\mathbf{y})=p(\mathbf{y})$, KL distance ξ vanishes. Error function in Eq. (4.17) consists of two terms. First one is the negative log-likelihood of the observed sensor signals given the model parameters **W**. The second term, sensor entropy, is independent of **W** and will be dropped. So, minimizing ξ is equivalent to maximizing the likelihood of the data with respect to the model.

To minimize the error in Eq. (4.17), classical gradient-descent method can be used but it results in slow learning. Instead of gradient-descent method, an Expectation-Maximization (EM) algorithm can be developed for learning the IF model.

4.2.2 The EM Algorithm

The EM algorithm is an iterative method to maximize the log-likelihood of the observed data with respect to the parameters of the generative model describing that

data [4]. Instead of the likelihood $E[\log p(\mathbf{y} \mid \mathbf{W})]$ of the observed sensor data, one may consider the likelihood $E[\log p(\mathbf{y}, \mathbf{x}, \mathbf{q} \mid \mathbf{W})]$ of the complete data, i.e., the unobserved source signals and states. In EM algorithm, each iteration consists of two steps [4]:

1. Calculating the expected value of the complete-data likelihood, given the observed data and the current model:

$$F(\mathbf{W}', \mathbf{W}) = -E[log\, p(\mathbf{q}, \mathbf{x}, \mathbf{y} \mid \mathbf{W})] + F_H(\mathbf{W}') \tag{4.18}$$

   For each observed **y**, the average in the first term on the r.h.s is taken over the unobserved **x** and **q** using the source posterior $p(\mathbf{x}, \mathbf{q} \mid \mathbf{y}, \mathbf{W}')$. $\mathbf{W'}$ are the model parameters obtained in the previous iteration. $F_H(\mathbf{W'})$ is the entropy of the posterior and since it is independent of **W**, it has no effect.

2. Minimizing $F(\mathbf{W'}, \mathbf{W})$ (maximizing the averaged likelihood) with respect to **W** to obtain new model parameters:

$$\mathbf{W} = \arg\min_{\mathbf{W}''} F(\mathbf{W}', \mathbf{W}'') . \tag{4.19}$$

In the following, the EM algorithm is developed for the IF model. Firstly it must be shown that F in Eq. (4.18) is bounded from below by the error in Eq. (4.17). By dropping the average over the observed **y** in the error function, it is obtained that:

$$\begin{aligned} \xi(\mathbf{W}) &= -\log p(\mathbf{y} \mid \mathbf{W}) = -\log \sum_{\mathbf{q}} \int p(\mathbf{q}, \mathbf{x}, \mathbf{y} \mid \mathbf{W})\, d\mathbf{x} \\ &\le -\sum_{\mathbf{q}} \int p'(\mathbf{q}, \mathbf{x} \mid \mathbf{y}) \log \frac{p(\mathbf{q}, \mathbf{x}, \mathbf{y} \mid \mathbf{W})}{p'(\mathbf{q}, \mathbf{x} \mid \mathbf{y})} d\mathbf{x} \equiv F \end{aligned} \tag{4.20}$$

where the second line follows from Jensen's inequality [4] and holds for any conditional density p'. In the derivation of the EM algorithm for IF model, source posterior computed using the parameters from the previous iteration is chosen as p' :

$$p'(\mathbf{q}, \mathbf{x} \mid \mathbf{y}) = p(\mathbf{q}, \mathbf{x} \mid \mathbf{y}, \mathbf{W}') \tag{4.21}$$

which is obtained from Eq. (4.13) with $\mathbf{W} = \mathbf{W}'$. After the previous iteration, an approximate error function $F(\mathbf{W}', \mathbf{W})$ is obtained due to the Markov Property of IF model in Eq. (4.13):

$$\xi(\mathbf{W}) \le F(\mathbf{W}', \mathbf{W}) = F_v + F_B + F_T + F_H \tag{4.22}$$

where

$$\begin{aligned} F_V(\mathbf{W}', \mathbf{H}, \mathbf{\Lambda}) &= -\int p(\mathbf{x} \mid \mathbf{y}, \mathbf{W}') \log p(\mathbf{y} \mid \mathbf{x})\, d\mathbf{x}\,, \\ F_B(\mathbf{W}', \{\mu_{i,qi}, \nu_{i,qi}\}) &= -\sum_{i=1}^{L} \sum_{qi=1}^{n_i} p(q_i \mid \mathbf{y}, \mathbf{W}') \int p(x_i \mid q_i, \mathbf{y}, \mathbf{W}') \log p(x_i \mid q_i)\, dx_i\,, \\ F_T(\mathbf{W}', \{w_{i,qi}\}) &= -\sum_{i=1}^{L} \sum_{qi=1}^{n_i} p(q_i \mid \mathbf{y}, \mathbf{W}') \log p(q_i), \end{aligned} \tag{4.23}$$

and the last one is the negative entropy of the source posterior:

$$F_H(\mathbf{W}') = \sum_{\mathbf{q}} \int p(\mathbf{q}, \mathbf{x} \mid \mathbf{y}, \mathbf{W}') \log\, p(\mathbf{q}, \mathbf{x} \mid \mathbf{y}, \mathbf{W}')\, d\mathbf{x}. \tag{4.24}$$

Note that to get F_B in Eq. (4.23), $p(\mathbf{q} \mid \mathbf{x}) p(\mathbf{x} \mid \mathbf{y}) = p(\mathbf{q} \mid \mathbf{y}) p(\mathbf{x} \mid \mathbf{q}, \mathbf{y})$ is used. It can be obtained from Eq. (4.13).

When $\mathbf{W} = \mathbf{W}'$, because of the choice in Eq. (4.21), Eq. (4.22) becomes an equality:

$$F(\mathbf{W}', \mathbf{W}') = \xi(\mathbf{W}')\,. \tag{4.25}$$

Next $F(\mathbf{W}', \mathbf{W})$ is considered and it is minimized with respect to $\mathbf{W}$ and then new parameters are obtained from Eq. (4.19) to satisfy:

$$\xi(\mathbf{W}) \le F(\mathbf{W}', \mathbf{W}) \le F(\mathbf{W}', \mathbf{W}') = \xi(\mathbf{W}') \tag{4.26}$$

which proves that the current EM step does not increase the error [4]. After the derivation of EM algorithm for IF model, learning rules for the mixing matrix and noise covariance are found as (see Appendix C):

$$\begin{aligned}\mathbf{H} &= E\left\{\mathbf{y}\left\langle \mathbf{x}^T \mid \mathbf{y}\right\rangle\right\}\left(E\left\{\left\langle \mathbf{x}\mathbf{x}^T \mid \mathbf{y}\right\rangle\right\}\right)^{-1} \\ \mathbf{\Lambda} &= E\left\{\mathbf{y}\mathbf{y}^T\right\} - E\left\{\mathbf{y}\left\langle \mathbf{x}^T \mid \mathbf{y}\right\rangle\right\}\mathbf{H}^T\end{aligned} \tag{4.27}$$

and learning rules for source MOG parameters are:

$$\begin{aligned}\mu_{i,qi} &= \frac{E\left\{p(q_i \mid \mathbf{y})\left\langle x_i \mid q_i, \mathbf{y}\right\rangle\right\}}{E\left\{p(q_i \mid \mathbf{y})\right\}} \\ \nu_{i,qi} &= \frac{E\left\{p(q_i \mid \mathbf{y})\left\langle x_i^2 \mid q_i, \mathbf{y}\right\rangle\right\}}{E\left\{p(q_i \mid \mathbf{y})\right\}} - \mu_{i,qi}^2 \\ w_{i,qi} &= E\left\{p(q_i \mid \mathbf{y})\right\}.\end{aligned} \tag{4.28}$$

where $\left\langle \mathbf{x} \mid \mathbf{y}\right\rangle$ is an $L\times 1$ vector and it denotes the conditional mean of sources given the sensors $\mathbf{y}$. $\left\langle \mathbf{x}\mathbf{x}^T \mid \mathbf{y}\right\rangle$ is the $L\times L$ matrix denoting the source covariance conditioned on the sensors. Lastly $\left\langle x_i \mid q_i, \mathbf{y}\right\rangle$ denotes the mean of sensor i conditioned on both the hidden states q_i of this source and the observed sensors.

In the BSS problem, the sources are recovered only within an order permutation and scaling. The effect of arbitrary permutation of sources can be solved by a corresponding permutation of the columns of **H**, leaving **y** unchanged. Scaling source x_i by a factor σ_j would not affect **y.** If the j'th column of **H** is scaled by $1/\sigma_j$ at the same time, then the following scaling transformation is performed at every iteration [4]:

$$\begin{aligned}&\sigma_j^2 = \sum_{qj=1}^{n_j} w_{j,qj}\left(\nu_{j,qj} + \mu_{j,qj}^2\right) - \sum_{qj=1}^{n_j}\left(w_{j,qj}\mu_{j,qj}\right)^2, \\ &\mu_{j,qj} \to \frac{\mu_{j,qj}}{\sigma_j}, \quad \nu_{j,qj} \to \frac{\nu_{j,qj}}{\sigma_j^2}, \quad H_{ij} \to H_{ij}\sigma_j.\end{aligned} \tag{4.29}$$

It can be shown that scaling transformation does not change the error function [4].

4.3 Recovering the Sources

After IF generative model parameters have been estimated, the sources can be reconstructed from the sensor signals. One method of recovering the sources is the least mean squares (LMS) estimation.

4.3.1 LMS Estimator

The optimal estimate in the least squares sense is the conditional mean of the sources given the observed sensor outputs [4]:

$$\hat{\mathbf{x}}^{LMS}(\mathbf{y}) = \langle \mathbf{x} \mid \mathbf{y} \rangle = \int \mathbf{x}\ p(\mathbf{x} \mid \mathbf{y}, \mathbf{W})\, d\mathbf{x} \tag{4.30}$$

where source posterior $p(\mathbf{x} \mid \mathbf{y}, \mathbf{W}) = \sum_{q} p(\mathbf{q} \mid \mathbf{y}) p[\mathbf{x} \mid \mathbf{q}, \mathbf{y}]$ depends on the generative parameters. The calculation of Eq. (4.30) is given in Appendix D. It is given by a weighted sum of terms that are linear in the data **y**:

$$\hat{\mathbf{x}}^{LMS}(\mathbf{y}) = \sum_{\mathbf{q}} p(\mathbf{q} \mid \mathbf{y})(A_{\mathbf{q}}\mathbf{y} + b_{\mathbf{q}}) \tag{4.31}$$

where $A_{\mathbf{q}} = \sum_{\mathbf{q}} \mathbf{H}^{\mathrm{T}} \mathbf{\Lambda}^{-1}$, $b_{\mathbf{q}} = \sum_{\mathbf{q}} \mathbf{V}_{\mathbf{q}}^{-1} \boldsymbol{\mu}_{\mathbf{q}}$, $\sum_{\mathbf{q}} = \left(\mathbf{H}^{\mathrm{T}} \mathbf{\Lambda}^{-1} \mathbf{H} + \mathbf{V}_{\mathbf{q}}^{-1}\right)^{-1}$.

4.3.2 Simulation Results

The performance of IFA is demonstrated on mixtures of sources corrupted by Gaussian noise (variance of noise=0.02). The source signals have α-stable distribution where the shape parameter of the distribution α, has the value 1.8. Detailed

information on α-stable distributions can be found in Appendix E. There are two source signals and the mixture is formed multiplying the 2×1 source vector by a random 2×2 mixing matrix. In the simulation, each source density was modeled by n_i=3 state MOG which provided a sufficiently accurate description of the signals.

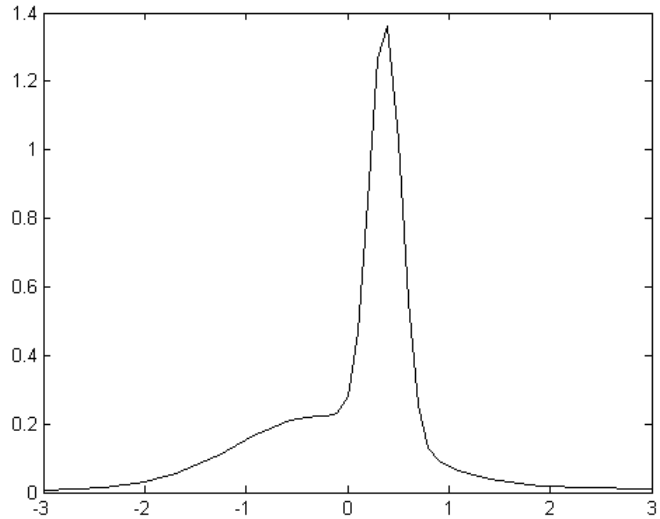

Figure 4.2 The estimated pdf of first source

The length of the observation is 1000 samples for each sensor. In Figure 4.2, the estimated pdf of the first source is given. The 3 components of the pdf of the first source can be seen in Figure 4.3.

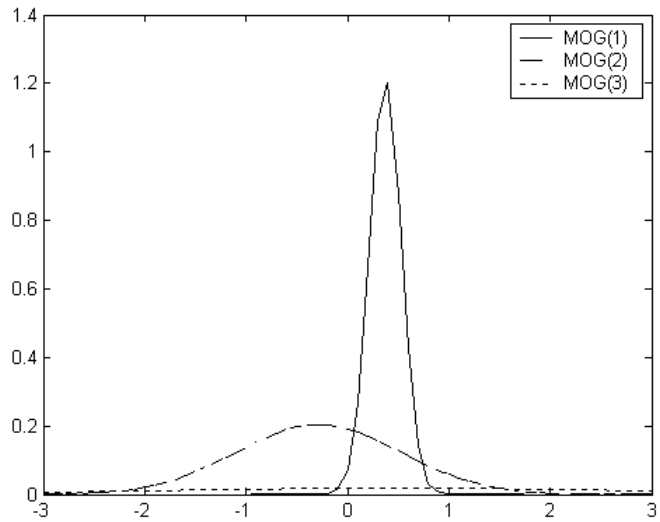

Figure 4.3 Estimated components of the pdf of the first source

In Figure 4.4, the signal estimated by IFA and the source signal is given. As it is seen IFA can estimate the source signal successfully.

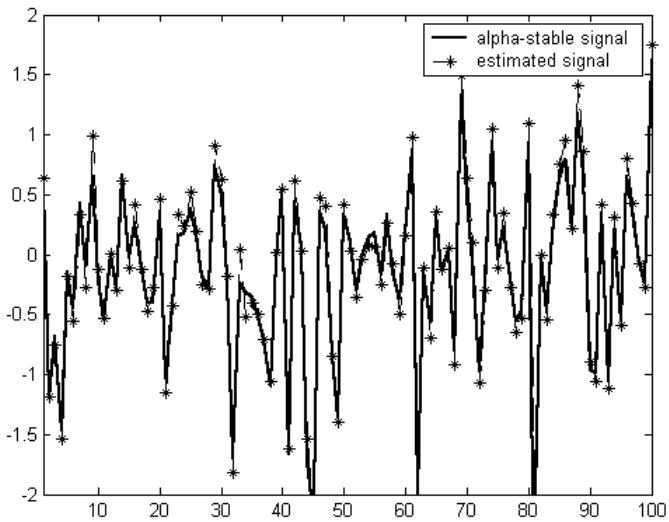

Figure 4.4 Estimation of the signal

In the second experiment, the source signals are chosen as binary signals and they were mixed again by a 2×2 mixing matrix. The mixtures of sources were corrupted by a Gaussian noise that has 0.2 variance. IFA were used to separate sources from noisy mixtures. In Figure 4.5 the estimated and the original binary signals are seen. As the figure depicts IFA can separate binary numbers successfully.

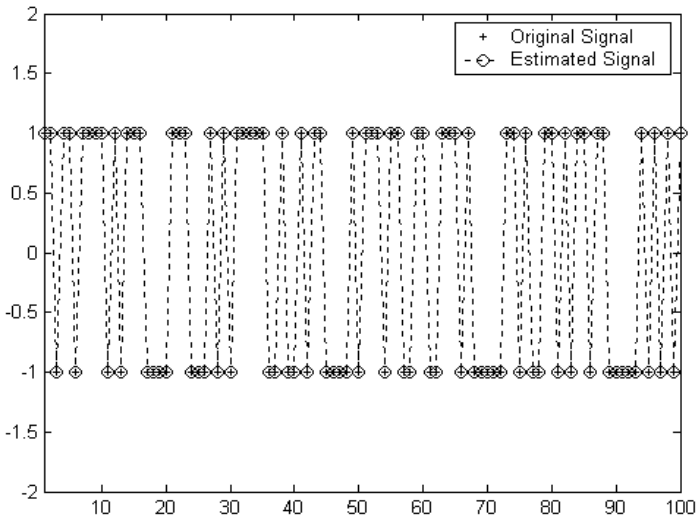

Figure 4.5 Estimated and original binary signals

Also Figure 4.6 and 4.7 show the estimated pdf and its components of binary data which were estimated by IFA.

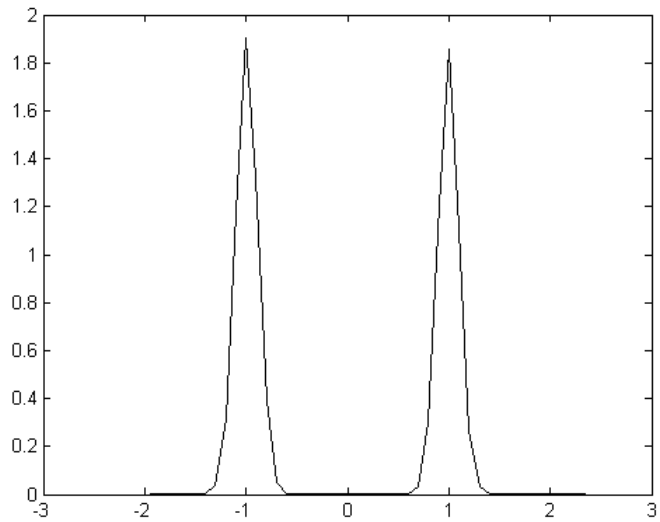

Figure 4.6 Estimated pdf of the second source

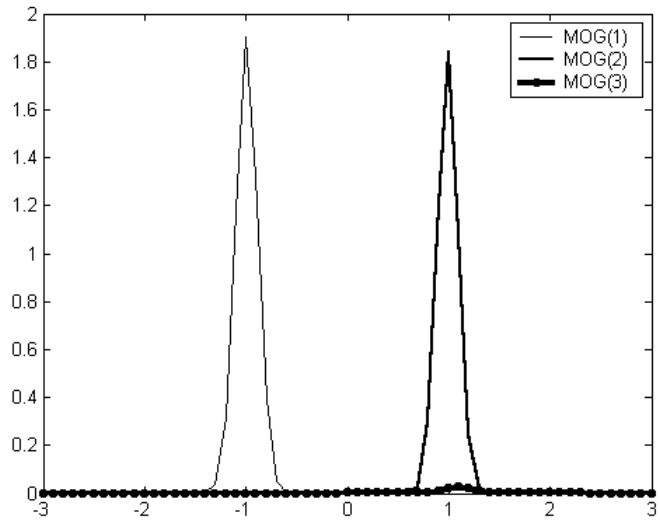

Figure 4.7 Components of the pdf of the second source

Chapter 5

ICA AND IFA APPLICATIONS IN CDMA

In CDMA communication, especially in CDMA downlink problem, e.g. from base station to mobile receiver, each user must separate the received signal and detect its own symbols since the mobile receiver does not know the other users' code sequences [23]. This separation is a BSS problem. That's why, there are many works on ICA applications in CDMA systems. In this chapter, application of ICA to CDMA downlink problem is considered and performance of systems is analyzed with simulation results. In addition to this, a new technique of BSS methods called IFA presented in the previous chapter is applied to a basic CDMA downlink model.

First, CDMA models for different channel scenarios are given in section 5.1.

5.1 CDMA Downlink Models used in ICA Applications

5.1.1 Basic K-user DS-CDMA Downlink Model

In this model, channel is assumed to be ideal and there is no multipath fading. The antipodal K-user baseband DS-CDMA downlink system is modeled as [24, 25]:

$$r(t) = \sum_{k=1}^{K} \sum_{i=-M}^{M} A_k b_k[i] s_k(t - iT - \tau_k) + \sigma n(t) \tag{5.1}$$

where $r(t)$ is the received signal, A_k is the k'th user's received signal amplitude, $b_k[i]$ is the i'th antipodal data bit of the k'th user, $s_k(t)$ is the k'th user's signature waveform with normalized energy ($\|s_k(\cdot)\|=1$), T is the symbol period, τ_k is the relative offset of the k'th user, σ^2 is the noise power, $n(t)$ is the AWGN with unit power spectral density. In this model $2M+1$ symbols are observed. The simplified synchronous one-shot CDMA model is:

$$r(t) = \sum_{k=1}^{K} A_k b_k s_k(t) + \sigma n(t) \tag{5.2}$$

After chip matched filtering and sampling, the received signal **r** can be expressed in $C\times1$ vector form:

$$\mathbf{r} = \sum_{k=1}^{K} A_k b_k \mathbf{s}_k + \sigma \mathbf{n} \tag{5.3}$$

where C is the processing gain or spreading gain, $\mathbf{s_k}$ is the $C\times1$ vector of the k'th user signature sequence and **n** is the $C\times1$ AWGN vector. Eq. (5.3) can be rewritten in matrix form as:

$$\mathbf{r} = \mathbf{Gb} + \mathbf{n} \tag{5.4}$$

where **r** is the $C\times1$ observed data vector, **G** is the $C\times K$ unknown full rank mixing matrix, **b** is the $K\times1$ source data vector and **n** is $C\times1$ noise vector as mentioned before. Here **G=SA** where $\mathbf{S}=[\mathbf{s_1}\ \mathbf{s_2}.....\mathbf{s_K}]$ and $\mathbf{A}=\mathrm{diag}(A_1....A_K)$ [24]. As it was given before in Chapter 3, model in Eq. (5.4) is a classical noisy ICA model. This K-user CDMA model is a simple model under the assumption of ideal channel models with no multipath fading and delays.

5.1.2 DS-CDMA Downlink Models in Multipath Fading Channels

5.1.2.1 Model in Slow Fading Case

As the CDMA downlink model in slow fading case, the model in Eq. (2.11) is adopted. In this model, the length of the chip sequence is C and then C-vectors $\mathbf{r}_m$ are collected from subsequent discretized equispaced data samples $r(n)$ as

$$\mathbf{r}_m = \left[r(mC)\ r(mC+1)\ \ldots\ r((m+1)C-1)\right]^T \tag{5.5}$$

and Eq. (5.5) has the form [26]:

$$\mathbf{r}_m = \sum_{k=1}^{K}\left[b_{k,m-1}\sum_{l=1}^{L} a_l \mathbf{g}_{kl}^{E} + b_{k,m}\sum_{l=1}^{L} a_l \mathbf{g}_{kl}^{L} \right] + \mathbf{n}_m \tag{5.6}$$

where $\mathbf{g}_{kl}^{E}$, $\mathbf{g}_{kl}^{L}$ and $\mathbf{n}_m$ denote the "early" and the "late" parts of code vectors and the noise vector, respectively. In Eq. (5.6), $\mathbf{g}_{kl}^{E}$ and $\mathbf{g}_{kl}^{L}$ are:

$$\mathbf{g}_{kl}^{E} = \left[s_k(C-d_l+1) \ldots s_k(C) \;\; 0\cdots 0 \right]^T, \tag{5.7}$$

$$\mathbf{g}_{kl}^{L} = \left[\; 0\cdots 0 \quad s_k(1)\cdots s_k(C-d_l) \right]^T. \tag{5.8}$$

In Eqs. (5.7) and (5.8), d_l is a discrete delay and generally it is assumed that $d_l \in \{0,....,(C-1)/2\}$. In a more compact form, Eq. (5.6) can be represented as:

$$\mathbf{r}_m = \mathbf{G}\mathbf{b}_m + \mathbf{n}_m \tag{5.9}$$

where $\mathbf{G} = \left[\mathbf{g}_1 \cdots \mathbf{g}_{2K} \right]$ is a $C\times 2K$ matrix and contains the basis vectors and fading terms

$$\mathbf{G} = \left[\sum_{l=1}^{L} a_l \mathbf{g}_{1l}^{E} \;\; \sum_{l=1}^{L} a_l \mathbf{g}_{1l}^{L} \;\cdots\cdots\; \sum_{l=1}^{L} a_l \mathbf{g}_{Kl}^{E} \;\; \sum_{l=1}^{L} a_l \mathbf{g}_{Kl}^{L} \right] \tag{5.10}$$

and the $2K$-vector b_m contains the transmitted symbols

$$\mathbf{b}_m = \left[b_{1,m-1} \quad b_{1,m} \quad \cdots \quad b_{K,m-1} \quad b_{K,m} \right]^T. \tag{5.11}$$

For N transmitted symbols Eq. (5.9) can be rewritten in a pure matrix form as

$$\mathbf{X} = \mathbf{G}\mathbf{B} + \mathbf{N} \tag{5.12}$$

where

$$\begin{aligned}\mathbf{X} &= \begin{bmatrix}\mathbf{r_1} & \cdots & \mathbf{r_N}\end{bmatrix},\\ \mathbf{B} &= \begin{bmatrix}\mathbf{b_1} & \cdots & \mathbf{b_N}\end{bmatrix},\\ \mathbf{N} &= \begin{bmatrix}\mathbf{n_1} & \cdots & \mathbf{n_N}\end{bmatrix}.\end{aligned} \tag{5.13}$$

5.1.2.2 Model in Fast Fading Case

In the fast fading case the CDMA downlink model is given in Eq. (2.12). The derivation of the previous model in section 5.1.2.1 can be developed for this model given in Eq. (2.12). After sampling and chip matched filtering, C-vectors $\mathbf{r}_m$ given in Eq. (5.5)

$$\mathbf{r}_m = \left[r(mC)\ r(mC+1)\ \ldots\ r((m+1)C-1)\right]^T \tag{5.14}$$

can be written as [27]:

$$\mathbf{r}_m = \sum_{k=1}^{K}\sum_{l=1}^{L}\left[a_{l,m-1}b_{k,m-1}\mathbf{g}_{kl}^{E} + a_{l,m}b_{k,m}\mathbf{g}_{kl}^{L}\right] + \mathbf{n}_m \tag{5.15}$$

where $\mathbf{g}_{kl}^{E}$ and $\mathbf{g}_{kl}^{L}$ as they are given in Eqs (5.7) and (5.8), respectively.

The matrix $\mathbf{R} = \begin{bmatrix}\mathbf{r}_1 & \cdots & \mathbf{r}_N\end{bmatrix}$ can be represented in compact form with N transmitted symbols as:

$$\mathbf{R} = \mathbf{GF} + \mathbf{N} \tag{5.16}$$

where $C\times 2KL$ matrix $\mathbf{G}$ contains the basis vectors:

$$\mathbf{G} = \begin{bmatrix}\mathbf{g}_{11}^{E} & \mathbf{g}_{11}^{L} & \cdots & \mathbf{g}_{KL}^{E} & \mathbf{g}_{KL}^{L}\end{bmatrix} \tag{5.17}$$

and the $2KL\times N$ matrix $\mathbf{F} = \begin{bmatrix}\mathbf{f}_1 & \cdots & \mathbf{f}_N\end{bmatrix}$ contains the symbol and fading terms:

$$\mathbf{f}_m = \begin{bmatrix}a_{1,m-1}b_{1,m-1} & a_{1,m}b_{1,m} & \cdots & a_{L,m-1}b_{K,m-1} & a_{L,m}b_{K,m}\end{bmatrix}. \tag{5.18}$$

5.2 Multiuser Detection of DS-CDMA Signals Using FastICA

5.2.1 Fast ICA Application to DS-CDMA Signals

CDMA downlink models in Eqs. (5.4), (5.12) and (5.16) are just the linear ICA models. The goal is solving the symbol process directly without knowing the codes or any channel parameters like delays or fading. FastICA algorithm is the simplest and fastest linear ICA algorithm as mentioned in Chapter 3. For FastICA applications in CDMA models, firstly data is sphered or whitened [26, 18, and 28]. Whitening is a common preprocessing task which normalizes the component variances and filters some additive noise simultaneously. It is performed by using SVD or eigenvalue decomposition. In this application, the model in Eq. (5.12) is considered. After the whitening operation the new data has the form:

$$\mathbf{Y} = \mathbf{\Lambda}_S^{-1/2}\mathbf{U}_S^T\mathbf{X} \tag{5.19}$$

where $\mathbf{\Lambda}_S$ and $\mathbf{U}_S$ corresponds to the *2K* eigenvalues and eigenvectors of the correlation matrix estimate $\mathbf{XX}^T/N$.

Then the FastICA algorithm using the fourth order statistic kurtosis has the following form [29]:

1. Initialize $\mathbf{w}(0)$ with a random value of unity norm.
2. $\mathbf{w}(k) = \frac{1}{N}\mathbf{Y}\left[\mathbf{Y}^T\mathbf{w}(k-1)\right]^3 - 3\mathbf{w}(k-1)$
3. $\mathbf{w}(k+1) \leftarrow \mathbf{w}(k) / ||\mathbf{w}(k)||$
4. Repeat steps 2 and 3 until $\left|\mathbf{w}(k)^T\mathbf{w}(k-1)\right|$ is sufficiently close to 1.

In step 2, $(.)^3$ corresponds to an elementwise operation. The final vector $\mathbf{w}(k)$ separates one of the non-Gaussian source signals and an important property of this algorithm is that a very small number of iterations is needed for convergence.

This procedure is the fixed-point rule for estimating one basis vector. In CDMA application for model in Eq. (5.12), there are *2K* independent components, so, the algorithm must be run *2K* times to find *2K* independent components. For this

operation, let **C** be the mixing matrix which is the orthonormal basis of the sphered data [26]. This simply means that in the noiseless case of Eq. (5.12):

$$\mathbf{C} = \mathbf{\Lambda}_S^{-1/2}\mathbf{U}_S^T\mathbf{G}\,. \tag{5.20}$$

Then the estimated vector $\mathbf{w}(k)$ equals one of the columns of **C**. Several independent components can be estimated by using an orthonormal projection in the beginning of step 3:

$$\mathbf{w}(k) = \mathbf{w}(k-1) - \widetilde{\mathbf{C}}\widetilde{\mathbf{C}}^T\mathbf{w}(k) \tag{5.21}$$

where $\widetilde{\mathbf{C}}$ is a matrix whose columns are previously found.

5.2.2 Simulation Results

In simulation experiments, CDMA downlink model for slowly fading channel in subsection 5.1.2.1 is considered. Gold codes are used as a signature waveform and length of the code is C=31. N=500 symbols are observed. The number of paths is L=3. In simulations only the real part of the data is used where the same simulations can be done for imaginary part of the signal, too. In the first experiment, it is assumed that there are no path losses.

For a relatively high SNR of 20 dB FastICA can estimate all of the transmitted symbols successfully when the number of users is five. Result of five users' case is given in Figure 5.1

Figure 5.2 shows 500 estimated symbols as a function of number of users, K, where the SNR is a parameter and takes values of 5, 10 and 20 dB. In this simulation, number of paths, L, is 3 and they have 0dB, -5dB and -5dB path losses. For 12 users, the number of correctly estimated symbols decreases for three SNR values.

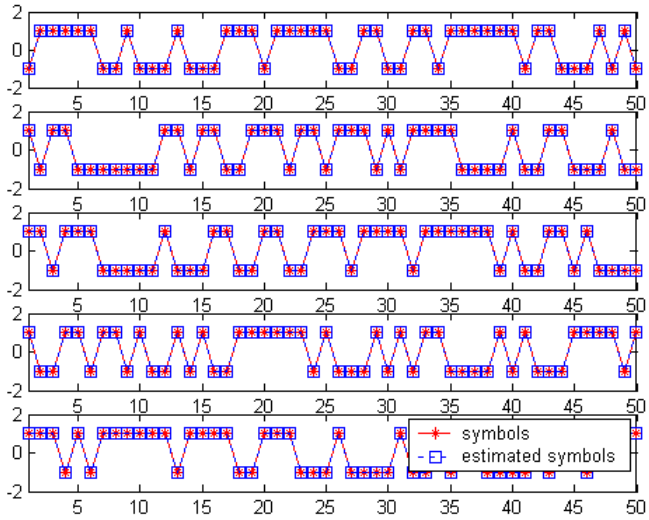

Figure 5.1 Estimation of CDMA symbols with FastICA for 5-users

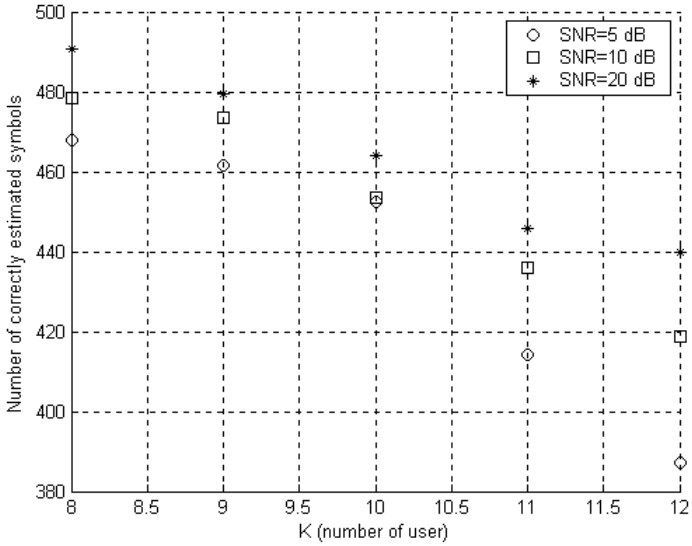

Figure 5.2 The number of correctly estimated symbols versus *K*

In Figure 5.3, the performance of FastICA detector is given for different number of users as a function of SNRs. The number of observation is 5000 and it is applied for 100 independent simulations. Also in Figure 5.4, MF is used for symbol detection and in this case there is no delay information. From Figures 5.3-5.4 it is clearly seen that FastICA outperforms matched filter detectors. At every SNR values, even when there are 12 users FastICA has a good performance and can detect symbols successfully. If

number of user increases, for low SNR values, FastICA detector's performance decreases but still it performs better than matched filter.

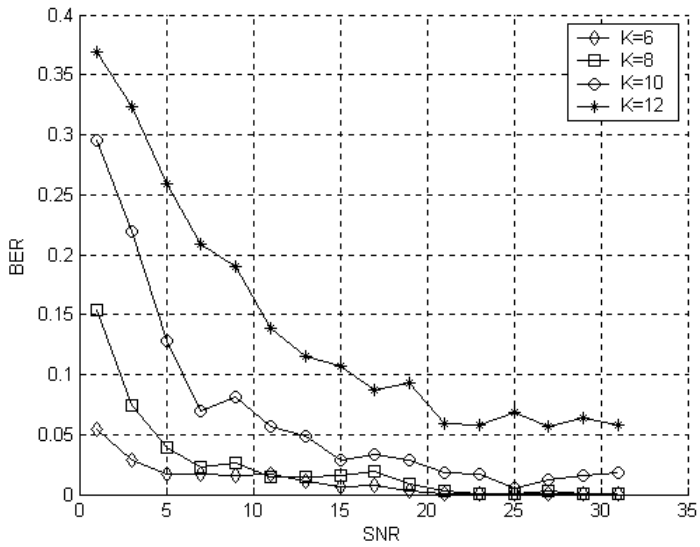

Figure 5.3 BER as a function of SNR with *K* as a parameter for FastICA

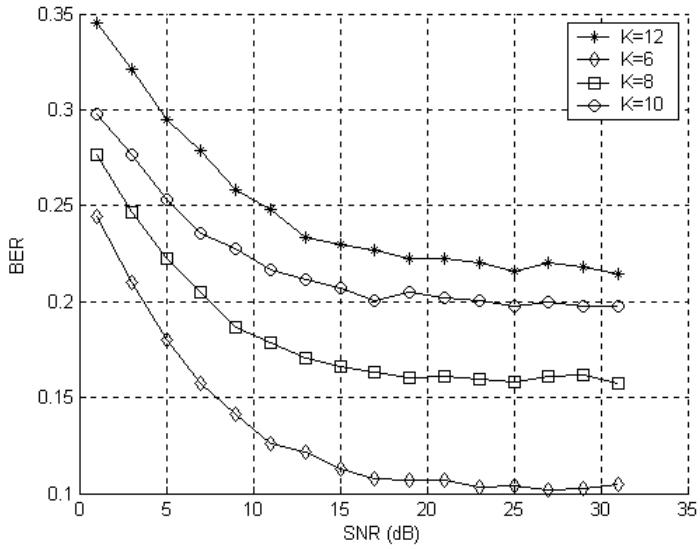

Figure 5.4 BER as a function of SNR with K as a parameter for MF

For the ideal case in which delays are assumed to be known, single user detector (SUD) estimates symbols for the first user with $\hat{\mathbf{b}}_m = sign\left[\mathbf{G}_{\text{firstuser}}^T \mathbf{r}_m\right]$ where $\mathbf{G}$ is

given in Eq. (5.10) and $\mathbf{G}_{\text{first user}}$ is the corresponding column of the first user. In Figure 5.5, FastICA, SUD and MMSE (Minimum Mean Square Error) are compared.

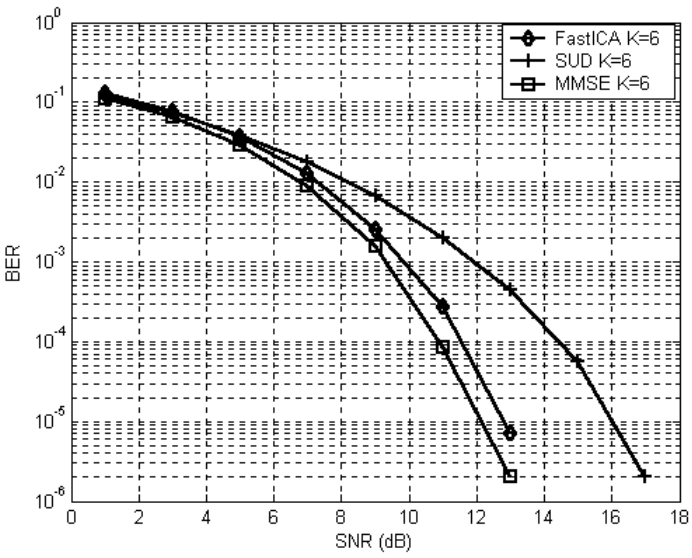

Figure 5.5 BER as a function of SNR for FastICA, SUD and MMSE

In this simulation, FastICA does not have any information about multipath delays or fading but it is assumed that the SUD and MMSE detector have perfect channel information. FastICA, SUD and MMSE detector are applied for 6 users and 4 paths which have 0 dB, -5 dB, -5 dB and -5 dB losses. The observation number is 5000 and 100 independent simulations are realized. It is seen from Figure 5.5 that FastICA outperforms SUD without channel information. For 10^{-5} BER value, SUD needs 4 dB more power compared to Fast ICA. Also it is seen that MMSE detector is superior to FastICA. For 10^{-5} BER value, FastICA needs 1 dB more power compared to MMSE detector but considering that FastICA reaches this performance without channel information, this performance degradation can be accepted.

5.3 Delay Estimation Using FastICA Algorithm in CDMA Systems

In CDMA systems, timing acquisition is an important subject for finding the synchronization information. FastICA algorithm can also be used for simultaneous demodulation of the symbol sequence and finding the synchronization information

[18]. In this application, the derivation of the compact matrix form of the downlink model is different from the derivations in section 5.1.

5.3.1 Signal Model in Delay Estimation for Slowly Fading Multipath Channels

The CDMA data and the *C*-vectors from subsequent discretized equispaced data samples have again the forms in Eq. (2.11) and (5.5), respectively. If a 2*C*-vector is defined as $\mathbf{q}_m = \left[\mathbf{r}_m \mathbf{r}_{m+1}^T\right]^T$, this vector can be shown by:

$$\mathbf{q}_m = \sum_{k=1}^{K}\left[b_{k,m-1}\sum_{l=1}^{L} a_l \mathbf{g}_{kl}^E + b_{k,m}\sum_{l=1}^{L} a_l \mathbf{g}_{kl}^F + b_{k,m+1}\sum_{l=1}^{L} a_l \mathbf{g}_{kl}^L\right] + \mathbf{n}_m \tag{5.22}$$

where $\mathbf{n}_m$ denotes the noise vector and $\mathbf{g}_{kl}^E, \mathbf{g}_{kl}^F, \mathbf{g}_{kl}^L$ denote the "early","full" and "late" parts of the code vectors with size *2C*, respectively, given as:

$$\begin{aligned}
\mathbf{g}_{kl}^E &= \left[s_k(C-d_l+1)\cdots s_k(C)0\cdots 0\right]^T, \\
\mathbf{g}_{kl}^F &= \left[0\cdots 0\ s_k(1)\cdots s_k(C)\,0\cdots 0\right]^T, \\
\mathbf{g}_{kl}^L &= \left[0\cdots 0\ s_k(1)\cdots s_k(C-d_l)\right]^T.
\end{aligned} \tag{5.23}$$

Here $\mathbf{g}_{kl}^F$ contains the original chip sequence of user *k*, padded by d_l zeros in the beginning and C-d_l zeros in the end [18]. Then the matrix $\mathbf{Q}=[\mathbf{q}_1.....\mathbf{q}_{N-1}]$ can be represented in compact form as:

$$\mathbf{Q} = \mathbf{GB} + \mathbf{N} \tag{5.24}$$

where $2C\times 3K$ mixing matrix

$$\mathbf{G} = \left[\sum_{l=1}^{L} a_l \mathbf{g}_{1l}^E \ \sum_{l=1}^{L} a_l \mathbf{g}_{1l}^F \ \sum_{l=1}^{L} a_l \mathbf{g}_{1l}^L \ \cdots \sum_{l=1}^{L} a_l \mathbf{g}_{Kl}^E \ \sum_{l=1}^{L} a_l \mathbf{g}_{Kl}^F \ \sum_{l=1}^{L} a_l \mathbf{g}_{Kl}^L\right] \tag{5.25}$$

contains the code vectors as column basis vectors and the attenuation coefficients and $3K\times(N-1)$ matrix $\mathbf{B}=[\mathbf{b}_1....\mathbf{b}_{N-1}]$ contains the symbols transmitted by the user with

$$\mathbf{b}_m = \left[b_{1,m-1}\ b_{1,m}\ b_{1,m+1} \cdots b_{K,m-1}\ b_{K,m}\ b_{K,m+1}\right]^T . \tag{5.26}$$

5.3.2 Synchronization Algorithm

Assuming that transmitted symbol sequences belonging to different users are independent, FastICA algorithm is applied to the CDMA data model Eq. (5.24) to estimate the unknown matrix **G** that contains code sequences and to estimate symbols simultaneously.

As it was mentioned in Chapter 3, firstly the data is whitened to normalize the component variances. At the same time some part of the noise is filtered out [18]. The new whitened data **Y** is given by

$$\mathbf{Y} = \mathbf{\Lambda}_S^{-1/2}\mathbf{U}_S^T\mathbf{Q} \tag{5.27}$$

where $\mathbf{\Lambda}_S$ and $\mathbf{U}_S$ corresponds to the *3K* eigenvalues and eigenvectors of correlation matrix estimate $\mathbf{QQ}^T/(N-1)$, respectively [18]. Then FastICA algorithm is applied where kurtosis is used as the contrast function. Since in CDMA downlink only one signal is of interest, estimating only one basis vector would be desirable. For this reason, initial vector $\mathbf{w}(0)$ has to be given with a good enough initial guess. This can be provided by training sequence $\mathbf{w}(0)=\mathbf{Y}_P[b_{11}....b_{1P}]^T$ where first *P* symbols of the user of interest are known and $\mathbf{Y}_P$ is the corresponding part of **Y** [18]. The procedure of applying FastICA algorithm was shown section 5.2.1. The only difference here is initialization of $\mathbf{w}(0)$ in step 1 since in this application, initialization of $\mathbf{w}(0)$ is not random and it depends on the first *P* symbols of the user of interest. Then the corresponding basis vector of the mixing matrix **G** in the non-whitened space will be found by a reverse transform [18]:

$$\mathbf{v} = \mathbf{U}_S\mathbf{\Lambda}^{1/2}\mathbf{w} . \tag{5.28}$$

In fact **v** is an estimate of the column of **G** corresponding to b_{1m}:

$$\mathbf{v} = \sum_{l=1}^{L} a_l \mathbf{g}_{1l}^{F} \ . \tag{5.29}$$

Let the index of the strongest path be 1 and d_1 be the corresponding delay which is needed to be found. The magnitudes of the fading terms of the secondary paths are smaller than the ones of the main path and it is assumed that the delay d_1 of the main path is smaller than other delays $\{d_l, l=2....L\}$. Under these assumptions, the estimate of **v** will contain values close to zero in absolute value, i.e. $|v_i|\approx 0$ for $i=1,....,d_1$, so the delay of the main path d_1 can be found by finding first index i for which $|v_i|>t$ where t is a threshold that can be found by using experimental results and it depends on channel noise and strength of the other paths [18].

5.3.3 Simulation Results

FastICA algorithm is tested for CDMA model in Eq. (2.11) with Gold codes of length C=31. The channel paths' powers are -5,-5 and 0dB for every user and SNR of the main path changes from 10 dB to 30 dB. The number of samples of the observed signal is N=300. The length of the training sequences for assuring convergence to the desired user of FastICA algorithm is P=20. In simulations, probability of acquisition (POA) which is the ratio of correctly estimated delays to total estimation number is analyzed. Optimum threshold values that are also used in simulations can be found from Figure 5.6. These threshold values are found through extensive simulations by trial. In Figure 5.7, the relation between probability of acquisition and number of user K is given for SNR=10dB and 20dB. It shows that up to 6 users, FastICA technique can estimate time delay successfully for both SNR values. So, the POA performance of FastICA is compared to MF and Constrained Minimum Output Energy (CMOE) detectors when K=2, 4 and 6 in Figures 5.8, 5.9 and 5.10. CMOE minimizes the output energy of the interfering sources and leaves the desired source undistorted. By using this method delay estimation can be found by using [18]

$$\hat{d} = \arg\min_{d} \mathbf{g}(d)^T \hat{\mathbf{C}}_{xx}^{-1} \mathbf{g}(d)$$

where $\mathbf{g}(d) = [s_1(C-d+1)\cdots s_1(C)\ s_1(1)\cdots s_1(C-d)]$ and $\hat{\mathbf{C}}_{xx}^{-1}$ is estimate of the autocorrelation matrix of the observed data.

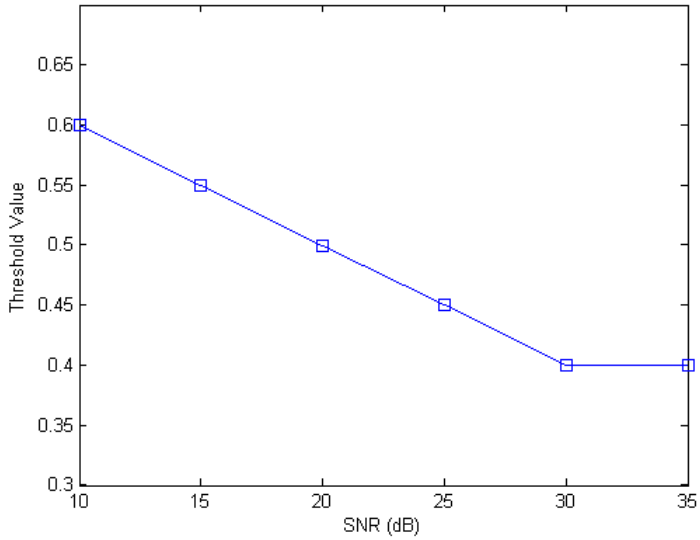

Figure 5.6 Choosing the threshold value

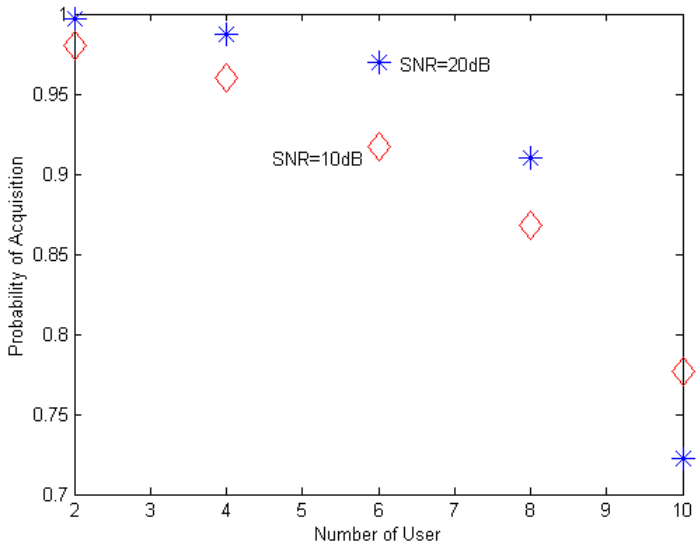

Figure 5.7 Probability of acquisition as a function of *K* for SNR=10 and 20dB

Figure 5.8 depicts that FastICA is superior to both MF and CMOE detectors when the number of users, *K*=2, 4 and 6. For *K*=2, FastICA can reach perfect POA at SNR=20 dB whereas MF detectors and CMOE detector cannot reach this performance. For *K*=6, FastICA detector can reach POA=0.95 where MF and the CMOE detectors can only achieve POA values of 0.9 even if the SNR is increased without bound.

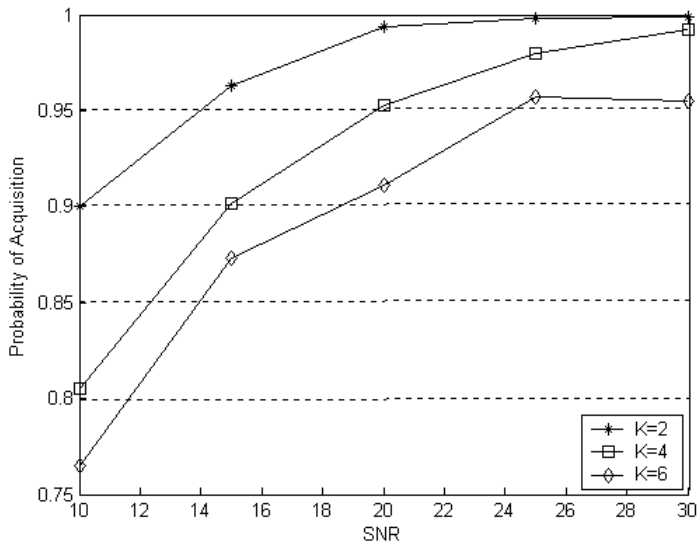

Figure 5.8 Probability of acquisition as a function of SNR and *K* for FastICA

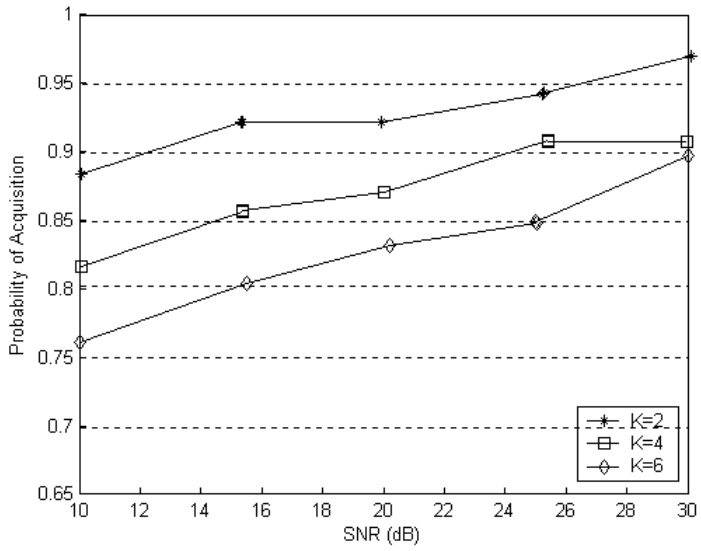

Figure 5.9 Probability of acquisition as a function of SNR and *K* for Matched Filter

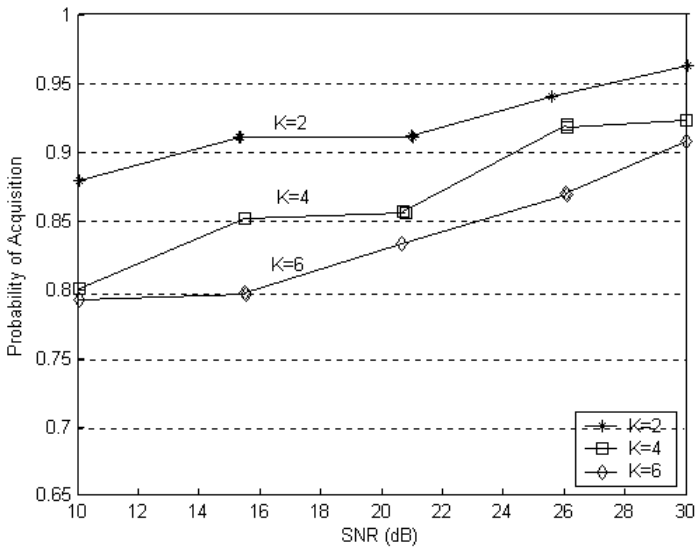

Figure 5.10 Probability of acquisition as a function of SNR and *K* for CMOE

The number of the training symbols is an important point in delay estimation and it affects the probability of acquisition. In simulation *P* is chosen as 10 and it can be seen from Figure 5.11 that even for two users and for high SNR values FastICA cannot estimate all delays correctly. In Figure 5.12, for two *P* values performances are compared. If *P* is chosen as 5, it is seen that performance of delay estimation fails and it cannot reach to 0.9 value of POA even for two users.

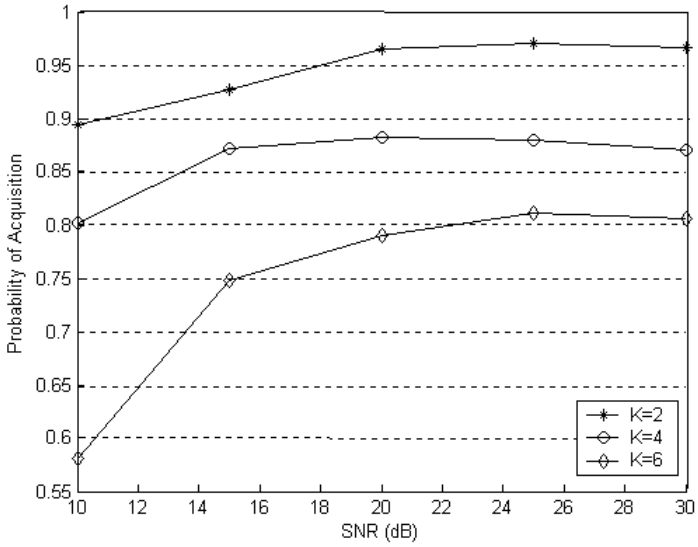

Figure 5.11 Probability of acquisition as a function of SNR and *K* for FastICA (*P*=10)

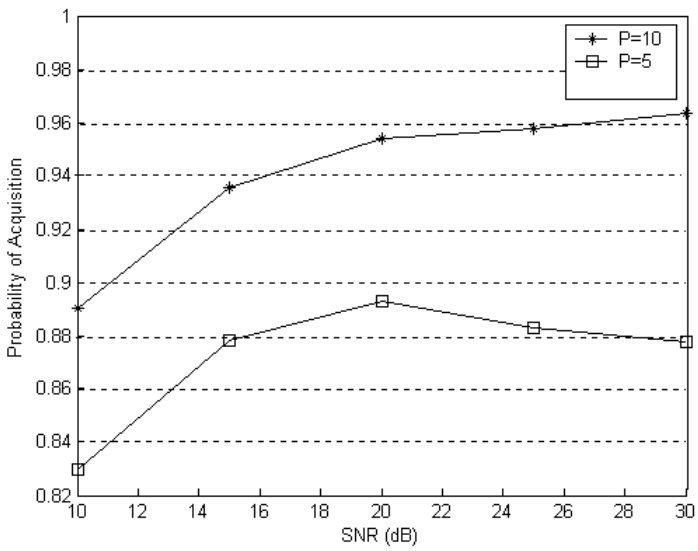

Figure 5.12 Probability of acquisition as a function of SNR and *P* for FastICA (*K*=2)

FastICA algorithm based on BSS for timing acquisition is noticed to perform better than matched filter or other conventional algorithms and also this technique needs only a few training symbols for estimating delays of the channels.

5.4 Blind Detection of DS-CDMA Signals based on Regularized ICA

Classical ICA algorithms are effective in separating non-Gaussian signals but they don't incorporate any information of the weighting matrix. For instance in CDMA problems, those algorithms do not use the signature sequence. Regularized ICA (Reg. ICA) is a new method that combines a contrast function and a regularization function to integrate the information of the user's sequence [24]. The traditional ICA has a problem in distinguishing which extracted source is the desired signal and determining its polarity. The new ICA detector solves this problem and it is flexible, robust and provides good interference suppression, fast convergence and low BER performance when compared with a MF. The only information required is the signature and timing of the desired user. However MMSE detector and decorrelating detector require the complete knowledge of all the users' signature sequences.

For this method's application, basic CDMA model with ideal channel assumption given in section 5.1.1. To goal is to estimate filter weight **w** such that filtered output can be used to estimate the data symbol.

5.4.1 Blind Reg. ICA detector

For this method, a new one unit cost function $J(\mathbf{w})$ is proposed to evaluate the filter weight $\mathbf{w}$ [24]:

$$J(\mathbf{w}) = \frac{1}{2}\{E[F(\mathbf{y})] - E[F(\mathbf{v})]\}^2 + \eta C(\mathbf{w}) \tag{5.30}$$

where $F(\cdot)$ is a smooth even function, $\mathbf{v}$ is a normalized Gaussian distributed random variable, η is the regularization parameter and $C(\mathbf{w})$ is the regularization functional. The contrast function is the measure of non-Gaussianity which is an approximation to the negentropy that measures the mutual information among the sources. For estimating the desired user's signals, "$C(\mathbf{w})$" regularization functional is used that incorporates prior information of the desired user's signature [24]. The filter weight estimation can be described mathematically:

$$\begin{aligned} \mathbf{w} &= \arg\max_{\mathbf{w}\in\Omega}\{J(\mathbf{w})\} \\ &= \arg\min_{\mathbf{w}\in\Omega}\left\{-\frac{1}{2}\{E[F(\mathbf{y})] - E[F(\mathbf{v})]\}^2 - \eta C(\mathbf{w})\right\} \end{aligned} \tag{5.31}$$

where Ω is the solution space for filter weight $\mathbf{w}$. By using stochastic gradient descent method to optimize Eq. (5.31), the detector weight update algorithm is given by:

$$\begin{aligned} \mathbf{w}(n+1) &= \mathbf{w}(n) - \gamma\left\{\frac{\partial[-J(\mathbf{w})]}{\partial\mathbf{w}}\right\} \\ &= \mathbf{w}(n) + \gamma\alpha(\mathbf{w})f(\mathbf{w}(n)^T\mathbf{r}(n))\mathbf{r}(n) + \lambda\frac{\partial C(\mathbf{w})}{\partial\mathbf{w}} \end{aligned} \tag{5.32}$$

where

$$\alpha(\mathbf{w}) = E[F(\mathbf{w}^T\mathbf{r})] - E[F(\mathbf{v})]. \tag{5.33}$$

In Eq. (5.32), $f(\cdot)$ is the derivative of $F(\cdot)$, γ is the learning step size for contrast function, $\lambda=\gamma\eta$ is the effective learning step size for the regularization functional. If

$F(\mathbf{y})=\mathbf{y}^4$ is chosen, $\alpha(\mathbf{w})$ becomes the kurtosis of $\mathbf{y}$ that ensures the identifiability and convergence of the detector. In this case:

$$\alpha(\mathbf{w}) = E\left[F(\mathbf{w}^T\mathbf{r})\right] - E\left[F(\mathbf{v})\right] = E\left[\mathbf{y}^4\right] - E\left[\mathbf{v}^4\right] \tag{5.34}$$

If the output $\mathbf{y}$ is normalized that $E\left[(\mathbf{w}^T\mathbf{r})^2\right] = E\left[\mathbf{y}^2\right] = 1$ and $E\left[\mathbf{v}^2\right] = 1$ since Gaussian random variable is normalized, $\alpha(\mathbf{w})$ is equal to the kurtosis of $\mathbf{y}$:

$$\begin{aligned}\alpha(\mathbf{w}) &= E\left[\mathbf{y}^4\right] - 3\left[E\left(\mathbf{y}^2\right)\right]^2 - \left\{E\left[\mathbf{v}^4\right] - 3\left[E\left(\mathbf{v}^2\right)\right]^2\right\} \\ &= \mathrm{kurt}(\mathbf{y}) - \mathrm{kurt}(\mathbf{v}) \\ &= \mathrm{kurt}(\mathbf{y})\end{aligned} \tag{5.35}$$

Both the data sources and filtered output are sub-Gaussian distributed which means that kurt(b_i)<0 and kurt($\mathbf{y}$)<0. The kurtosis of $\mathbf{y}$ can be found as:

$$\mathrm{kurt}(\mathbf{y}) = \mathrm{kurt}\left(\mathbf{w}^T\mathbf{r}\right) = \mathrm{kurt}\left(\mathbf{w}^T\mathbf{G}\mathbf{b}\right) = \sum_i \left(\mathbf{G}^T\mathbf{w}\right)^4 \mathrm{kurt}(b_i). \tag{5.36}$$

Since kurt(b_i)<0 for binary antipodal modulation, kurt($\mathbf{y}$)<0. In order to achieve good results, it is sufficient to estimate $\alpha(\mathbf{w})$'s polarity since it is part of the learning factor. If positive scalar β is introduced as a substitute for $-\alpha(\mathbf{w})$, Eq. (5.32) can be rewritten as:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu f(\mathbf{w}(n)^T\mathbf{r}(n))\mathbf{r}(n) + \lambda\frac{\partial C(\mathbf{w})}{\partial \mathbf{w}} \tag{5.37}$$

where $\mu = \gamma\beta$. $C(\mathbf{w})$ should be chosen as a meaningful, differentiable function with the optima at $\mathbf{w}=\boldsymbol{s}_j$ under the constraint $||\mathbf{w}||=1$ [24]. If a second order cost function is used for regularization functional [24]:

$$C(\mathbf{w}) = \frac{1}{2}\left(\mathbf{w}^T\mathbf{s}_j\right)^2 = \frac{1}{2}\mathbf{w}^T\mathbf{S}_j\mathbf{w} \tag{5.38}$$

where $\mathbf{S}_j = \mathbf{s}_j \mathbf{s}_j^T$.

If Eq. (5.38) and $F(\mathbf{y}) = \frac{1}{4}\mathbf{y}^4$ is substituted into Eq. (5.37), regularized ICA detector weight estimation algorithm is obtained as:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\left(\mathbf{w}(n)^T \mathbf{r}(n)\right)^3 \mathbf{r}(n) + \lambda(\mathbf{s}_j^T \mathbf{w}(n))\mathbf{s}_j . \tag{5.39}$$

5.4.2 Simulation Results

In simulations, a CDMA system of K=5 users and C=31 Gold codes is used where the modulation is binary phase shift keying (BPSK) modulation. Signal to interference noise ratio (SINR) of the j'th user at n'th iteration is defined as [24]:

$$SINR(n) = \frac{\sum_{r=1}^{S} A_j^2 \left(\mathbf{s}_j^T \mathbf{w}(n)\right)^2}{\sum_{r=1}^{S}\left[\sum_{k \neq j}^{K} A_k^2 \left(\mathbf{s}_k^T \mathbf{w}(n)\right)^2 + \mathbf{w}(n)^T \mathbf{R}_{nn} \mathbf{w}(n)\right]} \tag{5.40}$$

where S is number of independent simulations and $\mathbf{R}_{nn}$ is the noise autocorrelation matrix.

In the first experiment, signal power of the desired user is 10dB and other users' signal powers are 30dB. In Figure 5.13 the SINR of desired user is given. It is observed that after 400 symbols, SINR is found to approach the optimum value of 10dB, on the other hand, matched filter does not offer any SINR improvement.

In the second experiment, again all the interfering signal powers are 30dB and desired user's signal power varies from 4dB to 12dB. After the steady state has been achieved BER is computed by 1000 symbols in Figure 5.14. It is observed from the figure that Regularized ICA detector offers better BER performance over the MF detectors for basic CDMA model under the ideal channel assumptions.

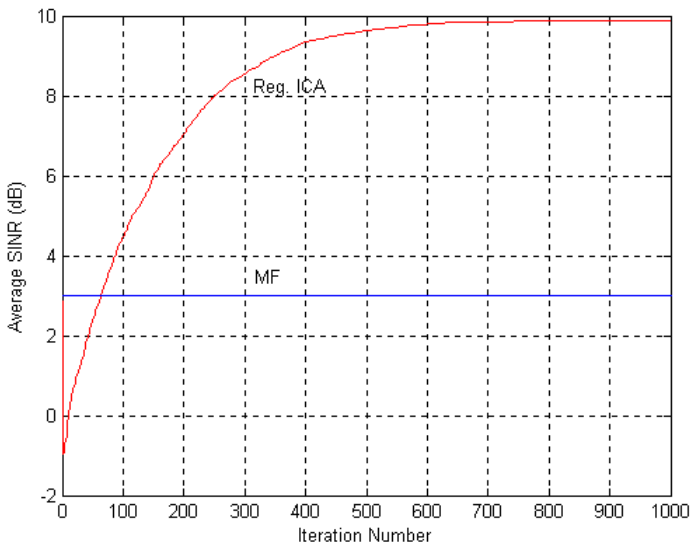

Figure 5.13 Average SINR performance as a function of iteration number of MF and ICA ($\lambda=\mu=2\times10^{-5}$)

Since this technique integrates the user's signature sequence to address the order and polarity indeterminacy of classical ICA algorithms, there is no need to use training symbols to estimate desired user's symbols. Also with this technique, MAI can be suppressed as it is shown in Figure 5.13. This system is applied for basic CDMA model but then, adaptation of this system to multipath CDMA models can be worked and analyzed as a future work.

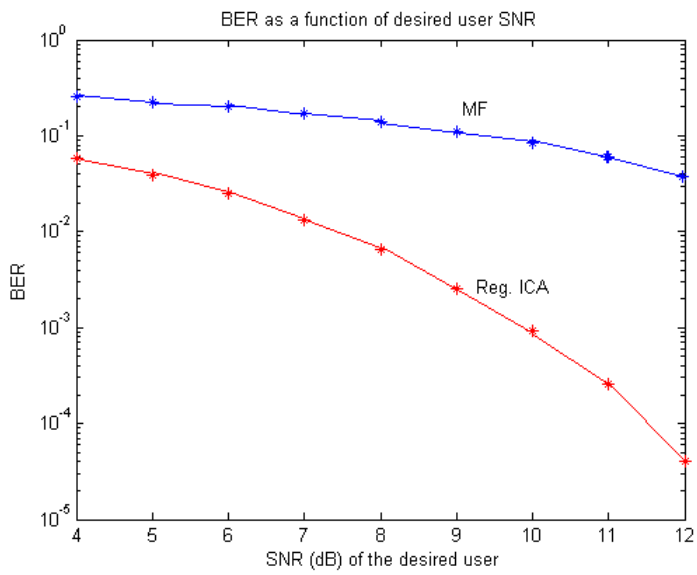

Figure 5.14 BER performances of ICA and MF as a function of user SNR (*K*=5, for the other users SNR= 30 dB)

5.5 IFA Detector for Basic CDMA Downlink Model

IFA is a new algorithm for blind seperation of noisy mixtures with non-square matrix. As it was mentioned in Chapter 4, for high noisy situations, IFA can outperform ICA. In this section, IFA is applied to the basic CDMA downlink model in Eq. (5.1) and compared with Reg. ICA given in previous section. For these applications of CDMA, a matrix form of the basic CDMA model in Eq. (5.4) is used.

If Eq. (5.4) is rewritten in open form:

$$\mathbf{r}_{C\times N} = [\mathbf{s_1}.....\mathbf{s_K}]_{C\times K}\,\mathrm{diag}[A_1.....A_K]\begin{bmatrix}\mathbf{b}_1\\ \vdots \\ \mathbf{b}_K\end{bmatrix}_{K\times N} + \begin{bmatrix}\mathbf{n}_1\\ \vdots \\ \mathbf{n}_C\end{bmatrix}_{C\times N} \tag{5.41}$$

where $\mathbf{s}_j$ are code sequences of users, A_i are amplitudes of users, $\mathbf{b}_i$ are transmitted symbols. In simulations, IFA is applied to the model in Eq. (5.4) and results are obtained to compare ICA and IFA in symbol detection for the CDMA downlink model.

5.5.1 Simulation Results

In simulations, a CDMA system of K=2 users with C=31 Gold codes is considered. In the first experiment signal power of the desired user is varied from 1 to 12 dB while signal power of the interfering user is fixed to 30 dB. IFA, ICA and matched filter detectors are applied for 10 independent simulations and 1000 transmitted symbols.

In Figure 5.15, the number of correctly estimated symbols is given for ICA, IFA and matched filter detector. It can be seen clearly from the figure that IFA can estimate all transmitted symbols correctly even though there is high noise (variance of noise is 0.7) and the desired user's signal power is low. On the other hand, Reg. ICA can estimate all symbols correctly after the 6dB of desired user's power.

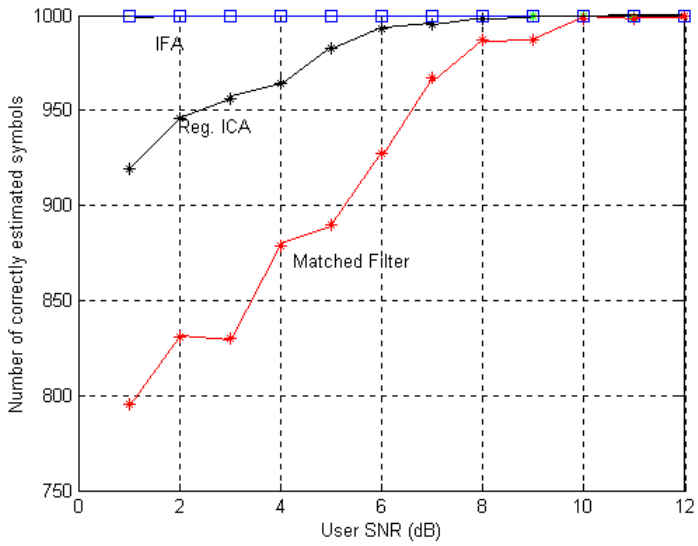

Figure 5.15 Number of correct estimated symbols as a function of User SNR for IFA Reg. ICA and MF

Chapter 6

Estimating Non-Gaussian Fading Channels in CDMA Systems

In this chapter, PS-ICA is used for estimating non-Gaussian fading channels. Considering some fading channel measurements showing that the fading channel coefficients may have an impulsive nature, these coefficients are modeled with an α-stable distribution whose shape parameter α takes values between 1.8 and 1.9. These α values show that the distribution resembles a Gaussian distribution but has a more impulsive nature. In this chapter, also IFA method is applied for estimating fading channels in a basic CDMA model in order to achieve faster convergence.

6.1 Non-Gaussian Fading Channels in CDMA Communication

In mobile communication systems, multipath fading channels are modeled by Gaussian distribution. This Gaussian assumption for the in phase and quadrature components of the received signal leads to Rayleigh and Rice distributions for the envelope of the signal and to a uniform distribution for the phase of the signal [12]. Many systems are designed with Gaussian assumption of the multipath channels in mobile communication. However, empirical results show that Gaussian model does not hold for a large variety of wireless fading channels every time [13]. Many measurements performed in Europe over the GSM system show that the received in-phase and quadrature components are non-Gaussian and so the envelope of the signal cannot be Rayleigh or Rician distributed. This deviation of channel statistics from the "Gaussianity" assumption leads to a degradation in system performance [14]. In the literature, distributions which have more impulsive characteristics than Gaussian distribution as Middleton Class A and Gauss-Laplace models were used for modeling channel coefficients and also SIRP (spherically invariant random process) was used for non-Gaussian channel models [12, 13, 14, 30].

In this work, fading channel coefficients are modeled with an α-stable distribution whose "shape parameter" or "characteristic exponent" α takes values between 1.8 and 1.9. These values show that the distribution resembles a Gaussian distribution but has a more impulsive nature [31, 32]. α-stable distribution is given in more detail in Appendix E.

6.2 Non-Gaussian Fading Channel Estimation by PS-ICA

PS-ICA is a BSS method that minimizes mutual information using a PS-based parametric model. When source distributions are nearly Gaussian or have same kurtosis with the Gaussian distribution PS-ICA can separate sources faster than a conventional ICA method. In Chapter 3, Pearson System based ICA is given with more details.

This method is used in estimating non-Gaussian fading coefficients of the multipath channel in CDMA communication. As it is explained in the beginning of the chapter, the coefficients of fading channels are usually modeled by a Gaussian distribution but sometimes other cases exist where non-Gaussian distributions model these channels more successfully. In the model, the fading channel is assumed to be nearly Gaussian. To model this case, an α-stable distribution is used. The channel in the CDMA downlink model used in this application is assumed to be the fast fading multipath channel model given in Eq. (2.12). The fading channel coefficients $\{a_{lm}; m=1, ...,N; l=1,\ldots,L\}$ have an α-stable distribution. The ICA formulation of the CDMA downlink signal in a fast fading multipath channel scenario was given in section 5.1.2.2.

In this system, it is assumed that channel coefficients are constant for some number of bits. With the help of training symbols $b_{\mathrm{km}}=1$, **F** becomes containing only channel coefficients. After separating the received signal into its real and imaginary parts as shown in Figure 6.1, PS-ICA is applied and coefficients of fading channels are estimated.

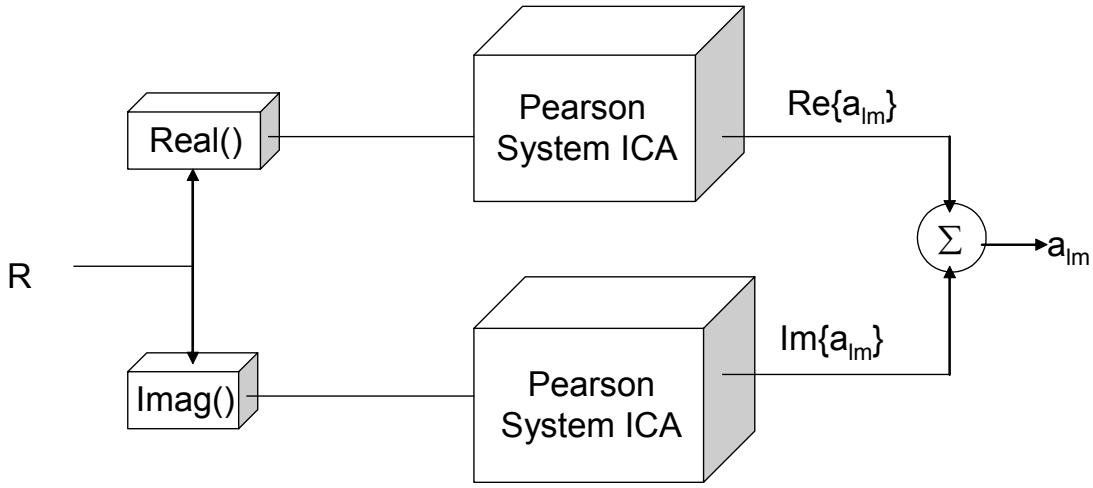

Figure 6.1 Separation of received signal into the real and the imaginary parts

6.2.1 Simulation Results

In simulations, CDMA downlink model given in Eq.(2.12) is used for K=2 users and L=3 paths. 1000 symbols are assumed to be transmitted and only real part of the received signal is considered. In Figure 6.2, 1.85-stable fading channel is estimated and the system follows channel coefficients. For the same model ICA can estimate fading coefficients (Figure 6.3) but PS-ICA performs faster than ICA.

If α is chosen as 1.95 which corresponds to a stable distribution very close to Gaussian, PS-ICA can still estimate channel but classical ICA cannot give the same performance for the same number of iterations (Figures 6.4-6.5).

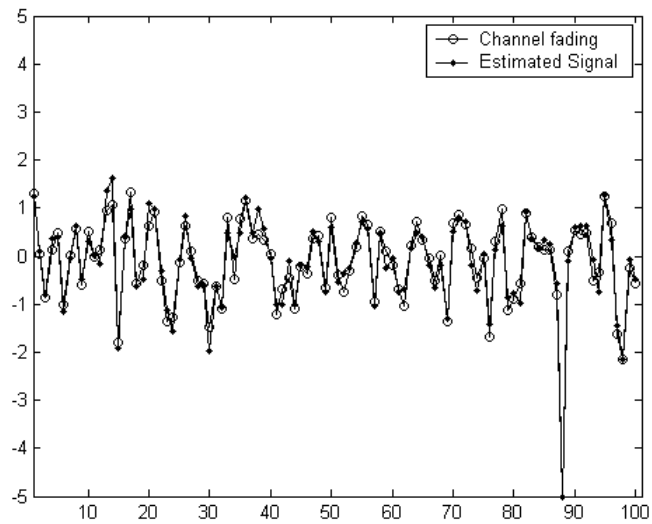

Figure 6.2 Fading channel estimation by Pearson System based ICA (alpha=1.85)

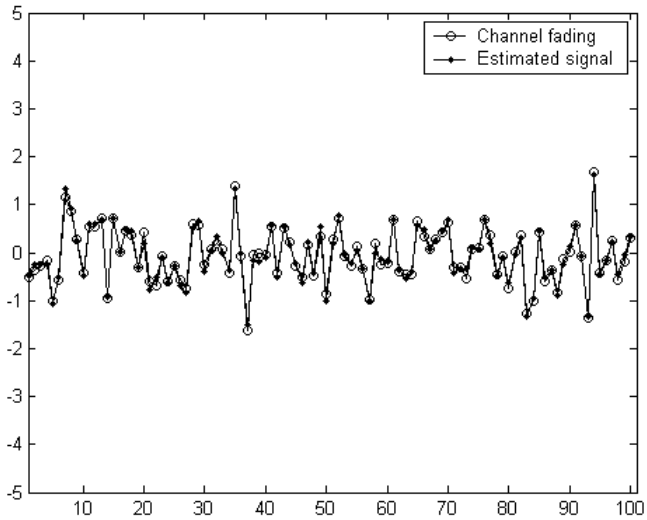

Figure 6.3 Fading channel estimation by classical ICA (alpha=1.85)

As a result it can be said that PS-ICA can be chosen for estimating non-Gaussian fading channels for faster convergence when the channel fading coefficients have a distribution close to Gaussian. Also it is more robust than conventional ICA to variations in channel statistics since PS can model a wide class of distributions like Rayleigh and log-normal distributions.

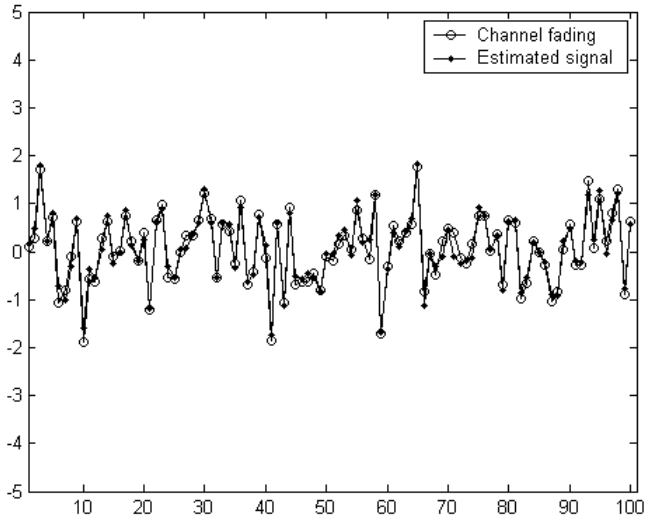

Figure 6.4 Fading channel estimation by Pearson System based ICA (alpha=1.95)

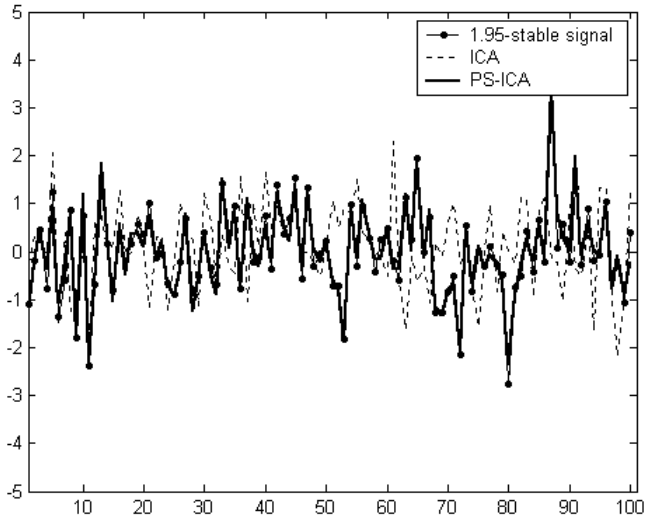

Figure 6.5 Fading channel estimation by classical ICA and PS-ICA (alpha=1.95)

Figure 6.6 shows estimation error as a function of SNR for PS-ICA and classical ICA. In 10-30 dB SNR interval, for PS-ICA, estimation error stays in acceptable values where classical ICA has worse performance than PS-ICA when alpha is 1.90.

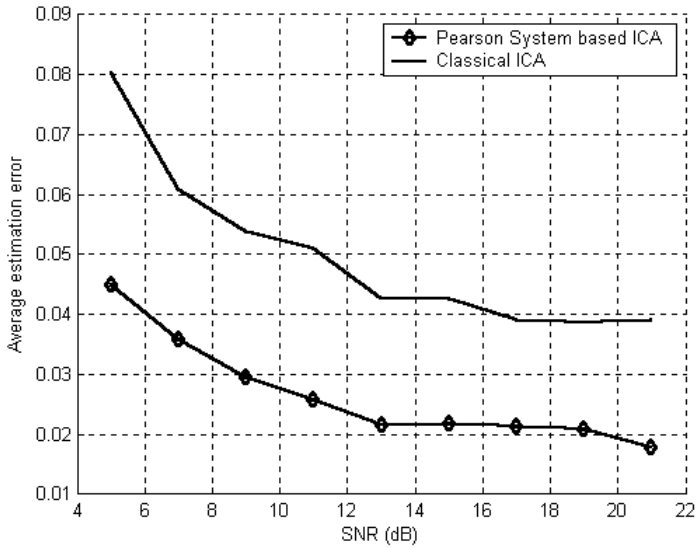

Figure 6.6 Estimation errors as a function of SNR (α=1.90)

Lastly, for different numbers of training symbols, channel estimation performances are analyzed where SNR is 6 dB. Like the other methods numbers of training symbols affects channel estimation which is given in Figure 6.7. The figure depicts that to estimate fading coefficients with acceptable average estimation error interval, at least 1000 observation symbols are needed.

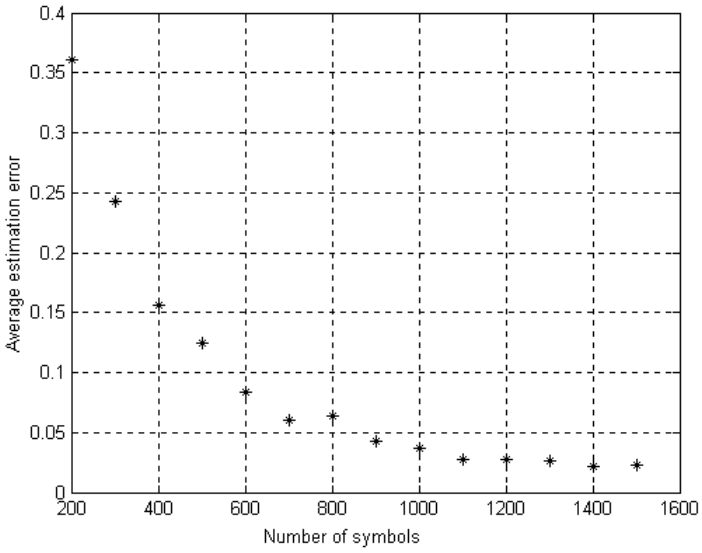

Figure 6.7 Average estimation error as a function of number of symbols

6.3 Channel Estimation by IFA in CDMA System

As explained in Chapter 4, in IFA method, each source distribution is modeled by MOG. Because of this, IFA can separate sources successfully for wide class of distributions. Especially increasing the number of Gaussians in modeling sources can make more successful separations but on the other hand this decreases the convergence speed.

In this section IFA is applied to CDMA system for estimating channel coefficients. These fading coefficients are modeled by an alpha-stable distribution that has alpha values close to 2. This makes the distribution close to Gaussian. The convergence speed problem of IFA exists again in this application. So, to overcome this problem, again the basic model is used as it is given in simulation part 6.3.1.

In this application there is an advantage of using IFA in channel estimation. In this case, in contrast to the channel estimation by PS-ICA, instead of choosing all users' symbols equal to 1, only one user's symbols can be chosen 1. So there is no need to assume constant channel coefficients for a known number of symbols. Even when the coefficients vary at every symbol period, IFA can estimate fading coefficients.

6.3.1 Simulation Results

In the simulations, fast fading CDMA model in Eq. (5.16) is used. In the experiments, K=2 and L=1 are chosen to limit the number of sources to 4 to overcome the convergence problem of IFA. Fading coefficients are modeled by 1.85-stable distribution. In Figures 6.8 and 6.9, it is seen that IFA estimator tracks channel coefficients for both noise variances 0.1 and 0.01. According to simulation results, it must be pointed that both PS-ICA and IFA can estimate channel if channel coefficients have a distribution close to Gaussian. In the case of IFA method using pilot symbols for only a single user is enough to estimate the channel where ICA method requires pilots for every user.

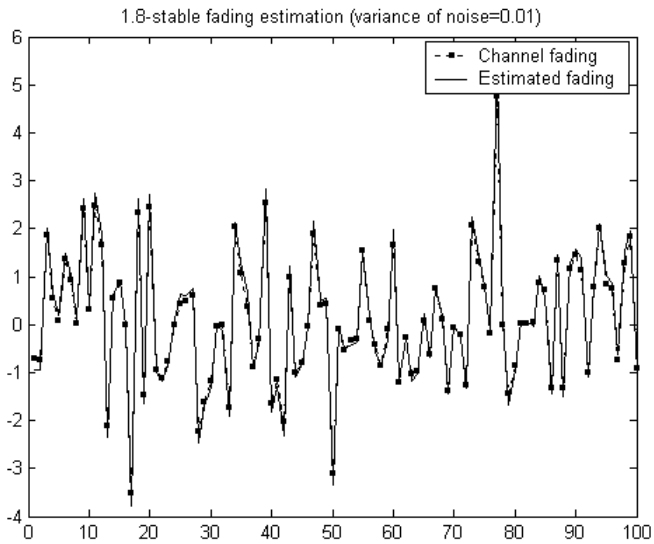

Figure 6.8 Fading channel estimation (variance of noise = 0.01)

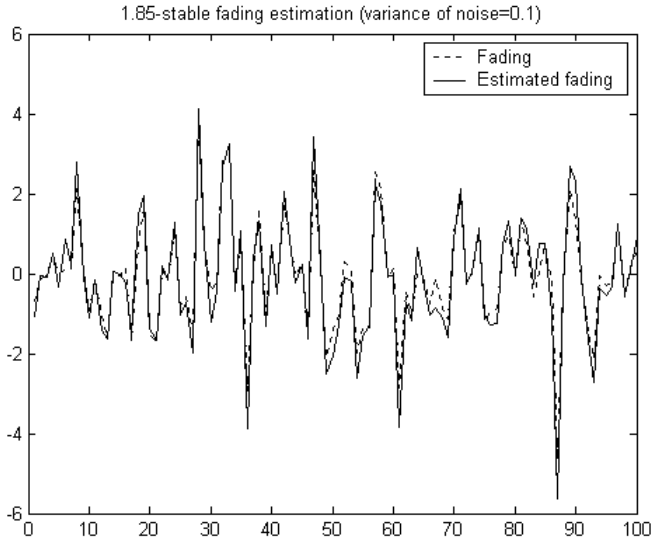

Figure 6.9 Fading channel estimation (variance of noise = 0.1)

Estimating transmission channel with IFA is a new method and in this work, to overcome convergence speed problem of IFA, the CDMA system with two users and one path is considered but for more practical CDMA systems IFA is needed to speed up by introducing modifications into its formulation.

Chapter 7

CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

CDMA is a multiple access technique where all users use the same frequency and transmit simultaneously. In CDMA, spread spectrum modulation is used and each user has its own unique spreading code. The receiver side of CDMA systems is more complex than other multiaccess system receivers. Because of non-orthogonal code sequences, simple correlator receivers which handle the MAI like the additional noise cannot achieve good performances. So in CDMA downlink transmission, multiuser detection algorithms continue to be a very active area of research.

In the CDMA downlink problem, the receiver is interested in only one of the information sources transmitted by the base station. The desired user has not got any information about other users' code sequences. Because of this, separation of desired user's signal from other users' signals is a typical blind source separation problem.

In the BSS, source signals are mixed together by some linear transformation corrupted by noise and observation data measured by sensors arise from that mixture of sources. In BSS problem, those source signals are recovered by using observations. One of the techniques used in BSS is ICA. The sources can be found using ICA provided that all sources are independent and non-Gaussian.

In this thesis, examples of ICA applications in CDMA downlink problem are provided. First of all CDMA downlink model that is used in ICA applications are described. Two downlink models are used. First one is a basic CDMA downlink model in which ideal channel without multipath fading is considered and the second one is DS-CDMA downlink model in a multipath fading channel. These models are expressed in matrix form which is similar to classical noisy ICA models. Using FastICA algorithm, symbols are detected. Simulation results are provided to demonstrate that without any information about code sequences, symbols of the desired user can be detected successfully even in low signal-to-noise ratios. In the case

that FastICA algorithm is used for detecting symbols in CDMA systems, there is no need to estimate other channel parameters by using extra systems.

In CDMA systems, time acquisition is an important subject. Simultaneously, ICA algorithm can detect symbols and multipath delays for time acquisition successfully under some assumptions. Simulation results show that the delay estimation capability seems to be best with FastICA algorithm. Also this technique needs only few training symbols in delay estimation. However with this method, only delays which are integer times of chip duration can be estimated but as a future work, standard delay profiles and chip pulse shaping can be used to consider more realistic channels and non-integer delays.

Also blind regularized ICA detector is shown. In this method, prior knowledge which is the code sequence of the desired user, is used in ICA to detect desired user's symbols correctly. Since this technique integrates the user's signature sequence to address the order and polarity indeterminacy of classical ICA algorithms, there is no need to use training symbols to estimate desired user's symbols. It is also shown that multiple access interference can be suppressed by regularized ICA successfully. As it is mentioned, this technique is applied for basic CDMA model in which there is no multipath fading or delays. As a future work, adaptation of this technique to a multipath CDMA model that is more realistic than the model used in this application can be studied and analyzed.

A new method of BSS, IFA is also applied for basic CDMA downlink model and simulation results are given to compare regularized ICA and IFA. It is shown that for small number of users, IFA can give good detection performance even when the desired user has low power with respect to other users. In the cases with many users, IFA has disadvantages and can fail. It has time and speed problem.

In this thesis, the non-Gaussian fading coeffients of a CDMA transmission channel are estimated by PS-ICA. In mobile communication systems, multipath fading channels are modeled by Gaussian distribution and many receiver systems are designed for Gaussian assumption of the fading characteristics. However many measurements show that sometimes fading characteristics can deviate from Gaussian distribution. For these cases existing estimators can fail in getting channel information. In this thesis, to model fading that deviate from Gaussian, alpha-stable distribution is used. For alpha values close to 2, the alpha stable distribution resembles to a Gaussian

distribution. Because of this 1.8-1.95-stable distribution is used for modeling multipath fading characteristic. Since PS-ICA is a fast technique for separating sources that have distribution close to a Gaussian distribution, PS-ICA is applied for estimating fading coefficients. Simulation results show that our system can estimate fading coefficients in acceptable error margins.

Future research should focus on IFA and IFA applications in CDMA systems. Since IFA has a better performance than ICA in very noisy cases, it can separate symbols in low signal to noise ratios. Also in IFA since sources are modeled by MOG, it can estimate fading coefficients correctly regardless of the fading distributions and it can be more robust to variations in fading characteristics. For IFA applications, number of users or paths is very important since in many sources cases, IFA is very slow and complex. To solve this problem some approximations can be done as a future work. After solving this problem, applying IFA to CDMA systems will be easier. For these reasons, IFA applications in CDMA will be a very active area of research.

REFERENCES

[1] J. S. Lee and L. E. Miller. "*CDMA Systems Engineering Handbook*," Artech House: Boston, 1998.

[2] "*Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband spread spectrum Cellular System,*" TIE / EIA Interim Standard 95, Washington DC: Telecommunication Industry Association, July 1993.

[3] S. Verdu. "*Multiuser Detection,*" Cambridge University Press, 1998.

[4] H. Attias. "*Independent Factor Analysis,*" Neural Computation, 11:803-855, 1999.

[5] A. Hyvarinen. "*One-unit contrast functions for independent component analysis: a statistical analysis,*" Proc. IEEE Workshop on Neural Networks for Signal Processing: pp. 388-397, Sept 1997.

[6] K. N. Leach. "*A survey paper on independent component analysis*," System Theory: pp. 239-242, March 2002.

[7] E. Lawrey. "*The suitability of OFDM as a modulation technique for wireless telecommunication with a CDMA comparison*" master thesis, James Cook University, Australia, 1997.

[8] Mobile Communication, Chapter 3: Media Access available at www.ipd.uka.de/~koenig/MODIS/17.10.01.2.ppt

[9] J. Meel. "*Spread Spectrum,*" De Nayer Institute, 1999.

[10] T. S. Rappaport. "*Wireless Communication,*" Prentice Hall, 1996.

[11] M. C. Jeruchim, P. Balaban and K. Sam Shanmugan. "*Simulation of Communication Systems: Modeling, Methodology and Techniques*, " Kluwer Academic, Plenum Publisher, New York, 2000.

[12] C. Chayawan, V. A. Aalo. "*Performance study of MRC systems with multiple cochannel interferers in a non-Gaussian multipath fading environment*," Vehicular Technology Conference, 2002. Proceedings. VTC 2002-Fall. 2002 IEEE 56th: Volume: 3, pp. 1720-1724, vol.3, 24-28 Sept. 2002.

[13] A. Abdi, H. A. Barger, M. Kaveh. "*Signal Modeling in wireless fading channels using spherically invariant process*," IEEE International Conference on Acoustics, Speech and Signal Processing: vol. 4, pp. 2297-3000, 2000.

[14] K.H. Biyari, W.C. Lindsey. "*Error performance of DPSK mobile communication systems over non-Rayleigh fading channels*," Vehicular Technology, IEEE Transactions on, Volume: 44, Issue: 2, pp. 211 – 219, May 1995.

[15] A. Hyvarinen. "*A family of fixed point algorithms for independent component analysis*," IEEE Int. Conf. on Acoustic, Speech and Signal Processing: vol. 5, pp. 3917-3920, April 1997.

[16] A. Hyvärinen. "*Survey on Independent Component Analysis*," Neural Computing Surveys 2, pp. 94-128, 1999.

[17] S. Van Vaerenbergh. "*Ica and FastICA*", March 1997.

[18] R. Critescu, T. Ristaniemi, J. Joutsensalo, J. Karhunen. "*CDMA delay estimation using FastICA algoritm,*" IEEE Int. Symp. on Personal, Indoor and Mobile Radio Comm:, Vol. 2, pp. 1117-1120, Sept. 2000.

[19] J. Karvanen, V. Koivunen. "*Blind separation methods based on Pearson System and its extensions*," Signal Processing, Vol. 82, pp. 663-667, April 2002.

[20] J. Karvanen, J. Eriksson and V. Koivunen. "*Pearson System based method for blind separation*," ICA 2000, pp. 585-590, 2000.

[21] D. T. Pham. "*Blind separation of instantaneous mixture of sources via independent component analysis*," IEEE Trans. Signal Processing, pp. 2768-2779, 1996.

[22] J. Karvanen, V. Koivunen. Pearson System-Based ICA Code available at http://www.tsi.enst.fr/icacentral/algos.html.

[23] J. G. Proakis, "*Digital Communications*", McGraw-Hill, Inc, New York, 3rd edition, 1995.

[24] Kim-Hui Yap, L. Guan and J. Evans. "*Blind adaptive detection for CDMA systems based on regularized independent component analysis*", GLOBECOM 2001, vol. 1, pp. 249-253, Nov. 2001.

[25] A. J. Viterbi. "*CDMA: Principles of Spread Spectrum Communication*", Addisson Wesley Longman, Inc., 1995.

[26] J. Joutsensalo, T. Ristaniemi. "*Learning algorithms for blind multiuser detection in CDMA downlink*", IEEE Int. Symp on Personal, Indoor and Mobile Radio Comm., vol. 3, pp. 1040-1044, Sept. 1998.

[27] R. Critescu, J. Joutsensalo, T. Ristaniemi, "Fading channel estimation by mutual information minimization for Gaussian stochastic processes", 2000 IEEE Int. Conf. on Comm., Vol.1, pp 56-59, 18-22 June 2000.

[28] S. E. El-Khamy, M. Lotfy, A. S. Badaway and A. Farid. "*On the blind multiuser detection of DS-CDMA signals using the independent component analysis* ", 20. National Radio Science Conference, March 2003.

[29] A. Hyvarinen. FastICA code at http://www.cis.hut.fi/projects/ica/fastica/.

[30] K.H.Biyari, W.C.Lindsey. "*A quasi-moment approach for the analysis of non coherent communications over complex non-Gaussian fading channels*" Global Telecommunications Conference, 1993, including a Communications Theory Mini-Conference. Technical Program Conference Record, IEEE in Houston. GLOBECOM'93, IEEE, pp.429-1432, vol.3, 29 Nov.-2 Dec. 1993.

[31] M. Shao, C. L. Nikias. "*Signal Processing with fractional lower order moments: Stable Processes and their applications*," Proc. IEEE, July 1993.

[32] G. Samaradotnitsky, M. S. Taqqu. "*Stable non-Gaussian random processes*," Chapman and Hall, 2000

APPENDIX A

Definition of Kurtosis

Consider a scalar random variable of zero mean, say x, whose characteristic function is denoted by $\hat{f}(t)$ [16]:

$$\hat{f}(t) = E\{\exp(itx)\} \tag{A.1}$$

Expanding the logarithm of the characteristic function as a Taylor series, one obtains

$$\log \hat{f}(t) = \kappa_1 (it) + \kappa_2 (it)^2 / 2 + \cdots + \kappa (it)^r / r! + \cdots \tag{A.2}$$

where the κ_r are some constants. These constants are called the cumulants of x. In particular, the first three cumulants (for zero mean variables) have simple expressions:

$$\begin{aligned} \kappa_1 &= E\{x\} = 0 \\ \kappa_2 &= E\{x^2\} \\ \kappa_3 &= E\{x^3\} \end{aligned} \tag{A.3}$$

Fourth-order cumulant, called kurtosis, can be expressed as:

$$\mathrm{kurt}(x) = E\{x^4\} - 3\left(E\{x^2\}\right)^2 \tag{A.4}$$

Kurtosis can be considered a measure of the non-Gaussianity of x. For a Gaussian random variable, kurtosis is zero.

APPENDIX B

Derivation of FastICA Algorithm

The iterative optimization algorithm used here to estimate w, is called *Newton's method*.

Given a cost function F(w), Newton's method applies the iteration [17]

$$\mathbf{w} \leftarrow \mathbf{w} - \left[\frac{\partial^2 F(\mathbf{w})}{\partial \mathbf{w}^2} \right]^{-1} \frac{\partial F(\mathbf{w})}{\partial \mathbf{w}} \tag{B.1}$$

to estimate its maxima.

The measure of non-Gaussianity will be the approximation of negentropy $J(\mathbf{w}^{\mathrm{T}}\mathbf{z})$. The maxima of this approximation are typically obtained at certain optima of $E\{G(\mathbf{w}^{\mathrm{T}}\mathbf{z})\}$. The function G can be any of the nonquadratic functions.

The optima of $E\{G(\mathbf{w}^{\mathrm{T}}\mathbf{z})\}$ under the constraint $E\{(\mathbf{w}^{\mathrm{T}}\mathbf{z})^2\}=||\mathbf{w}||^2=1$ are obtained using its Lagrangian

$$L(\mathbf{w}) = E\{G(\mathbf{w}^T\mathbf{z})\} + \beta(\|\mathbf{w}\|^2 - 1) \tag{B.2}$$

They are found where the gradient of the Lagrangian is zero:

$$\frac{\partial L(\mathbf{w})}{\partial \mathbf{w}} = E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\} + \beta\mathbf{w} = 0 \tag{B.3}$$

The function g is the derivative of G. Eq. (B.3) can be solved applying Newton's method on the Lagrangian. The second-order gradient of the Lagrangian is

$$\frac{\partial^2 L(\mathbf{w})}{\partial \mathbf{w}^2} = E\{\mathbf{z}\mathbf{z}^T g'(\mathbf{w}^T\mathbf{z})\} + \beta\,\mathbf{I} \tag{B.4}$$

The Newton method supposes an inversion of this matrix. In order to avoid this computational load in every iteration, an approximation of the second-order gradient is made. Since the data is whitened, we can write:

$$E\{\mathbf{z}\mathbf{z}^T g'(\mathbf{w}^T\mathbf{z})\} \approx E\{\mathbf{z}\mathbf{z}^T\}E\{g'(\mathbf{w}^T\mathbf{z})\} = E\{g'(\mathbf{w}^T\mathbf{z})\}\mathbf{I} \tag{B.5}$$

The iteration of the Newton method given in Eq. (B.1) is then obtained as

$$\mathbf{w} \leftarrow \mathbf{w} - \frac{E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\} + \beta\mathbf{w}}{E\{g'(\mathbf{w}^T\mathbf{z})\} + \beta}. \tag{B.6}$$

After multiplying both sides of Eq. (B.6) by $E\{g'(\mathbf{w}^T\mathbf{z})\} + \beta$ and simplifying the expression, basic iteration in FastICA is obtained:

$$\mathbf{w} \leftarrow E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\} - E\{g'(\mathbf{w}^T\mathbf{z})\}\mathbf{w}. \tag{B.7}$$

APPENDIX C

Derivation of Learning Rules in IFA

To derive the EM learning rules, $F(\mathbf{W}', \mathbf{W})$ must be minimized with respect to **W.** This can be done by first computing its gradient $\partial F / \partial \mathbf{W}$ layer by layer. For the visible layer parameters [4]:

$$\begin{aligned} \frac{\partial Fv}{\partial \mathbf{H}} &= \mathbf{\Lambda}^{-1}\mathbf{y}\left\langle \mathbf{x}^T \mid \mathbf{y} \right\rangle - \mathbf{\Lambda}^{-1}\mathbf{H}\left\langle \mathbf{x}\mathbf{x}^T \mid \mathbf{y} \right\rangle \\ \frac{\partial Fv}{\partial \mathbf{\Lambda}} &= -\frac{1}{2}\mathbf{\Lambda}^{-1} + \frac{1}{2}\mathbf{\Lambda}^{-1}\left(\mathbf{y}\mathbf{y}^T - 2\mathbf{y}\left\langle \mathbf{x}^T \mid \mathbf{y} \right\rangle \mathbf{H}^T + \mathbf{H}\left\langle \mathbf{x}\mathbf{x}^T \mid \mathbf{y} \right\rangle \mathbf{H}^T\right)\mathbf{\Lambda}^{-1} \end{aligned} \tag{C.1}$$

whereas for the bottom hidden layer

$$\begin{aligned} \frac{\partial F_B}{\partial \mu_{i,qi}} &= -\frac{1}{v_{i,qi}} p(q_i \mid \mathbf{y})\left(\left\langle x_i \mid q_i, \mathbf{y} \right\rangle - \mu_{i,qi}\right), \\ \frac{\partial F_B}{\partial v_{i,qi}} &= -\frac{1}{2v_{i,qi}^2} p(q_i \mid \mathbf{y})\left(\left\langle x_i^2 \mid q_i, \mathbf{y} \right\rangle - 2\left\langle x_i \mid q_i, \mathbf{y} \right\rangle \mu_{i,qi} + \mu_{i,qi}^2 - v_{i,qi}\right), \\ \frac{\partial F_T}{\partial \omega_{i,qi}} &= -p(q_i \mid \mathbf{y}) + w_{i,qi}. \end{aligned} \tag{C.2}$$

APPENDIX D

Derivation of the EM Algorithm

To obtain F in terms of IF model parameters $\mathbf{W}$, Eq. (4.12) is substituted in Eq. (4.23) and after a bit of algebra [4]

$$F_V = \frac{1}{2}\log|\det\mathbf{\Lambda}| + \frac{1}{2}Tr\mathbf{\Lambda}^{-1}(\mathbf{y}\mathbf{y}^T - 2\mathbf{y}\langle\mathbf{x}^T \mid \mathbf{y}\rangle\mathbf{H}^T + \mathbf{H}\langle\mathbf{x}\mathbf{x}^T \mid \mathbf{y}\rangle\mathbf{H}^T) \tag{D.1}$$

The integration over the sources $\mathbf{x}$ required to compute F_V appears in Eq. (D.1) via the conditional mean and covariance of the sources given the observed sensor signals, defined by

$$\langle m(\mathbf{x}) \mid \mathbf{y}, \mathbf{W}'\rangle = \int d\mathbf{x} p(\mathbf{x} \mid \mathbf{y}, \mathbf{W}') m(\mathbf{x}) \tag{D.2}$$

where $m(\mathbf{x}) = \mathbf{x}$ or $m(\mathbf{x}) = \mathbf{x}\mathbf{x}^T$; note these conditional averages depend on the parameters $\mathbf{W}'$.

To get F_B, $p(x_i \mid q_i) = \varsigma(x_i - \mu_{i,q_i}, v_{i,q_i})$ is substituted in Eq. (4.23):

$$F_B = \sum_{i=1}^{L}\sum_{q_i=1}^{n_i} p(q_i \mid \mathbf{y}, \mathbf{W}')\left[\frac{1}{2}log\, v_{i,q_i} + \frac{1}{2v_{i,q_i}}\left(\langle x_i^2 \mid q_i, \mathbf{y}\rangle - 2\langle x_i \mid q_i, \mathbf{y}\rangle\mu_{i,q_i} + \mu_{i,q_i}^2\right)\right] \tag{D.3}$$

where the integration over the source x_i indicated in F_B in Eq. (4.23) enters via the conditional mean and variance of this source given both the observed sensor signals and the hidden state of this source, defined by

$$\langle m(x_i) \mid q_i, \mathbf{y}, \mathbf{W}'\rangle = \int dx_i\, p(x_i \mid q_i, \mathbf{y}, \mathbf{W}') m(x_i) \tag{D.4}$$

and $m(x_i) = x_i$ or $m(x_i) = x_i^2$.

Starting from the Eq. (4.13), it is straightforward to show that, had both the sensor signals and the state from which each source is drawn been known, the sources would have a Gaussian density,

$$p(\mathbf{x} \mid \mathbf{q}, \mathbf{y}) = \varsigma(\mathbf{x} - \boldsymbol{\rho}_q(\mathbf{y}), \Sigma_q) \tag{D.5}$$

with covariance matrix and mean given by

$$\boldsymbol{\Sigma}_q = (\mathbf{H}^T \boldsymbol{\Lambda}^{-1} \mathbf{H} + \mathbf{V}_q^{-1})^{-1}, \quad \boldsymbol{\rho}_q(\mathbf{y}) = \boldsymbol{\Sigma}_q (\mathbf{H}^T \boldsymbol{\Lambda}^{-1} \mathbf{y} + \mathbf{V}_q^{-1} \boldsymbol{\mu}_q) \tag{D.6}$$

Note that the mean depends linearly on the data.

The posterior probability of the source states given the sensor data can be obtained from Eqs. (4.10-16) via

$$p(\mathbf{q} \mid \mathbf{y}) = \frac{p(\mathbf{q}) p(\mathbf{y} \mid \mathbf{q})}{\sum_{q'} p(q') p(\mathbf{y} \mid q')} \tag{D.7}$$

From Eq. (D.5) it is obtained that:

$$\langle \mathbf{x} \mid \mathbf{q}, \mathbf{y} \rangle = \boldsymbol{\rho}_q(\mathbf{y}), \quad \langle \mathbf{x}\mathbf{x}^T \mid \mathbf{q}, \mathbf{y} \rangle = \boldsymbol{\Sigma}_q + \boldsymbol{\rho}_q(\mathbf{y}) \boldsymbol{\rho}_q(\mathbf{y})^T \tag{D.8}$$

To obtain the conditional averages given in Eq. (D.2), Eq. (D.8) is summed over the states $\boldsymbol{q}$ with probabilities in Eq. (D.7) to get

$$\langle m(\mathbf{x}) \mid \mathbf{y} \rangle = \sum_q p(\mathbf{q} \mid \mathbf{y}) \langle m(\mathbf{x}) \mid \mathbf{q}, \mathbf{y} \rangle \tag{D.9}$$

taking $m(\mathbf{x}) = \mathbf{x}, \mathbf{x}\mathbf{x}^T$. Individual source averages in Eq. (D.4) together with the corresponding state posterior and their product is given by summing over all the other sources with Eqs. (D.7) and (D.8),

$$p(q_i \mid \mathbf{y})\langle m(x_i) \mid q_i, \mathbf{y}\rangle = \sum_{\{qj\}j\neq i} p(\mathbf{q} \mid \mathbf{y})\langle m(x_i) \mid \mathbf{q}, \mathbf{y}\rangle \tag{D.10}$$

Finally, the individual state posterior is obtained from Eq. (D.7):

$$p(q_i \mid \mathbf{y}) = \sum_{\{qj\}j\neq i} p(\mathbf{q} \mid \mathbf{y}) \tag{D.11}$$

APPENDIX E

α-stable Distribution

An important class of non-Gaussian phenomena encountered in practice can be characterized by its impulsive nature [31]. Signals in this class have sharp spikes. Their pdfs decay in the tails less rapidly than the Gaussian density function.

α-stable distribution is a result of generalization of the Gaussian distribution and it includes Gaussian distribution as a limiting case.

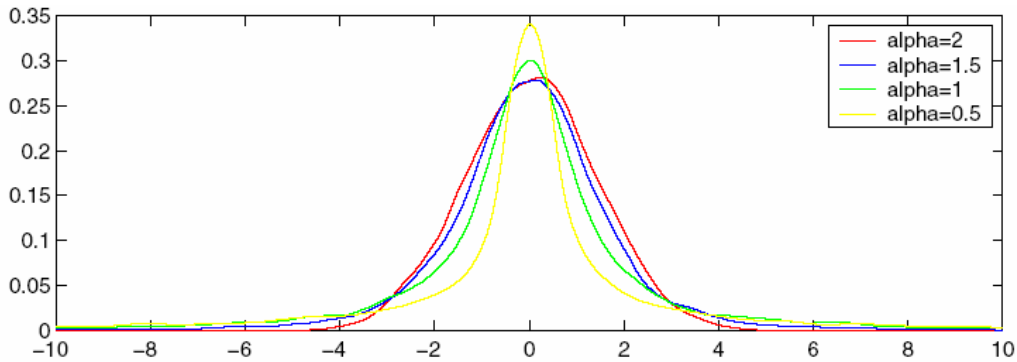

Figure E.1 Stable densities corresponding to alpha values

The main difference between non-Gaussian and Gaussian distribution is the tails of density functions. Tails of the stable density are heavier than those of the Gaussian density (Figure E.1). It has a parameter α ($0 < \alpha \leq 2$) called "characteristic exponent" that controls the heaviness of its tails. A value of α close to 2 indicates a more Gaussian type of behavior. A small value of α indicates impulsiveness of distributions (Figure E.2).

If α=2, α-stable distribution is reduced to Gaussian distribution. It is known that for α-stable distribution with characteristic exponent α, only moments of order less than α are finite so the second order moment (variance) of a stable distribution with α<2 does not exist [31]. As a result of this property, many statistical signal-processing tools give misleading results.

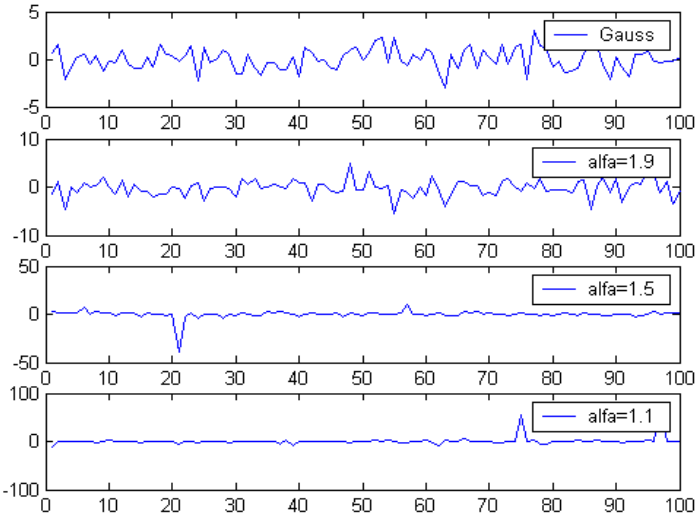

Figure E.2 Stable distributed signals for different α values.

E.1. Basic Properties of the α-stable Distribution

Stability property and Generalized Central Limit theorem are the important properties of α-stable distribution.

Theorem 1 Stability Property: A random variable X is stable if and only if for any independent random variables X_1, X_2 with same distributions as X and for arbitrary constants a_1, a_2, there exist constants a and b such that:

$$a_1 X_1 + a_2 X_2 \overset{d}{=} aX + b$$

where the notation $X \overset{d}{=} Y$ means that X and Y have the same distribution [31].

Theorem 2 Generalized Central Limit Theorem: X is the limit in distribution of normalized sums of the form [31]:

$$S_n = (X_1 + ... + X_n) / a_n - b_n$$

where X_1, X_2, ..., are i.i.d and $a_n \to \infty$, if and only if X is stable.

Proposition 1: If X is α-stable random variable and $0 < \alpha < 2$ then:

$$E|X|^P = \infty \text{ if } p \geq \alpha \text{ and}$$

$$E|X|^P < \infty \text{ if } 0 \leq p < \alpha$$

if α=2 then

$$E|X|^P < \infty \text{ for all } p \geq 0$$

so for $0 < \alpha \leq 1$, stable distributions have no first or higher order moments, for $1<\alpha<2$, all the fractional moments of order p where $p < \alpha$ exist. For α=2 all moments exist.

The stable distribution can be described by its characteristic function. A distribution function F(x) is stable iff its characteristic function has the following form [32]:

$$\varphi(t) = exp\left\{ja\,t - \gamma\,|\,t\,|^{\alpha}\,\left[1 + j\beta\,Sign(t)w(t,\alpha)\right]\right\} \tag{E.1}$$

where

$$w(t,\alpha) = \begin{Bmatrix} \tan\dfrac{\alpha\pi}{2} & \alpha \neq 1 \\ \dfrac{2}{\pi}\log|\,t\,| & \alpha = 1 \end{Bmatrix} \tag{E.2}$$

and

$$sign(t) = \begin{Bmatrix} 1 & t > 0 \\ 0 & t = 0 \\ -1 & t < 0 \end{Bmatrix}$$

In Eq. (E.1), four parameters determine the stable distributions:

1. a: The location parameter ($-\infty<a<\infty$)
2. γ: Dispersion –Scale parameter ($\gamma>0$)
3. β: The index of skewness ($-1<\beta<1$)
4. α: The characteristic exponent ($0<\alpha\leq2$)

A stable distribution is called “standard” if a=0, γ=1. When β=0, the distribution is symmetric about the center and with characteristic function “α”, stable distribution is called symmetric α-stable or SαS [31, 32].

No closed form expression exists for α-stable distribution except for the Gaussian (α=2), Cauchy (α=1, β=0) and Pearson (α=1/2, β=-1) distributions.