SOURCE SEPARATION OF SEISMOLOGICAL SIGNALS USING BLIND DECONVOLUTION AND INDEPENDENT COMPONENT ANALYSIS TECHNIQUES

by

Yalçın Kılıç

B.S. in E.E., Boğaziçi University, 2003

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Electrical and Electronic Engineering

Boğaziçi University

2006

ACKNOWLEDGEMENTS

I am very thankful to my thesis supervisors Ayşın Baytan Ertüzün and Mustafa Aktar for their help, interest and kindness during the preparation of this dissertation. I would like to mention their patience that motivated me to accomplish this task in even very hopeless situations.

I would like to also thank to the jury members for participating in the jury and their guiadance in my undergraduate and graduate studies.

Lastly, I thank my family and my friends for their support and confidence in me.

ABSTRACT

SOURCE SEPARATION OF SEISMOLOGICAL SIGNALS USING BLIND DECONVOLUTION AND INDEPENDENT COMPONENT ANALYSIS TECHNIQUES

Blind source separation and independent component analysis techniques have been implemented in many fields to analyze signals, in order to reveal the hidden factors and components that underlie. Analyzing some basic problems in seismology, it is seen that these techniques can play an important role in analyzing field data.

This work aims to apply current natural gradient signal processing techniques on seismological signals to separate independent components and separate the source and medium effects present in a seismological observation. Two algorithms are implemented for this purpose with a small modification and corrections proposed to one.

The results obtained indicate that instantaneous independent component analysis methods can be employed to preprocess seismological observations to separate the effects of remote and local seismic events. In addition, single channel blind deconvolution techniques may be used to separate the effects of propagation medium from the seismic source.

ÖZET

SİSMOLOJİK İŞARETLERİN KÖR TERS EVRİŞİM VE BAĞIMSIZ BİLEŞENLER ANALİZİ YÖNTEMLERİ KULLANILARAK AYRIŞTIRILMASI

Kör kaynak ayrışımı ve bağımsız bileşenler analizi yöntemleri, pek çok alanda sinyalleri işlemek ve bu sinyalleri oluşturan saklı etkenleri ve bileşenleri ortaya çıkarmak için kullanılmaktadır. Sismolojide karşılaşılan bazı problemler incelendiğinde, bu yöntemlerin elde edilen gözlemleri incelemede önemli rol oynayabileceği görülmektedir.

Bu çalışma, işaret işlemede kullanılan mevcut doğal düşüm yöntemlerinin, sismolojik işaretlerdeki bağımsız bileşenleri ve bu işaretlerdeki kaynak ve ortam etkilerini birbirinden ayrıştırmak için uygulanmasını amaçlamaktadır. Bu amaç doğrultusunda iki algoritma uygulanmış, bu algortimalardan birisi için küçük değişiklikler öne sürülmüş ve hataları düzeltilmiştir.

Elde edilen sonuçlar, anlık bağımsız bileşenler analizi yöntemlerinin, sismolojik gözlemlerdeki uzak ve yakın sismik olaylardan kaynaklanan etkilerin birbirinden ayrıştırılması için bir önişleme yöntemi olarak kullanılabileceğini göstermektedir. Buna ilave olarak, tek kanallı kör ters evrişim yöntemleri, kaynak ve yayılma ortamı etkilerini birbirinden ayrıştırmak için kullanılabilir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

$\mathbf{A}$	Mixing matrix
$\underline{\mathbf{A}}(z)$	System of all-pole filters
$\mathbf{B}_p$	Deconvolution matrix at time index p
$\underline{\mathbf{B}}(z,k)$	Deconvolution system
$\underline{\mathbf{C}}(z,k)$	Overall system
$\mathbf{D}$	Diagonal matrix
$\mathbf{D}_p$	Linear prediction matrix at time index p
$\underline{\mathbf{D}}(z)$	Linear prediction system
$\underline{\mathbf{D}_\mathbf{g}}(z)$	Diagonal nonsingular matrix of filters
$d\mathbf{y}(k)$	Differential operator for $\mathbf{y}$
$d\underline{\mathbf{W}}(z,k)$	Differential operator for $\underline{\mathbf{W}}(z,k)$
$d\underline{\mathbf{X}}(z,k)$	Modified coefficient differential matrix
$d\mathbf{X}_p$	Modified coefficient differential matrix at time index p
$d\phi$	Differential of the loss function
$E\{\}$	Expectation operator
E_1	Performance index
$f_i(y_i)$	Activation function for i^{th} estimate signal
$\mathbf{f}(\mathbf{y}(k))$	Vector of activation functions
$\mathbf{g}$	The output vector of the deconvolution system $\underline{\mathbf{B}}(z,k)$
$\underline{\mathbf{H}}(z)$	Convolving system
$\mathbf{H}_{ij}$	Impulse response from the j^{th} source to the i^{th} observation
$\mathbf{H}_p$	Mixing matrix at time delay p
$kurt()$	Kurtosis operator
L	Order of the filters in the deconvolution system
M	Order of the filters in the convolution system
m	Number of observations

n	Number of sources
$p_i(y_i)$	Probability density function of the i^{th} estimate signal
$\mathbf{P}$	Permutation matrix
$\mathbf{r}$	Intermediate signal used in linear prediction process
$\mathbf{s}$	Vector of source signals
$\mathbf{s}(k-p)$	Source vector at time delay p
s_i	i^{th} source signal
$tr(.)$	Matrix trace operator
$\mathbf{u}$	Vector of intermediate signals
$\mathbf{W}$	Demixing matrix
$\underline{\mathbf{W}}(z,k)$	Deconvolution system at time index k
$\underline{\mathbf{W}}^{-1}(z,k)$	Inverse of the decovoltion system $\underline{\mathbf{W}}(z,k)$
$\mathbf{W}_m$	Demixing matrix at time index m
$\mathbf{wh}$	Vector of white signals
$\mathbf{x}$	Vector of observation signals
x_i	i^{th} observation signal
$\mathbf{y}$	Vector of estimate signals
z^{-1}	Unit time delay operator
z^{-p}	Time delay operator of delay p
$\underline{0}$	System of all 0 coefficients
$(.)^H$	Matrix hermitian operator
ϕ	Loss function
μ	Stepsize parameter
AR	Autoregressive
BD	Blind deconvolution
BSI	Blind system identification
BSS	Blind source separation
FIR	Finite impulse response
HOS	Higher order statistics

ICA	Independent component analysis
ICI	Interchannel interference
ISI	Intersymbol interference
i.i.d.	Independent and identically distributed
MIMO	Multiple-input multiple-output
Mw	Moment magnitude
p.d.f.	Probability density function
SCA	Sparse component analysis
UTC	Coordinated universal time

1. INTRODUCTION

1.1. Problem Statement: Blind Source Separation

Blind source separation (BSS) and independent component analysis (ICA) techniques have been a popular area of study especially for the last two decades. The techniques have been implemented in many fields including processing mixtures of speech signals of different speakers, revealing information from biomedical observations obtained through many sensors, separating interfered radio signals and even finding hidden factors in time series of financial data [1].

ICA is the general term used for the statistical techniques developed mainly to analyze random signals in order to reveal the hidden factors and components that underlie. In ICA model, the observed multivariate data are assumed to be the mixtures of the source signals, and the mixing system is not known as well as the source signals. The main motivation for the ICA techniques is to find good estimates for the independent source signals observing the statistical properties of the mixtures. The basic assumptions of the model are the mutual independence of the source signals, nongaussianity of these independent components and the invertibility of the mixing system.

Regarding how the generative mixing system is modeled, ICA problems can be classified to be as either instantaneous or convolutive. In the instantaneous ICA case, there is no information about the timing of the signals and it is assumed that each signal reaches the observation points at the same time. In the convolutive case however, the mixing system is modeled to be made up of filters which act on the input signals and this is much more challenging than the former one.

Blind deconvolution (BD) method, which is also called channel equalization in communications terminology, is a special application of ICA developed with the motivation to separate signals which are mixed both temporally and spatially. The method is blind in the sense that the original sources and the mixing system are not known.

Existing methods used in BD can be classified as cumulant-based, Bussgang-type and information-theoretic methods such as mutual information minimization and infomax maximization [2]. As the name implies, cumulant-based methods make use of the higher order statistics (HOS) of the observation signals, and this is not very efficient as the estimation of these statistics is very sensitive to noise and outliers than that of second order statistics [3].

The Bussgang-type algorithms employ stochastic gradient descent methods to minimize a non-convex cost function evaluated using the output of the BD system [4]. These techniques have been used for a long time, and considering the robustness and the smallest mean-square error after convergence, the Godard algorithm presented in [5] is the best performing Bussgang type algorithm.

The main disadvantage of stochastic gradient optimization methods for parameterized systems is their slow convergence. To solve this problem, Amari *et al.* offered to use the well-known information-theoretic method, the natural gradient search method and it was proved in [6] that the natural gradient search method outperforms ordinary gradient search method as the parameter space has a Riemannian metric structure rather than Euclidian in many cases where these methods are to be used.

The natural gradient search method is a modified gradient search where the standard stochastic gradient search direction is altered according to the local Riemannian structure of the parameter space [7], and the resulting natural gradient search direction is guaranteed to be invariant to the statistical relationships between the parameters of the model, and providing statistically-efficient learning performance [6].

1.2. Motivation for the Thesis

As already mentioned, BSS and BD techniques have been widely studied in several fields varying from speech separation (the well-known cocktail party problem) to analyzing financial data in order to find underlying hidden information in observed time series. Hence, it seems to be a good idea to employ these techniques in any domain where

an array of m receivers picks up linear mixtures of n source signals [8], to separate the mixtures into more meaningful signals.

Seismology is the science which studies the generation and the propagation of elastic deformations occurring in the earth in response to disturbances such as earthquakes and explosions. In practice, sensors such as seismometers or geophones capture the elastic deformations and transform them into electrical signals which are then analyzed using approaches identical to those in speech, image or biomedical processing. The seismic signal itself is generally modeled as the output of a simple linear system which is, in general, a good approximation for the useful frequency (0.01-50 Hz) and disturbance (10^{-9}-10 mm) ranges. The input to the system is assumed to be the time varying disturbance created at the source (earthquake rupture or nuclear explosion, etc.), while the Earth Crust, where the wave propagates, represents the time invariant linear system itself. Most models also include the noise (often additive, but also others types) as the third component of all natural observations systems.

Establishing an analogy between the well known blind system identification (BSI), BD and ICA problems in communications, and the problems defined in geophysics, one can reasonably argue that the methods developed to analyze and solve the communications problems, can also be used in geophysics in order to get information about the characteristics of the source signal production mechanism or the wave propagation medium. Based on these reasoning, there have been recent efforts to define and solve the problem with BD and ICA methods as in [9].

Applications of BSS and BD methods are slowly finding their way in earthquake seismology. One typical example is the estimation of the properties of the wave propagation medium in the close vicinity of the observation point. This is an important problem from the practical point of view because the effects of the earthquakes on the engineered structures are closely related to the geophysical structure of the site. It is well known that the local site effects can significantly amplify seismic motions and increase the potential for damage and collapse of buildings [9]. Motivated with this phenomenon, researchers focused on explaining site characteristics with investigating recorded seismological data. In the efforts that try to estimate the seismological source signal and

the site response characteristics, the geophysical data observed at the recording stations are considered to be the result of the convolution of the source signal (seismic wavelet) with the site responses (transfer functions or channels in communications terminology) representing the characteristics of the transmission path of the waves from the source location(s) to each of the recording station [9], and this problem is a suitable candidate to be treated with BD methods.

Several other methods were also implemented to analyze the field data and the site response characteristics, and examples of these methods can be found in [10].

Based on these justifications, the main motivation for this study is to employ the BSS and BD techniques on geophysical observations, gather information about the source mechanisms and the propagation medium that act together to produce these geophysical data and separating the observed data into more meaningful signals. Two potential applications of BSS and BD techniques are investigated.

The first application is the separation of two different signals generated from two different earthquake occurrences. In seismological observations, a common problem is to have signals from a distant earthquake to mix with a local one, or vice versa. In other words, an observation focused on a local earthquake sequence can often and easily be blurred by signals coming from another earthquake which occurred in another part of the earth. Since earthquakes are very common and seismic waves travel efficiently across the earth, this problem is not uncommon. In this situation it is desired to separate the signals coming from the two independent sources. In most cases, since the signals are recorded at more than one component and also at more than one station recording simultaneously, we can assume that they are instantaneously mixed. In this application the aftershock sequence for the Urla earthquakes (17 October 2005 at 05:45 UTC, Mw 5.4; 17 October at 09:46 UTC, Mw 5.8 and 20 October at 21:40 UTC, Mw 5.8) recorded at a local station described above is artificially mixed with teleseismic signals coming from Sumatra Earthquake (26 December 2004 at 00:58 UTC, Mw 8.9). Since the observation study was intended to analyze Urla aftershocks, disturbances from the Sumatra Earthquake are suppressed.

The second application is the separation of the effect of the propagation medium from the source. For this purpose, the continuous recordings of the same aftershock activity at Urla are used. The sequences include a train of earthquake recordings generated by a multitude of small size earthquakes located roughly on the same area. Since all earthquakes are all individual events and have different source properties, they can be assumed to be independent processes. However since these earthquakes are all recorded at one common station, the propagation medium is common to all, which means that the transfer function of the site is unique and can be determined using BD method.

1.3. Scope of the Thesis

In this study, it is aimed to implement the natural gradient BSS and BD techniques, and apply these methods to the geophysical data in order to separate the effects of independent seismological events form these data, and reveal information about the source signals and the site responses which act together to produce these data. The natural gradient search method and the algorithms proposed in [7] and [11] are implemented, a small modification to the algorithm in [11] is proposed with corrections on the algorithm, and simulation results performed in MATLAB are presented to discuss their performances and analyze their characteristics.

In part 2, the basic theory for ICA, BSS and BD is presented. Here the natural gradient algorithm presented in [6] is provided as well as a modified version of it discussed in [11]. A small modification and corrections proposed to the latter one are included in this part as well. In part 3, some basic knowledge about seismic signal processing is provided. In part 4, the simulation results for the algorithms presented in the preceding parts are provided. Several simulations performed with the modified algorithm are included here as well. The simulations include the ones carried out with random signals, and seismological signals. In the last part, the advantages and the disadvantages of the presented approaches are stated and the obtained results are discussed. Several recommendations regarding the future geophysical applications of ICA methods are given.

2. INDEPENDENT COMPONENT ANALYSIS

2.1. Independent Component Analysis and Blind Source Separation

As mentioned before, ICA is the general term used for the computational methods used for analyzing multivariate random observation signals in order to find the random independent source signals of interest that underlie. Recently, ICA techniques have gained popularity because of their potential applications in signal processing. Given only the observation signals, the main goal of ICA is to recover the independent sources, which are known or assumed to be linearly mixed, and gather information about the mixing system.

In ICA problem, the unknown different and independent sources are assumed to be linearly mixed through a mixing system, where the mixing procedure is defined as

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k), \tag{2.1}$$

with $\mathbf{s}(k) = [s_1(k) \cdots\cdots s_n(k)]^T$ being the vector of sources at time index k, $\mathbf{A}$ is the m-by-n mixing matrix and is the resultant observed vector of mixtures at time index k [12]. The basic requirements for a solution of the described problem to exist can be stated as the independence of each of the sources constructing $\mathbf{s}(k)$ and the invertibility of the mixing system $\mathbf{A}$. The structure of the mixing system $\mathbf{A}$ in (2.1) identifies the type of the ICA problem. If $\mathbf{A}$ is a matrix consisting of coefficients scaling the source signals, then the problem is said to be an instantaneous ICA problem, if it is a matrix of filters instead of coefficients and the operation is convolution instead of multiplication, then the problem is a convolutive ICA problem.

Although BSS and ICA are two terms that are used in literature interchangeably, it can be said that BSS is an application of the ICA model just like BD. The aim of the BSS schemes is to separate the independent sources up to a scaling and permutation factors using the observed data. In other words, the BSS techniques try to find a matrix $\mathbf{W}$ for which the equation

$$\mathbf{WA} = \mathbf{PD}\,, \tag{2.2}$$

holds true, where $\mathbf{P}$ is a permutation matrix that accounts for the re-ordering of the estimates of the original sources and $\mathbf{D}$ is a diagonal matrix that accounts for scaling [12]. Hence if the BSS scheme is successful, the estimate signals

$$\mathbf{y}(k) = \mathbf{Wx}(k) = \mathbf{PDs}(k)\,, \tag{2.3}$$

are as same as the original source signals up to permutation and scaling factors [13].

In the theory of the instantaneous ICA problem, no information about the timing of the signals is included and it is assumed that each signal reaches the observation points at the same time. The signals are multiplied by constants and results are added on each other, the observed signals are assumed to be formed in this way. However, most of the phenomena analyzed with these techniques can not be modeled with such assumptions as delayed and convolved versions of the sources may appear at the observations in most of the time for real-life problems. At least, the propagation velocities of the signals and the location of the sensors make the problem nearly impossible to solve without containing any information about the timing of each signal.

2.2. Blind Deconvolution

In the preceding section, it was shown that the conventional instantaneous ICA techniques do not contain any information about the timing of the signals and discard the fact that delayed and convolved versions of the source signals may appear at the observation points. For example, in the well-known cocktail party problem, where the sensors (microphones) receive many speech signals from many sources (speakers), the propagation of the signals through the medium is not instantaneous; as a result, there will be differences in the time of arrival between the sources in the mixtures [3]. The multipath arrival of the signals is not accounted for in equation (2.1) as well. These important facts motivated researchers to define and solve the problem with a more realistic model, and BD model has arisen as a consequence. The general model for multiple-input multiple-output (MIMO) BD is illustrated in Figure 2.1.

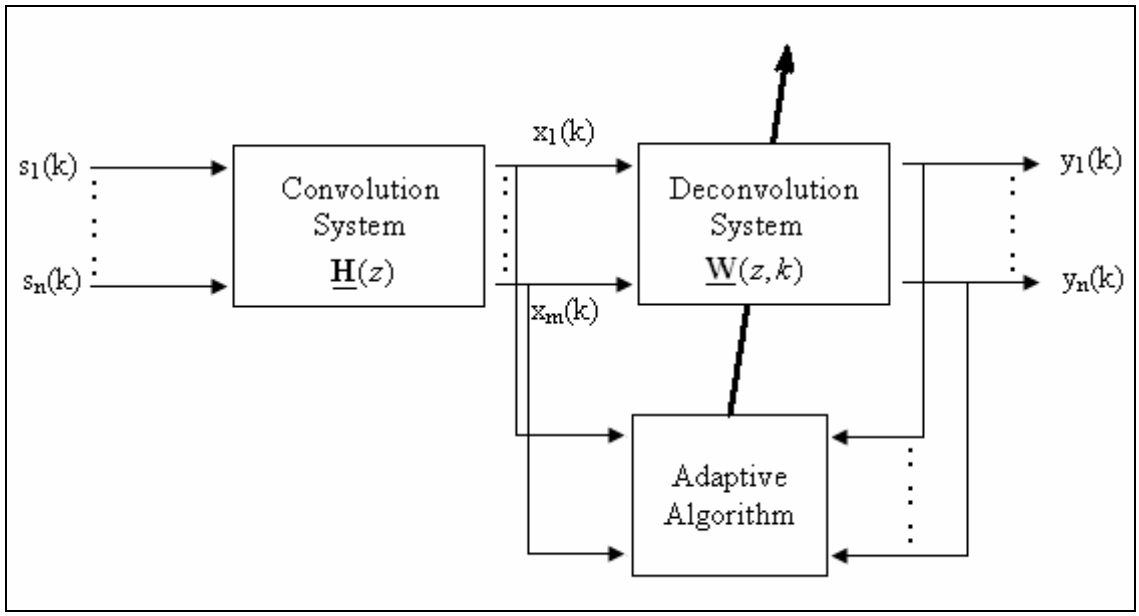

Figure 2.1. General blind deconvolution system model

When the mixing system $\mathbf{A}$ in equation (2.1) consists of filters instead of constant coefficients and the multiplication operation is replaced with the convolution operation, i.e. the observations contain delayed and convolved samples of the source signals, the problem is named the convolutive ICA problem and it can be handled with BD methods.

The instantaneous mixture case may be thought to be a special case of the described problem with the corresponding filter orders equal to 0. To define the problem and formulate it, we may introduce the unit time delay operator z^{-1}, and with $\mathbf{s}(k) = [s_1(k) \cdots\cdots s_n(k)]^T$ representing the source vector and $\mathbf{x}(k) = [x_1(k) \cdots\cdots x_m(k)]^T$ representing the observation vector as previously, we have the notation

$$z^{-1}[\mathbf{s}(k)] = \mathbf{s}(k-1), \tag{2.4}$$

for the source signals, and

$$z^{-1}[\mathbf{x}(k)] = \mathbf{x}(k-1), \tag{2.5}$$

for the observation signals.

Following the above notation and the notation of the preceding section, the convolving (mixing) system $\underline{\mathbf{H}}(z)$, which is a matrix of filters instead of constant coefficients now, can be denoted as

$$\underline{\mathbf{H}}(z) = \sum_{p=0}^{M} \mathbf{H}_p z^{-p}, \tag{2.6}$$

with each $\mathbf{H}_p$ being an m-by-n matrix, the elements of these matrices account for the filter coefficients at time delay p and the maximum filter order of the mixing system is *M*. Adapting the above notation, the observed mixtures $\mathbf{x}(k)$ can now be written as

$$\mathbf{x}(k) = \underline{\mathbf{H}}(z)[\mathbf{s}(k)] = \sum_{p=0}^{M} \mathbf{H}_p \mathbf{s}(k-p). \tag{2.7}$$

Then, the Finite Impulse Response (FIR) deconvolution system $\underline{\mathbf{W}}(z,k)$ can be denoted as

$$\underline{\mathbf{W}}(z,k) = \sum_{m=0}^{L} \mathbf{W}_m(k) z^{-m}, \tag{2.8}$$

with $\underline{\mathbf{W}}(z,k)$ being an n-by-m de-mixing matrix at time index *k,* and the time index is introduced because of the adaptive iteration procedure and time dependence of the deconvolution system.

The BD system produces estimates $\mathbf{y}(k)$ of the original source signals $\mathbf{s}(k)$ according to the system equation

$$\mathbf{y}(k) = \underline{\mathbf{W}}(z,k)[\mathbf{x}(k)] = \underline{\mathbf{W}}(z,k)\underline{\mathbf{H}}(z)[\mathbf{s}(k)]. \tag{2.9}$$

With an analogy to the equation (2.2), the deconvolution system $\underline{\mathbf{W}}(z,k)$ should be adjusted in such a manner that the equation

$$\lim_{k\to\infty} \underline{\mathbf{W}}(z,k)\underline{\mathbf{H}}(z) = \mathbf{P}\underline{\mathbf{D}_{\mathbf{g}}}(z), \tag{2.10}$$

holds true where $\mathbf{P}$ is an n-by-n permutation matrix with a single unity entry in any of its rows and columns and $\underline{\mathbf{D}_{\mathbf{g}}}(z)$ is a diagonal nonsingular scaling matrix of filters with a single non-zero coefficient in each filter [7].

2.3. The Natural Gradient

Till today, many methods have been implemented to describe and solve the BD problem, which was mainly introduced in the previous section. One of the reasonable methods proposed is to redefine the problem in the frequency domain, based on the fact that the convolution in the time domain corresponds to multiplication in the frequency domain and well known BSS techniques can easily be translated into the frequency domain to solve the problem as in [14] and [15]. However, the permutation problem in the frequency domain is much more serious and challenging than the time domain counterpart, and any approach that is proposed on this fact must address this issue as in [16]. Many techniques based on using HOS have also been addressed in literature, and some of these techniques can be founded in [17] and [18].

Besides these, most of the recent studies have been trying to solve the problem with the natural gradient approach proposed by Amari *et al.* in [19]. The algorithm proposed alters the gradient search direction according to the local Riemannian structure of the parameter space [20] and performs BD in the time domain in an iterative manner. The basic assumptions of the method are;

- There are as many observations as there are the sources (i.e. $m = n$)
- Source signals are statistically-independent
- Each source signal is independent and identically-distributed (i.i.d.)
- Source distributions are non-gaussian
- Source signals are stationary

To form the basic theory of the method, we may introduce the loss (cost) function

$$\phi(\underline{\mathbf{W}}(z,k)) = -\sum_{i=1}^{n} \log p_i(y_i(k)) - \frac{1}{2\pi j}\oint \log\left|\det \underline{\mathbf{W}}(z,k)\right| z^{-1} dz\,, \tag{2.11}$$

following the steps in [7] with $p_i(y_i)$ defined as the probability density function (p.d.f.) of the i^{th} estimate and the second term is used for ensuring that $\underline{\mathbf{W}}(z,k) = \underline{0}$ is not a minimizing point. Now, the goal of the algorithm is to minimize $E\{\phi(\underline{\mathbf{W}}(z,k))\}$ with respect to $\underline{\mathbf{W}}(z,k)$.

Introducing

$$f_i(y_i) = -\frac{d}{dy_i}\log p_i(y_i)\,, \tag{2.12}$$

and a vector valued function

$$\mathbf{f}(\mathbf{y}(k)) = \left[f_1(y_1(k)\cdots\cdots f_m(y_m(k))\right], \tag{2.13}$$

one obtains

$$d\left(-\sum_{i=1}^{n}\log p_i(y_i(k))\right) = \mathbf{f}^T(\mathbf{y}(k))d\mathbf{y}(k)\,. \tag{2.14}$$

Observing equation (2.9), it follows that

$$d\mathbf{y}(k) = d\underline{\mathbf{W}}(z,k)\underline{\mathbf{W}}^{-1}(z,k)\mathbf{y}(k)\,. \tag{2.15}$$

Defining a modified coefficient differential

$$d\underline{\mathbf{X}}(z,k) = \sum_{p=1}^{n} d\mathbf{X}_p z^{-p} = d\underline{\mathbf{W}}(z,k)\underline{\mathbf{W}}^{-1}(z,k)\,, \tag{2.16}$$

as in [20], one gets

$$d\left(\frac{1}{2\pi j}\oint \log\left|\det \underline{\mathbf{W}}(z,k)\right| z^{-1} dz\right) = tr(d\mathbf{X}_0(k)), \tag{2.17}$$

where the $tr(.)$ operator stands for the matrix trace operation. Putting these together, one obtains

$$d\phi(\mathbf{W}(z,k)) = \mathbf{f}^T(\mathbf{y}(k)) d\mathbf{X}(z,k)\left[\mathbf{y}(k)\right] - tr(d\mathbf{X}_0(k)). \tag{2.18}$$

After some mathematical manipulations and arrangement, details of which can be found in [7] and Appendix A, the natural gradient update equations

$$\mathbf{y}(k) = \sum_{p=0}^{L} \mathbf{W}_p(k)\mathbf{x}(k-p), \tag{2.19}$$

$$\mathbf{u}(k) = \sum_{q=0}^{L} \mathbf{W}_{L-q}^H(k)\mathbf{y}(k-q), \tag{2.20}$$

and

$$\mathbf{W}_p(k+1) = \mathbf{W}_p(k) + \mu(k)\left(\mathbf{W}_p(k) - \mathbf{f}(\mathbf{y}(k-L))\mathbf{u}^H(k-p)\right), \tag{2.21}$$

are obtained, where L is the order of the filters in $\underline{\mathbf{W}}(z,k)$ and $(.)^H$ is the Hermitian operation. As discussed in [19] and [21], the choice of the nonlinearities

$$f_i(y_i) = -\frac{d}{dy_i}\log p_i(y_i), \tag{2.22}$$

effects the convergence behavior of the algorithm, and these choices are performed based on the statistics of the source signals. Defining the gaussianity measure operator kurtosis

$$kurt(x) = E\{x^4\} - 3\left(E\{x^4\}\right)^2, \tag{2.23}$$

as in [1], it can be easily shown that super-gaussian signals have positive kurtosis, gaussian signals have zero kurtosis and sub-gaussian signals have negative kurtosis. Based on this justification, typical choices for the nonlinearity $f_i(y_i)$ in (2.22) are

$$f_i(y_i) = |y_i|^2 y_i, \tag{2.24}$$

for sub-gaussian signals, and

$$f_i(y_i) = \tanh(\gamma y_i), \tag{2.25}$$

with $\gamma \succ 2$ for super-gaussian signals [22]. In cases, where the source signals contain both super and sub-gaussian signals, the functions $f_i(y_i)$ should be chosen in an alternative manner [23].

2.4. Natural Gradient Algorithm Based on Autoregressive Modeled Inputs

In the development of the natural gradient BD process, it is assumed that the source signals are i.i.d. Therefore, the algorithm can be used efficiently in communication problems for blind channel equalization where the source signals may be reasonably assumed to be i.i.d without loss of generality. However, for most of the real-world problems such as speech processing, the whiteness assumption is not valid. The algorithm may also be used with this fact in mind, but it must be noted that the whitened versions of the sources will be found at the output of the deconvolving system instead of the original ones.

Based on this observation, there have been many attempts to make the natural gradient algorithm work for colored sources as well. One of these attempts proposed in [11] by Douglas *et al.* tries to define how the source signals are colored. The phenomenon is explained using Auto-Regressive (AR) process generators and the modified model make use of linear prediction together with the natural gradient BD process to obtain estimates for the original colored sources. The basic idea and the graphical illustration for the system is illustrated in Figure 2.2.

The model is developed based on the fact that, the natural gradient algorithm tries to whiten its inputs both temporally and spatially. Thus, white signals, at least signals that are whiter than the observations, are obtained at the output of the system. However, this is not desired especially in cases where it is known that the original source signals are not white. In an effort to make a more realistic model, Douglas *et al.* offered a modification on the original algorithm in [11]. The basic idea is to model how the source signals are colored and integrate this model with the original natural gradient algorithm. It is assumed that, white signals

$$\mathbf{wh}(k) = \left[wh_1(k) \cdots\cdots wh_n(k)\right]^T \tag{2.26}$$

are processed by a system $\underline{\mathbf{A}}(z)$ of all-pole filters (AR process generators). Following the notation of the previous section, the source signals $\mathbf{s}(k) = \left[s_1(k) \cdots\cdots s_n(k)\right]^T$ can be thought to be colored according to the system equation

$$\mathbf{s}(k) = \underline{\mathbf{A}}(z)\left[\mathbf{wh}(k)\right]. \tag{2.27}$$

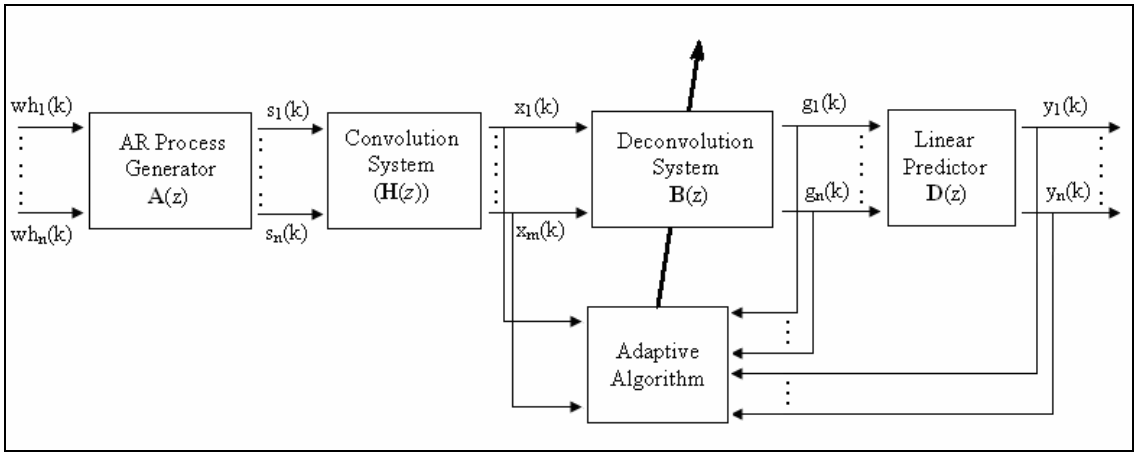

Figure 2.2. Blind deconvolution system for AR modeled inputs

The observation signals $\mathbf{x}(k) = \left[x_1(k) \cdots\cdots x_m(k)\right]^T$ are again generated according to the equation

$$\mathbf{x}(k) = \underline{\mathbf{H}}(z)[\mathbf{s}(k)] = \sum_{p=0}^{M} \mathbf{H}_p \mathbf{s}(k-p). \tag{2.28}$$

Dividing the overall deconvolution system $\underline{\mathbf{W}}(z,k)$ into two subsystems $\underline{\mathbf{B}}(z,k)$ and $\underline{\mathbf{D}}(z)$, we obtain the relation

$$\mathbf{W}_p(k) = \mathbf{D}_p * \mathbf{B}_p(k) \tag{2.29}$$

or equivalently

$$\underline{\mathbf{W}}(z,k) = \underline{\mathbf{D}}(z)\underline{\mathbf{B}}(z,k). \tag{2.30}$$

In equation (2.30), the system $\underline{\mathbf{B}}(z,k)$ accounts for the deconvolution process and the system $\underline{\mathbf{D}}(z)$ accounts for the inversion of the AR process, i.e. linear prediction process.

With these relations in hand, we may define

$$\mathbf{y}(k) = \sum_{p=0}^{L} \mathbf{W}_p(k)\mathbf{x}(k-p) \tag{2.31}$$

and

$$\mathbf{u}(k) = \sum_{q=0}^{L} \mathbf{W}_{L-q}^{H}(k)\mathbf{y}(k-q) \tag{2.32}$$

with $\mathbf{W}_p$ previously defined.

Re-writing the equation in (2.21) as

$$\mathbf{D}_p * \mathbf{B}_p(k+1) = \mathbf{D}_p * \mathbf{B}_p(k) + \mu(k)\left(\mathbf{D}_p * \mathbf{B}_p(k) - \mathbf{f}(y(k-L))\mathbf{u}^T(k-p)\right), \tag{2.33}$$

noting that $\underline{\mathbf{A}}(z)$ is the inverse of the system $\underline{\mathbf{D}}(z)$ and convolving both sides with $\underline{\mathbf{A}}(z)$, one obtains

$$\mathbf{B}_p(k+1) = \mathbf{B}_p(k) + \mu(k)\left(\mathbf{B}_p(k) - \underline{\mathbf{A}} * \mathbf{f}(y(k-L))\mathbf{u}^T(k-p)\right). \tag{2.34}$$

With these relations and the model, we may re-write the update equations of the algorithm as

$$\mathbf{g}(k) = \sum_p \mathbf{B}_p(k)\mathbf{x}(k-p) \tag{2.35}$$

$$\mathbf{y}(k) = \sum_p \mathbf{D}_p(k)\mathbf{g}(k-p) \tag{2.36}$$

$$\mathbf{u}(k) = \sum_p \mathbf{W}_{L-p}(k)\mathbf{y}(k-p) \tag{2.37}$$

$$\mathbf{r}(k) = \underline{\mathbf{A}} * \mathbf{f}(y(k-L)) \tag{2.38}$$

$$\mathbf{B}_p(k+1) = \mathbf{B}_p(k) + \mu(k)\left(\mathbf{B}_p(k) - \mathbf{r}((k))\mathbf{u}^T(k-p)\right). \tag{2.39}$$

Here, a modification to the algorithm is done introducing the equation (2.37) instead of the original two equation version in [11] to decrease the number of operations to be done at each iteration and simplify the overall algorithm. In addition, the error of time indexes in (2.38) and (2.39) were corrected in the implementation of the algorithm. One can easily verify that the equation (2.35) through (2.39) are identical to the ones in (2.19) through (2.21) if the inputs are i.i.d signals, i.e. the order of the AR process generators are 0, and the new set of equations form a more general basis for the algorithm.

3. INDEPENDENT COMPONENT ANALYSIS OF SEISMIC SIGNALS

3.1. Blind Source Separation of Seismic Signals

Everyday, there are several earthquakes on all over the world and these earthquakes are strong enough to be felt locally [24]. Although the effects felt by people increase going closer to the center of the seismic event such as an earthquake or an explosion occurred in a mine, each seismological source produce waves propagating through the solid structure of the earth and causes some changes that can be recorded at even very far away locations. The majority of these changes are too week to be felt, however they are easily recorded at many recording stations set all over the globe as seen in Figure 3.1. The recordings obtained in a station can illustrate the effects of a strong seismic event that is even thousands of kilometers away from the station as the illustration implies.

While investigating the effects and the mechanism of local earthquakes, recordings obtained through local recording stations are studied. However, as previously mentioned, these recordings contain the effects of remote seismic events, and for a reasonable investigation, it would be better to remove these effects with a preprocessing method.

Generally, the effects of earthquakes are recorded by many recording stations, and as these stations are close to each other, the effects of seismic events at these stations are more or less similar to each other. Assuming that the recordings are the mixtures of independent distant and local seismic events, and with the traditional ICA model

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k), \tag{3.1}$$

in hand, the ICA methods can be used to separate the effects of the distant and local seismic events.

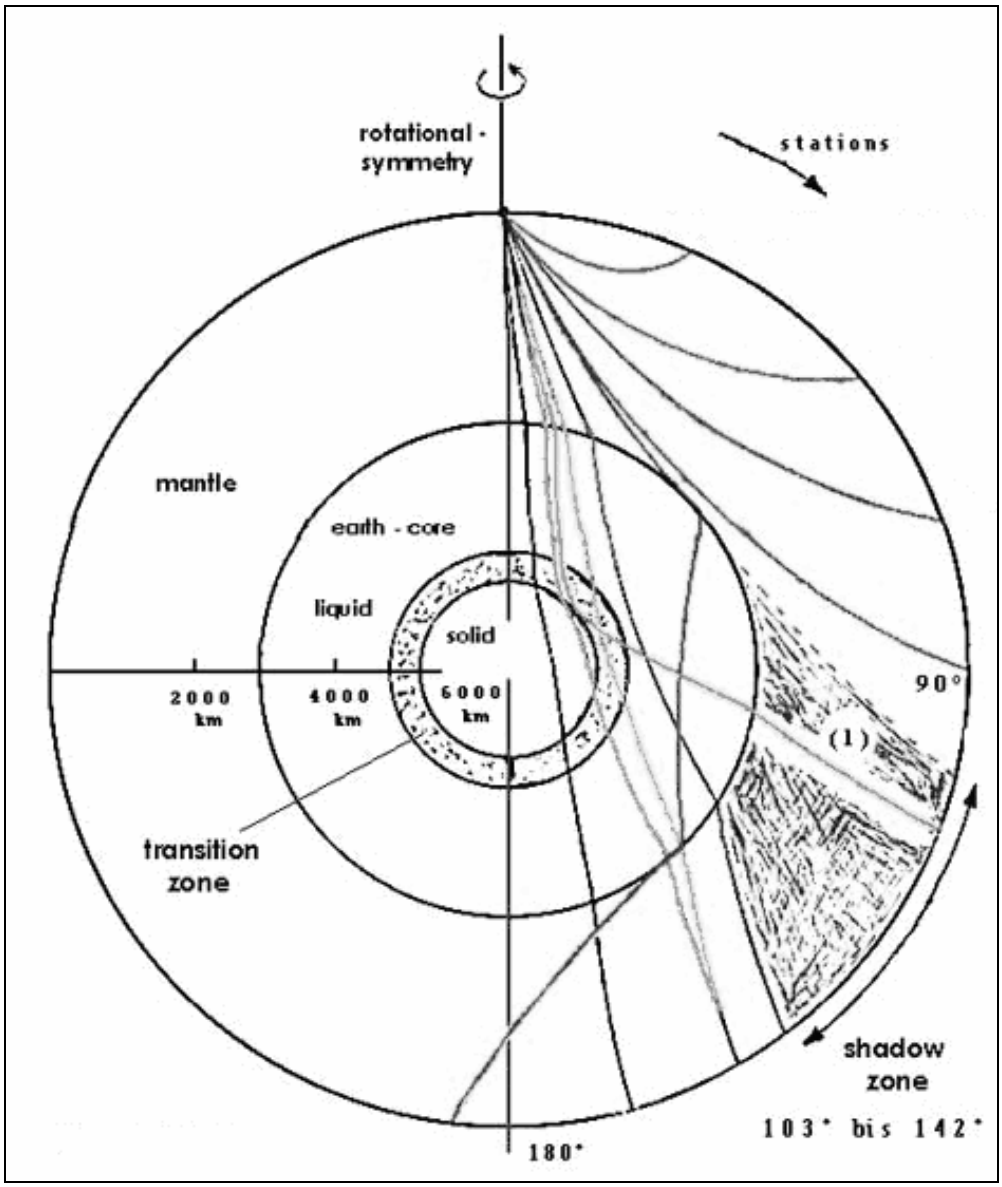

Figure 3.1. Propagation of seismic waves [25]

In (3.1), $\mathbf{s}(k) = [s_1(k) \cdots\cdots s_n(k)]^T$ is the vector of source signals at time index k, $\mathbf{A}$ is the mixing matrix and $\mathbf{x}(k) = [x_1(k) \cdots\cdots x_m(k)]^T$ is the vector of resultant observed vectors obtained at the recording stations. Adapting this notation and model, it can be reasonably tried to preprocess the recording to remove the effects of the undesired seismic signals.

3.2. Blind Deconvolution of Seismic Signals

The effects of the earthquakes on the engineered structures are closely related to the geophysical structure of the site. It is well known that the local site effects can significantly amplify seismic motions and increase the potential for damage and collapse of buildings [9]. This phenomenon motivated researchers to investigate the internal structure of the earth to reveal information about how a site and the engineered structures constructed atop will react to a seismic event.

In recent studies about the subject, it is often assumed that observed seismic data series are the result of a source time function convolving with the impulse response from the source to the observation point, which may also be called the Green's function[10]. However, neither the source time function nor the Green's function is known for analysis and finding a suitable estimate for both time series is a challenging task for the seismologists.

In literature, recovering the impulse response of the media and the source time function from a given seismic trace is called the seismic deconvolution. Many studies were done to address the issue and solve the problem like ones in [26], [10], [27] and [9].

To put the seismic deconvolution problem in mathematical terms, it can be said that the seismic recordings $x_i(t)$ consist of the convolution of a source time function (or seismic wavelet) $s(t)$ with the impulse response of the path from the source signal to the observation station (the Green's function) $h_i(t)$ and additional noise $n_i(t)$ as illustrated in Figure 3.2 [28]. Thus, we have

$$x_i(t) = s(t) * h_i(t) + n_i(t) . \tag{3.2}$$

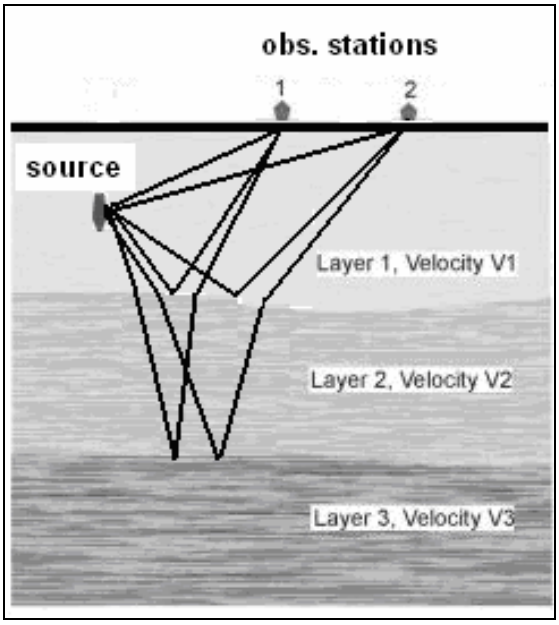

Figure 3.2. Production mechanism of a seismic trace

Observing equations (2.28) and (3.2), it is seen the latter one is a single channel realization of the former one with additional noise. Therefore, the seismological problem can be handled with BD methods previously developed for other purposes. Based on this justification, many studies were done till today. In [29], a previously developed blind deconvolution algorithm of [30] was employed to solve the problem. The method is mainly based on reconstruction of signals from information on their phase only and needs two or more recordings for reconstruction. Velis *et al.* made use of the fourth order cumulant function on seismic deconvolution in [31]. In another attempt, Pflug formed a relationship between the fourth order cumulant function on the application of fourth order deconvolution, the signal passband and the trispectral domain [32], while Santamaria *et al*. developed an iterative algorithm with the gaussian mixture model to identify the reflectivity sequence on seismic deconvolution in [27].

In this study, the natural gradient algorithm is applied to the seismological signals in order to estimate the source time function and the impulse response for an observed data series.

4. SIMULATIONS

4.1. Simulations for the Natural Gradient Algorithm

In this section, results for two single-channel and multi-channel blind deconvolution simulations are presented to illustrate the performance for the natural gradient approach and its drawbacks. For the simulations, the update equations in (2.19), (2.20) and (2.21) were used, and the algorithm was implemented in MATLAB. The initial conditions were chosen for $\underline{\mathbf{W}}(z,k)$ in a manner so that the filter(s) on the diagonal have a unit magnitude impulse at the center tap, and the non-diagonal filters have all zero taps as in [7] to ensure that the estimates are delayed versions of the original sources in case of no mixing. The performance measures used are the intersymbol interference (ISI) and the interchannel interference (ICI) which may be defined as

$$ISI_i = 1 - \frac{(\max(\mathbf{C}_{ii}))^2}{\sum_j C_{iij}^2}, \tag{4.1}$$

and

$$ICI_{ik} = \frac{\left(\sum_{j,i\neq k} C_{ikj}^2\right)}{\max(\underline{\mathbf{C}}(z,k))^2}, \tag{4.2}$$

for real coefficient filters, where $\underline{\mathbf{C}}(z,k)$ is the overall matrix of filters composed by the convolution of the mixing and the de-mixing system, i.e.

$$\underline{\mathbf{C}}(z,k) = \underline{\mathbf{W}}(z,k)\underline{\mathbf{H}}(z). \tag{4.3}$$

In (4.3) $\mathbf{C}_{ii}$ is the filter on the i^{th} diagonal of the overall system function $\underline{\mathbf{C}}(z,k)$, ISI_i is the ISI on the i^{th} channel and ICI_{ik} is the ICI from the k^{th} signal to the i^{th} estimation.

In the first simulation scenario, a single channel deconvolution task is performed using the natural gradient algorithm defined. The algorithm is used to recover a sub-gaussian signal sequence $s_1(k)$ of 100000 samples length, where $s_1(k) \in \{-2,-1,0,1,2\}$ $for\ \forall k$ with equal probability and the used mixing system is randomly chosen to be

- $\mathbf{H}_{11}(z) = 1.0285 - 0.3854z^{-1} - 0.5364z^{-2} + 0.6451z^{-3} + 0.2262z^{-4}$

The order of the inverse filter in $\underline{\mathbf{W}}(z,k)$ is set to be *L=47* and the function $f_1(y_1)$ is chosen to be $f_1(y_1) = |y_1|^2 y_1$ since it is known that the distribution of $s_1(k)$ is sub-gaussian. The results (the impulse response of overall system and the ISI) obtained in this scenario are illustrated in Figure 4.1 and Figure 4.2.

Observing the results obtained as a result of the simulation, one sees that the overall system function $\underline{\mathbf{C}}(z,k)$ has a dominant impulse at $k = 24$ and impulses with little or nearly no power for $k \neq 24$ and the ISI_1 is as low as -25 dB at iteration number 25000. Results show that the ISI_1 is lowered from -5 dB to about -27 dB and therefore the obtained estimate is a delayed and scaled version of the original source signal, as it should be for a successful BD procedure and simulations with super-gaussian signals produced similar results.

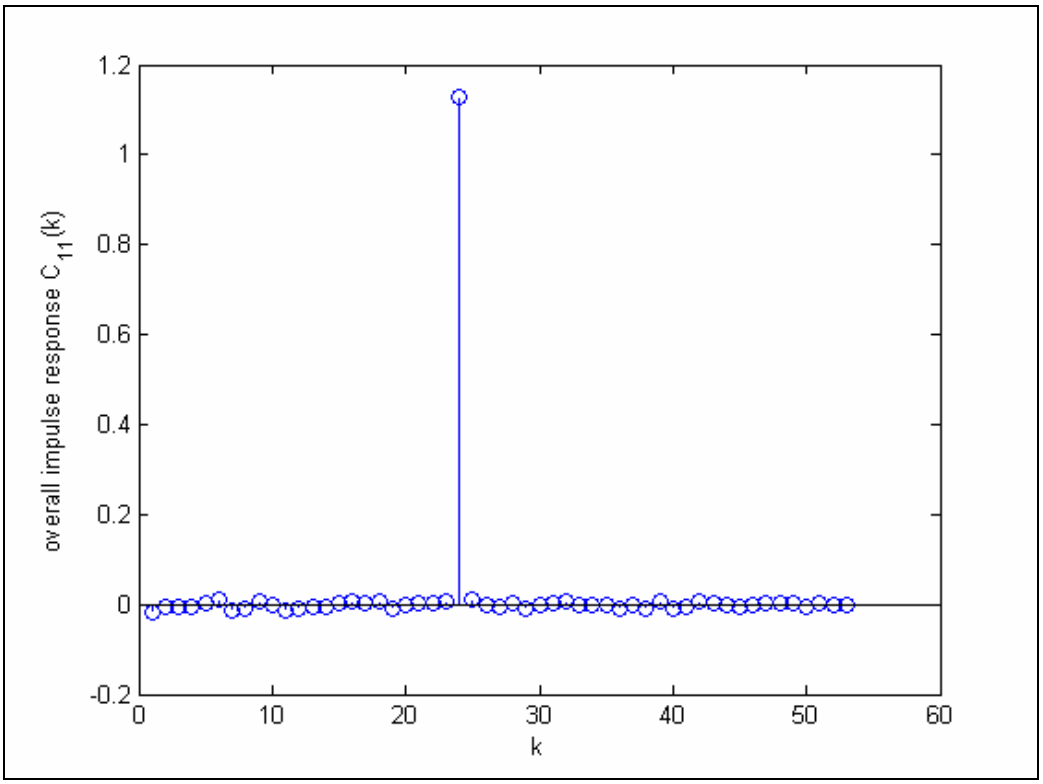

Figure 4.1. Overall impulse response for the first simulation scenario

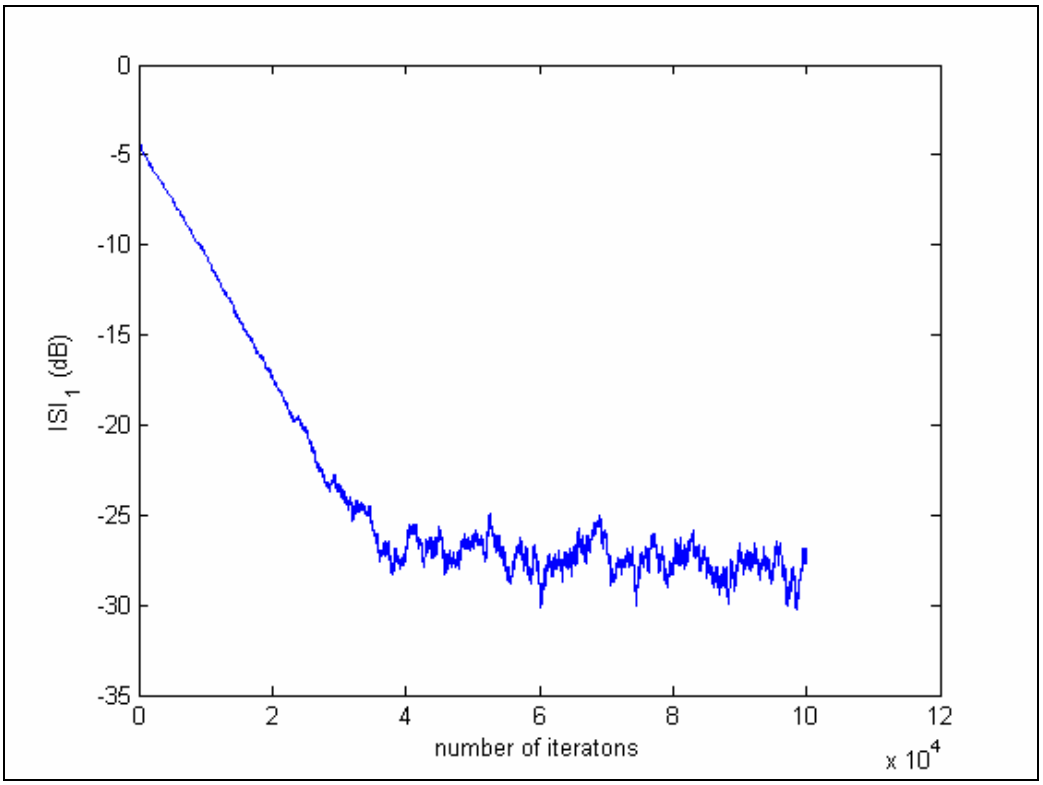

Figure 4.2. ISI_1 for the first simulation scenario

In the second scenario, the natural gradient algorithm is employed to perform a two-channel multichannel BD task. This time, the source signals are chosen to be super-gaussian ones, namely $s_1(k)$ and $s_2(k)$. Both $s_1(k)$ and $s_2(k)$ are the realizations of a random process where each element of $s_1(k)$ and $s_2(k)$ is from a gamma distribution produced using the MATLAB's gamrnd(.) function using 1 for both the parameters in A and B. The convolving system is randomly generated as

- $\mathbf{H}_{11}(z) = 1.0285 - 0.3854z^{-1} - 0.4364z^{-2} + 0.5451z^{-3}$
- $\mathbf{H}_{12}(z) = 0.5 + 0.2008z^{-1} - 0.1705z^{-2}$
- $\mathbf{H}_{21}(z) = 0.5 + 0.1748z^{-1} - 0.1890z^{-2}$
- $\mathbf{H}_{22}(z) = 0.9541 + 0.4853z^{-1} + 0.2362z^{-2} - 0.3452z^{-3}$

The order of the inverse filters in $\underline{\mathbf{W}}(z,k)$ are set to be *L=63* and the functions $f_1(y_1)$ and $f_2(y_2)$ was chosen to be $f_1(y_1) = \tanh(3y_1)$ and $f_2(y_2) = \tanh(3y_2)$. The results obtained in this scenario are illustrated in Figure 4.3 and Figure 4.4.

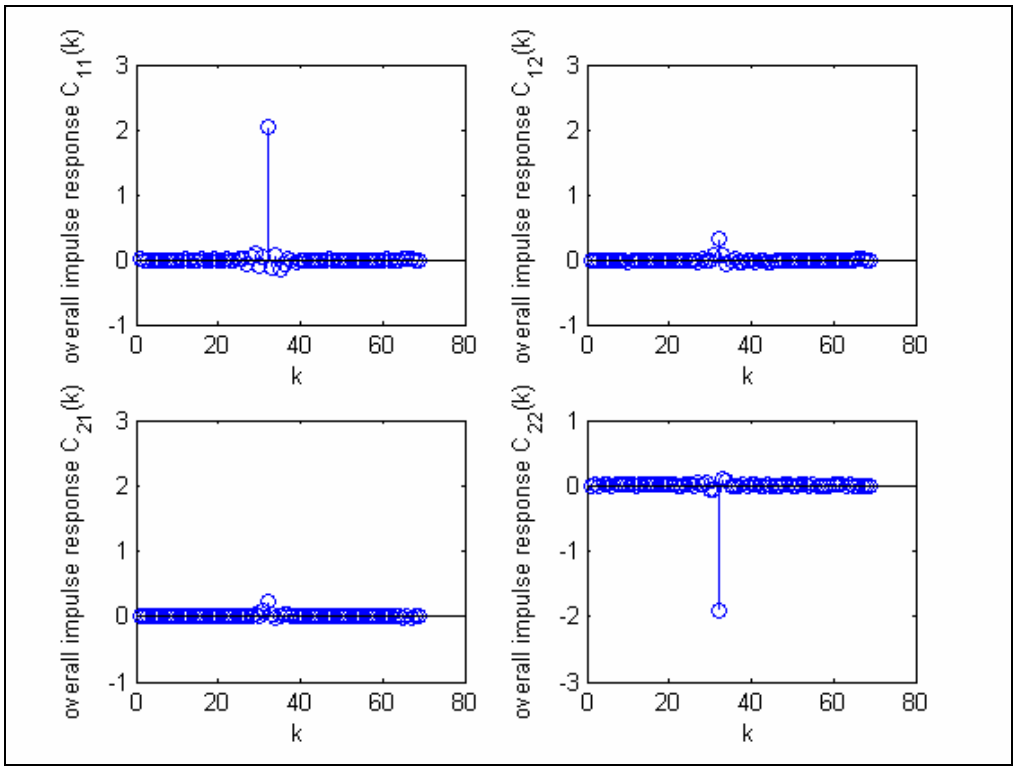

Figure 4.3. Overall impulse responses for the second simulation scenario

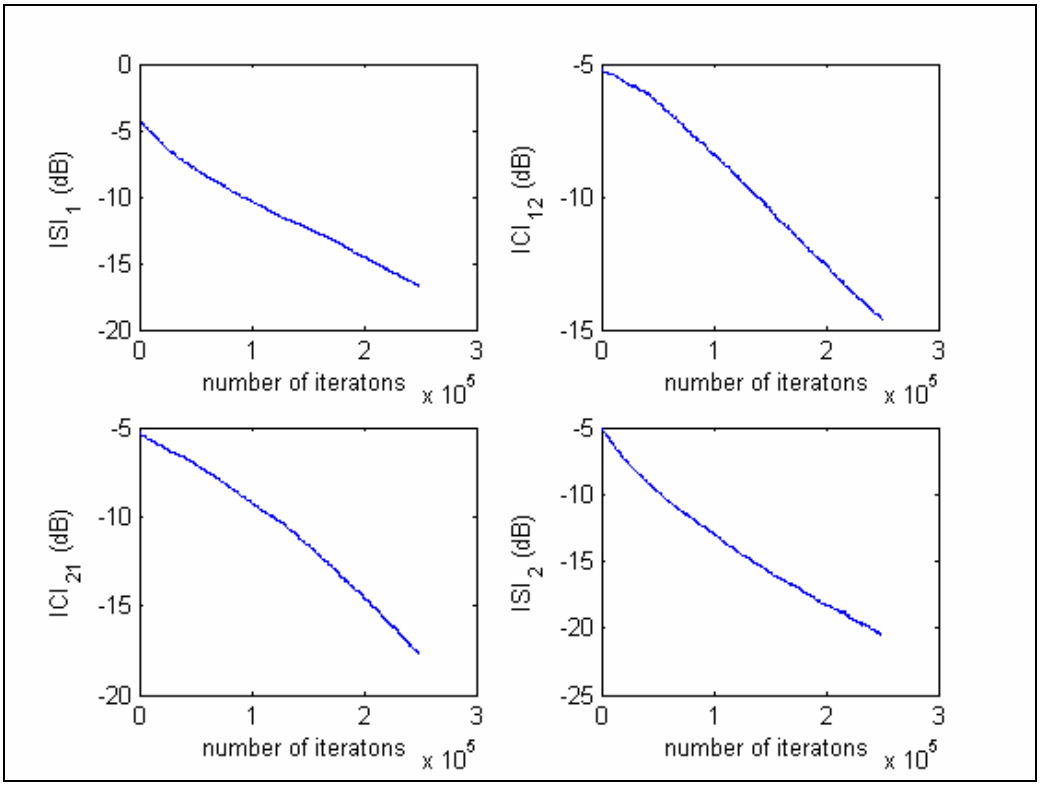

Figure 4.4. ISI_1, ICI_{12}, ICI_{21} and ISI_2 for the second simulation scenario

Examining Figure 4.3 and Figure 4.4, it is again seen that the overall system function matrix $\underline{\mathbf{C}}(z,k)$ consist of dominant delayed impulses on the diagonal, and the impulse responses on the off-diagonals nearly adds nothing to the output of the deconvolution system, which implies that the mixed signals are separated and the ISI and ICI factors are nearly completely eliminated at iteration number 250000. Similar results were obtained in the simulations performed with sub-gaussian signals.

Besides these desired accuracy properties of the algorithm, it was observed that the algorithm converges slowly which implies that large data sets are required for efficient deconvolution of the observation signals. In addition, it was observed the algorithm fails in case of forward systems which have deep notches in their passbands.

4.2. Simulations for the Modified Natural Gradient Algorithm

In this section, results for simulations of two multichannel BD cases with the modified natural gradient algorithm are presented. In the simulations, equations (2.35)

through (2.39) are used and the algorithm is implemented in MATLAB. As in the previous section, the initial conditions were chosen for $\underline{\mathbf{B}}(z,k)$ and $\underline{\mathbf{D}}(z)$ in a manner so that the filter(s) on the diagonal of $\underline{\mathbf{B}}(z,k)$ have a unit magnitude impulse at the center tap, and only the center matrix of $\underline{\mathbf{D}}(z)$ equals to unity matrix, with a similar reasoning. The ISI and the ICI are evaluated using the equations (4.1) and (4.2) with $\underline{\mathbf{C}}(z,k)$ defined as

$$\underline{\mathbf{C}}(z,k) = \underline{\mathbf{W}}(z,k)\underline{\mathbf{H}}(z) \tag{4.4}$$

In the third scenario, the modified natural gradient algorithm is used to recover the mixtures of two AR process $s_1(k)$ and $s_2(k)$ of 150000 samples length. The AR processes are generated according to the equations

- $s_1(k) = -0.6s_1(k-1) + wh_1(k)$
- $s_2(k) = -0.5s_2(k-1) - 0.1s_2(k-2) + wh_2(k)$

The random processes $wh_1(k)$ and $wh_2(k)$ are generated according to $wh_m(k) \in \{-2,-1,0,1,2\}$ *for* $\forall k$ *and* $m \in \{1,2\}$ and the mixing system parameters are set as

- $\mathbf{H}_{11}(z) = -0.6436 + 0.3803z^{-1} - 1.0091z^{-2} - 0.3195z^{-3} + 0.6125z^{-4}$
- $\mathbf{H}_{12}(z) = -0.6 + 0.5z^{-1} + 0.2z^{-2}$
- $\mathbf{H}_{21}(z) = 0.5 - 0.3z^{-1} - 0.2z^{-2}$
- $\mathbf{H}_{22}(z) = 0.6686 + 1.1908z^{-1} - 1.2025z^{-2} - 0.0198z^{-3}$

The nonlinearities $f_1(y_1)$ and $f_1(y_1)$ are chosen as $f_1(y_1) = |y_1|^2 y_1$ and $f_2(y_2) = |y_2|^2 y_2$ and the order of the inverse filters are set to be 64. The results for the simulation are illustrated in Figure 4.5 and Figure 4.6.

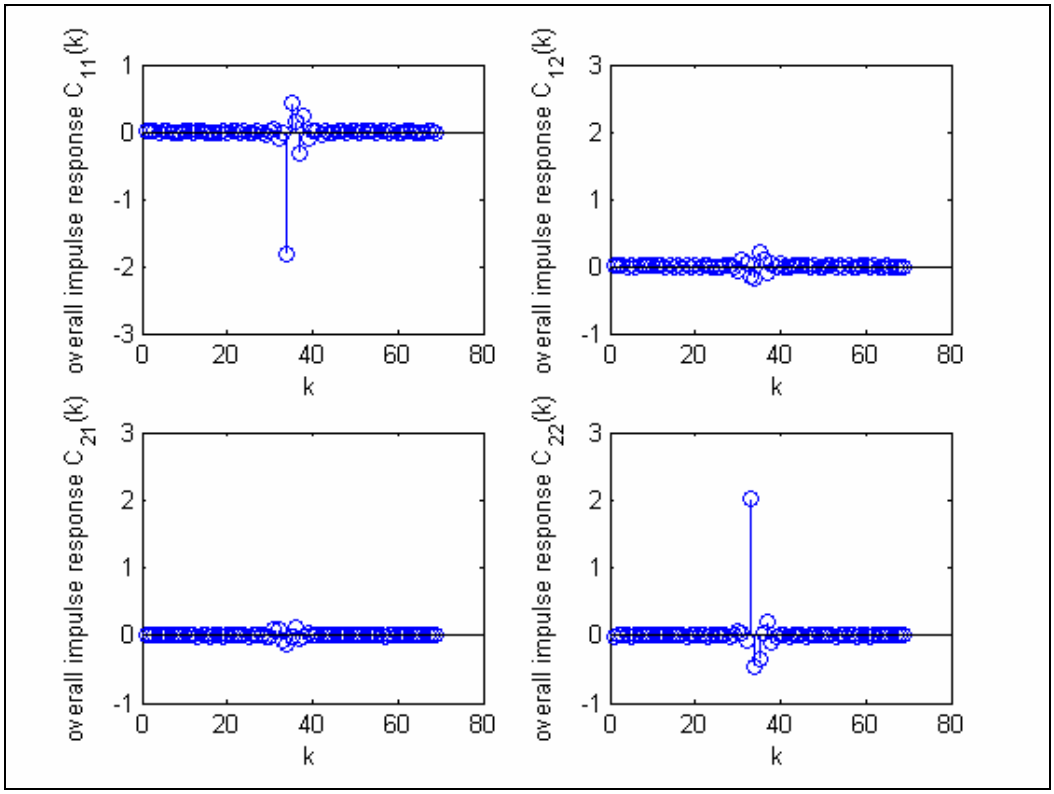

Figure 4.5. Overall impulse responses for the third simulation scenario

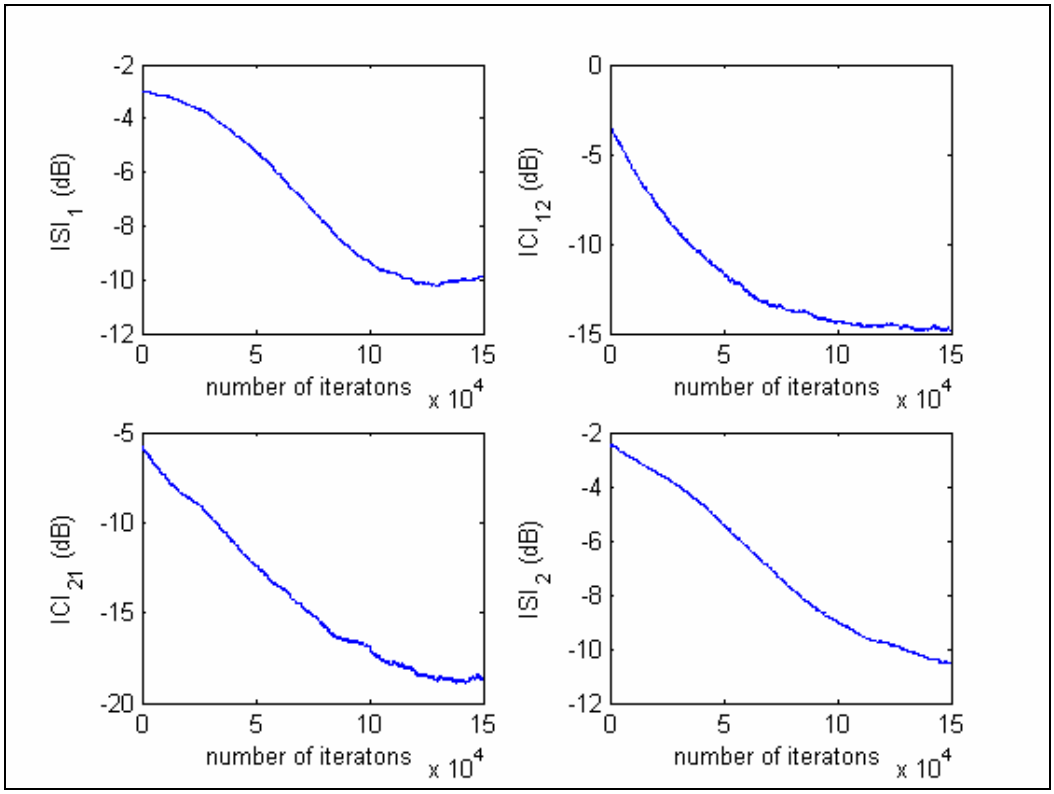

Figure 4.6. ISI_1, ICI_{12}, ICI_{21} and ISI_2 for the third scenario

In the fourth simulation scenario, the algorithm is used for BD of the mixtures of two AR process $s_1(k)$ and $s_2(k)$ of 150000 samples length. The signals wh_1 and wh_2 that are used to generate the AR processes are both the realizations of a random process with its each element being from a gamma distribution an generated with MATLAB's gamrnd(.) function with parameters A and B set equal to 1. The AR process coefficients, the mixing system and the order of the inverse system coefficients are the same as the ones used in the previous scenario. The used nonlinearities are $f_1(y_1) = \tanh(3y_1)$ and $f_2(y_2) = \tanh(3y_2)$. The simulation results are illustrated in Figure 4.7 and Figure 4.8.

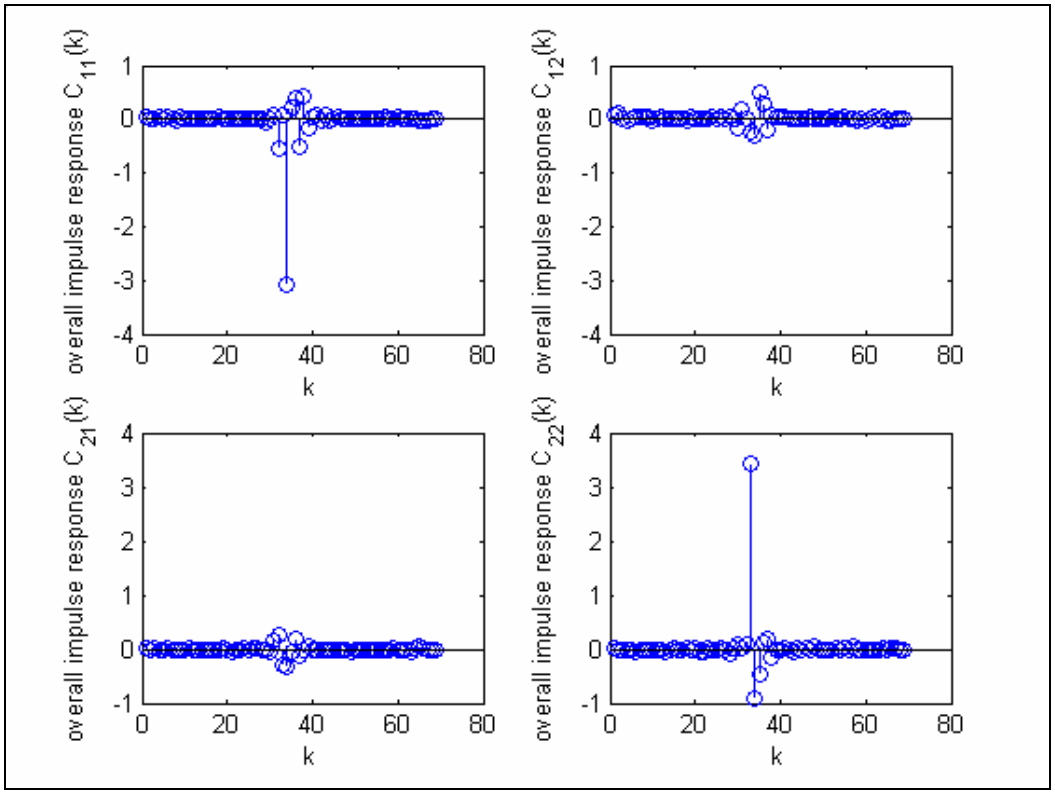

Figure 4.7. Overall impulse responses for the fourth scenario

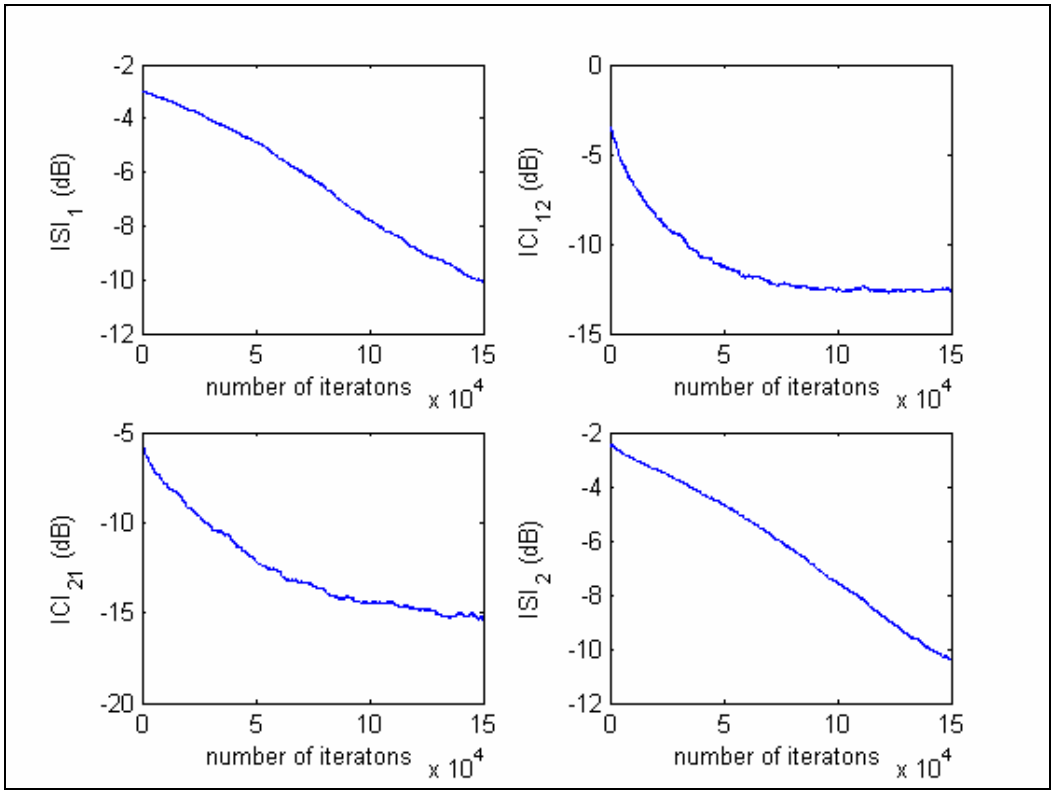

Figure 4.8. ISI_1, ICI_{12}, ICI_{21} and ISI_2 for the fourth scenario

Based on the theory of the algorithm and the simulation results obtained in this section, it is seen that the algorithm has acceptable convergence properties for both random sub-gaussian and super-gaussian signals which can be modeled to be AR processes.

Comparing with the original natural gradient algorithm, it is observed that the modified natural gradient algorithm has worse convergence properties than the former one which assumes that the source signals are white. This may be due to the fact that the linear prediction process is stopped at iteration number=20000 in the simulations for performance reasons. The convergence behavior may be improved with an alternative linear prediction approach.

In addition, the modified natural gradient algorithm forms a more general basis for BD in the sense that it may also be used for white signals with the order of the linear prediction process equal to 0.

Besides these, in some simulations which are not included here, it is observed that the algorithm may fail to separate mixtures that are formed with systems which have deep notches in their passbands as in the original natural gradient algorithm. This implies that it must be guaranteed that the forward system has smooth passbands for successful BD of the mixtures.

4.3. Simulations with Seismological Signals

In this section, the simulation results for the seismological signals are presented. For the simulations, the basic theory presented in the previous part is used together with the ICA model.

4.3.1. Blind Source Separation of Observations of Independent Earthquake Signals

In this part, the results of two simulations performed with artificially mixed seismological signals of different characteristics are presented. The signals are instantaneously mixed and instantaneous ICA was applied to the mixture signals.

In the first seismological simulation scenario, the recordings from a local earthquake ($s_1(k)$) and a remote earthquake ($s_2(k)$) are instantaneously mixed with the random mixing matrix

- $\mathbf{A} = \begin{bmatrix} 0.7003 & 0.8137 \\ 0.5377 & 0.5280 \end{bmatrix}$

using the system equation (2.1) to obtain the mixture signals $x_1(k)$ and $x_2(k)$. The source signals, the mixture signals and the estimated source signals are illustrated in Figure 4.9, Figure 4.10 and Figure 4.11 respectively.

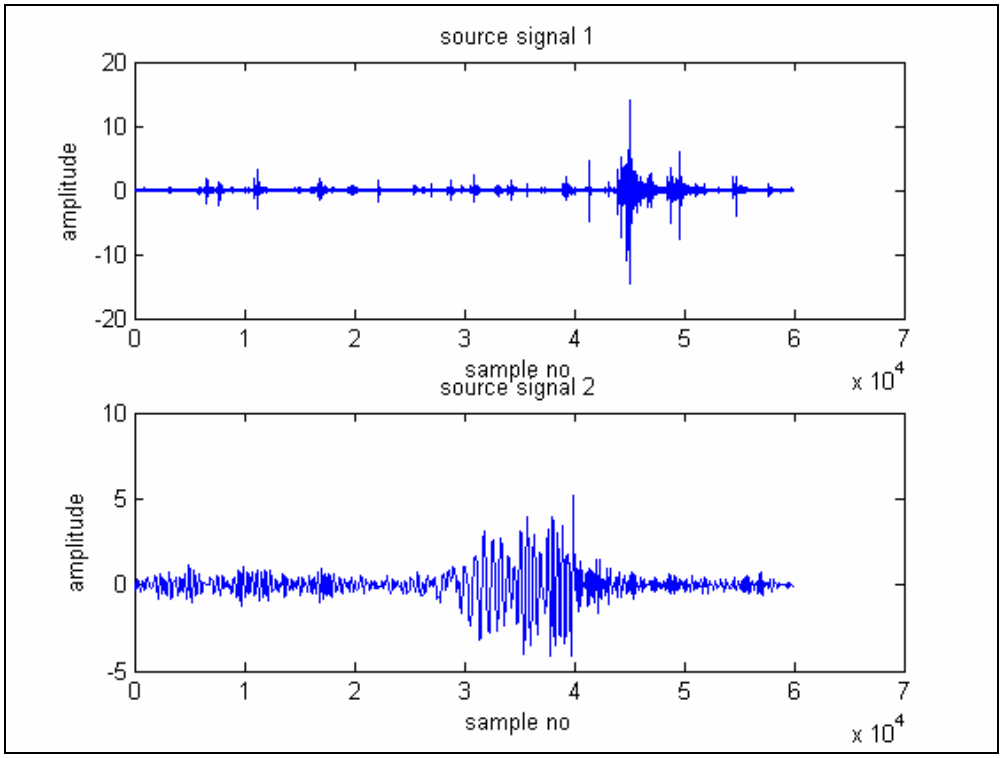

Figure 4.9. Source signals for the first seismological simulation scenario (last 60000 samples)

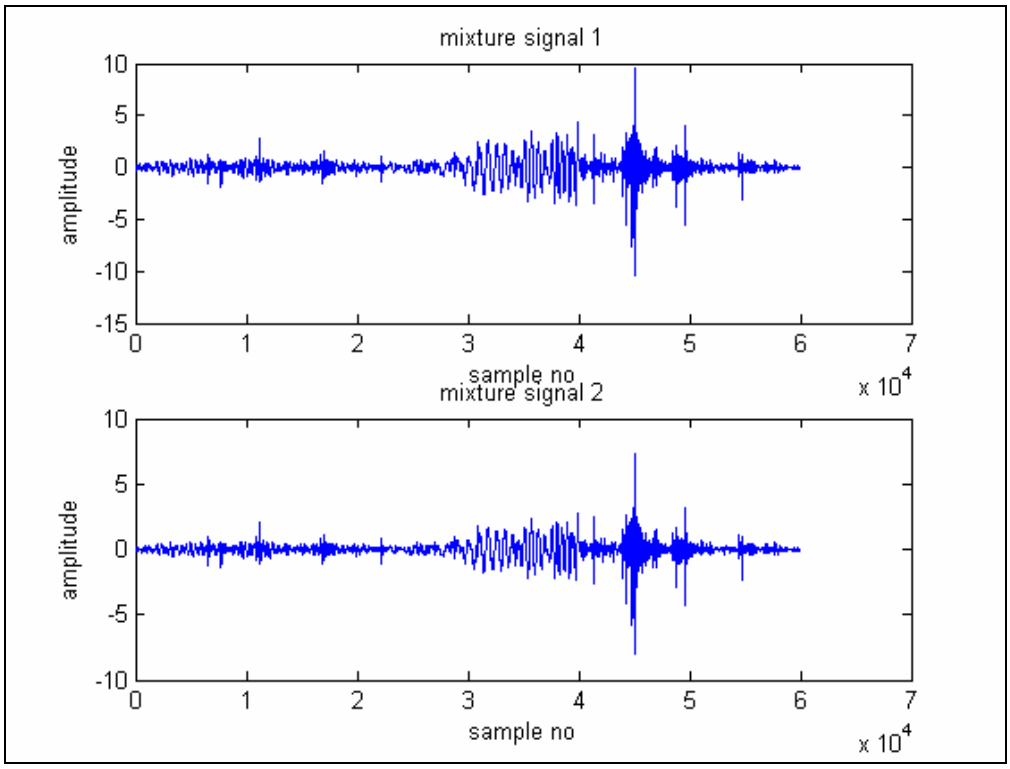

Figure 4.10. Mixture signals for the first seismological simulation scenario (last 60000 samples)

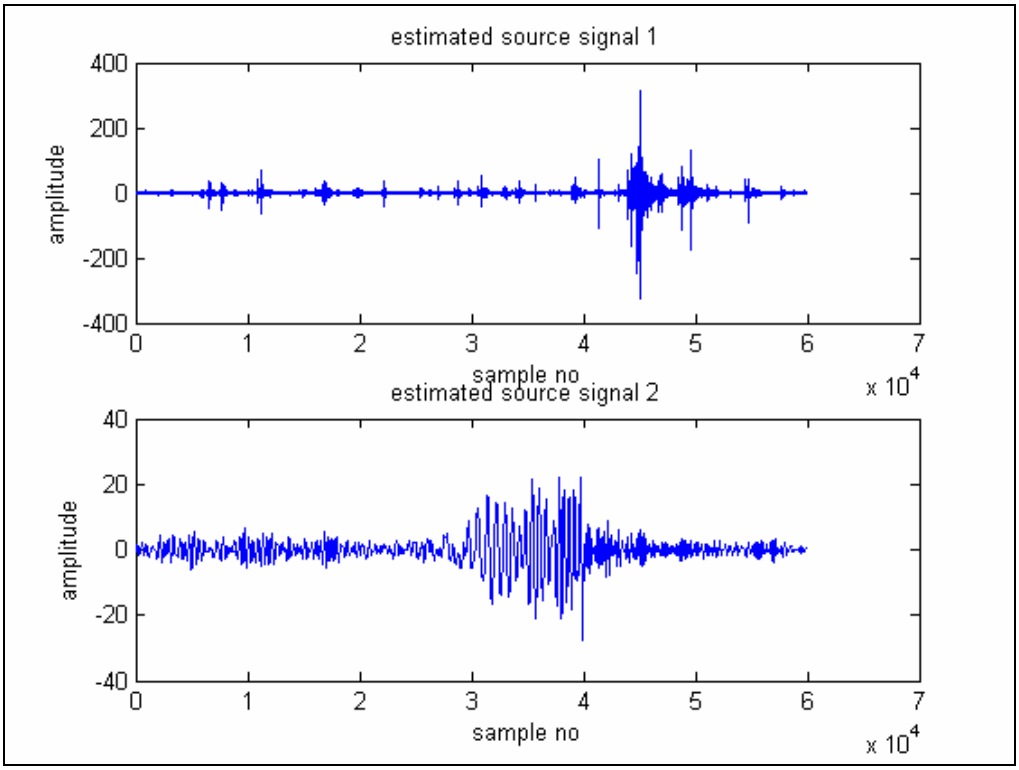

Figure 4.11. Estimated source signals for the first seismological simulation scenario (last 60000 samples)

The inverse system matrix

- $\mathbf{W} = \begin{bmatrix} -179.5 & 275.4 \\ -39.6 & 51.0 \end{bmatrix}$

and the overall system matrix

- $\mathbf{WA} = \begin{bmatrix} 22.4 & -0.6 \\ -0.3 & -5.3 \end{bmatrix}$

is obtained for the simulation.

To measure the performance of the algorithm and the efficiency of the separation, we may introduce the performance index

$$E_1 = \sum_{i=1}^{m}\left(\sum_{j=1}^{m}\frac{|p_{ij}|}{\max_k |p_{ik}|}-1\right)+\sum_{j=1}^{m}\left(\sum_{i=1}^{m}\frac{|p_{ij}|}{\max_k |p_{kj}|}-1\right) \tag{4.5}$$

as in [1] where p_{ij} is the ij^{th} element of the $m\times m$ matrix $\mathbf{WA}$. It is clear that the performance index E_1 attains its minimum value zero for an ideal permutation matrix and perfect separation. The larger the E_1 is, the poorer the separation of the algorithm [1]. The performance index for this simulation of the algorithm is calculated as

- $E_1 = 0.2154$.

The performance index and the results illustrated show that the source signals are almost perfectly recovered, and the mixing is eliminated efficiently.

In the second seismological simulation scenario, the recordings from a local earthquake ($s_1(k)$) and a remote earthquake ($s_2(k)$) are again instantaneously mixed with the random mixing matrix

- $\mathbf{A} = \begin{bmatrix} 0.7826 & -0.6871 \\ 0.5242 & -0.5630 \end{bmatrix}$

using the system equation (2.1) to obtain the mixture signals $x_1(k)$ and $x_2(k)$. The local signal used in this scenario is different than the one used in the previous scenario, however the remote signal is common to both. The source signals, the mixture signals and the estimated source signals are illustrated in Figure 4.12, Figure 4.13 and Figure 4.14 respectively.

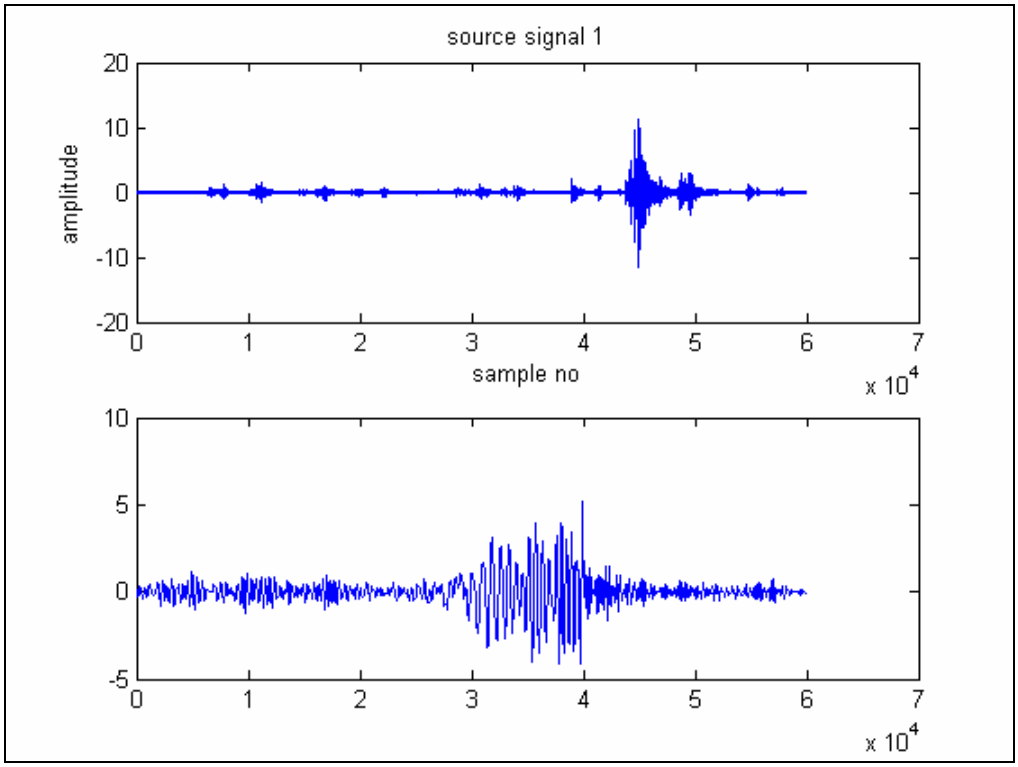

Figure 4.12. Source signals for the second seismological simulation scenario (last 60000 samples)

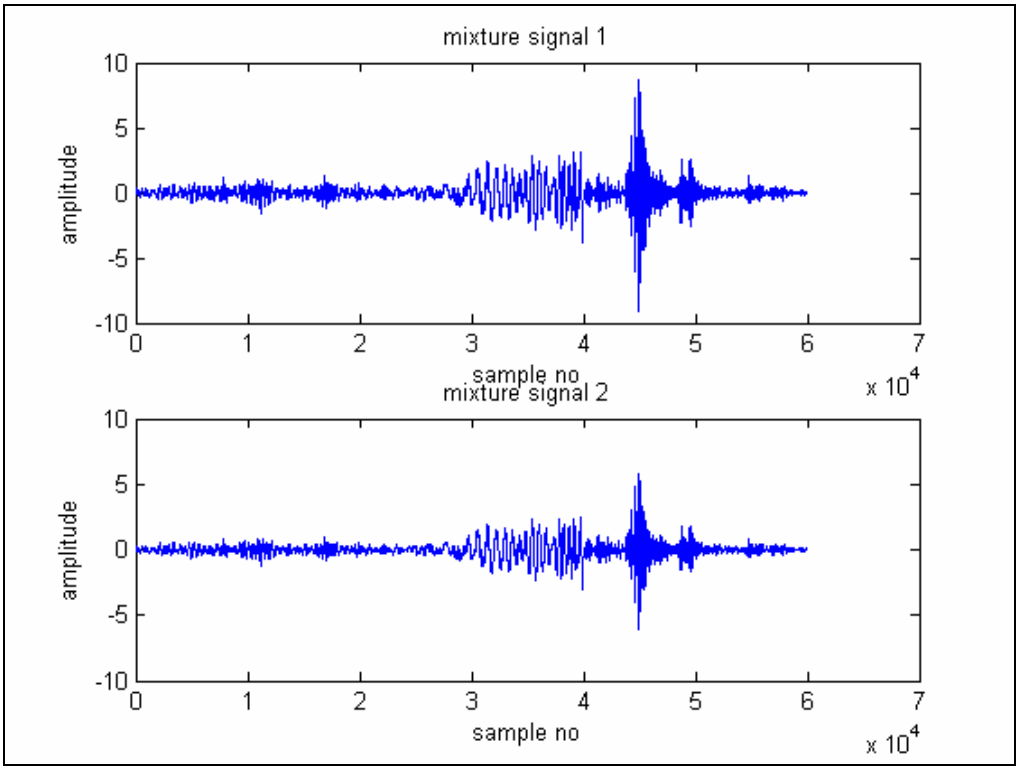

Figure 4.13. Mixture signals for the second seismological simulation scenario (last 60000 samples)

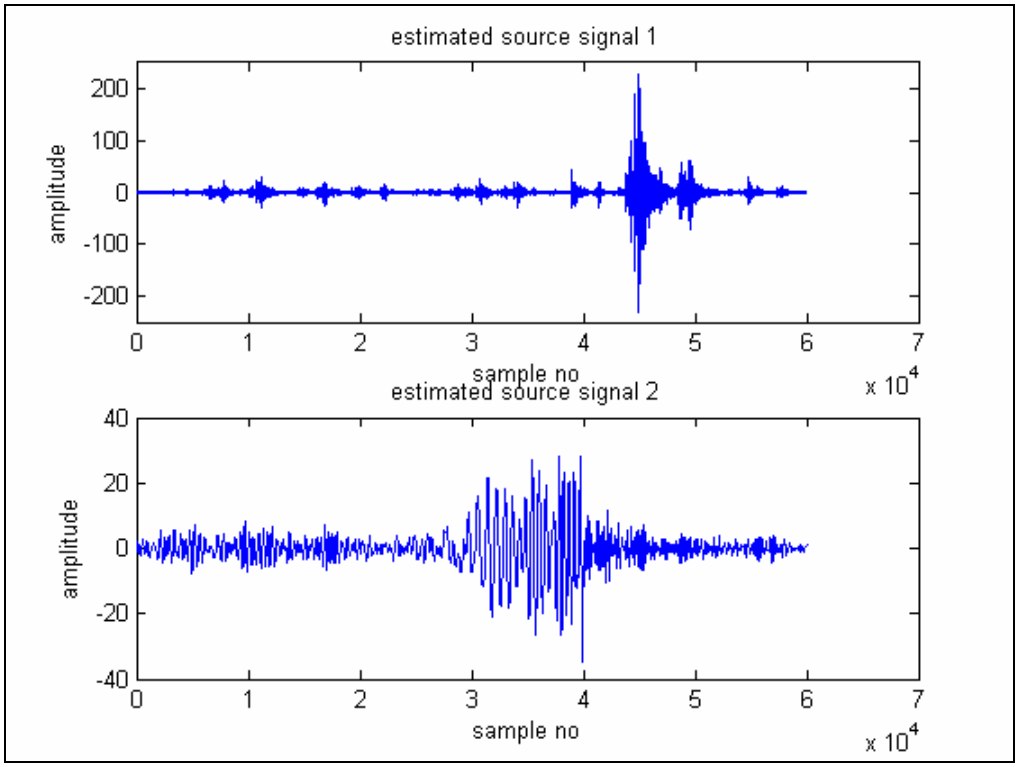

Figure 4.14. Estimated source signals for the second seismological simulation scenario (last 60000 samples)

The inverse system matrix

- $\mathbf{W} = \begin{bmatrix} 225.9 & -275.9 \\ -66.1 & 92.6 \end{bmatrix}$

and the overall system matrix

- $\mathbf{WA} = \begin{bmatrix} 20.1 & -0.1 \\ -0.4 & -6.7 \end{bmatrix}$

is obtained for the simulation.

The performance index for this simulation of the algorithm is calculated as

- $E_1 = 0.0885$.

Again the performance index and the results illustrated show that the source signals are efficiently separated.

Observing the obtained results in the two simulations, it is seen that the artificially instantaneously mixed independent seismological signals were efficiently separated using the instantaneous ICA approach. It can be argued the real-life counterpart of the problem is a convolutive one since the signals recorded at each station have different travel paths and have different delays. However signals coming from thousands of kilometers away can be assumed to be arriving at the stations simultaneously. In addition, the instantaneous mixture assumption holds true if the stations are roughly at equal distance to the local source, which is typical. In cases where it's absolutely known that the truth and the assumptions are not consistent, the observation signals may be aligned with respect to a reference point in time introducing delays.

4.3.2. Single Channel Blind Deconvolution of Seismological Signals

In this part, the results of two simulations performed with the modified natural gradient algorithm applied on seismological signals are presented. The algorithm was used for single-channel BD of seismological signals in order to separate the effects of the propagation medium from the original source signal. The algorithm was separately applied on the different continuous recordings of the same aftershock activity at Urla.

In the third seismological simulation scenario, the modified natural gradient BD method is applied to the recording of an observation station. The goal of the process is to find the source time function and the impulse response of the propagation media. The original observation signal depicted in Figure 4.15 is preprocessed and the signal portions with negligible power are removed from the signal set as Amari *et al.* did in [7] for the BD of speech signals, and the signal illustrated in Figure 4.16 is obtained.

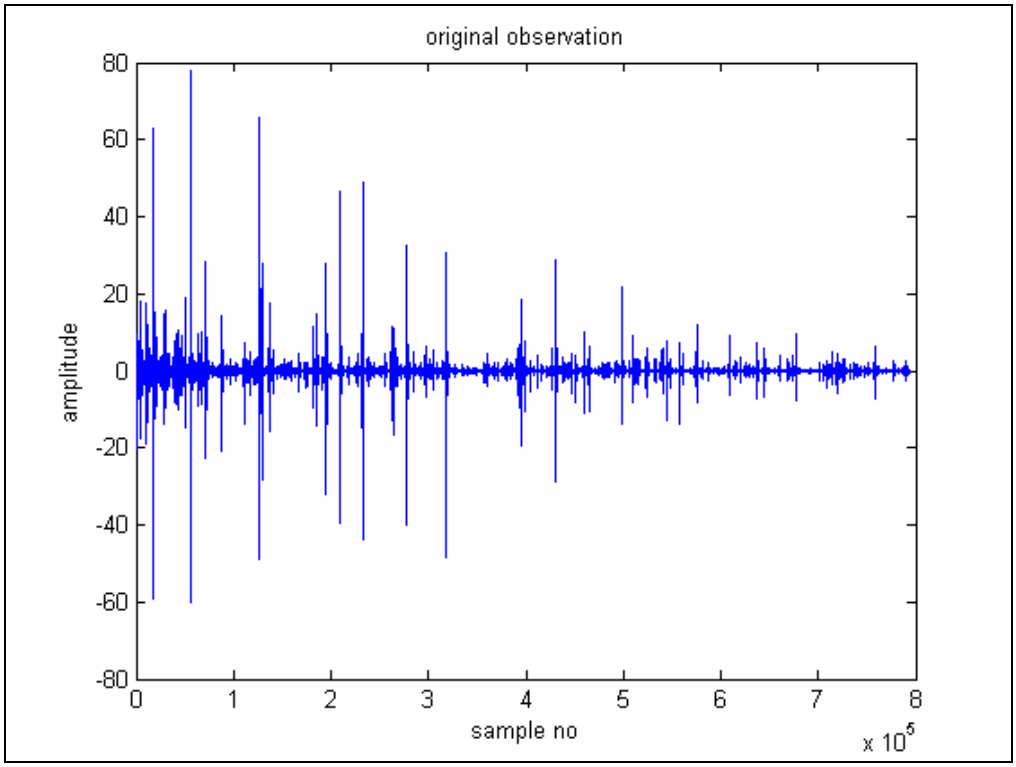

Figure 4.15. Original observation signal for the third seismological simulation scenario

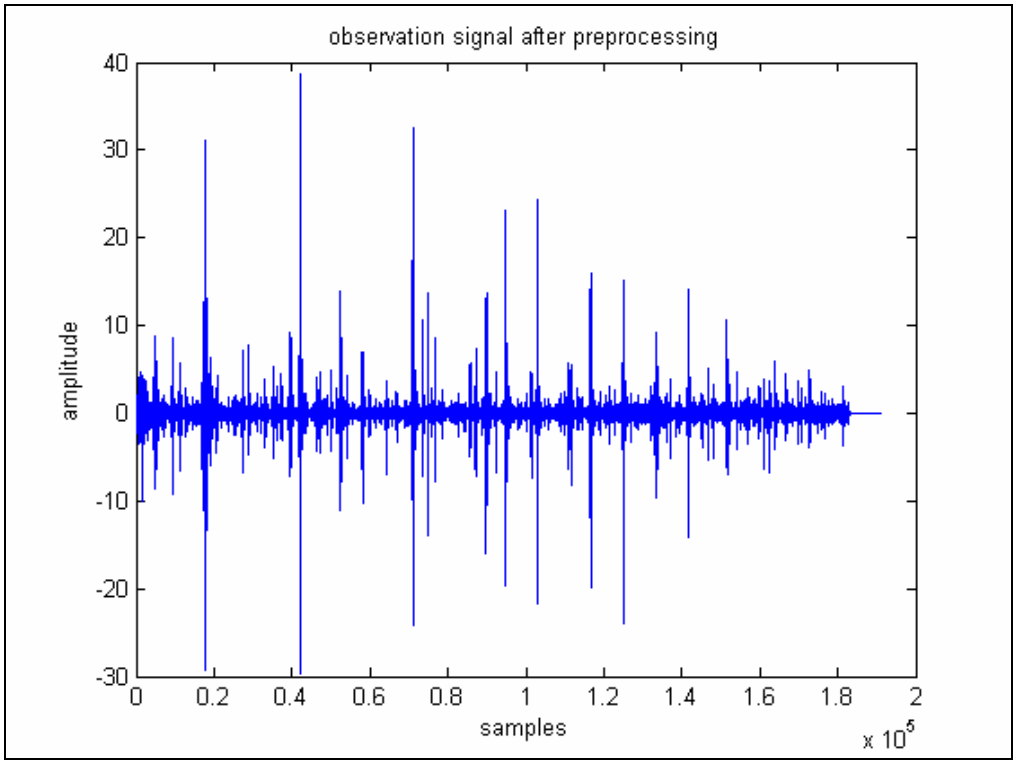

Figure 4.16. Observation signal after preprocessing for the third seismological simulation scenario

The signal obtained after the preprocessing procedure is blindly deconvoluted using the modified natural gradient BD algorithm with the deconvolution filter order set to 700. As a result of the process the signal estimate and the impulse response estimate depicted in Figure 4.17 and Figure 4.18 are obtained respectively. The estimate of the impulse response is obtained using the simple time domain inversion algorithm invimplms.m presented in [33].

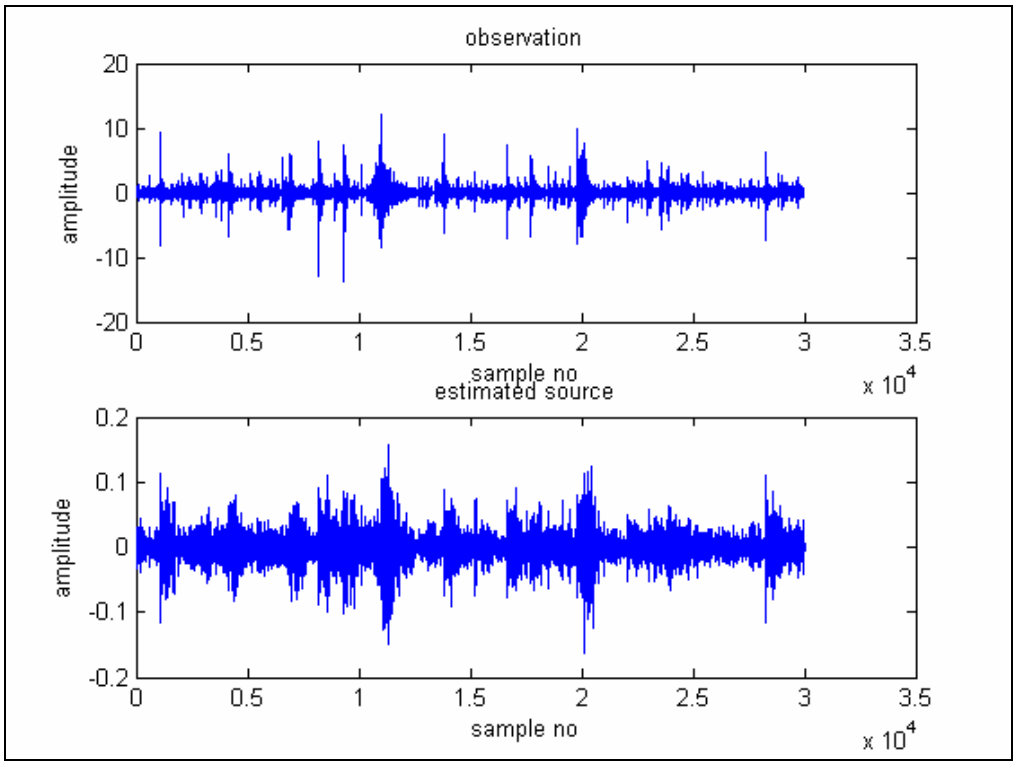

Figure 4.17. The observation signal and the estimated source signal for the third seismological simulation scenario (last 30000 samples)

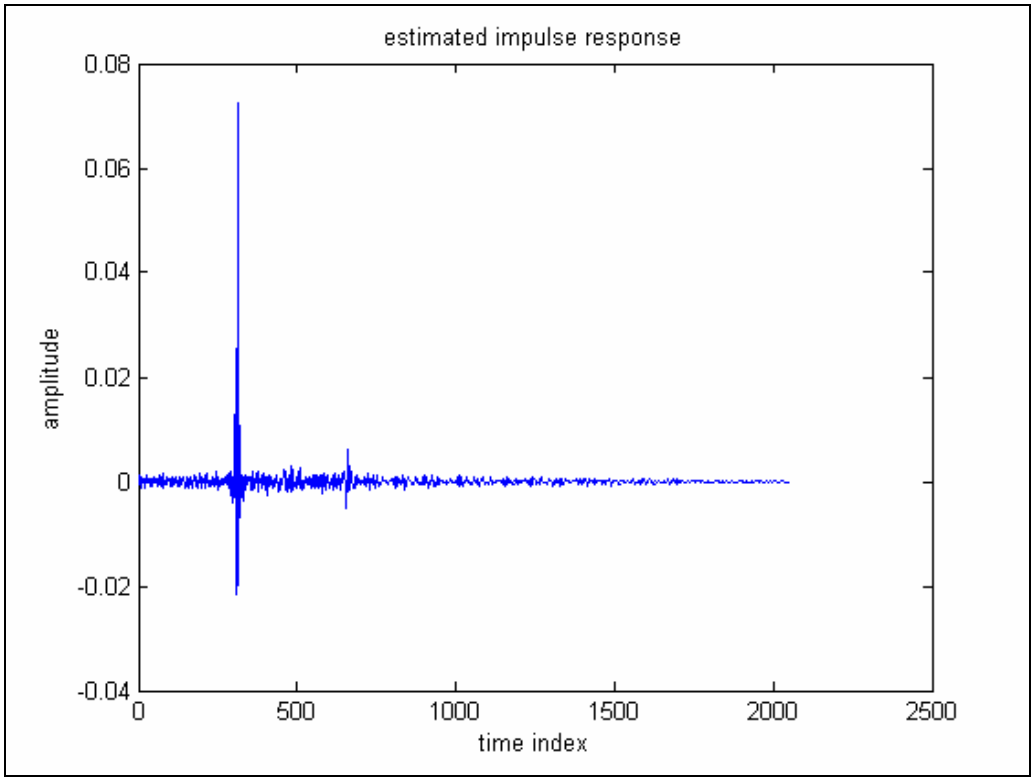

Figure 4.18. Estimated impulse response for the third seismological simulation scenario

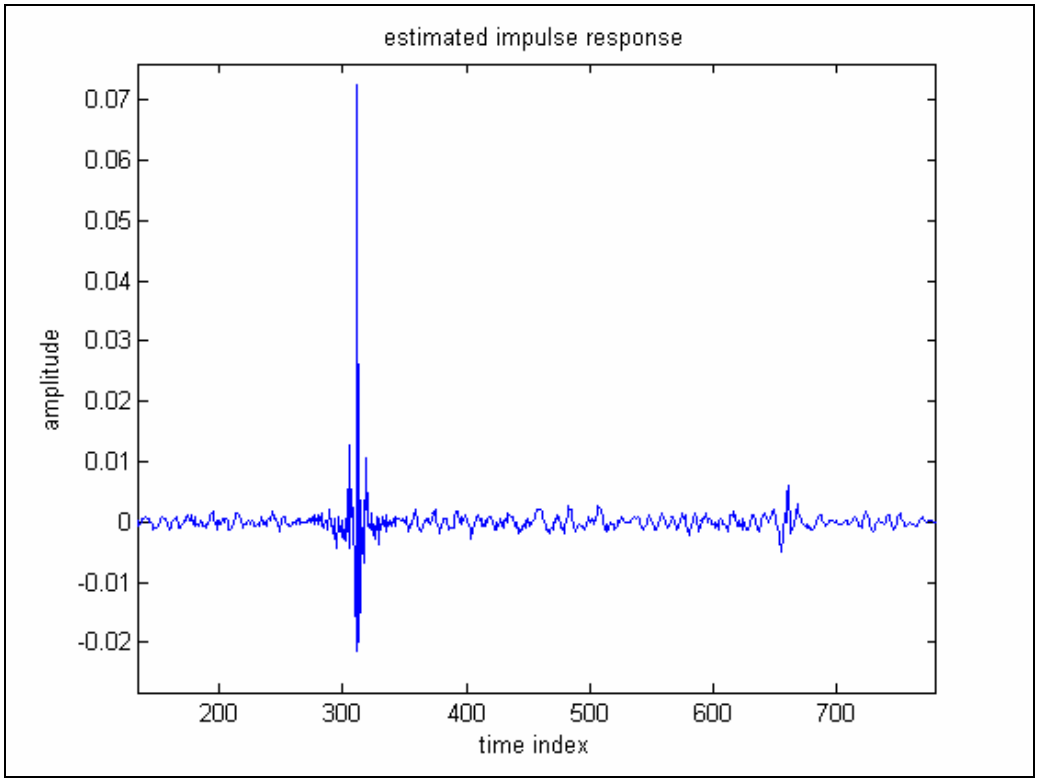

Figure 4.19. Estimated impulse response for the third seismological simulation scenario (zoomed in time)

In the fourth seismological simulation scenario, the modified natural gradient BD method is again applied to the recording of an observation station. As stated before, this recording is from a station that is simultaneously recorded with the observation in the previous scenario. The original observation signal depicted in Figure 4.20 is preprocessed and the signal portions with negligible power are removed from the signal set as previously and the signal illustrated in Figure 4.21 is obtained.

The signal obtained after the preprocessing procedure is blindly deconvoluted using the modified natural gradient BD algorithm with the deconvolution filter order set to 700. As a result of the process the signal estimate and the impulse response estimate depicted in Figure 4.22 and Figure 4.23 are obtained respectively. Again the estimate of the impulse response is obtained using the simple time domain inversion algorithm invimplms.m presented in [33].

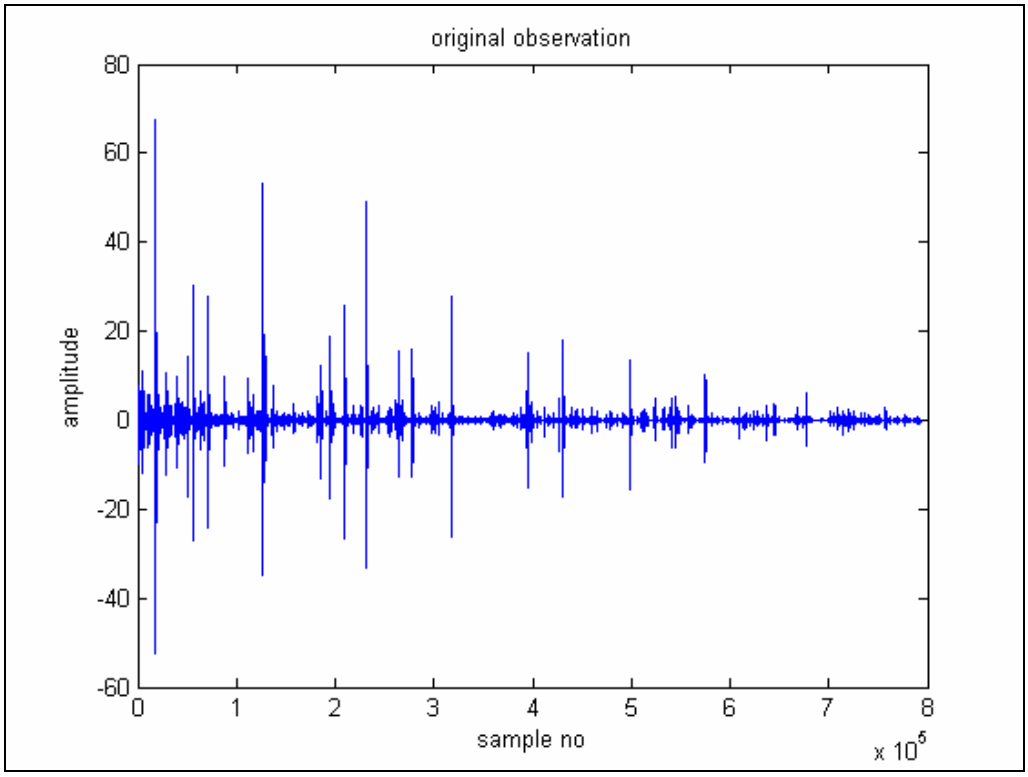

Figure 4.20. Observation signal for the fourth seismological simulation scenario

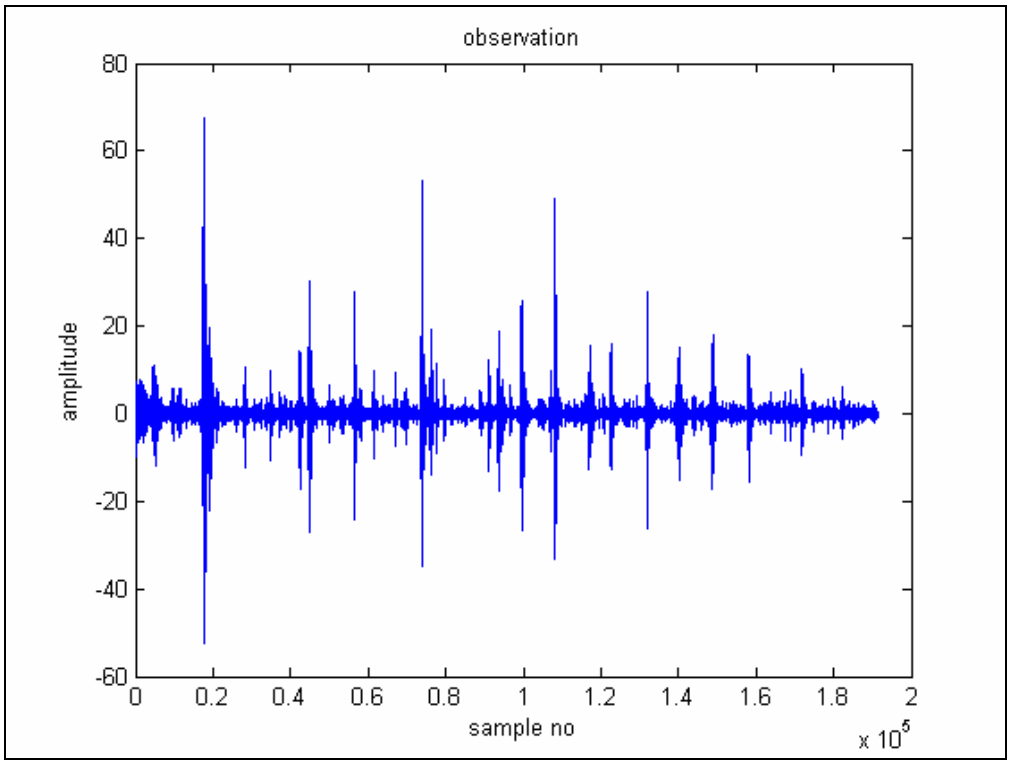

Figure 4.21. Observation signal after preprocessing for the fourth seismological simulation scenario

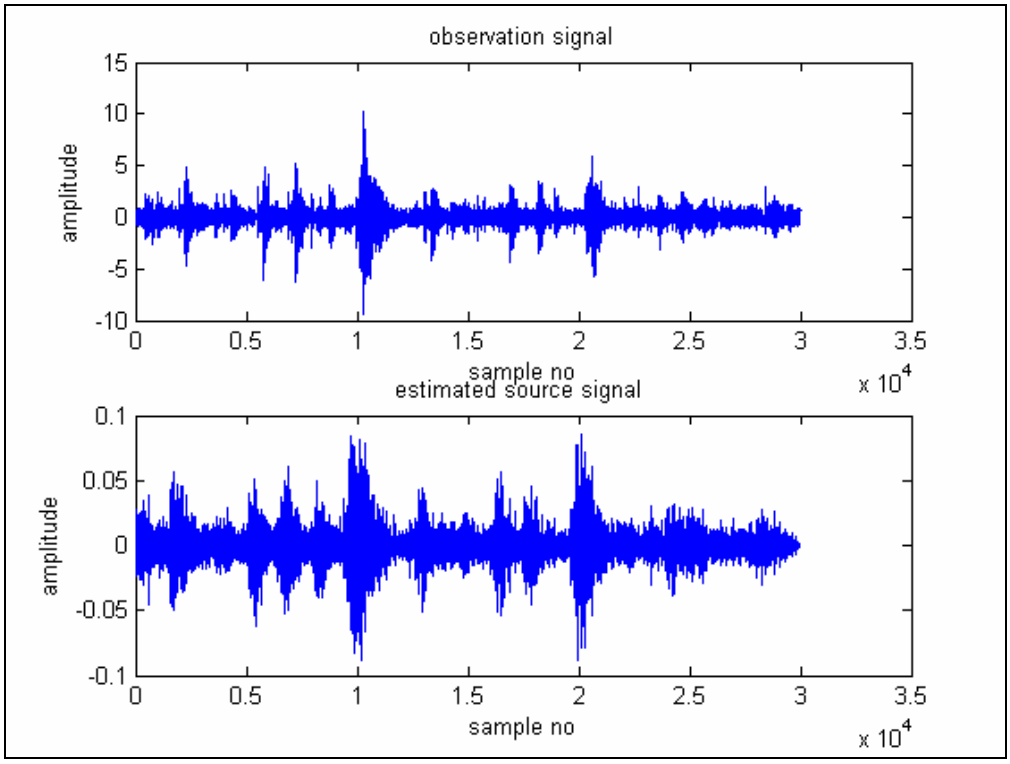

Figure 4.22. The observation signal and the estimated source signal for the fourth seismological simulation scenario (last 30000 samples)

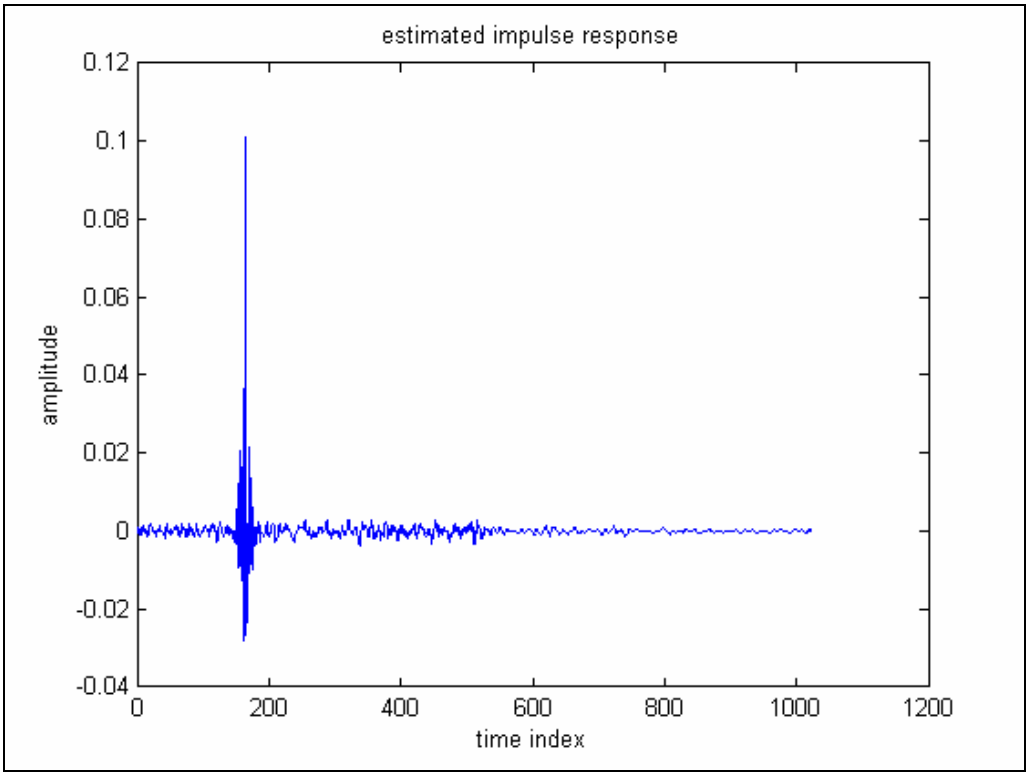

Figure 4.23. Estimated impulse response for the fourth seismological simulation scenario

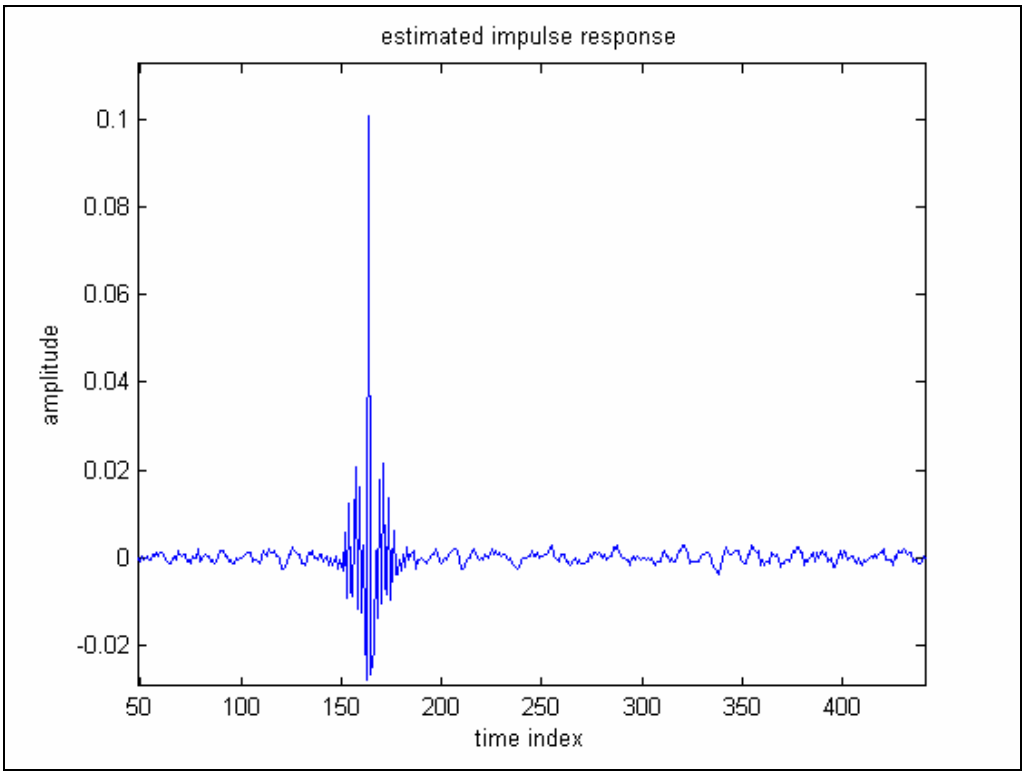

Figure 4.24. Estimated impulse response for the fourth seismological simulation scenario (zoomed in time)

Observing the obtained figures, namely Figure 4.18, Figure 4.19, Figure 4.23 and Figure 4.24, we see that the obtained estimates for the forward channels have similar characteristics with some differences. Both estimates have dominant impulses with more or less symmetrical and considerably little coefficients around them. Although we have no means to test whether these estimates exactly corresponds to the correct impulse responses in real life, they may provide information about the characteristics of the propagation medium of the seismological waves.

In real-life, what is needed in most site response analyses is the dominant frequency in the impulse response. The dominant frequency is found by selecting the frequency which corresponds to the largest peak of the Fourier amplitude spectrum of the impulse response. In most practical applications, we are only interested to know the thickness of the soft sediments layer that overly the hard rock at bottom. This is because most of the site amplification that is expected to occur at certain location is only due to the soft sediment layer (often between 3-200 meter) and the properties of the hard rock below the soft sediments have practically no influence on the overall amplification. For simple "soft sediment over hard rock model" (or one-layer over half-space model), it is well known that the depth of the low velocity soft sediment is related to the dominant frequency by the following relation:

$$th_s = \frac{v_w}{f_d \times 4} \tag{4.8}$$

where th_s is the thickness of sediment, v_w is the velocity of the seismic wave and f_d is the dominant frequency of the impulse response.

One can see that if the velocity of seismic wave is known, the thinkness of the soft sediment can easily be found using the dominant frequency of the estimated impulse response. We therefore conclude that the present BD method is useful for estimating the thickness of the upper soft layer. However in more complicated situations where a full detailed multi-layer model of the ground is needed, the time domain information will also be required which cannot done by the present BD algorithm.

5. CONCLUSIONS

BSS and ICA techniques have been implemented in many fields varying from separating mixtures of speeches of different speakers to analyzing time series of financial data. Due to this huge potential of the techniques, new methods and application areas arise day-by-day.

In this work, it is aimed to apply the BSS and BD methods on the seismological signals in order to solve two problems. The first problem is the separation of the different signals generated by two independent and distant earthquake occurrences. Instantaneous ICA method was used to model and solve the problem. The second problem is the separation of the effects of the propagation medium from the seismological source. Single channel BD methods were employed to solve the problem with a convolutive ICA approach. For the simulations, the natural gradient BD and modified gradient BD based on AR modeled inputs proposed in [7] and [11] implemented in MATLAB. Based on the results obtained in part 4, it can generally be stated that the natural gradient algorithm works efficiently for the single-channel and multichannel blind deconvolution process, and the algorithm has desired properties.

The first seismological application of ICA consisted of separating two components generated from two independent seismological sources performed well in the sense that the two signals were separated efficiently using the instantaneous ICA methods. Since the mixture was created artificially, the evaluation of the separation performance was simply a matter of comparing the outcomes with the original unmixed components. The separation was nearly complete in the sense that the local earthquakes with short impulsive character were well isolated from sustained oscillations of large period that originate from the teleseismic event. The example studied is a typical case of signal from local earthquakes at a distance of about 40 km, being contaminated by distant ones, about 3000 km. However, a wide range of similar situations can be encountered where the distance between the two sources are not as large as this case. In this situation a proportional degradation of the separation performance should naturally be expected. A second point to consider is to what

degree the instantaneous assumption would be valid. Since the signals recorded at different stations, signals arriving at each location would have a different travel path, therefore a different delay which is contrary to the instantaneous assumption. However, for stations used in local studies, i.e., with distance less than 100 km from each other, signals coming from few thousands of km can easily be assumed to be arriving simultaneously. However for the second component which includes the local earthquakes, this assumption only holds for stations that are roughly at equal distance to the local epicenters. It is clear that the method can only be applied to stations that are installed on a circle around the source, which is the typical geometry for studying local earthquakes. Finally, the simulation assumes that perfect replica of each component are recorded at all observation points which may not the exact case in real life situation. In practice, since signals are assumed to be recorded at two different observation points, the components will not be exactly the same but will include slight differences. Furthermore a noise term generated by the observational system will also be present. In conclusion, a limited degradation of the performance should be expected when operating in real life conditions, as compared to the idealistic case used in the simulations.

The second application of ICA on seismological signals is the separation of the effects of the propagation medium from the seismological sources using the convolutive ICA approach to find the soft sediment thickness. Single channel modified natural BD method was applied on the continuous recordings of the aftershock activities after the Urla earthquakes (17 October 2005 at 05:45 UTC, Mw 5.4; 17 October at 09:46 UTC, Mw 5.8 and 20 October at 21:40 UTC, Mw 5.8). Observing the corresponding simulation results, it is seen that the convolutive ICA method can also be used to obtain estimates about the source time function and the impulse response of the propagation media for any observed seismological trace. And using these data, the soft sediment thickness can be estimated using the simple relationship described. However, the sparsity of the observations is a challenging issue for this task and it must be handled as done in this work.

5.1. Future Work

As a future work, multichannel BD techniques can be applied to the seismological observations to gather information about the source time functions and the impulse response of the propagation media for the data processed with instantaneous ICA method. However, the need for large data sets with the natural gradient algorithm is a challenging task for this purpose and it must be handled by some means. In addition, sparse component analysis (SCA) methods can be employed to process seismological signals, since it is known the seismological source signals and the impulse responses are both sparse.

REFERENCES

1. Hyvarinen A., J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley and Sons Ltd., New York, 2001.

2. Haykin S., *Unsupervised Adaptive Filtering*, Vol. 1, John Wiley and Sons Ltd., New York, 2000.

3. Torkkola K., "Blind Separation of Delayed and Convolved Sources", in S. Haykin (ed.), *Unsupervised Adaptive Filtering*, Vol. 1, pp.321-275, John Wiley and Sons Ltd., New York, 2000.

4. Cichocki A. and S. Amari, *Adaptive Blind Signal and Image Processing-Learning Algorithms and Applications*, John Wiley and Sons Ltd., West Sussex, 2002.

5. Godard D. N., "Self-recovering equalization and carrier tracking in two-dimensional data communication systems", *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1867-1875, 1980.

6. Amari S., "Natural gradient works efficiently in learning", *Neural Computation*, Vol. 10, pp. 251-276, 1998.

7. Amari S., S. C. Douglas, A. Cichocki, and H. H. Yang, "Novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient approach", *in Proc. SYSID*, Kitakyushu, Japan, July 8-11, 1997, pp. 1057-1062.

8. Bell A.J. and T.J. Sejnowski, "An information maximization approach to blind seperation and blind deconvolution", *Neural Computation*, vol. 7, pp. 1129-1159, 1995.

9. Zerva A., A. Petropulu and P.Y. Bard, “Blind deconvolution methodology for site response evaluation exclusively from ground surface seismic recordings”, *Soil Dynamcics and Earthquake Engineering*, vol. 18, pp. 47-57, 1999.

10. Liao B. and H. Huang, “Estimation of the source time function based on blind deconvolution with gaussian mixtures”, *Pure and Applied Geophysics*, vol. 162, Issue 3, pp.479-494, Mar. 2005.

11. Sun X. and S. Douglas, “A natural gradient convolutive blind source separation algorithm for speech mixtures”, *in Proc. 3rd Int. Conf. ICA Source Separation*, San Diego, CA, Dec. 2001.

12. Torkkola K., “Blind separation of convolved sources based on information maximization”, *in Proc. IEEE Workshop on Neural Networks and Signal Processing*, Kyota, Japan, 1996.

13. Comon P., “Independent component analysis - a new concept?”, *Signal Processing*, vol. 36, pp. 287-314, 1994.

14. Araki S., S. Makino, T. Nishikawa, and H. Saruwatari, “Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech”, *In ICASSP*, pp 2737--2740, 2001.

15. Larue A.,J.I. Mars and C. Jutten,“Frequency-Domain Blind Deconvolution Based on Mutual Information Rate”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 54, Issue: 5, pp. 1771- 1781, May 2006.

16. Sawada H., R. Mukai, S. Araki, and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation”, *IEEE Transactions on Speech and Audio Processing*, vol.12, pp.530–538, Sept. 2004.

17. Yellin D. and E. Weinstein, “Multichannel signal separation: methods and analysis”, *IEEE Transactions on Signal Processing*, vol. 44, pp. 106-118, Jan. 1996.

18. Tugnait J. K., “Adaptive blind separation of convolutive mixtures of independent linear signals”, *Signal Processing*, vol. 73, pp. 139-152, 1999.

19. Amari S., A. Cichocki, and H.H. Yang, “A new learning algorithm for blind signal separation”, *Advances in Neural Information Processing Systems*, vol. 8, pp. 757-763, 1996.

20. Amari S., S. C. Douglas, A. Cichocki, and H. H. Yang, “Multichannel blind deconvolution and equalization using the natural gradient”, *in Proc. SPAWC, Paris, France*, Apr. 16-18, 1997, pp. 101-104.

21. Cardoso J.F. and B. H. Laheld, “ Equivariant adaptive source separation”, *IEEE Transactions on Signal Processing*, vol. 44, pp. 3017-3030, 1996.

22. Amari S., T.P. Chen and A. Cichocki, “Stability analysis of learning algorithms for blind source separation”, *Neural Networks*, vol. 10, pp. 1345–1351, Nov. 1997.

23. Douglas S. C., A. Cichocki, and S. Amari, “Multichannel blind separation and deconvolution of sources with arbitrary distributions”, *in IEEE Workshop on Neural Networks for Signal Processing*, Almelia Island Plantation, FL, pp. 436-445, Sep. 1997.

24. Peter M. Shearer, *Introduction to Seismology*, Cambridge University Press, Cambridge, 1999.

25. St. Michael Gymnasium Monschau, *On the seismogram as a 'Journey through the interior of the earth*,
http://www.mgm.monschau.de/seismic/english/artikel/artikel_2b.php, 2005

26. Baig A. M., M. G Bostock and J.-P. Mercier, “Spectral reconstruction of teleseismic P Green's functions”, *Journal of Geophysical Research*, vol. 110, Aug. 2005.

27. Santamaría I., C. Pantaleón, J. Ibáñez and A. Artés, "Deconvolution of Seismic Data Using Adaptive Gaussian Mixtures", *IEEE Transactions. on Geoscience and Remote Sensing,* vol. 37, no.2, pp. 855-858, 1999.

28. Larue A., M. V. D. Baan, J. I. Mars and C. Jutten, "Sparsity or whiteness: what criterion to use for blind deconvolution of seismic data?" *SEG Technical Program Expanded Abstracts,* pp. 1642-1645, 2005.

29. Zerva A., A.P. Petropulu and P.-Y. Bard, "Blind System Identification", *Proceedings of the 8th ASCE Joint Specialty Conference on Probabilistic Mechanics and Structural Reliability*, PMC2000, University of Notre Dame, Indiana, 2000.

30. Pozidis H. and A.P. Petropulu, "Cross-Spectrum Based Multichannel Blind Identification", *IEEE Transactions on Signal Processing*, vol. 45, no. 12, pp. 2977-2993, Dec. 1997.

31. Velis D.R. and T.J. Ulrych, "Simulated annealing wavelet estimation via fourth-order cumulant matching", *Geophysics*, vol. 61, no. 6, pp.1939-1948, 1996.

32. Pflug L. A., "Principal domains of the trispectrum, signal bandwidth, and implications for deconvolution", *Geophysics*, vol. 65, issue 3, pp. 958-969, 2000.

33. Cain B., http://www.musicdsp.org/showone.php?id=188, last reached 09 Sept. 2006

APPENDIX A: THE NATURAL GRADIENT ALGORITHM

In multichannel blind deconvolution, an m-dimensional observation vector $\mathbf{x}(k) = [x_1(k) \cdots\cdots x_m(k)]^T$ is modeled to be formed by the system equation

$$\mathbf{x}(k) = \underline{\mathbf{H}}(z)[\mathbf{s}(k)] = \sum_{p=0}^{M} \mathbf{H}_p \mathbf{s}(k-p), \tag{A.1}$$

where $\mathbf{s}(k) = [s_1(k) \cdots\cdots s_n(k)]^T$ is the n-dimensional vector of sources at time index k, $\underline{\mathbf{H}}(z)$ is the convolutive mixing system, and $\mathbf{H}_p$ is the matrix of filter coefficients at time index p.

Defining the mixing process as in (A.1), the multichannel blind deconvolution schemes try to find a deconvolution system, such that the estimation signals

$$\mathbf{y}(k) = \underline{\mathbf{W}}(z,k)[\mathbf{x}(k)] = \underline{\mathbf{W}}(z,k)\underline{\mathbf{H}}(z)[\mathbf{s}(k)], \tag{A.2}$$

are equal to the original source signals up to permutation, scaling and delay factors where

$$\underline{\mathbf{W}}(z,k) = \sum_{m=-\infty}^{\infty} \mathbf{W}_m(k) z^{-m} \tag{A.3}$$

is the z-transform of the deconvolution system and $\underline{\mathbf{H}}(z)$ is the z-transform of the mixing system.

To adaptively find an estimate for the matrix system function, the natural gradient algorithm may be used introducing the loss (cost) function for the matrix system

$$\phi(\underline{\mathbf{W}}(z,k)) = -\sum_{i=1}^{n} \log p_i(y_i(k)) - \frac{1}{2\pi j} \oint \log\left|\det \underline{\mathbf{W}}(z,k)\right| z^{-1} dz, \tag{A.4}$$

where $p_i(y_i)$ is defined as the p.d.f. of the i^{th} estimate and $j = \sqrt{-1}$. Minimization of this criterion corresponds minimizing the mutual information of the extracted source signals in $\mathbf{y}(k)$ [7].

To minimize the expected value of the $\phi(\underline{\mathbf{W}}(z,k))$, the steepest descent direction $d\underline{\mathbf{X}}(z,k)$ will be used. The term

$$d\phi(\underline{\mathbf{W}}(z,k)) = \phi((\underline{\mathbf{W}}(z,k)) + d\underline{\mathbf{W}}(z,k)) - \phi(\underline{\mathbf{W}}(z,k) \tag{A.5}$$

is the gradient of the loss function [7].

Introducing

$$f_i(y_i) = -\frac{d}{dy_i}\log p_i(y_i), \tag{A.6}$$

and a vector valued function

$$\mathbf{f}(\mathbf{y}(k)) = \left[f_1(y_1(k)\cdots\cdots f_m(y_m(k))\right], \tag{A.7}$$

one obtains

$$d\left(-\sum_{i=1}^{n}\log p_i(y_i(k))\right) = \mathbf{f}^T(\mathbf{y}(k))d\mathbf{y}(k). \tag{A.8}$$

Observing equation (2.9), it follows that

$$d\mathbf{y}(k) = d\underline{\mathbf{W}}(z,k)\underline{\mathbf{W}}^{-1}(z,k)\mathbf{y}(k). \tag{A.9}$$

Defining a modified coefficient differential

$$d\underline{\mathbf{X}}(z,k) = \sum_{p=1}^{n} d\mathbf{X}_p z^{-p} = d\underline{\mathbf{W}}(z,k)\underline{\mathbf{W}}^{-1}(z,k), \tag{A.10}$$

as in [20], and combining with (A.5) and (A.6) one gets

$$d\left(\frac{1}{2\pi j}\oint \log\left|\det \underline{\mathbf{W}}(z,k)\right| z^{-1} dz\right) = tr(d\mathbf{X}_0(k)), \tag{A.11}$$

where the $tr(.)$ operator stands for the matrix trace operation. Combining these, one obtains

$$d\phi(\mathbf{W}(z,k)) = \mathbf{f}^T(\mathbf{y}(k)) d\mathbf{X}(z,k)[\mathbf{y}(k)] - tr(d\mathbf{X}_0(k)). \tag{A.12}$$

As stated in [7], the natural gradient algorithm as defined in the space spanned by $d\underline{\mathbf{X}}(z,k)$ is

$$\Delta\underline{\mathbf{X}}(z,k) = -\mu(k)\frac{d\phi(\mathbf{W}(z,k))}{d\underline{\mathbf{X}}(z,k)}. \tag{A.13}$$

Rewriting the algorithm in terms of $\underline{\mathbf{W}}(z,k)$, one obtains

$$\mathbf{W}_p(k+1) = \mathbf{W}_p(k) - \mu(k)\frac{d\phi(\mathbf{W}(z,k))}{d\underline{\mathbf{X}}(z,k)}(\mathbf{W}(z,k)). \tag{A.14}$$

Defining

$$\mathbf{u}_p(k) = \sum_{q=-\infty}^{\infty} \mathbf{W}_q^H(k)\mathbf{y}(k-p+q), \tag{A.15}$$

the update equation is obtained as

$$\mathbf{W}_p(k+1) = \mathbf{W}_p(k) + \mu(k)\left(\mathbf{W}_p(k) - \mathbf{f}(\mathbf{y}(k))\mathbf{u}_p^H(k)\right) \tag{A.16}$$

for $-\infty < p < \infty$. However, the infinite filter $\underline{\mathbf{W}}(z,k)$ is approximated by the FIR filter given by

$$\mathbf{y}(k) = \sum_{p=0}^{L} \mathbf{W}_p(k)\mathbf{x}(k-p)\,, \tag{A.17}$$

And then we have

$$\mathbf{u}(k) = \sum_{q=0}^{L} \mathbf{W}_{L-q}^{H}(k)\mathbf{y}(k-q)\,, \tag{A.18}$$

and

$$\mathbf{W}_p(k+1) = \mathbf{W}_p(k) + \mu(k)\left(\mathbf{W}_p(k) - \mathbf{f}(\mathbf{y}(k-L))\mathbf{u}^H(k-p)\right). \tag{A.19}$$

as update equations of the algorithm [7].