Blind Separation of Noisy Multivariate Data Using Second-Order Statistics

by

Keith Herring

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
[February 2005]
December 2005

Author .....
Department of Electrical Engineering and Computer Science
December, 2005

Certified by 12/15/04
David H. Staelin
Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Blind Separation of Noisy Multivariate Data Using Second-Order Statistics

by

Keith Herring

Submitted to the Department of Electrical Engineering and Computer Science
on December, 2005, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

A second-order method for blind source separation of noisy instantaneous linear mixtures is presented and analyzed for the case where the signal order k and noise covariance $\mathbf{G}\mathbf{G}^H$ are unknown. Only a data set $\mathbf{X}$ of dimension $n > k$ and of sample size m is observed, where $\mathbf{X} = \mathbf{A}\mathbf{P} + \mathbf{G}\mathbf{W}$. The quality of separation depends on source-observation ratio $\frac{k}{n}$, the degree of spectral diversity, and the second-order non-stationarity of the underlying sources. The algorithm estimates the Second-Order separation transform $\mathbf{A}$, the signal Order, and Noise, and is therefore referred to as SOON. SOON iteratively estimates: 1) k using a scree metric, and 2) the values of $\mathbf{A}\mathbf{P}$, $\mathbf{G}$, and $\mathbf{W}$ using the Expectation-Maximization (EM) algorithm, where $\mathbf{W}$ is white noise and $\mathbf{G}$ is diagonal. The final step estimates $\mathbf{A}$ and the set of k underlying sources $\mathbf{P}$ using a variant of the joint diagonalization method, where $\mathbf{P}$ has k independent unit-variance elements.

Tests using simulated Auto Regressive (AR) gaussian data show that SOON improves the quality of source separation in comparison to the standard second-order separation algorithms, i.e., Second-Order Blind Identification (SOBI) [3] and Second-Order Non-Stationary (SONS) blind identification [4]. The sensitivity in performance of SONS and SOON to several algorithmic parameters is also displayed in these experiments. To reduce sensitivities in the pre-whitening step of these algorithms, a heuristic is proposed by this thesis for whitening the data set; it is shown to improve separation performance. Additionally the application of blind source separation techniques to remote sensing data is discussed. Analysis of remote sensing data collected by the AVIRIS multichannel visible/infrared imaging instrument shows that SOON reveals physically significant dynamics within the data not found by the traditional methods of Principal Component Analysis (PCA) and Noise Adjusted Principal Component Analysis (NAPCA).

Thesis Supervisor: David H. Staelin
Title: Professor of Electrical Engineering

Acknowledgments

I'd like to thank my thesis advisor Professor David Staelin for his help in this effort. Additionally these are some names of good people: Tom, Phyllis, Paul, Harold, Ines, Dale, Brad, Cleave, Wayne, Pat, Jay, Lucy, Keith C, Blanch, Evelyn, Richard, Todd, Gentzy, Richard, Orvel, Gentz, Jerri, Ralph, Connie, Chris, Randal, Courtney, Brian, Kyle, Chris, Zach, Loren, Darren, Scott, Margaret V. Stringfellow (TLOML)(MTL)

Contents

1 Introduction **15**

1.1 Motivation for Research 15

1.2 Blind Signal Separation (BSS) Background 16

1.2.1 Generalized Blind Signal Separation Problem 16

1.2.2 System Models 16

1.2.3 Prior Work and Solution Methods 18

1.2.4 Problem Statement 20

1.3 Thesis Outline 21

2 Second-Order separation Order-Noise estimation (SOON) **23**

2.1 SOBI 23

2.1.1 The SOBI algorithm 23

2.1.2 Limitations of SOBI 25

2.2 SONS 27

2.2.1 The SONS algorithm 27

2.2.2 Limitations of SONS 29

2.3 SOON 30

2.3.1 Order Estimation 30

2.3.2 Noise Estimation 31

2.3.3 Sensitivity of Robust Whitening 36

2.3.4 Data Partitioning 39

2.3.5 Algorithm Summary 41

3 Evaluation of SOON **43**

3.1 Problem Space 43

3.1.1 Source-to-Observation Ratio $\frac{k}{n}$ 43

3.1.2 Data Signal-to-Noise Ratio (SNR_X) 44

3.1.3 Sample Size (m) 44

3.1.4 Angle between Columns of Mixing Transform ($\angle_A$) 44

3.2 Performance Metric 45

3.3 SOON/SOBI/SONS Comparison 45

3.3.1 Test 1: Stationary-Colored Sources 46

3.3.2 Test 2: Non-stationary-White Sources 47

3.3.3 Results 49

3.4 Stability over Problem Space 50

3.4.1 Testing 50

3.4.2 Results 51

4 Application to Remote Sensing and Financial Data **53**

4.1 Remote Sensing Background 53

4.2 Principal Component Analysis (PCA) 54

4.2.1 The Method 54

4.2.2 Limitation of PCA 54

4.3 Noise Adjusted Principal Component Analysis (NAPCA) 55

4.3.1 The Method 55

4.3.2 Limitation of NAPCA 56

4.4 Blind Signal Separation Approach 56

4.4.1 BSS Model 56

4.4.2 Solution using SOON 57

4.4.3 Applicability of SOON 57

4.5 Case Study: AVIRIS Data 58

4.5.1 The AVIRIS instrument 58

4.5.2 Quantitative Comparison 58

4.5.3 Qualitative Comparison 61
4.6 Financial data 65
4.6.1 Adjusted Stock Returns Experiment 65
4.6.2 Results 65

5 Higher Order Extension **69**
5.1 Higher Order Architecture 69
5.2 Maximum Information/Maximum Likelihood 71
5.3 Cumulants 71

6 Conclusions and Suggested Further Work **73**
6.1 Suggestions for further work 73
6.1.1 Higher Order Statistics 73
6.1.2 Partitioning Schemes 74
6.1.3 Application to Other Fields 74
6.2 Conclusions 74

A **77**
A.1 Joint Diagonalization 77
A.2 Global Convergence Algorithm 78

Bibliography **81**

List of Figures

2-1 Typical Scree Plot which plots the ordered log-magnitude eigenvalues of data matrix $\mathbf{X}$. Here the number of underlying signals, k, is 20 while the number of observations, n, is 55. The horizontally orientated dotted lines represent the line fitting the plateaus of the scree plots and the vertical dotted lines represent the intersection of the two curves in which the order estimation is obtained. 32

3-1 Test 1, which compares the performance of SOON to that of SOBI and four variants of SONS. The data set here is comprised of stationary sources which have diverse correlation functions. 48

3-2 Test 2, which compares the performance of SOON to that of SOBI and two variants of SONS. The data set here is comprised of white second-order non-stationary sources. 49

4-1 Data collected over Moffet Field, CA by the AVIRIS instrument, channels 20, 50, 80, 110, 140, 170, and 200 are shown from left to right. . 59

4-2 This graph displays the noise metric defined above for each of the components found by SOON and NAPCA. The components are sorted here in ascending order of this metric. By inspection the higher ranking components of SOON, those containing the most information content, are less noisy than the corresponding NAPCA components 60

4-3 Here are the top halves of the visually most interesting components found by SOON and NAPCA. In the top row are the NAPCA components, and the bottom row presents the SOON components. 61

4-4 Here are the bottom halves of the visually most interesting components found by SOON and NAPCA. In the top row are the NAPCA components, and the bottom row presents the SOON components. . . 62
4-5 Here are the normalized channel coefficients for the 5^{th} SOON component of figure 4-4. Here it can be seen that the majority of energy is weighted in the higher frequency, visible wavelength channels of the instrument, channels 1-40. 63
4-6 Here are the normalized channel coefficients for the 6^{th} SOON component of figure 4-4. Here it can be seen that the majority of energy is weighted in the near-infrared channels, channels 40-110. 64
4-7 Here the scree plot is shown for a data set consisting of the adjusted returns for 500 US equities over the ten year period starting in 1994 through 2003. 66
4-8 Four of the highest variance components found using the SOON algorithm from the financial data set. 67
4-9 The complete 10 year data set projected onto four of the components found by SOON trained over the first seven years of data. 68

5-1 Architecture of the generalized SOON algorithm 70

List of Tables

2.1 The SOBI Algorithm 26

2.2 The SONS Algorithm 29

2.3 The Expectation-Maximization Algorithm 37

2.4 The SOON algorithm 42

3.1 SONS variants 47

3.2 SONS variants 48

3.3 Nominal Test Parameters 51

3.4 SNR (dB) and order error performance of SOON on colored stationary gaussian test data 51

3.5 SNR (dB) and order error performance of SOON on white non-stationary gaussian test data 51

Chapter 1

Introduction

1.1 Motivation for Research

As digital storage has become cheaper the number of ever larger data sets has increased accordingly. It is often desirable to reduce the dimensionality of these data sets while retaining their unique, non-redundant information. Benefits of this process include an increased understanding of the underlying dynamics present in the data, more efficient use of storage space, and a decrease in the computational processing time needed for the analysis of the transformed data set. Additionally in many research areas such as remote sensing, telecommunications, neurobiology, manufacturing, finance, and speech processing, sensor arrays produce data sets that can be characterized as mixtures of some set of interesting phenomena and unwanted noise. Here it is often desirable to obtain each interesting component separately given little information as to the underlying system and signal characteristics. The field of Blind Signal Separation (BSS) offers approaches and methods for dealing with such problems. This thesis extends and improves existing BSS algorithms.

1.2 Blind Signal Separation (BSS) Background

1.2.1 Generalized Blind Signal Separation Problem

The general problem of Blind Signal Separation (BSS) can be described as follows: Given a vector comprising n sensor output signals,

$$\mathbf{X}(t) = [x_1(t), x_2(t), ..., x_n(t)] \quad 0 \leq t \leq T \tag{1.1}$$

where each sensor output signal, x_i(t), is the output of some unknown system F_i, whose inputs are from a set of k source signals,

$$\mathbf{P}(t) = [p_1(t), p_2(t), ..., p_k(t)] \quad 0 \leq t \leq T \tag{1.2}$$

and l noise signals,

$$\mathbf{W}(t) = [w_1(t), w_2(t), ..., w_l(t)] \quad 0 \leq t \leq T \tag{1.3}$$

$$\Rightarrow x_i(t) = F_i(\mathbf{P}(t), \mathbf{W}(t)) \quad 1 \leq i \leq n \tag{1.4}$$

find an inverse system or transform such that the source signals, p_i(t) $1 \leq i \leq k$, can be obtained. In such problems, some assumptions must be made as to the structure of the systems, F_i's, for the problem to be well posed. Additionally it is also often necessary to have restrictions on the statistical characteristics of the source signals, $\mathbf{P}$(t), and noise, $\mathbf{W}$(t).

1.2.2 System Models

Although many models are possible for the systems, F_i's, there are a few special cases that have been heavily studied [4] which serve as good representations of typical systems. In each the noise model is most often additive.

Linear Convolutive Model

Here the sensor output signals, $\mathbf{X}(t)$ are modeled as the output of a linear system,

$$x_i(t) = \sum_{j=1}^{k} \int_{-\infty}^{\infty} a_{ij}(t,\tau) p_j(\tau) d\tau + \sum_{j=1}^{l} \int_{-\infty}^{\infty} g_{ij}(t,\tau) w_j(\tau) d\tau \quad 1 \leq i \leq n \tag{1.5}$$

where

$$\mathbf{A}(t,\tau) = \begin{pmatrix} a_{11}(t,\tau) & \dots & a_{1k}(t,\tau) \\ \vdots & \ddots & \vdots \\ a_{n1}(t,\tau) & \dots & a_{nk}(t,\tau) \end{pmatrix} \tag{1.6}$$

is the matrix of linear systems that the source signals pass through and,

$$\mathbf{G}(t,\tau) = \begin{pmatrix} g_{11}(t,\tau) & \dots & g_{1l}(t,\tau) \\ \vdots & \ddots & \vdots \\ g_{n1}(t,\tau) & \dots & g_{nl}(t,\tau) \end{pmatrix} \tag{1.7}$$

is the matrix of linear systems that the noise signals pass through.

Linear Instantaneous Mixture Model

An important special case of the linear convolutive model is the instantaneous linear mixture model. Here the linear systems of $\mathbf{A}(t,\tau)$ and $\mathbf{G}(t,\tau)$ are restricted to be memoryless, causal, and time invariant systems. The system equations then become,

$$\begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix} = \begin{pmatrix} a_{11}p_1(t) + \ldots + a_{1k}p_k(t) \\ \vdots \\ a_{n1}p_1(t) + \ldots + a_{nk}p_k(t) \end{pmatrix} + \begin{pmatrix} g_{11}w_1(t) + \ldots + g_{1l}w_l(t) \\ \vdots \\ g_{n1}w_1(t) + \ldots + g_{nl}w_l(t) \end{pmatrix} \tag{1.8}$$

$$\Rightarrow \mathbf{X}(t) = \mathbf{A}\mathbf{P}(t) + \mathbf{G}\mathbf{W}(t) \tag{1.9}$$

Non-Linear Models

Additionally, many systems where Blind Separation is applicable contain nonlinearities [4]. Models and solutions exist for these types of systems, however because of the increased complexity involved, these systems are often approximated by their linear counterparts.

1.2.3 Prior Work and Solution Methods

An assumption in blind signal separation problems regarding the relationship between source signals, $\mathbf{P}$, is that they are mutually independent. This assumption is necessary in order to avoid ambiguity in trying to distinguish signal content that is probabilistically related. The problem can be described as given sensor outputs, $\mathbf{X}$, find $\mathbf{P}$ such that:

$$\mathbf{P} = \min_{\mathbf{P}'} g(\mathbf{P}') \tag{1.10}$$

where g(•) is a cost function measuring independence between source signals in $\mathbf{P}$. Many techniques have been proposed for solving the BSS problem under many different models and sets of assumptions. Although many methods exist, the approaches they take can be roughly divided into four general categories based on the type of cost function used for measuring independence.

Second Order Statistics

In general, the blind separation problem involves solving for approximately n^2 unknown parameters which correspond to how each unknown source signal is mixed within each sensor output. In order for a unique solution for the mixing transform to be attainable there must be at least n^2 independent sources of information for the dependencies between sensor outputs. The zero-lag covariance matrix of the data for instance provides $\frac{n^2}{2}$ independent sources of information. Given that stationary white gaussian noise is completely specified by its zero-time lag covariance matrix,

separating these signals uniquely is not possible. However if the underlying sources have non-white independent auto-correlation functions, this second-order source of dependence can be used to unmix the sources. Some popular methods that take advantage of second-order statistics include, factor analysis [1], Bayesian BSS [2], SOBI [3], and SONS [4]. Additionally a second-order method was developed in [17] that incorporated SOBI for use on low SNR data sets and cases where the signal order is unknown. In each the cost function signifies the second-order dependence of the source signals.

Higher Order Statistics

If the underlying sources are non-gaussian, they will not be uniquely specified by their first and second moments. In these cases, the higher order moments of the sources can be used to specify dependence between sensor outputs. Many possibilities for cost functions exist in this domain, including kurtosis, mutual information, and other non-gaussian features. Some widely used methods for higher-order based source separation include contrast functions, cumulant matching [5][6][7], and independent component analysis [8]. In each the goal is to minimize some cost function measuring the higher-order dependencies between source signals.

Non-Stationarity

An additional potential characteristic of the source signals that methods attempt to use is the non-stationarity of their probabilistic behavior. If the probability distributions for each source signal vary independently of one another then dependence of the sensor outputs will correspondingly vary. Dependencies between the sensor outputs at different times then give rise to independent sources of information. Cost functions are derived based on finding source signals that are independent over each of these time windows [4].

Time-Frequency Diversity/Orthogonality

Signal separation is also carried out by taking advantage of any orthogonality between the source signals. If the source signals are known a-priori to fall within non-overlapping frequency bands then separation can be as simple as band-pass filtering the data set. Additionally if the sources fall in non-overlapping time segments, the separation is again straight forward.

1.2.4 Problem Statement

Solution methods for BSS that utilize higher order statistics are computationally demanding and require large sample sizes in order to obtain statistically accurate results. Higher order statistics are not applicable to gaussian sources. Methods falling into the category of Time-Frequency Diversity require a restrictive set of assumptions. We therefore restrict our attention in this thesis to the second-order solution domain of BSS. Here the problem we are interested in is that of developing a BSS method that uses all possible information available from second-order statistics, which is also robust to noise. In particular we adopt for this thesis the instantaneous linear mixture model for which we have m samples of the sensor output vector $\mathbf{X}$(t):

$$\begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix} = \mathbf{A} \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{km} \end{pmatrix} + \mathbf{G} \begin{pmatrix} w_{11} & \cdots & w_{1m} \\ \vdots & \ddots & \vdots \\ w_{l1} & \cdots & w_{lm} \end{pmatrix} \tag{1.11}$$

$$\Rightarrow \mathbf{X} = \mathbf{AP} + \mathbf{GW} \tag{1.12}$$

In order to make the solution unique we make the standard assumption that the mixing transform, $\mathbf{A}$, is of full column rank. Here we also assume the noise, $\mathbf{W}$, to be temporally white, i.e.,

$$E[W_i W_j^H] = \delta_{i-j}\mathbf{I} \tag{1.13}$$

where W_i represents the i^{th} column of $\mathbf{W}$. Additionally the noise covariance matrix, $\mathbf{G}\mathbf{G}^H$, is assumed to be diagonal. This provides an accurate model for any instrument or measurement noise that may be present in the observations. While $\mathbf{G}\mathbf{G}^H$ is restricted to be diagonal it should be noted that the diagonal elements can take on any value, allowing the noise variances to vary widely across sensor outputs in $\mathbf{X}$. Here we also fix the source signals to have unit variance,

$$E[P_i P_i^H] = 1 \tag{1.14}$$

where P_i denotes the i^{th} row of $\mathbf{P}$. This assumption removes any scaling ambiguity in the separation of $\mathbf{A}$ and $\mathbf{P}$. The goal is to under this model come up with a solution method for best estimating the source signals, $\mathbf{P}$, based exclusively on knowledge of $\mathbf{X}$.

1.3 Thesis Outline

Chapter 2 outlines the prior work in second-order blind signal separation, and introduces a new algorithm in this problem domain, SOON. SOON extends the set of second-order algorithms by addressing some of the limitations associated with other techniques in this set. Section 2.1 introduces the popular second-order separation algorithm, SOBI. Here the theoretical justification for the two-stage approach used in SOBI, namely pre-whitening and joint diagonalization, is outlined. Additionally the limitations of the SOBI algorithm are discussed. In Section 2.2 a more recently proposed second-order algorithm, SONS, is introduced. Here it is shown how SONS improves upon SOBI along with noting the limitations left unaddressed by SONS. Section 2.3 introduces the SOON algorithm proposed in this thesis. This section shows how SOON improves upon SOBI and SONS through a more robust approach to order and noise estimation. Additionally in this section the importance of several algorithmic parameters is discussed along with the proposal of several heuristics for working with these parameters.

In Chapter 3 the performance of SOON is compared to that of SOBI and SONS over simulated data. In Section 3.1 the problem space of BSS problems is characterized by discussing several degrees of freedom to which separation performance is sensitive. Section 3.2 introduces the performance metrics used in this thesis to gauge algorithmic performance. In Section 3.3 a test using simulated colored stationary sources is presented and the results are displayed. Similarly in Section 3.4 a second test with simulated white nonstationary sources is proposed, and the results are then analyzed. Section 3.5 explores the sensitivity of the SOON algorithm to the degrees of freedom discussed in Section 3.1.

Chapter 4 examines the application of blind signal separation to remote sensing data. Section 4.1 motivates this study by justifying the desire for blind signal separation in remote sensing. Section 4.2 introduces the classic Principal Components Algorithm (PCA) which is often used in remote sensing for component analysis. Section 4.3 then outlines the Noise Adjusting extension of PCA (NAPCA) along with discussing its drawbacks. Next in Section 4.4 a two-dimensional blind signal separation model is presented. Finally in Section 4.5 an experiment using remote sensing images collected by the AVIRIS 224-channel instrument is performed to compare SOON to PCA and NAPCA.

Chapter 5 discusses the extension of the SOON algorithm to include higher order statistics. In Section 5.1 a generalized architecture is presented that allows for the separation of signals based on a general cost function. In Section 5.2 the Information Maximization and Maximum Likelihood approaches to signal separation are introduced and discussed. Section 5.3 examines the cumulant based approach to higher-order separation.

Chapter 6 outlines possible extensions to this work and summarizes the work presented in this thesis. In Section 6.1 the higher order extensions introduced in Chapter 5 are further discussed. Additionally Section 6.1 elaborates on further work in partitioning schemes for taking advantage of non-stationarity in data sets. Lastly in this section there is a discussion on how this work can be applied to data sets collected in many other fields. Section 6.2 concludes by summarizing this work.

Chapter 2

Second-Order separation Order-Noise estimation (SOON)

2.1 SOBI

2.1.1 The SOBI algorithm

The SOBI algorithm, which was introduced in [3], is a second-order separation method that relies on there being some level of spectral diversity among the underlying sources. The algorithm starts with pre-whitening the observation matrix $\mathbf{X}$ by finding a whitening transform $\mathbf{H}$ that whitens the signal part of $\mathbf{X}$:

$$E[\mathbf{HAP}(\mathbf{HAP})^H] = \mathbf{HAR_P}(0)(\mathbf{HA})^H = \mathbf{HA}(\mathbf{HA})^H = \mathbf{I} \tag{2.1}$$

where $\mathbf{R_P}$ is the source covariance matrix and $\mathbf{I}$ is the identity matrix. In order to find the whitening transform, $\mathbf{H}$, it is necessary that either the noise variances, σ_i^2's, be known or that an estimate of them can be made. In SOBI a solution is offered for the case where the noise variances across observations are equal in magnitude. With this assumption the whitening transform $\mathbf{H}$ is obtained through either an eigenvalue decomposition (EVD) of the sample covariance matrix of $\mathbf{X}$, $\widehat{\mathbf{R}}_X$, or of the covariance

matrix $\mathbf{R}_X$ if it is known a-priori:

$$\mathbf{H} = [(\lambda_1 - \widehat{\sigma}^2)^{(1/2)}\mathbf{h_1}, ..., (\lambda_k - \widehat{\sigma}^2)^{(1/2)}\mathbf{h_k}]^H \tag{2.2}$$

where $\widehat{\sigma}^2$ is the estimated variance of the noise, λ_i denotes the i^{th} largest eigenvalue, and $\mathbf{h}_i$ is the corresponding eigenvector of either $\widehat{\mathbf{R}}_X$ or $\mathbf{R}_X$. From (2.1), $\mathbf{U}$=$\mathbf{HA}$ can be seen to be a unitary matrix. The desired mixing transform $\mathbf{A}$ can then be expressed in terms of $\mathbf{H}$ and $\mathbf{U}$ as

$$\mathbf{A} = \mathbf{H}^{\#}\mathbf{U} \tag{2.3}$$

where # denotes the Moore-Penrose pseudoinverse. The last step of SOBI then is to find this unitary transform, $\mathbf{U}$, so that $\mathbf{A}$ may be obtained from (2.3). This is achieved first by noting that the time delayed covariance matrices of the whitened data, $\mathbf{HX}$, adhere to the following relation:

$$\mathbf{R_{HX}}(\tau) = \mathbf{HAR_P}(\tau)(\mathbf{HA})^H \qquad \forall\tau \neq 0 \tag{2.4}$$

$$\Rightarrow \mathbf{R_{HX}}(\tau) = \mathbf{UR_P}(\tau)\mathbf{U}^H \qquad \forall\tau \neq 0 \tag{2.5}$$

By the relation in (2.5) it is then possible to obtain the desired unitary factor, $\mathbf{U}$, as any unitary matrix that diagonalizes $\mathbf{R}_{HX}(\tau)$ for some time lag τ. As pointed out in [3] it follows from the spectral theorem for normal matrices[1] [9] that the existence of a unitary matrix $\mathbf{V}$ is guaranteed such that for a nonzero time lag τ,

$$\mathbf{V}^H\mathbf{R_{HX}}(\tau)\mathbf{V} = diag\{d_1, ..., d_k\} \tag{2.6}$$

[1]A normal matrix $\mathbf{M}$, i.e. $\mathbf{MM}^H$=$\mathbf{M}^H\mathbf{M}$, by the spectral theorem is unitarily diagonalizable, i.e. there exists a unitary matrix $\mathbf{U}$ and diagonal matrix $\mathbf{D}$ such that $\mathbf{M}$=$\mathbf{UDU}^H$.

However in order for $\mathbf{V}$ to be essentially equal [2] to $\mathbf{U}$, making $\mathbf{U}$ directly attainable from $\mathbf{V}$, the condition:

$$\rho_i(\tau) \neq \rho_j(\tau) \quad \forall i \neq j \tag{2.7}$$

where

$$\rho_i(\tau) = E[\mathbf{P}_i(t+\tau)\mathbf{P}_i^*(t)] \tag{2.8}$$

and $\mathbf{P}_i$ denotes the $i'th$ source, must hold, which is not true in general for an arbitrary nonzero time lag τ. The idea used in SOBI is to find a unitary matrix, $\mathbf{V}$, such that (2.6) holds for a set of nonzero time lags, $\{\tau_i \mid i = 1, ..., L\}$, rather than just one lag τ. Now $\mathbf{U}$ will be directly attainable from $\mathbf{V}$ so long as (2.7) holds for at least one τ_i. This decreases the probability of not being able to identify $\mathbf{U}$, and subsequently $\mathbf{A}$, as it will be less likely for (2.7) not to hold for all $\tau_i's$ in this set. An algorithm overview of SOBI appears in Table 2.1.

2.1.2 Limitations of SOBI

Commonly in BSS algorithms, as is the case with SOBI, it is computationally desirable to have a pre-whitening step that first finds a whitening transform, $\mathbf{H}$, that when applied to the mixing transform, $\mathbf{A}$, forms a unitary matrix, i.e.

$$(\mathbf{HA})(\mathbf{HA})^H = diag\{\lambda_1, .., \lambda_n\} \tag{2.9}$$

This allows the search space for the mixing transform to be reduced to that of unitary matrices. In order to accurately find such a transform, however, it is necessary to remove the noise bias present in the EVD of the data covariance matrix. One limitation of SOBI is that in order to remove this bias and obtain a whitening transform that satisfies (2.9) it assumes that the noise covariance matrix is either known a-priori or

[2] Two matrices $\mathbf{M}$ and $\mathbf{N}$ are essentially equal if there exists a matrix $\mathbf{P}$, such that $\mathbf{M}= \mathbf{PN}$, where $\mathbf{P}$ has exactly one nonzero entry in each row and column, where these entries have unit modulus [sobi].

Table 2.1: The SOBI Algorithm

Step	
1.	Calculate the sample covariance matrix, $\widehat{\mathbf{R}}_X(0)$, of $\mathbf{X} = \mathbf{AP} + \mathbf{GW}$ $\Rightarrow \widehat{\mathbf{R}}_X(0) \approx \mathbf{A}\mathbf{R}_P(0)\mathbf{A}^H + \mathbf{G}\mathbf{R}_W(0)\mathbf{G}^H$ $\approx \mathbf{A}\mathbf{A}^H + \mathbf{G}\mathbf{G}^H$ Let λ_1,...,λ_k denote the **k** largest eigenvalues and $\mathbf{h}_1$,...,$\mathbf{h}_k$ the corresponding eigenvectors of $\widehat{\mathbf{R}}_X(0)$
2.	Assuming white noise, an estimate $\widehat{\sigma}^2$ of the noise variance is the average of the **n - k** smallest eigenvalues of $\widehat{\mathbf{R}}_X(0)$. Calculate the whitening matrix, **H**, as $\mathbf{H} = [(\lambda_1 - \widehat{\sigma}^2)^{(1/2)}\mathbf{h_1}, ..., (\lambda_k - \widehat{\sigma}^2)^{(1/2)}\mathbf{h_k}]^H$
3.	Let the whitened data be denoted as $\mathbf{Z} = \mathbf{HX}$ Calculate the sample covariance matrices, $\widehat{\mathbf{R}}_Z(\tau)$, for a fixed set of time lags $\tau\epsilon\{\tau_i \mid i = 1, ..., L\}$
4.	Find **U** as the joint diagonalizer (see Appendix A.1) of the set $\{\widehat{\mathbf{R}}_Z(\tau_i) \mid$ $i = 1, ..., L\}$, recalling $\widehat{\mathbf{R}}_Z(\tau_j) \approx \mathbf{U}\mathbf{R}_P(\tau_i)\mathbf{U}^T$.
5.	Estimate **A** and **P**: $\widehat{\mathbf{A}} = \mathbf{W}^{\#}\mathbf{U}$ ($^{\#}$ indicates the Moore-Penrose pseudo-inverse) $\widehat{\mathbf{P}} = (\mathbf{U}^T\mathbf{H})\mathbf{X}$

that the noise variances are of approximately equal magnitude over the observations. In the case where the covariance matrix is unknown it is assumed that the noise covariance matrix is a scaled identity matrix. With this assumption an estimate is then made of the unknown scaling factor by averaging over the last few eigenvalues of the data covariance matrix, i.e.

$$\mathbf{G}\mathbf{G}^H = \sigma^2\mathbf{I} \tag{2.10}$$

where

$$\sigma^2 = \frac{\lambda_{n-k+1} + ... + \lambda_n}{n-k} \tag{2.11}$$

$$\mathbf{R_X}(0) = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H \tag{2.12}$$

with λ_i being the i^{th} smallest diagonal element of $\mathbf{\Lambda}$. If this equal-energy assumption does not hold, which is commonly the case, the whitening step will be biased by noise. More specifically the condition of $\mathbf{HA}$ being unitary will not generally hold for the calculated whitening transform $\mathbf{H}$. This in turn will lead to degraded separation performance for low SNR data as the exact new mixing transform, $\mathbf{U} = \mathbf{HA}$, will not be contained in the search space comprised of unitary matrices.

An additional limitation of SOBI is its restriction to the separation of stationary sources. If any sources happen to be non-stationary, which is common for real data sets, the performance of SOBI will suffer as sample statistics will lose accuracy from averaging over regions of the data set with different statistics.

2.2 SONS

2.2.1 The SONS algorithm

A more recently developed second-order separation algorithm, SONS, introduced in [4], attempts to address these limitations associated with SOBI. First the noise bias that is often present in the pre-whitening step of SOBI is addressed through a procedure they introduce called robust whitening. The zero-time-lag data covariance matrix is expressed as:

$$\mathbf{R_X}(0) = \mathbf{A}\mathbf{R_P}(0)\mathbf{A}^H + \mathbf{G}\mathbf{R_W}(0)\mathbf{G}^H \tag{2.13}$$

$$= \mathbf{A}\mathbf{A}^H + \mathbf{G}\mathbf{G}^H \tag{2.14}$$

The bias originates from the noise covariance term $\mathbf{GG}^H$. This term must be known or estimated in order for the signal part of the data to be accurately whitened. To avoid this bias, robust whitening uses a linear combination of time delayed data covariance matrices to obtain a whitening transform as opposed to using the zero-time-lag covariance matrix. This involves first starting with an arbitrary weight vector, α, for the coefficients of the sample covariance matrices used in the linear

combination, i.e.

$$\mathbf{R} = \alpha_1 \widehat{\mathbf{R}}_X(\tau_1) + ... + \alpha_J \widehat{\mathbf{R}}_X(\tau_J) \quad \tau_i \neq 0 \; \forall i \tag{2.15}$$

where

$$\widehat{\mathbf{R}}_X(\tau_i) \approx \mathbf{A}\mathbf{R}_\mathbf{P}(\tau_i)\mathbf{A}^H \tag{2.16}$$

follows from the white noise assumption. The desired whitening transform can then be found using the EVD of $\mathbf{R}$:

$$\mathbf{R} = \mathbf{V}_R \mathbf{\Lambda}_R \mathbf{V}_R^H \tag{2.17}$$

$$\Rightarrow \mathbf{H} = \mathbf{\Lambda}_R^{-\frac{1}{2}} \mathbf{V}_R^H \tag{2.18}$$

However because $\mathbf{R}$ is not necessarily positive definite, the whitening transform $\mathbf{H}$ may not be valid. Therefore to ensure that $\mathbf{R}$ is positive definite the next step of the procedure is to use the finite-step global convergence algorithm[3] [10] to adapt the initial weight vector in such a way that the resulting $\mathbf{R}$ is positive definite. Robust whitening helps performance by increasing the likelihood that $\mathbf{HA}$ will be approximately calculated as a unitary matrix.

Additionally SONS allows for the separation of not only spectrally diversified sources, but also second-order non-stationary sources. In SONS the data is partitioned such that sample covariance matrices are calculated separately over each partition. This new set of covariance matrices is then used in the same joint diagonalization step implemented in SOBI so that any change in source statistics over different partitions can be exploited. The SONS algorithm appears in Table 2.2.

[3]see A.2 for algorithm description

2.2.2 Limitations of SONS

One potential source of separation performance loss associated with SONS is in the robust whitening step. The performance of this procedure in whitening the data is sensitive to the initial weight vector chosen. Additionally in the case of white non-stationary data, the robust whitening procedure is not applicable because of the lack of time-correlations among the sources. For these data sets the separation performance will be degraded by any noise present, as is the case with SOBI.

Another source of sensitivity in performance involves how the data set is partitioned. If the data partitions in SONS are made arbitrarily as opposed to roughly matching the non-stationarity of the data, sample statistics will be poor causing performance in estimating the sources to suffer. This sensitivity is shown empirically in the tests conducted in Chapter 3 of this thesis.

Table 2.2: The SONS Algorithm

Step	
1.	Calculate the set of sample covariance matrices, $\{\widehat{\mathbf{R}}_X(\tau_i) \mid i = 1, ..., J\}$ where $\mathbf{X} = \mathbf{AP} + \mathbf{GW}$. Choose any non-zero initial vector of weights $\alpha = [\alpha_1, ..., \alpha_J]$. Use the global convergence algorithm to iteratively update α till $\mathbf{R} = \alpha_1\widehat{\mathbf{R}}_X(\tau_1) + ... + \alpha_J\widehat{\mathbf{R}}_X(\tau_J)$ is positive definite. Let λ_1,...,λ_k denote the $\mathbf{k}$ largest eigenvalues and $\mathbf{h}_1$,...,$\mathbf{h}_k$ the corresponding eigenvectors of $\mathbf{R}$
2.	Calculate the whitening matrix, $\mathbf{H}$, as $\mathbf{H} = [(\lambda_1)^{(1/2)}\mathbf{h_1}, ..., (\lambda_k)^{(1/2)}\mathbf{h_k}]^H$
3.	Let the whitened data be denoted as $\mathbf{Z} = \mathbf{HX}$ Divide $\mathbf{Z}$ into L non-overlapping blocks (time windows T_i) and calculate the sample covariance matrices $\widehat{\mathbf{R}}_Z(T_i, \tau_j)$ for i=1..L and j=1....M
4.	Find the unitary matrix $\mathbf{U}$ as the joint diagonalizer of the set $\{\widehat{\mathbf{R}}_Z(T_i, \tau_j)\})$ using the joint diagnalization method.
5.	Estimate $\mathbf{A}$ as: $\widehat{\mathbf{A}} = \mathbf{H}^{\#}\mathbf{U}$ ($^{\#}$ indicates the Moore-Penrose pseudo-inverse)

2.3 SOON

The SOON algorithm, presented in this thesis, seeks to improve upon SONS and SOBI by carefully addressing the limitations associated with those algorithms. In particular SOON works to improve separation under noisy conditions by iteratively estimating both the number of sources k and the noise covariance matrix $\mathbf{G}\mathbf{G}^H$. More specifically, SOON uses a scree metric to estimate the signal order, and then the EM algorithm to estimate the noise covariance matrix. This first step of SOON iterates between order and noise estimation to improve the estimates of each parameter. Using this method an estimate of the noise can be constructed and used to improve the SNR of the data matrix. Robust whitening is then performed on the noise adjusted data by using a heuristic proposed in this thesis for choosing the initial weight vector. The Joint Diagonalization method is then used to find the unknown unitary mixing matrix of the transformed data set.

2.3.1 Order Estimation

The quality of performance of any separation algorithm in obtaining the desired source signals depends on the knowledge of the number of sources within the data. Most algorithms either assume that this number is known or make restrictive assumptions with respect to the noise so that an estimate of this number can be easily obtained. This is true for SOBI and SONS. For the case in which the noise covariance matrix $\mathbf{G}\mathbf{G}^H$ is an identity matrix multiplied by some scalar and this covariance matrix is known, the problem of estimating the number of underlying sources is trivial. The number of sources in this case is simply the multiplicity of the smallest eigenvalue of the data covariance matrix:

$$\widehat{k} = \rho([\lambda_1, .., \lambda_n], \lambda_{min}) \tag{2.19}$$

where $\widehat{k}$ is the source order estimate, λ_i is the i^{th} eigenvalue of $\mathbf{R_X}$, λ_{min} is the smallest such eigenvalue, and $\rho(\mathbf{a}, b)$ denotes the number of elements in vector $\mathbf{a}$ that

have value b. What makes estimation of the number of sources difficult in practice is that the noise covariance is often not a scaled identity matrix as in equation (2.10). Additionally the exact data covariance matrix, $\mathbf{R_X}$, is not usually known so that the order estimate must be calculated using the finite sample covariance matrix, $\widehat{\mathbf{R}}_\mathbf{X}$. The eigenvalues of finite sample covariance matrices are guaranteed to be distinct so that using Equation (2.11) will not work in practice regardless of the structure of $\mathbf{R_X}$. Information theoretic approaches to this problem have been studied in [11],[12] and other methods in [13],[14]. In each, however, it has been assumed that either $\mathbf{GG}^H$ is a scaled identity matrix as in (2.10), or that $\mathbf{GG}^H$ is known a priori. To be robust in this estimation, so that we may use the less restrictive noise model of this thesis, we adopt a method outlined in the ION algorithm [16] which utilizes a scree plot. The scree plot used here displays the log magnitude-ordered eigenvalues of the sample correlation matrix for the data matrix $\mathbf{X}$ as a function of eigenvector number. In Figure 2-1, we see a typical scree plot, where the plateau to the right of the break generally represents pure noise components. Here the true order of the underlying process is 20, while the data matrix $\mathbf{X}$ is of rank 55. Multivariate data with additive noise yields eigenvalues representing the noise plus signal energy. Therefore with noise-normalized signals, the eigenvalues corresponding to pure noise have approximately equal amplitudes and form a plateau of length n-k. The estimated number of sources is the number of eigenvalues that lie above the extrapolated plateau. In iterations where the distribution of eigenvalues is such that an accurate estimate cannot be made, SVD is used by estimating the number of signals as the number of singular values that lie above some pre-determined threshold. In general however, the use of a scree plot allows for a more robust order estimation to be made as opposed to solely using SVD.

2.3.2 Noise Estimation

The ability of a separation algorithm to effectively remove or reduce the noise signals will ultimately determine the accuracy of the source signal estimation of that algorithm. Models that assume the noise covariance to be a scaled identity matrix

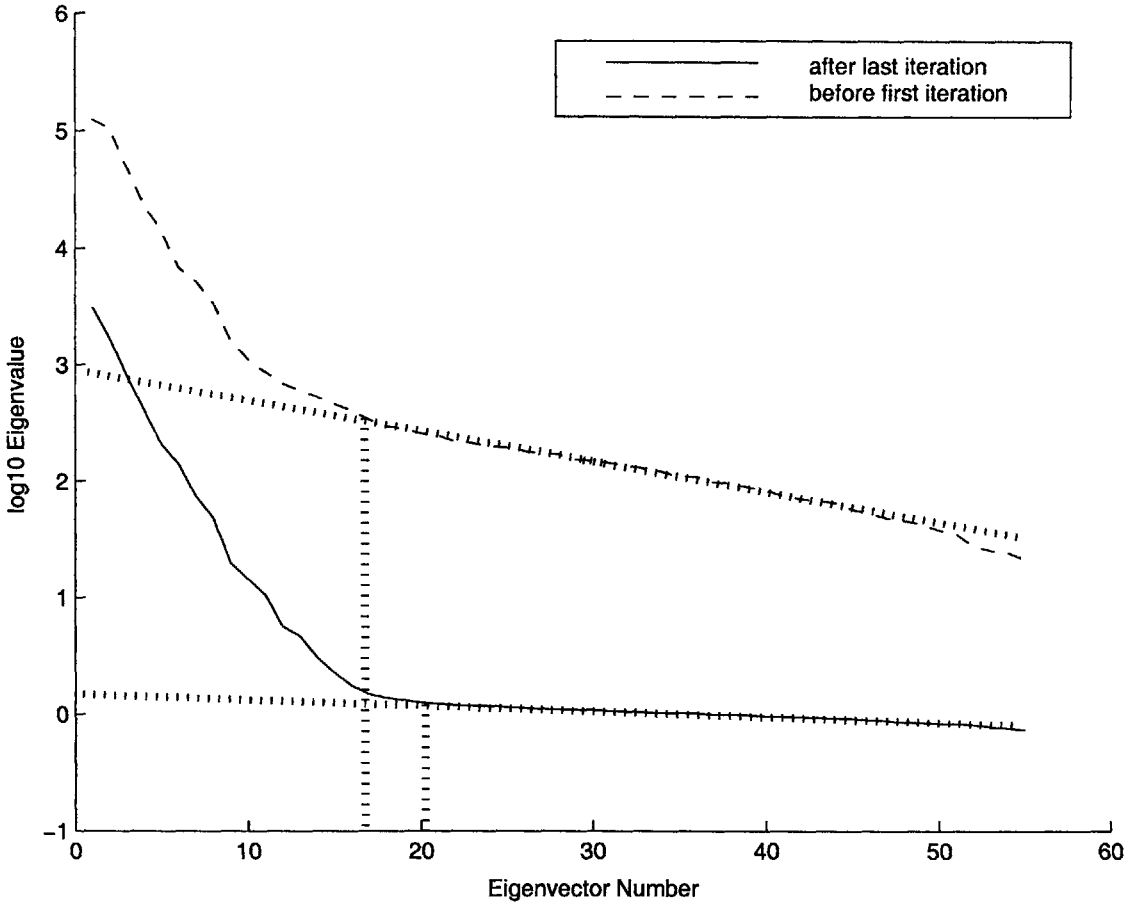

Figure 2-1: Typical Scree Plot which plots the ordered log-magnitude eigenvalues of data matrix $\mathbf{X}$. Here the number of underlying signals, k, is 20 while the number of observations, n, is 55. The horizontally orientated dotted lines represent the line fitting the plateaus of the scree plots and the vertical dotted lines represent the intersection of the two curves in which the order estimation is obtained.

are restrictive as many data sets do not have equal variance across sensor outputs. To account for this possibility we use a variant of the EM method to obtain a better estimate of these variances, namely the parameter $\mathbf{G}$, than is obtained in equation (2.10). This allows for a noise estimate to be made that can then be used to increase the SNR of the data, which improves separation performance.

The Expectation Maximization (EM) algorithm is a two-step iterative technique for finding the Maximum Likelihood (ML) estimate for an unknown parameter of interest. The problem generally involves finding a value for this unknown parameter that maximizes the conditional probability of some observed and unknown stochastic

parameters given the unknown parameter of interest, i.e. find:

$$\psi_{ML} = \max_{\psi} f(\Phi_{known}, \Phi_{unknown}|\psi) \tag{2.20}$$

where ψ is the unknown stochastic parameter of interest, Φ_{known} is a set of known parameters that depend on ψ, and $\Phi_{unknown}$ is the corresponding set of unknown parameters. In order to find the value of ψ that maximizes (2.20) it is necessary to obtain the parameters contained in $\Phi_{unknown}$. However, these parameters are not directly available. The idea then incorporated in the EM algorithm is to replace $\Phi_{unknown}$ with:

$$E[\Phi_{unknown}|\psi_{ML}, \Phi_{known}] \tag{2.21}$$

The ML estimate of ψ using the EM algorithm is then obtained by iteratively obtaining the expression in (2.21) and maximizing (2.20) by substituting (2.21) for $\Phi_{unknown}$.

In the ION algorithm [16], a variant of the EM method was derived to estimate the noise variances in the case where the observation matrix, $\mathbf{X}$ is temporally white, i.e.

$$E[X_i X_j^H] = \delta_{i-j}\mathbf{I} \tag{2.22}$$

where X_i represents the i^{th} column of $\mathbf{X}$. In the problem we are considering here though this is generally not the case, so we derive a variant of the EM algorithm for estimating these variances for data sets that may or may not contain time correlations. In (1.12) $\mathbf{A}$ and $\mathbf{G}$ are unknown but fixed parameters, whereas $\mathbf{P}$ and $\mathbf{X}$ are stochastic matrices. In the framework of the EM algorithm then we have,

$$\psi = \{\mathbf{A}, \mathbf{G}\} \tag{2.23}$$

$$\Phi_{known} = \{\mathbf{X}\} \tag{2.24}$$

$$\Phi_{unknown} = \{\mathbf{P}\} \tag{2.25}$$

The maximum likelihood estimates of $\mathbf{A}$ and $\mathbf{G}$ then are,

$$\{\mathbf{A}_{ML}, \mathbf{G}_{ML}\} = \max_{\mathbf{A},\mathbf{G}} f(x_1, .., x_m, p_1, .., p_m | \mathbf{A}, \mathbf{G}) \tag{2.26}$$

where

$$\mathbf{X} = [\ x_1 \quad x_2 \quad .... \quad x_m\ ] \qquad x_i \epsilon \mathbb{C}^{nx1} \tag{2.27}$$

$$\mathbf{P} = [\ p_1 \quad p_2 \quad .... \quad p_m\ ] \qquad p_i \epsilon \mathbb{C}^{kx1} \tag{2.28}$$

Using Bayes' rule, (2.26) can be written as

$$f(p_1, .., p_m | \mathbf{A}, \mathbf{G}) f(x_1, .., x_m | p_1.., p_m, \mathbf{A}, \mathbf{G}) \tag{2.29}$$

Here $\mathbf{P}$ does not depend on $\{\mathbf{A},\mathbf{G}\}$, so the term $f(p_1, .., p_m | \mathbf{A}, \mathbf{G})$ can be dropped from (2.29) as it does not effect the maximization. Additionally we have the following relation

$$x_i = \mathbf{A}p_i + \mathbf{G}w_i \quad \forall i \tag{2.30}$$

where w_i is the i^{th} time sample of the stochastic matrix $\mathbf{W}$. Therefore given $\mathbf{P}$, $\mathbf{A}$, and $\mathbf{G}$, the x_i's are independent as they depend stochastically only on the w_i's which are assumed independent. Thus (2.26) can be written as

$$\max_{\mathbf{A},\mathbf{G}} f(x_1 | p_1, \mathbf{A}, \mathbf{G}) ... f(x_m | p_m, \mathbf{A}, \mathbf{G}) \tag{2.31}$$

$$= \max_{\mathbf{A},\mathbf{G}} \prod_{i=1}^{m} \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{GG})^H}} \exp^{-B_1} \tag{2.32}$$

where

$$B_1 = (x_i - \mathbf{A}p_i)^H(\mathbf{G}\mathbf{G}^H)^{-1}(x_i - \mathbf{A}p_i)/2 \tag{2.33}$$

After taking the log of (2.32) and changing the sum to be over the rows of $\mathbf{X}$ as opposed to the columns we arrive with the final expression for the ML estimate of $\mathbf{A}$ and $\mathbf{G}$:

$$\{\mathbf{A}_{ML}, \mathbf{G}_{ML}\} = \min_{\mathbf{A},\mathbf{G}}(\sum_{i=1}^{n} mlog(G_i) + \frac{1}{G_i}B_2) \tag{2.34}$$

where

$$B_2 = \mathbf{X}_i X_i^H - 2X_i\mathbf{P}^H A_i^H + A_i\mathbf{P}\mathbf{P}^H A_i^H \tag{2.35}$$

A_i is the i^{th} row of $\mathbf{A}$, X_i is the i^{th} row of $\mathbf{X}$, and G_i is the i_{th} diagonal element of $\mathbf{G}\mathbf{G}^H$. However, the terms $\mathbf{P}^H$ and $\mathbf{P}\mathbf{P}^H$ in (2.35) are not known, so that (2.24) can not be directly minimized. So the EM algorithm proposes that we replace these terms with their expectations given $\mathbf{X}$ and our current estimates of $\mathbf{A}$ and $\mathbf{G}$ in what is referred to as the expectation step of the algorithm. To obtain these expectations as in [16] we can first solve for the distribution $f(p_i|x_i, \mathbf{A}, \mathbf{G})$, which by Bayes' rule can be expressed as:

$$f(p_i|x_i, \mathbf{A}, \mathbf{G}) = \frac{f(x_i|p_i, \mathbf{A}, \mathbf{G})f(p_i|\mathbf{A}, \mathbf{G})}{f(x_i|\mathbf{A}\mathbf{G})} \tag{2.36}$$

Noting that x_i, p_i, and w_i are all jointly gaussian, we can solve for (2.36) as

$$f(p_i|x_i, \mathbf{A}, \mathbf{G}) = N(\mathbf{S}^{-1}\mathbf{A}^H(\mathbf{G}\mathbf{G}^H)^{-1}x_i, \mathbf{S}^{-1}) \tag{2.37}$$

where

$$\mathbf{S} = \mathbf{A}^H(\mathbf{G}\mathbf{G}^H)^{-1}\mathbf{A} + \mathbf{I} \tag{2.38}$$

which implies

$$E[\mathbf{P^H}|\mathbf{X},\mathbf{A},\mathbf{G}] = \mathbf{X}^H(\mathbf{G}\mathbf{G}^H)^{-1}\mathbf{A}\mathbf{S}^{-1} \tag{2.39}$$

Additionally we have the following relation,

$$E[p_i^H p_i] = \mathbf{R}_{p_i^H|x_i,\mathbf{A},\mathbf{G}} + E[p_i^H|x_i,\mathbf{A},\mathbf{G}]E[p_i^H|x_i,\mathbf{A},\mathbf{G}]^H \tag{2.40}$$

where $\mathbf{R}_{p_i^H|x_i,\mathbf{A},\mathbf{G}}$ denotes the covariance of p_i given $\mathbf{X}$, $\mathbf{A}$, and $\mathbf{G}$. From (2.36) this can be expressed as:

$$\mathbf{R}_{p_i^H|x_i,\mathbf{A},\mathbf{G}} = \mathbf{S}^{-1} \tag{2.41}$$

This then implies that the second unknown quantity, $E[\mathbf{P}\mathbf{P}^H|\mathbf{X},\mathbf{A},\mathbf{G}]$, can be expressed as,

$$E[\mathbf{P}\mathbf{P}^H|\mathbf{X},\mathbf{A},\mathbf{G}] = m\mathbf{S}^{-1} + \mathbf{S}^{-1}\mathbf{A}^H(\mathbf{G}\mathbf{G}^H)^{-1}\mathbf{X}\mathbf{X}^H(\mathbf{G}\mathbf{G}^H)^{-1}\mathbf{A}\mathbf{S}^{-1} \tag{2.42}$$

which completes the expectation step of the algorithm. The maximization step of the algorithm then requires optimizing over (2.34) by substituting for the unknown quantities their corresponding estimated values found in (2.39) and (2.42). An analytical solution for the unknown parameters is obtainable by taking the partial derivative of (2.34) with respect to $\mathbf{A}$ and $\mathbf{G}$. The derivation for these maximization equations is independent of the spectral characteristics of the data. Therefore this derivation is identical to that found in [16] and can be referenced there. The maximization equations along with the entire iterative EM algorithm derived for use in SOON can be seen in Table 2.3.

2.3.3 Sensitivity of Robust Whitening

The robust whitening procedure utilized by the SONS algorithm, requires first choosing an initial set of weights for the set of covariance matrices used in the whitening

Table 2.3: The Expectation-Maximization Algorithm

Step	
1.	Initialization: Iteration index $j = 1$; number of iterations j_{max} $\mathbf{GG}_1^H = 0.5\mathbf{I}$ $\mathbf{A}_1 = \{$ unit-variance n x $\widehat{k}_i$ array of g.r.n. $\}$
2.	Expectation Step: $\mathbf{S}_j = \mathbf{A}_j^H(\mathbf{GG}_j^H)^{-1}\mathbf{A}_j + \mathbf{I}$ $\mathbf{C}_j = E[\mathbf{P^H}\|\mathbf{X}, \mathbf{A}_j, \mathbf{G}_j] = \mathbf{X}^H(\mathbf{GG}_j^H)^{-1}\mathbf{A}_j\mathbf{S}_j^{-1}$ $\mathbf{D}_j = E[\mathbf{PP}^H\|\mathbf{X}, \mathbf{A}_j, \mathbf{G}_j]$ $= m\mathbf{S}_j^{-1} + \mathbf{S}_j^{-1}\mathbf{A}_j^H(\mathbf{GG}_j^H)^{-1}\mathbf{XX}^H(\mathbf{GG}_j^H)^{-1}\mathbf{A}_j\mathbf{S}_j^{-1}$
3.	Maximization Step: $\mathbf{A}_{j+1} = \mathbf{XC}_j\mathbf{D}_j^{-1}$ $G_{q,j+1} = ((\mathbf{X}^H)_q^H(\mathbf{X}^H)_q - \mathbf{A}_{q,j+1}\mathbf{C}_j^H(\mathbf{X}^H)_q)/m;$ where q= 1, ..., n indexes the columns of $\mathbf{X}^T$ and $\mathbf{A}$; G_q is the q^{th} diagonal element of $\mathbf{GG}^H$,
4.	If $j < j_{max}$, increment j and go to Step 2 for further iteration.
5.	After last iteration, output $\mathbf{G}_{i+1}$ and $\widehat{\mathbf{X}}_{i+1} = [\mathbf{X}^H(\mathbf{GG}_{j+1}^H)^{-1}\mathbf{A}_{j+1}\mathbf{S}_{j+1}^{-1}]\mathbf{A}_{j+1}^H$

process. These matrices are the delayed sample covariance matrices, $\widehat{\mathbf{R}}_X(\tau)$, which are calculated using a finite sample size with upper bound m. Here the $(i,j)^{th}$ entry for $\widehat{R}_X(\tau)$ is calculated as:

$$\widehat{\mathbf{R}}_X(\tau)_{i,j} = \frac{x_{i1}x_{j(\tau+1)} + \ldots + x_{i(m-\tau)}x_{jm}}{m-\tau} \tag{2.43}$$

When making use of delayed sample covariance matrices of the form, $\widehat{\mathbf{R}}_X(\tau)$, we would like to use those that have high signal-to-noise ratio (SNR) so that a whitening matrix will be found that accurately whitens the signal portion of the data. Given the white noise assumption of the model used in this thesis, the noise intrinsic to the data should be roughly negligible in the calculated time-delayed covariance matrices. The noise we are concerned about then in this context is the sample noise from calculating $\widehat{\mathbf{R}}_X(\tau)$ from a finite sample size. The SNR of a given calculated sample covariance

matrix can be described as:

$$SNR\{\widehat{\mathbf{R}}_X(\tau)\} \equiv \frac{E_\tau}{\Psi_\tau} \tag{2.44}$$

where,

$$E_\tau \equiv \Sigma_{i,j}|\widehat{\mathbf{R}}_X(\tau)_{ij}| \tag{2.45}$$

$$\Psi_\tau \equiv \Sigma_{i,j}E[\varepsilon_{ij}] \tag{2.46}$$

The expected sample error, ε_{ij}, incurred by approximating $\mathbf{R}_X(\tau)$ by $\widehat{\mathbf{R}}_X(\tau)$ can then be calculated as the standard deviation of $\widehat{\mathbf{R}}_X(\tau)_{ij}$ in (2.43):

$$E[\varepsilon_{ij}] = \sqrt{Var[\frac{x_{i1}x_{j(\tau+1)} + \ldots + x_{i(m-\tau)}x_{jm}}{m-\tau}]} \tag{2.47}$$

$$= \sqrt{E[(\frac{x_{i1}x_{j(\tau+1)} + \ldots + x_{i(m-\tau)}x_{jm}}{m-\tau})^2] - E[\frac{x_{i1}x_{j(\tau+1)} + \ldots + x_{i(m-\tau)}x_{jm}}{m-\tau}]^2} \tag{2.48}$$

$$= O(\sqrt{\frac{E[\widehat{\mathbf{R}}_X(\tau)_{ij}]}{m-\tau}}) \tag{2.49}$$

where (2.49) follows from a gaussian assumption on $\mathbf{X}$, which is consistent with the second-order theme of this thesis. Therefore the SNR of $\widehat{\mathbf{R}}_X(\tau)\}$ grows proportional to the square root of its energy, i.e.,

$$SNR\{\widehat{\mathbf{R}}_X(\tau)\} \sim sqrt[E_\tau] = \sum_{i,j}\sqrt{|\widehat{\mathbf{R}}_X(\tau)_{ij}|} \tag{2.50}$$

A good strategy then is to weight the delayed covariance matrices proportional to the square root of their energy in the initial linear combination rather than arbitrarily choosing these weights. Here we adopt this strategy in SOON by first calculating the

set of covariance matrices and then assigning the initial weight vector α accordingly. Namely, calculate the set of sample covariance matrices,

$$\{\widehat{\mathbf{R}}_X(\tau_i) \mid i = 1, ..., J\} \tag{2.51}$$

then calculate the energy of $\widehat{\mathbf{R}}_X(\tau_i)$ as

$$E_i \equiv \sum_{j,k} (\widehat{\mathbf{R}}_X(\tau_i)_{jk})(\widehat{\mathbf{R}}_X(\tau_i)_{jk})^* \quad \forall i \tag{2.52}$$

finally initializing the weight vector α as

$$\alpha = \{\sqrt{E_1}, \sqrt{E_2}, ..., \sqrt{E_J}\} \tag{2.53}$$

This heuristic as compared to arbitrary initialization increases the probability the calculated whitening transform $\mathbf{H}$ has high SNR.

2.3.4 Data Partitioning

An important contribution of the SONS algorithm was to add the capability of using the potential second-order non-stationarity of the underlying source signals as a means for separating them. The heuristic suggested by SONS to accomplish this is to partition the whitened data set, $\mathbf{HX}$, into L equally sized segments,

$$\mathbf{HX} = \{\mathbf{HX}_1, \mathbf{HX}_2, ..., \mathbf{HX}_L\} \tag{2.54}$$

With this strategy, sets of time delayed covariance matrices can be calculated over each segment separately so that any change in source statistics between segments can be taken advantage of, i.e. calculate the set:

$$\{\mathbf{R}_{HX_i}(\tau_j)\} \qquad 1 \leq i \leq L, \quad 1 \leq j \leq M \tag{2.55}$$

The technique used here is once again the joint diagonalization method which states that there exists a unitary transform, $\mathbf{V}$, that is essentially equal to the desired unknown unitary mixing factor, $\mathbf{U}$, that jointly diagonalizes the set of delayed covariance matrices from each data partition:

$$\mathbf{V}^H \mathbf{R}_{HX_i}(\tau)\mathbf{V} = diag\{d_1, ..., d_n\} \tag{2.56}$$

The ability of this method to accurately separate source signals though depends on the covariance matrices, $\mathbf{R}_{HX_i}(\tau)$, being good representations of the stochastic relationships in the data. In order for this to be true the source signals must be approximately stationary within each data partition. In cases where the source signal statistics vary continuously, i.e.,

$$\sigma^2_{\mathbf{P}_i}(t) = f(t) \tag{2.57}$$

where $\sigma^2_{\mathbf{P}_i}(t)$ is the variance of the i^{th} source at time t and f(t) is a continuous function in time, there exists a tradeoff in that the windows over which sample covariance matrices are calculated need to be large enough that the statistics are accurate but small enough that the second order statistics are fairly constant over the window. Additionally when statistics change at discrete points in time, arbitrary data segmentation can often lead to an averaging effect that also degrades the statistical validity of the computed covariance matrices. In order to gain insight into which partitioning scheme is appropriate it is necessary to approximately characterize the non-stationarity of the data set. The method applicable for this characterization depends on the type of data being analyzed. For time series data many heuristics exist for partitioning the data into stationary segments [18]. An algorithm that computes exact partitioning of the data into stationary segments scales[4]as $O(2^{T-1})$, where T is the number of

[4]Consider a time series with T samples, then there exists T-1 positions between samples in which a partition can be placed. Since in each position a partition can either be placed or not placed the total number of ways to partition the time series is 2^{T-1}. Hence the problem of finding which of those 2^{T-1} possibilities is an optimal partitioning of the time series into stationary segments will scale as $O(2^{T-1})$

data samples. This makes exact partitioning impractical, causing most heuristics to seek approximate solutions to the segmentation problem. For other data sets, such as images found in remote sensing, the process of finding a proper partitioning scheme can be as simple as manually inspecting the image for visibly distinct regions, or using standard classification techniques to define the regions. The purpose of this section is to note that for any given data set, $\mathbf{X}$, there is likely a better heuristic for partitioning $\mathbf{X}$ than simply dividing it into an arbitrary number of equally sized partitions. The heuristic chosen will greatly depend on any a-priori information regarding the non-stationarity of the data set.

2.3.5 Algorithm Summary

The SOON algorithm combines the order estimation, noise estimation, joint diagonalization method, and improved robust whitening outlined in this chapter in order to produce a robust second-order separation algorithm. A complete algorithm description of SOON can be seen in Table 2.4.

Table 2.4: The SOON algorithm

Step	
1.	Optionally normalize rows of $\mathbf{X}$ to zero mean and unit variance. Initialize $\widehat{\mathbf{G}}_0 = \mathbf{I}$.
2.	Noise-normalize $\mathbf{X}$: $\mathbf{X}_i = \widehat{\mathbf{G}}_{i-1}^{-1}\mathbf{X}$
3.	Estimate signal order k_i using a scree plot of $\mathbf{X}_i$ and SVD
4.	Estimate $\mathbf{G}_i\mathbf{G}_i^H$ and noise reduced data, $\widehat{\mathbf{X}}_i$ using the EM algorithm and k_i
5.	Check for iteration termination conditions; if none, increment the index i and return to Step 2.
6.	Calculate the set $\{\widehat{R}_{\widehat{X}}(\tau_i)\}$ for i=1...L. Initialize whitening vector α according to energy of this set. Use the global convergence algorithm to iteratively update α till $\mathbf{R} = \alpha_1\widehat{\mathbf{R}}_X(\tau_1) + ... + \alpha_J\widehat{\mathbf{R}}_X(\tau_J)$ is positive definite. Let λ_1,...,λ_k denote the $\mathbf{k}$ largest eigenvalues and $\mathbf{h_1}$,...,$\mathbf{h}_k$ the corresponding eigenvectors of $\mathbf{R}$
7.	Calculate the whitening matrix, $\mathbf{H}$, as $\mathbf{H} = [(\lambda_1)^{(1/2)}\mathbf{h_1}, ..., (\lambda_k)^{(1/2)}\mathbf{h_k}]^H$
8.	Characterize the non-stationarity of $\widehat{X}$ and construct partitioning set $\{T_i\}$ for i=1...J according to the proper heuristic.
9.	Calculate the sample covariance matrices $\widehat{\mathbf{R}}_Z(T_i, \tau_j)$ for i=1..L and j=1....M
10.	Find the unitary matrix $\mathbf{U}$ as the joint diagonalizer of the set $\{\widehat{\mathbf{R}}_Z(T_i, \tau_j)\}$) using the joint diagnalization method.
11.	Estimate $\mathbf{A}$ as: $\widehat{\mathbf{A}} = \mathbf{H}^{\#}\mathbf{U}$ ($^{\#}$ indicates the Moore-Penrose pseudo-inverse)

Chapter 3

Evaluation of SOON

Now that SOON has been introduced and described relative to the standard second-order algorithms SOBI and SONS, it is desirable to see how they compare in terms of some measure of performance over typical data sets in this domain.

3.1 Problem Space

In order to come up with typical data sets it is necessary to outline the important degrees of freedom in the problem space to which the separation performance is particularly sensitive. Data sets can be generated that have these degrees of freedom fixed at what would be considered typical values.

3.1.1 Source-to-Observation Ratio $\frac{k}{n}$

One degree of freedom to which separation performance is sensitive is the ratio between the number of source signals k and the number of sensor output observation signals, n. Clearly for the problem to be well posed, this ratio, $\frac{k}{n}$, must be less than or equal to 1. Additionally for a fixed number of sources if the number of noisy observations is increased the separation performance should be helped as new information is added. Therefore we would expect the performance of a separation algorithm to increase as $\frac{k}{n} \rightarrow 0$, and decrease as $\frac{k}{n} \rightarrow 1$. In fact, the scree-plot method for estimating

source order begins to fail when k $\gtrsim \frac{n}{3}$ [16].

3.1.2 Data Signal-to-Noise Ratio (SNR_X)

Another degree of freedom that effects separation performance is the SNR of the observation matrix **X**,

$$SNR_{\mathbf{X}} = \frac{E_{\mathbf{AP}}}{E_{\mathbf{GW}}} \tag{3.1}$$

where,

$$E_{\mathbf{Z}} = \sum_{i,j} |\mathbf{Z}_{ij}| \tag{3.2}$$

We should expect separation performance to be better for high SNR data as the noise will be less likely to bias our estimate of a whitening matrix and the number of source signals.

3.1.3 Sample Size (m)

In addition the sample size, m, of the data set will affect how well the signals can be separated. Given that the performance of all three algorithms being compared depends on estimating the probabilistic structure of the data through sampling statistics, the performance of this process will be degraded for smaller sample sizes.

3.1.4 Angle between Columns of Mixing Transform ($\angle_A$)

The i^{th} column of the mixing transform, **A**, projects the i^{th} source signal into the observation space. Clearly the more orthogonal these columns are the more orthogonal the contributions of each signal will be in this space and the system will be less mixed and more easily separable. This degree of freedom is defined as,

$$\angle_{\mathbf{A}} = \frac{\mathbf{A}_i^H \mathbf{A}_j}{\|\mathbf{A}_i\| \|\mathbf{A}_j\|} \tag{3.3}$$

3.2 Performance Metric

In order to compare the performance of these different algorithms we must first define a performance metric. A clear choice for this is how well each estimates the mixing transform $\mathbf{A}$, as this will determine how accurately each method separates the source signals. The metric we use here for gauging the quality of this estimate is an inverse residual metric,

$$\widehat{r}_{\mathbf{A}} \equiv \frac{Tr\{\mathbf{A}_S\}}{Tr\{\mathbf{A}_N\}} \tag{3.4}$$

where

$$\mathbf{A}_S = E[\mathbf{A}\mathbf{A}^H] \tag{3.5}$$

$$\mathbf{A}_N = E[(\mathbf{A} - \widehat{\mathbf{A}})(\mathbf{A} - \widehat{\mathbf{A}})^H] \tag{3.6}$$

and $Tr\{\bullet\}$ is the standard trace operator on square matrices. Here a larger value of $\widehat{r}_{\mathbf{A}}$ indicates a better estimate of $\mathbf{A}$.

3.3 SOON/SOBI/SONS Comparison

The two potential characteristics of a data set that second-order separation methods utilize for separation purposes are second-order non-stationarity and spectral diversity. Use of these algorithms is therefore restricted to data sets that possess at least one of these characteristics. For the purpose of comparing the performance of SOON to that of SOBI and SONS, two test data sets were constructed, one for testing performance over each characteristic. In both data sets the number of observations and sources were fixed at 55 and 20 respectfully. The mixing matrix, $\mathbf{A}$, was constructed so that its columns had 45 degrees of separation between them. In addition the

columns of $\mathbf{A}$ were normalized,

$$\mathbf{A}_i^H \mathbf{A}_i = \mathbf{A}_j^H \mathbf{A}_j \qquad \forall i, j \tag{3.7}$$

so that no single source would dominate the performance of the algorithms. Additionally the noise variances were independently and randomly chosen from a normalized exponential distribution, i.e.

$$f_{\mathbf{G}_{ii}}(x) = e^{-x} \quad x \geq 0 \tag{3.8}$$

where $f_{\mathbf{G}_{ii}}$ represents the distribution function of the noise variance for the i^{th} observation. The performance for each test was calculated over varying SNR of the observation matrix $\mathbf{X}$. In each test both SOBI and SONS were given the correct number of sources whereas SOON was not.

3.3.1 Test 1: Stationary-Colored Sources

In the first data set we illustrate the performance of SOON on stationary time correlated gaussian sources. Here the underlying signals are modelled using a standard Auto-Regressive (AR) model:

$$\mathbf{P}_{i,j} = \rho_i \mathbf{P}_{i,j-1} + \epsilon_{i,j} \tag{3.9}$$

where $\mathbf{P}_{ij}$ is the j^{th} time sample of the i^{th} source signal, ρ_i is the correlation coefficient of the i_{th} source, and $\epsilon_{i,j}$ is a random gaussian number. The $\rho_i's$ where chosen independently and uniformly in (0,1). Using this test set we compare the performance of SOON to that of SOBI and four variants of the SONS algorithm. The variants of SONS vary in two degrees of freedom, namely how the initial weight vector is chosen in the whitening step and how the data is partitioned for the calculation of covariance matrices. The initial weight vector was either chosen proportional to the square root of energy as described in Chapter 2, or equally weighted. The data was then either not partitioned, which is optimal for this stationary case, or divided into ten segments

of equal size. The variants of SONS are outlined in Table 3.1 and the results for the first test can be seen in Figure 3-1.

Table 3.1: SONS variants

Variant Number	Initial Weight Vector	Partitioning Scheme
1	proportional to energy	1 segment
2	proportional to energy	10 equal size segments
3	equally weighted	1 segment
4	equally weighted	10 equal size segments

3.3.2 Test 2: Non-stationary-White Sources

With the second test set we tested the performance of SOON on non-stationary white gaussian sources. Here the underlying sources are white gaussian signals with time varying second moments. In particular the second moments of each source vary independently of one another at discrete points in time that are distributed as a Poisson process, i.e.

$$\mathbf{P}_{i,j} = \rho_{i,t}\epsilon_{i,j} \tag{3.10}$$

where $\mathbf{P}_{i,j}$ is the jth time sample of the ith source, $\epsilon_{i,j}$ is a normalized random gaussian number, and

$$\rho_{i,t} = \sigma_{ij} \quad t_{ij} \leq t < t_{i(j+1)} \tag{3.11}$$

$$f_{\sigma_{ij}} = e^{-x} \quad x \geq 0 \tag{3.12}$$

$$f_{t_{ij}}(x) = \frac{\lambda^j x^{j-1} e^{-\lambda x}}{(j-1)!} \quad x \geq 0 \tag{3.13}$$

Here we have set λ to be $\frac{10}{T}$, where T is the number of time samples in the data set, so that we should expect the sources to on average be characterized as 10 equally sized

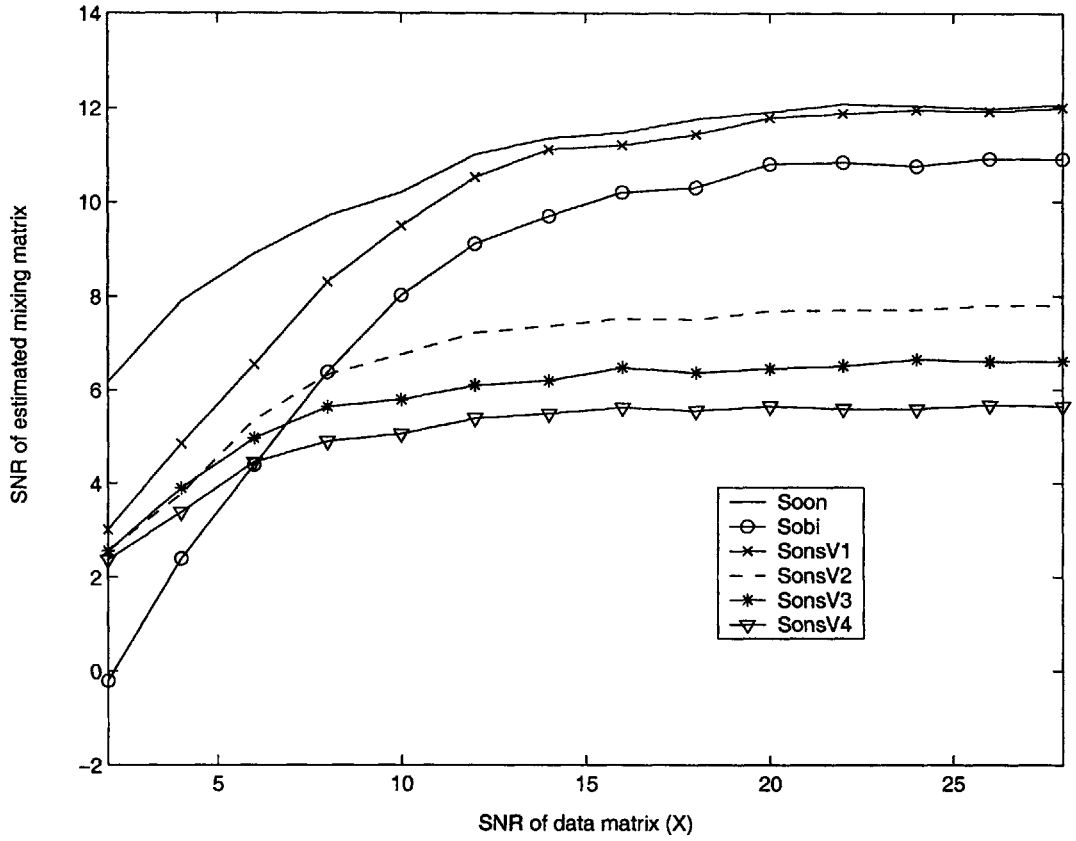

Figure 3-1: Test 1, which compares the performance of SOON to that of SOBI and four variants of SONS. The data set here is comprised of stationary sources which have diverse correlation functions.

segments with different second moments. Here robust whitening is not applicable since there exists no time correlation in the underlying sources. We therefore test SOON with SOBI and two variants of the SONS algorithm that each use traditional data whitening. In one variant the partitioning is done such that the non-stationarity of the data set is matched exactly whereas in the other the data set is partitioned into 5 equal sized partitions. The two variants are outlined in Table 3.2 and the results for this test can be seen in Figure 3-2.

Table 3.2: SONS variants

Variant Number	Partitioning Scheme
1	matches non-stationarity
2	5 equal size segments

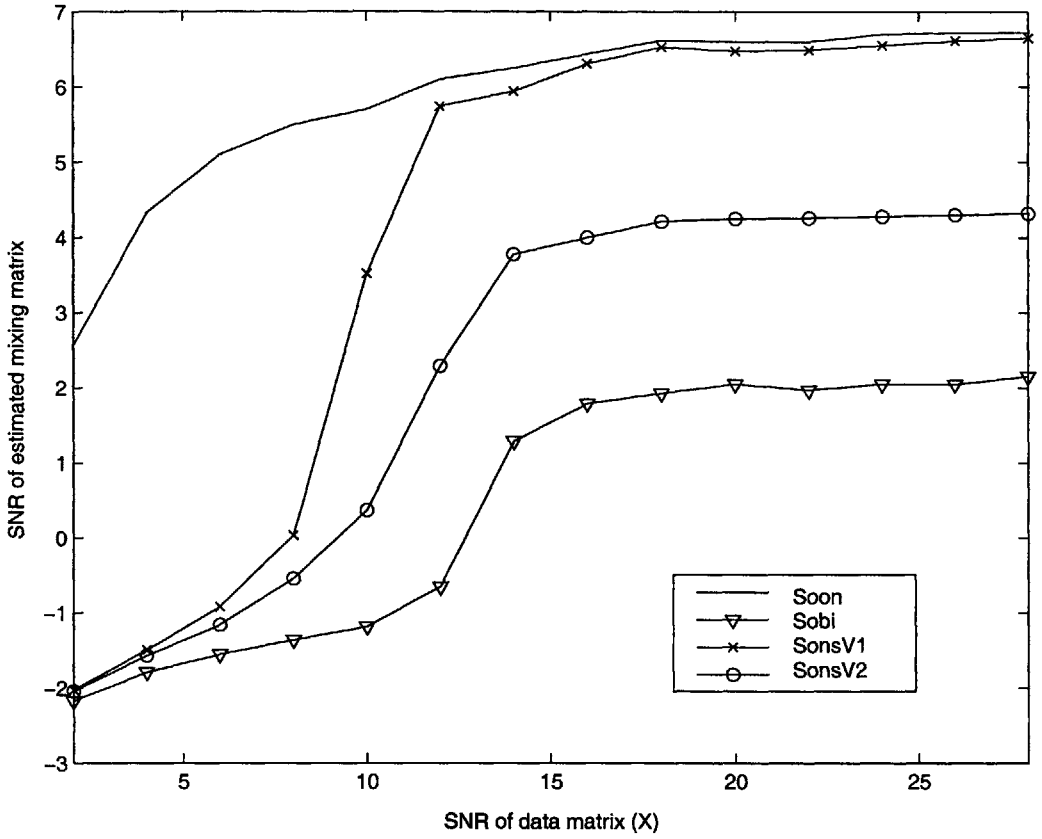

Figure 3-2: Test 2, which compares the performance of SOON to that of SOBI and two variants of SONS. The data set here is comprised of white second-order non-stationary sources.

3.3.3 Results

By inspection it can be seen that SOON performs better at the estimation of the mixing transform, **A**, when compared to both SOBI and SONS over the entire range of data SNR. As expected, this improvement is larger for low SNR data. The performance gain is particularly large in this range for the non-stationary/white test case. This can be attributed to the inability of SONS to use robust whitening for purely non-stationary white data. Also through this test we can see how sensitive performance is to properly characterizing the non-stationarity of the data. Each successive improvement in data segmenting from SOBI to SONS results in a large improvement in separation quality. Additionally in the colored stationary test, Figure 3-1, the sensitivity of performance to the initial weight vector choice is visible. Note in Figure

3-1 the performance gain achieved by choosing the initial weight vector according to the heuristic introduced in Chapter 2 was not particularly sensitive to the data SNR given that the inaccuracies present in the delayed sample covariance matrices are predominately due to error incurred by using a finite sample size.

3.4 Stability over Problem Space

3.4.1 Testing

Now that SOON has been shown to give increased separation performance in comparison to other standard second-order methods over typical data sets, we would like to display its sensitivity to the outlined degrees of freedom within the problem space. Here again we show the results for the two types of sources used in the above tests. For both we start from a nominal test case whose parameters can be seen in Table 3.3. Each additional test case then differers from the nominal case in one degree of freedom to display the sensitivity of SOON to that parameter. The parameters explicitly explored here are the source-observation ratio $\frac{k}{n}$, the number of data samples m, the data SNR, and the angle between columns of the mixing transform $\angle_A$. The performance for the estimates of $\mathbf{A}$, $\mathbf{P}$, $\mathbf{G}$, and $\mathbf{W}$ are given using the previously defined residual metric in Equation (3.4). Performance for the order estimation is then given by:

$$\mu_k = E[\frac{k - k'}{k}] \tag{3.14}$$

$$\sigma_k = \sqrt{E[(\frac{k - k'}{k} - \mu_k)^2]} \tag{3.15}$$

where k' and k are the estimated and actual number of sources respectfully. The results of these tests can be seen in Tables 3.4 and 3.5.

Table 3.3: Nominal Test Parameters

Parameter	Value
n	55
k	$0.36n$ $(k = 20)$
m	m=10,000
SNR of $\mathbf{X}$	16 dB
Mean angle between columns of $\mathbf{A}$	$\angle_A = 45°$

Table 3.4: SNR (dB) and order error performance of SOON on colored stationary gaussian test data

Exp.	Nom.	k = 0.5n	m = 1000	$\mathbf{SNR}_X$= 3 dB	$\angle_A$= 20°
$\widehat{r}_{\mathbf{G}}$	37.3	19.8	32.4	32.7	33.7
$\widehat{r}_{\mathbf{W}}$	4.4	3.1	4.4	4.5	4.2
$\widehat{r}_{\mathbf{A}}$	11.9	11.8	5.4	6.5	10.1
$\widehat{r}_{\mathbf{P}}$	10.9	10.1	3.7	3.3	5.3
μ_k	-0.002	-0.006	-0.002	-0.004	-0.002
σ_k	0.01	0.03	0.01	0.02	0.01

3.4.2 Results

By inspection we can see that SOON performs well in the estimation of all parameters in the nominal test case. Here it is also evident that an increase in $\frac{k}{n}$ significantly affects only the performance of the noise estimation. A decrease in separation performance, estimation of A and P, is noticeable when the sample size of the data set was decreased. This results from a loss in statistical accuracy in the estimated covariance

Table 3.5: SNR (dB) and order error performance of SOON on white non-stationary gaussian test data

Exp.	Nom.	k = 0.5n	m = 1000	$\mathbf{SNR}_X$= 3 dB	$\angle_A$= 20°
$\widehat{r}_{\mathbf{G}}$	36.8	19.8	31.9	32.2	33.9
$\widehat{r}_{\mathbf{W}}$	4.4	3.1	4.5	4.3	4.6
$\widehat{r}_{\mathbf{A}}$	6.2	6.6	2.2	3.3	6.3
$\widehat{r}_{\mathbf{P}}$	3.7	4.5	.3	.8	.5
μ_k	0	-0.008	-0.002	-0.002	-0.004
σ_k	0	0.04	0.01	0.01	0.02

matrices used in the whitening and joint diagonalization steps. Likewise we see that a decrease in data SNR, significantly affects only the separation performance, i.e. estimates of $\mathbf{A}$ and $\mathbf{P}$. The effect of decreasing the linear independence of the columns of the mixing transform A was limited primarily to the estimation of the source signals P. These observations are consistent over both data sets, with the only difference being that the absolute separation performance was better for the stationary colored sources than the non-stationary white sources.

Chapter 4

Application to Remote Sensing and Financial Data

4.1 Remote Sensing Background

Remote sensing typically characterizes the electromagnetic radiation emitted or reflected over some geographic area of interest. The instrument gathering the data often records energy over many different frequency channels, where the response for each channel depends on the physical properties of the surface being observed. For high resolution instruments with many channels there is often a high level of redundancy within the data collected. It is therefore often desirable to change the basis over which the data is stored in order to gain insight into the underlying dynamics of the data, and to facilitate greater data compression. The two classic methods most used in the remote sensing community for these purposes are Principal Component Analysis (PCA) and its extension, Noise Adjusted Principal Component Analysis (NAPCA) [19] [20].

4.2 Principal Component Analysis (PCA)

4.2.1 The Method

Principal component Analysis is a method for changing the coordinate system of a data set. Formally the process can be described as operating on a zero-mean data set $\mathbf{X}$, find a transform,

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{pmatrix} \tag{4.1}$$

such that,

$$\mathbf{b}_i = \max_{\|\mathbf{b}\|=1} \mathbf{b}\mathbf{X}^i(\mathbf{b}\mathbf{X}^i)^H \tag{4.2}$$

where,

$$\mathbf{X}^i = \mathbf{X} - \sum_{k=1}^{i-1} \mathbf{b}_k^H \mathbf{b}_k \mathbf{X} \tag{4.3}$$

The vector $\mathbf{b}_i$ is referred to as the i^{th} principal component and is contained in the subspace spanned by the data set which is orthogonal to $\mathbf{b}_1$, $\mathbf{b}_2$, ..., and $\mathbf{b}_{i-1}$. This vector is the vector in this space along which the data set has greatest variance. The Principal components transform has the special property that the first j principal components span the j-dimensional subspace of the data with the greatest total variance. This makes PCA a popular choice for transforming data sets, as higher variance is associated with higher information content.

4.2.2 Limitation of PCA

The problem associated with PCA is that the method ranks components by variance without considering whether this variance corresponds to interesting phenomena or

noise. For noisy data sets, the PCA result may be undesirable as the high ranking components may have lower SNR than lower ranked components. A natural correction to this leads to an extension of PCA, Noise-Adjusted Principal Components.

4.3 Noise Adjusted Principal Component Analysis (NAPCA)

4.3.1 The Method

NAPCA works to improve PCA by eliminating the effect of noise in the ranking process. It proceeds by finding a transform that normalizes the noise within the data set so that the noise adjusted principal components can then be found as the principal components of the transformed data set. More formally, find a transform $\mathbf{H}$ on the data,

$$\mathbf{HX} = \mathbf{HAP} + \mathbf{HGW} \tag{4.4}$$

such that:

$$\mathbf{HG}(\mathbf{HG}^H) = \mathbf{I} \tag{4.5}$$

with:

$$\mathbf{b}_i = \max_{\|\mathbf{b}\|=1} \mathbf{bHX}^i(\mathbf{bHX}^i)^H \tag{4.6}$$

where:

$$\mathbf{HX}^i = \mathbf{HX} - \sum_{k=1}^{i-1} \mathbf{b}_k^H \mathbf{b}_k \mathbf{HX} \tag{4.7}$$

where the $\mathbf{b}_i$'s now represent the noise-adjusted principal components. By transforming the data in this way the noise contributes equal variance for any vector in the

observation space. This then produces components ranked by the variance of their signal contribution.

4.3.2 Limitation of NAPCA

The drawback with NAPCA is the same as with SOBI and SONS. In order for the process to be applicable the noise covariance matrix must either be known or an accurate estimate must be achievable. If this is not the case then the noise cannot be normalized.

4.4 Blind Signal Separation Approach

4.4.1 BSS Model

An alternative approach to this problem that we explore here is to model remote sensing data using a two-dimensional extension of the blind signal separation framework used by this thesis. Namely, given n two-dimensional images, $\mathbf{X}$:

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n] \tag{4.8}$$

with each image, $\mathbf{X}_i$, of size m_1 by m_2, and expressible in terms of k surfaces,

$$\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, \ldots \mathbf{P}_k] \tag{4.9}$$

and a normalized white noise component, $\mathbf{W}_i$, as,

$$\mathbf{X}_i = \alpha_{i1}\mathbf{P}_1 + \alpha_{i2}\mathbf{P}_2 + \ldots + \alpha_{ik}\mathbf{P}_k + G_i\mathbf{W}_i \tag{4.10}$$

where $\mathbf{G}_i$ is a positive scalar, find the set of k underlying surfaces, $\mathbf{P}$. Here the data set, $\mathbf{X}$, is transformed to a new basis, whose components are the $\mathbf{P}_i$'s.

4.4.2 Solution using SOON

In the BSS model, it is assumed that the response images, $\mathbf{X}_i$'s, recorded at different frequencies will be comprised of a linear combination taken from a set of underlying surfaces, $\mathbf{P}_i's$, that are in some way independent. This model can be justified by the fact that electromagnetic energy will respond differently based on the physical properties of the surface from which it emanated. However in order for SOON to be applicable, the set of surfaces, $\mathbf{P}$, that the electromagnetic energy responds to, must have diverse spatial correlations or independently varying spatial correlations. Namely at least one of the following conditions must hold,

Condition I

The spatial correlation of each surface is distinct:

$$\mathbf{R}_{\mathbf{P}_i} \neq \mathbf{R}_{\mathbf{P}_j} \quad \forall i, j \tag{4.11}$$

where $\mathbf{R}_{\mathbf{P}_i}$ represents the two-dimensional auto-correlation function of the surface $\mathbf{P}_i$.

Condition II

The spatial correlations of each surface are non-stationary:

$$\exists t_{i1}, t_{i2}, \tau_{i1}, \tau_{i2} \quad s.t. \quad \mathbf{R}_{\mathbf{P}_i}(\tau_{i1}, \tau_{i2}) \neq \mathbf{R}_{\mathbf{P}_i}(t_{i1} + \tau_{i1}, t_{i2} + \tau_{i2}) \quad \forall i \tag{4.12}$$

4.4.3 Applicability of SOON

The above conditions do not heavily restrict the applicability of SOON to remote sensing data. Justifying this statement is the fact that many data sets are collected over regions that contain a diverse set of surface features that are comprised of varying physical components. The location of these components within the surface area will often be non-uniform leading to non-stationary responses. Additionally their locations will often be independent of one another resulting in a diverse set of spatial

correlations.

4.5 Case Study: AVIRIS Data

4.5.1 The AVIRIS instrument

To compare the performance of SOON to that of PCA and NAPCA we have chosen a data set collected with the AVIRIS instrument used at the Jet Propulsion Laboratory [21] [22]. Aviris is 224 channel instrument that records frequencies ranging from 370 to 2500 nm which spans the visible to the mid-infrared bands. The images used here have been collected over Moffet Field, CA and are particularly interesting in that the upper halves of these images contain natural terrain whereas the bottom halves contain mostly man-made structures. The images collected by channels 20, 50, 80, 110, 140, 170, and 200 can be seen in figure 4-1.

4.5.2 Quantitative Comparison

Component Ranking Metric

Given that the mixing matrix, $\mathbf{A}$, is not known for this data set a new quantitative metric is needed in order to compare the performances of SOON, PCA, and NAPCA. A natural selection for this metric is the information content of the components found by each method. Here we would like to rank the significance of a component as being low if it resembles white noise and high if it contains some level of meaningful structure. To do this we take the two-dimensional Fast Fourier Transform (FFT) of each component and sum over the weights in its high frequency region, i.e.,

$$F_i \equiv FFT(\mathbf{P}_i)$$

$$\zeta_i = \sum_{j,k:j+k>\tau} |F_{ijk}|$$

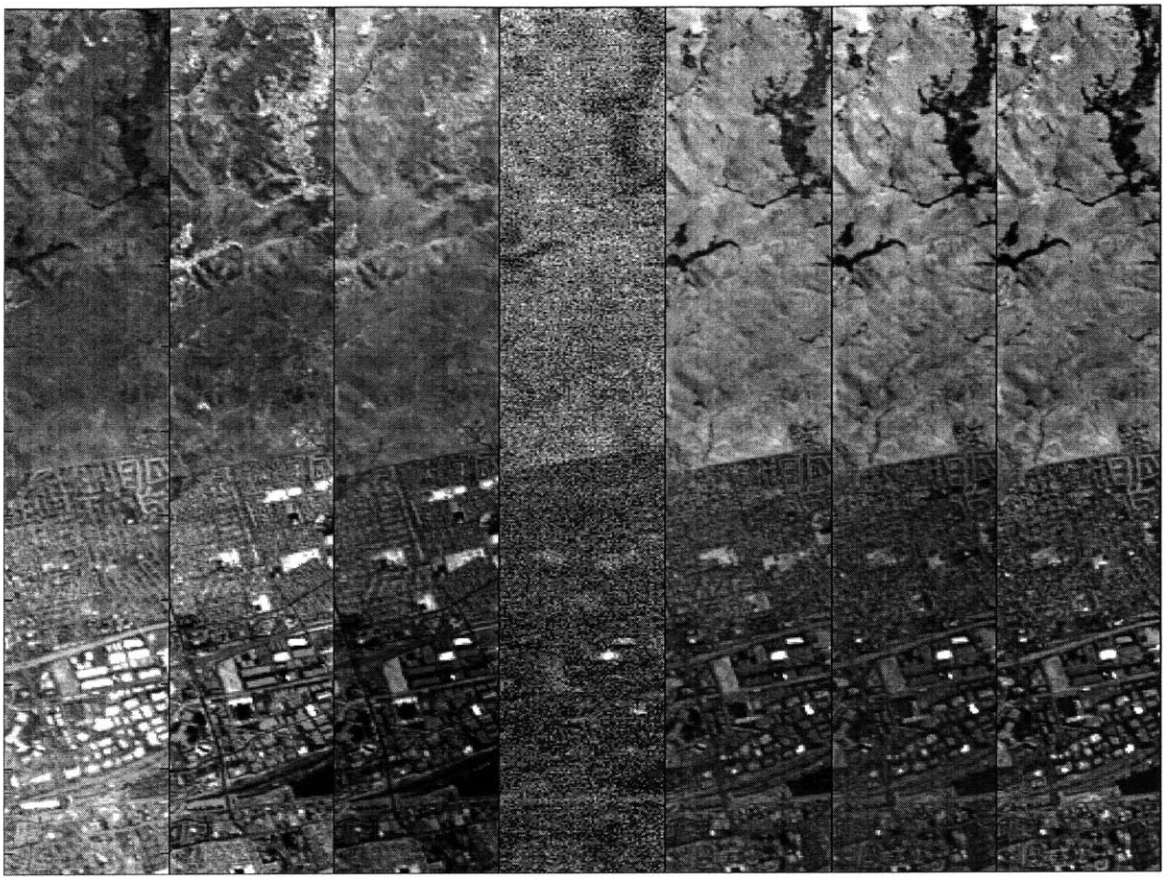

Figure 4-1: Data collected over Moffet Field, CA by the AVIRIS instrument, channels 20, 50, 80, 110, 140, 170, and 200 are shown from left to right.

where $\mathbf{P}_i$ is the i^{th} component found by the given method and τ is the threshold for being considered high frequency. In particular for an image with pixel dimensions x_1 and x_2, we have made the selection of τ as:

$$\tau = min\{\frac{x_1}{2}, \frac{x_2}{2}\}$$

so as to give a non-dimensionally biased result. Those components with a relatively large ζ_i are then considered to be less interesting as they more closely resemble white noise.

Test and Results

Here we have tested the performance of SOON and NAPCA in terms of this metric over the AVIRIS data set. The normalization step of NAPCA has been carried out by using the estimate of the noise covariance matrix provided by SOON. The results of using PCA have been omitted as they were nearly identical to those of NAPCA. This is due to the noise across frequency channels being fairly normalized on the instrument. In figure 4-2 we see a graph displaying the value of this metric for the components found by each method ranked in ascending order. For the noise threshold

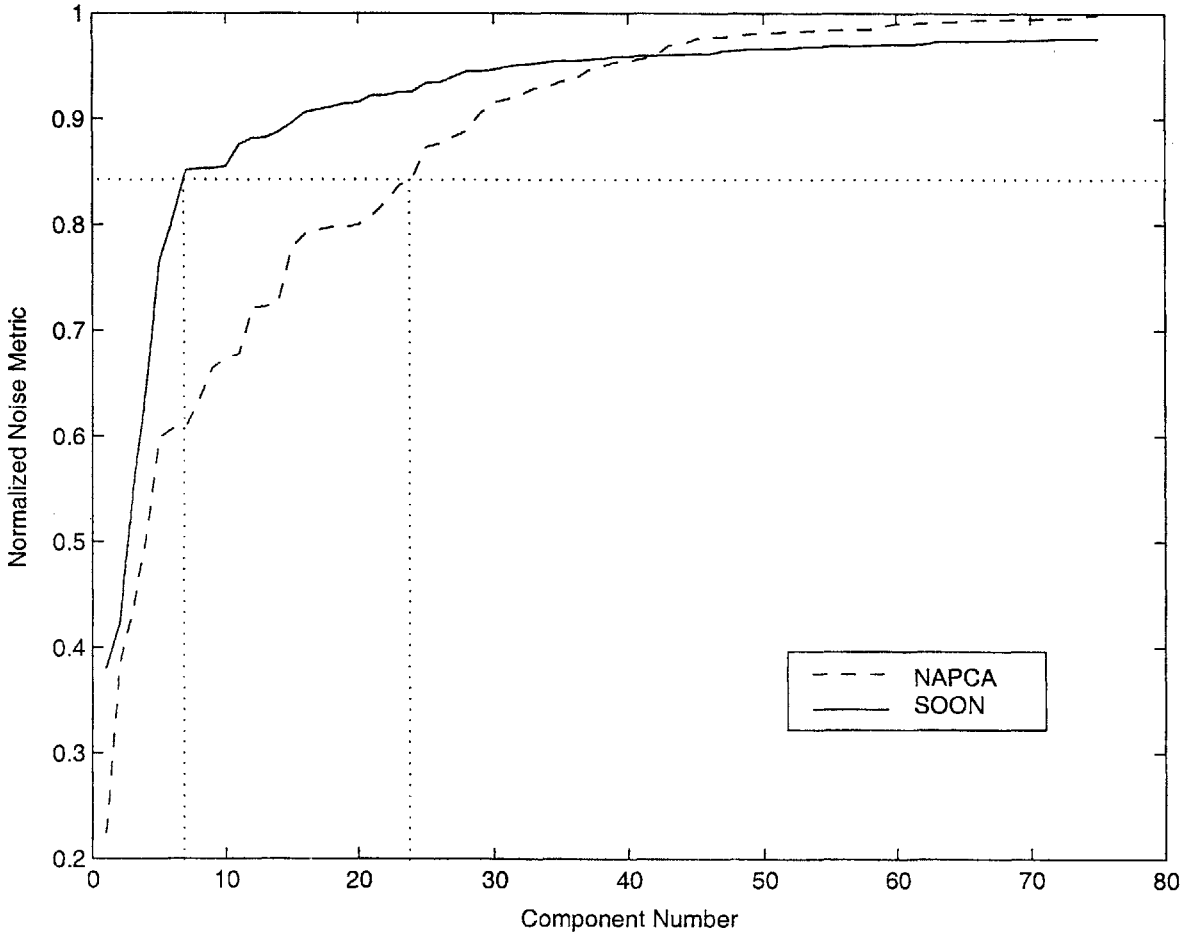

Figure 4-2: This graph displays the noise metric defined above for each of the components found by SOON and NAPCA. The components are sorted here in ascending order of this metric. By inspection the higher ranking components of SOON, those containing the most information content, are less noisy than the corresponding NAPCA components

in figure 4-2, the number of components found using SOON and NAPCA was 24 and 7 respectfully. Here it is clearly shown that the highest ranking components found by SOON were less noisy than the corresponding components found using NAPCA.

4.5.3 Qualitative Comparison

Remote Sensing data, such as that collected by the AVIRIS instrument, often consists of images that can be visually inspected for information content. Here we further compare SOON to NAPCA, by visually inspecting the components found by each method to see how either might be more informative than the other. Below in figure 4-3 are the top halves (natural terrain) of the visually most interesting images from each method. Similarly, in figure 4-4 the most interesting bottom halves (urban terrain) found by each method are displayed.

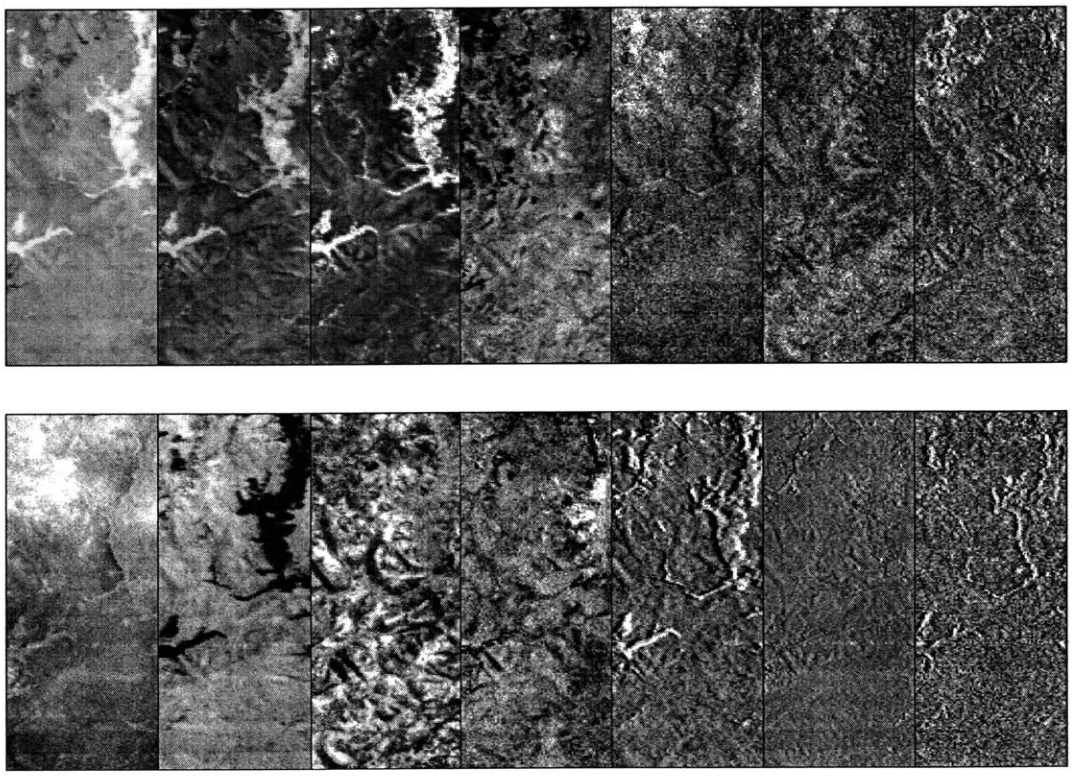

Figure 4-3: Here are the top halves of the visually most interesting components found by SOON and NAPCA. In the top row are the NAPCA components, and the bottom row presents the SOON components.

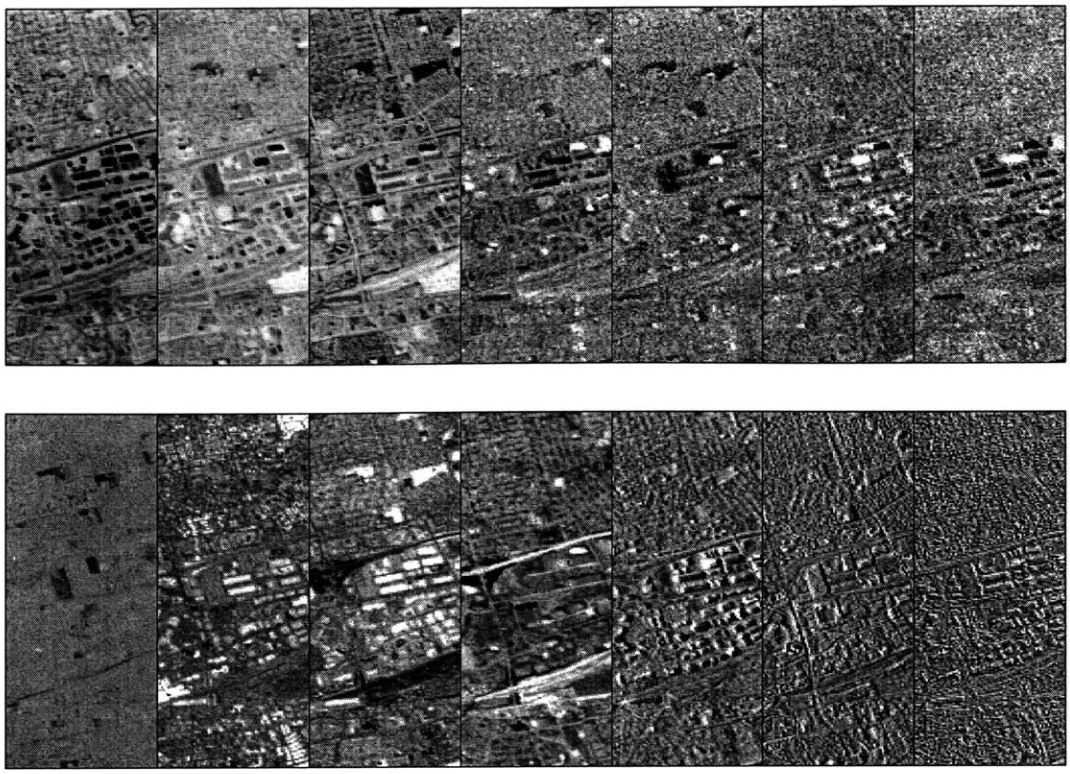

Figure 4-4: Here are the bottom halves of the visually most interesting components found by SOON and NAPCA. In the top row are the NAPCA components, and the bottom row presents the SOON components.

Component Comparison

By inspection we can see that the components picked up by each method are visually quite different. The NAPCA components seem to capture redundant information. This is particularly noticeable in the first three NAPCA components of Figure 4-3. In contrast every component found by SOON is quite different. Additionally, the components found by NAPCA appear to drop into the noise much faster. This can be verified by inspecting the last few components found by each method in Figures 4-3 and 4-4.

Physical Interpretation

When using these methods to change the basis of the data set, it is often desirable to find components that have some sort of physical significance which aid in the

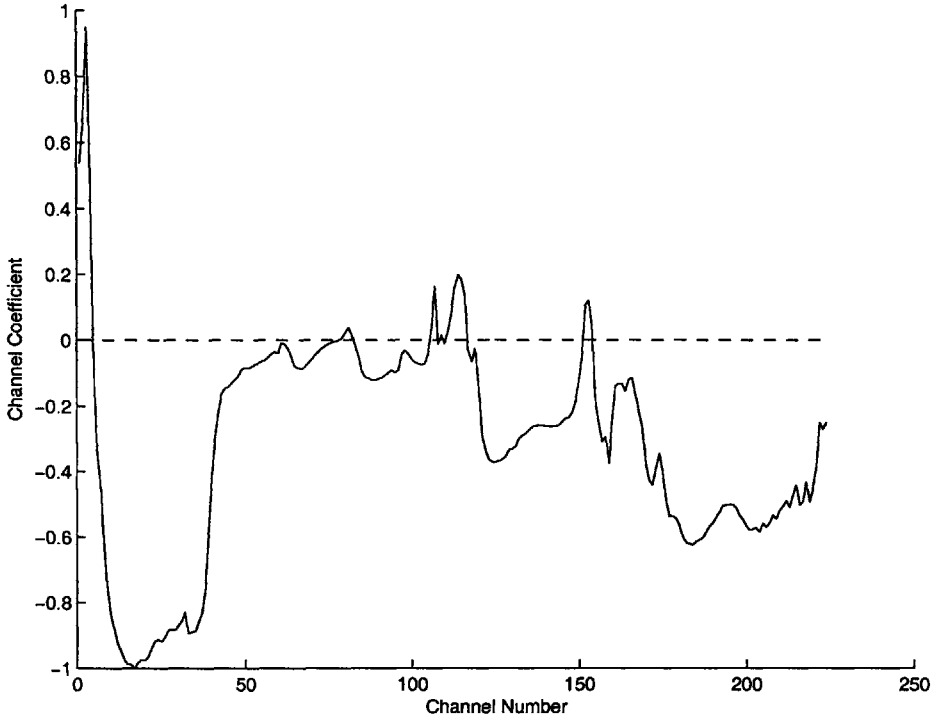

Figure 4-5: Here are the normalized channel coefficients for the 5^{th} SOON component of figure 4-4. Here it can be seen that the majority of energy is weighted in the higher frequency, visible wavelength channels of the instrument, channels 1-40.

understanding of the geophysics of the surface being observed. An example of this can be seen in the shadowing features present in some of the SOON components that aren't present in the NAPCA components. In particular, shadowing can be seen in the 5^{th}, 6^{th}, and 7^{th} SOON components of Figures 4-2 and 4-3. These images exhibit three distinct shadowing signatures that vary according to their level of spatial correlation. In the 5^{th}, 6^{th}, and 7^{th} SOON components of Figure 4-3, the shadowing can be seen to become finer in each successive image. Theoretically SOON will separate surface features that have diverse spatial correlations, which explains the separation of these shadowing signatures into components based on spatial correlation. The normalized channel weights for the 5^{th} SOON component of Figure 4-3 are shown in Figure 4-4. By inspection of this component it looks like a shadowing signature cast by buildings and other objects blocking sunlight. Looking at Figure 4-3, it can be seen that the majority of the energy of this component comes from the high frequency, visible

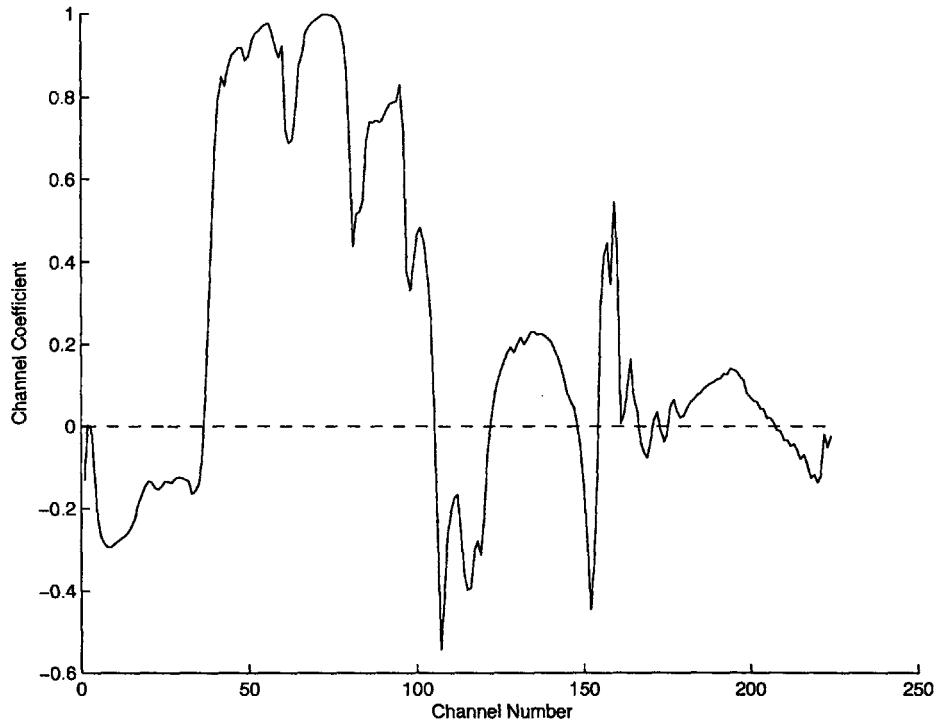

Figure 4-6: Here are the normalized channel coefficients for the 6^{th} SOON component of figure 4-4. Here it can be seen that the majority of energy is weighted in the near-infrared channels, channels 40-110.

wavelength channels of the instrument, which supports this observation. Additionally, the channel weights for the 6^{th} SOON component of Figure 4-3 are displayed in Figure 4-5. Here, within this component, there is a large contrast in magnitude between the left and right sides of the buildings in the image. A possible explanation for this phenomena could be that this component contains the thermal signature cast by the sun heating one side of the objects in the image. Looking at Figure 4-5, the majority of the energy of this component comes from the shorter wavelength infrared channels of the instrument, which supports this claim.

4.6 Financial data

4.6.1 Adjusted Stock Returns Experiment

Additionally blind signal separation techniques can be applied to financial data, such as equity returns in order to potentially resolve interesting phenomena underlying market behavior. A short experiment was carried out to test the applicability of SOON in this difficult domain. In particular the data set was comprised of daily stock returns for 500 companies traded on US markets over the ten-year period 1994-2003. The returns were normalized with respect to dividends, splits, merges, and all other cash flows to share holders so that such artifacts would not factor in the resulting pricing behavior. The data was obtained from the Center for Research in Security Prices [27] and provided by Wharton Research Data Services [28]. Here the objective was to obtain the independent components detected by SOON and establish their utility in predicting future pricing behavior and explaining market dynamics. A scree plot containing the eigenvalues of the covariance matrix for the pricing data can be seen in Figure 6-1. Additionally a collection of independent components found by SOON within this data set are shown in Figure 6-2.

4.6.2 Results

By inspection of Figure 6-1 it can be seen that the space spanned by the highest ordered eigenvectors contain most of the variance in the data set. This seems to suggest that a large percentage of the movement in the market is explained by a relatively small number of components. Additionally the components in Figure 6-2 are sinusoids and imply that perhaps there exist predictable behavior in the market if the sample correlations hold over time. To test the predictability of these components an experiment was carried out where SOON was trained on the first seven years of the data set to obtain a set of independent components. The entire ten years of data was then projected onto these components. Four of the resulting signals can be seen in Figure 6-3. Figure 6-3 clearly shows that the components found using SOON are

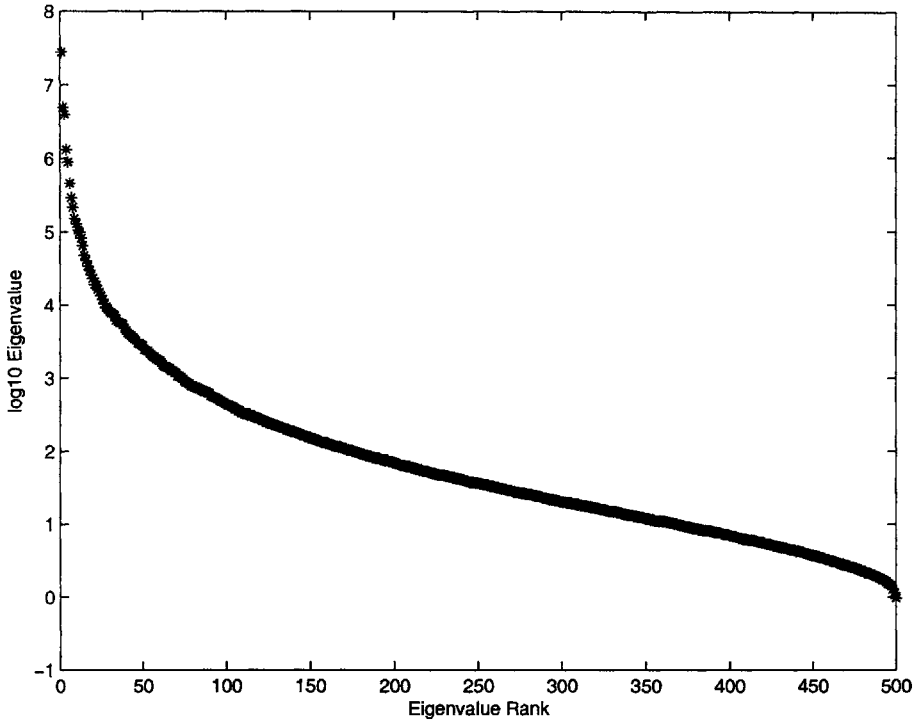

Figure 4-7: Here the scree plot is shown for a data set consisting of the adjusted returns for 500 US equities over the ten year period starting in 1994 through 2003.

not helpful in predicting the pricing behavior of the market. This can be explained by the fact that equity pricing behavior is inherently non-stationary so that correlations between different stocks change over time. The result of this market behavior is that the whitening transform and time delayed covariance matrices calculated in SOON over any particular time interval do not hold over future periods.

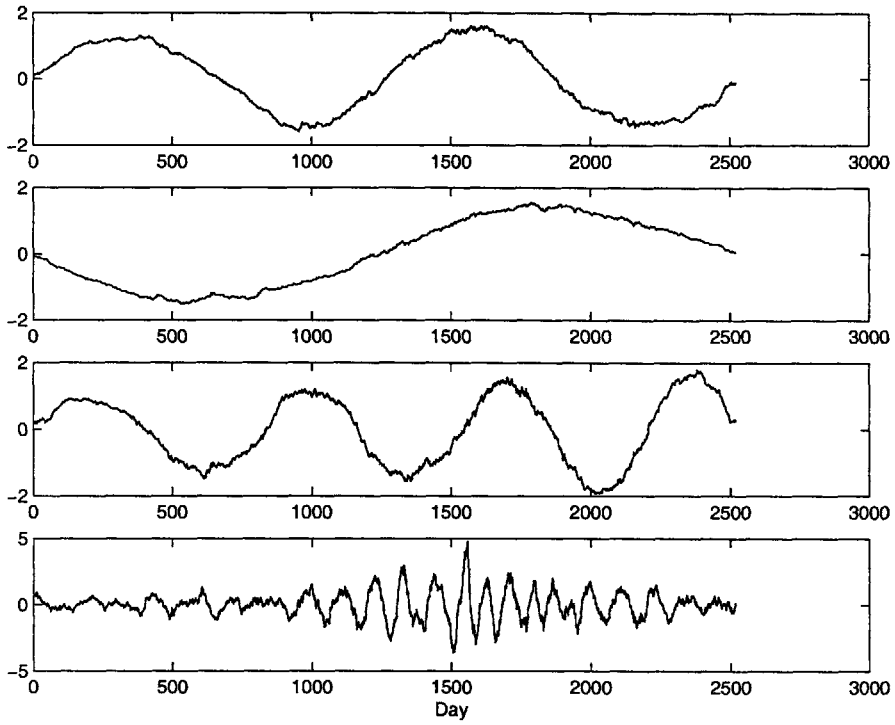

Figure 4-8: Four of the highest variance components found using the SOON algorithm from the financial data set.

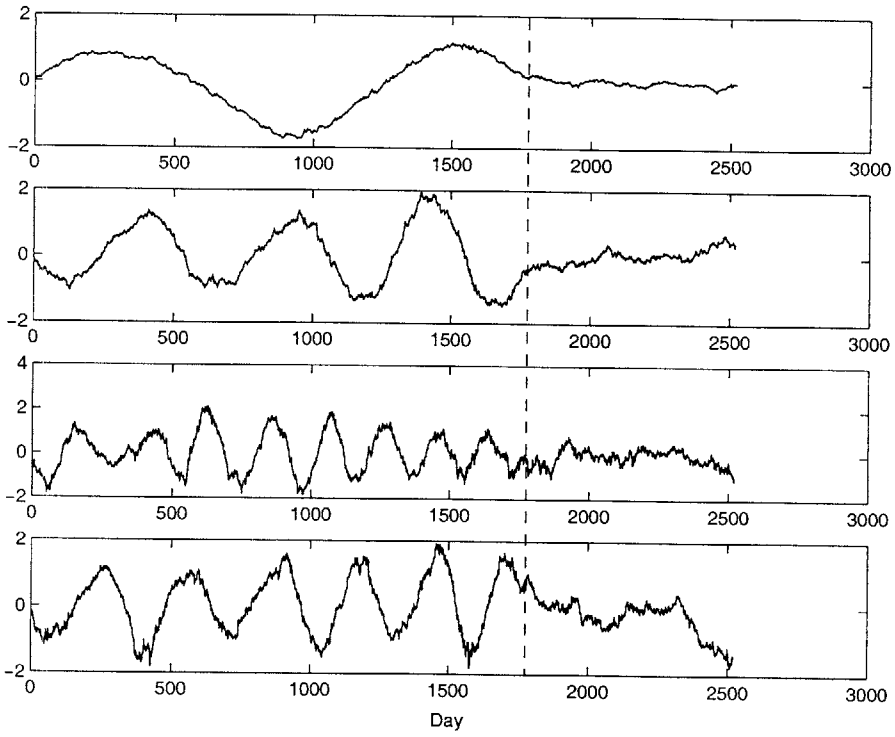

Figure 4-9: The complete 10 year data set projected onto four of the components found by SOON trained over the first seven years of data.

Chapter 5

Higher Order Extension

5.1 Higher Order Architecture

The SOON algorithm presented in this thesis works within the domain of second-order separation by taking advantage of second-order diversity and non-stationarity in the underlying source signals. Sources with higher-order statistics offer additional opportunities for source separation. The present second-order architecture of the SOON algorithm is constructed to estimate first the number of sources and the noise through an iterative procedure using scree plots and the Expectation Maximization (EM) algorithm. Next the noise-reduced data is whitened through a heuristically modified robust whitening step. The last step finds the transform on the data set that minimizes a joint diagonalization criteria for a set of second-order matrices. This architecture could be modified to take advantage of higher order statistics by replacing the joint diagonlization step of the algorithm with a new cost function, $g(\bullet)$, that evaluates the cost of the estimated signals based on some higher order statistic, i.e. find $\widehat{\mathbf{P}}$ such that:

$$\widehat{\mathbf{P}} = \widehat{\mathbf{A}}^{\#}\widehat{\mathbf{X}} \tag{5.1}$$

with:

$$\widehat{\mathbf{A}} = \min_{\mathbf{A}} g(\mathbf{A}^{\#}\widehat{\mathbf{X}}) \tag{5.2}$$

where $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{A}}$ are estimates of $\mathbf{P}$ and $\mathbf{A}$, and $\widehat{\mathbf{X}}$ is the noise-reduced data set[1]. A block diagram for this generalized SOON architecture can be seen in Figure 6-1. This new cost function could include minimizing the joint higher order moments

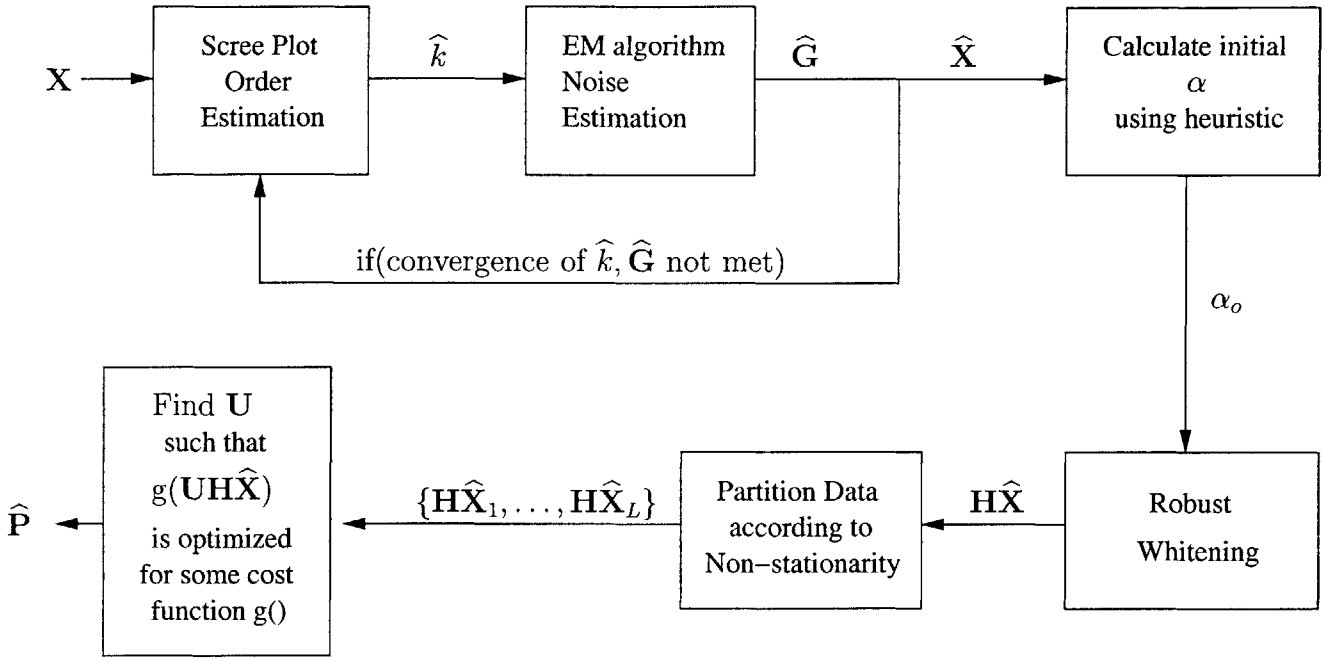

Figure 5-1: Architecture of the generalized SOON algorithm

of the source signals or joint information content. Many cost functions for defining independence exist in the higher order domain, the appropriate choice of which would depend on the data model.

[1]here # indicates the Moore-Penrose pseudo-inverse

5.2 Maximum Information/Maximum Likelihood

One possibility for the cost function, $g(\bullet)$, is the popular infomax cost function introduced in [23]. Here we maximize with respect to $\widehat{\mathbf{A}}$,

$$g(\widehat{\mathbf{A}}) = H(h(\widehat{\mathbf{A}}^{\#}\widehat{\mathbf{X}})) \tag{5.3}$$

where $H(\mathbf{y})$ is the differential entropy for the vector of random processes $\mathbf{y}$,

$$H(\mathbf{y}) = -\int_{\mathbf{y}'} log(f_{\mathbf{y}}(\mathbf{y}'))f_{\mathbf{y}}(\mathbf{y}')d\mathbf{y}' \tag{5.4}$$

and $h() : \mathbb{R}^{n\times m} \rightarrow \mathbb{R}^{n\times m}$ is a squashing function for restricting the entropy to be finite. An equivalent approach [24] is to find $\widehat{\mathbf{P}}$ as the maximum likelihood [25] [26] estimate of $\mathbf{P}$ given $\mathbf{X}$, i.e.

$$\widehat{\mathbf{P}} = \max_{\mathbf{P}} f(\mathbf{X}|\mathbf{P}) \tag{5.5}$$

In both cases the distribution of the sources are assumed known in advance and the estimate $\widehat{\mathbf{P}}$ of the sources is found as the estimate that most resembles this distribution.

5.3 Cumulants

Another higher order approach for the cost function, $g(\bullet)$, are cumulant based methods [5]. Often these methods restrict their attention to fourth-order cumulants for computational reasons, as is the case in the popular JADE [7] separation algorithm. The fourth-order cumulant for the random variables $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$ is defined as:

$$Cum(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) = c_1 - c_2 \tag{5.6}$$

where:

$$c_1 = E[\mathbf{X}_1'\mathbf{X}_2'\mathbf{X}_3'\mathbf{X}_4'] \tag{5.7}$$

$$c_2 = E[\mathbf{X}_1'\mathbf{X}_2']E[\mathbf{X}_3'\mathbf{X}_4'] + E[\mathbf{X}_1'\mathbf{X}_3']E[\mathbf{X}_2'\mathbf{X}_4'] + E[\mathbf{X}_1'\mathbf{X}_4']E[\mathbf{X}_2'\mathbf{X}_3'] \tag{5.8}$$

and $\mathbf{X}_i' = \mathbf{X}_i - E[\mathbf{X}_i]$. Following directly from this definition, the fourth-order cumulant for independent random variables $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$ is 0. Using fourth-order cumulants then for blind signal separation would involve finding a transform $\widehat{\mathbf{A}}$ on the whitened data $\widehat{\mathbf{X}}$ such that the fourth-order cross cumulants of the transformed data set are minimized, i.e. find $\widehat{\mathbf{A}}$ such that:

$$\sum_{i \neq j \neq k \neq l} Cum(\widehat{\mathbf{A}}^{\#}\widehat{\mathbf{X}}_i, \widehat{\mathbf{A}}^{\#}\widehat{\mathbf{X}}_j, \widehat{\mathbf{A}}^{\#}\widehat{\mathbf{X}}_k, \widehat{\mathbf{A}}^{\#}\widehat{\mathbf{X}}_l) \tag{5.9}$$

is minimized, where $\widehat{\mathbf{A}}^{\#}\widehat{\mathbf{X}}_i$ refers to the i^{th} row of the transformed data set, and Cum() is the sample fourth-order cumulant. The number of cumulants grows large with the size of the data set so often only a subset is used in the optimization. This subset depends on the nature of the dependencies present in the data.

Chapter 6

Conclusions and Suggested Further Work

6.1 Suggestions for further work

6.1.1 Higher Order Statistics

The SOON algorithm presented in this thesis works within the domain of second-order separation by taking advantage of second-order diversity and non-stationarity in the underlying source signals. The experiments of this thesis compared SOON to other second-order methods on both simulated and real data sets. A proposed extension of the SOON architecture for separation based on a general cost function was introduced in Chapter 5. An interesting extension of this work would be to implement this architecture to separate signals based on their fourth-order cumulants. A comparison then could be made between this fourth order version of SONS and the standard algorithms in this class, including the JADE algorithm. Additionally, data sets collected in the remote sensing community are most often non-gaussian in nature. It would be interesting to see how this fourth-order SONS compares to the results of Chapter 4.

6.1.2 Partitioning Schemes

This thesis mentioned the significance of correct data partitioning when taking advantage of non-stationarity within the data set. Additionally, this significance was shown explicitely in the simulated testing of Chapter 3. As noted, many heuristics exist for partitioning data sets according to their non-stationarity. An interesting study would be to analytically or quantitatively characterize the sensitivity of separation performance to the accuracy of the partitioning scheme used. In this way a tradeoff between computational burden and performance accuracy could be resolved giving insight into what heuristics are appropriate.

6.1.3 Application to Other Fields

In this thesis we have illustrated the potential for using BSS algorithms, in particular SOON, on remote sensing data in order to observe interesting dynamics within such data sets. In doing so, the experiment of Chapter 4 revealed interesting phenomena within the data that PCA and NAPCA did not detect. It is quite likely that similar discoveries could be found in data sets collected in other fields. An interesting extension to this work would then be to run experiments on a collection of data sets to see the range of applicability of this software to other fields such as atmospheric sciences, demographic data, and biological data.

6.2 Conclusions

This thesis has presented a blind signal separation algorithm that estimates the number of sources, noise levels, the mixing transform, and the separate signals and noises themselves. The performance of this algorithm, SOON, has been shown to be robust with respect to separation using second-order statistics, namely making use of second-order non-stationary and diversity in time correlation. In addition we have shown that SOON gives increased separation performance as compared the traditional second-order methods, SOBI and SONS, over a typical data sets. Similarly

this performance was shown to be robust to the nominal degrees of freedom of blind separation problems. Also the application of SOON to remote sensing data has been established. This included an experiment that revealed SOON's ability to detect interesting phenomena that traditional PCA and NAPCA analysis did not find.

Appendix A

A.1 Joint Diagonalization

Here we outline the joint diagonalization [3] method that is used in the SOBI, SONS, and SOON algorithms. The problem here can be formulated as: Given a set of L matrices,

$$\mathbf{M} = \{\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_L\} \tag{A.1}$$

find a unitary matrix $\mathbf{V}$ such that the joint diagonalization for $\mathbf{M}$, $\mathbf{V}$ is:

$$JD(\mathbf{M}, \mathbf{V}) \equiv \Sigma_{i=1}^{L} off(\mathbf{V}^H \mathbf{M}_i \mathbf{V}) \tag{A.2}$$

is minimized, where for a matrix $\mathbf{N}$ we define off($\mathbf{N}$) as:

$$off(\mathbf{N}) \equiv \sum_{i \neq j} | \mathbf{N}_{ij} |^2 \tag{A.3}$$

If it is the case that:

$$\mathbf{M}_i = \mathbf{U}\mathbf{D}_i\mathbf{U}^H \quad 1 \leq i \leq L \tag{A.4}$$

for some unitary matrix $\mathbf{U}$, and

$$\mathbf{D}_i = diag[d_{1i}, d_{2i}, \ldots, d_{ni}] \tag{A.5}$$

then any unitary matrix $\mathbf{V}$ that minimizes (A.2) will be essentially equal to $\mathbf{U}$ if and only if,

$$\exists k, 1 \leq k \leq L \quad s.t. \quad d_{ik} \neq d_{jk} \quad \forall 1 \leq i \neq j \leq n \tag{A.6}$$

In the blind signal separation context (A.4) will hold for the set of delayed covariance matrices of the whitened data. However because these covariance matrices are calculated using finite sample approximations, the Joint Diagonalization method will find a unitary matrix that approximately diagonalizes this set of matrices by minimizing (A.2).

A.2 Global Convergence Algorithm

The global convergence algorithm [4] is used by SONS and SOON in the robust whitening step for finding a whitening matrix that is positive definite. The procedure can be outlined as: Given a matrix $\mathbf{R}$ that can be expressed as:

$$\mathbf{R} = \sum_{i=1}^{K} \alpha_i \mathbf{R}_i \tag{A.7}$$

where each $\mathbf{R}_i$ is a symmetric positive definite matrix, and

$$\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_K] \tag{A.8}$$

is a non-zero vector, iteratively update α until the matrix $\mathbf{R}$ is positive definite. At each step of the procedure the matrix $\mathbf{R}$ is checked for positive definiteness. If $\mathbf{R}$ is not positive definite, α is updated as:

$$\alpha = \alpha + \delta \tag{A.9}$$

where:

$$\delta = \frac{[\mathbf{u}^H\mathbf{R}_1\mathbf{u}, \mathbf{u}^H\mathbf{R}_2\mathbf{u}, \ldots, \mathbf{u}^H\mathbf{R}_K\mathbf{u}]^H}{\|\mathbf{u}^H\mathbf{R}_1\mathbf{u}, \mathbf{u}^H\mathbf{R}_2\mathbf{u}, \ldots, \mathbf{u}^H\mathbf{R}_K\mathbf{u}\|} \tag{A.10}$$

and $\mathbf{u}$ represents the eigenvector of $\mathbf{R}$ with smallest eigenvalue.

Bibliography

[1] Attias, H. *Independent factor analysis.* Neural Computation, vol. 11: pp. 803-851, (1998).

[2] Rowe, Daniel B. *A Bayesian approach to blind source separation.* Journal of Interdisciplinary Mathematics, 5(1): 49-76, (2002).

[3] Belouchrani, A., K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. *A blind source separation technique using second-order statistics.* IEEE Transactions on Signal Processing, 45(2): 434 - 444, (1997).

[4] A. Cichocki and S. ichi Amari. *Adaptive Blind Signal and Image Processing.* west sussex, England: John Wiley Sons, LTD

[5] Cardoso, J.-F., S. Bose, and B. Friedlander. *On optimal source separation based on second and fourth order cumulants.* In Proc. IEEE Workshop on SSAP, Corfou, Greece (1996).

[6] D. T. Pham, "Blind separation of instantaneous mixtures of sources based on order statistics," *IEEE Transactions on Signal Processing*, vol. 48, pp. 363-375, 2000.

[7] J. F. Cardoso, "Iterative technique for blind source separation using fourth order cumulants," *Signal Processing VI: Theory and Application*, pp. 739-742, 1992.

[8] Comon, P. *Independent Component Analysis - A New Concept?* Signal Processing, 36: pp. 287-314, (1994).

[9] R. Horn and C. Johnson, *Matrix Analysis.* Cambridge, U.K.: Cambridge Univ. Press, 1985.

[10] Y. I. L. Tong and R. Liu, "A finite-step global convergence algorithm for the parameter estimation of multichannel ma processes."

[11] R. R. K. L. C. Zhao and Z. D. Bai, "On detection of the number of singals in presence of white noise," *Journal of Multivariate Analysis*,vol. 20(1).

[12] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing.* vol. 33(2).

[13] H. Akaile, "A new look at the statistical model identification," *ieeetaucon*, vol. 19

[14] G. Schwartz, "Estimating the dimension of a model," *4nn. Stat.*, vol. 6, 1978.

[15] Lee, J. *Blind noise estimation and compensation for improved characterization of multivariate processes.* PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. (March 2000).

[16] Lee, J. and D.H. Staelin. *Iterative signal-order and noise estimation for multivariate data.* Electronics Letters, 37(2): pp. 134-135, (2001).

[17] Mueller, A. *Iterative Blind Separation of Gaussian Data of Unknown Order.* MEng thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. (June 2003).

[18] H. E. S. Kensuke Fukuda and L. A. N. Amaral, "Heuristic segmentation of a nonstationary time series," *Physical Review E,* vol. 69, 2004.

[19] A. A. Green, B. Berman, P. Switzer, and M. D. Craig. A transformation for ordering multipsectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing,* 26:65-74, 1988.

[20] J. B. Lee, A. S. Woodyatt, and M. Berman. Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing,* 28:295-304, 1990.

[21] *AVIRIS.* Jet Propulsion Laboratory, California Institute of Technology. 2000-2001 <http://aviris.jpl.nasa.gov/>. Accessed Feb.-May, 2003.

[22] Vane, G. *1987: Airborne Visible/Infrared Imaging Spectrometer (AVIRIS).* JPL Publication 87-36: p.97, (1987).

[23] Bell, A.J. and T.J. Sejnowski. *An information-maximization approach to blind separation and blind deconvolution.* Neural Computation, Vol. 7: pp. 1129-1159, (1995).

[24] Cardoso, J. F. (1997). Infomax and maximum likelihood for blind source separation. IEEE Signal Processing Letters, 4(4):112-114.

[25] Belouchrani, A. and J.-F. Cardoso. *Maximum likelihood source separation by the expectation-maximization technique: Deterministic and stochastic implementation.* In Proceedings of 1995 International Symposium on Non-Linear Theory and Applications. Las Vegas, NV: pp. 49-53, (1995).

[26] Dempster, A., N. Laird, and D. Rubin. *Maximum likelihood estimation from incomplete data.* Journal of the Royal Statistical Society (B): 39(1), (1977).

[27] Center for Research in Security Prices. <http://gsbwww.uchicago.edu/research/crsp/>. Accessed Nov., 2004.

[28] Wharton Research Data Services <https://wrds.wharton.upenn.edu/>. Accessed Nov., 2004.