High Resolution Algorithms for Spectral Analysis and Array Processing

by

Weixiu Du
B.S., Shandong Polytechnical University, China, 1983

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

We accept this dissertation as conforming to the required standard

Dr. R. Lynn Krin, Supervisor (Department of ECE)

Dr. A. Zielinski, Departmental Member (Department of ECE)

Dr. P. Driesen, Departmental Member (Department of ECE)

Dr. R. R. Davidson, Outside Member (Department of Mathematics)

Dr. R. Vandieck, Additional Member (Department of ECE)

Dr. A. R. Moghaddam, External Examiner (U. of Wisconsin)

© WEIXIU DU, 1992

University of Victoria

All rights reserved. Dissertation may not be reproduced in whole or in part. by photocopying or other means, without the permission of the author.
ABSTRACT

In this dissertation a novel covariance matrix estimator has been proposed for the covariance-matrix based high resolution spectral estimators. The proposed covariance matrix estimator can fully exploit cross correlations existent among distinct sets of random vectors drawn from different processes to obtain a more stable estimate of the covariance matrix from a short data record corrupted by additive noise. This estimator is derived from a theorem on the Least Squares Linear Prediction of one random vector from another random vector. The theorem can also be interpreted as estimating the auto-covariance matrix of the first random vector from the the cross-correlation matrix between the two random vectors and the auto-covariance matrix of the second random vector such that a given optimal criterion is satisfied. Applying this method in conjunction with a high resolution algorithm results in performance improvement of the spectral estimator.

The new covariance matrix estimator has been applied in the following three areas:

1. Spatial smoothing for the direction of arrival estimation in the presence of coherent signals.
2. Covariance enhancement by utilizing temporal correlations between array snapshot vectors.

Simulations show that the expected performance improvement can be achieved in terms of resolution, estimation errors and SNR threshold.

In addition to the covariance matrix estimator, we also present some other research results in array processing and seismic signal processing. A general transformation
matrix based on the vector p-norm has been proposed. This new transformation matrix provides options to satisfy different design specifications in array processing. Finally, the velocity estimation problem in seismic signal processing is discussed. The conventional semblance method is found to be the conventional beamforming method for a fixed two-way time. An optimal velocity estimator is proposed based on the Linearly Constrained Minimum Variance beamformer. The optimal velocity estimator demonstrates the high discrimination power (resolution) both for noise-free data and noisy data. When the new covariance estimator is used in conjunction with the optimal velocity estimator, we can achieve a resolution with deeper notch. This fact once more demonstrates the advantages of the proposed covariance matrix estimator.

Dr. R. Lynny Kirlin, Supervisor (Department of ECE)

Dr. A. Zelniski, Departmental Member (Department of ECE)

Dr. P. Driessen, Departmental Member (Department of ECE)

Dr. R. R. Davidson, Outside Member (Department of Mathematics)

Dr. R. Vahldieck, Additional Member (Department of ECE)

Dr. A. R. Moghaddasjoo, External Examiner (U. of Wisconsin)
Contents

Title Page i

Abstract ii

Table of Contents iv

List of Principle Symbols and Uses xi

List of Tables xiv

List of Figures xv

Acknowledgments xviii

1 Introduction 1

1.1 Spectral Estimation and Array Processing 1

1.1.1 Applications and Performance Measurement 2

1.2 Role of the Covariance Matrix 3

1.3 Summary of Other Work 4
CONTENTS

1.3.1 Sample Covariance Matrix ......................................................... 4
1.3.2 Maximum Likelihood Estimate of the Structured Covariance Matrix ........................................... 5
1.3.3 Periodicity Constrained Estimator ............................................. 5
1.3.4 Coherent Signal-subspace Method ............................................ 5
1.3.5 Steered Covariance Matrix ....................................................... 6
1.4 Contributions of the Dissertation ................................................... 6
1.5 Organization of the Dissertation .................................................... 7

2 Historic Review

2.1 Why High Resolution Spectral Estimators .................................... 9
  2.1.1 Classical Spectral Estimators .................................................. 9
  2.1.2 Limitation of Classical Spectral Estimators ............................... 10
2.2 Data Modeling and Assumptions .................................................... 11
  2.2.1 Array Processing ................................................................. 11
  2.2.2 Spectral Estimation .............................................................. 13
  2.2.3 Summary of Assumptions ....................................................... 14
  2.2.4 Sample Covariance Matrix .................................................... 14
2.3 High Resolution Spectral Estimators ............................................. 15
  2.3.1 Conventional Beamforming .................................................... 16
  2.3.2 Maximum Entropy Method (MEM) .......................................... 16
CONTENTS

2.3.3 Minimum Variance Method (MV) ............................................... 18
2.3.4 Pisarenko Method ................................................................. 18
2.3.5 MUSIC ................................................................................. 19
2.3.6 Root MUSIC ........................................................................ 20
2.3.7 Minimum Norm Method ....................................................... 21
2.3.8 ESPRIT and Matrix Pencil ............................................... 22
2.3.9 Maximum Likelihood (ML) Method ....................................... 23

2.4 Classifications of Spectral Estimators .............................................. 25
  2.4.1 Classification by Chronology ............................................. 26
  2.4.2 Classification by Model ..................................................... 26
  2.4.3 Classification by Vector Space .......................................... 27
  2.4.4 Classification by Numerical Procedure ............................ 28

2.5 Determination of the Number of Signals ....................................... 28

3 Covariance Estimation .................................................................. 30
  3.1 Introduction and Statistical Background .................................. 30
    3.1.1 Nuisance Parameters .................................................... 32
    3.1.2 Estimation Improvement by Averaging Multiple Estimates ...... 33
  3.2 LSLP of Random Vectors ...................................................... 34
  3.3 Improving the Estimates of Covariance Matrices by Incorporating
      Cross-correlations ..................................................................... 36
  3.4 Applications to Covariance-based Spectral Estimators .............. 38
CONTEN.

4 Applications 42

4.1 Spatial Smoothing 42

4.1.1 Introduction 42

4.1.2 Spatial Smoothing for the Subspace-based Method 44

4.1.3 Improved Spatial Smoothing Methods 45

4.1.4 Simulation Results 53

4.2 Covariance Improvement by Taking Advantage of Temporal Correlations 55

4.2.1 Introduction 55

4.2.2 Estimating Covariance Matrices by Exploiting Temporal Correlations 57

4.2.3 Matrix Enhancement Using Temporal Correlations 59

4.2.4 Enhancement for Coherent Signals 60

4.2.5 Extension to Wideband Signals 61

4.2.6 Simulation Results 62

4.3 Spectral Estimation 64

4.3.1 Introduction 64

4.3.2 Matrix Telescope Series 67

4.3.3 Proposed Method 67

4.3.4 Order Selection 70

4.3.5 Simulation Results 72
CONTENTS

5 Design of Transformation Matrices ...................................................... 76
   5.1 Introduction ................................................................. 76
   5.2 Background ................................................................. 78
       5.2.1 Vector P-norm and Frobenius Norm ................. 78
       5.2.2 Sector Processing and Interpolated Arrays ....... 79
   5.3 P-norm Method ............................................................ 81
       5.3.1 Implementation .................................................. 82
   5.4 Simulation Results ..................................................... 84

6 Seismic Signal Processing ................................................................. 86
   6.1 Introduction ................................................................. 86
   6.2 Modeling Seismic Signals ............................................... 87
       6.2.1 Wavefront Models ............................................. 87
       6.2.2 Wave Shape Model ........................................... 89
   6.3 Velocity Estimation ..................................................... 90
       6.3.1 Current Techniques ............................................ 90
       6.3.2 Modifications and Improvements Over the Previous Methods 95
   6.4 The Relationship Between Semblance and Eigenstructure Velocity
       Estimators ................................................................. 97
       6.4.1 Semblance and Conventional Beamforming .......... 97
       6.4.2 Optimal Beamformers ......................................... 98
   6.5 Optimal Velocity Estimators ........................................ 99
CONTENTS

6.5.1 Some Special Considerations .................................................... 99
6.5.2 Enhancement of the Estimates of Covariance Matrices ... 100
6.5.3 LCMV Velocity Estimator ..........................................................101
6.6 Simulations .................................................................
   6.6.1 Comparison of Coherency Measure Threshold Discrimination 103
   6.6.2 Handling Noisy Seismic Data ........................................... 105

7 Summary and Conclusions ............................................................ 122
   7.1 Introduction ............................................................................. 122
   7.2 Summary of the Dissertation .................................................. 122
   7.3 Conclusions and Future Work ................................................. 124

A Matrix Derivative Rules ............................................................. 126
   A.1 Derivation of Vectors ............................................................ 126
      A.1.1 Definitions ....................................................................... 126
      A.1.2 Chain Rule .................................................................... 127
      A.1.3 Some Useful Formulas .................................................. 127
   A.2 Derivation of Matrices .......................................................... 127
      A.2.1 Definitions ....................................................................... 128
      A.2.2 Product and Chain Rules .............................................. 128
      A.2.3 Some Useful Formulas .................................................. 129

B Fisher Information Matrix .......................................................... 130
CONTENTS

B.1 Another Form for the Fisher Information Matrix .................. 130
B.2 Information Matrix for Gaussian Random Vectors ............... 130
B.3 Information Matrix for Array Processing .......................... 132

C Ricker Wavelets ..................................................... 133
List of Principle Symbols and Uses

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\partial}{\partial \theta_i}$</td>
<td>partial derivative w.r.t. $\theta_i$</td>
</tr>
<tr>
<td>$| \cdot |_F$</td>
<td>matrix Frobenius norm</td>
</tr>
<tr>
<td>$| \cdot |_p$</td>
<td>vector p-norm</td>
</tr>
<tr>
<td>$(\cdot)^#$</td>
<td>pseudo-inverse of a matrix</td>
</tr>
<tr>
<td>$(\cdot)^{-1}$</td>
<td>inverse of a matrix</td>
</tr>
<tr>
<td>$(\cdot)^*$</td>
<td>complex conjugate</td>
</tr>
<tr>
<td>$(\cdot)^H$</td>
<td>Hermitian transpose of a matrix</td>
</tr>
<tr>
<td>$\delta_{ij}$</td>
<td>Kronecker delta</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>sensor spacing</td>
</tr>
<tr>
<td>$\Delta R_i$</td>
<td>perturbation of $\hat{R}_i$</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>wavelength</td>
</tr>
<tr>
<td>$\lambda_i$</td>
<td>$i$th eigenvalue of covariance matrix $R$</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>eigenvalue matrix</td>
</tr>
<tr>
<td>$\mu$</td>
<td>mean vector</td>
</tr>
<tr>
<td>$\otimes$</td>
<td>Kronecker product</td>
</tr>
<tr>
<td>$\omega$</td>
<td>frequency in rad</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>noise variance</td>
</tr>
<tr>
<td>$\theta_i$</td>
<td>bearing angle of the $i$th signal source</td>
</tr>
<tr>
<td>$\varphi, \theta$</td>
<td>parameter to be estimated</td>
</tr>
<tr>
<td>$\mathbf{a}(\theta)$</td>
<td>steering vector</td>
</tr>
<tr>
<td>$\mathbf{A}(\theta)$</td>
<td>steering matrix</td>
</tr>
<tr>
<td>AR</td>
<td>auto-regressive</td>
</tr>
<tr>
<td>ARMA</td>
<td>auto regressive-moving average</td>
</tr>
<tr>
<td>$\mathbf{B}$</td>
<td>steering matrix for the virtual array</td>
</tr>
<tr>
<td>$c, v$</td>
<td>wave velocity of the medium</td>
</tr>
<tr>
<td>CDP</td>
<td>common depth point</td>
</tr>
<tr>
<td>CSM</td>
<td>coherent signal-subspace method</td>
</tr>
<tr>
<td>$d$</td>
<td>number of signals</td>
</tr>
<tr>
<td>$d(x, t)$</td>
<td>measured seismogram</td>
</tr>
<tr>
<td>$\text{det}[\cdot]$</td>
<td>determinate of a matrix</td>
</tr>
<tr>
<td>DFT</td>
<td>discrete Fourier transform</td>
</tr>
<tr>
<td>DOA</td>
<td>direction of arrivals</td>
</tr>
<tr>
<td>ESPRIT</td>
<td>Estimation of signal parameter via rotation invariance technique</td>
</tr>
<tr>
<td>$f_p$</td>
<td>peak frequency of the Ricker wave spectrum</td>
</tr>
<tr>
<td>FFT</td>
<td>fast Fourier transform</td>
</tr>
<tr>
<td>FIR</td>
<td>finite impulse response</td>
</tr>
</tbody>
</table>
### LIST OF SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$</td>
<td>identity matrix</td>
</tr>
<tr>
<td>$I_{M,N}(k)$</td>
<td>$[0_{M \times k}, I_M, 0_{M \times (N-M-k)}]$</td>
</tr>
<tr>
<td>$I_{M,N}(k)$</td>
<td>$[I_j, 0_{j \times M}, 0_{j \times (N-M-j)}]$</td>
</tr>
<tr>
<td>$J(\varphi_D)$</td>
<td>Fisher information matrix as a function of $\varphi_D$</td>
</tr>
<tr>
<td>$J_e$</td>
<td>reflection or exchange matrix</td>
</tr>
<tr>
<td>$K_c$</td>
<td>coherency measure by Key's method</td>
</tr>
<tr>
<td>LCMV</td>
<td>Linearly constrained minimum variance beamformer</td>
</tr>
<tr>
<td>LP</td>
<td>linear prediction</td>
</tr>
<tr>
<td>LS</td>
<td>least squares</td>
</tr>
<tr>
<td>$M$</td>
<td>number of sensors</td>
</tr>
<tr>
<td>MA</td>
<td>moving-average</td>
</tr>
<tr>
<td>MEM</td>
<td>Maximum entropy method</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td>MV</td>
<td>Minimum variance</td>
</tr>
<tr>
<td>MSE</td>
<td>mean squared error</td>
</tr>
<tr>
<td>$n(x,t)$</td>
<td>measurement noise in seismogram</td>
</tr>
<tr>
<td>$p$</td>
<td>root mean squared slowness</td>
</tr>
<tr>
<td>$P$</td>
<td>signal covariance matrix</td>
</tr>
<tr>
<td>$P_A$</td>
<td>projection operator of matrix $A$</td>
</tr>
<tr>
<td>$P(\omega)$</td>
<td>power spectrum density</td>
</tr>
<tr>
<td>rank[]</td>
<td>rank of a matrix</td>
</tr>
<tr>
<td>$r_i$</td>
<td>$i$th root of a polynomial</td>
</tr>
<tr>
<td>$r(k)$</td>
<td>autocorrelation function</td>
</tr>
<tr>
<td>$R$</td>
<td>covariance matrix</td>
</tr>
<tr>
<td>$R(t_0, v_s)$</td>
<td>in-gate covariance matrix</td>
</tr>
<tr>
<td>$\hat{R}^2(t_0, v_s)$</td>
<td>improved in-gate covariance matrix (squared)</td>
</tr>
<tr>
<td>$R_a$</td>
<td>covariance matrix of $x_a(t)$</td>
</tr>
<tr>
<td>$R_f$</td>
<td>forward spatially smoothed covariance matrix</td>
</tr>
<tr>
<td>$R_{fb}$</td>
<td>forward-backward spatially smoothed covariance matrix</td>
</tr>
<tr>
<td>$\hat{R}_f^2$</td>
<td>forward spatially smoothed covariance matrix (squared) with improvement</td>
</tr>
<tr>
<td>$\hat{R}_{fb}^2$</td>
<td>forward-backward spatially smoothed covariance matrix (squared) with improvement</td>
</tr>
<tr>
<td>$R_{ij}$</td>
<td>forward cross-covariance matrix of $x_i(t)$ and $x_j(t)$</td>
</tr>
<tr>
<td>$R_{ji}$</td>
<td>backward cross-covariance matrix of $x_i(t)$ and $x_j(t)$</td>
</tr>
<tr>
<td>$\hat{R}_{ii}$</td>
<td>improved estimate of $\hat{R}_{ii}$</td>
</tr>
<tr>
<td>$\hat{R}_{ii}^*$</td>
<td>estimate of $\hat{R}_{ii}$ (squared) from cross correlations</td>
</tr>
<tr>
<td>$\hat{R}_{ii}^{*r}$</td>
<td>improved estimate of $\hat{R}_{ii}$ (squared)</td>
</tr>
</tbody>
</table>
LIST OF SYMBOLS

- $s$ — signal vector
- $s_i(t)$ — $i$th incident signal
- span{·} — vector space spanned by columns of a matrix
- $S_c$ — semblance coefficients
- $S_{\text{enhanced}}$ — visually enhanced semblance coefficients
- $S$ — signal matrix
- SNR — signal to noise ratio
- $t_0$ — two-way travel time for zero-offset trace
- $t_x$ — two-way travel time for a trace of offset $x$
- $\text{tr}(\cdot)$ — trace of a matrix
- $T$ — linear transformation matrix
- $U(p, t_0)$ — hyperbolic transform coefficients
- $U, V$ — singular vector matrix
- $v_i$ — $i$th eigenvector of the covariance matrix $R$
- $v_{\text{rms}}$ — root mean squared velocity
- $V_s$ — signal eigenvector matrix
- $V_n$ — noise eigenvector matrix
- $W$ — noise matrix
- $x$ — trace offset
- $x(t), y(t)$ — observed data (vector)
- $x_a(t)$ — augmented sample vector
- $X$ — data matrix
List of Tables

6.1 Events parameters for noise-free data .......................... 105
6.2 Events parameters for noisy data ............................... 115
List of Figures

4.1 Translational equivalent subarrays .............................................. 43

4.2 Partitioning a uniform linear array into subarrays ......................... 49

4.3 MUSIC spectra using the conventional and the improved forward-
backward spatial smoothing methods .............................................. 54

4.4 Root MSE vs. SNR for the conventional and the improved spatial
smoothing methods ......................................................................... 56

4.5 Formation of augmented sample vectors ....................................... 58

4.6 MUSIC spectra using the conventional and the enhanced covariance
matrix estimation methods (using temporal correlations) ................. 63

4.7 Root MSE vs. SNR for the conventional and the enhanced covari-
ance matrix estimation methods (using temporal correlations) .......... 65

4.8 Root MSE vs. SNR for augmented sample vectors of different sizes 66

4.9 Telescoping series of a matrix ..................................................... 68

4.10 Pisarenko spectra using the conventional and the new covariance
matrix estimators (frequency estimation) ........................................ 73

4.11 Root MSE vs. SNR for the conventional and the new covariance
estimators (frequency estimation) ................................................ 74
LIST OF FIGURES

4.12 The average Root MSE vs. order of the covariance matrix........75

5.1 Convergence of the Newton’s method.................................85

6.1 A model for the reflection seismology system.................88
6.2 An hypothetical CDP gather........................................91
6.3 Forming a time gate for semblance analysis................93
6.4 Noise-free synthetic CDP gather for discrimination threshold test. 106
6.5 Velocity contour using semblance (clean data)...............107
6.6 Velocity contour for LCMV with spatial smoothing (clean data) . 107
6.7 Marginal plot for region 1 (clean data).......................108
6.8 Marginal plot for region 2 (clean data).......................109
6.9 Marginal plot for region 3 (clean data).......................110
6.10 Velocity contour using the display-enhanced semblance (clean data) 111
6.11 Marginal plot of the display-enhanced semblance velocity spectrum for region 1 (clean data) 111
6.12 Marginal plot of the display-enhanced semblance velocity spectrum for region 2 (clean data) 112
6.13 Marginal plot of the display-enhanced semblance velocity spectrum for region 3 (clean data) 112
6.14 Noisy synthetic CDP gather........................................115
6.15 Velocity contour for semblance (noisy data)................116
LIST OF FIGURES

6.16 Velocity contour for Key’s method (noisy data) ........................................ 116

6.17 Velocity contour for LCMV with spatial smoothing (noisy data) .... 117

6.18 Velocity contour for LCMV with the improved spatial smoothing
(noisy data) ............................................................................................................ 117

6.19 Marginal plot of the semblance velocity spectrum in region 3 (noisy
data) ...................................................................................................................... 118

6.20 Marginal plot of the velocity spectrum by the LCMV with the con­
ventional spatial smoothing in region 3 (noisy data) .............................. 118

6.21 Marginal plot of the velocity spectrum by the LCMV with the im­
proved spatial smoothing in region 3 (noisy data) ................................. 119

6.22 Velocity contour for the display-enhanced semblance (noisy data) .... 119

6.23 Marginal plot of the display-enhanced semblance velocity spectrum
for region 3 (noisy data) .................................................................................. 120

6.24 Overlaid marginal plot of the semblance velocity spectra in region
3 (noisy data) .................................................................................................... 120

6.25 Overlaid marginal plots of the velocity spectra by the LCMV with
the conventional spatial smoothing in region 3 (noisy data) .............. 121

6.26 Overlaid marginal plots of the velocity spectra by the LCMV with
the improved spatial smoothing in region 3 (noisy data) ................. 121
Acknowledgments

I wish to express my sincere gratitude to my supervisor, Dr. R. Lynn Kirlin for his constant encouragement, guidance, support and assistance throughout the dissertation research. His deep insight in array signal processing and invaluable suggestions made this research all possible.

I would also like to thank Drs. A. Zielinski, P. Driessen, R. R. Davidson and R. Vahldieck for their services as my supervisory committee members, and Dr. A. R. Moghaddamjoo for his service as the external examiner in my Ph. D. oral examination.

My acknowledgments are also due to all the faculty members and staff in the ECE department. The help they offered to me during my Ph. D. program is deeply appreciated.

This dissertation is dedicated to my parents, Mr. Xingwu Du and Mrs. Yishu Zhang, my wife Jessica, and my son Melvin. Their love and care made it worthwhile.

This research is partly supported by the University of Victoria Fellowship, and the Petch Research Fellowship.
Chapter 1

Introduction

1.1 Spectral Estimation and Array Processing

Spectral estimation or spectral analysis has for several decades been a traditional research area for statisticians. The introduction of the fast Fourier transform (FFT) algorithm two decades ago, however, has expanded the role of spectral estimation from one of research novelty to one of practical utility. The digital signal processing community has, as a result, taken an ever-increasing interest in spectral estimation research and applications. In the signal processing community a spectral estimator is referred to as any signal processing method which characterizes the frequency content of a measured signal. The practical problems in spectral analysis are often to obtain an estimate of the true spectrum from a finite interval of observation which is usually corrupted by additive noise.

Array processing deals with the processing of signals carried by propagating wave phenomena. The received signal is obtained by a sensor array located in the field of interest. The aim of array processing is from the sensed data to extract useful information about the received signal field (e.g. its signature, direction of arrival (DOA), speed of propagation). Although spectral estimation and array processing appear to be distinct, many of the algorithms proposed are equally ap-
applicable to both problems [1]. For this reason, two problems can often be discussed in a unified approach.

Several recent books by Kay [2] Marple [3], Haykin [4, 5] provide an excellent review on these topics.

1.1.1 Applications and Performance Measurement

Spectral analysis and array processing have a wide range of applications such as system identification, exploration seismology, sonar, radar, radio astronomy and tomography. The measurement of the performance for a spectral estimator depends on applications. Some useful statistical measurements of spectral estimators are bias, error covariance, resolution and signal to noise (SNR) threshold. Let $\varphi \in \mathbb{C}^{d \times 1}$ be the true parameter vector and $\hat{\varphi}$ be the estimate obtained using a spectral estimator from a finite observation corrupted with noise. Then we define bias, error covariance, resolution and SNR threshold as follows, with $E$ denoting ensemble or expected value and the superscript $H$ denoting matrix Hermitian transpose.

**Definition 1.1 (Bias):** The bias of an estimate by a spectral estimator is defined by the statistical average of the difference between the estimated parameter vector $\hat{\varphi}$ and the true parameter vector $\varphi$ or

$$\Delta \varphi = E[\hat{\varphi} - \varphi] = E[\hat{\varphi}] - \varphi.$$  \hfill(1.1)

**Definition 1.2 (Error Covariance):** The covariance of the estimation error associated with a spectral estimator is defined by the covariance of the error vector $\hat{\varphi} - \varphi$ or

$$Cov(\hat{\varphi}) = E\{(\hat{\varphi} - E(\hat{\varphi}))(\hat{\varphi} - E(\hat{\varphi}))^H\}$$ \hfill(1.2)
CHAPTER 1. INTRODUCTION

For the unbiased estimator, we have

\[ \text{Cov}(\hat{\varphi}) = E\{(\hat{\varphi} - \varphi)(\hat{\varphi} - \varphi)^H\} \]  

(1.3)

**Definition 1.3 (Resolution):** The capability of the spectral estimator to resolve very close frequency components is referred to as resolution, which is commonly quantified with its probability.

**Definition 1.4 (SNR Threshold):** The SNR above which the spectral estimator can achieve expected resolution is denoted the SNR threshold.

### 1.2 Role of the Covariance Matrix

The univariate Gaussian or normal distribution occupies a central position in the statistical theory of analyzing random variables since this statistical model is suitable for such a large number of cases when multiple measurements are treated. This situation is even more pronounced in multivariate analysis due to the paucity of analytically tractable multivariate distributions — one notable exception being the multivariate Gaussian distribution. Consider a complex Gaussian random vector \( x \in C^{M \times 1} \) with mean vector \( \mu = E[x] \) and nonsingular Hermitian covariance matrix \( R = E[(x - \mu)(x - \mu)^H] \). Then the density of the \( x \) is

\[
f(x) = \frac{1}{\pi^M \det(R)} \exp \left[ -((x - \mu)^H R^{-1}(x - \mu)) \right]
\]

(1.4)

For Gaussian random vectors, mean vectors and covariance matrices are sufficient statistics, which can be used to uniquely determine the distribution density functions. Further, in many practical situations, mean vectors of multivariate random processes are either zero or can be removed from the data vectors, thus covariance matrices become the only statistics to be estimated for Gaussian random processes.
Similarly, covariance matrices also play a central role in high resolution algorithms developed for spectral estimation and array processing since we usually use the Gaussian distribution to model random vectors obtained in these applications. As a matter of fact, many of these high resolution algorithms require a two step procedure. First an estimate of the covariance matrix is obtained from the given data set; then the spectral analysis algorithms are applied. The results of the second step are highly sensitive to the quality of the results of the first step. For stationary random processes, the estimate of the covariance matrix with required stability can be obtained by the sample covariance matrix using a sufficient long observation time period. In theory these estimates (sample covariance matrices) approach the true covariance matrices with probability one when the observation time is infinitely long, assuming that the random processes are stationary and ergodic. However, in practice the observation time is usually limited since either the random processes can only be treated as stationary in a limited time interval or the available data samples from the random process are finite. Thus a practical problem is to find the best possible estimator for a covariance matrix from a finite observation.

1.3 Summary of Other Work

1.3.1 Sample Covariance Matrix

The sample covariance matrix has been widely used in the estimation of the covariance matrix of multivariate random processes. It is shown that the sample covariance matrix is the maximum likelihood (ML) estimate of the true covariance for i.i.d. Gaussian multivariate random vectors assuming that the true covariance matrix does not have any specific structure other than symmetry [6]. This estimator is no longer optimal when additional a priori information regarding the structure of the covariance matrix or signal sources are available.
1.3.2 Maximum Likelihood Estimate of the Structured Covariance Matrix

In the case of a uniform linear array and uncorrelated sources, we have additional \textit{a priori} structural information of the covariance matrix. One such piece of information is the Toeplitz structure of the covariance matrix. To exploit this information, Burg \textit{et al} [7] derived an iterative algorithm to compute the ML estimate of a Toeplitz constrained covariance matrix.

1.3.3 Periodicity Constrained Estimator

Another piece of structural information is the periodicity of the spatial covariance function for a uniform linear array and uncorrelated sources. Ziskind and Wax [8] proposed a scheme to exploit the periodicity property of the spatial covariance function existent in these scenarios. When applying the periodicity constrained estimation method in conjunction with the high resolution algorithm, the resolution enhancement can be achieved especially in the threshold region.

1.3.4 Coherent Signal-subspace Method

In the scenarios of temporally wide-band sources Wang and Kaveh [9] proposed the Coherent Signal-subspace Method (CSM) to more effectively exploit larger source bandwidths. In CSM, focusing matrices are used to reduce multiple narrow-band covariance estimates made over the receiver bandwidth to a single (hopefully) focused covariance matrix. CSM offers a significant improvement over the incoherent methods in the wide-band setting.
1.3.5 Steered Covariance Matrix

Also in the wide-band setting, Krolik and Swingler [10] proposed to use steered covariance matrices. It is shown that the steered covariance matrix has the advantage that it can be estimated with much greater statistical stability for a given sample size and exhibits lower SNR threshold.

1.4 Contributions of the Dissertation

In this dissertation a novel covariance matrix estimator [11, 12] will be presented. This estimator takes advantages of cross correlations existent among distinct sets of random vectors drawn from different processes. For example, correlations can be those between outputs from different subarrays in spatial smoothing, or between snapshots of the array outputs or even between time sliced vectors from a data sequence in spectral estimation. Applying this method in conjunction with a given high resolution algorithm results in enhanced performance improvement for the spectral estimator. The improved covariance estimator is derived from the theorem on the least squares linear prediction of multivariate random processes; it is further simplified for covariance-based (and especially subspace-based) high resolution spectral estimators. It has applications in the following spectral analysis and array processing areas, and some of these applications will be studied in detail later:

1. Spatial smoothing in DOA estimation.
2. Spectrum estimation in time sequence analysis.
3. Exploiting temporal correlations in array processing
4. Beamforming
5. Velocity estimation in seismic processing
6. Focusing transformation design in CSM

In addition to the new covariance matrix estimator we will also present in this dissertation some other work we have done in the areas of spectral analysis and array processing.

1.5 Organization of the Dissertation

This dissertation is organized as follows:

- Chapter 2: provides a historic review of major classical and modern spectral estimation methods. A unified signal and noise model is established for the cases of both array processing and spectral analysis.

- Chapter 3: presents a covariance matrix estimation method which can fully exploit the cross correlations of multivariate random vectors. The estimator is further simplified for covariance-based high resolution spectral estimators, where signal subspace is of dominant importance. Asymptotically there is no difference between the new covariance matrix estimator and other estimators, but the proposed method shows its advantage over the sample covariance method when the observation time is short. As a consequence of applying the new method, the performance of the following high resolution algorithms can be enhanced. Even though for a specific spectral estimator a better covariance matrix estimate will yield simultaneously smaller error variance, lower SNR threshold, and higher resolution, any one of these measures justifies the method.

- Chapter 4: applies the theory developed in Chapter 3 to different areas of array processing and spectral analysis. For each application the theory of Chapter 3 is adapted to the specific problem; simulation results verify the achievement of the expected goal.
• *Chapter 5:* presents a transformation matrix method based on the vector $p$-norm which generalizes the commonly used Frobenius norm method.

• *Chapter 6:* is focused on the application of the modern array processing methodology to reflection seismic signal processing. Especially the velocity estimation problem has been studied.

• *Chapter 7:* summarizes and concludes this dissertation.
Chapter 2

Historic Review of Spectral Estimators

2.1 Why High Resolution Spectral Estimators

2.1.1 Classical Spectral Estimators

Classical digital spectral estimation methods are usually based on the Discrete Fourier Transform (DFT) or the Wiener-Khintchine theorem. Suppose we observe a stationary random process for a very long time, so that we obtain a time series $y(n)$ for $n = 1, 2, \ldots, N$, where $N$ is very large. Then the associated power spectrum can be obtained by the periodogram

$$P(\omega) = \frac{1}{N}|Y(\omega)|^2$$  \hspace{1cm} (2.1)

where $Y(\omega)$ is the DFT of $y(n)$

$$Y(\omega) = \sum_{n=1}^{N} y(n)e^{-j\omega n}.$$
Today $Y(\omega)$ is computed very rapidly by means of the Cooley-Tukey fast Fourier transform (FFT) [13].

Wiener in his 1930 paper [14] gave the following method, which was standard until the work of Tukey in 1949 [15]. Wiener’s method was intended for very long time series. It consisted of computing the autocorrelation function as the time average

$$r(k) = \frac{1}{N} \sum_{n} y^*(n)y(n + k),$$

with the asterisk star * denoting complex conjugate, for $-p \leq k \leq p$, where $p$ less than the data length $N$, and then computing the power spectrum $P(\omega)$ as the Fourier transform

$$P(\omega) = \sum_{k=-\infty}^{\infty} r(k)e^{-j\omega k}, \quad (2.2)$$

where $r(k)$ is truncated after $p$ lags. This Fourier transform relationship between the autocorrelation and the power spectrum is now called the Wiener-Khintchine theorem.

2.1.2 Limitation of Classical Spectral Estimators

In spite of the computational efficiency of the FFT, resolutions of these spectral estimators are limited to the reciprocal of temporal or spatial apertures of the recorded data. When high resolution spectral analysis is required, the data record length should be long in the case of time series analysis, or the sensor array should cover a large spatial range in the case of array processing. As we mentioned earlier, in practical reality large temporal or spatial aperture may be difficult to obtain. For example the sizes of sensor arrays may not be easily extended; thus
the spatial aperture is often much smaller than required to obtain desired high resolution using the conventional methods. Further, the resolution of a spectral estimator is in some sense closely related to the capability of detection of signals by the estimator. Increasing signal detection capability is a common need in radar, sonar and many other array processing applications. This need has been motivating the development of so-called high resolution spectral estimators. Discussions of the dissertation will be focused on recently developed high resolution spectral estimators.

2.2 Data Modeling and Assumptions

Since spectral estimation and array processing are equivalent in many aspects, the signal models used in both applications can be characterized in a unified approach. In this section the signal model for bearing (DOA) estimation problem in array processing is considered first; later it is shown that this signal model will be equally applicable to spectral estimation. To simplify the complexity of formulation, we mostly consider one-dimensional estimation problems from either temporally or spatially uniformly sampled data. The generalization of the one-dimensional and uniformly sampled case to multi-dimensional and non-uniformly sampled cases is a straightforward extension of the simplified case.

2.2.1 Array Processing

Consider a uniform linear sensor array composed of $M$ omni-directional sensors immersed in a homogeneous dispersive medium with wave propagation velocity $c$. Assume that $d$ narrow-band sources ($d < M$) centered at frequency $\omega_0$ are impinging on the array from directions $\theta_1, \theta_2, \ldots, \theta_d$ and are associated with signals $s_1(t), s_2(t), \ldots, s_d(t)$. Assume further that the signals emitted by the sources are ergodic and stationary narrow-band stochastic processes. Then the array output
vector is specified by the linear combination of steering vectors, which are mutually linearly independent

\[ x(t) = \sum_{k=1}^{d} s_k(t)a(\theta_k) + w(t), \]  

(2.3)

where \( w(t) \) is the additive noise vector and \( a(\theta_k) \) is the steering vector for the \( k \)th signal, which is given by

\[ a(\theta_k) = [1, e^{j\omega_0\tau_k}, e^{j2\omega_0\tau_k}, \ldots, e^{j(M-1)\omega_0\tau_k}]^T, \]  

(2.4)

where \( \tau_k = \Delta \sin \theta_k/c \) with \( c \) denoting the propagation speed of the signal in the media, with \( \Delta \) representing sensor spacing satisfying condition \( \Delta \leq \lambda/2 \), and with \( \lambda \) denoting the wavelength. We use superscripts \( T \) and \( -1 \) to represent respectively matrix transpose and inverse operations. \( a(\theta) \) is also called the direction vector or array manifold. We use the symbol \( \theta \) without a subscript to represent a possible direction of arrival and the subscripted symbol \( \theta_k, k = 1, 2, \ldots, d \), to represent the true direction of arrival in the noise-free data. Alternatively equation (2.3) can be written by a vector-matrix form

\[ x(t) = As(t) + w(t), \]  

(2.5)

where \( A = [a(\theta_1), a(\theta_2), \ldots, a(\theta_d)] \) is the steering matrix, and \( s(t) = [s_1(t), s_2(t), \ldots, s_d(t)]^T \). The additive noise is assumed to be a stationary, zero-mean random process that is temporally and spatially white and uncorrelated with the signals. With these assumptions we get the covariance matrix of the array output vectors

\[ R = E[x(t)x^H(t)] = APA^H + \sigma^2 I, \]  

(2.6)

where \( P = E[s(t)s^H(t)] \) is the signal covariance matrix and \( \sigma^2 I = E[w(t)w^H(t)] \) is the noise covariance matrix. When \( P \) is a \( d \times d \) diagonal matrix, the signals
are uncorrelated; when \( P \) is nonsingular and nondiagonal, the signals are partially correlated; and when \( P \) is singular, at least two signals are coherent, (that is, at least one source is a scaled and delayed version of the other source(s)).

In the above, \( t \) should be replaced with \( t_i, i = 1, 2, \ldots, N \) when the array output is digitized, and \( N \) denotes the number of array output vectors available. We also call \( N \) the number of snapshots.

### 2.2.2 Spectral Estimation

Let a scalar record of data sequence \( y(t_i), i = 1, 2, \ldots, N' \) be composed of a uniformly sampled sum of \( d \) cisoid signals. \( N = N' - M + 1 \) data vectors \( x(t_i) \in \mathbb{C}^{M \times 1} \) can be formed by

\[
x(t_i) = [y(t_i), y(t_i + 1), \ldots, y(t_i + M - 1)]^T = \sum_{k=1}^{d} a(\omega_k) s_k(t_i) + w(t_i),
\]

where \( a(\omega_k) \) and \( w(t_i) \) have the similar definition as in the case of array processing, i.e. \( a(\omega_k) = [1, e^{j\omega_k}, e^{j2\omega_k}, \ldots, e^{j(M-1)\omega_k}]^T \) with \( \omega_k \) denoting the frequency in \( \text{rad/s} \) of the \( k \)th cisoid, and \( w(t_i) \in \mathbb{C}^{M \times 1} \) is the additive noise vector formed in the same way as \( x(t_i) \). The matrix form of equation (2.7) is given by

\[
x(t_i) = As(t_i) + w(t_i),
\]

where \( A = [a(\omega_1), a(\omega_2), \ldots, a(\omega_d)] \) is the signal model matrix, and \( s(t_i) = [s_1(t_i), s_2(t_i), \ldots, s_d(t_i)]^T \). With the same assumptions on the additive noise, we have

\[
R = E[x(t_i)x^H(t_i)] = APA^H + \sigma^2 I.
\]

Comparing equations in the above two subsections, we conclude the DOA estimation and frequency estimation problems are equivalent in the form of signal
modeling and formulation. Due to the equivalence, either spectral estimation or the DOA estimation or both are used in problem formulation.

### 2.2.3 Summary of Assumptions

We summarize assumptions often imposed in array processing and spectral analysis, and/or in the previous signal and noise modeling:

- The number of signals is less than the number of sensors, namely, \( d < M \).
- The geometry of the sensor array is such that \( a(\theta_k) \) for \( k = 1, 2, \ldots, d \) are linearly independent. For uniform linear arrays and uniformly sampled time sequences, this assumption is automatically satisfied due to the Vandermonde structure of the \( A \) matrix.
- The noises \( w(t_i) \) are independent of signals \( s_k(t_i) \) and of each other, and are identically distributed complex, zero-mean, Gaussian vectors with covariance matrix \( \sigma^2 I \), where \( \sigma^2 \) is an unknown scalar.
- The sensor spacing \( \Delta \) is less than or equal to half of the smallest wavelength to avoid spatial aliasing effects; namely, \( \Delta \leq \lambda_{\text{min}}/2 \).

### 2.2.4 Sample Covariance Matrix

In a practical digital system a true covariance matrix of array output vectors is not available; only an estimate is obtained using \( N \) sampled output vectors \( x(t_i) \):

\[
\hat{R} = \frac{1}{N} \sum_{i=1}^{N} x(t_i)x(t_i)^H.
\]  

(2.10)

\( \hat{R} \) is termed the *sample covariance matrix* of \( x(t_i) \), where we use the caret sign "~" to distinguish the estimate from the true value.
CHAPTER 2. HISTORIC REVIEW

2.3 High Resolution Spectral Estimators

With the estimate of the array output covariance matrix given, we are able to present so-called high resolution spectral estimators in terms of the estimated covariance matrix. In his recent paper, Kirlin [16] formulated a unified presentation for several recently developed high resolution algorithms based on the eigen-decomposition of $\hat{R}$. It is well known that the covariance matrix $\hat{R}$ is positive (semi-)definite, thus it can be written in terms of its ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$ and associated eigenvectors $\hat{v}_i, i = 1, 2, \ldots, M$:

$$\hat{R} = \hat{V} \hat{\Lambda} \hat{V}^H = \sum_{i=1}^{M} \lambda_i \hat{v}_i \hat{v}_i^H,$$

(2.11)

where $\hat{V}$ is an $M \times M$ unitary matrix composed of $M$ eigenvectors

$$\hat{V} = [\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_M],$$

and $\hat{\Lambda}$ is a diagonal matrix formed with elements of $\lambda_i$

$$\hat{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M).$$

Similarly the inverse of $\hat{R}$ can be written as:

$$\hat{R}^{-1} = \sum_{i=1}^{M} \lambda_i^{-1} \hat{v}_i \hat{v}_i^H.$$

(2.12)

Based on the eigen-decomposition of $\hat{R}$, we can partition signal and noise subspace in terms of eigenvectors, or

$$\hat{R} = [\hat{V}_s \hat{V}_n] \begin{bmatrix} \hat{\Lambda}s & 0 \\ 0 & \hat{\Lambda}_n \end{bmatrix} \begin{bmatrix} \hat{V}_s^H \\ \hat{V}_n^H \end{bmatrix},$$

(2.13)
where \( \hat{V}_s \) composed of the first \( d \) eigenvectors forms the signal subspace and \( \hat{V}_n \) composed of the remaining eigenvectors forms the noise subspace, respectively. Similarly \( \hat{\Lambda}_s \) is the eigenvalues associated with signal subspace and \( \hat{\Lambda}_n \) associated with the noise subspace.

The caret sign \(^\wedge\) may be neglected for notation brevity in the remaining part of the dissertation if it can be inferred from context or if the estimate does not need to be distinguished from the true value.

### 2.3.1 Conventional Beamforming

The conventional method of mapping the monochromatic field power as a function of signal arriving angle is the beamforming operation, which employs a procedure known as delay-and-sum processing to steer a beam in a particular direction. Given measurements from a sensor array and the field of view \( \Theta \), the beamformer scans the region of interests in which sources may be present and calculates array output power

\[
P_{BF}(\theta) = a_H(\theta)R_a(\theta)\frac{a_H(\theta)a(\theta)}{a_H(\theta)a(\theta)}. \tag{2.14}
\]

The peaks of the spectrum represent the estimates of DOA. The conventional beamforming is computationally simple, but it has poor performance at low SNR or in presence of multiple sources.

### 2.3.2 Maximum Entropy Method (MEM)

The Wiener-Khintchine theorem is not suitable for short data records since the autocorrelation function is truncated after \( p \) lags. This results in a smoothing effect in the frequency domain. When the \( p \) autocorrelation sequence \( r(1), r(2), \ldots, r(p) \)
CHAPTER 2. HISTORIC REVIEW

is assumed known, then a question may be posed as to how the remaining unknown lags $r(p+1), r(p+2), \ldots$ should be specified instead of arbitrarily setting them to zero. Motivated by the results of information theory, Burg [17] argued that the extrapolation should be made in such a way as to maximize the entropy of the time series characterized by the extrapolated autocorrelation sequence. This time series would then be most random, in an entropy sense, of all series that have the known first $p$ lags. He proposed the maximum entropy method which begins with maximizing the entropy rate of a Gaussian random process

$$
\int_{-\pi/2}^{\pi/2} \ln P_{MEM}(f) df
$$

(2.15)

subject to the constraints

$$
\int_{-\pi/2}^{\pi/2} P_{MEM}(f)e^{i2\pi fk} df = r(k)
$$

(2.16)

over $0 \leq k \leq p$. The solution found by the Lagrange multiplier techniques can be written as follows:

$$
P_{MEM}(f) = \frac{1^T R^{-1} 1}{|1^T R^{-1} a(f)|^2},
$$

(2.17)

where $1^T = (1, 0, \ldots, 0, 0)$ and $R$ is the Toeplitz autocorrelation matrix of known lags.

It has been shown that for Gaussian random processes and a known autocorrelation sequence of uniform spacing the MEM spectrum is equivalent to that of order $p$ auto-regressive (AR) model based on linear prediction theory [3]. A good review on the relationship of MEM and AR model can be found in Papoulis [18].
2.3.3 Minimum Variance Method (MV)

The minimum variance method was developed by Capon [19]; it is also known as Capon's Maximum Likelihood method. In a filtering interpretation, the MV estimator is an Finite Impulse Response (FIR) digital filter which minimizes the variance of the array output (and thus interferences) under the constraint that its response to the specified direction is unity:

$$\min_{w} E[|w^H x(t)|^2] \quad \text{subject to} \quad w^H a(\theta) = 1.$$  \hspace{1cm} (2.18)

The optimal solution, easily found using Lagrange multiplier techniques, is

$$P_{MV}(\theta) = \frac{1}{a^H(\theta)k^{-1}a(\theta)}.$$  \hspace{1cm} (2.19)

2.3.4 Pisarenko Method

The Pisarenko method [20] is the first algorithm to employ a subspace of the estimated covariance matrix. In the Pisarenko method, we first calculate the \((d + 1) \times (d + 1)\) covariance matrix and \(v_{d+1}\), the eigenvector associated with the smallest eigenvalue. It is easy to show that first \(d\) eigenvectors form the same vector subspace as the steering matrix \(A\). Since \(v_{d+1}\) is orthogonal to all the first \(d\) eigenvectors and the signal subspace, it is also orthogonal to all direction vector \(a(\theta)\) at the true DOA. The orthogonality relationship is given by

$$a^H(\theta_k)v_{d+1} = 0 \quad \text{for} \quad k = 1, 2, \ldots, d.$$  \hspace{1cm} (2.20)

Then the extrema-searching algorithm of the Pisarenko method is given by

$$P_{Pisarenko}(\theta) = \frac{1}{||a^H(\theta)v_{d+1}||^2}.$$  \hspace{1cm} (2.21)
Ideally, at the true signal incident direction, \( P_{\text{Pisarenko}}(\theta) \) has a value of infinity. However, since we only have estimates of the eigenvector \( v_{d+1} \), the orthogonality relationship does not hold any more. Instead the cosine of the angle between \( a(\theta_k) \), \( k = 1, 2, \ldots, d \), and \( v_{d+1} \) will be probably close to zero, or \( P_{\text{Pisarenko}}(\theta) \) will have peaks at the estimates of DOA.

For a uniform linear array the Pisarenko method can be represented in a polynomial-rooting format

\[
D(z) = a^H(z)v_{d+1} = r_0 \prod_{i=1}^{d}(1 - r_i z^{-1}),
\]

(2.22)

where \( z = e^{j\omega \Delta \sin \theta / c} \). The directions of arrival are found from the angles of the \( d \) roots of the polynomial.

### 2.3.5 MUSIC

The MUSIC or MULTiple SIgnal Classification algorithm was first developed by Schmidt [21] and Bienvenu et al [22]. The MUSIC algorithm can be viewed as an extension of the Pisarenko method since it uses a covariance matrix of size \( M \times M \), where \( M \) can be any number as long as \( M > d \) is satisfied. The MUSIC algorithm, too, employs the orthogonality relationship of the noise subspace to steering vector at the true signal directions \( \theta_k \) for \( k = 1, 2, \ldots, d \) or

\[
a^H(\theta_k)V_n = 0. \quad (2.23)
\]

Thus the MUSIC power spectrum is written as

\[
P_{\text{MUSIC}}(\theta) = \frac{1}{||a^H(\theta)V_n||^2} = \frac{1}{\sum_{i=d+1}^{M} ||a^H(\theta)v_i||^2},
\]

(2.24)
with the peaks of the spectrum representing the DOA estimates. Like other subspace-based methods, MUSIC has the advantage that it can be applied to a sensor array of an arbitrary geometry, which is known or calibrated. However it will totally fail when coherent signals are present.

2.3.6 Root MUSIC

Root MUSIC, first suggested in [23], is a variation of MUSIC, and is applicable to uniform linear sensor arrays. The polynomial of Root MUSIC is given by

$$P_{RM}(z) = a^H(z) V_n V_n^H a(z)$$

$$= r_0 \prod_{i=1}^{M-1} (1 - r_i z^{-1})(1 - r_i^* z). \quad (2.25)$$

There are $2(M - 1)$ roots associated with the polynomial $P_{RM}$; the $d$ roots with the largest amplitude inside unit circle are chosen as signal zeros, and their angles are the DOA estimates. The advantage of Root MUSIC over Spectral MUSIC can be understood by considering the effects of an error $\Delta r_i$. Assuming the error $\Delta r_i$ is a random variable with mean $r_i$, it can be decomposed into radial and tangent components. Since the DOA estimates take angles of the $d$ roots with the largest amplitude inside unit circle, the radial component of $\Delta r_i$ will cause no error in the DOA estimate by Root MUSIC. However, such radial errors do affect the spectrum by Spectral MUSIC. In fact, the sharpness of the spectral peaks is determined by the radial component of $\Delta r_i$ or the radius distance between $r_i$ and the unit circle. This is particularly critical for closely spaced roots as it may result in only one peak causing an apparent loss in resolution. So Spectral methods always have less resolution compared to Root forms.
2.3.7 Minimum Norm Method

The Minimum Norm method [24] is also a subspace-based algorithm. It finds the \( M \times 1 \) vector \( d_{MN} \) with a unit first element, which is entirely in the noise subspace and has the minimum 2-norm. We partition the signal and noise subspace vectors as follows

\[
V_s = \begin{bmatrix} g^H \\ V'_s \end{bmatrix}, \quad V_n = \begin{bmatrix} c^H \\ V'_n \end{bmatrix},
\]

where \( g^H \) and \( c^H \) are the first rows of \( V_s \) and \( V_n \), respectively. Since \( d_{MN} \) lies in the range of \( V_n \), \( d_{MN} \) will be orthogonal to the columns of \( V_s \), i.e.,

\[
V_s^H d_{MN} = 0. \tag{2.26}
\]

Solving equation (2.26) via the least squares technique, we have

\[
d_{MN} = \begin{bmatrix} \frac{1}{V'_{g}^H g} \\ \frac{V'_{c}^H c}{1 - g^H g} \end{bmatrix}. \tag{2.27}
\]

Alternatively \( d_{MN} \) can be represented using noise eigenvectors as follows

\[
d_{MN} = \begin{bmatrix} \frac{1}{V'_{c}^H c} \\ \frac{V'_{c}^H c}{1 - g^H g} \end{bmatrix}. \tag{2.28}
\]

For a uniform linear array the polynomial rooting scheme for the DOA estimation using the Minimum Norm method is given by

\[
D(z) = a^H(z)d_{MN} = \prod_{i=1}^{M-1} (1 - r_iz^{-1}). \tag{2.29}
\]
Ideally $D(z)$ has $d$ zeros on the unit circle and the remaining zeros lie inside the unit circle. In the case of noisy data the $d$ roots closest to the unit circle are chosen as the signal roots; their angles are the DOA estimates.

The extrema-searching form of the Minimum Norm method is given by

$$P_{MN}(\theta) = \frac{1}{||a^H(\theta)d_{MN}||^2},$$

(2.30)

whose peaks represent the DOA estimates.

### 2.3.8 ESPRIT and Matrix Pencil

ESPRIT (Estimation of Signal Parameter via Rotation Invariance Technique) [25] is a technique that may be used to estimate frequency or direction of arrival. As compared to MUSIC, ESPRIT retains most of the essential features of the arbitrary array of sensors, but it achieves a significant reduction in computation complexity of MUSIC associated with extrema-searching procedures by imposing a constraint on the structure of the sensor array. In other words, there should be two identical subarrays available; one is a known-displacement $\delta$ (no rotation) version of the other. The output vectors from the two subarrays are

$$x(t) = A_x s(t) + n_x(t) = A s(t) + n_x(t);$$

(2.31)

$$y(t) = A_y s(t) + n_y(t) = A\Phi s(t) + n_y(t),$$

(2.32)

where the matrix $\Phi$ is a diagonal $d \times d$ matrix of the phase delays between the doublet sensors for the $d$ wavefronts:

$$\Phi = \text{diag}[e^{j\omega_0 \delta \sin \theta_1/c}, e^{j\omega_0 \delta \sin \theta_2/c}, \ldots, e^{j\omega_0 \delta \sin \theta_d/c}].$$

(2.33)
Let $E_{sx}$ and $E_{sy}$ be the signal subspace obtained from subarray output vector $x(t)$ and $y(t)$, respectively. We can form a matrix pencil as follows:

$$V_{sy} \Phi = V_{sx}. \quad (2.34)$$

When estimates of $V_{sx}$ and $V_{sy}$ are available, $\Phi$ can be determined using a least squares or a total least squares approach. The method using a least squares approach to estimate $\Phi$ is referred to as Least Squares (LS) ESPRIT. Asymptotically, Least Squares and Total Least Squares (TLS) ESPRIT will have the same performance [26]. For LS ESPRIT

$$\Phi = (V_{sy}^H V_{sy})^{-1} V_{sy}^H V_{sx}. \quad (2.35)$$

ESPRIT can also be characterized by the matrix pencil from linear algebra theory. It is shown [27] that the DOA estimation problem can be viewed as the generalized eigenvalue problem.

### 2.3.9 Maximum Likelihood (ML) Method

ML estimators possess salient properties of asymptotic unbiasedness and asymptotic efficiency. Assume that the noise vector $w(t)$ is a stationary and ergodic complex valued Gaussian process of zero mean and covariance matrix $\sigma^2 I$, where $\sigma^2$ is an unknown scalar. The noise samples $w(t_i), i = 1, 2, \ldots, N$ are statistically independent. If we pack $N$ independent snapshot vectors into an $M \times N$ matrix column by column

$$X = AS + W, \quad (2.36)$$

where $X$ and $W$ are the $M \times N$ matrices

$$X = [x(t_1), x(t_2), \ldots, x(t_N)],$$
\[
W = [w(t_1), w(t_2), \ldots, w(t_N)],
\]
and \( S \) is the \( d \times N \) matrix
\[
S = [s(t_1), s(t_2), \ldots, s(t_N)],
\]
then the joint distribution of the sampled data is given by
\[
f(X|\theta) = \prod_{i=1}^{N} \frac{1}{\pi^{M} \det[\sigma^2I]} \exp\{-\frac{1}{\sigma^2} \| [x(t_i) - As(t_i)]^H [\sigma^2I]^{-1} [x(t_i) - As(t_i)] \| \}.
\] (2.37)

The log likelihood, ignoring constant terms, is given by
\[
L = -NM \text{log}\sigma^2 - \frac{1}{\sigma^2} \sum_{i=1}^{N} \| x(t_i) - As(t_i) \|^2.
\] (2.38)

To maximize \( L \), we have to maximize the log likelihood with respect to all unknown parameters, though only the DOA estimates are of interest. To obtain the compressed likelihood [28], we maximize \( L \) with respect to \( \sigma^2 \) with \( A \) and \( s \) fixed and get the following
\[
\hat{\sigma}^2 = \frac{1}{MN} \sum_{i=1}^{N} \| x(t_i) - As(t_i) \|^2.
\] (2.39)

Substituting this result back into \( L \) and ignoring constant terms, we get the following minimization problem
\[
\min_{A,s} \left\{ \sum_{i=1}^{N} \| x(t_i) - As(t_i) \|^2 \right\}
\] (2.40)
Applying the compressed likelihood method once more by fixing $A$ and minimizing the above function with respect to $s$ yields

$$\hat{s}(t_i) = (A^H A)^{-1} A^H x(t_i).$$  \hspace{1cm} (2.41)

Substituting $\hat{s}(t_i)$ into the minimization function, we can obtain the final log-likelihood function

$$L(A) = L(\theta) = tr(P_A R),$$  \hspace{1cm} (2.42)

where $P_A = A(A^H A)^{-1} A^H$ is the projection operator of $A$, $R$ is the sample covariance matrix estimated from the given data and $tr[ ]$ is the trace of the bracketed matrix.

In spite of the optimality of ML method, maximizing the log-likelihood function (2.42) involves multivariate nonlinear programming problem, which is computationally demanding. Further, the log-likelihood function is non-concave; this requires a careful selection of the initial value for the maximization. In contrast, the subspace-based extrema-searching methods introduced previously just need the one-dimensional line search procedure, which is computationally much less complex than ML. Although several researchers [29] [30] [31] have proposed different fast numerical algorithms, the ML estimator at this stage is still more a theoretical standard to which the suboptimal estimators are compared than a practical tool.

### 2.4 Classifications of Spectral Estimators

There are many classification methods for spectral estimators depending on the classification criteria used. The following methods are suggested.


2.4.1 Classification by Chronology

Chronologically, we can roughly classify spectral estimators into two categories:

1. Classical Methods:
   Periodogram and the Wiener-Khintchine theorem typify the classical methods.

2. Modern Methods:
   All the high resolution methods reviewed previously in this chapter can be included in this category.

2.4.2 Classification by Model

Spectral estimators may or may not be model-based; that is, they may be parametric or non-parametric:

1. Parametric Methods:
   In parametric methods of spectral analysis a model is assumed in the formulation of the problem, and the parameters of the model are estimated from the limited observation interval. Although these models can take on a variety of different forms, a rational model with finite parameters is among the most popular in the contemporary spectral estimation. When employing a rational model, the power spectrum of the signal processes has the following form

   \[ P(e^{j\omega}) = \frac{b_0 + b_1 e^{-j\omega} + \ldots + b_m e^{-j(m-1)\omega}}{1 + a_1 e^{-j\omega} + \ldots + a_n e^{-j(n-1)\omega}}. \] (2.43)

   Herein, only \( m + n + 1 \) model parameters are to be determined. As we discussed previously, theoretically the Wiener-Khintchine theorem requires infinitely long data record. Then if there is only a short data record available,
this will inevitably yield a parameter mismatch dilemma; that is, the number of parameters to be estimated exceeds the number of data sample available. The finite number of parameters in this model then provides the mechanism for circumventing this parameter mismatch dilemma. Accordingly we distinguish an auto-regressive (AR) model by an all-pole rational function, a moving-average (MA) model by an all-zero rational function, and an auto-regressive-moving average (ARMA) model by a rational function with both poles and zeros. In this dissertation most of the spectral estimators discussed belong to the rational model category.

2. Non-parametric Methods:
Non-parametric methods of spectral analysis differ from parametric ones in that no specific model is presupposed in formulating the estimation problem. Typical methods in this category are the periodogram and the Wiener-Khintchine methods. The Wiener-Khintchine might also be classified as parametric MA with infinite parameters.

2.4.3 Classification by Vector Space

For covariance-based methods we can classify spectral estimators into non-subspace methods and subspace methods. In the latter signal and noise subspaces are separated using corresponding eigenvectors; the former does not commit this kind of subspace partition.

1. Non-subspace Methods:
   Maximum Entropy, Conventional Beamforming, Minimum Variance and Maximum Likelihood belong to this category.

2. Subspace Methods:
Pisarenko, MUSIC, Root MUSIC, ESPRIT and Matrix pencil belong to this category.
2.4.4 Classification by Numerical Procedure

Modern spectral estimators also fall into three classes designated by the numerical procedure.

1. Extrema-search:
   - One-dimensional Line Search: Beamforming, Pisarenko, Minimum Variance, Maximum Entropy, MUSIC and Minimum Norm.
   - Multi-dimensional Search: Maximum Likelihood.

2. Polynomial Rooting:
   Pisarenko, Root MUSIC and Minimum Norm.

3. Matrix-shifting:
   ESPRIT and Matrix pencil.

2.5 Determination of the Number of Signals

Subspace-based spectral estimators assume that the number of signals is known so that the vector space $C^{M \times M}$ can be correctly partitioned into signal and noise subspaces. When the number $(d)$ of signals is unknown and the $N > d$ data is noise-free, then the number of nonzero eigenvalues of the sample covariance matrix determines the number of signals. When the data is corrupted by noise, then the number of signals may be estimated by detecting the distribution of the eigenvalues of the sample covariance matrix. Two popular methods are based on the information theoretic criteria (AIC) for model order selection introduced in the context of linear prediction by Akaike [32, 33] and the minimum description length (MDL) criterion introduced by Schwartz [34] and Rissanen [35]. The number of signals is then determined as the value for which the criterion is minimized. Both
criteria were adapted to spectral analysis and array processing by Wax and Kailath [36]. The function to be minimized for the AIC criterion is

\[
AIC(k) = -2 \log \left( \left( \prod_{i=k+1}^{M} \lambda_i \right)^N \left/ \left( \frac{1}{M-k} \sum_{i=k+1}^{M} \lambda_i \right)^{(M-k)N} \right. \right) + 2k(2M - k) \tag{2.44}
\]

and for the MDL criterion,

\[
MDL(k) = - \log \left( \left( \prod_{i=k+1}^{M} \lambda_i \right)^N \left/ \left( \frac{1}{M-k} \sum_{i=k+1}^{M} \lambda_i \right)^{(M-k)N} \right. \right) + \frac{1}{2} k(2M - k) \log(N) \tag{2.45}
\]

The estimated number of sigmoids \( \hat{d} \) is taken to be that value of \( k \) which minimizes either \( AIC(k) \) or \( MDL(k) \). As for the detection performance of the two criteria, see [36].
Chapter 3

Covariance Estimation by Incorporating Cross-correlations

3.1 Introduction and Statistical Background

As shown in the Chapter 2, all modern spectral estimators mentioned here involve at the first step of processing the estimation of the covariance matrix from a short data record corrupted with noise. Assuming that observed random vectors are wide-sense stationary and ergodic, the time average operation is used in place of the ensemble average operation for the purpose of covariance matrix estimation. Due to the special situation of the short data records we are dealing with in spectral analysis and array processing, extra efforts have been exerted to find a better estimate of the covariance matrix by taking into account more information regarding signal sources and the inherent structure of the covariance matrices for certain cases. The following well known conclusions from statistics have been widely exploited by electrical engineers in detection and estimation problems: a better parameter estimate can always be obtained

- by properly utilizing more information regarding the parameter to be estimated and
CHAPTER 3. COVARIANCE ESTIMATION

• by using longer observation times.

To provide a more rigorous mathematical description of the above conclusions we consider two basic statistical problems. Before we start to elaborate on the problems we first review the concepts of the Fisher information matrix and the Cramer-Rao bound.

Let the conditional probability density function of random vector $x$ be given by $f(x|\varphi_D)$, where $\varphi_D = [\theta_1, \theta_2, \ldots, \theta_{D-1}, \theta_D]^T$ is the $D \times 1$ parameter vector. Then we can define the Fisher information matrix as follows:

**Definition 3.1 (Fisher Information Matrix):** The Fisher information matrix $J(\varphi_D)$ with elements $J_{ij}$ is defined by

$$J_{ij} = E \left[ \frac{\partial}{\partial \theta_i} \log f(x|\varphi_D) \frac{\partial}{\partial \theta_j} \log f(x|\varphi_D) \right].$$

(3.1)

**Theorem 3.1 (Cramer-Rao Bound):** The error covariance matrix for $\hat{\varphi}_D$, an unbiased estimator of $\varphi_D$, as defined by equation (1.3) is bounded by

$$\text{Cov}(\hat{\varphi}_D) \geq J^{-1}(\varphi_D),$$

(3.2)

provided $J(\varphi_D)$ is positive definite. That is, $\text{Cov}(\hat{\varphi}_D) - J^{-1}(\varphi_D)$ is positive (semi-)definite. The $i$th diagonal element of $\text{Cov}(\hat{\varphi}_D)$ denotes the mean-squared error of the estimate of $\theta_i$ and we have

$$\text{Var}(\hat{\theta}_i) \geq J_{ii}^{-1}.$$  

(3.3)

The exact Fisher information matrices for the Gaussian random vectors are given in Appendix B.
CHAPTER 3. COVARIANCE ESTIMATION

3.1.1 Nuisance Parameters

Definition 3.2 (Nuisance Parameters): The subset of parameters in $\varphi_D$ not of interest to us is referred to as nuisance parameters.

Let $\theta_D$ be a nuisance parameter in $\varphi_D$. Partitioning the parameter vector $\varphi_D$ and the Fisher information matrix as follows:

$$
\varphi_D = [\varphi_D^{T-1}, \theta_D]^T, \quad (3.4)
$$

$$
J(\varphi_D) = \begin{bmatrix}
J_{11}(\varphi_D) & j(\varphi_D) \\
j^T(\varphi_D) & \gamma(\varphi_D)
\end{bmatrix}, \quad (3.5)
$$

where $J_{11}(\varphi_D)$ is the Fisher information matrix for $\varphi_D^{T-1}$, the first $D-1$ elements of $\varphi_D$ when $\theta_D$ is specified. From the partitioned matrix inverse theorem [28], we get the inverse of the partitioned $J(\varphi_D)$

$$
J^{-1} = \begin{bmatrix}
J_{11}^{-1} + \rho J_{11}^{-1}jj^TJ_{11}^{-1} & -\rho J_{11}^{-1}j \\
-\rho j^TJ_{11}^{-1} & \rho
\end{bmatrix}, \quad (3.6)
$$

where $\rho = (\gamma - j^TJ_{11}^{-1}j)^{-1} > 0$ for the positive definite matrix $J$. The upper left block of $J^{-1}$ is the lower bound of the error covariance for the estimate $\hat{\varphi}_{D-1}$, with the nuisance parameter estimated simultaneously and written as

$$
[J^{-1}]_{D-1} = J_{11}^{-1} + \rho J_{11}^{-1}jj^TJ_{11}^{-1}, \quad (3.7)
$$

or

$$
[J^{-1}]_{D-1} - J_{11}^{-1} = \rho J_{11}^{-1}jj^TJ_{11}^{-1}. \quad (3.8)
$$

Since $\rho > 0$ and $J_{11}^{-1}jj^TJ_{11}^{-1}$ is a positive semi-definite matrix, we conclude that

$$
[J^{-1}]_{D-1} - J_{11}^{-1} \geq 0. \quad (3.9)
$$
The above equation shows that adding nuisance parameters into the estimation problem will increase the estimation error of the parameters of interest. This result can be extended to the general situation where the nuisance parameter is a complex vector.

This conclusion can be used to explain why the ML estimate of a Toeplitz structured covariance matrix [7] possesses an advantage over the sample covariance method. For a Toeplitz covariance matrix of size $M \times M$ there are $M$ unknown parameters to be estimated using Burg's ML estimation method. However, the sample covariance matrix method estimates $M(M+1)/2$ parameters, which is equivalent to adding $M(M-1)/2$ nuisance parameters in the estimation problem, thus causing larger estimation errors. The method of Ziskind and Wax [8] utilizes the periodicity relationship between the $M$ parameters and thus achieves a better performance.

3.1.2 Estimation Improvement by Averaging Multiple Estimates

Assume that there are $N$ estimates of the covariance matrix, $\hat{R}_i$ for $i = 1, 2, \ldots, N$, which is the sum of the true covariance matrix $R$ and the estimation error $\Delta R_i$, associated with the $i$th estimate.

$$\hat{R}_i = R + \Delta R_i.$$  \hspace{1cm} (3.10)

The average of the $N$ estimates is

$$\hat{R} = R + \frac{1}{N} \sum_{i=1}^{N} \Delta R_i = R + \Delta R.$$  \hspace{1cm} (3.11)

Define the "vector-packed" $M \times M$ matrix as an $M^2 \times 1$ vector formed by packing the matrix column by column into a vector. Let the vector-packed estimation error,
\( \Delta R_i \) for \( i = 1, 2, \ldots, N \) be independent and identically distributed as multivariate Gaussian random vectors with zero means and covariance \( Q \); then the vector-packed \( \Delta R \) will also have a multivariate Gaussian distribution with zero mean and covariance \( Q/N \). As \( N \) increases the covariance matrix of \( \Delta R \) approaches zero.

The CSM method by Wang and Kaveh [9] and the method to be presented in this chapter achieve the goal of estimation improvement based on the above argument. The CSM method obtains the multiple estimates of covariance matrices from the widebandness of signal sources; the method we are presenting obtains multiple estimates of covariance matrices from the cross-correlations between the sets of random processes, which can be, for example in subarray processing, the outputs of different subarrays. The purpose of the steered covariance matrices method proposed by Krolik and Swingier [10] is also to utilize the widebandness of incident signals to reduce the SNR threshold.

In the following sections the Least Squares Linear Prediction (LSLP) of one random vector from another random vector is reviewed, and a method for incorporating the cross-correlations is presented based on the review.

### 3.2 LSLP of Random Vectors

Let random vectors \( x_1 \in \mathbb{C}^{M_1} \) and \( x_2 \in \mathbb{C}^{M_2} \) have a covariance matrix \( R \in \mathbb{C}^{(M_1+M_2) \times (M_1+M_2)} \) defined as

\[
R = \text{Cov}(x_1, x_2) = \begin{bmatrix}
R_{11} & R_{12} \\
R_{21} & R_{22}
\end{bmatrix},
\]

(3.12)

where \( R_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)^H] \); \( i, j = 1, 2 \); and \( \mu_i = E[x_i] \) is the mean vector of the process \( x_i \); \( i = 1, 2 \). Now we consider the LSLP of one random vector by
Chapter 3. Covariance Estimation

another random vector; that is, we want to find a linear transformation\(^1\) \( f \), given by

\[ \hat{x}_1 = f(x_2) = Tx_2 + x_0, \]  

(3.13)
such that the expectation of the 2-norm of the error vector \( \hat{x}_1 = x_1 - \hat{x}_1 \) or

\[ J = E\{||\hat{x}_1||_2^2\} = E\{||x_1 - \hat{x}_1||_2^2\} = E\{||x_1 - (Tx_2 + x_0)||_2^2\} \]  

(3.14)
is minimized, where \( x_0 \in C^{M_1} \) and \( T \in C^{M_1 \times M_2} \) is a transformation matrix. The linear transformation which minimizes \( J \) is given in the following theorem.

Theorem 3.2 (LSLP of Random Vectors): For \( x_1 \) and \( x_2 \) as above,

\[ E\{||\hat{x}_1||_2^2\} = E\{||x_1 - (Tx_2 + x_0)||_2^2\} \geq \text{tr}\{\text{Cov}(\hat{x}_1)\}, \]  

(3.15)

and equality holds for \( T = R_{12}R_{22}^{-1} \) and \( x_0 = \mu_1 - R_{12}R_{22}^{-1}\mu_2 \). In the above, \( \text{Cov}(\hat{x}_1) \) is the covariance matrix of the prediction error vector \( \hat{x}_1 \) with the least 2-norm and is specified by

\[ \text{Cov}(\hat{x}_1) = R_{11} - \text{Cov}(\hat{x}_1) = R_{11} - R_{12}R_{22}^{-1}R_{21}, \]

where \( \text{Cov}(\hat{x}_1) = R_{12}R_{22}^{-1}R_{21} \) is the covariance matrix of the optimal prediction of \( x_1 \) by \( x_2 \).

Proof: see [37]

We do not try to use the theorem directly since we do not need the prediction vector \( \hat{x}_1 \); instead we are concerned with the estimation of the auto-covariance

\(^1\)Note: we are abusing the term linear transformation here. Formally, \( f \) is termed the affine function. The term linear transformation is reserved for those affine functions that map zero into zero. Since in many applications of signal processing, \( x_1 \) and \( x_2 \) often have zero means, the affine functions \( f \) degenerate to linear transformations.
matrix of \( x_1 \). Note that the above theorem not only provides the optimal prediction of the random vector \( x_1 \) from the random vector \( x_2 \), but also implies that the auto-covariance matrix of \( x_1 \), \( R_{11} \) can also be estimated from \( R_{12} \) and \( R_{22} \).

\[
\hat{R}_{11} = R_{12}R_{22}^{-1}R_{21}, \quad (3.16)
\]

where \( \hat{R}_{11} \) is in fact the covariance matrix of the optimal prediction of \( x_1 \) by \( x_2 \). It is equation (3.16) that shows us how the cross-correlations between \( x_1 \) and \( x_2 \) can be exploited to predict the auto-covariance matrices. In practice we must use estimates of \( R_{12} \) and \( R_{22} \).

For zero-mean random vectors \( x_1 \) and \( x_2 \) we can provide another interpretation of the above theorem. Let us rewrite the objective function \( J \) in equation (3.14) as

\[
J = \text{tr}\{E[(x_1 - Tx_2)(x_1 - Tx_2)^H]\} \\
= \text{tr}\{R_{11} - R_{12}T^H - TR_{21} + TR_{22}T^H\} \\
= \text{tr}\{R_{11} - f(R_{22}, R_{12})\}, \quad (3.17)
\]

where matrix function \( f(R_{22}, R_{12}) = R_{12}T^H + TR_{21} - TR_{22}T^H \) is an estimate of \( R_{11} \) in terms of \( R_{22} \) and \( R_{12} \) (or \( R_{21} \)). The above equation indicates that optimal transform obtained in the theorem also minimizes the trace of the difference matrix \( R_{11} - f(R_{22}, R_{12}) \).

### 3.3 Improving the Estimates of Covariance Matrices by Incorporating Cross-correlations

In the previous section a theorem on the LSLP of random vectors was presented using the true second central moments of the random vectors \( x_1 \) and \( x_2 \). In this section we will come back to reality, i.e., we have only finite observations of \( x_1 \).
CHAPTER 3. COVARIANCE ESTIMATION

and $x_2$. We need to estimate the auto-covariance matrix of $x_1$ from all the observations we have. When the two random vector processes are correlated and their observations have similar signal-to-noise ratios (SNR), we are able to improve the estimate of the auto-covariance matrix of $x_1$ by incorporating its estimate $\hat{R}_{11}$ with the prediction from the cross-correlations $\hat{R}_{12}$ in (3.16). There may be different ways to do this; one approach is to average the original estimate with its prediction directly:

$$\hat{R}_{11} = \frac{1}{2}(\hat{R}_{11} + \hat{R}_{12}). \quad (3.18)$$

In some applications, however, the improvement can be achieved only if there exist relatively high (cross)-correlations between the random processes $x_1$ and $x_2$, and if $x_2$ does not have a SNR much lower than that of $x_1$. Otherwise, the expected improvement cannot be guaranteed. If the SNR values of the random processes were known, we could use the proper weighted averaging to prevent the predicted auto-covariance matrix from deteriorating the estimate of $R_{11}$

$$\hat{R}_{11}^w = \frac{1}{2}(\hat{R}_{11} + w_{12}\hat{R}_{12}), \quad (3.19)$$

where $w_{12}$ is the weighting factor. For the purpose of choosing the weighting factor $w_{12}$, we can consider a normalized informative parameter, I, defined as

$$I = \frac{\text{tr}\{\text{Cov}(\hat{x}_1)\}}{\text{tr}\{R_{11}\}} = 1 - \frac{\text{tr}\{R_{12}R_{22}^{-1}R_{21}\}}{\text{tr}\{R_{11}\}},$$

which indicates the normalized energy of the prediction error. The weighting factor can thus be selected inversely proportional to the value of this parameter.
For example, a reasonable choice is

\[ w_{12} = \left( \frac{\text{tr}[R_{12}R_{22}^{-1}R_{21}]}{\text{tr}[R_{11}]} \right)^q, \]  

(3.20)

where \( q \) can be any positive number. Now let us consider two special cases. When \( I = 0 \) or the prediction error is zero, \( w_{12} \) has a value of one, and when \( I = 1 \) or the two random vectors are uncorrelated with \( R_{12} = R_{21} = 0 \), \( w_{12} \) has a value of zero. This is exactly the manner in which we expect \( w_{12} \) to vary. On the other hand for \( x_1 \) with a given SNR the prediction error usually decreases with the SNR value of \( x_2 \). Thus \( w_{12} \) will take a larger value for high SNR and smaller value for low SNR. In this way, it can effectively prevent noise propagation.

### 3.4 Applications to Covariance-based Spectral Estimators

The following discussions are based on the DOA estimation scenarios, but it should be understood that the results obtained equally apply to the spectral estimation problems. Let us consider two general sub-arrays of the identical array geometry; one is a displaced version of the other as discussed in the ESPRIT algorithm. Then the output vectors from the two subarrays are

\[
\begin{align*}
    x_1(t) &= A_1s(t) + n_1(t) = As(t) + n_1(t), \\
    x_2(t) &= A_2s(t) + n_2(t) = A\Phi s(t) + n_2(t).
\end{align*}
\]

(3.21)  

(3.22)

If, in addition to the assumptions imposed previously, we assume that additive noises from different arrays are uncorrelated, then we get the cross-covariance matrix of the sub-array outputs

\[ R_{ij} = E[x_i x_j^H] = A_i S A_j^H + \sigma^2 I \delta_{ij}, \text{ for } i, j, = 1, 2 \]

(3.23)
CHAPTER 3. COVARIANCE ESTIMATION

and

\[ \hat{R}_{11} = A_1 S A_2^H R_{22}^{-1} A_2 S A_1^H = A_1 \hat{S} A_1^H, \]  

(3.24)

where \( \hat{S} = S A_2^H R_{22}^{-1} A_2 S \). When the covariance-based algorithms, especially the subspace-based methods, are used for spectral analysis, the signal subspace structure and the non-singularity or the rank of the signal covariance matrix preoccupy our attention in the covariance matrix analysis. We do not care at all about the structure of the signal covariance matrix except for its rank. For example spatial smoothing techniques [38] will generally change the structure of the signal covariance matrix and increase its rank. From matrix theory, we know that the rank of the product of two matrices is always less than or equal to the lower rank of the two matrices, or

\[ \text{rank}(AB) \leq \min[\text{rank}(A), \text{rank}(B)]. \]  

(3.25)

Retaining \( R_{22}^{-1} \) in (3.24) generally will obviously not help to increase the rank of the signal covariance matrix for coherent signals. Excluding \( R_{22}^{-1} \) simply effects a different \( S \) matrix. So we have every reason to neglect \( R_{22}^{-1} \) in the expression for \( \hat{R}_{11} \). Then the predicted auto-covariance matrix is re-defined for the sub-array processing as follows:

\[ \hat{R}_{11}^{\text{e}} = w_{12} R_{12} R_{21}, \]  

(3.26)

where \( w_{12} \) is the weighting factor. Its function is twofold: it normalizes \( \hat{R}_{11}^{\text{e}} \) to cancel the scaling effects arising from neglecting \( R_{22}^{-1} \) and prevents noise from propagating from the second sub-array when its outputs have very low SNR. Similar to the definition of \( w_{12} \) in (3.20), we can define \( w_{12} \) by

\[ w_{12} = \left( \frac{\text{tr}(R_{12} R_{21})}{\text{tr}(R_{11} R_{22})} \right)^q. \]  

(3.27)
As another example we consider the special case of $x_1 = x_2$. According to equation (3.26), this special case, which is approximately true when $x_1$ and $x_2$ are highly correlated, gives

$$\hat{R}^{2e}_{11} = R_{11}R_{11}.$$ 

The above result implies that $R_{12}R_{21}$ can be approximately regarded as the prediction of $R_{11}R_{11}$, which in turn contains the same signal subspace as $R_{11}$. Omitting $R_{22}^{-1}$ in the expression of $\hat{R}^2_{11}$ not only reduces the computational requirements but also avoids the numerical errors introduced by the matrix inversion operation. Note again that in the above discussions we assume that true cross-correlations are known for the convenience of formulation. If only finite observations are available, the improved estimate of (squared) $R_{11}$ should be given by

$$\tilde{R}_1^2 = \frac{1}{2}(\hat{R}_{11}^2 + \hat{R}_{11}^{2e}),$$ \hspace{1cm} (3.28)

where $\hat{R}_{11}^{2e} = w_{12}\hat{R}_{12}\hat{R}_{21}$ and $\hat{R}_{ij}$ is the estimated cross-covariance matrix. Similarly, the improved estimate of (squared) $R_{22}$ is given by

$$\tilde{R}_2^2 = \frac{1}{2}(\hat{R}_{22}^2 + \hat{R}_{22}^{2e}).$$ \hspace{1cm} (3.29)

Although we have only considered the case of two sub-arrays, the above results apply to the cases of more than two sub-arrays. For example, if there are $L$ sub-arrays, the predicted covariance matrix of the $i$th sub-array from the cross-correlations will be

$$\hat{R}_{ii}^{2e} = \frac{1}{L-1} \sum_{j=1,j \neq i}^L w_{ij}\hat{R}_{ij}\hat{R}_{ji},$$ \hspace{1cm} (3.30)
where the role of $w_{ij}$ is to prevent noise propagation only, and $\tilde{R}_{ii}$ will be

$$\tilde{R}_{ii}^2 = \frac{1}{2}(\tilde{R}_{ii}^2 + \tilde{R}_{ii}^{2e})$$

(3.31)
Chapter 4

Applications

4.1 Spatial Smoothing

4.1.1 Introduction

The subspace-based methods have proven to be an effective means of obtaining the DOA estimates of multiple signals from the outputs of a sensor array; however, the performance of algorithms based on these methods will severely degrade when some of the signals are coherent or highly correlated. The spatial smoothing method for coherent signals was first proposed by Evans et al. [38] and later developed by Shan et al. [39], Williams et al. [40] and Pillai et al. [41]. The solution is based on a preprocessing scheme that averages covariance matrices of \( L \) translational equivalent subarrays (see figure 4.1). In the resulting covariance matrix the signal covariance \( P \) will possess full rank, and the subspace-based method can be applied effectively. The multiple translational equivalent subarrays can be \( L \) individual arrays of identical geometrical structure but displaced to a certain spatial position with respect to the reference array (usually the first array), or they can be obtained by partitioning the original uniform linear array. In both cases the rank increment is obtained at the cost of losing the effective spatial aperture. In order to achieve a larger effective aperture, the method of modified spatial smoothing, also called
forward-backward spatial smoothing, has been proposed by Evans et al. [38] and was extensively studied by Williams et al. [40] and Pillai et al. [41]. Using the forward-backward spatial smoothing method, a uniform linear array can resolve as many as $2M/3$ coherent signals (with $M$ representing the number of sensor elements), which coincides with the bounds of the number of coherent signals resolvable by such an array [42].

While many efforts have been made to increase the effective spatial aperture for the spatial smoothing scheme, any information in the cross-correlations of the outputs of the subarrays has been ignored. In the conventional spatial smoothing scheme only the auto-correlations of the outputs of individual subarrays are utilized to obtain the final estimate of the covariance matrix. Utilizing the covariance matrix enhancement method presented in Chapter 3, we propose new spatial smoothing methods to exploit the information otherwise lost by the conventional algorithms.
4.1.2 Spatial Smoothing for the Subspace-based Method

Assume that $d$ narrowband coherent signals centered at $\omega = \omega_0$ are received by $L$ translational equivalent subarrays, each containing $m$ sensors, and that the additive noises from the different subarrays are uncorrelated. Then it is easy to verify that the $m \times m$ cross-covariance matrix of the $i$th and $j$th subarrays is given by

$$R_{m}^{ij} = A_m D_i P D_j^H A_m^H + \sigma^2 I \delta_{ij}, \quad (4.1)$$

where $A_m$ is the $m \times d$ steering matrix of the reference subarray, the one with the reference sensor placed at the coordinate origin, and

$$D_i = \text{diag}[e^{j \omega \tau_1}, e^{j \omega \tau_2}, \ldots, e^{j \omega \tau_d}] \quad (4.2)$$

is a $d \times d$ diagonal matrix, where $\tau_k = (\delta_i - \delta_1) \sin \theta_k / c$, with $\delta_i$ denoting the displacement of the $i$th subarray relative to the origin of the coordinate system. To facilitate expressions for the backward smoothing, we form the $m \times m$ backward cross-covariance matrix

$$\hat{R}_{m}^{ij} = J_r (R_{m}^{ij})^* J_r, \quad (4.3)$$

where $*$ denotes the complex conjugate and $J_r$ is the $m \times m$ reflection or exchange matrix, which has unity elements along the cross or secondary diagonal and zeros elsewhere. The cross or secondary diagonal of a matrix is the diagonal proceeding from the lower left to the upper right of the matrix. The smoothed covariance matrices using the forward only and the forward-backward techniques are given respectively by

$$R_f = \frac{1}{L} \sum_{i=1}^{L} R_{m}^{ii}, \quad (4.4)$$

$$R_{fb} = \frac{1}{2L} \sum_{i=1}^{L} [R_{m}^{ii} + \hat{R}_{m}^{ii}] = \frac{1}{2} [\hat{R}_f + J_r \hat{R}_f J_r]. \quad (4.5)$$
It is easy to observe that the forward only and the forward-backward spatial smoothing utilize only the covariance matrices of the outputs of individual subarrays. All the cross-subarray covariance matrices, \( R_{ij}^m \) and \( R_{ji}^m \) (\( i \neq j \)), are ignored.

### 4.1.3 Improved Spatial Smoothing Methods

In order to take full advantage of the cross-correlations, we present improved spatial smoothing techniques which contain all \( R_{ij}^m \) for \( i, j = 1, \ldots, L \) in the final smoothed covariance matrix. They require two steps. First we apply equations (3.30) and (3.31) to obtain the improved estimate of the covariance matrix of each subarray and then implement the spatial smoothing, in which the average of all the improved subarray covariance matrices is calculated. Another requirement for the proposed methods is that they must guarantee the smoothed signal covariance to reach full rank using the same number of subarrays as in the conventional methods.

The improved covariance matrix for the \( i \)th subarray is given by

\[
\hat{R}_{ii}^2 = \frac{1}{L} \sum_{j=1}^{L} w_{ij} \hat{R}_{ij} \hat{R}_{ji},
\]

where each \( w_{ij} \) is set to ones assuming that the output of all subarrays have similar SNR values. Then the proposed forward only smoothing method and forward-backward smoothing method can be written as

\[
\hat{R}_{f}^2 = \frac{1}{L^2} \sum_{i=1}^{L} \sum_{j=1}^{L} R_{ij}^m R_{ji}^m,
\]

\[
\hat{R}_{fb}^2 = \frac{1}{2L^2} \sum_{i=1}^{L} \sum_{j=1}^{L} [R_{ij}^m R_{ji}^m + R_{ji}^m R_{ij}^m] = \frac{1}{2} [\hat{R}_{f} + J_r \hat{R}_{f} J_r]
\]

respectively. The following provides a proof that the proposed methods need the same number of subarrays to make the signal covariance matrix of full rank as do the conventional methods.
CHAPTER 4. APPLICATIONS

Forward Only Spatial Smoothing

Equation (4.1) can be used to rewrite (4.7) as

$$\tilde{R}_f = A_m \frac{1}{L^2} \left\{ \sum_{i=1}^{L} D_i \left\{ \sum_{j=1}^{L} P D_j^H A_m^H A_m D_j P \right\} + 2\sigma^2 P \right\} D_i^H A_m^H + \frac{1}{L} \sigma^4 I, (4.9)$$

and the $d' \times d$ smoothed signal covariance matrix is then

$$\tilde{P}_f = \frac{1}{L^2} \sum_{i=1}^{L} D_i \left\{ \sum_{j=1}^{L} P D_j^H A_m^H A_m D_j P \right\} D_i^H$$

$$= \frac{1}{L^2} \sum_{i=1}^{L} D_i P D_i^H, (4.10)$$

where

$$P = \left[ \sum_{j=1}^{L} P D_j^H A_m^H A_m D_j P \right] + 2\sigma^2 P. (4.11)$$

**Lemma 4.1:** The $\tilde{P}$ in (4.11) is a $d \times d$ positive semi-definite matrix, with at least one nonzero element in each row.

**Proof:** Since the original signal covariance matrix $P$ is positive semi-definite, it is straightforward to show that $\left[ \sum_{j=1}^{L} P D_j^H A_m^H A_m D_j P \right]$ is a positive semi-definite matrix. Then $\tilde{P}$, the sum of two positive semi-definite matrices is also positive semi-definite. The following two facts should be considered

- The positive semi-definite matrix has nonnegative diagonal elements.
- All the diagonal elements of $2\sigma^2 P$ are positive. If this were not true, it would imply that some of the signals would have zero energy, which conflicts with our assumptions.
It follows thereafter that \( P \) has positive diagonal elements; in other words, \( P \) has at least nonzero element in each row. Furthermore, \( P \) can be written as the product of its Hermitian square root:

\[
P = C^H C,
\]

(4.12)

where \( C \) is also positive semi-definite, with each row having at least one non-zero element.

**Theorem 4.1:** If the number of subarrays is greater than or equal to the number of signals, the forward-only spatially smoothed covariance matrix (4.7) corresponds to a full rank signal covariance matrix.

**proof:** We rewrite \( \tilde{P}_f \) as

\[
\tilde{P}_f = [D_1, D_2, \ldots, D_L] \left[ \begin{array}{ccc}
\frac{1}{L^2} P \\
& \ddots \\
& & \frac{1}{L^2} P
\end{array} \right] [D_1, D_2, \ldots, D_L]^H,
\]

(4.13)

where \( D_1 = I \) when the first subarray is taken as the reference array. \( \tilde{P}_f \) can be further simplified to

\[
\tilde{P}_f = \frac{1}{L^2} GG^H,
\]

(4.14)

where \( G \) is a \( d \times Md \) block matrix,

\[
G = [D_1 C, D_2 C, \ldots, D_L C].
\]

(4.15)

Obviously the rank of \( \tilde{P}_f \) is equal to the rank of \( G \). Next we will prove that when the subarray displacement is properly designed, the matrix \( G \) will have rank \( d \).
Recall that the rank of a matrix is unchanged by a permutation of its columns. It can be easily verified that

\[ \text{rank}\{G\} = \text{rank} \begin{bmatrix} c_{11}b_1 & c_{12}b_1 & \ldots & c_{1d}b_1 \\ c_{21}b_2 & c_{22}b_2 & \ldots & c_{2d}b_2 \\ \vdots & \vdots & \ddots & \vdots \\ c_{d1}b_d & c_{d2}b_d & \ldots & c_{dd}b_d \end{bmatrix}, \]

(4.16)

where \( c_{ij} \) is the \( ij \)th element of the matrix \( C \) and \( b_i \) is a \( 1 \times L \) vector,

\[ b_i = [e^{j\omega \tau_i \delta_1}, e^{j\omega \tau_i \delta_2}, \ldots, e^{j\omega \tau_i \delta_d}]. \]

(4.17)

Lemma 4.1 shows that \( C \) has at least one nonzero element in each row; thus the matrix \( G \) will be of full row rank if

1. the number of subarrays is greater than or equal to the number of signals, namely, \( L \geq d \), and
2. the displacements of the subarray are selected such that \( \{b_1, b_2, \ldots, b_d\} \) are linearly independent.

The second condition can be satisfied by properly selecting the displacement value for each subarray. For example, if we choose

\[ \delta_i = i\delta_1, \]

(4.18)
then \( \{b_1, b_2, \ldots, b_d\} \) can be embedded in a Vandermonde matrix, which will guarantee these vectors being linearly independent. In fact we can obtain this kind of subarray by partitioning a uniform linear array as described in [39]; that is, we form the first subarray with sensors \( \{1, 2, \ldots, m\} \), the second one with sensors \( \{2, 3, \ldots, m + 1\} \) and so on (see figure 4.2). The first condition then implies that a uniform linear array can at most resolve \( M/2 \) coherent signals by the improved forward only spatial smoothing method.
CHAPTER 4. APPLICATIONS

Forward-Backward Smoothing

There is a constraint on the geometrical structure of the subarrays for the forward-backward spatial smoothing method to apply.

**Lemma 4.2:** For the forward-backward spatial smoothing method the geometrical structure of subarrays must be such that

\[ J_rA_m^* = A_mT, \]  

(4.19)

where \( A_m \) is the steering matrix of the reference subarray and \( T \) is a \( d \times d \) diagonal matrix; that is, the subarrays must be symmetric.

**Proof:** The proof is a straightforward result from the fact that the ranges of \( \tilde{R}_f \) and \( \tilde{R}_b \) are required to be the same. The ranges of the two matrices are given by

\[ \mathcal{R}(\tilde{R}_f) = \text{span}\{A_m\}, \]
\[ \mathcal{R}(\tilde{R}_b) = \text{span}\{J_rA_m^*\}. \]

From \( \mathcal{R}(\tilde{R}_f) = \mathcal{R}(\tilde{R}_b) \), we have

\[ J_rA_m^* = A_mT_1, \]  

(4.20)

where \( T_1 \) is a \( d \times d \) nonsingular matrix. Since nondiagonal \( T_1 \) implies that the structure of the arrays is data dependent, we have to constrain \( T_1 \) to being a
nonsingular diagonal matrix, and thus we have proved equation (4.19). A little thought will convince the readers that equation (4.19) implies the symmetry of subarrays.

*Special case:* When the subarray is uniformly linear, we can write

\[ J_r A_n^* = A_m D^{1-m}, \]  

(4.21)

where

\[ D = \text{diag}[\epsilon^{j\omega_1 \tau_1}, \epsilon^{j\omega_2 \tau_2}, \ldots, \epsilon^{j\omega_n \tau_n}], \]

with \( \tau_k = \Delta \sin \theta_k/c \), and \( \Delta \) denoting the sensor spacing in each subarray.

Assume that the subarrays satisfy the condition on the structure given by the Lemma 4.2. We can simplify equation (4.8) using equations (4.9) and (4.10).

Then we obtain the improved forward-backward spatially smoothed signal covariance matrix:

\[ \hat{\Phi}_{fb} = \frac{1}{2} [A_m \hat{P}_f A_m^H + J_r A_n^* \hat{P}_f^* A_m^T J_r]. \]  

(4.22)

To simplify our proof, we assume the subarrays are uniform and linear with displacements satisfying the following equation:

\[ \delta_i = (i - 1) \Delta. \]  

(4.24)
thus can be further simplified to

\[ \hat{P}_{fb} = \frac{1}{2L^2} \left[ \sum_{i=1}^{L} D^{i-1} \hat{P}(D^{i-1})^H + D^{2-i-m} P^* (D^{2-i-m}) \right]. \tag{4.25} \]

**Theorem 4.2:** If the number of subarrays is greater than or equal to half of the number of signals, the forward-backward spatially smoothed covariance matrix (4.8) corresponds to a full rank signal covariance matrix.

**Proof:** Similar to what we have done in the forward only case, we rewrite \( \hat{P}_{fb} \) as

\[ \hat{P}_{fb} = [D^{2-L-m}, \ldots, D^{-m}, D^{1-m}, I, \ldots, D^{L-1}] \begin{bmatrix} \frac{1}{2L^2} P^* & \cdots & \frac{1}{2L^2} P \\ \frac{1}{2L^2} P^* & \ddots & \frac{1}{2L^2} P \\ \vdots & \ddots & \vdots \end{bmatrix} [D^{2-L-m}, \ldots, D^{-m}, D^{1-m}, I, \ldots, D^{L-1}]^H. \tag{4.26} \]

We continue to simplify \( \hat{P}_{fb} \) to

\[ \hat{P}_{fb} = \frac{1}{2L^2} HH^H, \tag{4.27} \]

where it is easy to verify that

\[ H = [D^{2-L-m} C^*, \ldots, D^{-m} C^*, D^{1-m} C^*, \ldots, D^{L-1} C^*]. \tag{4.28} \]

The rank of \( \hat{P}_{fb} \) is the same as the rank of \( H \). By expanding the right-hand side of
equation (4.28) and permuting the resulting columns, we have

\[
\tilde{H} = \begin{bmatrix}
    c_{11}d_1 & c_{12}d_1 & \cdots & c_{1d}d_1 & c_{11}b_1 & c_{12}b_1 & \cdots & c_{1d}b_1 \\
    c_{21}d_2 & c_{22}d_2 & \cdots & c_{2d}d_2 & c_{21}b_2 & c_{22}b_2 & \cdots & c_{2d}b_2 \\
    \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
    c_{d1}d_1 & c_{d2}d_2 & \cdots & c_{dd}d_d & c_{d1}b_d & c_{d2}b_d & \cdots & c_{dd}b_d 
\end{bmatrix} = [E[G]], \quad (4.29)
\]

where \( c_{ij} \) is the \( ij \)th element of the matrix \( C \), and \( a_i \) and \( b_i \) are \( 1 \times L \) vectors

\[
a_i = [e^{j(2-\nu-m)\omega_0\tau_1}, \ldots, e^{j(-\nu-m)\omega_0\tau_p}, e^{j(1-m)\omega_0\tau_1}], \quad (4.30)
\]

and

\[
b_i = [1, e^{j\omega_0\tau_1}, \ldots, e^{j(L-1)\omega_0\tau_1}]. \quad (4.31)
\]

We have proven that \( G \) has full row rank due to its Vandermonde structure and that \( E \) also has full row rank. Although it is possible that \( H \) is row-rank deficient, the possibility is very small in practice. Excluding some special cases we can conclude, assuming \( L \geq 2d \), that the improved forward-backward spatial smoothing method usually obtains a nonsingular spatial smoothed signal covariance matrix. For uniform linear arrays, as many as \( 2M/3 \) coherent signals can be resolved.

**Remark**

Note for the case of uniform linear arrays that the additive noises of different subarrays may not be uncorrelated any more. In this case, as proposed in [43], we can either ignore this kind of correlations when SNR is high or \( L \gg d \), or we can estimate the noise variance \( \sigma^2 \) and subtract the noise covariance before we apply the algorithm. Alternatively we can use another algorithm as will be presented later in the section 4.3 on spectral estimation.
4.1.4 Simulation Results

In this section simulation results are presented to illustrate the performance of the new spatial smoothing methods and to compare it with that of the conventional methods in terms of both resolution and the Root Mean Squared Errors (the Root MSE).

Resolution

In the simulation, we use a uniform linear array of 16 sensors; the wavenumber is defined as \( \pi \sin \theta \). Three fully correlated signals with equal powers are arriving from bearing angles 10°, 14°, and 80° respectively. The SNR, the signal to the noise energy ratio, is 15 dB; the number of snapshots is 64 and the size of the subarray is 8. For both the conventional and the proposed methods we use the forward-backward spatial smoothing scheme to obtain the smoothed covariance matrix and then apply the MUSIC algorithm to find the power spectra. Although the noise is correlated for subarrays obtained by partitioning a linear array, we ignore these noise correlations since the SNR is high for this case. Five typical runs using the conventional forward-backward spatial smoothing method are plotted in figure 4.3 (a). This figure shows that the two signals of bearing angles 10° and 14° are not resolved. The results for the improved forward-backward spatial smoothing proposed in this section are shown in figure 4.3 (b), where resolution of the two signals at close bearing angles is achieved. The simulation results show that the proposed methods produce a more stable estimate of the smoothed covariance matrix and thus higher resolution can be achieved.

Estimation Errors

In order to compare estimation errors produced by the new and conventional spatial smoothing methods, a similar experimental configuration has been designed. A
Figure 4.3: MUSIC spectra using (a) the conventional forward-backward spatial smoothing and (b) the improved forward-backward spatial smoothing. The two signals at 10° and 14° are not resolved in (a), but are well resolved in (b).
uniform linear array of 16 sensors is employed to receive two coherent signals with bearing angles 10° and 20°. We use two farther separated sources to exclude possible effects caused by the different resolution capabilities. Both the new and the conventional forward-backward spatial smoothing schemes have been applied for the estimation of spatial smoothed covariance matrices. The number of snapshots and the size of subarrays are the same as those used in the previous simulation. The Root MSE vs SNR has been calculated via Monte Carlo simulations (two hundred runs for each case) using the Root Minnorm based on the estimated covariance matrices by the new and conventional spatial smoothing algorithms. Figure 4.1 (a) and 4.4 (b) show that the new spatial smoothing algorithms outperform the conventional spatial smoothing algorithms in terms of the estimation errors. The advantage of the new method is particularly pronounced when the SNR value is near the SNR threshold region.

4.2 Covariance Improvement by Taking Advantage of Temporal Correlations

4.2.1 Introduction

In array processing, even though consecutive sample vectors are considered independent with respect to the noise, they generally are correlated in time with respect to the signal, since the multiple snapshots \( x(t_i) \) are usually obtained by continually sampling the continuous array output signals. In this case correlations between the snapshots do exist in most practical situations. How to exploit these correlations in the scenario of narrowband signals will be addressed in this section [44], and the extension to the wideband case will be briefly sketched.
Figure 4.1: Root MSE vs. SNR for (a) the signal with bearing angle 10° and (b) the signal with bearing angle 20°. The solid and dashed lines correspond to the new and the conventional forward-backward spatial smoothing algorithms respectively.
Covariance Matrix Estimation via Sample Covariance Method

In the first step of the signal parameter estimation problem the given array output vectors are put into an appropriate statistical framework, then the Maximum Likelihood principle provides a systematic way to obtain an estimator. There are two approaches for data characterization by the statistical model in array processing. One approach, called the deterministic method, is developed based on the assumption that the observation noise is i.i.d. zero mean Gaussian with signals being known. This approach has been used in deriving the ML DOA estimator in Chapter 2 and yields the joint Gaussian distribution in equation (2.27). The second approach, termed the stochastic method, is different from the first approach in that the stochastic method assumes the array output vectors are i.i.d. Gaussian with zero mean. Similarly, we can obtain the joint distribution of array output vectors as follows:

\[ f(X|\theta) = \prod_{i=1}^{N} \frac{1}{\pi^{M} \det[R]} \exp\{-x(t_i)^{H}R^{-1}x(t_i)\}. \]  

(4.32)

where \( X \) is defined in equation (2.36). The sample covariance matrix in equations (2.6) and (2.9) is derived via the Maximum Likelihood principle based on the stochastic model.

4.2.2 Estimating Covariance Matrices by Exploiting Temporal Correlations

The proposed method for covariance matrix estimation is a two step procedure. First the sample covariance matrix corresponding to the augmented \( KM \)-dimensional sample vectors is calculated, with \( K \) representing the number of consecutive snapshots packed into one augmented sample vector. Then the temporal correlations contained in the augmented covariance matrix are used to enhance the autocovari-
CHAPTER 4. APPLICATIONS

Augmented Sample Vectors

We assume that the temporal correlation time for the array output vectors \( x(t_i) \) is \((K-1)T\), where \( T \) is the temporal sampling period. The DOA estimation problem for stationary narrow-band signal sources with known temporal frequency spectra does not require specific temporal aperture location information due to the near-zero bandwidth of the signal sources. However, in order to incorporate the temporal correlations into the estimate of the covariance matrix, we add the temporal aperture by packing \( K \) consecutive narrow-band snapshots to form the \( KM \)-dimensional augmented sample vectors (see Figure 4.5):

\[
x_a(t_i) = [x^T(t_i), x^T(t_{i+1}), \ldots, x^T(t_{i+K-1})]^T.
\]

(4.33)

This vector form was suggested by Wax et al. [45] to do simultaneous spatio-temporal estimation for narrow-band sources. We have a different purpose; that
is, to include the temporal correlations of the array outputs in the augmented sample covariance matrix \( \hat{R}_a \), which is given by

\[
\hat{R}_a = \frac{1}{N - K + 1} \sum_{i=1}^{N-K+1} x_a(t_i) x_a^H(t_i).
\]  

(4.34)

Furthermore if we consider \( \hat{R}_a \) to be a \( K \times K \) block matrix with each block an \( M \times M \) matrix such that

\[
\hat{R}_a = \begin{bmatrix}
\hat{R}_{11} & \hat{R}_{12} & \cdots & \hat{R}_{1K} \\
\hat{R}_{21} & \hat{R}_{22} & \cdots & \hat{R}_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{R}_{K1} & \hat{R}_{K2} & \cdots & \hat{R}_{KK}
\end{bmatrix},
\]

where \( \hat{R}_{ij} \) denotes the temporal cross covariance matrix of the \( i \)th and \( j \)th packed output vectors in the augmented vectors, then the off-diagonal blocks contain the temporal correlations ignored by the conventional sample spatial covariance method where only the diagonal blocks are used. It is possible to incorporate the off-diagonal block information into the final estimate of the \( M \times M \) spatial covariance matrix \( \hat{R}_{11} \) using the method proposed in Chapter 2.

4.2.3 Matrix Enhancement Using Temporal Correlations

By applying equations (3.30) and (3.31), we have the improved estimate of the \( i \)th diagonal block in \( R_a \)

\[
\hat{R}_{ii}^2 = \frac{1}{K} \sum_{k=1}^{K} \hat{R}_{ik} \hat{R}_{ki}.
\]  

(4.35)

For relatively large \( N \) the covariance matrix \( \hat{R}_a \) is approximately block Toeplitz, or the blocks on each diagonal are approximately equal. Thus the final estimate
can be any of the $\hat{R}_{ii}$. For example, we choose

$$\hat{R}^2 = \hat{R}_{i_1}^2 = \frac{1}{K} \sum_{k=1}^{K} \hat{R}_{ik} \hat{R}_{ki},$$

$$= \frac{1}{K} \hat{R}(1 : M, 1 : KM)\hat{R}^H(1 : M, 1 : KM), \quad (4.36)$$

where $\hat{R}(1 : M, 1 : KM)$ is an $M \times MK$ matrix obtained by taking the first $M$ rows of $\hat{R}_a$.

### 4.2.4 Enhancement for Coherent Signals

The expectation of $\hat{R}_{ij}$ can be written as

$$R_{ij} = R_{i-j} = AP_{i-j}A^H + \sigma^2 I_{ij}, \quad (4.37)$$

where $P_{i-j} = E[s(t)s^H(t-(i-j))]]$ is the $d \times d$ cross signal covariance matrix. When the signal sources are coherent, the $P_{i-j}$ become singular. Simulations show that an efficient covariance enhancement requires the $P_{i-j}$ to have full rank. For a uniform linear array, spatial smoothing techniques [39] have been developed to accomplish this goal. Thus an enhancement procedure for coherent signal sources is proposed as follows:

1. Calculate $\hat{R}_a$ using $N-K+1$ augmented sample vectors.

2. Perform spatial smoothing for $R_{ik}$, $k=1,2, \ldots, K$, and produce the smoothed covariance matrix $\hat{R}_{ik}$ of the size $m \times m$, with $m$ ($d < m < M$) representing the size of the subarrays.

3. Calculate the enhanced covariance matrix using

$$\hat{R}^{2s} = \sum_{k=1}^{K} \hat{R}_{ik}^s \hat{R}_{ki}^s. \quad (4.38)$$
CHAPTER 4. APPLICATIONS

4.2.5 Extension to Wideband Signals

For the bearing estimation problem with the presence of wideband signals, we have recourse to either CSM [9] or Broad-Band Signal-Subspace Spatial-Spectrum (BASS-ALE) [46, 47]. Our covariance matrix enhancement algorithm can be used in conjunction with BASS-ALE for wideband signals.

**BASS-ALE Method**

BASS-ALE method consists of the following major steps [46, 47]:

- Determine the dimension \( D \) of the wideband source space by *time-bandwidth product* as defined in [46, 47].

- Form augmented sample vectors \( x_a^L(t_i) \in C^{LM \times 1} \) with \( LM > D \), and calculate the augmented covariance matrix \( R_a^L \).

- Find the spatial spectrum based on the eigen-decomposition of \( R_a^L \).

**Covariance Matrix Enhancement for Wideband Signals**

Assuming that the temporal source correlation time is \( s (K - 1)T \), with \( K > L \), we propose the covariance matrix enhancement procedure as follows:

- Form augmented sample vectors \( x_a^K(t_i) \in C^{KM \times 1} \) and calculate the augmented covariance matrix \( R_a^K \).

- Apply equation (4.36) for covariance enhancement.
4.2.6 Simulation Results

In this section simulation results are presented to illustrate the superiority of the new covariance matrix estimation method over the conventional sample covariance method in terms of the spatial spectral resolution and the Root MSE.

Resolution

In the simulation we consider two fully correlated narrowband signals with equal powers impinging a uniform linear array from bearing angles 10° and 14° respectively. The uniform linear array is composed of 16 sensors with half-wavenumber spacing, where the wavenumber is given by \( \lambda = \frac{f_0}{c} \), with the normalized frequency \( f_0 = 0.1 \) Hz. The SNR is 15 dB; the number of snapshots is 100 and the size of the subarray is 8. Augmented sample vectors of size \( 48 \times 1 \) (\( K = 3 \)) are produced. Since the signals are coherent, we use the forward-backward spatial smoothing scheme to obtain the smoothed covariance matrix for both the conventional and proposed method, and then we apply the MUSIC algorithm to find the spatial spectrum. MUSIC spectra from five typical runs using the conventional sample covariance matrix are plotted in figure 4.6 (a). This figure shows that the two signals are not resolved. The results using the new estimation method proposed in this section are shown in figure 4.6 (b), wherein the achievement of resolution of the two signals can be seen. The simulation results show that the proposed methods produce a more stable estimate of the covariance matrix and higher spatial resolution by taking into account the temporal correlations.

Estimation Errors

Since the old and new covariance estimators yield different resolution capabilities, we move the bearing angles of the two signals farther apart to 10° and 20°. We carry out two hundred runs using the Root Minnorm to calculate the Root MSE
Figure 4.6: MUSIC spectra using (a) the conventional sample covariance matrix method and (b) the enhanced covariance matrix estimation method. Two signals at 10° and 14° are not resolved in (a), but are well-resolved in (b).
of DOA estimates for different SNR ranging from -10 dB to 30 dB. The remaining simulation parameters are the same as in the previous simulation. Figure 4.7 (a) and 4.7 (b) show that the new covariance estimator is also preferable in terms of estimation errors especially when SNR is low.

An interesting observation is that, when we use larger $K$, the spatial spectra are very similar to what is obtained here for $K = 3$. This result implies that in this simulation the first and second order temporal correlations dominate the temporal cross correlations. In figure 4.8 (a) and 4.8 (b), the Root MSE for the conventional method and the new method with respect to $K = 2$ and $K = 3$ are plotted. Although cases with $K$ greater than 3 are not plotted, their simulation results are almost identical to the case of $K = 3$.

### 4.3 Spectral Estimation

#### 4.3.1 Introduction

Thus far we have considered only array processing and the associated covariance matrix of the array output vectors. We now discuss the modern high resolution temporal spectral estimation problem using the same covariance matrix framework. Given a short record of the data sequence $\tilde{y}(t)$ for $i = 1, 2, \ldots, N$, composed of uniformly spaced samples of exponential (cisoidal) signals, data vectors are formed using a sliding window on the data sequence. The covariance matrix is then calculated, and a spectral estimation algorithm is used to determine the values of the parameters of the exponential functions. Since the data vectors are overlapped, both the signal and noise components are correlated with their counterparts in the adjacent vectors. This fact deserves special considerations. In this section we will adapt the new covariance matrix estimation method to the spectral estimation problem.
CHAPTER 4. APPLICATIONS

Figure 4.7: Root MSE vs. SNR for (a) the signal with bearing angle 10° and (b) the signal with bearing angle 20°. The solid and dashed lines correspond to the new and the conventional covariance estimators respectively.
Figure 4.8: Root MSE vs. SNR for (a) the signal with bearing angle 10° and (b) the signal with bearing angle 20°.
4.3.2 Matrix Telescope Series

Assuming the size of the covariance matrix is $M$, then an $M \times M$ covariance matrix associated with $N - M + 1$ data vectors $x_M(t_i) \in \mathbb{C}^{M \times 1}$ is given by equation (2.10). In order to reveal the relationship between the conventional correlation estimation method and the method to be proposed next, we consider an alternative approach for calculating the covariance matrix $R$. Let $x_N = [y(t_1), y(t_2), \ldots, y(t_N)]^T$; and define $I_{M,N}(k)$ as an $M \times N$ ($N > M$) matrix whose first $k$ columns and last $N - M - k$ columns are null with an $M \times M$ identity matrix in between such that

$$I_{M,N}(k) = \begin{bmatrix} 0_{M \times k}, I_M, 0_{M \times (N-M-k)} \end{bmatrix},$$

where $I_M$ is the $M \times M$ identity matrix and $0_{M \times k}$ is the $M \times k$ zero matrix. Then the covariance matrix $R$ can be written

$$R = \frac{1}{N - M + 1} \sum_{i=0}^{N-M} I_{M,N}(i) R_N I_{M,N}^T(i). \quad (4.39)$$

where $R_N = x_N x_N^T$. In other words, in order to find the $M \times M$ covariance matrix $R$, we form the $N \times N$ covariance matrix $R_N$ first. $R$ is the average of $(N - M + 1)$ $M \times M$ diagonal blocks (matrix telescoping series) in $R_N$ obtained by using an $M \times M$ mask sliding along the diagonal of $R_N$. See figure 4.9 for a graphical interpretation of this procedure. From figure 4.9, we can see that all correlations out of the blocks along the diagonal band matrix are ignored in this method. A new method will be devised to utilize the correlations off the band matrix.

4.3.3 Proposed Method

We first build a set of $M' \times 1$ data vectors $x_{M'}(t_i)$, for $i = 1, 2, \ldots, N - M' + 1$, with $M \leq M' \leq N$. Then a covariance matrix $R'$ of order $M'$ is calculated using
CHAPTER 4. APPLICATIONS

Figure 4.9: Telescoping series of a matrix. Diagonal blocks are obtained by sampling $R_N$ using a sliding $M \times M$ mask.

equation (2.10). At this stage we can apply the result in the spatial smoothing section directly to determine $R$ if the one or more of the following assumptions are satisfied:

- SNR is high;
- $M' \gg M$;
- Noise variance is known or can be estimated.

since the cross covariance between noise vectors can be neglected in these cases. Otherwise we have to consider the random vectors with zero noise cross covariance or the random vectors must be non-overlapped. We also use the matrix telescoping series method to calculate the improved covariance matrix. Unlike the conventional method, in the new method the averaging operation of $M' - M + 1$ diagonal blocks is implemented after the estimate of each diagonal block has been improved according to equations (3.30) and (3.31). The improvement procedure can be
illustrated considering the \( j \)th block, \( R_{M'}^{ij} \):

\[
R_{M'}^{ij} = I_{M,M'}(j)R^T_{M,M'}(j)
\]

\[
= \frac{1}{M' - M + 1} \sum_{i=1}^{M'-M+1} I_{M,M'}(j)x_{M'}(t_i)x_{M'}^H(t_i)I_{M,M'}^T(j)
\]

\[
= \frac{1}{M' - M + 1} \sum_{i=1}^{M'-M+1} y_1(t_i)y_1^H(t_i).
\]  

(4.10)

where \( y_1(t_i) = I_{M,M'}(j)x_{M'}(t_i) \) is composed of a proper portion (the \( j \)th to the \( j + M - 1 \)th elements) of the data vector \( x_{M'}(t_i) \). If we consider \( y_1(t_i) \) as one random vector, the remaining portion of the data vector \( x_{M'}(t_i) \) can be regarded as another random vector \( y_2(t_i) \), with

\[
x_{M'}(t_i) = \begin{bmatrix} y_{21}^T, y_1^T(t_i), y_{22}^T \end{bmatrix}^T
\]

\[
y_2(t_i) = \begin{bmatrix} y_{21}^T, y_{22}^T \end{bmatrix}^T.
\]

We can also write \( y_2(t_i) = I_{M,M'}(j)x_{M'}(t_i) \), where \( I_{M,M'}(j) \in (\mathbb{C}^{M' - M} \times \mathbb{C}^M) \) is the complement of \( I_{M,M'}(j) \),

\[
\tilde{I}_{M,M'}(j) = \begin{bmatrix} I_j & 0_{j \times M} & 0_{j \times (M' - M - j)} \\
0_{(M' - M - j) \times j} & 0_{(M' - M - j) \times M} & I_{M' - M - j} \end{bmatrix},
\]

and satisfies

\[
I_{M,M'}^T(j)I_{M,M'}(j) + \tilde{I}_{M,M'}^T(j)\tilde{I}_{M,M'}(j) = I_{M'}.
\]  

(1.11)

Then \( R_{y_1y_2} \), the cross correlation of \( y_1(t_i) \) and \( y_2(t_i) \), is given by

\[
R_{y_1y_2} = I_{M,M'}(j)R^T\tilde{I}_{M,M'}(j).
\]  

(4.12)
Applying equations (3.30) and (3.31), we get the improved estimate of the \( j \)th diagonal block as

\[
\hat{R}_{M'}^{(j)} = I_{M,M'}(j)R_{M,M'}^T(j)R_{M,M'}I_{M,M'}(j) + I_{M,M'}(j)R_{M,M'}^T(j)I_{M,M'}(j)R_{M,M'}(j)
\]

\[= I_{M,M'}(j)(R^t)^2I_{M,M'}(j). \tag{4.43}\]

Averaging all the improved diagonal blocks yields

\[
\hat{R}^2 = \frac{1}{M' - M - 1} \sum_{i=0}^{M'-M} I_{M,M'}(i)(R^t)^2I_{M,M'}(i). \tag{4.44}\]

Obviously, the proposed method is a generalization of the conventional method. When \( M' = N \), the new estimate is the same as the conventional estimate except for a scalar term \( \hat{R}^2 = ||F_N||^2 R \). When \( M' = M \), the new estimate is the square of the conventional method or \( \hat{R}^2 = R^2 \).

### 4.3.4 Order Selection

Unlike the array processing, time sequence analysis for frequency estimation has more freedom in order selection for the correlation matrix (covariance matrix if random processes are zero mean). Whereas in array processing the order or size of the covariance matrix is determined by the number of sensors of the array. For given \( N \) data sample, using a covariance matrix of larger size means that the following spectral estimators can take advantage of a larger time aperture to obtain higher resolution. However as the size \( M \) of covariance matrix increases, the number \( N - M + 1 \) of data vectors available decreases accordingly, and the quality of estimate of the covariance matrix becomes worse or the stability of the estimate is reduced. Order selection is a trade-off between resolution and stability, but there is not a general approach available to do this. Order selection depends
on several factors such as the number of signals, SNR value, data record length, the closeness of frequency components, etc.

**Pisarenko Method**

When the Pisarenko method is used, the order of the covariance matrix estimated by the conventional method is determined by the number of signals; namely, the size of the covariance matrix is $d + 1$. However, the new method has to select a proper value for $M'$, which can take a value in the range of $d + 1 \leq M' \leq N$. As a rule of thumb, $M'$ should be selected such that the rank of $R'$ is equal to or greater than $d$, the maximum number of signals considered. For a specific case, the optimal order can be found via trial and error.

**Other Subspace-based Methods**

The other subspace-based methods, such as MUSIC and Minimum Norm apply to all covariance matrices of the size larger than the number of signals. Using the conventional method, the empirical choice for $M$ falls into the range of $N/3 \leq M \leq 2N/3$ for short data [48, 49]. The new method has to determine the orders of two covariance matrices $M'$ and $M$ to achieve the optimal performance. For a short data record our experiences are summarized as follows:

- The preliminary size $M'$ also falls into the range of $N/3 \leq M' \leq 2N/3$.
- The final size $M$ is usually smaller than both $M'$ and the optimal order in the conventional method.
4.3.5 Simulation Results

Using the Pisarenko Method

Frequency estimation involves the problem of order selection, but inherent in the Pisarenko method is determination of the size of the final covariance matrix. The Spectral Pisarenko method is used for resolution studies, and the Root Pisarenko method is used for estimation error analysis. The complex data sequence is composed of two closely spaced cisoids with frequencies $f_1 = 0.20$ Hz and $f_2 = 0.22$ Hz. The length of the data sequence is 64, and the SNR is 15 dB. For comparison both the conventional and new methods have been used to find the covariance matrix, then the spectral Pisarenko method is applied to each. Five typical runs using the estimates of a $(d+1) \times (d+1)$ sample covariance matrix are plotted in figure 4.10 (a), which shows that two cisoids at frequency 0.20 Hz and 0.22 Hz are either not resolved or very much biased. For the new method, we chose $M' = 40$ and the results are plotted in figure 4.10 (b). We can see that two sinusoids are well-resolved with very good accuracy.

As was done in previous sections, frequencies of the two cisoids are next further separated to 0.2 Hz and 0.3 Hz for the estimation error analysis. Two hundred runs using the Root Pisarenko method are conducted to calculate the Root MSE of the DOA estimates for different SNR ranging from -10 dB to 30 dB. The remaining simulation parameters are the same as in the previous simulation. Figure 4.11 (a) and 4.11 (b) show that except for extremely high or low SNR, the new covariance matrix estimator makes significant improvements over the sample covariance matrix method. As a matter of fact, in extreme situations either the conventional estimator is already good enough or the SNR has fallen below the threshold of the Pisarenko method. Thus applying the new estimator in these cases will not make any differences.
Figure 4.10: Pisarenko spectra using (a) the conventional sample covariance matrix and (b) the new covariance matrix estimators. Two cisoids at 0.20 Hz and 0.22 Hz are either not resolved or greatly biased in (a), but are accurately resolved in (b).
Figure 4.11: Root MSE vs. SNR for (a) the signal at 0.2 Hz and (b) the signal at 0.22 Hz.
CHAPTER 4. APPLICATIONS

Figure 4.12: The average Root MSE for both signals vs. the order of the covariance matrix. For the new method, the x-axis denotes the order of the first covariance matrix, and the order of the final covariance matrix is 10. SNR=10 dB, N=32.

Using the Root Minnorm Method

When Minimum Norm is used for the frequency estimation, we need to consider the effects of choosing the order of the covariance matrix on the estimation error. For the new covariance matrix estimator, we must choose both $M$ and $M'$. We design a simulation using two closely spaced sinusoids with frequencies $f_1 = .20$ Hz and $f_2 = .22$ Hz. The length of the data sequence is 32, and the SNR is 10 dB. Both the conventional and new method have been used to find the covariance matrix, then Minimum Norm is applied to each. Monte Carlo trials (200) show that for the conventional method, the Minimum Norm method achieves the minimum MSE at $M = 16$ (see figure 4.12). For the new method we chose $M = 10$ and the optimal value for $M'$ is 24 as shown in the figure. Clearly the Minimum Norm method with the new covariance estimator obtains more accurate estimates of spectral parameters.
Chapter 5

Design of Transformation Matrices for Array Processing via Vector P-norm

5.1 Introduction

This chapter is relatively independent of previous chapters. In this chapter we will discuss the transformation design problem, and proposed a new design method based on the vector p-norm.

Design of transformation matrices plays a central role in many array processing techniques such as interpolated arrays [50] and CSM [9]. Since Wang and Kaveh introduced the focusing transformation matrix in the CSM algorithm, several classes of transformation matrices have been proposed [9, 51]. The CSM algorithm assumes that a preliminary estimate of the direction of arrival (DOA) can be obtained by using a low resolution spectral estimator. Focusing matrices transform covariance matrices at different temporal frequencies into ones at a pre-specified frequency. The averaging of these transformed covariance matrices will enhance the estimates of covariance matrices by suppressing the perturbations, which are usually assumed to be uncorrelated and tend to cancel each other through the
averaging operation. As a consequence of the focusing transformation, the CSM method can achieve a lower SNR threshold for angular resolution. Due to the assumption that a preliminary estimate of DOA is available, focusing matrices can be designed using fewer angular constraints around the DOA estimate than the degrees of freedom (the number of sensors). If the angular constraints are imposed exactly at the true DOA's of incident signal sources, the focusing transformation forms a perfect mapping, (a mapping without any transformation error). Thus in [51] Hung and Kaveh proposed a class of focusing matrices based solely on reducing focusing loss, which is defined by the ratio of array SNR's after and before the focusing operation.

Without a priori information on the spatial spectrum, the task of transformation matrices becomes the transformation from one vector subspace into another vector subspace. The constraints should be imposed at each possible angle. Thus the number of constraints will generally be larger than the degrees of freedom, and a transformation error will inevitably be introduced. To reduce this error Friedlander proposed to use the technique of sector processing in interpolated arrays [50]; i.e. the overall field of view of the array is segmented into several sectors, and a transformation matrix is designed for each sector using selected vector samples in this sector. A typical solution to this kind of problem is the Frobenius norm method. The computation associated with this transformation matrix design is a major consideration for choosing the Frobenius norm, but it is not unique choice, especially when the sector processing technique is applied. In this situation transformation matrices can be calculated offline and stored in memory for later use. This paves the way for the application of other transformation matrix design methods, which are computationally more extensive than the Frobenius norm.

In the next section the concepts of the vector p-norm and the matrix Frobenius norm are briefly reviewed; sector processing and interpolated arrays, which will be used as a application example of the proposed method, are outlined.
5.2 Background

5.2.1 Vector P-norm and Frobenius Norm

A vector norm on $\mathbb{C}^N$ is a function from $\mathbb{C}^N$ into the nonnegative real numbers whose value in some sense measures the size of a vector in $\mathbb{C}^N$. The vector $p$-norm is defined by

$$
\|x\|_p = \left( \sum_{i=1}^{N} |x_i|^p \right)^{1/p}, \quad 1 \leq p < \infty,
$$

where $x = [x_1, x_2, \ldots, x_N]^T$ is a vector in $\mathbb{C}^N$. The concept of norm can be extended to matrices since the set of matrices $\mathbb{C}^{M \times N}$ is a vector space which is essentially identical with $\mathbb{C}^{MN}$. Consequently, any vector norm on $\mathbb{C}^{MN}$ corresponds to a matrix norm on $\mathbb{C}^{M \times N}$. For example the 2-norm on $\mathbb{C}^{MN}$ induces the Frobenius norm on $\mathbb{C}^{M \times N}$ defined by

$$
\|A\|_F = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} a_{ij}^2},
$$

where $A = \{a_{ij}\}$ is a matrix in $\mathbb{C}^{M \times N}$.

In the $p$-norm family, $\| \cdot \|_1$, $\| \cdot \|_2$ and $\| \cdot \|_\infty$ are the most popular vector norms. From the definition of the $p$-norm, it follows that as $p$ increases, the relative contributions of the large elements in the vector to the $p$-norm increase accordingly. As an extreme case, $\| \cdot \|_\infty$ only takes the value of the largest element of the vector $x$. This feature of the vector $p$-norm manifests itself in the optimization problem when the $p$-norm of the approximation error vector is chosen as the objective function to be minimized. Different $p$ will generally yield different distributions.
of the elements of the error vector after the optimal solution has been reached. Similar to the above discussion, as $p$ increases, the elements of the error vector tend to become equal at the final stage of the optimization. Taking advantage of these features, we usually take small $p$ to achieve the robustness of the algorithm to some unexpected outliers in statistical signal processing, while in digital filter design, $|| \cdot ||_{\infty}$ is commonly used to make sure that the maximum errors do not exceed the prescribed specifications. However, there is no general rule for choosing the optimal $p$; a good $p$ in one specific application might be a bad choice in another. Nevertheless in using the $p$-norm family a good compromise can always be possible, and it is flexible for many engineering applications. For example, in designing a transformation for interpolated arrays as will be discussed next, each element in the error vector corresponds to the approximation error of the direction vector at a certain direction. If the density of the arriving angles is uniform in one sector and we want the transformation errors to be equal for every direction, the Frobenius method obviously cannot satisfy this design requirement. However the generalized $p$-norm method will be able to accomplish this goal.

5.2.2 Sector Processing and Interpolated Arrays

Since we will use the design of a transform matrix for an interpolated array as an application example of the $p$-norm method, we will give here a brief outline of the concept of the interpolated array. Note well that the application of the proposed method is not limited to interpolated arrays; it can be applied to the CSM method when the sector processing technique is adopted.

Interpolated arrays were proposed by Friedlander and can be used to address a variety of issues related to DOA estimation as mentioned in [50]. The idea is to use the outputs of virtual arrays computed from the real array by a linear transformation. The procedures of designing the interpolated arrays using the sector processing technique are as follows:
1. Segment the field of view of the array into $K$ sectors. The $k$th sector is defined by the angle interval $[\theta_1^k, \theta_2^k]$.

2. Define a set of $N$ angles

\[ \Theta_k = [\theta_1^k, \theta_1^k + \Delta \theta, \theta_1^k + 2\Delta \theta, \ldots, \theta_2^k] \]

for each sector.

3. Compute the steering matrix $A_k \in \mathbb{C}^{M \times N}$ associated with the set $\Theta_k$ for the given array as

\[ A_k = [a(\theta_1^k), \ldots, a(\theta_2^k)]. \]

4. Compute the steering matrix of the virtual array $B_k \in \mathbb{C}^{M \times N}$ using the same angle set $\Theta_k$:

\[ B_k = [b(\theta_1^k), \ldots, b(\theta_2^k)] \]

5. Find a transformation matrix $T_k \in \mathbb{C}^{M \times M}$ such that the norm of the error matrix $\|E_k\| = \|T_k A_k - B_k\|$ is minimized.

6. Compute the transformation error. If the error is larger than desired, repeat step 1 with reduced sector interval size. Otherwise store the transformation matrices for future use.

In [50] the matrix Frobenius norm is used as the distance measure in the designing procedures, and the $k$th transformation matrix is given by

\[ T_k = B_k A_k^\# , \quad (5.3) \]

where the superscript $\#$ denotes matrix pseudo-inverse operation. Using the SVD of matrix $A_k$ we have

\[ A_k = V_k \left[ \begin{array}{cc} \Sigma_k & 0 \\ 0 & 0 \end{array} \right] U_k^H, \]
where \( V_k \) is an \( M \times M \) unitary matrix, \( U_k \) an \( N \times N \) unitary matrix and \( \Sigma_k = \text{diag}[\sigma_1, \sigma_2, \ldots, \sigma_r] \) with \( r \leq \min(M, N) \). Then equation (5.3) can be rewritten as

\[
T_k = B_k U_k \begin{bmatrix} \Sigma_k^{-1} & 0 \\ 0 & 0 \end{bmatrix} V_k^H. \tag{5.4}
\]

In the following sections we will present the p-norm methods for transformation design.

### 5.3 P-norm Method

Let us consider the problem of approximation of the matrix \( B \in \mathbb{C}^{M \times N} \) by the linear transformation of matrix \( A \in \mathbb{C}^{M \times N} \); i.e., we want to find a transformation matrix \( T \) such that the error matrix \( E = TA - B \) is minimized according to some measure. Let

\[
T = [t_1, t_2, \ldots, t_M]^H
\]

\[
B = [b_1, b_2, \ldots, b_M]^H
\]

\[
A = [a_1, a_2, \ldots, a_N].
\]

Then the error matrix \( E \) is given by

\[
E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix} = \begin{bmatrix} t_1^H A - b_1^H \\ t_2^H A - b_2^H \\ \vdots \\ t_M^H A - b_M^H \end{bmatrix}. \tag{5.5}
\]

If the matrix Frobenius norm is used, the optimal \( T \) minimizes the cost function

\[
L(T) = ||E||_F. \tag{5.6}
\]
CHAPTER 5. DESIGN OF TRANSFORMATION MATRICES

We know that the Frobenius norm is equivalent to the vector 2-norm if matrix $A$ is packed into one long vector row by row. Another way to implement the Frobenius norm method is to design the transformation matrix row by row to minimize the vector 2-norm of each corresponding row vector of the error matrix $E$. Similarly we propose to design the transformation matrix $T$ row by row using the vector $p$-norm with the cost function defined by [52]

$$L(t_i) = \left\{ \sum_{j=1}^{N} (e_{ij}^*e_{ij})^p \right\}^{1/p} = \left\{ \sum_{j=1}^{N} [(a_j^H t_i - b_{ij})^*(a_j^H t_i - b_{ij})]^p \right\}^{1/p} \text{ for } i = 1, 2, \ldots, M, \quad (5.7)$$

where $e_{ij}$ is the $j$th element of row error vector $e_i$ and $b_{ij}$ the $j$th element of vector $b_i$. When $p = 1$, the $p$-norm based method degenerates into the Frobenius norm method.

5.3.1 Implementation

The minimization of $L(t_i)$ with respect to $t_i$ can be accomplished using any one of the well-established unconstrained optimization algorithms in [53]. Here the Newton’s method is used to find the optimal $t_i$ and hence the optimal $T$ matrix. The Newton’s method is stated briefly as follows:

1. Pick an initial $t_i(0)$.

2. Calculate the gradient $g_i(k) = \nabla_{t_i(k)} L(t_i(k))$ and the Hessian matrix $H_i(k) = \nabla^2_{t_i(k)} L(t_i(k))$.

3. Update vector $t_i(k)$ such that

$$t_i(k + 1) = t_i(k) - H_i^{-1}(k)g_i(k). \quad (5.8)$$
4. If the relative error decreases, and \( \frac{[L(t_i(k)) - L(t_i(k+1))]}{L(t_i(k))} \) is less than a given positive small number \( \epsilon \), then \( t_i(k+1) \) is the solution; otherwise set \( k=k+1 \) and return to step 2.

To implement the Newton’s algorithm, we have to derive the gradient and the Hessian matrix of the cost function \( L(t_i) \) with respect to \( t_i \). After some algebraic operations, we have the gradient vector \( g_i \) and the Hessian matrix \( H_i \) as follows (Matrix derivative rules can be found in Appendix A):

\[
g_i = 2 \left[ \sum_{j=1}^{N} (e_{ij}^2)^p \right]^{\frac{1-p}{p}} \left[ \sum_{j=1}^{N} (e_{ij}^2)^{p-1} e_{ij} a_i \right]. \tag{5.9}
\]

For \( p > 1 \)

\[
H_i = 4(1-p) \left[ \sum_{j=1}^{N} (e_{ij}^2)^p \right]^{\frac{1-2p}{p}} \left[ \sum_{j=1}^{N} (e_{ij}^2)^{p-1} e_{ij} a_i \right] \left[ \sum_{j=1}^{N} (e_{ij}^2)^{p-1} e_{ij} a_i^H \right]^H + 4(p - 0.5) \left[ \sum_{j=1}^{N} (e_{ij}^2)^p \right]^{\frac{1-p}{p}} \left[ \sum_{j=1}^{N} (e_{ij}^2)^{p-1} e_{ij} a_i^H \right]. \tag{5.10}
\]

For \( p = 1 \)

\[ H = H_i = 2AA^H. \tag{5.11} \]

In the above, \( e_{ij}^2 = e_{ij}^* e_{ij} \).

In practice we need to consider more than the algorithm itself. One question is how to choose the initial value of \( t_i \); in the simulations we select the 2-norm solution as the initial value of \( t_i \). Another question is how to deal with the inverse of the Hessian matrix when \( H \) is almost singular. We use the regulation method to overcome the numerical problem caused by the singularity of the Hessian matrix;
i.e., we replace $H$ in the algorithm by $H + \delta I$, where $\delta$ is a small positive number to make $H$ invertible.

### 5.4 Simulation Results

In order to verify the implementation procedure proposed, the convergence of the $p$-norm based algorithm using Newton's method will be demonstrated here.

We want to use the $p$-norm method to design an interpolated array. The original uniform linear sensor array is composed of four omni-directional sensors with adjoint sensors spaced at half wavelength. The virtual array to be designed possesses the same geometrical shape except that it has a position shift of two wavelengths with respect to the original array. We consider the case of $p = 2$ and the sector interval $[0^\circ, 30^\circ]$ with $\Delta \theta = 1^\circ$. In figure 5.1 the normalized $l_i(t_i)$ vs. the number of iteration is plotted. It can be seen that after about 5 iterations the $t_i$ will converge to the optimal $p$-norm solution. When the number of sensors increases, more iterations are needed for $t_i$ to converge. We can obtain a similar convergence rate for relatively small $p$. The $\infty$-norm is not considered here due to the numerical problems of the Newton's method. Other algorithms such as Minimax and Iterative Reweighted Least Squares can be used for the $\infty$-norm optimization.
Figure 5.1: The plot of the normalized $L(t_i)$ vs. the number of iterations, $i = 1, 2, 3, 4$. The optimal solutions are obtained after about 5 iterations.
Chapter 6

Velocity Estimation in Exploration Seismic Signal Processing

6.1 Introduction

The ultimate goal of exploration seismic signal processing is to reconstruct as accurately as possible an image of the subsurface in which both structure and physical properties are delineated. The exploration process may be broken down into three large categories: data acquisition, data processing and interpretation. The field of exploration seismology offers especially fertile ground for digital signal processing, and it represents an extremely important area for applications in array processing. It is an area in which development of array processing procedures has been underway for some time. The most important variable in seismic prospecting is velocity, the propagation speed of the seismic wave, because it contains information on the properties of the propagation medium. The term velocity analysis, also called velocity estimation is currently used to denote the process of determining velocity from the stacking of the CDP (common depth point data) [54] [55]. In this chapter we will focus our discussions on the velocity estimate problem.
To some extent it is safe to say that the velocity analysis is equivalent to the DOA estimation and spectral estimation. However, high resolution spectral estimators have not been successfully applied in this area since the seismic reflection signals do not satisfy some important assumptions based on which these high resolution spectral estimators have been developed. High resolution spectral estimators are usually developed based on a model of a plane wavefront and stationary signals. The seismic reflection signals, however, do not fit this model; rather they possess the following features:

- the seismic signals have non-plane (hyperbolic) wavefronts, and
- they are transient or non-stationary random processes.

The above features of the seismic signal deserve some special considerations when the modern array processing technique is applied.

### 6.2 Modeling Seismic Signals

#### 6.2.1 Wavefront Models

For the single layer case as described in figure 6.1, the reflection time-distance curves are approximated well by a simple hyperbolic of the form

\[
t_x^2 = t_0^2 + \frac{x^2}{v^2},
\]

where \(t_x\) is the two-way reflection time for a trace of the offset distance \(x\), \(t_0\) is the two-way reflection time for the zero offset trace, and \(v\) is the compressional wave velocity of the medium.
Reflection times for a horizontal reflector below a sequence of $N$ horizontal layers with constant interval velocities can be described by an infinite power series of the form [54]

$$t_{x,N}^2 = C_1 + C_2 x^2 + C_3 x^3 + \ldots \quad (6.2)$$

A hyperbolic approximation analogous to the single layer case results from retention of the first two terms,

$$t_{x(N)}^2 = t_{0(N)}^2 + \frac{x^2}{v_{rms}^2}, \quad (6.3)$$

where

$$t_{0(N)} = \sum_{k=1}^{N} \frac{2Z_k}{v_k},$$

$$v_{rms} = \left[ \frac{1}{t_{0(N)}^2} \sum_{k=1}^{N} v_k t_k \right]^{1/2}.$$

Here $x$ is the offset distance, $N$ the number of layers overlying the reflecting horizon, $Z_k$ the thickness of the $k$th layer, $v_k$ the interval velocity of the $k$th layer, $t_k$ the
two-way travel time in the \( k \)th layer, \( t_0(N) \) the two-way travel time to the bottom of the \( N \)th layer for the normal incident trace, and \( v_{rms} \) the root-mean-square (rms) velocity.

With the help of the hyperbolic model the measured seismogram can be written as

\[
d(x, t) = \int \int U(p, t_0) \delta(t_0 - \sqrt{t^2 - p^2 x^2}) dt_0 d(p) + n(x, t)
\]  

(6.4)

where

\( J(x, t) \) = recorded seismogram at offset \( x \) and two-way time \( t \).

\( U(p, t_0) \) = hyperbolic transform coefficient at rms slowness \( p = 1/v_{rms} \) and zero-offset time \( t_0 \).

\( n(x, t) \) = measurement noise at offset \( x \) and two-way time \( t \).

### 6.2.2 Wave Shape Model

Although we can assume an impulse signal with certain slowness and zero-offset time as an ideal reflection signal, the earth propagation media spreads it out such that it can be approximated as a Ricker wavelet. This method has found widespread use in seismic signal processing. It is, in fact, the second order derivative of the Gaussian function. The normalized expression for the Ricker wave is given by

\[
R_n(t) = \left( \frac{t^2}{\sigma_r^2} - 1 \right) \exp \left( -\frac{t^2}{2\sigma_r^2} \right),
\]

(6.5)

where \( \sigma_r \) is the unique parameter of the Ricker wave. The normalized peak frequency, \( f_p \), of its power spectrum can be determined by the following relationship
between \( f_p \) and \( \sigma_r \) (the proof is provided in Appendix C)

\[
f_p = \frac{1}{\sqrt{2\pi\sigma_r^2}}.
\]

(6.6)

### 6.3 Velocity Estimation

Hyperbolic model characteristics of reflection time-distance curves provide a means for establishing the necessary velocity relationships. Based on the hyperbolic model, several methods have been developed for this purpose [54] [56] [57] [58]. In the next subsection, we will review three representative methods.

#### 6.3.1 Current Techniques

**Semblance**

Semblance is probably the most commonly used velocity estimation method. It consists of performing a stack across the CDP gather along various hyperbolic trajectories and calculating the reflection coherency. To illustrate the principle of the semblance velocity estimation procedure let us consider a hypothetical CDP gather (figure 6.2). Suppose that the noise free reflection in figure 6.2 form an exact hyperbola and that the zero offset time is \( t_0 \). Assume that the velocity analysis is to be carried out with respect to \( t_0 \) and the range of velocity to be covered by the analysis is \( v_{\text{min}} \) to \( v_{\text{max}} \). Then the velocity analysis by semblance method is carried out as follows:

1. An initial stacking velocity \( v_1 = v_{\text{min}} \) is assumed and a lag trajectory corresponding to the \( t_0 \) and \( v_1 \) is found. We take a time gate of \( N + 1 \) samples from each trace, symmetrically disposed about the aforementioned hyperbolic trajectory (figure 6.3).
Figure 6.2: An hypothetical CDP gather
2. The in-gate sample covariance matrix $R$ is determined using the $N + 1$ sample vectors obtained across the traces. Since the data is discrete, the lag trajectory may not coincide with data samples. One might want to use interpolation for the lag trajectory and the $N + 1$ sample vectors, or just take the value of the nearest data sample. In both cases, data errors will inevitably be introduced.

3. The degree of match (coherency) between the traces at this alignment is measured by the semblance coefficient

$$S_c = \frac{u^T R u}{M \text{tr}(R)},$$

where $M$ is the number of sensors and $u$ is a column vector consisting of all one’s.

4. The velocity is then incremented by an appropriate step, and the semblance coefficient corresponding to the new hyperbolic trajectory is calculated.

5. Step (4) is repeated until $v_{\text{max}}$ is reached.

6. The zero offset time is incremented from $t_0$. Steps (1)–(5) are repeated.

7. The above process is repeated until the appropriate range of time down the data record has been covered.

Note that $S_c$ is a Rayleigh quotient [59] normalized by the trace of $R$. From the following facts that

$$\lambda_M \leq \frac{u^T R u}{u^H u} \leq \lambda_1,$$

and

$$\text{tr}\{R\} = \sum_{i=1}^{M} \lambda_i,$$
it follows that the semblance coefficient $S_c$ can only take the value between 0 and 1 or

$$0 \leq S_c \leq 1. \quad (6.8)$$

Note that non-negative $\lambda_1$ and $\lambda_M$ mentioned above are respectively the largest and smallest eigenvalues of the in-gate sample covariance matrix $R$.

**Key’s Method**

Key [57] introduced a new velocity estimator based on the eigen-decomposition of the in-gate sample covariance matrix $R$. The coherency measure is defined as

$$K_c = \alpha \frac{\lambda_1 - \sigma^2}{\sigma^2}. \quad (6.9)$$
where $\lambda_1$ is the largest eigenvalue of $R$, $\sigma^2 = \sum_{i=2}^{M} \lambda_i/(M - 1)$ is an approximation of the noise variance and $\alpha$ is a scaling factor according to the log-likelihood ratio that $\lambda_1$ is different from $\sigma^2$. It is not difficult to show that $K_c$ actually denotes the scaled SNR ratio (by $\alpha$) if the noise is white and uncorrelated with the signal. The velocity estimation procedures are the same except that the coherency measure $S_c$ is replaced with $K_c$. Key’s method will, however, involve extensive computation to determine the eigenvalues if $M$, the number of traces, is large. To reduce computational requirements, Key proposed to stack neighboring traces into $m$ ($m$ often takes values between 4 to 6) stacked traces; the in-gate sample covariance matrix is then found using the stacked traces. Key claimed the stacking procedure has the advantage of enhancing SNR and reducing computation intensity.

**Thorson’s Method**

In Thorson’s method, also called the inverse velocity stacking method [58], the recorded seismic profile is modeled as a linear combination of simple hyperbolic events of constant amplitude. A set of weighting coefficients is derived such that the resulting model approximates the input profile in the least squares sense. Each hyperbolic event is characterized by two components, $t_0$ and $p$. The set of weighting coefficients can be interpreted directly to the energy associated with the various velocities at different zero-offset times. In this sense the coefficients form an idealized constant velocity stack and are thus useful for velocity analysis.

For a given seismogram $d(x, t)$, find the weights $U(p, t_0)$ that satisfy equation (6.4) while minimizing the noise term $n(x, t)$. Thorson and Claerbout showed that this problem can be solved as an overdetermined linear system. Discretizing $d(x, t)$, $U(p, t_0)$ and $n(x, t)$ at discrete values of $x, t, p$ and $t_0$, equation (6.4) can be reduced to the linear system

$$d = Lu + n,$$

(6.10)
where \( d \) is a vector containing all values in the measured seismogram, \( u \) is a vector containing all the weights, \( n \) is the vector containing all noise values and \( L \) is the matrix containing all possible hyperbolic events in the discrete form of the delta function of equation (6.4). The least squares (LS) solution to the above linear equation is

\[
u = (L^T L)^{-1} L^T d.
\]  

(6.11)

While this algorithm is optimal in the LS sense, it requires extensive computation to find the pseudo-inverse of \( L \) (typically of the order 150,000 x 150,000).

6.3.2 Modifications and Improvements Over the Previous Methods

Several modifications and improvements over the aforementioned methods have been proposed by Kirlin and Du in [60], [61], [62] and [63]. We will not detail these methods here, instead we just briefly summarize the modifications.

Modified Key's Method

Key's method is defined as the scaled signal eigenvalue to the noise eigenvalue ratio. Clearly it cannot be used in the noise free case. We also have noticed that this method is very sensitive to the estimation error of the noise eigenvalue, especially when the noise is colored or correlated with the signal. In [61] we proposed to use the signal eigenvalue itself as a coherency measure to eliminate the problem of sensitivity to noise eigenvalue estimation error. We also proposed a new trace grouping method to enhance the velocity resolution.
Total Least Squares Velocity Estimator

Thorson’s method is a velocity estimator based on the Least Squares fitting of the recorded seismic data to a hyperbolic-model based velocity space. Thorson’s method is optimal in the LS sense assuming that the hyperbolic-modeled velocity space can perfectly represent the real velocity space. However the hyperbolic model is not perfect due to inconsistent traces such as the ones caused by static shifts [54]. When the model space is erroneous, the Total Least Squares (TLS) fit should be used, since TLS minimizes the errors caused by both the observation noise and the model mismatch. In [60] the TLS velocity estimator is proposed. The tentative results using short synthetic data indicate that improvements over the LS method may be achieved.

Applying MUSIC-type Velocity Estimators

MUSIC-type high resolution methods have also been used in velocity estimation in a local region of the seismic data [62] [63]. In order to handle the widebandness of seismic data, the CSM method has been utilized to estimate the focused covariance matrix. We also proposed a focusing based the spatial smoothing algorithm to improve the estimate of the covariance matrix. Based on the velocity estimation results, an interference cancelling algorithm has been proposed using the LS method. Simulations show satisfactory results for some synthetic data. The same high resolution velocity estimator has been independently proposed by Biondi and Kostov [64]. However they did not solve the spatial smoothing problem for non-planar waves; we solved this problem using the focused spatial smoothing method assuming the preliminary estimate of the velocity is given. The major disadvantage of this method is that it can not be easily used to estimate both parameters of the hyperbolic data (two-way time and the velocity). In [62], [63] and [64], only slowness is considered assuming the two-way time is already known.
6.4 The Relationship Between Semblance and Eigenstructure Velocity Estimators

Because high-resolution spectral estimators have received so much interest over the past decade, in his recent paper [65] Kirlin has performed thorough comparison between conventionally used semblance and eigenstructure velocity estimators (such as MUSIC). Our work on velocity estimation to be presented in this chapter is a continuation of the research in this direction. Based on the comparison, we recommend that the statistically optimal array processing method be applied to the velocity estimation problem.

6.4.1 Semblance and Conventional Beamforming

Recall that the conventional beamforming method employs a procedure of delay-and-sum processing to steer a beam in a particular direction. For the given field of view, the beamformer scans all possible angles where sources may be present, and calculates the array output power. The array output power can be written as the following for narrow-band signals

\[ P_{BF}(\theta) = \frac{a^H(\theta)Ra(\theta)}{a^H(\theta)a(\theta)}. \]  

(6.12)

The peaks of the spectrum are interpreted as the estimates of DOA. We can rewrite the above equation as

\[ P_{BF}(\theta) = \frac{u^H(\theta)R(\theta)u}{M}, \]  

(6.13)

and \( R(\theta) \) is the steered covariance matrix defined as in [10]

\[ R(\theta) = T(\theta)RT(\theta)^H, \]
where \( T = \text{diag}[1, e^{j\omega \Delta \sin \theta / c}, \ldots, e^{j(M-1)\omega \Delta \sin \theta / c}] \). The covariance matrix steering operation can also be implemented in the time domain by adding appropriate delays to different traces. As a matter of fact the in-gate covariance matrix used by the semblance method is exactly the steered covariance matrix estimated from the data in the time gate. Comparing equation (6.13) with (6.7), we find that at each time step the semblance velocity algorithm is actually a normalized (by the trace of \( R \)) conventional beamformer.

### 6.4.2 Optimal Beamformers

The conventional beamformer is data independent since it is independent of the statistics of the array output. Recently several statistically optimal beamformers have been developed [66]. We can anticipate that applying these optimal beamformers to the velocity estimation problem will yield better velocity estimation algorithms. The following tables summarize four commonly used optimal beamformers:

1. **Multiple Sidelobe Canceller (MSC):**

   | Definitions | \( x_a \) — auxiliary data  
   |            | \( y_m \) — primary data  
   |            | \( r_{ma} = E[x_a y_m^*] \)  
   |            | \( R_a = E[x_a x_a^H] \)  
   | Array Output | \( y = y_m - w_a^H x_a \)  
   | Criterion | \( \min_{w_a} E[\|y_m - w_a^H x_a\|^2] \)  
   | Optimal Weights | \( w_a = R_a^{-1} r_{ma} \)

2. **Reference Signal (RS) Beamformer:**
Definitions | \( x \) — array data  
| \( y_d \) — desired signal  
| \( r_{xd} = E[xy_d^*] \)  
| \( R_x = E[xx^H] \)  

Array Output | \( y = w^H x \)  

Criterion | \( \min_w E\{ |y - w^H x|^2 \} \)  

Optimal Weights | \( w = R_x^{-1} r_{xd} \)  

3. Maximum SNR Beamformer:

Definitions | \( x = s + n \) — array data  
| \( s \) — signal component  
| \( n \) — noise component  
| \( R_s = E[ss^H] \)  
| \( R_n = E[nn^H] \)  

Array Output | \( y = w^H x \)  

Criterion | \( \max_w \frac{w^H R_x w}{w^H R_n w} \)  

Optimal Weights | \( R_n^{-1} R_s w = \lambda_{\text{max}} w \)  

4. The Linearly Constrained Minimum Variance (LCMV) Beamformer:

Definitions | \( x \) — array data  
| \( C \) — constraint matrix  
| \( f \) — response vector  
| \( R_x = E[xx^F] \)  

Array Output | \( y = w^H x \)  

Criterion | \( \min_w \{w^H R_x w\} \) under the constraint \( C^H w = f \)  

Optimal Weights | \( w = R_x^{-1} C [C^H R_x C]^{-1} f \)  

6.5 Optimal Velocity Estimators

6.5.1 Some Special Considerations

As we pointed out in the introduction to this chapter, seismic reflections are transient signals with short presence time. The duration of the steering time gate is
thus usually very short. A typical time gate used in velocity analysis is about 48 milli-seconds, thus only about 25 snapshots are available to estimate the steered or in-gate sample covariance matrix if the sampling period is 2 milli-seconds. The number of sensors, however, is usually much larger than this number. In order to have a relatively stable estimate of a covariance matrix of size $M$, the number of snapshots should be at least 5 to 10 times larger than the matrix size. We cannot obtain more time samples, but we can utilize the spatial smoothing technique to solve this problem. Although spatial smoothing originally is used to overcome the multipath problem in DOA estimation, it can also be used to improve the stability of the estimate of the covariance matrix by trading off the spatial aperture. Two conditions accompany the spatial smoothing techniques; first, the signals must have a plane wavefront, and second, the subarrays must have the same geometry. If the seismic reflection is perfectly matched by a hyperbolic trajectory, it does have a plane wavefront in this short time gate. Furthermore if we partition a linear array using the method shown in figure 4.2, the ideal covariance matrices associated with these subarrays are identical for an exactly matched reflection. On the other hand, if the seismic reflection is not matched, both the signal and the noise components of the subarray covariance matrices are combined incoherently. The resulting covariance matrix is more or less like a noise only covariance matrix, which will result in low array output power. The above analysis justifies the feasibility of the application of the spatial smoothing technique to the velocity estimation problem.

6.5.2 Enhancement of the Estimates of Covariance Matrices

Assume the current time gate is formed at zero offset time $t_0$ and steering velocity $v_s$. Adopting the same notation as used in spectral estimation (Chapter 4, section
4.3), the spatial smoothed covariance matrix is

$$ R(t_0, v_a) = \frac{1}{M - m + 1} \sum_{i=0}^{M-m} I_{m,M}(i) R_M(t_0, v_a) I_{m,M}^T(i), $$

(6.14)

where $R_M(t_0, v_a)$ is the $M \times M$ in-gate covariance matrix, with $M$ denoting the number of sensors, and $m$ is the size of the subarrays. The enhanced covariance matrix is given by

$$ \tilde{R}^2(t_0, v_a) = \frac{1}{M - m + 1} \sum_{i=0}^{M-m} I_{m,M}(i) \tilde{R}_M^2(t_0, v_a) I_{m,M}^T(i). $$

(6.15)

The spatially smoothed covariance matrix or the enhanced spatially smoothed covariance matrix can be directly utilized in the optimal beamformers.

### 6.5.3 LCMV Velocity Estimator

The optimal beamformers given previously each has advantages and disadvantages. The first three methods require a priori information, and access to this information depends on applications. Lack of this knowledge may prevent utilization of these methods. The LCMV beamformer, however, is a very general approach acceptable for many applications. Thus we will apply the LCMV method to the velocity estimation problem.

The constraint matrix $C$ is designed so that the desired signal is allowed to pass without any attenuation and known strong interferences are blocked by forming zero beamformer responses at their arrival angle (velocities). While we may not know the velocity information about the strong interferences, we do assume that the desired signal has the same velocity as the steering velocity. In other words the desired signal is assumed to be accurately flattened in the time gate. The constraint
that allows the desired signal to pass without attenuation can be translated into the following

\[ u^H u = 1, \]  

where \( u \) is the a column vector containing all ones. If this is the only constraint used, we obtain the optimal weights

\[ w = \frac{R^{-1}(t_0, v_s)u}{u^H R^{-1}(t_0, v_s)u}, \]

and the array output power is then

\[ P(t_0, v_s) = \frac{1}{u^H R^{-1}(t_0, v_s)u}, \]

which is used as the coherency measure for the velocity spectrum.

If information regarding the interferences is available, more constraints can be added into the constraint matrix such that the LCMV beamformer has zero response to these known interferences. Thus the interferences can be cancelled to a maximum extent.

Remarks

The LCMV velocity estimator and the high resolution velocity estimator proposed in [62],[63] and [64] both adopt the modern array processing methodology. However the LCMV method exhibits several advantages over the high resolution velocity estimator:

- The LCMV method utilizes the steering covariance matrix, which in turn yields a lower SNR threshold [10].
The LCMV method can obtain the estimates of both two-way time and velocity.

- Spatial smoothing can be performed without spatial focusing; in other words, we do not need the preliminary velocity estimate of the seismic data.

- The LCMV method can take into account available information regarding strong interferences by properly designing the constraint matrix $C$.

In addition to the basic form of the LCMV algorithm given here, we can also incorporate any performance enhancement methods proposed to the Minimum Variance spectral estimator, such as the signal component enhancement procedure given in [16].

### 6.6 Simulations

#### 6.6.1 Comparison of Coherency Measure Threshold Discrimination

As Neidell and Taner argued in their paper [56], noise present on the data channels affects coherency measures primarily through the apparent amplitude and shape diversity it creates. The precise character of the effects depends on the noise statistics and the signal-noise interactions. Therefore it is reasonable to perform an experiment on noise-free data to establish in some sense the discrimination or resolution of semblance and the optimal velocity estimator. Figure 6.4 depicts a synthetic CDP gather. These data are designed specially to test the resolving power and discrimination threshold of candidate coherency measures. Each region indicated in the data contains a doublet; the pulses are separated by 20 milliseconds in two-way time and the rms velocities differ by 200 $ft/s$. All events are Ricker wavelets with dominant frequency at 40 $Hz$. Since the objective of this computation is a resolution test rather than the modeling of physical reality, the
time separation and velocity increments for the doublets were chosen so that the trajectories tend to cross. The sensor array used is a uniform linear array of 32 sensors with sensor spacing 200 feet. The data is recorded starting from time 0.8 seconds until time 2.8 seconds, and the sampling period is 2 milli-seconds. All the calculations use a 10 milli-second time step and a 48 milli-second time gate but different coherency measures. Contoured results will be presented for final comparisons. Table 6.1 summarizes the data characteristics.

Figure 6.5 and 6.6 show the computed velocity spectra by the semblance method and the LCMV method with the conventional spatial smoothing. The final size of the covariance matrix for the LCMV method is 20. Since this is a noise-free case, the improvement of the spatial smoothing is not necessary. In each contour plot the coherency measure has been normalized. It is obvious that the LCMV method has better properties of resolution and parameter identification than semblance. In order to make the comparisons even clearer we find marginal velocity plots for the two methods at three regions. Figures 6.7, 6.8 and 6.9 show the marginal velocity spectra at the first, second and third region respectively. In the first and second regions semblance can hardly resolve the doublets, but the LCMV estimator resolves the doublet very well. For the third region semblance cannot resolve the doublet any more while the LCMV method still can resolve it. Since the data is noise free, the improvement on the spatial smoothed covariance matrix is not used here. We can observe some spurious peaks in the marginal velocity spectra obtained using both semblance and LCMV methods. These may be caused by data errors due to the mismatch between the lag trajectory and the data sample. In this simulation, when the above mismatch exists, the nearest data samples are used to form the sample vectors in the time gate. The high velocity events are more sensitive to these errors.

Kirklin proposed in [65] a display-enhanced semblance to sharpen the peaks in
CHAPTER 6. SEISMIC SIGNAL PROCESSING

<table>
<thead>
<tr>
<th>Event no.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero-offset time (in ms)</td>
<td>900</td>
<td>920</td>
<td>1500</td>
<td>1520</td>
<td>2200</td>
<td>2220</td>
</tr>
<tr>
<td>Velocity (in ft/s)</td>
<td>8000</td>
<td>8200</td>
<td>9000</td>
<td>9200</td>
<td>10000</td>
<td>10200</td>
</tr>
</tbody>
</table>

Table 6.1: Events in the noise-free synthetic data

the velocity spectra as follows:

$$S_{\text{enhanced}} = \frac{1}{1 - S_e}$$  \hspace{1cm} (6.19)

The display-enhanced semblance can increase the visual perception of peaks in the velocity spectrum. The contour and marginal plots of the display-enhanced semblance are given in figures 6.10, 6.11, 6.12, and 6.13 respectively. Although the spectral peaks are more visible in these plots, the peaks shown in the figures are very much biased.

6.6.2 Handling Noisy Seismic Data

To test the ability of the LCMV method to handle noisy seismic data, noisy synthetic data were generated as shown in figure 6.14. Band-passed noise (10 — 60 Hz) is added such that the data have 5 dB SNR, which is defined as follows:

$$\text{SNR} = \frac{A_{\text{ricker}}}{\sqrt{2}\sigma_{\text{bpm}}},$$

where $A_{\text{ricker}}$ is the amplitude of the Ricker wave and $\sigma_{\text{bpm}}$ is the standard deviation of the band-passed noise. There are six events present in the synthetic data and their parameters are given in table 6.2. Velocity analyses using semblance, $K\cdot v$'s method, the LCMV with the conventional spatial smoothing and the LCMV with
Figure 6.4: Noise-free synthetic CDP gather for discrimination threshold test.
Figure 6.5: Velocity contour plot using semblance (clean data).

Figure 6.6: Velocity contour plot using the LCMV velocity estimator with the conventional spatial smoothing (clean data).
Figure 6.7: Marginal velocity spectra using (a) semblance (b) the LCMV method (clean data). The doublet is barely resolved in (a), but is well resolved in (b).
Figure 6.8: Marginal velocity spectra for region 2 using (a) semblance (b) the LCMV method (clean data). The doublet is barely resolved in (a), but is well resolved in (b).
Figure 6.9: Marginal velocity spectra for region 3 using (a) semblance (b) the LCMV method (clean data). The doublet is not resolved in (a), but is well resolved in (b).
Figure 6.10: Velocity contour plot using the display-enhanced semblance (clean data).

Figure 6.11: Marginal plot of the display-enhanced semblance velocity spectrum for region 1 (clean data). The peaks are biased.
Figure 6.12: Marginal plot of the display-enhanced semblance velocity spectrum for region 2 (clean data). The peaks are biased.

Figure 6.13: Marginal plot of the display-enhanced semblance velocity spectrum for region 3 (clean data). The peaks are biased.
the improved spatial smoothing are compared. All coherency measures are normalized such that the maximum value is one. The contour plots of the velocity spectra using semblance and Key's method are shown in figures 6.15 and 6.16. The size of the final covariance matrix is chosen as 4 for Key’s method and 20 for the LCMV estimator. The contour plots of the velocity spectra using the LCMV with the conventional spatial smoothing and with the improved spatial smoothing are given in figures 6.17 and 6.18 respectively. From the contour plots we can see that Key’s method is quite unstable; it does not work at all in this situation. As compared with semblance, the LCMV method with both the conventional and the improved spatial smoothing yield very sharp contours. It is obvious that both semblance and the LCMV method can resolve the first two doublets. From the contour plots we can not tell whether the last doublet has been resolved. To investigate the resolution of the last doublet (in region 3) three marginal spectra have been plotted in figures 6.19, 6.20 and 6.21, which correspond to semblance, the LCMV with the conventional and the improved spatial smoothing, respectively. The marginal plots show that semblance does not resolve this doublet, but the LCMV method does. It also can be seen that the LCMV method with the improved spatial smoothing has a deeper notch between the two peaks. We therefore conclude that 1) optimal velocity estimators proposed in this chapters are clearly superior to the conventional semblance method, and 2) the improved covariance matrix estimator can be used to enhance the resolution power of the LCMV method.

Since the LCMV velocity estimator involves the matrix inversion operation, it requires more computation. When the covariance matrix is singular or nearly singular, the pseudo-inverse is used in place of the inverse operation.

The display-enhanced semblance is also tested in the case. Figures 6.22 and 6.23 show the contour plot of the velocity spectrum and the marginal plot for the third region. For the noisy data the display-enhanced semblance yields similar spectra to those by semblance. It cannot resolve the doublet in the third region.
Multiple Runs

The additive noise present in the noisy data is random, and so are the corresponding velocity spectra. Since the statistics of the noise is unknown, the statistical expectation of the velocity spectra cannot be determined analytically. However, Monte Carlo simulations provide an alternative approach to study the statistical properties of the velocity spectra. Because multiple runs of the experiment we conducted previously will take enormous computer CPU and disk space, we are not able to conduct extensive simulations. Only ten sets of noisy data are generated by choosing different noise generating seed numbers. The ten velocity spectra are found using semblance, the LCMV with the conventional spatial smoothing and the LCMV with the improved spatial smoothing. The overlaid marginal plots for region 3 are given in figure 6.24, 6.25 and 6.26, which correspond to semblance, the LCMV with the conventional spatial smoothing and the LCMV with the improved spatial smoothing respectively. Since semblance can produce similar velocity spectra for all ten runs, ten spectra are plotted in one figure. However, the velocity spectra by the LCMV methods vary dramatically from run to run. We put ten runs into four subplots so that the result for each run can also be seen from the plots. Comparing these marginal plots, we can reach the same conclusion as before; i.e., the semblance method can not resolve the doublet in the third region consistently; while the LCMV methods can resolve the doublet. Furthermore, the improvement of the covariance matrix yields resolutions with deeper notches between signal peaks.
CHAPTER 6. SEISMIC SIGNAL PROCESSING

<table>
<thead>
<tr>
<th>Event no.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero-offset time (in ms)</td>
<td>900</td>
<td>900</td>
<td>1600</td>
<td>1600</td>
<td>2200</td>
<td>2200</td>
</tr>
<tr>
<td>Velocity (in ft/s)</td>
<td>8000</td>
<td>10000</td>
<td>9000</td>
<td>10800</td>
<td>11000</td>
<td>12000</td>
</tr>
</tbody>
</table>

Table 6.2: Events in the noisy synthetic data

Figure 6.14: Noisy synthetic CDP gather, band-pass noise added (10-60 Hz).
Figure 6.15: Velocity contour plot using semblance (noisy data).

Figure 6.16: Velocity contour plot using Key's method (noisy data).
Figure 6.17: Velocity contour plot using the LCMV velocity estimator with spatial smoothing (noisy data).

Figure 6.18: Velocity contour plot using the LCMV velocity estimator with the improved spatial smoothing (noisy data).
Figure 6.19: Marginal plot of the semblance velocity spectrum for region 3 (noisy data). The doublet is not resolved.

Figure 6.20: Marginal plot of the velocity spectrum by the LCMV with the conventional spatial smoothing in region 3 (noisy data). The last doublet is resolved.
Figure 6.21: Marginal plot of the velocity spectrum by the LCMV with the improved spatial smoothing in region 3 (noisy data). The last doublet is resolved with a deeper notch.

Figure 6.22: Velocity contour plot using the display-enhanced semblance (noisy data).
Figure 6.23: Marginal plot of the display-enhanced semblance velocity spectrum for region 3 (noisy data). The doublet is not resolved.

Figure 6.24: Overlaid marginal plot of the semblance velocity spectra for region 3 (noisy data).
CHAPTER 6. SEISMIC SIGNAL PROCESSING

Figure 6.25: Overlaid marginal plots of the velocity spectra by the LCMV with the conventional spatial smoothing in region 3 (noisy data).

Figure 6.26: Overlaid marginal plots of the velocity spectra by the LCMV with the improved spatial smoothing in region 3 (noisy data).
Chapter 7

Summary and Conclusions

7.1 Introduction

This dissertation proposed several techniques for spectral estimation and array processing. The theoretical results were applied to several applications, and their performance was investigated via computer simulations. In this chapter we will summarize the results presented in this dissertation and present our final conclusions.

7.2 Summary of the Dissertation

Chapter 3 presented a covariance matrix estimator which can fully exploit the cross correlations between random vectors, for example the cross correlations between the output vectors of subarrays. The estimator was proposed based on the theory of the Least Squares Linear Prediction of one random vector from another random vector. This theorem was also stated as to estimate the auto-covariance matrix $R_{11}$ from the cross correlations $R_{12}$ and the auto covariance matrix $R_{22}$, such that the trace of error matrix $(R_{11} - R_{11}^c)$ is minimized, where the optimal $R_{11}^c$ is

$$R_{11}^c = R_{12}R_{22}^{-1}R_{21}. \quad (7.1)$$
CHAPTER 7. SUMMARY AND CONCLUSIONS

The improvement of the estimate of the covariance matrix $R_{11}$ was achieved by averaging the estimate from $z_1$ and the one from $R_{12}$ and $R_{22}$ or

$$\hat{R}_{11} = \frac{1}{2}(R_{11} + I_{11}). \quad (7.2)$$

For subspace-based high resolution spectral estimators, the $R_{11}$ was simplified to

$$R_{11}^e = R_{12}R_{21}. \quad (7.3)$$

The proposed covariance matrix estimator demonstrated its significance in some spectral estimation and array processing applications presented in Chapter 4. As a consequence of applying the new covariance matrix estimator, the performance of the high resolution algorithms was enhanced. The new covariance matrix estimator was shown to yield simultaneously smaller error variance, lower SNR threshold, and higher resolution.

Although the proposed covariance matrix estimator can be widely used in spectral analysis and array processing, we considered only three applications in Chapter 4. Spatial smoothing was the first application example we studied, where the cross correlations between subarrays were exploited. It was shown that the new spatial smoothing algorithms can resolve the same number of coherent signals as the conventional spatial smoothing methods. The second example applied the new covariance matrix estimator to utilize the temporal correlations between adjacent snapshots for both narrowband and wideband signals. This kind of temporal correlations existed if the array vectors were obtained by continually sampling the output of the sensor array. The last application example was estimating frequencies of sinusoids embedded in white noise. In this case data vectors were obtained by a sliding window sampling the time sequence. This problem was similar to the spatial smoothing method using a linear array except that there was an order selection problem. For all the applications using the new covariance matrix estimation algorithms yielded higher resolution and smaller estimation errors.
Chapter 5 was relatively independent from Chapters 3 and 4. In this chapter a transformation matrix design method based on the vector p-norm was proposed. It was a generalization of the conventional Frobenius norm method, and provided options to satisfy different design specifications. The new matrix design method was shown to be useful, for example, in CSM and the interpolated arrays.

Chapter 6 discussed the velocity estimation problem in seismic signal processing. The conventional semblance method was found to be the conventional beamformer method. After the review of the optimal beamformers, we proposed an optimal velocity estimator based on the LCMV beamformers. In order to handle the transient property of the seismic data we considered the effects of spatial smoothing in a steered time gate and applied smoothing to improve the estimate of the steered covariance matrix. Simulations were conducted to compare the discrimination threshold of the conventional semblance method and the optimal velocity estimator. It was shown the optimal velocity estimator performs much better than the conventional semblance in discrimination of close events.

7.3 Conclusions and Future Work

The performance analysis of the high resolution spectral estimators has already been studied by many researchers, for example in [67] and [68]. The emphasis of this dissertation was given to the application aspects of spectral analysis and array processing. Our efforts in applying the proposed covariance matrix estimator and the modern array processing techniques to several areas of array processing, spectral estimation, and seismic velocity estimation constituted the major portion of this dissertation. Our research can be further continued in the following directions:

- developing a new estimator for structured covariance matrices with cross correlations considered.
- exploring more applications of the proposed covariance matrix estimator.
• analyzing the performance of the optimal velocity estimator.
Appendix A

Matrix Derivative Rules

There is a real need for matrix calculus in fields such as multivariate analysis and multidimensional signal processing. In this appendix, we will summarize some important derivation rules for matrix calculus; the interested readers can refer to [69], [70] and [71] for more complete coverage of these topics.

A.1 Derivation of Vectors

A.1.1 Definitions

Let \( x \) and \( y \) be vectors of order \( n \) and \( m \) respectively and \( z \) be a scalar. The derivation of vectors is defined as follows:

1. The derivatives of a scalar with respect to a vector:

\[
\frac{\partial z}{\partial x} = \left[ \frac{\partial z}{\partial x_1}, \frac{\partial z}{\partial x_2}, \ldots, \frac{\partial z}{\partial x_n} \right]^T. \tag{A.1}
\]

2. The derivatives of a vector with respect to a scalar:

\[
\frac{\partial x}{\partial z} = \left[ \frac{\partial x_1}{\partial z}, \frac{\partial x_2}{\partial z}, \ldots, \frac{\partial x_n}{\partial z} \right]. \tag{A.2}
\]
3. The derivatives of a vector with respect to a vector:

\[ \frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix}. \]  

(A.3)

A.1.2 Chain Rule

Let \( x, y \) and \( z \) be column vectors of order \( n, r \) and \( m \) respectively. We then have the chain rule:

\[ \frac{\partial z}{\partial x} = \frac{\partial y}{\partial x} \frac{\partial z}{\partial y}. \]  

(A.4)

A.1.3 Some Useful Formulas

Let \( A \) be an \( n \times n \) matrix and \( x \) an \( n \) vector. We assume \( A \) is not a function of \( x \). Then we have the following useful derivatives:

\[ \frac{\partial Ax}{\partial x} = A^T, \]  

(A.5)

\[ \frac{\partial x^T A}{\partial x} = A, \]  

(A.6)

\[ \frac{\partial x^T x}{\partial x} = 2x, \]  

(A.7)

\[ \frac{\partial x^T A x}{\partial x} = 2A x. \]  

(A.8)

A.2 Derivation of Matrices

Let \( X, Y \) and \( z \) be an \( m \times n \) matrix, a \( p \times q \) matrix and a scalar respectively. We have the following definitions for the derivation of matrix:
APPENDIX A. MATRIX DERIVATIVE RULES

A.2.1 Definitions

1. The derivatives of a scalar with respect to a matrix:

The derivative of $z$ with respect to $X$, denoted by $\frac{\partial z}{\partial X}$, is defined as the following matrix of order $m \times n$

$$
\frac{\partial z}{\partial X} = \begin{bmatrix}
\frac{\partial z}{\partial x_1} & \frac{\partial z}{\partial x_2} & \cdots & \frac{\partial z}{\partial x_n} \\
\frac{\partial z}{\partial x_{m1}} & \frac{\partial z}{\partial x_{m2}} & \cdots & \frac{\partial z}{\partial x_{mn}}
\end{bmatrix}.
$$

(A.9)

2. The derivatives of a matrix with respect to a scalar:

$$
\frac{\partial X}{\partial z} = \begin{bmatrix}
\frac{\partial x_1}{\partial z} & \frac{\partial x_2}{\partial z} & \cdots & \frac{\partial x_n}{\partial z} \\
\frac{\partial x_{m1}}{\partial z} & \frac{\partial x_{m2}}{\partial z} & \cdots & \frac{\partial x_{mn}}{\partial z}
\end{bmatrix}.
$$

(A.10)

3. The derivatives of a matrix with respect to a matrix:

$$
\frac{\partial Y}{\partial X} = Y \otimes \frac{\partial}{\partial X} = \left\{ \frac{\partial y_{ij}}{\partial x_{kl}} \right\},
$$

where $\otimes$ denotes Kronecker product.

A.2.2 Product and Chain Rules

Let $X$, $Y$ and $Z$ be matrices of order $m \times n$, $n \times v$ and $p \times q$ respectively, then we have the product rule

$$
\frac{\partial (XY)}{\partial Z} = \frac{\partial X}{\partial Z} (I_q \otimes Y) + (I_p \otimes X) \frac{\partial Y}{\partial Z},
$$

(A.12)
and the chain rule

$$\frac{\partial Z}{\partial X} = \sum_{\alpha,\beta} \left( \frac{\partial y_{\alpha\beta}}{\partial X} \otimes I_p \right) \left( I_n \otimes \frac{\partial Z}{\partial y_{\alpha\beta}} \right).$$  \hspace{1cm} (A.13)

### A.2.3 Some Useful Formulas

Let $X$, $A$, and $B$ be matrices of such orders that the operation in all the following equations are defined. Furthermore, $A$ and $B$ are not a function of $X$. Under these assumptions, we have the following useful matrix derivatives

\[
\frac{\partial \log |X|}{\partial X} = (X^{-1})^T, \hspace{1cm} (A.14)
\]

\[
\frac{\partial |X|^r}{\partial X} = r|X|^{r-1}(X^{-1})^T, \hspace{1cm} (A.15)
\]

\[
\frac{\partial \text{tr}(AX)}{\partial X} = A^T, \hspace{1cm} (A.16)
\]

\[
\frac{\partial \text{tr}(AX^T)}{\partial X} = A, \hspace{1cm} (A.17)
\]

\[
\frac{\partial \text{tr}(X^TAXB)}{\partial X} = AXB + A^TXB^T, \hspace{1cm} (A.18)
\]

\[
\frac{\partial \text{tr}(X^TX)}{\partial X} = 2X, \hspace{1cm} (A.19)
\]

\[
\frac{\partial \text{tr}(X^n)}{\partial X} = n(X^{n-1})^T, \hspace{1cm} (A.20)
\]

\[
\frac{\partial \text{tr}(AX^{-1}B)}{\partial X} = -(X^{-1}BAX^{-1})^T. \hspace{1cm} (A.21)
\]
Appendix B

Calculation of Fisher Information Matrix

B.1 Another Form for the Fisher Information Matrix

Lemma B.1: Subject to regularity conditions that permit the interchange of the order of differentiation and integration, the $ij$th element of the Fisher information matrix may be cast in the following form

$$J_{ij} = -E \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x|\varphi) \right]. \quad (B.1)$$

Proof: See [28].

B.2 Information Matrix for Gaussian Random Vectors

In this appendix, we will give the formula to calculate the exact information matrix for Gaussian random processes. Let the real Gaussian random vector $x$ of order
\( M \) have the logarithm of the joint density function \( f(x) \) as

\[
\log f(x) = -\frac{M}{2} \log 2\pi - \frac{1}{2} \log \text{det} R(\theta) - \frac{1}{2} [x - m(\theta)]^T R^{-1}(\theta) [x - m(\theta)], \tag{B.2}
\]

where \( \theta \) is the real parameter vector of order \( p \). Differentiation with respect to \( \theta_k \) yields

\[
\frac{\partial \log f(x)}{\partial \theta_k} = -\frac{1}{2} \text{tr} \left\{ R^{-1}(\theta) \frac{\partial R^{-1}(\theta)}{\partial \theta_k} \right\} + \frac{1}{2} [x - m(\theta)]^T R^{-1}(\theta) \frac{\partial R^{-1}(\theta)}{\partial \theta_k} [x - m(\theta)]
\]

\[
+ \left[ \frac{\partial m(\theta)}{\partial \theta_k} \right]^T R^{-1}(\theta) [x - m(\theta)]. \tag{B.3}
\]

Multiplying \( \frac{\partial \log f(x)}{\partial \theta_k} \) by \( \frac{\partial \log f(x)}{\partial \theta_l} \) and taking expectation yields

\[
[J(\theta)]_{k,l} = -\frac{1}{4} \text{tr} \left\{ R^{-1}(\theta) \frac{\partial R(\theta)}{\partial \theta_k} \right\} \text{tr} \left\{ R^{-1}(\theta) \frac{\partial R(\theta)}{\partial \theta_l} \right\}
\]

\[
+ \left[ \frac{\partial m(\theta)}{\partial \theta_k} \right]^T R^{-1}(\theta) \left[ \frac{\partial m(\theta)}{\partial \theta_l} \right]
\]

\[
+ \frac{1}{4} E \left\{ [x - m(\theta)]^T R^{-1}(\theta) \frac{\partial R(\theta)}{\partial \theta_k} R^{-1}(\theta) \frac{\partial R(\theta)}{\partial \theta_l} \right\}
\]

\[
+ \frac{1}{4} E \left\{ [x - m(\theta)]^T R^{-1}(\theta) \frac{\partial R(\theta)}{\partial \theta_l} R^{-1}(\theta) [x - m(\theta)] \right\}. \tag{B.4}
\]

After some algebraic manipulation [72], we obtain

\[
[J(\theta)]_{k,l} = \frac{1}{2} \text{tr} \left\{ R^{-1}(\theta) \frac{\partial R(\theta)}{\partial \theta_k} R^{-1}(\theta) \frac{\partial R(\theta)}{\partial \theta_l} \right\}
\]
If we take \( N \) independent random vectors, the corresponding Fisher information matrix is given by

\[
\mathbf{J}(\theta)_{k,l} = \frac{N}{2} \text{tr} \left\{ R^{-1}(\theta) \frac{\partial R(\theta)}{\partial \theta_k} R^{-1}(\theta) \frac{\partial R(\theta)}{\partial \theta_l} \right\}
+ N \left[ \frac{\partial m(\theta)}{\partial \theta_k} \right]^T R^{-1}(\theta) \left[ \frac{\partial m(\theta)}{\partial \theta_l} \right].
\] (B.6)

### B.3 Information Matrix for Array Processing

For array processing, the Fisher Information Matrix is given as [67]

\[
J = \frac{\sigma^2}{N} \left\{ \sum_{i=1}^{N} \text{Re}[S^H(t_i)D^H[I - A^H A]^{-1}A]DS(t_i)] \right\},
\] (B.7)

where

\[
S(t_i) = \text{diag}[s_1(t_i), s_2(t_i), \ldots, s_d(t_i)],
\]

\[
D = [d(\theta_1), d(\theta_2), \ldots, d(\theta_d)],
\]

with \( d(\theta_i) = da(\theta_i)/d\theta_i \).
Appendix C

Ricker Wavelets

Let the Gaussian function $g(t)$ be given by

$$g(t) = e^{-t^2/2\sigma_t^2}.$$  \hspace{1cm} (C.1)

Taking the second order derivative of $g(t)$ with respect to $t$ yields the expression for the ricker wave,

$$R(t) = \frac{1}{\sigma_r^2} \left( \frac{t^2}{\sigma_r^2} - 1 \right) e^{-t^2/(2\sigma_r^2)}.$$  \hspace{1cm} (C.2)

Normalizing its peak to $-1$, we have the normalized version of the ricker wave as follows:

$$R_n(t) = \left( \frac{t^2}{\sigma_r^2} - 1 \right) e^{-t^2/(2\sigma_r^2)}.$$  \hspace{1cm} (C.3)

In order to find the relation between the peak frequency of the power spectrum of the ricker wave and the parameter $\sigma_r$, the Fourier transform of $R_n(t)$, $F(\omega)$, is calculated:

$$F(\omega) = \sqrt{2\pi} \sigma_r^3 \omega^2 e^{-\omega^2 / 2\sigma_r^2}.$$  \hspace{1cm} (C.4)
Taking the first order derivative of $F(\omega)$ with respect to $\omega$ and setting it to zero we have

$$\sigma_r^2 \omega^3 - 2\omega = 0. \quad (C.5)$$

Neglecting the negative and zero solutions, we find the normalized peak frequency, $f_p$, is given by

$$f_p = \frac{1}{\sqrt{2\pi}\sigma_r}. \quad (C.6)$$
Bibliography


