Distance Measures Among Classes of Deconvolved Sea-bottom Sonar Responses

By

Jeonghun Park

UNIVERSITY OF VICTORIA

Distance Measures Among Classes of Deconvolved Sea-bottom Sonar Responses

By Jeonghun Park
B.S., Chungbuk National University, Korea, 2000

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

Master of Applied Science
in the Department of Electrical and Computer Engineering

Supervisors: Dr. R. L. Kirlin and Dr. A. Zielinski

ABSTRACT

Efforts to extract information regarding the surficial composition of the ocean bottom have increased as acoustic echosounders have developed and the availability of computing power has corresponded with advances in signal processing techniques. With enhanced classification of bottom types from echo soundings as our objective, this thesis focuses on two major technical steps, deconvolution and class distance measure, to facilitate the objective.

First, since the sea-bottom sonar response is the result of convolution of the bottom's impulse response and the source ping, and the noise and delay component are observed in the response signals from the unknown sea bottom, we separate the source ping, which is independent of depth, through deconvolution. We present the processing details of least-square estimation of deconvolved sea-bottom sonar responses and also enhance the results using *singular value decomposition* (SVD) to produce rank-reduced (noise reduced) signals. The deconvolution algorithms resulting from this research are applied to three distinct real-world data sets which were collected.

Secondly, because direct minimization of the error probability to determine the optimum signal is often difficult to calculate, the distance measure is a simple alternative to the error probability. We discuss two common distance measures for class separation: divergence and Bhattacharyya. Further, we are also interested in measuring the distance between class pairs as data rank is reduced using SVD. We pay particular attention to the advantage of having knowledge of the transmitted carrier phase and determine its influence on the resulting distance between classes.

Abstract

Efforts to extract information regarding the surficial composition of the ocean bottom have increased as acoustic echosounders have developed and the availability of computing power has corresponded with advances in signal processing techniques. With enhanced classification of bottom types from echo soundings as our objective, this thesis focuses on two major technical steps, deconvolution and class distance measure, to facilitate the objective.

First, since the sea-bottom sonar response is the result of convolution of the bottom's impulse response and the source ping, and the noise and delay component are observed in the response signals from the unknown sea bottom, we separate the source ping, which is independent of depth, through deconvolution. We present the processing details of least-square estimation of deconvolved sea-bottom sonar responses and also enhance the results using *singular value decomposition* (SVD) to produce rank-reduced (noise reduced) signals. The deconvolution algorithms resulting from this research are applied to three distinct real-world data sets which were collected.

Secondly, because direct minimization of the error probability to determine the optimum signal is often difficult to calculate, the distance measure is a simple alternative to the error probability. We discuss two common distance measures for class separation: divergence and Bhattacharyya. Further, we are also interested in measuring the distance between class pairs as data rank is reduced using SVD. We pay particular attention to the advantage of having knowledge of the transmitted carrier phase and determine its influence on the resulting distance between classes.

Table of Contents

List of Symbols

$B = \mu(1/2)$	Bhattacharyya Distance
$\mathbf{C}_y$	Sample covariance matrix of **y**
d	divergence
E_{lse}	Minimum least-squares error
E_{nlse}	Normalized least-squares error
E_i	Average energy in class i
e	Error vector between the measurement and the estimate of the measurement
$\|\mathbf{e}\|^2$	Euclidean norm or magnitude of error vector **e**
h	Reflection coefficients
$\mathbf{h}_{rank}$	Rank reduced coefficients
$\mathbf{h}_{real}$	Real coefficients
$\mathbf{h}_{image}$	Imaginary coefficients
$\mathbf{h}_{complex}$	Complex estimated coefficients
$I(2,1)$	Average discriminating information for class ω_2
L	Length of the transmitted signal
M	Dimensionality of a raw vector
M_a	Mahalobis distance
$\mathbf{m}_i$	Mean vector of class i
N	Number of vectors in a data set

P	Dimensionality of a coefficients vector
P_i	Priori probability of class i
$q_i(\mathbf{x})$	Posteriori probability of class i
$\mathbf{R}$	Covariance matrix
$\mathbf{S}$	Diagonal matrix of the eigenvalue matrix of $\mathbf{X}$
$\mathbf{S}_s$	Non-zero diagonal matrix of the eigenvalue matrix of $\mathbf{X}$
$\mathbf{U}$	Singular vectors of $\mathbf{X}$
$\mathbf{U}_s$	Nonzero singular vectors of $\mathbf{X}$
$\mathbf{U}_n$	Zero singular vectors of $\mathbf{X}$
$\mathbf{V}$	Eigenvectors of $\mathbf{R}$
$\mathbf{V}_s$	Signal subspace eigenvectors
$\mathbf{V}_n$	Noise subspace eigenvectors
$\mathbf{v}_i$	Eigenvectors of $\mathbf{R}$
$\mathbf{w}$	Independent zero mean white Gaussian noise vector
$\mathbf{X}$	Known transmitted waveform
$\mathbf{X}^H$	Conjugate transpose of $\mathbf{X}$
$\mathbf{x}$	Transmitted signal vector
$\tilde{\mathbf{x}}$	Hilbert transform of $\mathbf{x}$
$\mathbf{x}_a$	Analytic transmitted signal vector
$\mathbf{y}$	Zero-mean data vector (sonar response)
$\hat{\mathbf{y}}$	Noiseless estimated output
$\hat{\mathbf{y}}_{noise}$	Noisy output
$\hat{\mathbf{y}}_{rank}$	Rank-reduced output
$\mathbf{y}_{align}$	Aligned reflected data

λ_i	Eigenvalues of **R**
$\mathbf{\Lambda}$	Diagonal matrix of the eigenvalue matrix of **R**
σ_n^2	Noise powers
ω_i	Class i
ε	Bayes error
ε_u	Upper bound of error
$\mu(s)$	Chernoff distance
$\mathbf{\Sigma}_i$	Diagonal element of the diagonal covariance matrix for class i

List of Tables

List of Figures

Acknowledgements

I really thank my supervisor, Dr. R. Lynn Kirlin for his guidance, support, endurance and encouragement during the course of this thesis work. His undivided motivation and endeavour have been the major factors in the success of this thesis.

I thank my co-supervisor, Dr. Adam Zielinski for his valuable comments during the progress of this research.

I would like to thank Quester Tangent Corporation, in particular Dr. Jon Preston, Karl Rhynas and Bloomer Steve for providing us with data sets and technical helps.

I thank Dr. Colin Bradley and Dr. Fayez Gebali for their services as my supervisory committee members, and Dr. Ross Chapman for his valuable comments for our research.

I am very grateful to all the friends at DSP lab for creating the friendly atmosphere in which I have had the pleasure to work. In particular, I wish to thank Zian Wang and Lu Xiaoli for their unforgettable help and suggestions.

Finally, I thank my parents for their support, motivation and unyielding belief that I would finish. Also, I would like to thank my fiancée, Eunhee, for giving me love and support.

Chapter 1

Introduction

The ability to extract information about the ocean bottom has proved elusive since it was first actively pursued in the mid-seventies, when oceanographic studies were initially undertaken to model the physical process. However, with the availability of rapidly increasing processing power, many advances have occurred.

1.1 Motivation

Knowledge of the ocean bottom has always been important to human life. Thus mankind has developed appropriate exploration technologies. For the first several thousand years, simple shallow water bathymetry were developed for navigational safety. But more recently, systematic surveys began to be conducted and a demand for new technologies arose.

A fundamental change in survey technology occurred with the development of the acoustic echo-sounder in the 1920's. This technique was revolutionary in that it was non-invasive and permitted large-scale surveys to occur. In the 1970's electronic hardware became sufficiently inexpensive that the idea of processing the data to extract bottom information characteristics became feasible. By using new low-cost computing facilities in the mid-1980's some groups made another significant forward leap in seabed classification.

The current motivation for seabed classification is both military and commercial. Decreasing world-wide fish stocks have typically required that increasingly sophisticated technology be employed to locate the schools. With the development of interpretation of seabed classification, the ship captains can predict good fishing areas through habitat identification.

1.2 Historical development

Early papers tended to focus on analysis of the raw data of ping-to-ping coherence measures. In Milligan, LeBlanc, and Middleton [1]'s works, the basic premise is that one can interpret a set of data and extract the ensemble statistics, without relying on a *priori* models. They generated a covariance matrix for an ensemble of data, and used eigenanalysis to obtain the principal components of the data set. A reduced number of components can be used to obtain a minimal dimension when a received signal is reconstructed. However, because the concept was developed in the mid-seventies, a major shortfall of this work was a lack of modern computational capability.

However, in Caughey's thesis [2], feature reduction via principal component analysis was re-examined, and numerical evaluation of feature suitability was done in Kavli [32]. LeBlanc [3] used a chirp sonar for sediment classification, and he used fully digital system, including the support for the associated digital signal processing.

Additionally, LeBlanc is one of the first to discuss the deconvolution of the source signal from the return signal. "Blind deconvolution of echosounder envelopes [4]" describes the deconvolution of the digitized envelope of the sea-bottom response to

determine the impulse response. An alternate approach to the determination of the impulse response is through least mean-square estimation of the complete impulse response given the raw response data rather than just its envelope. It is this latter approach that we investigate in this thesis.

1.3 Thesis outline

The remainder of this thesis is divided into five chapters.

Chapter 2 is a technical chapter dealing with the processing details of least-square estimation of deconvolved sea-bottom sonar responses using singular value decomposition (SVD) to produce rank-reduced signals [6, 31]. Since the sea-bottom sonar response is the result of convolution of the bottom's impulse response and the source ping, we might separate the source ping, which is independent of depth, through deconvolution. This chapter lays out the procedure for the deconvolution and simulates such a process. Because a rank-deficient data covariance matrix is involved we also investigate the effects of rank reduction on the raw data before further processing.

Complex deconvolution is involved in order to catch all the coherent information in a phase-shifted reflection in a real environment.

Chapter 3 is the second of two technical chapters dealing with the measurement of distance between classes. Distance measures are studied because it is generally difficult to carry out a direct determination of the error probability. In the first part we discuss two common distance measures for class separation: divergence and Bhattacharyya. In the second part we pay particular attention to the use of knowledge of the transmitted

carrier phase and determine its influence on the resulting distance between classes.

In Chapter 4 because we wish to deconvolve the transmitted signal from the response data, we utilize a record of the transmitted signal that has been obtained in a laboratory water tank.

In a later section of this chapter the deconvolution algorithms resulting from this research are applied to three distinct real-world data sets which were collected. We explore the enhancement of raw data and estimated coefficients by using principal component analysis (PCA). The results at each stage are discussed in context with the issues raised in chapter 2.

Chapter 5 presents the results of the measured values of divergence and Bhattacharyya distances between class pairs from among the preclassified data provided. Comparisons of cluster pair distances are made for the various processings that might be applied before classification. The raw data reflected from the sea bottom after transmitting a signal downward and the estimated coefficients obtained by deconvolving the transmitted signal from the raw data have been pre-categorized according to their associated transmitted frequency, bottom type and incident angle. The first major section of this chapter discusses the measures of class distance taken only from envelopes of raw data and magnitudes of the estimated coefficients. The section also discusses the distances obtained after suppressing the noise component in raw data by using SVD and obtaining low rank approximations.

The tabulated measures are indicative of the relative ease of discrimination between classes and demonstrate the utility of various preprocessings such as deconvolution and/or raw data rank reduction to enhance SNR.

In the second part, we determine the distance measures after complex deconvolution has

been applied. The tabulated distance measures of complex coefficients are compared with those that use only magnitude of the estimated coefficients.

Chapter 6 presents the conclusions of this research and makes recommendations for further work.

Lastly, appendixes are provided, including a discussion of start time estimation error problems.

1.4 Trademarks and copyrights

The following names are used throughout this thesis.

- "QTC", "QTC VIEW", and "ISAH-S" are trademarks of Quester Tangent Corp.
- "MATLAB" is a trademark of The MathWorks, Inc.

Chapter 2

Least-Square Estimation of Deconvolved Sea-bottom Sonar Responses

We are interested in determining the characteristic of sonar responses of the sea-bottom and comparing variations not only with bottom class but also with frequency, beamwidth, angle of incidence and depth. In order to better estimate and compare these variations, we might need to deconvolve the transmitted signal from the response. This chapter lays out the procedure for the deconvolution and simulates such a process. Because a rank-deficient matrix is involved we investigate the effects of rank reduction by using singular value decomposition (SVD). This method is verified through several simulated examples.

We are also interested in variations in start times of reflections in real environment causing phase shift error in deconvolution. Complex deconvolution is introduced for incorporating the coherent information that might be determined from phase-shift or delay in the reflections.

We begin with a discussion of properties of source measurement environment and the vector subspace which is related to the noise suppression.

2.1 Properties of source measurement environment

When an echo sounder transmits to the sea bottom, the transmitted wave is reflected.

Before we process the reflection from sea bottom we discuss several properties of echo sounder and reverberation mechanisms in the sea. In this section we briefly review several phenomena of the returned signal such as *beam pattern* and properties of reverberation.

2.1.1 Type of reverberation

When an acoustic signal goes under the sea there are re-radiations of the acoustic energy from the bottom boundary and inhomogeneities of many kinds such as schools of fish or tiny particles of dust that give the deep sea its blue color. Reverberation is the total sum of the scattering contribution from all scatterers.

The reverberation in the sea is of two basically different classes. The marine life and the inhomogeneous structure of the sea itself are the examples of scatterers of volume scattering. Another type of scatterer is *bottom reverberation*. It originates at scatterers on or near the sea bottom.

For sea bottom backscattering, although particle size of bottom materials serves as a means of classifying bottoms in terms of acoustic backscattering, the roughness of the sea bottom can be the dominant characteristic for backscattering.

Another important role in the scattering process is the penetration of sound into the bottom and its subsequent reradiation as backscattering at low frequency and high grazing angles, where reflection and scattering from subsurface layers and other inhomogeneities may occur [17].

Figure 2.1 shows bottom acoustic scattering mechanisms.

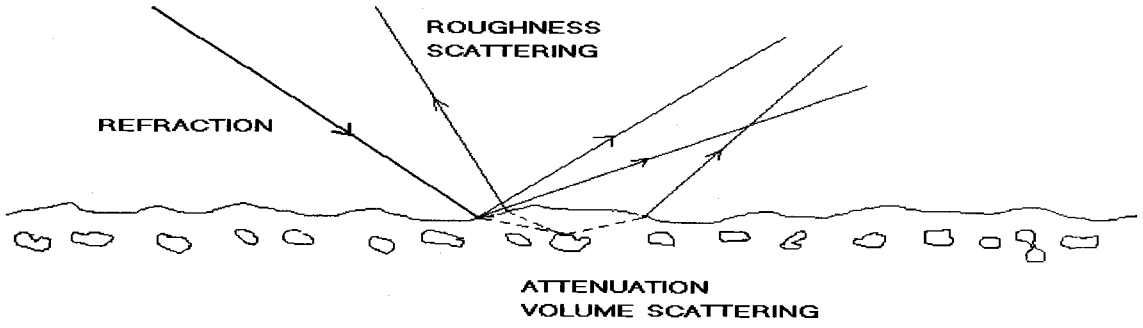


Figure 2.1: Bottom acoustic scattering mechanisms, including reflection and scattering at the water-bottom interface and attenuation and scattering in the sediment[20]

2.1.2 Bottom backscattering with angle and bottom type

The influence of particle size on the reverberation from the seabed was greater over rocky bottoms than over mud bottoms. We realize that the size of the particles comprising a sedimentary bottom is one of the indirect indicators of acoustic scattering. Therefore the bottom backscattering strength on a smooth bottom such as mud or silt is less than that on a rough bottom.

As for the influence of incident angle, a Lambert's law relationship between backscattering strength and angle appears to provide a good approximation of reflections, where in the backscattering strength varies as the square of the sine of the grazing angle [17].

In small incident angles the main contribution of backscattering at high frequency is surface scattering from roughness [20, 23, 24]. But at low frequency sound penetration of the sediment increases, and the effects of volume scattering and reflection from the basement become important.

In intermediate angles the volume scattering is the major backscattering at both high

and low frequency [22].

2.1.3 Properties of different frequencies and beamwidth

There are several properties of acoustic signal factors such as beamwidth and frequency [8]. Firstly beamwidth is a measure of the size of the conical shaped field of the transmitted echo pulse. The size of the seabed acoustic footprint is a function of the beamwidth and the water depth. Secondly frequency affects the attenuation and reflectivity of the acoustic pulse for the various media and determines the spatial resolution of the data coming to the sounder.

High frequency signals greater than 100 kHz typically provide greater resolution, suffer greater attenuation in the water column and penetrate only centimeters into the seabed.

Low frequency signals, 10 kHz to 100 kHz, provide less resolution than the high frequencies, smaller signal losses in the water and penetrate tens of centimeters into the seabed. Low frequency transducers generally have larger beamwidths.

2.1.4 Acoustic data acquisition

For data collection we are grateful to Dr M. Preston, Senior Scientist, Quester Tangent Corporation, who has encouraged QTC to provide us with our data sets. Sounder echoes of 24 and 200 kHz carrier from the same sites in bays and harbors around southern east Vancouver Island were recorded by a digital acquisition system (ISAH-S) and QTC VIEW seabed classification system developed from Quester Tangent Corporation. QTC VIEW systems typically use the signal from a single frequency echo sounder. The system is connected in parallel with the echo sounder transducer and digitally extracts

the echo trace. The digitized data were stored in a computer with an easily accessible file format, which made extraction of data trivial using MATLAB.

The 24 kHz transmitted signal was characterized by a 0.167 ms pulse length and a 20° beam width. The 200 kHz transmitted signal was characterized by a 0.16ms pulse length,4° beam width.

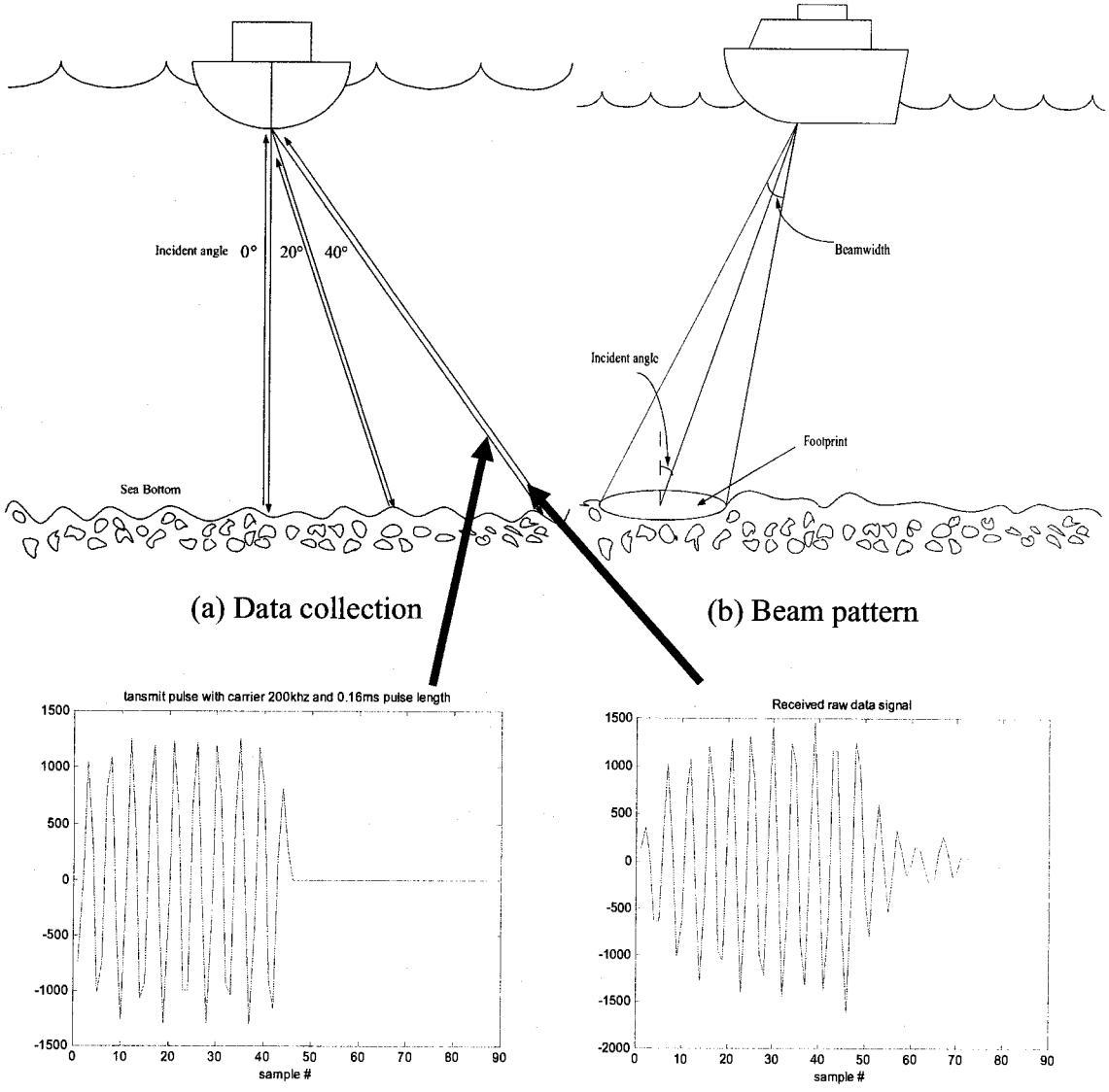


Figure 2.2: (a) paths of transmission and reflection (b) the beam pattern, (c) typical transmitted signal and (d) reflected signal convolved with bottom response plus noise.

Furthermore, the incident angles of the transducer were also run at 0°, 20° and 40°. Figure 2.2 shows the beam pattern and collection of the reflected data in the real environment.

2.2 Vector subspaces

Over the past decade much research has been devoted to the understanding and application of signal subspace and noise subspace processing. Based on the linear statistical model for vector data, all data vectors are linear combinations of their signal and noise components. The signal and noise components can be partitioned into two vector subspaces. Following subsections discuss vector subspaces in detail.

2.2.1 Covariance matrix structure

The linear statistical model assumes that the zero-mean data vector (sonar response) **y** is

$$\mathbf{y} = \mathbf{Xh} + \mathbf{w} \tag{2.1}$$

where **h** is an $P \times 1$ reflection coefficient vector, **X** is an $M \times P$ matrix of the known transmitted waveform, **Xh** is the estimate $\hat{\mathbf{y}}$ of **y** and **w** is an $M \times 1$ vector of independent zero mean white Gaussian noise elements.

When the N samples of **y, h** and **w** are arranged into the columns of **Y**, **H** and **w** , the sample covariance matrix is

$$\begin{aligned} \mathbf{C}_{\mathbf{y}} &= \frac{1}{N}\mathbf{Y}\mathbf{Y}^{H} \\ &= (\mathbf{XHH}^{H}\mathbf{X}^{H} + \mathbf{XHW}^{H} + \mathbf{WH}^{H}\mathbf{X}^{H} + \mathbf{WW}^{H})/N \end{aligned} \tag{2.2}$$

Because of the independence of **W** with **XH**, the expectation of the product terms with **H** are zero.

Therefore, the sample covariance matrix $\mathbf{C}_y$ of **y** is

$$\mathbf{C}_y = (\mathbf{XHH}^H\mathbf{X}^H + \mathbf{WW}^H)/N \tag{2.3}$$

The covariance matrix **R** can be obtained as

$$\mathbf{R} = E\{\mathbf{C}_y\} = \mathbf{X}E\{\mathbf{HH}^H\}\mathbf{X}^H + \sigma_n^2\mathbf{I} \tag{2.4}$$

where $E\{\mathbf{WW}^H\} = \sigma_n^2\mathbf{I}$.

2.2.2 Eigenstructure and subspaces

There are several properties regarding the eigenstructure of a data covariance matrix [5, 13]. The eigenvectors $\mathbf{v}_i$ of **R** are assigned to the columns of **V**, and its associated eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_M$ have the following properties:

1) The r (r<M) eigenvectors that are associated with the largest r eigenvalues span the same vector space as the columns of **X**, the signal subspace.
2) The smallest $M-r$ eigenvalues of **R** are all equal to σ_n^2.
3) The eigenvectors associated with 2) span the so-called noise subspace.
4) All eigenvectors in the signal subspace are orthogonal to those in the noise subspace.

(Note that the noise subspace term is not strictly correct because noise has equal power σ_n^2 in all dimensions including those of the signal subspace.)

Thus **R** can be rewritten,

$$\mathbf{R} = \sum_{i=1}^{r} \lambda_i \mathbf{v}_i \mathbf{v}_i^H + \sigma_n^2 \sum_{i=r+1}^{M} \mathbf{v}_i \mathbf{v}_i^H \tag{2.5.a}$$

$$= \mathbf{V}_s \mathbf{\Lambda}_s \mathbf{V}_s^H + \mathbf{V}_n \mathbf{\Lambda}_n \mathbf{V}_n^H \tag{2.5.b}$$

$$= [\mathbf{V}_s \mathbf{V}_n] \begin{bmatrix} \mathbf{\Lambda}_s & 0 \\ 0 & \mathbf{\Lambda}_n \end{bmatrix} [\mathbf{V}_s \mathbf{V}_n]^H \tag{2.5.c}$$

$$= \mathbf{V} \mathbf{\Lambda} \mathbf{V}^H \tag{2.5.d}$$

where the eigenvalues $\lambda_1, \lambda_2 ... \lambda_M$ of the covariance matrix have been partitioned to give $\mathbf{\Lambda_s}$ and $\mathbf{\Lambda_n}$,diagonal eigenvalues matrices of size r and $M-r$ respectively, and the eigenvector matrix $\mathbf{V}$ has been partitioned into signal subspace eigenvectors $\mathbf{V_s}$ and noise subspace eigenvectors $\mathbf{V_n}$.

From the four eigenstructure properties above, if an eigenvalue of $\mathbf{X}E\{\mathbf{H}\mathbf{H}^H\}\mathbf{X}^H$ is λ_k (k=0, 1,... , r), then an eigenvalue of **R** in signal subspace is $\lambda_k + \sigma_n^2$ and an eigenvalue of **R** in noise subspace is σ_n^2 [5].

Now we discuss the data covariance matrix which is used to estimate the two subspaces. When SNR is sufficiently high and sample size sufficiently large, then $M-r$ dimensions of noise power can be removed from the data, allowing processing to proceed with higher SNR data.

2.3 Principal component analysis (PCA)

Principal components can be found in two ways. The first is through the eigenanalysis of the covariance matrix of the data set. The second is through SVD of the raw data. Both will yield similar results for normal data (the squared singular values divided by

the vector length are equal to the eigenvalues of the covariance matrix.)

PCA identifies the major directions of variation in the data. PCA achieves this by computing the eigenvectors and eigenvalues of the covariance matrix of the data. Keeping only a few eigenvectors corresponding to the largest eigenvalues, PCA can be also used as a tool to reduce the dimensions of the data while retaining the major variation of the data. Equivalently, from SVD we can factor rectangular matrices into orthogonal components. The process allows rectangular 2-D data to be separated into major and minor energy.

The number of dimensions of the reduced feature space depends on the data. Obviously, reducing the data to just one dimension is always possible, but the resultant classifier will probably have high rates of misclassifications. On the other hand, if a large number of dimensions are used then the less significant axes will contain mainly contributions from noise, and will actually degrade system classification rates. Therefore examination of the sorted eigenvalues gives an idea of the suitable number of dimensions. At the level at which the eigenvalues have decayed a couple of orders of magnitude, their corresponding dimensions are unlikely to provide much information [26, 27].

2.4 Use of the SVD in computing the least-squares solution

The error vector between the measurement and the estimate of the measurement using the deconvolution model is defined as

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{h} \tag{2.6}$$

where $\mathbf{h}$ is an $P \times 1$ reflection coefficient vector, $\mathbf{X}$ is an $M \times P$ matrix of the known

transmitted waveform, $\mathbf{y}$ is the data (sonar response) vector and $\mathbf{Xh}$ is the estimate $\hat{\mathbf{y}}$ of $\mathbf{y}$.

Deconvolution is to determine the vector $\mathbf{h}$ that minimizes the squared norm of $\mathbf{e}$.

A problem that may occur is that the signal matrix $\mathbf{X}$ is rank deficient or poorly conditioned and not directly invertible [6]. A solution is to use a lower rank approximation to $\mathbf{X}$ that allows inversion. Thus we look into SVD. Applying SVD to $\mathbf{X}$ in section 2.1.1 gives

$$\mathbf{e} = \mathbf{y} - \mathbf{USV}^H\mathbf{h} \tag{2.7}$$

Since $\mathbf{U}$ is an orthogonal matrix, $\|\mathbf{e}\|^2 = \|\mathbf{U}^H\mathbf{e}\|^2$, and the vector $\mathbf{h}$ that minimizes $\|\mathbf{e}\|^2$ also minimizes $\|\mathbf{U}^H\mathbf{e}\|^2$

Now partition the orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$ partitioned as

$$\mathbf{U} = [\mathbf{U}_s \quad \mathbf{U}_n]$$

$$\mathbf{V} = [\mathbf{V}_s \quad \mathbf{V}_n] \tag{2.8}$$

where $\mathbf{U}_s$ and $\mathbf{V}_s$ contain the column vectors associated with the nonzero singular values of $\mathbf{X}$, and $\mathbf{U}_n, \mathbf{V}_n$ contain the column vectors associated with the zero singular values of $\mathbf{X}$.

We can obtain

$$\mathbf{U}^H\mathbf{e} = \begin{bmatrix} \mathbf{U}_s^H \\ \mathbf{U}_n^H \end{bmatrix}\mathbf{y} - \begin{bmatrix} \mathbf{S}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix}\mathbf{h}$$

$$= \begin{bmatrix} \mathbf{U}_s^H\mathbf{y} - \mathbf{S}_s\mathbf{V}_s^H\mathbf{h} \\ \mathbf{U}_n^H\mathbf{y} \end{bmatrix} \tag{2.9}$$

where $\mathbf{S}_s$ is the non zero diagonal matrix of the eigenvalue matrix of $\mathbf{X}$.

Hence, $\|\mathbf{U}^H\mathbf{e}\|^2 = \|\mathbf{U}_s^H\mathbf{y} - \mathbf{S}_s\mathbf{V}_s^H\mathbf{h}\|^2 + \|\mathbf{U}_n^H\mathbf{y}\|^2$ (2.10)

The minimum of $\left\|\mathbf{U}^H\mathbf{e}\right\|^2$ with respect to $\mathbf{h}$ is

$$\min_{\mathbf{h}}\left\|\mathbf{U}^H\mathbf{e}\right\|^2 = \left\|\mathbf{U}_n{}^H\mathbf{y}\right\|^2 \tag{2.11}$$

which occurs when

$$\mathbf{U}_s^H\mathbf{y} = \mathbf{S}_s\mathbf{V}_s^H\mathbf{h} \tag{2.12}$$

By transforming (2.12),

$$\mathbf{V}^H\mathbf{h} = \begin{bmatrix}\mathbf{S}_s^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U}_s^H\\ \mathbf{U}_n^H\end{bmatrix}\mathbf{y} = \mathbf{S}^{-1}\mathbf{U}^H\mathbf{y} \tag{2.13}$$

Therefore, the rank-reduced solution is written

$$\mathbf{h} = \mathbf{V}_s\mathbf{S}_s{}^{-1}\mathbf{U}_s{}^H\mathbf{y} = \hat{\mathbf{X}}^+\mathbf{y} \tag{2.14}$$

where $\hat{\mathbf{X}}^+ = \mathbf{V}_s\mathbf{S}_s{}^{-1}\mathbf{U}_s{}^H$ is the rank-reduced pseudoinverse of $\mathbf{X}$.

On the other hand, if $\mathbf{X}^H\mathbf{X}$ is non-singular we may directly find the LSE solution

$$\mathbf{h} = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{y} \tag{2.15}$$

The resulting minimum least-squares error is also given by Eq. (2.15)

$$E_{lse} = \mathbf{y}^H\mathbf{y} - \mathbf{y}^H\mathbf{X}\mathbf{h} \tag{2.16}$$

The normalized least-squares error is obtained by

$$E_{nlse} = \frac{\mathbf{y}^H\mathbf{y} - \mathbf{y}^H\mathbf{X}\mathbf{h}}{\mathbf{y}^H\mathbf{y}} \tag{2.17}$$

2.5 Simple examples of deconvolution

2.5.1 Input signal and output data with varied signal-to-noise ratio

Suppose that we have a linear time-invariant system whose characteristics are unknown.

By probing the system with an input sequence **x** and observing the output sequence **y**, we can find the unknown system by the deconvolution method.

In this simple example we model the unknown system as an all-zero (no poles) system. The problem then reduces to estimating the parameters of the model.

In figure 2.3, both the unknown system and the model are excited by the same input sequence **x**. Let **y** be the observed output of the unknown system and $\hat{\mathbf{y}}$ be the output of the model.

The difference between the output **y** of the unknown system, and the output $\hat{\mathbf{y}}$ of the model, is defined as the error sequence **e.** The parameters **h** of the unknown system can be obtained by minimizing the sum of the squared-error in the method of least squares. In our simulation, the coefficients **h** of FIR system model are set to 20 random numbers.

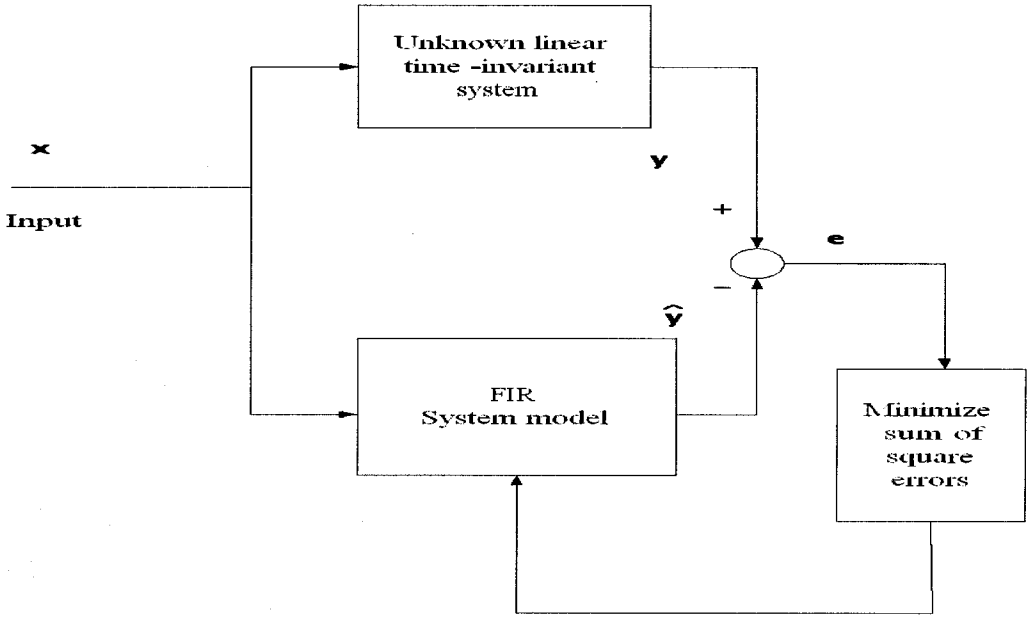


Figure 2.3: System identification based on FIR system

We use as the transmitted signal **x** for our simulation which has 175 points. A sinusoid pulse sequence having a 200 kHz carrier consists of three pulses, and each pulse has 25 points in figure 2.4.

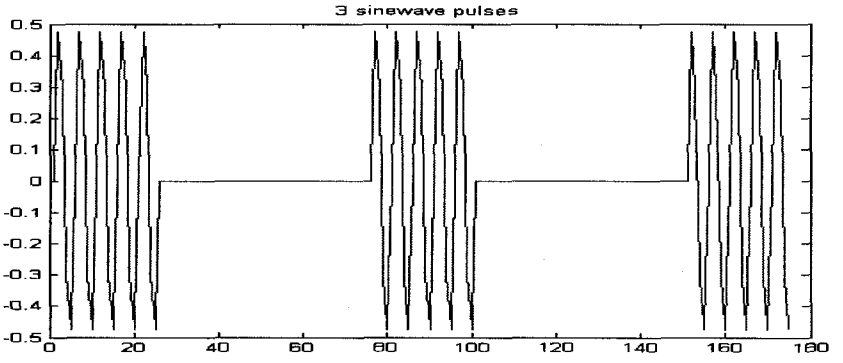


Figure 2.4: Input sequences

The least-squares optimization problem can now be presented in matrix form.

We put the input sequence into the FIR filter to get the output of the model, $\hat{\mathbf{y}}(\mathbf{n})$.

$$\hat{\mathbf{y}}(\mathbf{n}) = \sum_{k=0}^{P-1} \mathbf{h}(k)\mathbf{x}(n-k) \qquad \text{n} = 1,2,\ldots,M \tag{2.18}$$

where $M > P$. In our simulation, the known transmitted signal $\mathbf{X}$ is an $M \times P$ matrix, and coefficients of the unknown system $\mathbf{h}$ is an $P \times 1$ vector.

We define the estimated output data vector $\hat{\mathbf{y}}$ as

$$\hat{\mathbf{y}} = \mathbf{Xh} \tag{2.19.a}$$

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(P) \\ \vdots \\ \vdots \\ y(M) \end{bmatrix} = \begin{bmatrix} x(1) & 0 & \ldots & 0 \\ x(2) & x(1) & 0.. & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x(P) & x(P-1) & x(P-2)\ldots & x(1) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x(M) & x(M-1) & \vdots & x(M-P+1) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ \vdots \\ h(P-2) \\ h(P-1) \end{bmatrix} \tag{2.19.b}$$

In the real experiment, the noise and delay component are observed in the response signals from the unknown sea bottom. In order to be more representative of the true problem, a white Gaussian noise component is added after the convolution. We obtain

the noisy convolution

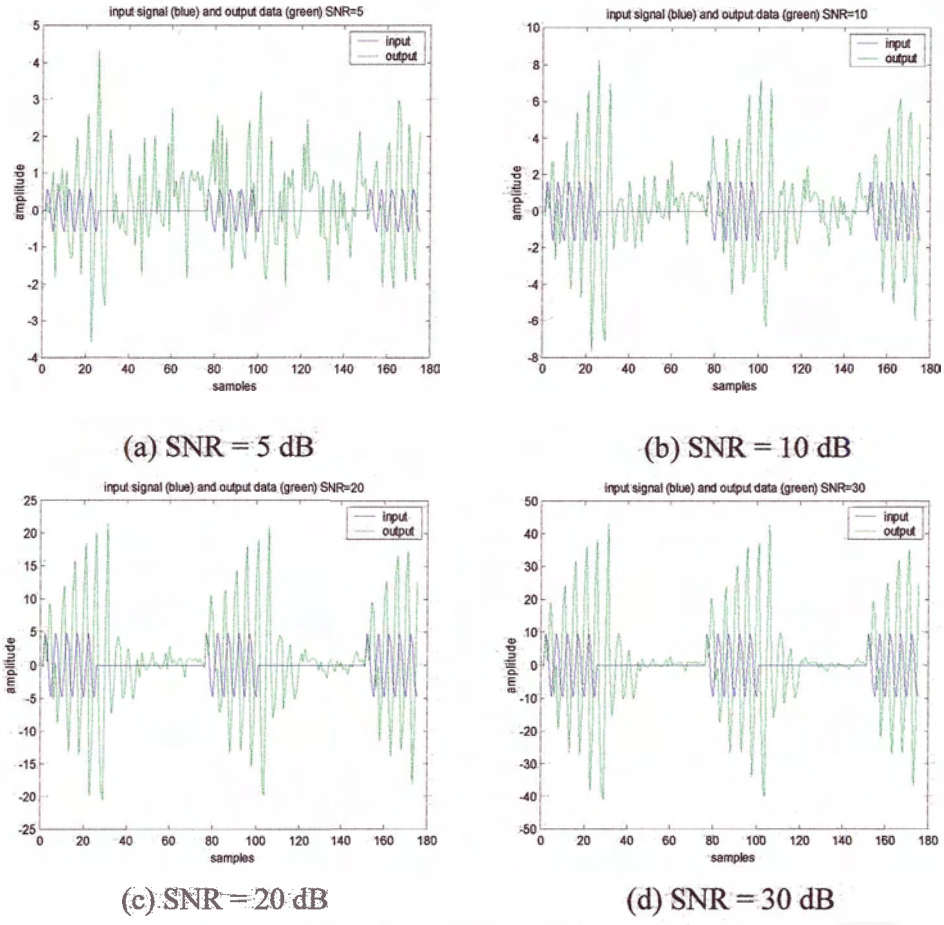
$$\hat{\mathbf{y}}_{\mathbf{noise}} = \hat{\mathbf{y}} + \mathbf{w} \tag{2.20}$$

where $\mathbf{w}$ is a white Gaussian noise.

In our simulation, we discuss the effect of decreasing or increasing SNR on the normalized least-square error (NLSE).

The error in estimating the coefficients by deconvolving $\hat{\mathbf{y}}_{\mathbf{noise}}$ varies with SNR also.

Figure 2.5 shows the transmitted signal sequence $\mathbf{x}$ and the noisy output data obtained by convolving the input sequence with P = 20 random values of an FIR filter and adding to that the white Gaussian random sequence.

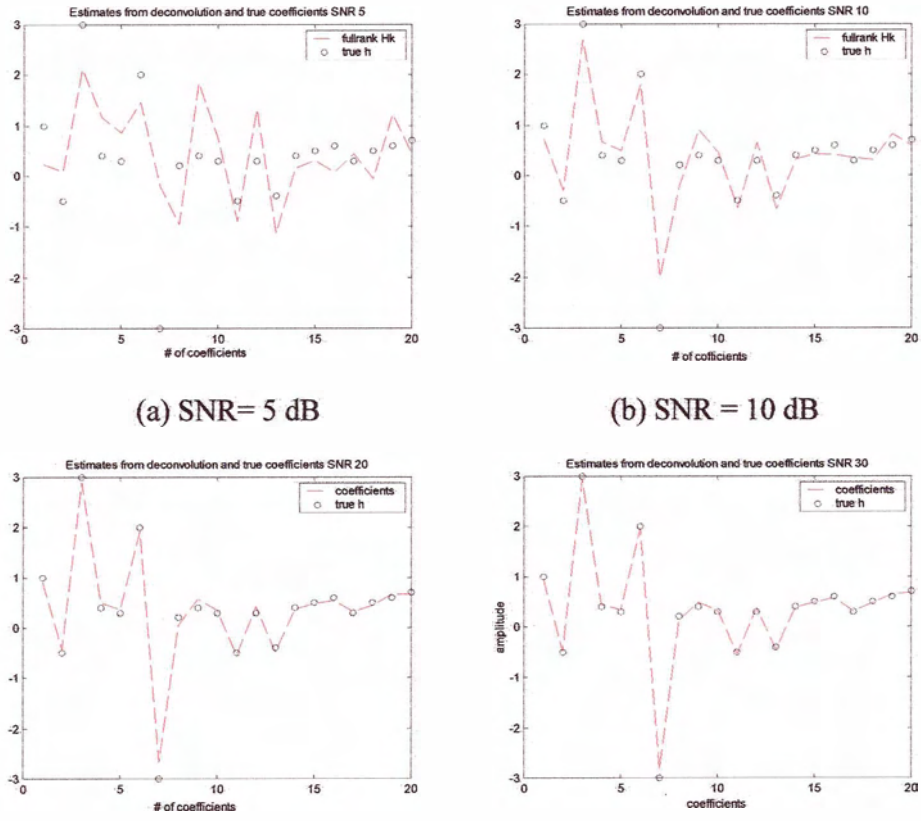


(a) SNR = 5 dB (b) SNR = 10 dB

(c) SNR = 20 dB (d) SNR = 30 dB

Figure 2.5: the observed output $\mathbf{y}$ and input sequence $\mathbf{x}$

When we deconvolve the transmitted signal from the received data in figure 2.5.a, the estimated coefficients in figure 2.6.a are in error. However, the results of deconvolution in figure 2.6.d are close to the original coefficients in figure 2.6.d. Consequently, the estimated coefficients approach the true coefficients as SNR of the received data increases.

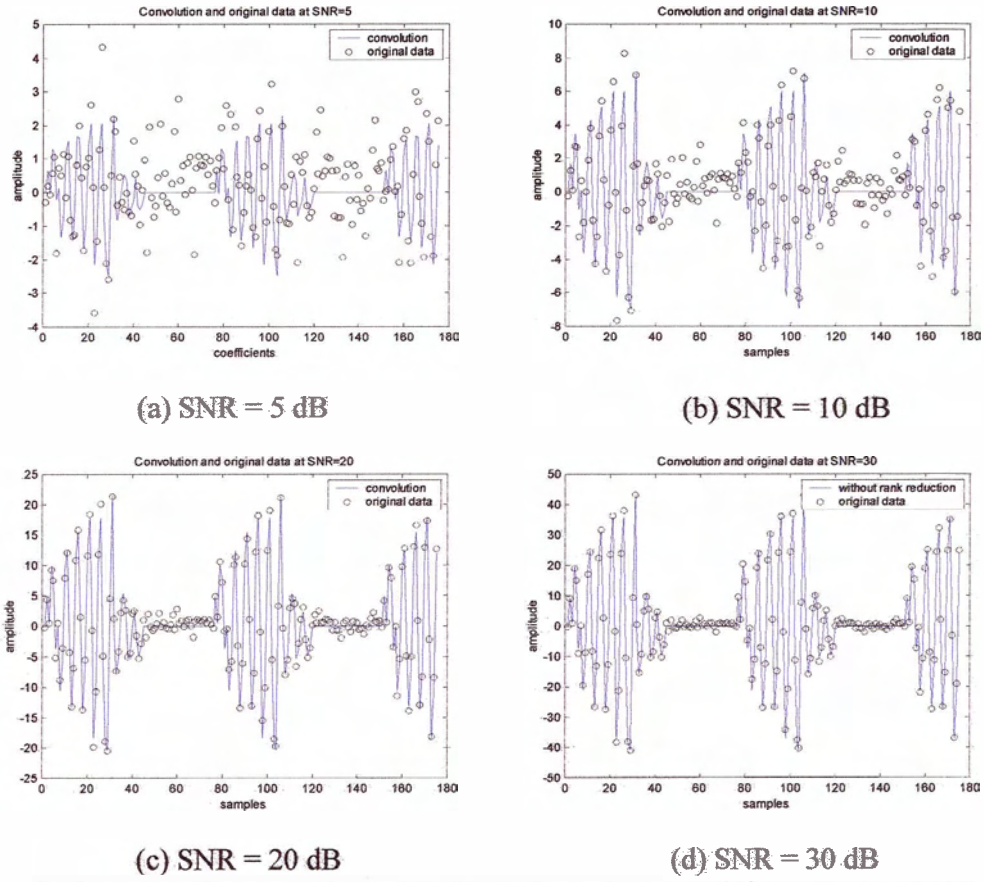


(a) SNR= 5 dB (b) SNR = 10 dB

(c) SNR = 20 dB (d) SNR = 30 dB

Figure 2.6: Estimated coefficients from deconvolution and true coefficients

The following figures show the comparison of the convolution of the input sequence **x** with the estimated coefficients **h** from deconvolution with respect to SNR and the noiseless received data.



(a) SNR = 5 dB (b) SNR = 10 dB

(c) SNR = 20 dB (d) SNR = 30 dB

Figure 2.7: Comparison with the original output **y** and convolution $\hat{\mathbf{y}}(\mathbf{n})$

From Table 2.1, it is obvious that normalized least-square error (NLSE) is decreasing or increasing with SNR. The NLSE is around 40% when SNR is 5 dB, but the error is only 0.35% at SNR 30dB.

Table 2.1: NLSE vs. SNR and deviation

Rank	*SNR*(dB)	*NLSE* (%)	Deviation of *NLSE*
full	5.96	39.37	0.3937
	13.14	13.09	0.1309
	19.16	3.68	0.0368
	29.61	0.34	0.0034

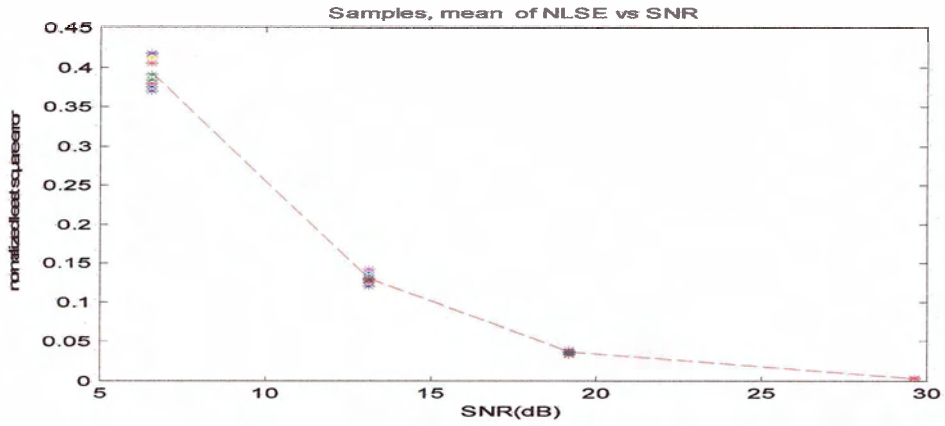


Figure 2.8: Normalized least-square error (*NLSE*) vs. *SNR*, the stars,"*" indicate the random variable, whereas the dotted line shows the mean of *NLSE* at each *SNR*

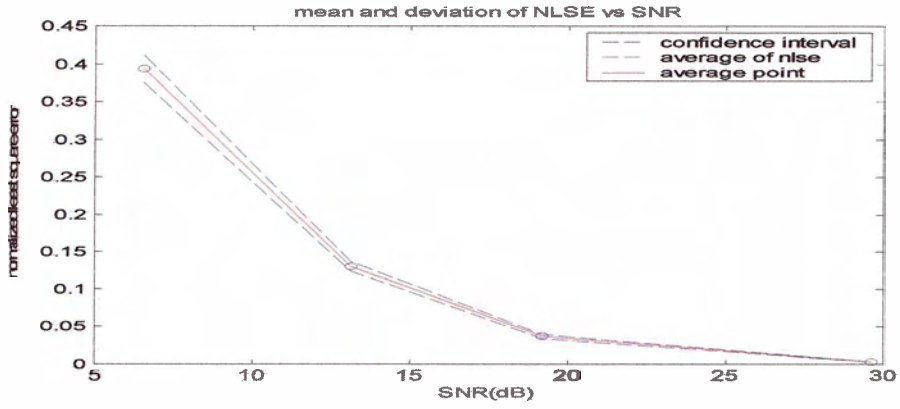


Figure 2.9: Mean and deviation of NLSE vs SNR, the blue dashed line shows the confidence interval of *NLSE* at each *SNR*

From figure 2.9, the deviation is also decreasing as long as SNR is increasing.

2.5.2 Simulation for enhancing SNR by using principal component analysis

In section 2.3 we introduced the fact that only the largest few eigenvalues contain most of energy in covariance matrix of observed data. Therefore, we can use the first few eigenimages and discard the others associated with noisy component in $\hat{\mathbf{y}}_{\mathbf{noise}}$. This

constitutes rank reduction of the data before processing further.

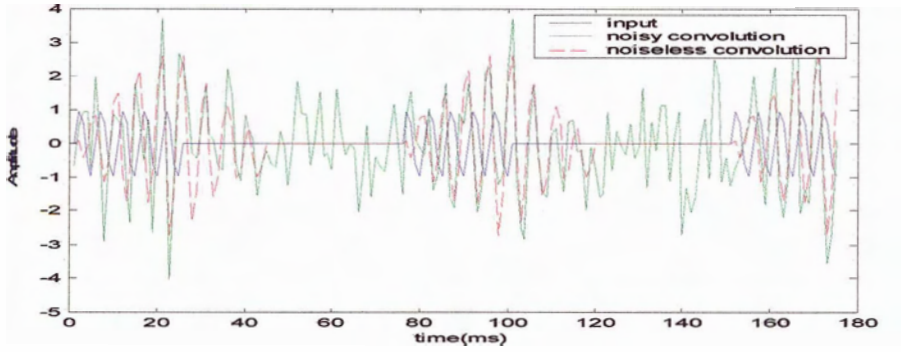


Figure 2.10: Input sequence $\mathbf{x}$, noisy convolution $\hat{\mathbf{y}}_{noise}$, and noiseless convolution $\hat{\mathbf{y}}$

In order to be more representative of the true problem, a white Gaussian noise component is added after the convolution $\hat{\mathbf{y}}$ (red dashed line). We obtain the noisy convolution $\hat{\mathbf{y}}_{noise}$ (green line) by adding a white Gaussian noise $\mathbf{w}$.

In order to reduce the rank on the vectors within $\hat{\mathbf{y}}_{noise}$, we need to partition the signal subspace and noise subspace by using eigenanalysis or the SVD method on the covariance matrix.

For reducing the rank of each data vector, we first remove its mean in order to get its deviation, and then reduce the rank of the deviation vector, and then add the mean back in.

In this simulation a covariance matrix having dimension 175×175 is produced and the eigenvalues and eigenvectors are calculated by the eigenanalysis method.

As we discussed in section 2.2, eigenvectors are a set of orthogonal basis vectors spanning the covariance matrix which can be used in conjunction with the eigenvalues to account for the energy in the covariance matrix. From figure 2.11 we have determined that with around 3 out of 175 eigenvectors we can typically account for 95% of the covariance.

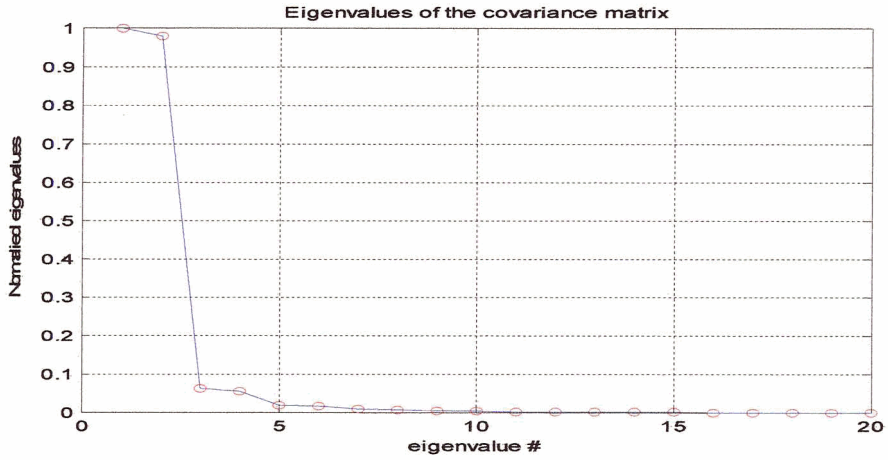


Figure 2.11: Eigenvalues of the covariance matrix

PCA provides us with a set of reduced feature vectors which contain most of the covariance energy.

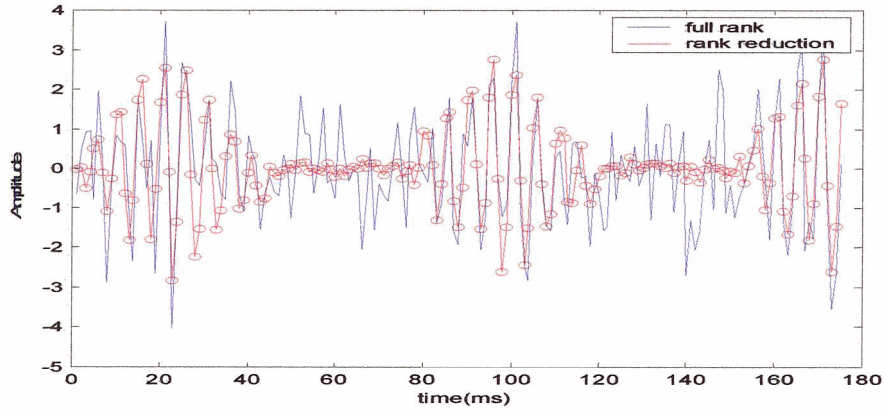


Figure 2.12: Output with Gaussian noise and rank reduced output

From figure 2.12 we show the noisy output $\hat{\mathbf{y}}_{\text{noise}}$ (blue line) and the rank-reduced output $\hat{\mathbf{y}}_{\text{rank}}$ (red line). The display shows some of the noise components in $\hat{\mathbf{y}}_{\text{noise}}$ have been suppressed. The following figure shows the difference between the rank-reduced output $\hat{\mathbf{y}}_{\text{rank}}$ and the noiseless filter output, $\hat{\mathbf{y}}$.

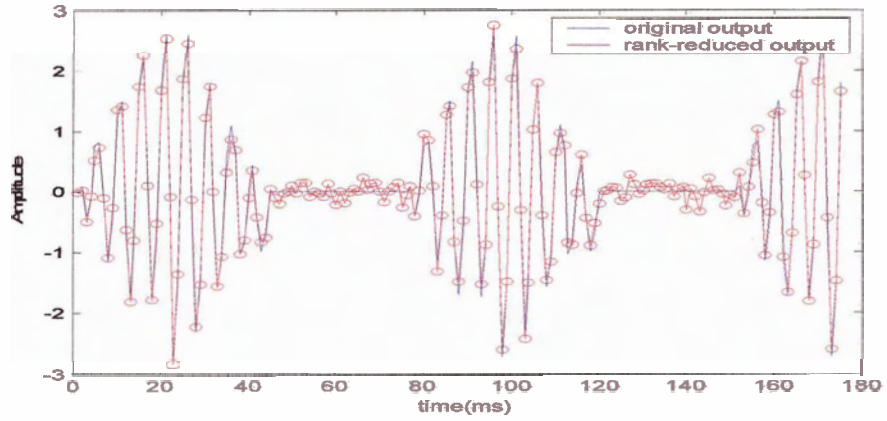


Figure 2.13: Noiseless output and rank reduced output

The normalized mean-square error of $\hat{\mathbf{y}}_{\text{rank}}$ and $\hat{\mathbf{y}}$ is less than 2%.

If a large number of dimensions are used then those with less energy will contain mainly noise, and will actually degrade system classification performance.

Consequently, instead of increasing power we can achieve enhancement of SNR by suppressing the noise signal with PCA method.

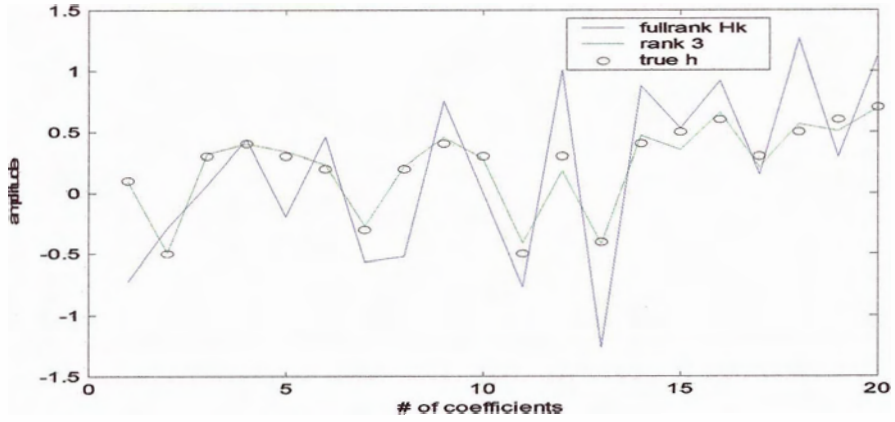


Figure 2.14: Coefficients of the unknown system and original FIR filter

From the *least-square estimation* (LSE) method in section 2.3, we can obtain the coefficients of an unknown system. Since our interest is in how we can get coefficients very close to the originals, we might apply rank-reduction prior to deconvolution. In this

simulation we have randomly chosen 20 original coefficients (circle "o") of FIR filter. Both full rank and rank reduced coefficients are obtained by

$$\mathbf{h} = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\hat{\mathbf{y}}_{noise} \tag{2.21.a}$$

$$\mathbf{h}_{rank} = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\hat{\mathbf{y}}_{rank} \tag{2.21.b}$$

where $\mathbf{X}$ is the input signal, $\hat{\mathbf{y}}_{noise}$ is noisy output and $\hat{\mathbf{y}}_{rank}$ is rank-reduced noisy output.

After the rank of the noisy output $\hat{\mathbf{y}}_{noise}$ is reduced, the coefficients $\mathbf{h}_{rank}$ from Eq. (2.21.b) are close to the original coefficients in figure 2.14.

The *Mean-Square Error (MSE)* between the original coefficients and the rank reduced coefficients $\mathbf{h}_{rank}$ is less than 5%, but *MSE* between the original coefficients and $\mathbf{h}$ from our simulation results before rank reduction is more than 20 %.

Consequently, reducing the rank of the data gives good efficiency for finding coefficients close to the original, and from these simulation results, principal component analysis (PCA) can be applied prior to deconvolution for sea bottom classification.

2.6 Achieving coherent deconvolution

2.6.1 Alignment of the reflection

Because the transmitted wave is identical from pulse to pulse, it can be recorded and to some extent coherently deconvolved from the return waveform. For coherent deconvolution the start time of the reflections must be well aligned so that the phase of the transmit pulse is identical for all responses within a class. Such coherent alignment is only practical or useful when the distinctive features among seabottom classes are

large with respect to the wavelength of the transmitted waveform, while the variations in those features within classes are smaller than the wavelength. Thus a flat sea bottom upper layer characterized by a distinctive thickness greater than a wavelength and having upper and lower boundary variations much less than a wavelength would be ideal for coherent processing. See Appendix A for further explanation of time alignment considerations.

Most of the variations in start times are associated with variations in distance to targets and need to be minimized for coherent processing, assuming distance variation within class is not a distinguishing feature. Less than perfect minimization leads to effective randomness in the phase of the transmitted signal. That randomness in turn produces at least a partially incoherent case, somewhere between total coherence and total incoherence. The minimized variations might be obtained by finding a best phase shift factor to apply to each vector within a class to make it as close as possible to a vector that is typical of the class but chosen arbitrarily to be the reference. The correction factors are found by crosscorrelating the each vector within a class with the reference.

2.6.2 Complex deconvolution

The analytic transmitted signal vector is defined as

$$\mathbf{x}_a = \mathbf{x} + j\tilde{\mathbf{x}} \tag{2.22}$$

where $\mathbf{x}$ is the transmitted signal vector and $\tilde{\mathbf{x}}$ is the Hilbert transform of $\mathbf{x}$. Since the two parts in Eq. (2.22) are at 90° relative phase shift at each frequency, they contain all the coherency information in a phase-shifted or delayed reflection.

In order to obtain a complex deconvolution, each of the real and imaginary parts in Eq. (2.22) is separately deconvolved from the class' aligned reflections. We use the LSE

algorithm in order to obtain estimated coefficients, and the real and imaginary coefficients are defined as

$$\mathbf{h}_{real} = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{y}_{align} \tag{2.23.a}$$

$$\mathbf{h}_{image} = (\tilde{\mathbf{X}}^H\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^H\mathbf{y}_{align} \tag{2.23.b}$$

where $\mathbf{X}$ is the known real transmitted signal, $\tilde{\mathbf{X}}$ is the Hilbert transform of $\mathbf{X}$ and $\mathbf{y}_{align}$ is the aligned reflected data.

From the deconvolution results of the real and imaginary parts in Eq. (2.23), the complex estimated coefficients are defined as

$$\mathbf{h}_{complex} = \mathbf{h}_{real} + j\mathbf{h}_{image} \tag{2.24}$$

2.7 Summary

When we know the transmitted signal and have the received data from the unknown but linear system, the characteristic of the system can be obtained by using the deconvolution method. In our project we used the least-square estimation method for deconvolution.

In the real environment there are not only the signal components which we want to obtain, but also the random Gaussian noise component. When SNR is increasing, the mean-square error (MSE) between the desired signal and the estimated signal decreases. However, we cannot obtain large SNR by arbitrarily increasing power. Also, much of the "noise" is simply random variation within classes that is not class distinctive. Instead we can use PCA from covariance matrix analysis to find the high SNR signal components of the received data. We take only few dimensions of signal component

instead of including the other dimensions which contain mostly noise components. From our simulation results we have found that the estimated coefficients from the deconvolution method are almost the same as the original coefficients when we use only 3 dimensions for signal.

Consequently, the deconvolution method can be useful to find the unknown system function which we are really interested in.

In the real environment there are variations in start times of reflections associated with variations in distance to targets. The aligned reflected data might be obtained by finding a best phase shift factor which can be found by finding the shift index of maximum crosscorrelation of each vector within a class with the reference vector for that class. Complex deconvolution utilizes the coherent information in a phase-shifted or delayed reflection and obtains improved coefficient estimation.

Chapter 3.

Methods for Measuring Pairwise Class Distance

In the last chapter we studied the deconvolution method in order to find the unknown system model of a bottom reflector and presented some results for simulated signals. For distinguishing classes of reflectors, it seems reasonable to deconvolve the transmitted signal from any reflection and compare deconvolved results for classification with those of the raw reflections. From the deconvolution result we can more directly obtain information about the reflector in which we are primarily interested for each class.

In this chapter we will discuss measurement of distance between classes, because direct minimization of the error probability to determine the optimum signal is often difficult to carry out, and even if it can be found, the expression may be too complicated for analytical or numerical minimization. The distance measure is a simpler and practical alternative to the error probability.

In the first part we discuss two common distance measures for class separation: divergence and Bhattacharyya.

In the second part we pay particular attention to the use of knowledge of the transmitted carrier phase and determine its influence on the resulting distance between

classes.

3.1 The Bayes error

When $\mathbf{y}$ is an observation vector, we want to determine which class $\mathbf{y}$ belongs to. This determination degenerates to determining pairwise comparisons, whether $\mathbf{y}$ belongs to ω_1 or ω_2 [11].

We utilize the criterion that if the probability of ω_1 given $\mathbf{y}$ is larger than the probability of ω_2, then $\mathbf{y}$ is classified to ω_1 and vice versa.

We can use *Bayes theorem* to obtain the decision rule.

$$q_1(\mathbf{y}) \underset{\omega_1}{\overset{\omega_2}{\lessgtr}} q_2(\mathbf{y}) \tag{3.1}$$

where $q_i(\mathbf{y})$ is the posteriori probability of ω_i

$$q_i(\mathbf{y}) = \frac{P_i p_i(\mathbf{y})}{p(\mathbf{y})} \tag{3.2}$$

$p(\mathbf{y})$ is the mixture density function, P_i is a priori probability and $p_i(\mathbf{y}) = p(\mathbf{y}/\omega_i)$ is conditional density.

The decision rule of Eq. (3.1) can be expressed as

$$P_1 p_1(\mathbf{y}) \underset{\omega_1}{\overset{\omega_2}{\lessgtr}} P_2 p_2(\mathbf{y}) \tag{3.3}$$

or

$$l(\mathbf{y}) = \frac{p_1(\mathbf{y})}{p_2(\mathbf{y})} \underset{\omega_1}{\overset{\omega_2}{\lessgtr}} \frac{P_2}{P_1} \tag{3.4}$$

In general, the decision rule of Eq. (3.4) does not lead to perfect classification. In order to evaluate the performance of a decision rule, we should calculate the probability of error, that is, the probability that a sample is assigned to the wrong class.

The conditional error given $\mathbf{y}$, $r(\mathbf{y})$, due to the decision rule of Eq. (3.1) is either $q_1(\mathbf{y})$ or $q_2(\mathbf{y})$ whichever smaller. That is

$$r(\mathbf{y}) = \min[q_1(\mathbf{y}), q_2(\mathbf{y})] \tag{3.5}$$

The Bayes error, total error, is computed by $E\{r(\mathbf{y})\}$.

$$\begin{aligned} \varepsilon = E\{r(\mathbf{y})\} &= \int r(\mathbf{y}) p(\mathbf{y}) d\mathbf{y} \\ &= \int \min[P_1 p_1(\mathbf{y}), P_2 p_2(\mathbf{y})] d\mathbf{y} \end{aligned} \tag{3.6}$$

We are going to use Eq. (3.6) in order to find the upper bounds.

3.2 Upper bounds on the Bayes error

We know that the calculation of the error probability is, in general, a difficult task. Instead of finding the error probability, we seek either an approximate expression for the error probability, or an upper bound on the error probability.

3.2.1 The Chernoff distance

An upper bound of the integral may be obtained by making use of the fact that

$$\min[a,b] \le a^s b^{1-s} \qquad 0 \le s \le 1 \tag{3.7}$$

for $a, b \ge 0$ [11]. This statement can be proved as follows. If $a < b$, the left side of Eq. (3.7) is a, and the right side can be rewritten as $a \times (b/a)^{1-s}$. Since $b/a > 1$ and $1-s \ge 0$ for $0 < s < 1$, the right side becomes larger than the left side. Likewise, if $a > b$, the left side of Eq. (3.7) is b, and the right side is rewritten as $b \times (a/b)^s$, which is larger than b because $(a/b) > 1$ and $s \ge 0$. Using the inequality of Eq. (3.7), the error can be bounded by

$$\varepsilon_u = P_1^s P_2^{1-s} \int p_1^s(\mathbf{y}) p_2^{1-s}(\mathbf{y}) d\mathbf{y} \tag{3.8}$$

where ε_u indicates an upper bound of error; it is called the *Chernoff bound.*

Suppose that we have two pattern classes characterized by two *n*-variant normal probability densities function as a Gaussian distribution [14] :

$$p_1(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}_1|^{1/2}} \exp[-\frac{1}{2}(\mathbf{y}-\mathbf{m}_1)^H \mathbf{C}_1^{-1}(\mathbf{y}-\mathbf{m}_1)] \tag{3.9}$$

and

$$p_2(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}_2|^{1/2}} \exp[-\frac{1}{2}(\mathbf{y}-\mathbf{m}_2)^H \mathbf{C}_2^{-1}(\mathbf{y}-\mathbf{m}_2)] \tag{3.10}$$

where $\mathbf{m}_1, \mathbf{m}_2$ are mean vectors and $\mathbf{C}_1, \mathbf{C}_2$ are covariance matrix.

We can find the optimum *s* by minimizing Eq. (3.11). That is

$$\int p_1^s(\mathbf{y}) p_2^{1-s}(\mathbf{y}) d\mathbf{y} = e^{-\mu(s)} \tag{3.11}$$

where

$$\mu(s) = \frac{s(1-s)}{2}(\mathbf{m}_2-\mathbf{m}_1)^H [s\mathbf{C}_1 + (1-s)\mathbf{C}_2]^{-1}(\mathbf{m}_2-\mathbf{m}_1) + \frac{1}{2}\ln\frac{|s\mathbf{C}_1 + (1-s)\mathbf{C}_2|}{|\mathbf{C}_1|^s |\mathbf{C}_2|^{1-s}} \tag{3.12}$$

This expression of $\mu(s)$ is the *Chernoff distance.*

3.2.2 Bhattacharyya distance

When *s* of Eq. (3.12) is selected as *1/2,* then, the upper bound is [11]

$$\varepsilon_u = \sqrt{P_1 P_2} \int \sqrt{p_1(\mathbf{y}) p_2(\mathbf{y})} d\mathbf{y} = \sqrt{P_1 P_2}\, e^{-\mu(1/2)} \tag{3.13}$$

where

$$\mu(1/2) = \frac{1}{8}(\mathbf{m}_2-\mathbf{m}_1)^H [\frac{\mathbf{C}_1+\mathbf{C}_2}{2}]^{-1}(\mathbf{m}_2-\mathbf{m}_1) + \frac{1}{2}\ln\frac{\left|\frac{\mathbf{C}_1+\mathbf{C}_2}{2}\right|}{\sqrt{|\mathbf{C}_1||\mathbf{C}_2|}} \tag{3.14}$$

The term $\mu(1/2)$ is called the *Bhattacharyya Distance*, a commonly used measure of the separability of two distributions.

We can see the first term gives the class separability due to the mean-difference, while the second term gives the class separability due to the covariance difference [11, 15, 21].

3.3 Divergence distance

Divergence is another measure of distance between two classes, and it was first introduced by Jeffreys [9]. In this section we discuss the concept of divergence and its application.

When the probability of occurrence of pattern **y**, given that it belongs to class ω_1, be $p_1(\mathbf{y})$ and the probability of occurrence of pattern **y**, given that it belongs to class ω_2, be $p_2(\mathbf{y})$, the discriminating information for class ω_1 versus ω_2 may be measured by the logarithm of the likelihood ratio and by the average discriminating information for class ω_1 is given by [10]

$$I(1,2) = \int_{\mathbf{y}} p_1(\mathbf{y}) \ln \frac{p_1(\mathbf{y})}{p_2(\mathbf{y})} d\mathbf{y} \tag{3.15}$$

The average discriminating information for class ω_2 is also given by

$$I(2,1) = \int_{\mathbf{y}} p_2(\mathbf{y}) \ln \frac{p_2(\mathbf{y})}{p_1(\mathbf{y})} d\mathbf{y} \tag{3.16}$$

The total average information for discriminating class ω_i from class ω_j is referred to as *divergence*, which is given by

$$\begin{aligned} d_{12} &= I(1,2) + I(2,1) \\ &= \int_{\mathbf{y}} [p_1(\mathbf{y}) - p_2(\mathbf{y})] \ln \frac{p_1(\mathbf{y})}{p_2(\mathbf{y})} d\mathbf{y} \end{aligned} \tag{3.17}$$

The average information for discrimination between classes is

$$I(1,2) = \int_{\mathbf{y}} p_1(\mathbf{y}_1, \mathbf{y}_2 \ldots, \mathbf{y}_n) \ln \frac{p_1(\mathbf{y}_1, \mathbf{y}_2 \ldots, \mathbf{y}_n)}{p_2(\mathbf{y}_1, \mathbf{y}_2 \ldots, \mathbf{y}_n)} d\mathbf{y}_1 d\mathbf{y}_2 \ldots, d\mathbf{y}_n$$
$$= \frac{1}{2} \ln \frac{|\mathbf{C}_2|}{|\mathbf{C}_1|} + \frac{1}{2} tr[\mathbf{C}_1(\mathbf{C}_2^{-1} - \mathbf{C}_1^{-1})] + \frac{1}{2} tr[\mathbf{C}_1(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^H] \quad (3.18)$$

Hence, we can obtain the divergence for two classes from Eq. (3.18)

$$d_{12} = \frac{1}{2} tr[(\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1})(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^H] + \frac{1}{2} tr[(\mathbf{C}_1 - \mathbf{C}_2)(\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1})] \quad (3.19)$$

Therefore, the divergence is obtained by using the expected values of the log-likelihood ratio for ω_i and ω_j [10].

3.4 Relation of divergence and Bhattacharyya distance

The first terms in both Eq. (3.14) and (3.19) are measures of distance between the two class means, while the second terms are measures of difference in covariance matrices. If the two classes have identical covariances, the first terms are both proportional to each other and to the Mahalobis distance

$$M_a = (\mathbf{m}_1 - \mathbf{m}_2)^H \mathbf{C}^{-1} (\mathbf{m}_1 - \mathbf{m}_2) \quad (3.20)$$

Whether or not the covariances of classes are equal, we can establish the inequality

$$d_{ij} \geq 8B_{ij} \quad (3.21)$$

where d_{ij} is divergence and B_{ij} is Bhattacharyya distance between ω_i and ω_j.

The proof of Eq. (3.21) is found in Kailath [12].

In order to compare Bhattacharyya and divergence distances, we use the ratio of the latter to the former.

When $B_{12}, B_{13}, ..., B_{1N}$ are the results of Bhattacharyya distance, they can be divided by B_{12}, and we can obtain the normalized Bhattacharyya distance as $1, \frac{B_{13}}{B_{12}}, \frac{B_{14}}{B_{12}} ..., \frac{B_{1N}}{B_{12}}$. With the same method for normalization, we can get the normalized divergence, $1, \frac{d_{13}}{d_{12}}, \frac{d_{14}}{d_{12}} ..., \frac{d_{1N}}{d_{12}}$. In order to find out which measure is better, we can compare the normalized Bhattacharyya distance and the normalized divergence,

$$\frac{B_{1N}}{B_{12}} \lessgtr \frac{d_{1N}}{d_{12}} \tag{3.22}$$

3.5 Class distance and gain with raw return vector data

From Kirlin [33], when the raw data is the convolution of the transmitted signal with the reflector's impulse response, the raw data vector in the frequency domain given a class ω_i is $\mathbf{Y}_i$, where the k^{th} element of $\mathbf{Y}_i$ is $Y(f_k) = X(f_k)H(f_k), k = 1,2,...,q$.

We can obtain $\mathbf{Y}_i$ by multiplying the transmitted waveform $X(f)$ by the transfer function $H_i(f)$. We consider the transmitted waveform $X(f)$ as coherent $X(f)$ and incoherent $X(f)$ with random phase over [-π,π), which has mean $\bar{X}(f)$ and variance $\sigma_x^2(f)$, and $H_i(f)$ has mean $\bar{H}_i(f)$ and variance $\sigma_{Hi}^2(f)$

3.5.1 Distance between means of classes

The means appear in a quadratic function of their difference, $\boldsymbol{\delta}_{ij} = \bar{\mathbf{Y}}_i - \bar{\mathbf{Y}}_j$. The k^{th} element of $\boldsymbol{\delta}_{ij}$ is

$$\boldsymbol{\delta}_{ij}(k) = E\{X(f_k)\}\left[\bar{H}_i(f_k) - \bar{H}_j(f_k)\right]$$

$$= \begin{cases} X(f_k)\left[\bar{H}_i(f_k) - \bar{H}_j(f_k)\right], & X \text{ coherent} \quad (3.23.a) \\ \bar{X}(f_k)\left[\bar{H}_i(f_k) - \bar{H}_j(f_k)\right], & X \text{ incoherent} \quad (3.23.b) \end{cases}$$

If the phase of the carrier is uniformly random over (-π, π), then $\bar{X}(f_k) = 0$. Therefore, Eq. (3.23.b) becomes zero, and the first terms of the Bhattacharyya distance Eq. (3.14) and divergence Eq. (3.19) become zero. The only course is to use only reflection envelopes.

3.5.2 Covariance of vector form $\mathbf{Y}_i$

Given the independence of the elements of $\mathbf{Y}_i$, and independence of $X(f_k)$ and $H_i(f_k)$, we can write the k^{th} diagonal element of the diagonal covariance matrix $\boldsymbol{\Sigma}_i$ of $\mathbf{Y}_i$ as

$$\boldsymbol{\Sigma}_i(k) = E\left\{\left|X(f_k)H_i(f_k) - E\{(X(f_k)H_i(f_k))\}\right|^2\right\}$$

$$= \begin{cases} |X(f_k)|^2 \sigma^2_{Hi}(f_k), & X \text{ coherent} \quad (3.24.a) \\ E\left\{\left|X(f_k)H_i(f_k) - E\{X(f_k)\}E\{H_i(f_k)\}\right|^2\right\}, & X \text{ incoherent} \quad (3.24.b) \end{cases}$$

The second term in Eq. (3.24.b) will be zero because of $\bar{X}(f_k) = 0$ when the phase of carrier is uniform random phase. Therefore, Eq. (3.24.b) will be

$$\begin{aligned} \boldsymbol{\Sigma}_i(k) &= E\left\{\left|X(f_k)\right|^2\right\} E\left\{\left|H_i(f_k)\right|^2\right\} \\ &= \left\{\sigma^2_{X(fk)} + \left|\bar{X}(f_k)\right|^2\right\} E\left\{\left|H_i(f_k)\right|^2\right\} \quad (3.25) \\ &= \sigma^2_{X(fk)} E\left\{\left|H_i(f_k)\right|^2\right\}, X \text{ incoherent} \end{aligned}$$

Because of $\bar{X}(f_k) = 0$ when the phase of carrier is uniform random phase, the covariance matrix has diagonal elements $\boldsymbol{\Sigma}_i(k) = \sigma^2_{X(fk)} E\left\{\left|H_i(f_k)\right|^2\right\}$.

3.6 Class distance by divergence and Bhattacharyya distance

In the last section we have found the mean and covariance of the raw returned data vector in coherent and incoherent forms. We will apply these means and covariances to obtain both coherent and non-coherent divergence and Bhattacharyya distance.

We consider divergence and Bhattacharyya distance in the frequency domain instead of the time domain because the transformation to the frequency domain gives diagonal covariance matrices and simplifies the formulation and the two distance measures can be related to a sum over the distances at each frequency.

3.6.1 Coherent and incoherent divergence

At each frequency we can obtain the divergence Eq. (3.19) as

$$d_{ij}(k)=\frac{1}{2}\left|\boldsymbol{\delta}_{ij}(k)\right|^2\left(\boldsymbol{\Sigma}_i^{-1}(k)+\boldsymbol{\Sigma}_j^{-1}(k)\right)+\frac{1}{2}\left(\boldsymbol{\Sigma}_i^{-1}(k)\boldsymbol{\Sigma}_j(k)+\boldsymbol{\Sigma}_j^{-1}(k)\boldsymbol{\Sigma}_i(k)-2\right) \tag{3.26}$$

We can consider coherent and incoherent divergence by applying the means and covariance which we obtained in the section 3.5.

3.6.1.1 Coherent divergence

By applying Eq. (3.23.a) and (3.24.a), we find coherent divergence to be

$$d_{ij}(k)_{coherent}=\frac{1}{2}\left|\bar{H}_i(f_k)-\bar{H}_j(f_k)\right|^2\left(\frac{1}{\sigma_{Hi}^2(f_k)}+\frac{1}{\sigma_{Hj}^2(f_k)}\right)+\frac{1}{2}\left(\frac{\sigma_{Hj}^2(f_k)}{\sigma_{Hi}^2(f_k)}+\frac{\sigma_{Hi}^2(f_k)}{\sigma_{Hj}^2(f_k)}-2\right) \tag{3.27}$$

3.6.1.2 Incoherent divergence

The incoherent divergence can be obtained by applying Eq. (3.23.b) and (3.25) in Eq. (3.26):

$$\mathbf{d}_{ij}(k)_{incoherent} = \frac{1}{2}\left|\bar{X}(f_k)\left[\bar{H}_i(f_k) - \bar{H}_j(f_k)\right]\right|^2 \left(\frac{1}{\sigma^2_{X(fk)} E\left\{\left|H_i(f_k)\right|^2\right\}} + \frac{1}{\sigma^2_{X(fk)} E\left\{\left|H_j(f_k)\right|^2\right\}} \right) + \frac{1}{2}\left(\frac{\sigma^2_{X(fk)} E\left\{\left|H_j(f_k)\right|^2\right\}}{\sigma^2_{X(fk)} E\left\{\left|H_i(f_k)\right|^2\right\}} + \frac{\sigma^2_{X(fk)} E\left\{\left|H_i(f_k)\right|^2\right\}}{\sigma^2_{X(fk)} E\left\{\left|H_j(f_k)\right|^2\right\}} - 2 \right) \tag{3.28}$$

Because $\bar{X}(f_k) = 0$ if the phase of the carrier is uniformly random over [-π, π), the first term of Eq.(3.28) is zero and the second term simplifies to

$$d_{ij}(k)_{incoherent} = \frac{1}{2}\left(\frac{E\left\{\left|H_j(f_k)\right|^2\right\}}{E\left\{\left|H_i(f_k)\right|^2\right\}} + \frac{E\left\{\left|H_i(f_k)\right|^2\right\}}{E\left\{\left|H_j(f_k)\right|^2\right\}} - 2 \right) \tag{3.29}$$

3.6.2 Coherent and incoherent Bhattacharyya distance

At each frequency, we can obtain Bhattacharyya distance Eq. (3.14) as

$$\mathbf{B}_{ij}(k) = \frac{1}{8}\left|\boldsymbol{\delta}_{ij}(k)\right|^2 \left(\frac{\boldsymbol{\Sigma}_i(k) + \boldsymbol{\Sigma}_j(k)}{2} \right)^{-1} + \frac{1}{2}\ln \frac{\left| \frac{\boldsymbol{\Sigma}_i(k) + \boldsymbol{\Sigma}_j(k)}{2} \right|}{\sqrt{\left|\boldsymbol{\Sigma}_i(k)\right|\left|\boldsymbol{\Sigma}_j(k)\right|}} \tag{3.30}$$

We also consider coherent and incoherent Bhattacharyya distance at each frequency.

3.6.2.1 Coherent Bhattacharyya

By applying coherent distance of means Eq. (3.23.a) and covariance (3.24.a) in Eq. (3.30),

The coherent Bhattacharyya distance is

$$\mathbf{B}_{ij}(k)_{coherent} = \frac{1}{4}\left|\bar{H}_i(f_k) - \bar{H}_j(f_k)\right|\left(\sigma^2_{Hi}(f_k) - \sigma^2_{Hj}(f_k)\right)^{-1} + \frac{1}{2}\ln\left(\frac{\sigma^2_{Hi}(f_k) + \sigma^2_{Hj}(f_k)}{2\sigma_{Hi}(f_k)\sigma_{Hj}(f_k)}\right) \tag{3.31}$$

3.6.2.2 Incoherent Bhattacharyya

In order to get incoherent Bhattacharyya distance, we also need to apply Eq. (3.23.b) and (3.25) in Eq. (3.30), resulting in

$$\begin{aligned}\mathbf{B}_{ij}(k)_{incoherent} = &\frac{1}{4}\left|\bar{X}(f_k)\left[\bar{H}_i(f_k) - \bar{H}_j(f_k)\right]\right|^2\left(\sigma^2_{X(fk)}E\left\{\left|H_i(f_k)\right|^2\right\} + \sigma^2_{X(fk)}E\left\{\left|H_j(f_k)\right|^2\right\}\right)^{-1} \\ &+\frac{1}{2}\ln\left(\frac{\sigma^2_{X(fk)}E\left\{\left|H_i(f_k)\right|^2\right\} + \sigma^2_{X(fk)}E\left\{\left|H_j(f_k)\right|^2\right\}}{2\left(\sigma^2_{X(fk)}E\left\{\left|H_i(f_k)\right|^2\right\}\right)^{1/2}\left(\sigma^2_{X(fk)}E\left\{\left|H_j(f_k)\right|^2\right\}\right)^{1/2}}\right)\end{aligned} \tag{3.32}$$

Because $\bar{X}(f_k) = 0$ if the phase of the carrier is uniformly random over [-π, π), the first term of Eq. (3.32) is zero and the second term simplifies to

$$\mathbf{B}_{ij}(k)_{incoherent} = \frac{1}{2}\ln\left(\frac{E\left\{\left|H_i(f_k)\right|^2\right\} + E\left\{\left|H_j(f_k)\right|^2\right\}}{2\left(E\left\{\left|H_i(f_k)\right|^2\right\}\right)^{1/2}\left(E\left\{\left|H_j(f_k)\right|^2\right\}\right)^{1/2}}\right) \tag{3.33}$$

We have considered coherent and incoherent transmitted signals, $X(f)$. For the coherent case, since the measures are not a function of $X(f)$, we can treat $X(f)$ as scaling factor which does not effect either information or detection probability. With incoherency from random carrier phase, the distance decreases and probability of classification error increases.

3.6.3 Gain in distance measure due to coherent deconvolution

For the distance measures of divergence and Bhattacharyya, the convolution with the transmitted signal $X(f)$ and impulse response $H(f)$ is applied in Eq. (3.26) and (3.30). If we have ideal deconvolution the distance measures between classes will be the same as the coherent measures because $X(f)$ has disappeared from the expressions. By comparing the coherent results of Eq. (3.27) and (3.31) to the incoherent results of Eq. (3.29) and (3.33), we can compare class distance with ideal deconvolution to that with random phase uniform in [-π, π).

From the gain factors we can determine the best that deconvolution can produce. However, in fact this gain does not require deconvolution, but simply gives the gain due to coherence, in which case deconvolution is applicable.

Divergence deconvolution gain factor is obtained by the ratio of Eq. (3.27) and (3.29).

$$G_D = \frac{\sum_{k=1}^{q}\left(\left|\bar{H}_i(f_k)-\bar{H}_j(f_k)\right|^2\left(\frac{1}{\sigma_{Hi}^2(f_k)}+\frac{1}{\sigma_{Hj}^2(f_k)}\right)+\left(\frac{\sigma_{Hj}^2(f_k)}{\sigma_{Hi}^2(f_k)}+\frac{\sigma_{Hi}^2(f_k)}{\sigma_{Hj}^2(f_k)}-2\right)\right)}{\sum_{k=1}^{N}\left(\frac{E\left\{\left|H_j(f_k)\right|^2\right\}}{E\left\{\left|H_i(f_k)\right|^2\right\}}+\frac{E\left\{\left|H_i(f_k)\right|^2\right\}}{E\left\{\left|H_j(f_k)\right|^2\right\}}-2\right)} \tag{3.34}$$

Bhattacharyya distance deconvolution gain factor is obtained by the ratio of Eq. (3.31) and (3.33)

$$G_B = \frac{\sum_{k=1}^{q}\left\{\frac{1}{2}\left|\bar{H}_i(f_k)-\bar{H}_j(f_k)\right|^2\left(\sigma_{Hi}^2(f_k)+\sigma_{Hj}^2(f_k)\right)^{-1}+\ln\left(\frac{\sigma_{Hi}^2(f_k)+\sigma_{Hj}^2(f_k)}{2\sigma_{Hi}(f_k)\sigma_{Hj}(f_k)}\right)\right\}}{\sum_{k=1}^{N}\ln\left(\frac{E\left\{\left|H_i(f_k)\right|^2\right\}+E\left\{\left|H_j(f_k)\right|^2\right\}}{2\left(E\left\{\left|H_i(f_k)\right|^2\right\}\right)^{1/2}\left(E\left\{\left|H_j(f_k)\right|^2\right\}\right)^{1/2}}\right)} \tag{3.35}$$

Total distances between class pairs are found by summing the $G_D(k)$ or $G_B(k)$ over the q frequencies.

From the gain results we expect class separation to increase when coherent deconvolution can be applied to raw data vectors of sonar, seismic or radar responses. Transformation to the frequency domain gives diagonal covariance matrices and simplifies formulation, but it is not required. Note again that deconvolution per se does not affect the distance measure, but it can be applied when the transmitted signal and the initial time of responses of the class are known and coherent. Deconvolution is necessary to determine parameters of the bottom response, and those are of interest in of themselves.

3.7 Summary

Direct minimization of the error probability to determine optimum processing is often difficult to carry out. Even if it can be found, the expression may be too complicated for analytical or numerical minimization. We have discussed two common distance measures as simple alternatives to the error probability.

Because Bhattacharyya distance is derived from the upper bound on the error probability, it has a closer relation with the Bayes error. Divergence evaluates the class separability by the difference of the information of two classes. However, if the two classes have identical covariances, the first terms of Eq. (3.14) and (3.19) are both proportional to each other and to the Mahalobis distance.

In section 3.5 we used knowledge of the transmitted carrier phase and determined its influence on the resulting distance between classes. We considered the coherent and incoherent transmitted waveforms in the frequency domain in order to find the gain factors of both distance measures which can allow formulation of the comparative performance of deconvolution. When the covariance matrices at each frequency are identical for both classes and the phase is uniform over $(-\pi, \pi)$, no mean difference remains between class means and only covariance information (in the envelopes of the responses) is left to give distance.

Coherent deconvolution can be achieved if the phase of the transmit pulse is identical for all responses and the responses themselves have some coherence within their class.

Chapter 4

Results of Least-Square Estimation of Deconvolved Sea-bottom Sonar Responses

Incorporating the deconvolution algorithm studied in chapter 2, this chapter outlines some of results of the application of the research into deconvolution in a real environment.

In the first section of this chapter data representation are briefly presented. And since the reflected data are the result of convolution of the transmitted signal with the bottom reflector, the transmitted signal should be generated by the same transducer as we used in the ocean. The collection of the transmitted signal is also presented.

In the second section of this chapter the deconvolution algorithms we have earlier presented are applied to three distinct real-world data sets. Enhancements of raw data and coefficients estimated by using *Principal component analysis*(PCA) are presented. The results at each stage are discussed in context with the issues raised in chapter 2.

4.1 Implementation overview

4.1.1 Data representation

In this chapter, we use the following variables consistently

M the dimensionality of a raw vector

P the dimensionality of a coefficients vector

N the number of vectors in a data set

L the length of the transmitted signal

Individual records of raw data are stored as $M \times 1$ column vectors. The records can be combined into raw data sets by forming $M \times N$ matrices. The coefficients are obtained from deconvolving the transmitted signal with raw data. Because we consider that the raw data has resulted from convolution of the transmitted signal with the bottom reflector, the dimensionality of a coefficient vector is calculated as $P = M - L + 1$ where L is the length of the transmitted signal.

Class statistics are represented for the raw data and their coefficients estimated from deconvolution. Firstly the mean of the raw data is represented as a $M \times 1$ vector, and the covariance of raw data is represented as $M \times M$. Secondly the mean of estimated coefficients is represented as $P \times 1$, and the covariance of the estimated coefficients is also represented as $P \times P$. Since the dimension of reflected raw data is different from others, the length of the raw data can be truncated to match that of the shortest record.

4.1.2 Collection of transmitted signals

For our real experiment we should use the same transducer as we used for the real experiment in the ocean. Figure 4.1 shows the method of collection of the transmitted signal by the hydrophone in the water tank. The measured and the transmitted signals are also shown in figure 4.2.

Although the measured signal in figure 4.2.a contains two pulses, only the first pulse is taken as the transmitted signal because the first pulse is collected by hydrophone directly came from the transducer; the second pulse is the multi-path signal, a reflection from a tank wall or bottom.

From figure 4.2 we can obtain that each transmitted pulse with 24 and 200 kHz carrier have 108 and 45 points duration. The following subsection discusses the raw data and its deconvolution.

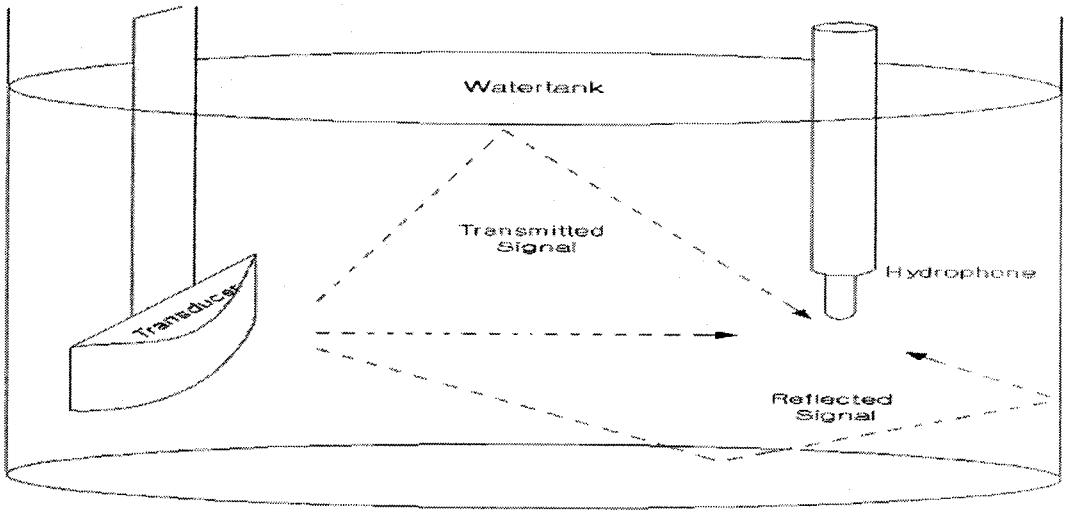
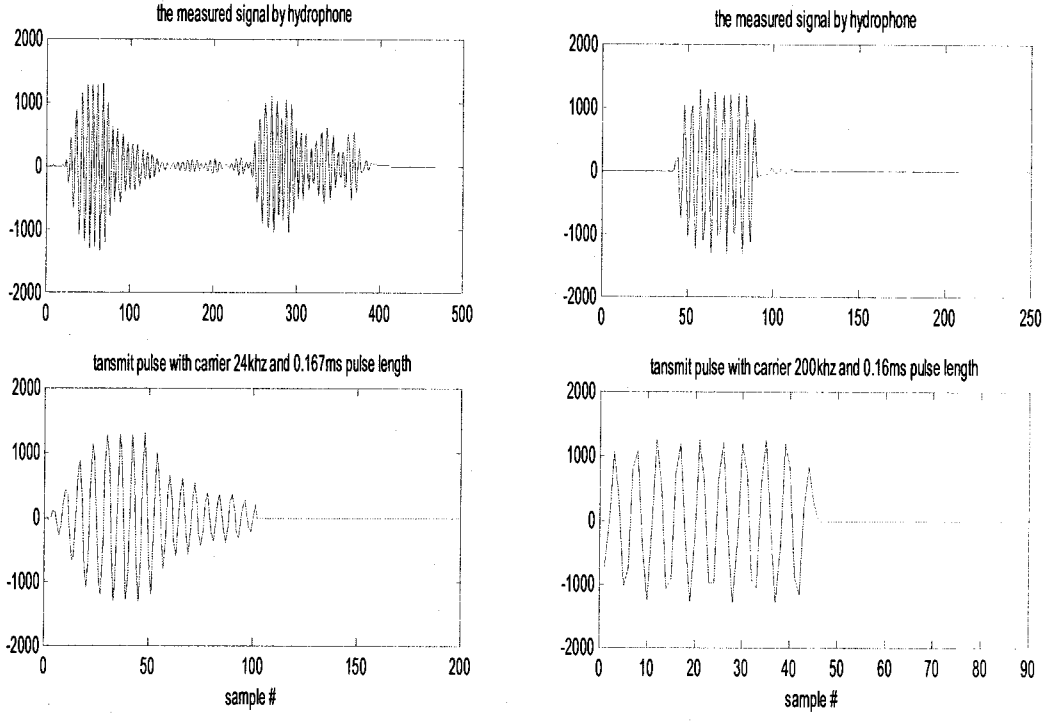


Figure 4.1: Water tank for collecting transmitted signal



(a) 24 kHz (b) 200 kHz

Figure 4.2: the measured signal by hydrophone and the transmitted signal picked from the measured signal in tank simulation

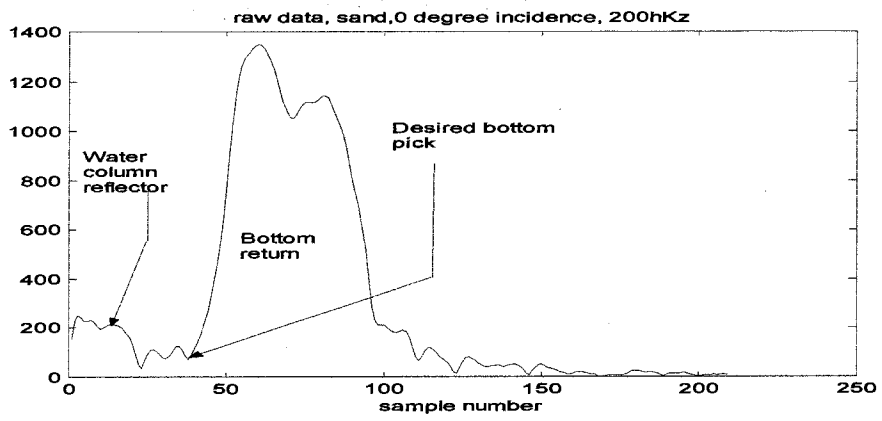
4.2 Result of deconvolution and principal component analysis

Since the received data from sea bottom is convolved with the transmitted signal and the bottom impulse response, we are particularly interested in the bottom impulse response which has all of the information. The results of bottom response from deconvolution are shown in following section.

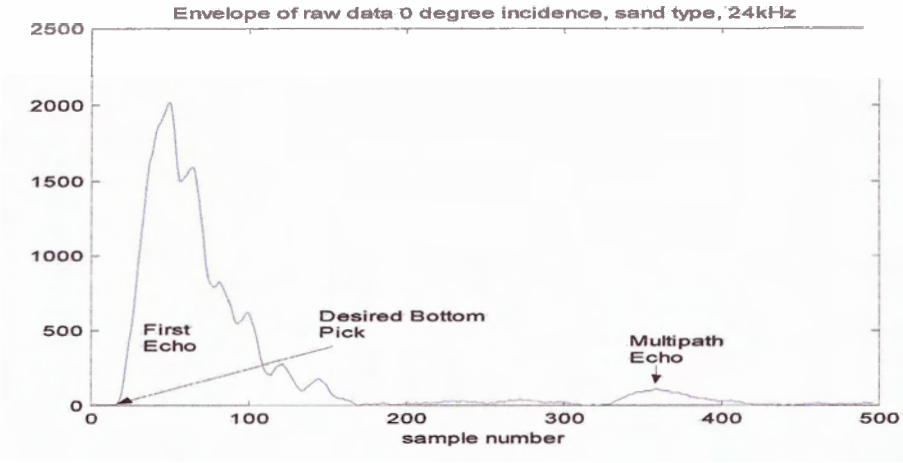
4.2.1 Collection of Reflection data and results of deconvolution

By using the transmitted signal for the real survey, the reflection data from the sea bottom is extracted. There is unexpected data coming from under the sea such as water column reflection and multi-path echoes. However, since we are only interested in the data from the sea bottom these artifacts can be discarded [15, 18].

The following figures show envelopes of a real reflected data vector with low and high frequency at sand type and 0° incidence.



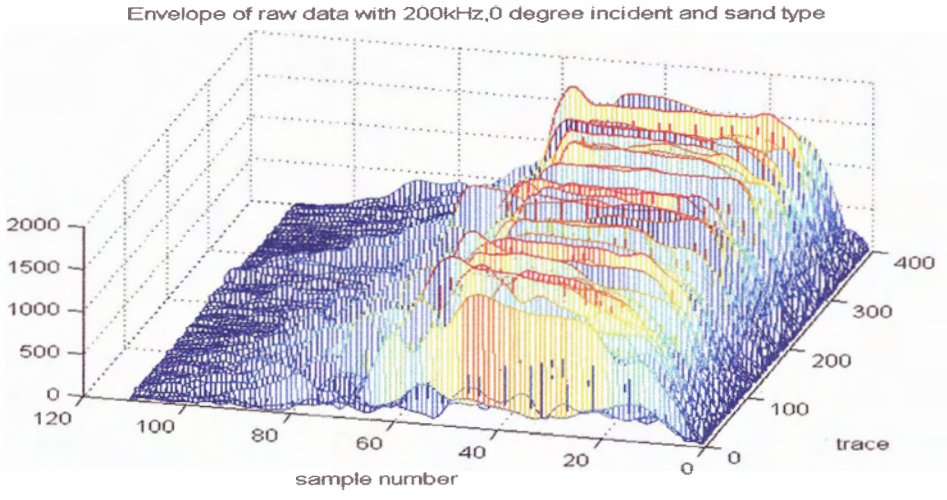
(a) Envelope of echo 200 kHz, 0° incidence and sand type



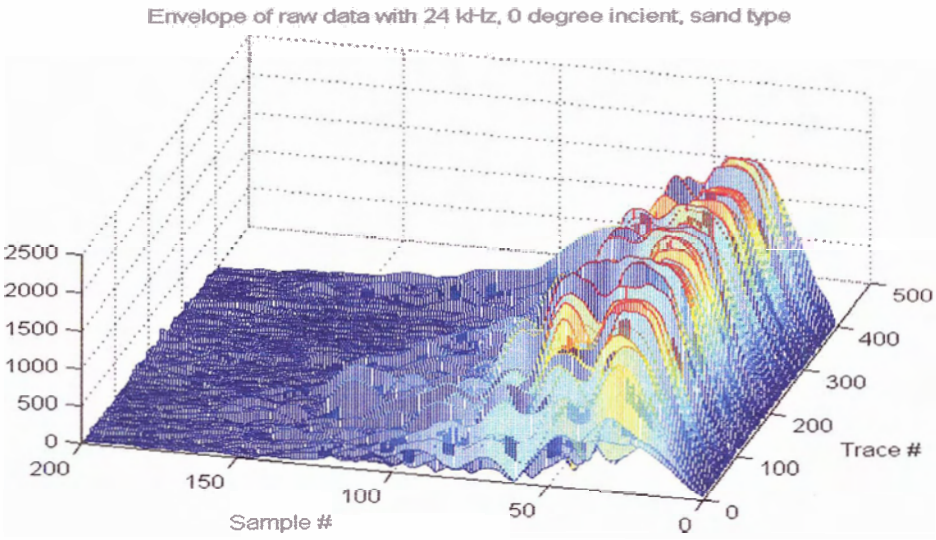
(b) Envelope of echo 24 kHz, 0° incidence and sand type

Figure 4.3: Envelope of echo high and low frequencies at 0° incident and sand type

In the above figures we can extracted the useful echo from sea bottom; the water column reflector in figure 4.3.a and the multi-path echo in figure 4.3.b. can be discarded. A 3-D view of multiple traces of raw data with the same parameters as in figure 4.4 is shown in following figures.



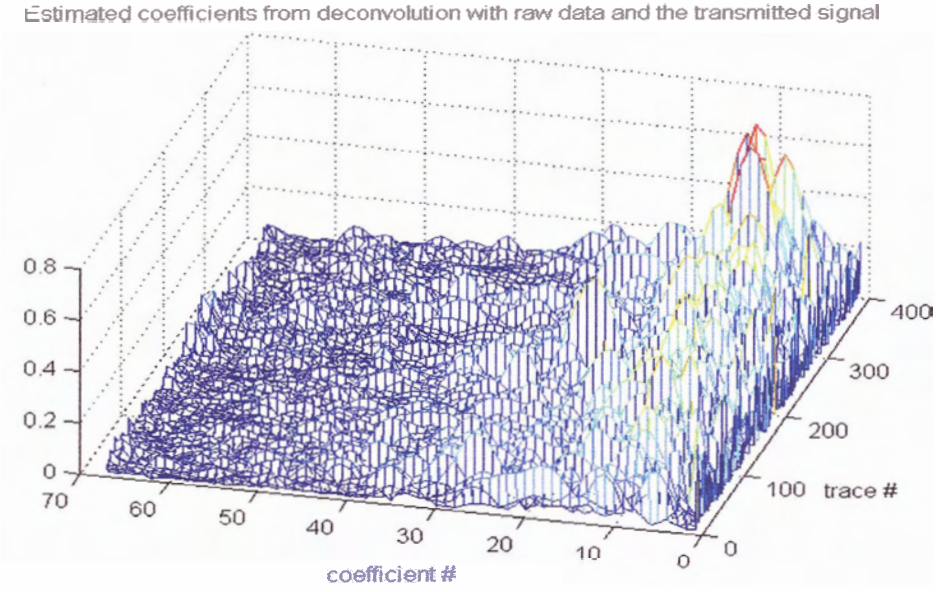
(a) 200kHz, 0° incidence and sand type



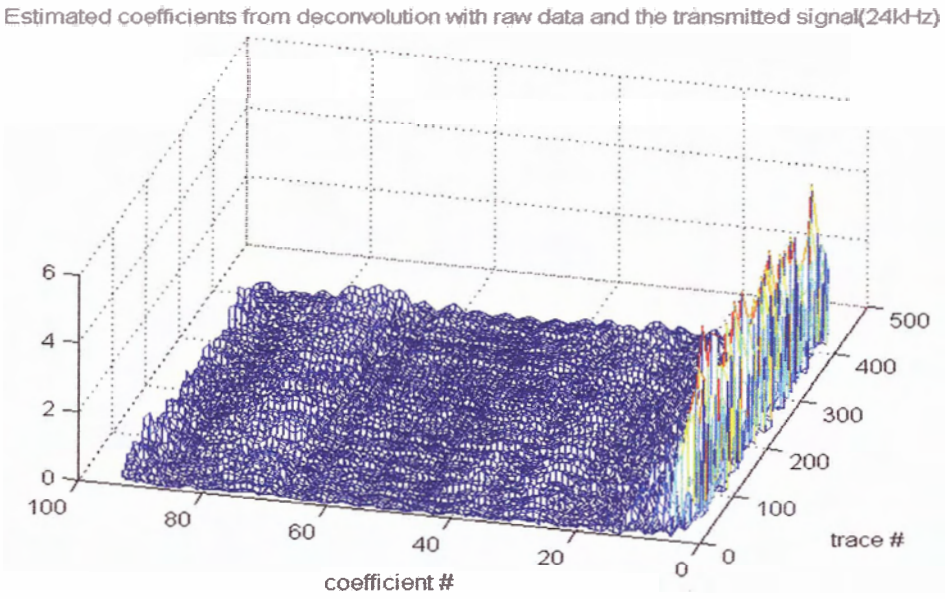
(b) 24 kHz, 0° incidence and sand type

Figure 4.4: Envelope of raw data vectors at 0° incidence and sand type

By using least-square estimation we discussed in chapter 2, we can obtain the results of deconvolution with the raw data and the transmitted signal.



(a) 200kHz sounder



(b) 24 kHz sounder

Figure 4.5: Estimated coefficients from deconvolution with raw data and the transmitted signal

In our example, the apparent dimension M_{low} of the raw reflection vector at 24 kHz is 198 samples, whereas the length L_{low} of the transmitted signal is 108. Therefore we can obtain the dimension P_{low} of the estimated coefficients for low frequency as

$$\begin{aligned} P_{low} &= M_{low} - L_{low} + 1 \\ &= 198 - 108 + 1 = 91 \end{aligned} \tag{4.1}$$

Similarly the dimension M_{high} of the raw data vector at 200 kHz is 110 samples, whereas the length L_{high} of the transmitted signal is 45 points. Therefore the dimension P_{high} of estimated coefficients for high frequency is

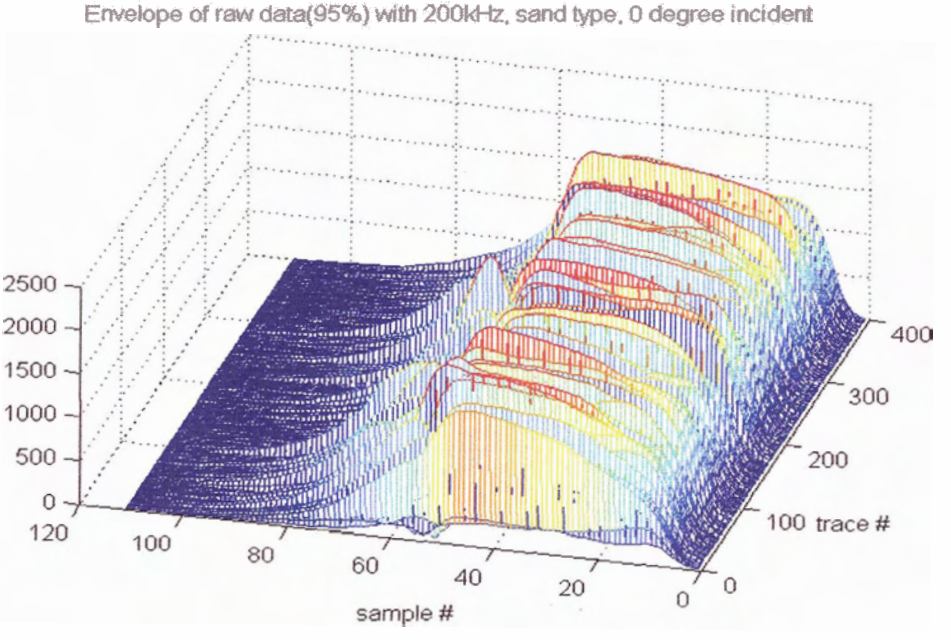
$$\begin{aligned} P_{high} &= M_{high} - L_{high} + 1 \\ &= 110 - 45 + 1 = 66 \end{aligned} \tag{4.2}$$

Figure 4.5 shows the magnitude of the estimated coefficients with 24 and 200 kHz carriers. Similar preprocessing procedures are needed for the other classes.

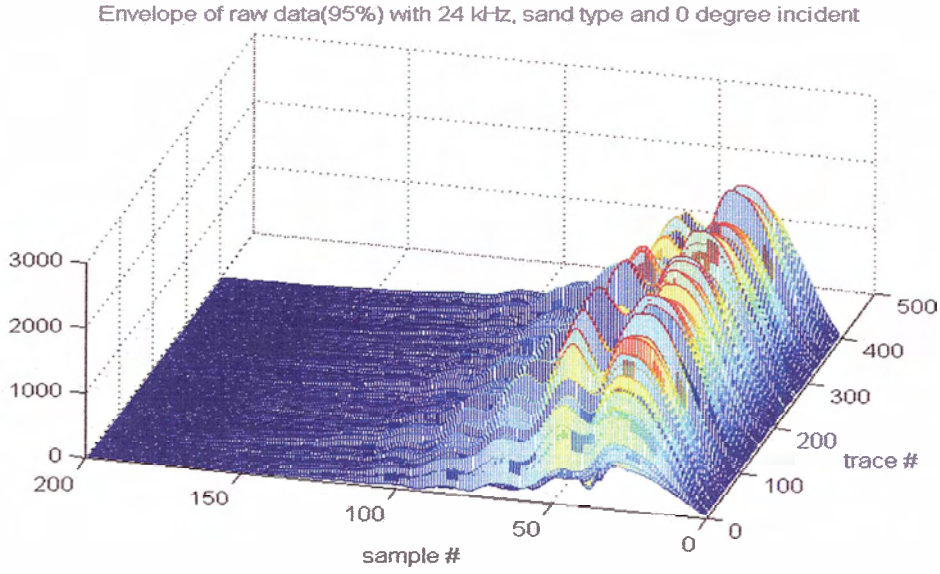
4.2.2 Enhancement of raw data and estimated coefficient by using PCA

The results we obtained in the last section consist of full-length feature vectors, and as seen in figure 4.4, the clustering is confounded by the fact that the majority of records contain noise or variability at each sample time that is not useful in the classification. Therefore it is desirable to reduce the number of dimensions in the feature vectors by using PCA discussed in chapter 2. The following results show the resulting enhancement of raw data and estimated coefficients.

We only take 4 dimensions for 24 kHz and 5 dimensions for 200 kHz in the full feature vectors of raw data to get 95% of covariance matrix's energy of raw data by using PCA.



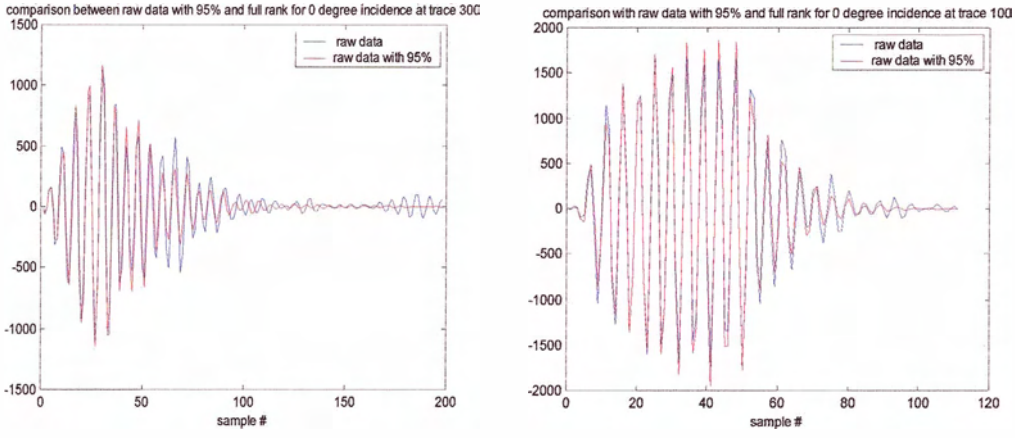
(a) 200kHz, 0° incidence and sand type



(b) 24 kHz, 0° incidence and sand type

Figure 4.6: Envelope of raw data with 95% of covariance matrices in low and high frequencies

By comparing figure 4.4 and 4.6, it is obvious that the raw data with 95% of covariance matrix has enhancement of noise compression more than those with full rank. In order to see the difference in detail, the following figure shows the comparison of raw data.



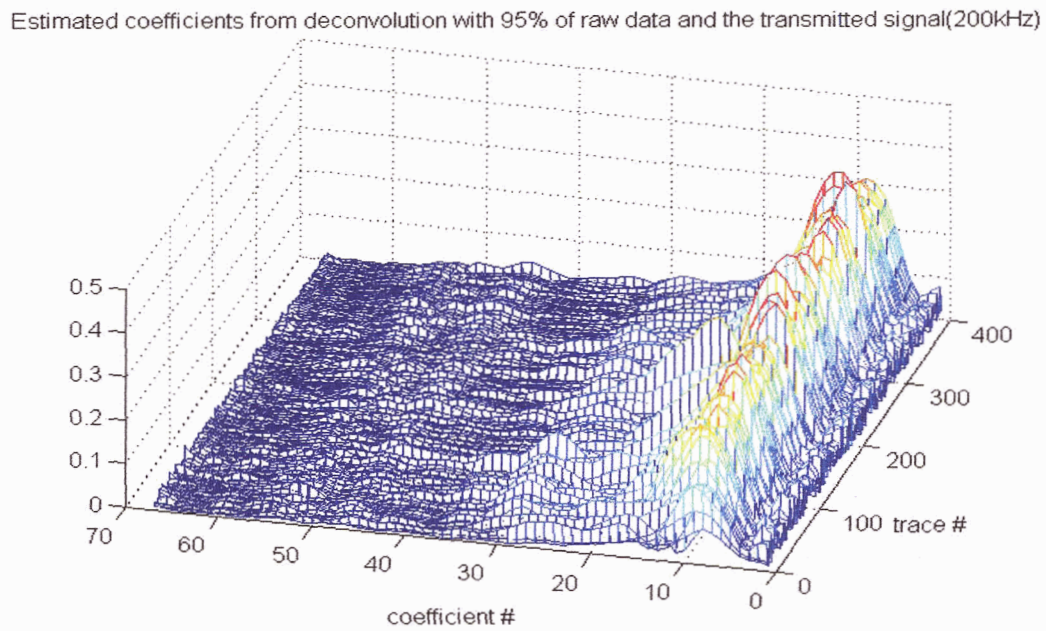
(a) 24 kHz (b) 200 kHz

Figure 4.7: Comparison between raw data with full rank and those with 95% covariance matrices in low and high frequencies, 0° incidence and sand type

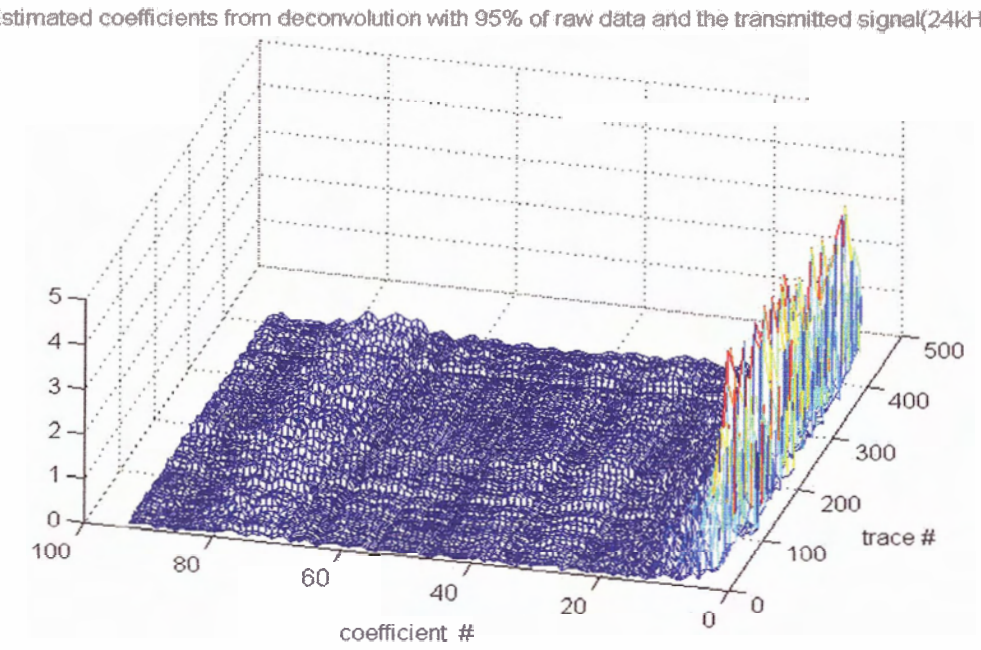
From figure 4.7 it is obvious that the tails of the data vector with 95% of covariance matrix fade away because the first rank reduced data vector contains most of the information of the original. The deconvolution method can be used to obtain the estimated coefficients of the rank reduced raw data, retaining 95% of covariance matrix energy,.

By comparing figure 4.5 and 4.8 it is also obvious that the estimated coefficients of the rank reduced raw data has produced much lower noise results.

We have obtained results from the other cases such as 20°, 40° incidence and sand, mud and gravel type that are similar to those of deconvolution for 0° incidence and sand type.



(a) 200 kHz, 0° incidence and sand type



(c) 24 kHz, 0° incidence and sand type

Figure 4.8: Magnitude of the estimated coefficients obtained by deconvolution of raw data with 95% of covariance matrices and the transmitted signal with 24 and 200 kHz

4.3 Summary

Using the transmitted signal obtained from the water tank test, the sea-bottom reflection model coefficients are found with the deconvolution algorithm. However, because the majority of dimensions of the reflected data contains noise component, we have reduced the number of dimensions in the feature vectors by using PCA, and the enhancement of raw data and estimated coefficient by using PCA have been verified by comparing results between the full rank raw and rank-reduced data.

The following chapter will discuss the distance measures among sonar data classes that have been deconvolved as we have just demonstrated.

Chapter 5
Divergence and Bhattacharyya Distance between Class Pairs

For the measure of "distance" or dissimilarity between classes we have utilized divergence and Bhattacharyya distance as discussed in chapter 3. In this chapter we exhibit the results of the application of divergence and Bhattacharyya distance to the preclassified data provided. Both the reflection raw data and their estimated coefficients are categorized according to their associated transmitted frequency, bottom type and incident angle.

This chapter is divided into two major sections. The first discusses the results of class distance of envelopes of raw data and magnitudes of the estimated coefficients. Further, we are interested in suppressing the noise component in raw data, and this is done by using singular values decomposition (SVD) and obtaining low rank approximations.
In the second part we use our knowledge of the transmitted carrier phase to deconvolve and determine the effect on the resulting distance between classes. We consider the frequency domain forms of the data in order to find the gain factors of improvement through deconvolution for both distance measures. The following tables contain the results for divergence and Bhattacharyya distance.

The first results are of normalized class divergences and Bhattacharyya distance for a 24 kHz carrier. These include 7 classes, according to bottom types and incident angles. Mud and sand type have 3 different incident angles, but for gravel type, only 0° incidence is processed.

We also show the results of normalized class divergences and Bhattacharyya distance for a 200 kHz carrier. These include 5 classes, according to bottom types and incident

angles. Mud has 3 different incident angles, but sand has only 2 because the data for 20° incidence angle has a different sample rate.

The tabulated measures are indicative of the relative ease of discrimination between classes and demonstrate the utility of preprocessing with deconvolution and or raw data rank reduction to enhance SNR.

5.1 Result of normalized divergence and Bhattacharyya distance

For divergence and Bhattacharyya distance measures, we would like the lengths of envelopes of raw data and vectors of magnitudes of the estimated coefficients to be identical among all classes but they are not. Therefore in order to compute a distance between classes having different lengths, we truncate the length of the longer records. This procedure allows the comparison among all the classes, but ignores the fact that if one class' records are longer, the class is clearly distinct on that basis alone. However, because classification information is often found in the early part of the record, the same-length measure is not altogether inappropriate

5.1.1 Normalization method for divergence and Bhattacharyya distance

In order to compare divergences and Bhattacharyya distances, we need somehow to normalize the data. Otherwise just a scale factor could create a new class of data from a single class. Before applying the divergence and Bhattacharyya distance measure, we choose to normalize by forcing average trace energy all classes to be equal. The average energy in each class is found from

$$E_i = |\mathbf{m}_i|^2 + tr(\mathbf{C}_i) \tag{5.1}$$

where $\mathbf{m}_i$ is the mean vector of a class and $\mathbf{C}_i$ is the covariance matrix of class i. The normalized covariance matrix is calculated by dividing covariance matrix by E_i and the mean vector by $\sqrt{E_i}$. Doing this makes every class the same total energy so that we can compare their distributions on location (means) and shape (covariance matrix) without scale being a factor.

5.1.2 Divergence of raw data envelopes and coefficient magnitudes

In tables 5.1 through 5.2, entries with " * " denote truncation of the raw data record length to 196 points to match that of the shortest raw data type record, which is for mud type at 0° incidence. Entries with " ** " denote truncation of the coefficient vector length to 91 points to match that of the shortest coefficient record, which is for mud type at 0° incidence for 24KHz carrier.

Table 5.1: Divergence of envelope of normalized raw data, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types.

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	112	122	4737	516	184	371
	20°*		0	42	704	96	38	180
	40°*			0	2961	337	122	180
Sand	0°*				0	50	138	517
	20°*					0	14	70
	40°*						0	50
Gravel	0°*							0

Table 5.2: Divergence of magnitudes of normalized estimated coefficients, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°**	40°**	0°**	20°**	40°**	0°**
Mud	0°	0	14.28	22.38	54.92	44.31	32.13	61.72
	20°**		0	15.60	64.15	39.87	23.44	54.52
	40°**			0	76.71	41.55	22.51	46.67
Sand	0°**				0	62.97	70.79	127.15
	20°**					0	8.52	18.61
	40°**						0	15.05
Gravel	0°**							0

From table 5.1 and 5.2, the divergence between classes of the same bottom type is smaller than that between different bottom types even if the incident angles of the transmitted signal are different. Typically, when we obtain divergence of envelope of raw data between class of 0° mud type and the other classes, the largest value is the divergence of mud and sand type at 0° incidence for 24 kHz transmitted signal because of the penetration of sound into the bottom and its subsequent reradiation as backscattering at low frequency and high grazing angles, where reflection and scattering from subsurface layers and other inhomogeneities may occur.

In the tables 5.3 through 5.4, entries with " * " denote truncation of the raw data record length to 87 points to match that of the shortest raw data type record, which is for mud type at 0° incidence. Entries with " ** " denote truncation of the coefficient vector length to 43 points to match that of the shortest coefficient record, which is for mud type at 0° incidence. In table 5.1 and 5.2, the divergence of mud and sand type at 0° incidence for 200 kHz transmitted signal is smaller than the others because there is no penetration of sound into sand type bottom and most of the downed sound are reflected

from surface of sand type bottom as like as mud type bottom.

Table 5.3: Divergences for classes' envelopes of normalized raw data for 200 kHz sounder and 0.16ms pulse length.

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°*	40°*	0°*	40°*
Mud	0°	0	5036	11641	195	979
	20°*		0	324	1309	152
	40°*			0	6623	214
Sand	0°*				0	489
	40°*					0

Table 5.4 : Divergence for classes' magnitudes of normalized estimated coefficients from deconvolution raw data with transmitted signal of 200 kHz sounder and 0.16ms pulse length

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°**	40°**	0°**	40°**
Mud	0°	0	52.50	85.57	25.01	87.08
	20°**		0	13.04	20.83	14.28
	40°**			0	34.61	5.88
Sand	0°**				0	34.61
	40°**					0

5.1.3 Bhattacharyya distance of classes' envelope of raw data and magnitude of the estimated coefficients

In order to compute Bhattacharyya distance between classes having different lengths, we truncate the length of the longer records to that of the shortest of the mud type as we did for the divergence measure.

For 24 kHz, the length of raw data also was truncated to 196 points and the length of

estimated coefficients to 91 points the same as the points for divergence.

From table 5.5 and 5.6, the largest value in the first column is mud and sand type at 0° incidence for 24 kHz transmitted signal because low frequency sound penetrates into the bottom and reradiates from the sub-bottom. The results are similar to those of divergence.

Table 5.5: Bhattacharyya distance for classes' envelope of normalized raw data, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	4.41	5.58	8.17	4.09	3.93	5.62
	20°*		0	2.30	6.74	4.01	2.26	3.85
	40°*			0	9.41	6.34	3.79	3.25
Sand	0°*				0	3.02	4.67	8.06
	20°*					0	1.36	5.75
	40°*						0	4.61
Gravel	0°*							0

Table 5.6: Bhattacharyya distance for classes' magnitudes of normalized estimated coefficients from deconvolution raw data with transmitted signal, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°**	40°**	0°**	20°**	40°**	0°**
Mud	0°	0	1.53	2.20	3.83	3.89	2.99	4.87
	20°**		0	1.62	4.26	3.52	2.37	4.34
	40°**			0	5.07	3.46	2.24	3.82
Sand	0°**				0	5.08	5.46	7.55
	20°**					0	0.95	1.91
	40°**						0	1.56
Gravel	0°**							0

For 200 kHz, the length of raw data also was truncated to 87 points, and the length of estimated coefficients to 43 points the same as for divergence

Table 5.7: Bhattacharyya distance for classes' envelope of raw data for 200 kHz sounder, 0.16ms pulse length

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°*	40°*	0°*	40°*
Mud	0°	0	15.97	22.09	7.23	15.44
	20°*		0	9.09	11.63	6.91
	40°*			0	17.51	8.64
Sand	0°*				0	11.75
	40°*					0

Table 5.8: Bhattacharyya distance for classes' magnitude of estimated coefficients from deconvolution raw data with transmitted signal for 200 kHz sounder and 0.16ms pulse length

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°**	40°**	0°**	40°**
Mud	0°	0	4.33	6.10	2.45	6.31
	20°**		0	1.42	2.14	1.53
	40°**			0	3.29	0.68
Sand	0°**				0	3.31
	40°**					0

When we compare distance measures of the envelope of the raw data with those of the magnitude of the estimated coefficients from our results, the former may be better than the latter because we have considerable incoherence of reference waveform phase from trace to trace causing a loss of the coefficients mean term in distance, and also the

deconvolution process is noisy.

5.2 Reduced rank pre-processing raw data for enhanced classification

5.2.1 Bhattacharyya distance of 200 kHz envelope of raw data

We can enhance the SNR by extracting principal components from the raw data using singular value decomposition (SVD).

Tables 5.9.a through 5.9.c contain Bhattacharyya class distances from envelope of raw data reduced to rank 30 and 20 respectively, where rank reduction is applied to enhance signal to noise (random variation)

In tables 5.9.a through 5.9.c for 200KHz carrier, entries with " * " denote truncation of the raw data vector length to 43 points to match that of the shortest raw data record, which is for mud type at 0° incidence.

We see that the values in table 5.9.a are larger (better) than those in table 5.7 because the rank reduced raw data are closer to those of the noiseless signal, thus the distinction and therefore the distance between different classes is enhanced. The values of Bhattacharyya distance with full rank in table 5.7 are smaller than those with rank-reduction in tables 5.9.a through 5.9.c, demonstrating that probability of error is decreased by applying rank reduction to the raw data.

Table 5.9: Bhattacharyya distance for classes' envelope of rank reduced raw

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°*	40°*	0°*	40°*
Mud	0°	0	18.26	23.69	9.67	19.28
	20° *		0	8.88	14.51	8.32
	40° *			0	20.37	7.72
Sand	0°*				0	16.80
	40°*					0

(a) rank 30

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°*	40°*	0°*	40°*
Mud	0°	0	18.28	24.74	9.88	20.11
	20°*		0	10.09	14.61	9.28
	40°*			0	20.66	7.93
Sand	0°*				0	16.83
	40°*					0

(b) rank 25

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°*	40°*	0°*	40°*
Mud	0°	0	20.45	28.96	11.21	22.51
	20°*		0	12.66	14.64	9.38
	40°*			0	23.20	9.05
Sand	0°*				0	16.25
	40°*					0

(c) rank 20

5.2.2 Bhattacharyya distance of 200 kHz estimated coefficients

Tables 5.10.a through 5.10.c contain class Bhattacharyya distances from magnitudes of coefficients after deconvolution of raw data reduced to rank 30 and 20 respectively, where rank reduction is applied to enhance signal to noise (random variation) .

In tables 5.10.a through 5.10.c for 200 kHz carrier, entries with " ** " denote truncation of the coefficient vector length to 43 points to match that of the shortest coefficient record, which is for mud type at 0° incidence

Because the estimated coefficients with rank reduced raw data are closer to those of the noiseless signal, the distance of estimated coefficients between different classes is enhanced. The values of Bhattacharyya distance of coefficients with full rank in table 8 are smaller than those with rank-reduction in tables 5.10, demonstrating that probability of classification error is decreased by applying rank reduction to the raw data.

Table 5.10: Bhattacharyya distance of magnitude of normalized estimated coefficients from deconvolution rank-reduced raw data with transmitted signal

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°**	40°**	0°**	40°**
Mud	0°	0	4.45	6.23	2.75	6.46
	20°**		0	1.92	2.49	1.72
	40°**			0	3.83	1.46
Sand	0°**				0	3.64
	40°**					0

(a) Rank 30

Chapter 5
Divergence and Bhattacharyya Distance between Class Pairs

For the measure of "distance" or dissimilarity between classes we have utilized divergence and Bhattacharyya distance as discussed in chapter 3. In this chapter we exhibit the results of the application of divergence and Bhattacharyya distance to the preclassified data provided. Both the reflection raw data and their estimated coefficients are categorized according to their associated transmitted frequency, bottom type and incident angle.

This chapter is divided into two major sections. The first discusses the results of class distance of envelopes of raw data and magnitudes of the estimated coefficients. Further, we are interested in suppressing the noise component in raw data, and this is done by using singular values decomposition (SVD) and obtaining low rank approximations.
In the second part we use our knowledge of the transmitted carrier phase to deconvolve and determine the effect on the resulting distance between classes. We consider the frequency domain forms of the data in order to find the gain factors of improvement through deconvolution for both distance measures. The following tables contain the results for divergence and Bhattacharyya distance.

The first results are of normalized class divergences and Bhattacharyya distance for a 24 kHz carrier. These include 7 classes, according to bottom types and incident angles. Mud and sand type have 3 different incident angles, but for gravel type, only 0° incidence is processed.

We also show the results of normalized class divergences and Bhattacharyya distance for a 200 kHz carrier. These include 5 classes, according to bottom types and incident

angles. Mud has 3 different incident angles, but sand has only 2 because the data for 20° incidence angle has a different sample rate.

The tabulated measures are indicative of the relative ease of discrimination between classes and demonstrate the utility of preprocessing with deconvolution and or raw data rank reduction to enhance SNR.

5.1 Result of normalized divergence and Bhattacharyya distance

For divergence and Bhattacharyya distance measures, we would like the lengths of envelopes of raw data and vectors of magnitudes of the estimated coefficients to be identical among all classes but they are not. Therefore in order to compute a distance between classes having different lengths, we truncate the length of the longer records. This procedure allows the comparison among all the classes, but ignores the fact that if one class' records are longer, the class is clearly distinct on that basis alone. However, because classification information is often found in the early part of the record, the same-length measure is not altogether inappropriate

5.1.1 Normalization method for divergence and Bhattacharyya distance

In order to compare divergences and Bhattacharyya distances, we need somehow to normalize the data. Otherwise just a scale factor could create a new class of data from a single class. Before applying the divergence and Bhattacharyya distance measure, we choose to normalize by forcing average trace energy all classes to be equal. The average energy in each class is found from

$$E_i = |\mathbf{m}_i|^2 + tr(\mathbf{C}_i) \tag{5.1}$$

where $\mathbf{m}_i$ is the mean vector of a class and $\mathbf{C}_i$ is the covariance matrix of class i. The normalized covariance matrix is calculated by dividing covariance matrix by E_i and the mean vector by $\sqrt{E_i}$. Doing this makes every class the same total energy so that we can compare their distributions on location (means) and shape (covariance matrix) without scale being a factor.

5.1.2 Divergence of raw data envelopes and coefficient magnitudes

In tables 5.1 through 5.2, entries with " * " denote truncation of the raw data record length to 196 points to match that of the shortest raw data type record, which is for mud type at 0° incidence. Entries with " ** " denote truncation of the coefficient vector length to 91 points to match that of the shortest coefficient record, which is for mud type at 0° incidence for 24KHz carrier.

Table 5.1: Divergence of envelope of normalized raw data, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types.

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	112	122	4737	516	184	371
	20°*		0	42	704	96	38	180
	40°*			0	2961	337	122	180
Sand	0°*				0	50	138	517
	20°*					0	14	70
	40°*						0	50
Gravel	0°*							0

Table 5.2: Divergence of magnitudes of normalized estimated coefficients, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°**	40°**	0°**	20°**	40°**	0°**
Mud	0°	0	14.28	22.38	54.92	44.31	32.13	61.72
	20°**		0	15.60	64.15	39.87	23.44	54.52
	40°**			0	76.71	41.55	22.51	46.67
Sand	0°**				0	62.97	70.79	127.15
	20°**					0	8.52	18.61
	40°**						0	15.05
Gravel	0°**							0

From table 5.1 and 5.2, the divergence between classes of the same bottom type is smaller than that between different bottom types even if the incident angles of the transmitted signal are different. Typically, when we obtain divergence of envelope of raw data between class of 0° mud type and the other classes, the largest value is the divergence of mud and sand type at 0° incidence for 24 kHz transmitted signal because of the penetration of sound into the bottom and its subsequent reradiation as backscattering at low frequency and high grazing angles, where reflection and scattering from subsurface layers and other inhomogeneities may occur.

In the tables 5.3 through 5.4, entries with " * " denote truncation of the raw data record length to 87 points to match that of the shortest raw data type record, which is for mud type at 0° incidence. Entries with " ** " denote truncation of the coefficient vector length to 43 points to match that of the shortest coefficient record, which is for mud type at 0° incidence. In table 5.1 and 5.2, the divergence of mud and sand type at 0° incidence for 200 kHz transmitted signal is smaller than the others because there is no penetration of sound into sand type bottom and most of the downed sound are reflected

from surface of sand type bottom as like as mud type bottom.

Table 5.3: Divergences for classes' envelopes of normalized raw data for 200 kHz sounder and 0.16ms pulse length.

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°*	40°*	0°*	40°*
Mud	0°	0	5036	11641	195	979
	20°*		0	324	1309	152
	40°*			0	6623	214
Sand	0°*				0	489
	40°*					0

Table 5.4 : Divergence for classes' magnitudes of normalized estimated coefficients from deconvolution raw data with transmitted signal of 200 kHz sounder and 0.16ms pulse length

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°**	40°**	0°**	40°**
Mud	0°	0	52.50	85.57	25.01	87.08
	20°**		0	13.04	20.83	14.28
	40°**			0	34.61	5.88
Sand	0°**				0	34.61
	40°**					0

5.1.3 Bhattacharyya distance of classes' envelope of raw data and magnitude of the estimated coefficients

In order to compute Bhattacharyya distance between classes having different lengths, we truncate the length of the longer records to that of the shortest of the mud type as we did for the divergence measure.

For 24 kHz, the length of raw data also was truncated to 196 points and the length of

estimated coefficients to 91 points the same as the points for divergence.

From table 5.5 and 5.6, the largest value in the first column is mud and sand type at 0° incidence for 24 kHz transmitted signal because low frequency sound penetrates into the bottom and reradiates from the sub-bottom. The results are similar to those of divergence.

Table 5.5: Bhattacharyya distance for classes' envelope of normalized raw data, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	4.41	5.58	8.17	4.09	3.93	5.62
	20°*		0	2.30	6.74	4.01	2.26	3.85
	40°*			0	9.41	6.34	3.79	3.25
Sand	0°*				0	3.02	4.67	8.06
	20°*					0	1.36	5.75
	40°*						0	4.61
Gravel	0°*							0

Table 5.6: Bhattacharyya distance for classes' magnitudes of normalized estimated coefficients from deconvolution raw data with transmitted signal, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°**	40°**	0°**	20°**	40°**	0°**
Mud	0°	0	1.53	2.20	3.83	3.89	2.99	4.87
	20°**		0	1.62	4.26	3.52	2.37	4.34
	40°**			0	5.07	3.46	2.24	3.82
Sand	0°**				0	5.08	5.46	7.55
	20°**					0	0.95	1.91
	40°**						0	1.56
Gravel	0°**							0

For 200 kHz, the length of raw data also was truncated to 87 points, and the length of estimated coefficients to 43 points the same as for divergence

Table 5.7: Bhattacharyya distance for classes' envelope of raw data for 200 kHz sounder, 0.16ms pulse length

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°*	40°*	0°*	40°*
Mud	0°	0	15.97	22.09	7.23	15.44
	20°*		0	9.09	11.63	6.91
	40°*			0	17.51	8.64
Sand	0°*				0	11.75
	40°*					0

Table 5.8: Bhattacharyya distance for classes' magnitude of estimated coefficients from deconvolution raw data with transmitted signal for 200 kHz sounder and 0.16ms pulse length

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°**	40°**	0°**	40°**
Mud	0°	0	4.33	6.10	2.45	6.31
	20°**		0	1.42	2.14	1.53
	40°**			0	3.29	0.68
Sand	0°**				0	3.31
	40°**					0

When we compare distance measures of the envelope of the raw data with those of the magnitude of the estimated coefficients from our results, the former may be better than the latter because we have considerable incoherence of reference waveform phase from trace to trace causing a loss of the coefficients mean term in distance, and also the

deconvolution process is noisy.

5.2 Reduced rank pre-processing raw data for enhanced classification

5.2.1 Bhattacharyya distance of 200 kHz envelope of raw data

We can enhance the SNR by extracting principal components from the raw data using singular value decomposition (SVD).

Tables 5.9.a through 5.9.c contain Bhattacharyya class distances from envelope of raw data reduced to rank 30 and 20 respectively, where rank reduction is applied to enhance signal to noise (random variation)

In tables 5.9.a through 5.9.c for 200KHz carrier, entries with " * " denote truncation of the raw data vector length to 43 points to match that of the shortest raw data record, which is for mud type at 0° incidence.

We see that the values in table 5.9.a are larger (better) than those in table 5.7 because the rank reduced raw data are closer to those of the noiseless signal, thus the distinction and therefore the distance between different classes is enhanced. The values of Bhattacharyya distance with full rank in table 5.7 are smaller than those with rank-reduction in tables 5.9.a through 5.9.c, demonstrating that probability of error is decreased by applying rank reduction to the raw data.

Table 5.9: Bhattacharyya distance for classes' envelope of rank reduced raw

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°*	40°*	0°*	40°*
Mud	0°	0	18.26	23.69	9.67	19.28
	20° *		0	8.88	14.51	8.32
	40° *			0	20.37	7.72
Sand	0°*				0	16.80
	40°*					0

(a) rank 30

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°*	40°*	0°*	40°*
Mud	0°	0	18.28	24.74	9.88	20.11
	20°*		0	10.09	14.61	9.28
	40°*			0	20.66	7.93
Sand	0°*				0	16.83
	40°*					0

(b) rank 25

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°*	40°*	0°*	40°*
Mud	0°	0	20.45	28.96	11.21	22.51
	20°*		0	12.66	14.64	9.38
	40°*			0	23.20	9.05
Sand	0°*				0	16.25
	40°*					0

(c) rank 20

5.2.2 Bhattacharyya distance of 200 kHz estimated coefficients

Tables 5.10.a through 5.10.c contain class Bhattacharyya distances from magnitudes of coefficients after deconvolution of raw data reduced to rank 30 and 20 respectively, where rank reduction is applied to enhance signal to noise (random variation) .

In tables 5.10.a through 5.10.c for 200 kHz carrier, entries with " ** " denote truncation of the coefficient vector length to 43 points to match that of the shortest coefficient record, which is for mud type at 0° incidence

Because the estimated coefficients with rank reduced raw data are closer to those of the noiseless signal, the distance of estimated coefficients between different classes is enhanced. The values of Bhattacharyya distance of coefficients with full rank in table 8 are smaller than those with rank-reduction in tables 5.10, demonstrating that probability of classification error is decreased by applying rank reduction to the raw data.

Table 5.10: Bhattacharyya distance of magnitude of normalized estimated coefficients from deconvolution rank-reduced raw data with transmitted signal

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°**	40°**	0°**	40°**
Mud	0°	0	4.45	6.23	2.75	6.46
	20°**		0	1.92	2.49	1.72
	40°**			0	3.83	1.46
Sand	0°**				0	3.64
	40°**					0

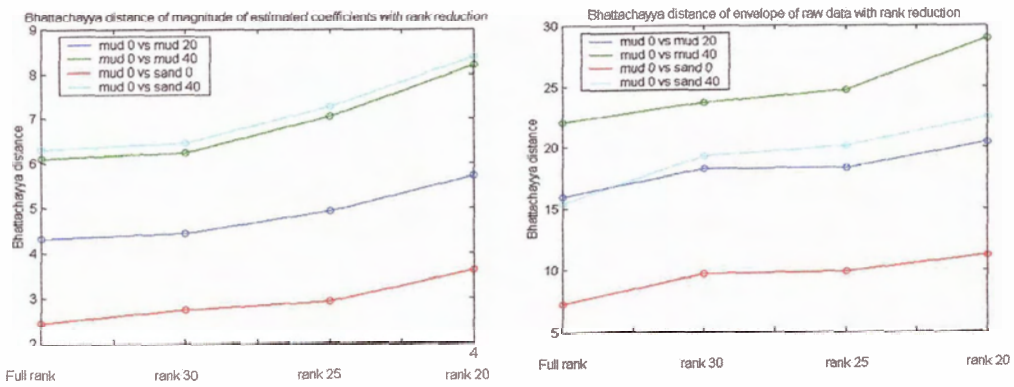
(a) Rank 30

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°**	40°**	0°**	40°**
Mud	0°	0	4.93	7.05	2.94	7.27
	20°**		0	2.48	2.69	1.96
	40°**			0	4.64	2.13
Sand	0°**				0	4.29
	40°**					0

(b) rank 25

200 kHz 0.16ms pulse		Mud			Sand	
		0°	20°**	40°**	0°**	40°**
Mud	0°	0	5.73	8.19	3.62	8.38
	20° **		0	3.80	3.12	2.58
	40° **			0	5.67	2.51
Sand	0°**				0	5.08
	40°**					0

(c) Rank 20



(a) Magnitude of estimated coefficients (b) Envelope of raw data

Figure 5.1: Bhattacharyya distance of magnitude of normalized estimated coefficients and envelope of raw data with rank reduction

Figure 5.1 shows Bhattacharyya distance is getting larger when the ranks are reduced to 20. It is demonstrating that probability of error is decreased by applying rank reduction to the raw data.

5.2.3 Bhattacharyya distance of 24 kHz envelope of raw data

In tables 5.11.a through 5.11.c (rank 30,25 and 20 respectively) for 24 kHz carrier, entries with " * " denote truncation of the raw data length to 196 points to match that of the shortest coefficient record, which is for mud type at 0° incidence.

We can also see that the values in tables 5.10.a become much larger than those of full rank in table 5.5. However, as the rank is reduced, the effect becomes smaller for enhancing distance between the classes of envelope of raw data for 24 kHz from table 5.5.b and c because the noise component might be already suppressed around rank 30.

Table 5.11: Bhattacharyya distance for classes' envelope of rank reduced raw

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	7.19	7.76	9.55	6.46	6.11	6.88
	20°*		0	3.61	9.09	5.88	4.07	7.59
	40°*			0	9.63	5.60	3.41	6.05
Sand	0°*				0	5.05	6.96	8.02
	20°*					0	2.19	4.60
	40°*						0	4.88
Gravel	0°*							0

(a) Rank 30

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	7.30	7.60	9.54	6.20	5.46	6.57
	20°*		0	2.93	8.22	5.30	3.50	6.68
	40°*			0	9.42	5.64	3.22	6.02
Sand	0°*				0	4.82	6.91	7.50
	20°*					0	2.10	4.24
	40°*						0	4.55
Gravel	0°*							0

(b) Rank 25

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	7.35	7.63	10.01	6.28	5.47	6.31
	20°*		0	3.30	7.29	4.31	2.97	7.40
	40°*			0	9.54	5.51	3.57	5.79
Sand	0°*				0	4.60	7.49	10.00
	20°*					0	2.39	7.53
	40°*						0	7.21
Gravel	0°*							0

(c) Rank 20

5.2.4 Bhattacharyya distance of 24 kHz magnitude of estimated coefficients

In tables 5.12.a through c(30, 25 and 20 respectively) for 24 kHz carrier, entries with " ** " denote truncation of the estimated coefficients length to 91 points to match that of the shortest coefficients record, which is for mud type at 0° incidence.

We see that the values in table 5.11.a through c become larger as rank is reduced because the rank-reduced raw data is closer to the noiseless signal, enhancing the

distance between different classes, demonstrating that probability of error is decreased by applying rank reduction to the raw data.

Table 5.12: Bhattacharyya distance for classes' magnitude of normalized estimated coefficients from deconvolution of the transmitted signal from the raw data, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types as rank is reduced

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°**	40°**	0°**	20°**	40°**	0°**
Mud	0°	0	4.06	4.83	5.60	3.73	3.28	7.53
	20°**		0	4.93	7.50	6.25	5.26	10.25
	40°**			0	8.48	6.40	4.94	8.51
Sand	0°**				0	4.95	6.97	10.06
	20°**					0	1.71	4.10
	40°**						0	3.33
Gravel	0°**							0

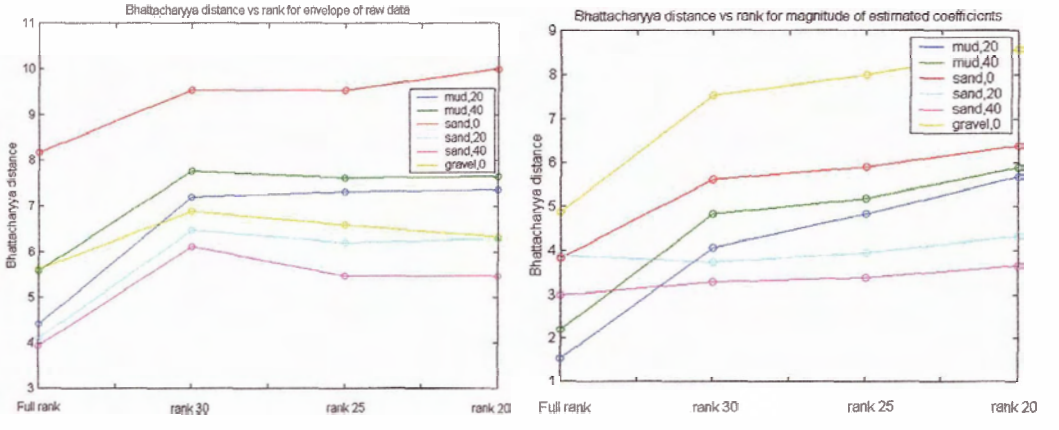
(a) Rank 30

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°**	40°**	0°**	20°**	40°**	0°**
Mud	0°	0	4.82	5.17	5.89	3.94	3.39	8.01
	20°**		0	5.53	8.38	7.16	6.01	11.04
	40°**			0	9.04	6.93	5.24	8.79
Sand	0°**				0	5.24	7.32	10.59
	20°**					0	1.90	4.75
	40°**						0	3.81
Gravel	0°**							0

(b) Rank 25

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°**	40°**	0°**	20°**	40°**	0°**
Mud	0°	0	5.66	5.88	6.39	4.31	3.66	8.57
	20°**		0	6.28	9.44	8.36	6.84	12.16
	40°**			0	9.67	7.36	5.40	9.18
Sand	0°**				0	5.63	7.77	11.56
	20°**					0	2.18	5.29
	40°**						0	4.45
Gravel	0°**							0

(c) Rank 20



(a) Envelope of raw data (b) Magnitude of estimated coefficients

Figure 5.2: Bhattacharyya distance vs Rank of envelope of raw data and magnitude of estimated coefficients 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

From figure 5.2 it is obvious that the Bhattacharyya distance becomes greater as the rank is reduced to 30 from full rank. However, for envelopes of raw data, the slope of distance becomes almost zero and for magnitude of estimated coefficients, the slope of distance becomes slightly larger as the rank is getting smaller. Therefore we might choose about rank 20 for the distance measure.

5.3 Results of measuring Gaussian mixture distance between classes

In sections 5.1 and 5.2, we have discussed the classification distance between classes of envelopes of raw data and magnitude of estimated coefficients from deconvolution. However, since the mean vectors of the envelopes of reflected raw data in each class do not become zero, because envelopes are nonlinearly extracted and random phases do not affect them, the classification distance of envelope of raw data is greater than that of magnitude of the estimated coefficients, which are linearly extracted from data and thus are much affected by random phase..

In chapter 3, when the raw data is the convolution of the transmitted signal with the reflector impulse response, we found coherent and incoherent distance measures by considering the coherent $X(f)$ and incoherent $X(f)$ with random phase over [-π, π), which has mean $\bar{X}(f)$ and variance $\sigma_X^2(f)$. In both coherent and incoherent distance measures we found that $X(f)$ has disappeared in the distance measures.

From the theoretical results on coherent and incoherent distance measures, we can expect to see the best class separation results when coherent deconvolution is applied to raw data vectors of sonar responses. The coherence is due only to an assumed accurate pick of start times of the reflections, allowing use of the mean vectors. However it is certain that considerable error in start time picks and uncertain phases have degraded potentially ideal coherence in the reflections.

In the following section we use Eq. (3.27) and (3.31) for coherent and Eq. (3.29) and (3.33) for incoherent to compare the results of coherent and incoherent divergence and Bhattacharyya distance between classes.

At higher frequencies the coefficients are not as well defined because 200 kHz does not penetrate the bottom as well as at 24 kHz. Also, it is more difficult to get phase coherency from aligning the traces since the high frequency period in time is much shorter. For this reason we only consider the results of the lowest frequency, 24 kHz.

5.3.1 Results of coherent and incoherent divergence

From preclassified narrow beam sonar shallow sea bottom reflections we have obtained the following measures of coherent processing over incoherent for divergence and Bhattacharyya distance between several sea bottom class pairs due.

In the following tables entries with " * " denote truncation of the coefficient vector length to 91 points to match that of the shortest coefficient record, which is for mud type at 0° incidence for 24 kHz carrier.

Table 5.13: Coherent and incoherent divergence at 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	169	619	1450	976	872	2459
	20°*		0	208	1155	485	421	1430
	40°*			0	771	166	134	578
Sand	0°*				0	500	563	1408
	20°*					0	15	415
	40°*						0	415
Gravel	0°*							0

(a) Coherent

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	156	607	866	935	854	2408
	20°*		0	200	697	467	416	1378
	40°*			0	368	142	128	546
Sand	0°*				0	276	260	938
	20°*					0	57	352
	40°*						0	352
Gravel	0°*							0

(b) incoherent

In order to get the coherent and incoherent Bhattacharyya distance, we are using Eq. (3.31) for the coherent and Eq. (3.33) for the incoherent first. The coherent Bhattacharyya distance is also found by assuming that the start times of the responses have been estimated closely.

Table 5.14: Coherent and incoherent Bhattacharyya distance at 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	10.14	34.82	71.86	52.20	47.41	111.91
	20°*		0	12.60	54.34	28.15	24.72	72.63
	40°*			0	42.01	10.00	8.04	32.66
Sand	0°*				0	28.18	31.89	72.31
	20°*					0	0.97	24.37
	40°*						0	26.13
Gravel	0°*							0

(a) coherent

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°*	40°*	0°*	20°*	40°*	0°*
Mud	0°	0	9.33	34.16	44.65	50.18	46.45	109.94
	20°*		0	12.12	32.53	27.16	24.42	70.24
	40°*			0	19.86	8.54	7.68	30.927
Sand	0°*				0	15.43	14.53	48.37
	20°*					0	0.35	20.76
	40°*						0	23.97
Gravel	0°*							0

(b) incoherent

From tables 5.13 and 5.14 the distance results between the same sea bottom types are smaller than those of the different sea bottom types. When we compare the values in tables 5.13.a and 5.14.a to 5.13.b and 5.14.b we find that, in general, the coherent results are greater than those of incoherent results.

5.3.2 Results of gain in distance measure due to coherent deconvolution

By comparing the coherent results of table 5.13.a and 5.14.a to the incoherent results of 5.13.b and 5.14.b, we can compare class distance with ideal (coherent) deconvolution to that with random phase uniform in $[-\pi, \pi)$. Gains greater than 1 imply that.

Table 5.15: Gain in divergence and Bhattacharyya distance at 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°	40°	0°	20°	40°	0°
Mud	0°	0	1.08	1.02	1.67	1.04	1.02	1.02
	20°		0	1.04	1.66	1.04	1.01	1.04
	40°			0	2.10	1.17	1.05	1.05
Sand	0°				0	1.81	2.16	1.50
	20°					0	2.72	1.18
	40°						0	1.18
Gravel	0°							0

(a) Divergence

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°	40°	0°	20°	40°	0°
Mud	0°	0	1.08	1.02	1.60	1.04	1.02	1.02
	20°		0	1.04	1.67	1.04	1.01	1.03
	40°			0	2.11	1.17	1.05	1.06
Sand	0°				0	1.83	2.19	1.50
	20°					0	2.72	1.17
	40°						0	1.19
Gravel	0°							0

(b) Bhattacharyya distance

From table 5.15 we recognize that entries in both tables are almost unity except those between sand type, 0° incidence and other classes. We conclude that a fairly coherent deconvolution can be obtained from the classification distance between sand type, 0° incidence and other sea bottom classes.

5.4 Result of Bhattacharyya distance for complex deconvolution

As we discussed in section 2.5, coherent deconvolution can be ideally achieved when the start time of the reflections is aligned precisely so that the phase of the transmit pulse is identical for all responses. Complex deconvolution can be used to obtain all the coherent information in a phase-shifted or delayed reflection. The following tables compare Bhattacharyya distances of real coefficients, magnitude of coherent complex coefficients and the coherent complex.

Table 5.16: Bhattacharyya distance of real estimated coefficients after alignment by crosscorrelation, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°**	40°**	0°**	20°**	40°**	0°**
Mud	0°	0	3.15	12.81	8.44	6.12	6.59	5.36
	20°**		0	13.00	10.30	7.07	6.86	5.98
	40°**			0	10.70	12.44	9.81	9.59
Sand	0°**				0	8.40	7.58	9.31
	20°**					0	3.23	7.85
	40°**						0	7.66
Gravel	0°**							0

Table 5.16 contains Bhattacharyya distances for real estimated coefficients obtained by deconvolution after the reflections are aligned by crosscorrelation. Tables 5.17 shows Bhattacharyya distance of magnitude of complex coefficients obtained by complex deconvolution after the reflections are aligned by crosscorrelation.

Table 5.17: Bhattacharyya distance of magnitude of complex coefficients aligned by crosscorrelating, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°**	40°**	0°**	20°**	40°**	0°**
Mud	0°	0	3.48	10.52	7.35	4.27	4.86	5.45
	20°**		0	9.73	7.50	4.97	5.53	5.45
	40°**			0	9.33	9.09	9.54	8.26
Sand	0°**				0	6.90	8.90	7.19
	20°**					0	2.79	5.94
	40°**						0	7.23
Gravel	0°**							0

Table 5.18: Bhattacharyya distance of complex coefficients after alignment by crosscorrelation, 24 kHz sounder, 0.167ms pulse length, 3 different bottom types

24 kHz 0.167ms pulse		Mud			Sand			Gravel
		0°	20°**	40°**	0°**	20°**	40°**	0°**
Mud	0°	0	3.12	16.60	12.24	9.46	10.50	7.96
	20°**		0	16.79	13.13	10.19	10.64	8.12
	40°**			0	13.45	13.66	11.64	14.22
Sand	0°**				0	7.67	7.67	13.31
	20°**					0	3.33	11.06
	40°**						0	11.24
Gravel	0°**							0

Table 5.18 shows results of coherent complex coefficients obtained by complex deconvolution after the reflections are aligned by crosscorrelation.

Since the coherent complex deconvolution contains all the coherent information, the values in table 5.18 are larger than those in table 5.16, 5.17. Further, the results in table 5.16 are larger than those in table 5.6, indicating that the alignment among the reflection vectors in a class plays an important role in obtaining coherent deconvolution and that complex demodulation and deconvolution can yield significant improvement in probability of classification error..

5.5 Summary

In sections 5.1 and 5.2 the classification distance between classes of envelopes of raw data and magnitudes of estimated coefficients from deconvolution are discussed because the means of the envelope of reflected raw data in each class do not become zero. The classification distance of envelope of raw data is better than that of magnitude of the estimated coefficients in our results because we have considerable incoherence of reference waveform phase from trace to trace causing a loss of the coefficients mean term in distance, and also the deconvolution process is noisy.

By suppressing the noise component in raw data by using singular values decomposition (SVD) and using a lower rank approximation to the data, distance measures are increased because the rank reduced raw data are closer to those of the noiseless signal.

In sections 3 through 6 we have considered the cases of both coherent and incoherent transmitted waveforms, comparing theoretically in the frequency domain. Knowledge of the transmitted carrier phase (coherent case) allows coherent deconvolution. We have obtained the experimental gain factors, the ratio of coherent to incoherent with random

phase uniform in $[-\pi, \pi)$ distance measures. Our results indicate that a fairly coherent deconvolution can be obtained from the classification distance between sand type, 0° incidence and other sea bottom classes.

Coherent deconvolution in practical problems has been achieved when the start time of the reflections is aligned precisely so that the phase of the transmit pulse is identical for all responses. The results of coherent deconvolution are larger than those for incoherent deconvolution. Since coherent complex deconvolution contains all the information in the phase shifted reflection, we can obtain better distance measure with this process than with any of the others we have considered.

Chapter 6

Summary and Recommendations

6.1 Least-square estimation of deconvolved sea-bottom sonar responses using SVD rank-reduced data

In order to determine the character of sonar responses of the sea-bottom, we compared variations not only with bottom class but also with frequency, beamwidth, angle of incidence and depth. Instead of using the raw sonar responses, we might get better estimates of the distinctive features of these variations by deconvolving the transmitted signal out of the response. For deconvolution we have utilized *Least-Square Estimation (LSE)* to find the bottom's impulse responses. However in the real environment there is a random noise component in addition to the signal components which we want to obtain. We have modeled that noise as additive, white, zero mean Gaussian. In order to get the desired information from bottom responses we can suppress the noise component by applying PCA to better estimate the signal component and the signal covariance matrix of the received data, because most of energy is concentrated in just a few dimensions. By using both deconvolution and PCA we obtain a better estimate of the unknown sea-bottom reflectors that are our real interest.

Consequently, the deconvolution method can be useful to find the unknown system function which we are really interested in.

6.2 Methods for measuring Gaussian mixture distance between classes

Because direct minimization of the error probability to determine the optimum signal is

often difficult to carry out, we measure the distance between classes. The distance measure is a simple alternative to the error probability. In our thesis we have utilized two common distance measures for class separation: divergence and Bhattacharyya. Bhattacharyya distance is derived from the upper bound on the error probability, which has a closer relation with the Bayes error and divergence equates the class separability to the difference of the information between the two classes. However, if the two classes have identical covariances, both distance measures become the Mahalobis distance.

We have used knowledge of the transmitted carrier's phase and have determined its influence on the resulting distance between classes. Using a frequency domain formulation, we have found the gain in distance due to knowledge of the transmitted waveform relative to distance in the incoherent case. In theory, coherence allows exact deconvolution. Coherent deconvolution in a practical problem has been achieved to the extent 1) that the start time of the reflections has been determined precisely so that the phase of the transmit pulse is identical for all responses and 2) that the reflector's class features include some coherence in the reflections.

6.3 Results of LSE of deconvolved sea-bottom sonar responses using SVD rank-reduced signals and class distance measures

The research associated with this thesis has utilized two software packages: MATLAB and the complete seabed classification system (QTC VIEW) developed by Quester Tangent Corporation. The processing techniques discussed in Chapter 2 and 3 have been applied to three world data sets. Those and the transmitted signal records from the simulation tank were provided by the Quester Tangent Corporation.

In chapter 4 the deconvolution algorithms we have presented are applied to three distinct real-world data sets. The transmitted signal was generated by the same transducer as was used in the ocean, and it was collected in a noiseless water tank environment. The reflection was preprocessed by only taking a few dimension of the received data containing most of the energy. The raw data and their estimated coefficients have been compared with coefficients from the rank reduced data. Rank reduction enhances the SNR and increases distance between classes.

In chapter 5 two classification distance measures were studied: First we discussed the results of the class pair distance measures using 1) envelopes of raw data and 2) magnitudes of estimated coefficients from deconvolution. Next, we measured the gain of coherent processing over incoherent for divergence and Bhattacharyya distances between several sea bottom raw class pairs. Third, we obtained coherent complex coefficients from complex deconvolution after alignment of reflection by cross-correlation. The results support the use of complex deconvolution when coherence is a factor for class discrimination and when care is taken to align start times.

6.4 Recommendations for further research

Although we have investigated class pair distances, other suggested future work would compare blind classification (clustering) before and after processing with any of our tested methods (SVD noise reduction, envelopes, estimated coefficients, etc) to determine error probabilities for comparison.

Bibliography

[1] Milligan, S.D., LeBlanc, L.R., and Middleton, F.H. Statistical grouping of acoustic reflection profiles. *Journal of the Acoustical Society of America.* 64(3), pp.795-807, Sept., 1978.

[2] Caughey, D.A. Seabed classification from acoustic echosounder returns, University of Victoria, 1996.

[3] LeBlanc, L.R., Mayer, L., Rufino, M., Schock, S.G., and King, J. Marine sediment classification using the chirp sonar. *Journal of the Acoustical Society of America.* 91(1), pp.107-115, Jan., 1992.

[4] Caughey, D.A., Kirlin, R.L. Blind deconvolution of echosounder envelopes ICASSP-96., Vol. 6, pp.7-10, May, 1996

[5] Kirlin, R.L., and Done, W.J., Covariance Analysis for Seismic Signal Processing, Geophysical Developments No.8, Society of Exploration Geophysicists, 1999. [ch 4]

[6] Prokis, J.G, and Rader, Algorithms for statistical signal processing, Prentice-Hall,2002.[ch 4]

[7] Freire, S.L.M., Ulrych, T., Application of singular value decomposition to vertical seismic profiling, Geophysics 53, pp.778-785, 1988.

[8] Galloway, J.L, Collins, W.T, Dual frequency acoustic classification of seafloor habitat using the QTC VIEW ,OCEANS '98 Conference Proceedings ,vol.3, 1998

[9] Jeffrey, H., An invariant form for the prior probability in estimation problems, Proc. Poy. Soc. A., vol 186, pp. 453-461, 1946.

[10] Tou, J.T., and Gonzalez, R.C, Pattern recognition principles, Addison-Wesley, 1974, [ch 7].

[11] Fukunaga, K., Introduction to statistical pattern recognition, 2nd ed, Academic press, 1990, [ch 3].

[12] Kailath, T., The divergence and Bhattacharyya distance measures in signal election, IEEE Trans. Communications, No. 1, pp. 52-80, Feb, 1967,.

[13] Kirlin, R.L. and Done, W.J. Coherent noise suppression in seismic data using eigenvalue/eigenvector decomposition, ICASSP'88, pp.908 -911, April, 1988.

[14] Li, X.Q. and King, I. Gaussian mixture distance for information retrieval, Neural Networks, IJCNN '99, pp 2544 -2549, 1999.

[15] Comaniciu, D., Meer, P., Kun Xu, Tyler, D., Retrieval performance improvement through low rank corrections, CBAIVL '99 Proceedings IEEE Workshop, June, 1999.

[16] Preston, J.M., Collins, W.T., Bottom classification in very shallow water by high-speed data acquisition, Proceedings of Oceans'00, Providence, RI, 2000.

[17] Urick. R.J, Principles of underwater sound, 3rd, McGraw-Hill, 1983,[ch8].

[18] Prager, B.T., Caughey, D.A., and Poeckert, R.H., Bottom classification: operational results from QTC VIEW, *Proceedings of Oceans'95*, Vol. 3, Oct, 1995.

[19] Kim, H.J, Chang, J.K, Seabed classification from acoustic profiling data using the similarity index, *Journal of the Acoustical Society of America.* 91(1), Jan, 1992.

[20] Jackson, D.R., Winebrenner, D.P., and Ishimaru, A., Application of the composite roughness model to high-frequency bottom backscattering, *Journal of the Acoustical Society of America,* 79(5), May, 1986.

[21] Lee, C.H., Choi, E.S, Bayes error evaluation of the Gaussian ML classifier, IEEE transaction on geosicence and promote sensing, 31(3), May. 2000.

[22] Godin, O.A., Mechanisms and statistics of high frequency ocean bottom backscatter, Final project report for QTC, Sep. 1999.

[23] Jackson, D.R., and Briggs, K.B., High-frequency bottom backscattering: roughness versus sediment volume scattering, Journal of the Acoustical Society of America, 92(2), pp. 962-977, Aug, 1992.

[24] Mourad, P.D., and Jackson, D.R. High-frequency sonar equation models for bottom backscatter and forward loss, Proceedings of Oceans'89, pp.1168-1175, NY, 1989.

[25] Jackson, D.R., Baird, A.M., Crisp, J.J., and Thomson, P.A.G., High-frequency bottom backscatter measurements in shallow water, Journal of the Acoustical Society of America, pp. 1188-1199, 1986.

[26] Mourad, P.D., and Jackson, D.R., A model/data comparison for low-frequency bottom backscatter, Journal of the Acoustical Society of America, pp.344-358, July, 1993.

[27] Prager, B.T., Caughey, D.A. and Poeckert, R.H., Bottom classification operational results from QTC VIEW, Proceedings of Oceans'95., San Diego, CA, 1995.

[28] Kim, H.J., Chang, J.K., Jou, H.T., Park, G.T., Seabed classification from acoustic profiling data using the similarity index, Journal of the Acoustical Society of America, 111(2), pp. 794-799, Feb. 2002.

[29] Galloway, J.L., Collins, W.T., Dual frequency acoustic classification of seafloor habitat using the QTC VIEW, Proceedings of Oceans'98., Nice, France, 1998.

[30] Peelbes, P.Z., Probability, random variables, and random signal principles, 3rd ed, McGraw-Hill, 1993, [ch 6].

[31] Scharf, L.L, Statistical signal processing: detection, estimation, and time series analysis, Addison-Wesley, 2002, [ch 9].

[32] Kavli, T., Carlin, M., and Weyer, E., Real-time seabed classification using multi-

frequency echo sounders, Proceeding of Oceanology Internation 94, pp.157-164. Institute of Acoustic, Bath, U.K., 1993.

[33] Kirlin R.L., Deconvolution for enhancement of separation of multivariate Gaussian clusters, University of Victoria, 2002.

Appendix A

START TIME JITTER

A.1 Example Two-impulse train reflection coefficients with timing error

If the true reflection coefficient train is $h(t) = \delta(t) - \delta(t - D)$ and we have one correct reflection start time and one start time estimate that is wrong by D sec, then we get a mean reflection coefficient train equal to

$$\hat{h}(t) = \left(\left[\delta(t) - \delta(t - D)\right] + \left[\delta(t - D) - \delta(t - 2D)\right]\right)/2$$
$$= \frac{1}{2}\delta(t) + 0\delta(t - D) - \frac{1}{2}\delta(t - 2D)$$

For the pulse at D sec has been averaged to zero just because we got the start time wrong on one of the reflections. In addition, we have a smaller coefficient at time 0 and a reflection at time 2D where we shouldn't.

A.2 General start time variations

So this is why start times are so important. Since we have been rather crude in picking start times, the coefficients we are working with are a mix of delays of the true ones for each reflection. From Kirlin's personal communication, we are probably getting something like the true coefficient train mean vector for each class, but it is smeared, essentially like a blurred picture from a moving camera or a moving object in the

picture. If the distance (time) of blur is not too large, we are probably getting a reasonable measure of the coefficients and for lower frequency carrier they may be moderately accurate and non-zero mean.

Appendix B

Sequential decimation and sampling frequency determination for the echo-sounder data

B.1 Decimation

Figure B.1.(a) shows the continuous-time Fourier Transform of $x_c(t)$, and Figure B.1.(b) shows the discrete-time Fourier transform of the sequence $x[n] = x_c(nT)$, when $2\pi / T = 4\Omega_N$. Thus, $\omega_N = \Omega_N T = \pi / 2$. If we downsample by a factor of M=3, we obtain the sequence $x_d[n] = x[3n] = x_c(n3T)$ whose discrete-time Fourier transform is plotted in Figure B.1.(c) with normalized frequency $\omega = \Omega T'$. Note that because $M\omega_N = 3\pi / 2$, which is greater than π, aliasing occurs. In general, to avoid aliasing in downsampling by a factor of M requires $\omega_N M < \pi$. If this condition does not hold, aliasing occurs. In order to downsample without aliasing, we can reduce the bandwidth of the signal $x[n]$ before downsampling. Thus, if $x[n]$ is filtered by an ideal lowpass filter with cutoff frequency π / M, then the output $\tilde{x}[n]$ can be downsampled without aliasing, as illustrated in Figure B.1.(d),(e), and (f).

Figure B.2 shows a general system for downsampling by a factor of M which is called a decimator, and downsampling by lowpass filtering followed by compression has been termed decimation [1].

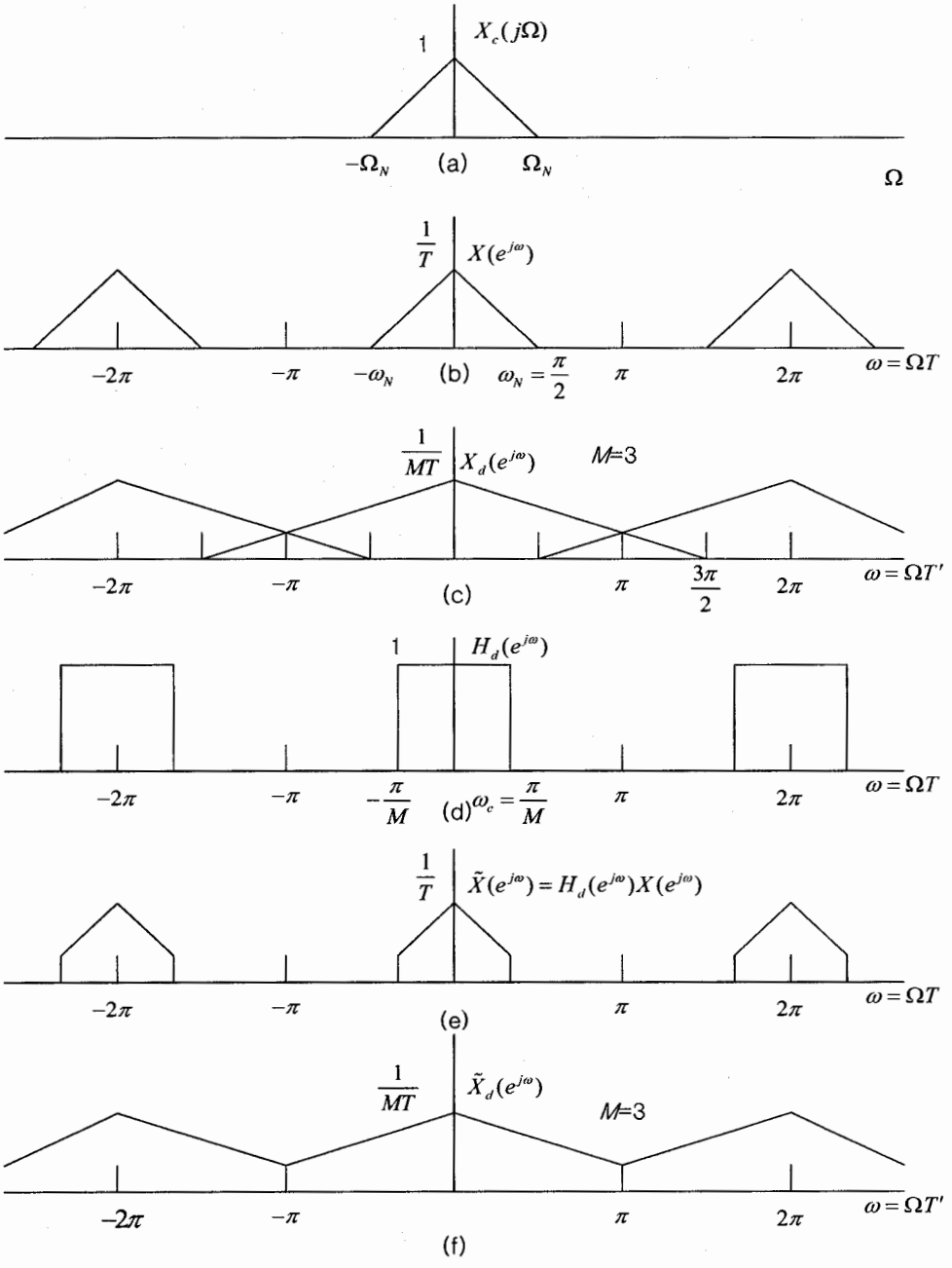


Figure B.1 (a)-(c) Downsampling with aliasing. (d)-(f) Downsampling with prefiltering to avoid aliasing

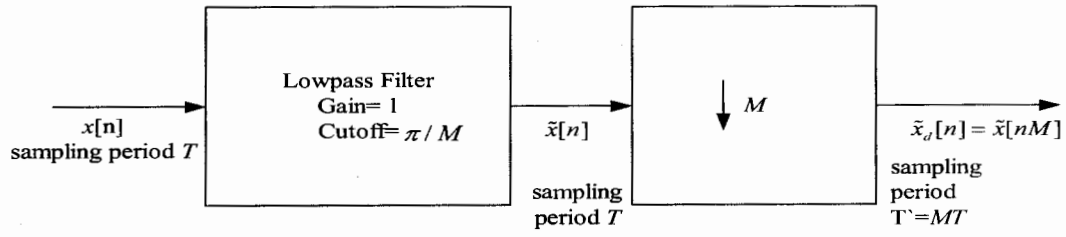


Figure B.2 General system for sampling rate reduction by M

B.2 Two decimations for echo-sounder application

The typical sonar signal in an echo-sounder application is described by two parameters; the sonar frequency and the pulse-length. Both parameters have to be known approximately to determine the sampling frequency [2]. A general description of the spectral properties of this particular sonar signal can be derived. The bandwidth is inversely proportional to the pulse-length, and generally an estimate of the width of the frequency band within the main lobe's –3dB points is

$$\Delta\omega = \frac{1}{\Delta T} \tag{B.1}$$

where ΔT is the pulse length of the transmit signal.

Since linear theory applies, the return signal will be constrained to the same frequency band. To widen the margin for error as well as for the estimation of the carrier frequency, assume that the return signal spectrum is contained in a frequency band twelve times as wide, i.e.

$$f_c - 6\Delta\omega \le \omega \le f_c + 6\Delta\omega \tag{B.2}$$

The upper and lower bandlimits are thus

$$f_L = f_c - \frac{6}{\Delta T} \quad \text{and} \quad f_H = f_c + \frac{6}{\Delta T} \tag{B.3}$$

Assuming that the full waveform signal is digitized with a sampling frequency f_s, which according to the Nyquist sampling theorem, will correctly capture the information contained in a frequency band

$$0 \le \omega < f_{ny}, \text{ with } f_{ny} = \frac{1}{2} f_s, \text{ the Nyquist frequency.}$$

In theory, two consecutive decimation operations are sufficient to extract the information-carrying band of an echo return.

The first decimation factor d_1 can be obtained from

$$d_1 = \frac{f_{ny}}{f_H} \tag{B.4}$$

where f_H is the upper bandlimit and f_{ny} is the Nyquist frequency.

Additionally the band $0 < \omega < f_L$ does not contain signal components carrying information. To discard this frequency band, a second decimation can be carried out.

The transformation to a negative frequency axis is justified by the way the anti-aliasing filters are implemented. On the negative frequency axis the second anti-alias filter is also a low-pass filter with cut-off $\hat{f}_L$ (note that $\hat{f}_L$ is in fact negative). The actual transformation to negative frequencies is a straightforward application of the Fourier frequency shift theorem.

$$\hat{f}_L = f_L - \hat{f}_{Ny} \tag{B.5}$$

where

$$\hat{f}_{ny} = \frac{f_{ny}}{d_1} \tag{B.6}$$

Therefore, second decimation factor d_2 is

$$d_2 = \frac{\hat{f}_{ny}}{\hat{f}_L} \tag{B.7}$$

For example, we used $f_s = 1.1\,\text{MHz}$ to sample an echo from a 200kHz sounder with a transmit pulse length of 0.16 ms. The upper bandlimit is $f_H = 237.5\,\text{kHz}$ from Eq. (B.3) and with $f_{Ny} = 550\,\text{kHz}$ the signal was oversampled about 2 times. For the first decimation, an anti-alias filter is applied as a proper low-pass filter with a cut-off frequency of 237.5 kHz, we may then resample this signal at a 2 times lower rate, thereby decimating the time series and reducing the number of samples needed to span a certain time by a factor of $d_1 = 2$.

After the first decimation was carried out, the Nyquist-frequency from Eq. (B.6) has become

$$\widehat{f}_{Ny} = \frac{550 \cdot 10^3}{2} = 275 \text{ kHz} \ ,$$

after shifting to negative frequencies from Eq. (B.5)

$$\widehat{f}_L = 162.5\text{kHz} - 275\text{kHz} = -112.5\text{kHz}$$

Therefore, second decimation factor $d_2 = \frac{-275}{-112.5} \approx 2$

Both times f_{ny} and $\widehat{f}_{ny}$ are the cut-off frequencies for the anti-alias filters applied in each decimation.

Finally, the overall decimation rate is $d_1 \cdot d_2 = 4$. The final sampling frequency is 1.1MHz/4 = 275 kHz for our research.

[1] Oppenheim, A.V., Schafer, R.W, Discrete-time signal processing, 2nd,Prientice-Hall, NJ, 1999, ch4.

[2] A. Rosenberger, Draft: High-speed data-acquisition core signal processing, QTC Technical Report, Sep, 2003.