ESTIMATION OF LARGE SPECTRAL FUNCTION AND ITS APPLICATION

A Dissertation

by

YUAN QU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jianhua Huang
Committee Members,	Mohsen Pourahmadi
	Michael Longnecker
	Ke-Li Xu
Head of Department,	Simon Sheather

August 2014

Major Subject: Statistics

Copyright 2014 Yuan Qu

ABSTRACT

It is of fundamental interest to routinely monitor waves and currents in the nearshore seas both scientifically and to the general public, because they play an important role in coastline erosion and they have a significant effect in the nearshore recreational activities. In this work, we show the way to estimate both wave height and wave direction with the data observed from a bottom-mounted, upward-looking Acoustic Doppler Current Profiler.

One of the most challenging works is to estimate the wave-number spectra using all gathered observations of receiving antennas. The frame of observed data is 100-dimensional time series with $T = 2399$. Due to the fact that there is only one realization of this multivariate time series, the conventional methods are either applicable for univariate time series or appropriate in low dimensional setting. In this work, we propose a new regularization estimator for wave-number spectral density with three merits: positive definite, smoothness and sparsity. This method can also be used to regularize any complex/real tensor in order to gain a resulting estimator with the above three merits. We describe and prove the convergence of our proposed algorithm, and compare our proposed estimator with the sample wave-number spectra and the other two regularization estimators: banding and extended tapering. The numerical results show that the estimation performance of our proposed approach is overwhelming better than other estimators. The proposed estimator and the extended tapering estimator are comparable in smoothness and positive definiteness. Unlike other estimators, our approach can produce a sparse estimator which would massively reduce the computation complexity for further study.

ACKNOWLEDGEMENTS

I would have never been able to complete my dissertation without the guidance of all my committee members, the help from my colleagues and friends, and the support from department and my family.

I would firstly express my sincere gratitude to my Ph.D. advisor Professor Jianhua Huang for his persevering encouragement and support of my postgraduate study and research as well as for his motivation and passion in the research area of machine learning. His guidance and training helped me gain enough knowledge and skills to produce my research and achieve this dissertation.

I would like to thank the rest of my thesis committee members: Professor Mohsen Pourahmadi, Professor Michael Longnecker, Professor Ke-Li, Xu and Professor Faming Liang (substitute for Professor Mohsen Pourahmadi) for their insightful comments and suggestions. My special thank goes to Dr. Ellen Toby for guiding me how to teach STAT 302 as a new instructor with full responsibilities in 2011.

Besides, I want to give many thanks of my fellow postgraduate students in our department: Dr. Saijuan Zhang, Dr. Ganggang Xu, Dr. Xiaolei Xun, Dr. Lin Zhang, Dr. Xuan Wang, Dr. Rubin Wei, Dr. Kun Xu, Dr. Ranye Sun, Dr. Yanqing Wang, Dr. Fang-Yu Lin, Shuo Feng, Jia You, Bohai Zhang, Senmao Liu, Nan Zhang, Ya Su and visiting scholars: Mengmeng Guo, Zhuqing Yu, for their kindness, help and friendship.

I would also like to thank the Department of Statistics and the Institute of Applied Mathematics and Computational Science for providing the support and equipment I needed to produce and finish my dissertation.

Last but not the least, I want to express my special thank to my parents, my

husband and my grandparents for their unconditional love, support and encouragement.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION: OCEAN WAVE THEORY AND MULTIVARIATE SPECTRUM ANALYSIS

1.1 Introduction to Ocean Wave Theory

It is of fundamental interest to routinely monitor waves and currents in the nearshore seas both scientifically and to the general public, because they play an important role in coastline erosion and they have a significant effect in the nearshore recreational activities. In this work, we show the way to estimate both wave height and wave direction with the data observed from a bottom-mounted, upward-looking Acoustic Doppler Current Profiler (ADCP). Before addressing ADCP data, we introduce the fundamentals of the ocean wave theory and the conventional spectral analysis.

In the actual analysis of wave data, it is almost impossible to evaluate the properties of ocean waves on a wave-by-wave basis in the time domain. For example, Figure (1.1) shows three portions of wave profiles recorded in shallow water, middle-level water and deep water respectively. However, if we regard the random waves as a stochastic process, it is possible to evaluate the statistical properties of waves in the frequency domain. This is the main reason that we apply spectral analysis to the modeling of ocean wave data.

In the view of stochastic process, waves are considered as a Gaussian process, that is, the probability distribution of wave displacement from the mean value is normally distributed. As examples, Figure (1.2) show the histograms of three portions of ocean wave in shallow water, middle-level water and deep water. As seen in these examples, these profiles of waves are well represented by the normal distributions. Therefore it is appropriate to predict statistical properties of ocean waves using conventional

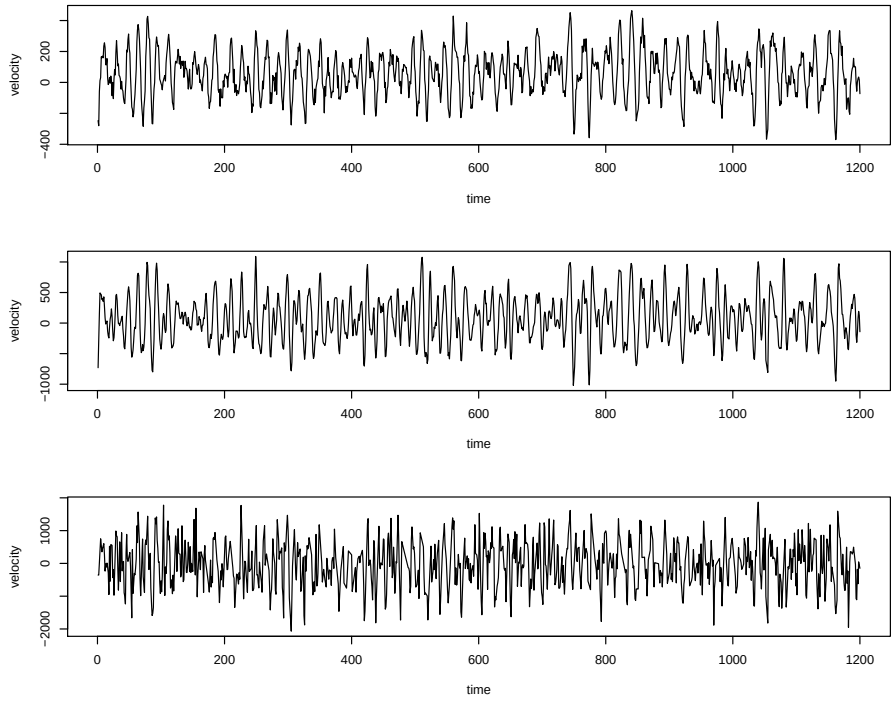

Figure 1.1: Wave profile of random seas for three different depth levels. Top plot: deep water; Middle plot: Middle-depth water; Bottom plot: shallow water.

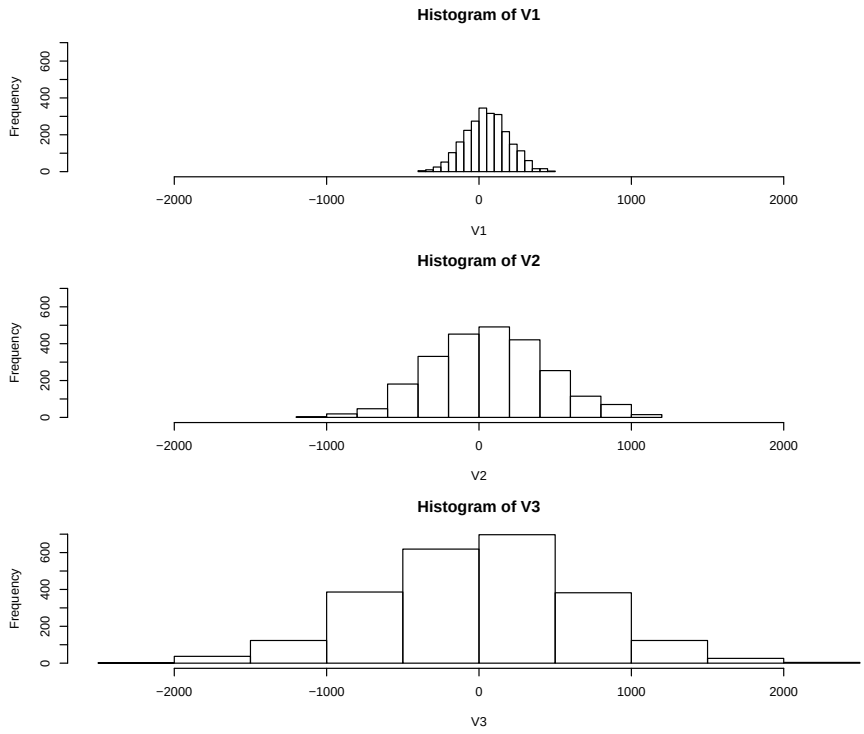

Figure 1.2: Comparing histograms of wave profile at different depth levels. Top plot: deep water; Middle plot: Middle-depth water; Bottom plot: shallow water.

theories of stochastic process.

1.1.1 Representing Ocean Waves as a Gaussian Random Process

Based on the central limit theorem in probability theory, the reason that we can consider ocean waves as a Gaussian process can be explained as follows.

Let η denote the profile of random wave at a fixed time t, which is a random variable defined in the sample space $(-\infty, \infty)$. We assume that η is the sum of a crowd of random components X_i, $i = 1, \cdots, n$, that is,

$$\eta = X_1 + X_2 + \ldots + X_n.$$

where the X_i, $i = 1, \cdots, n$ are statistically independent and identical distributed. Let the mean value of X_i be zero and the variance be σ^2. Because the X_i are statistically independent, the probability distribution of η has mean zero and variance $n\sigma^2$. Usually, n is very large.

By standardizing the random variable η, we can rewrite it as the new random variable Z given by

$$Z = \frac{\eta}{\sqrt{n}\sigma} = \sum_{j=1}^{n} \frac{X_j}{\sqrt{n}\sigma}.$$

Let $\phi_X(t)$ denote the characteristic function of X. Then referring to the properties of the characteristic function, the characteristic function of standardized random variable $\frac{X}{\sqrt{n}\sigma}$ can be given as $\phi_X(\frac{t}{\sqrt{n}\sigma})$. The characteristic function of Z can be represented as

$$\phi_Z(t) = \{\phi_X(\frac{t}{\sqrt{n}\sigma})\}^n.$$

On the other hand, we can expand the characteristic function as follows,

$$\phi_X(t) = 1 + itE[X] - \frac{t^2}{2}E[X^2] + \ldots$$

Since the mean and the second moment about zero of standardized variable are 0 and 1 respectively, $\phi_X(\frac{t}{\sqrt{n}\sigma})$ can be rewritten as

$$\phi_X(\frac{t}{\sqrt{n}\sigma}) = 1 - (\frac{t^2}{2n}) + o(\frac{t^2}{n}).$$

and we get that

$$\begin{aligned} \phi_Z^{(n)}(t) &= \{1 - (\tfrac{t^2}{2n}) + o(\tfrac{t^2}{n})\}^n \\ &= \{1 - \tfrac{1}{2n/t^2}\}^{(2n/t^2)\times\frac{t^2}{2}}. \end{aligned} \tag{1.1}$$

By letting n go to infinity, Equation (1.1) yields that

$$\phi_Z(t) = e^{(-1)\times\frac{t^2}{2}} = exp\{-\frac{t^2}{2}\}. \tag{1.2}$$

Equation 1.2 is the usual characteristic function of standardized normal distribution, and this implies that the random variable Z follows the normal distribution with mean zero and variance one. Therefore, η is normally distributed with mean zero and variance $n\sigma^2$. We finish proving that that the waves can be regarded as a Gaussian random process.

Pierson [24] shows an explanatory sketch indicating the above conclusion that ocean waves consist of an infinite number of sinusoidal waves having the same height with different frequencies and directional angles. Figure (1.3) shows Pierson's idea about this ocean waves' composition.

1.1.2 Energy Representation of Random Seas

In the view of spectral analysis, it is known that the wave spectral density can represent the potential and kinematic energies of waves or signal process. It also

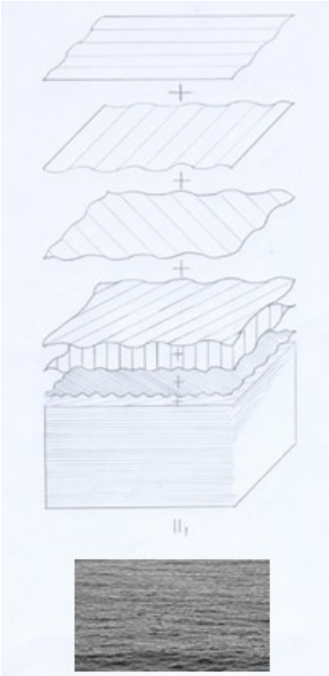

Figure 1.3: Structure of random seas.

plays a fundamental role in the evaluation of statistical properties of estimators.

In the conventional spectral analysis, the spectral density express the magnitude of energy as a function of frequency which is measured by cycles in this work. Because random waves in seas are not necessarily spreading in the same direction, the spectra representing the wave energy must consist of an argument denoting different spreading directions. Therefore, we consider a directional spectral density to represent ocean energy, denoted by $S(\omega, \theta)$, and the degree of sea severity is defined as the area under the directional spectral density function.

Let $\frac{1}{2}\rho g a_j^2$ denote the time average of wave energy at any frequency interval $\Delta\omega$ and any directional angle interval $\Delta\theta$ of the random sea, where a_j is a positive

random variable, we can give the following energy representation,

$$S(\omega,\theta)\Delta\omega\Delta\theta = \frac{1}{2}a_j^2. \tag{1.3}$$

Then the time average of the total energy of waves coming from various directions and including all frequencies is given by

$$\int_{-\pi}^{\pi}\int_{0}^{\infty} S(\omega,\theta)d\omega d\theta = \sum_{\Delta\omega}\sum_{\Delta\theta}\frac{1}{2}a_j^2. \tag{1.4}$$

This is the basis of the stochastic description of waves in seas.

1.1.3 Mathematical Presentation of Random Waves

In this section, we describe a mathematical representation of ocean wave profile based on Equation (1.3).

According to the principle of ocean wave theory, a 2-dimension wave with X-axis propagation can be expressed as a harmonic function, that is,

$$\zeta = a \times \sin(kx - \omega t).$$

where $k = \frac{2\pi}{\lambda}$ and $\omega = \frac{2\pi}{T}$. Now we consider a wave profile in random seas using the coordinate system (X, Y, Z) fixed in space, and let θ denote the angle taken in a counterclockwise direction with respect to the X-axis. The velocity of wave propagation may be written as

$$c = \frac{gT}{2\pi} \tag{1.5}$$

$$c = \frac{\omega}{k} \tag{1.6}$$

By summing Equation (1.5) and Equation (1.6), we get

$$k = \frac{2\pi\omega}{gT} = \frac{\omega^2}{g}.$$

The propagation distance, denoted by d, can be written as

$$d = x\cos\theta + y\sin\theta$$

We can give the profile of simple harmonic wave traveling at an angle θ with the X-axis as

$$\eta(x, y, t) = a\ cos\{\frac{\omega^2}{g}(xcos\theta + ysin\theta) - \omega t + \epsilon\}. \tag{1.7}$$

where a denotes the amplitude of this simple harmonic wave, ω denotes the frequency, ϵ denotes the phase.

Now we consider the profile of incidental ocean wave at time t with an infinite number of sinusoidal wave components consisting of different amplitudes a_j, coming from divergent directions θ_j, with various frequencies ω_j. a_j, θ_j and ω_j are random variables with the range $0 < a_j < \infty$, $-\pi < \theta_j < \pi$ and $0 < \omega_j < \infty$, respectively. The phase ϵ is also a random variable distributed uniformly over the range $-\pi < \epsilon < \pi$, and its magnitude depends on the frequency and angle. Now we can write the profile of this incidental ocean wave as

$$\eta(x, y, t) = \sum_j a_j \cos\{\frac{\omega_j^2}{g}(x\cos\theta_j + y\sin\theta_j) - \omega_j t + \epsilon_j\}.$$

The amplitude a_j satisfies the condition given in Equation(1.3) for any frequency $\Delta\omega$ and directional interval $\Delta\theta$.

When the number of discrete component waves given in Eq. (1.3) goes extremely

large and the frequency as well as directional angle components become extremely small, the summation can be replaced as a double integration with respect to ω and θ. Referring to [28] and [23], we can rewrite the profile of the random sea surface by the following stochastic integral representation

$$\begin{aligned}\eta(x,y,t) = \int_{-\pi}^{\pi}\int_{0}^{\infty} \quad & \cos\{\frac{\omega^2}{g}(x\cos\theta + y\sin\theta) - \omega t + \epsilon(\omega,\theta)\} \\ & \times\sqrt{2S(\omega,\theta)}d\omega d\theta. \end{aligned} \tag{1.8}$$

For the ease of presentation, we can express the wave profile given in Equation (1.8) in the form of a vector with new definition $k = \omega^2/g$ and the followings,

$$\begin{aligned} \boldsymbol{r} &= x\boldsymbol{i} + y\mathbf{j}; \\ \boldsymbol{k} &= k\cos\theta\boldsymbol{i} + k\sin\theta\mathbf{j} \\ &= k_x\boldsymbol{i} + k_y\mathbf{j} \end{aligned}$$

Then Equation (1.8) can be rewritten as

$$\begin{aligned} \eta(\boldsymbol{r},t) &= \int_{-\infty}^{\infty}\int_{-\pi}^{\pi}\cos(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t + \epsilon)dA(\omega,\theta) \\ &= Re\int_{-\infty}^{\infty}\int_{-\pi}^{\pi}e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t+\epsilon)}dA(\omega,\theta) \\ &= \int_{-\infty}^{\infty}\int_{-\pi}^{\pi}e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t+\epsilon)}dA(\omega,\theta) \end{aligned} \tag{1.9}$$

Where

$$E[dA(\boldsymbol{k},\omega dA^*(\boldsymbol{k},\omega)] = S(\boldsymbol{k},\omega)d\boldsymbol{k}d\omega \tag{1.10}$$

$S(\boldsymbol{k},\omega)$ is known as the wave-number frequency spectra.

1.2 Multivariate Spectral Analysis

As mentioned in Section 1.1, spectral analysis is natural and widely used to analyze wave profiles. In this work, all topics are discussed in frequency domain. Thereby in this section, we review the fundamental spectral analysis in Section 1.2.1 and the multivariate spectral analysis in Section 1.2.2 for ocean waves.

1.2.1 Fundamental Spectral Analysis

In this section, we describe the way to measure the wave displacement at a certain location $\eta(t)$. The auto-correlation function, denoted by $R(\tau)$, is defined as follows,

$$R(\tau) = E[\eta(t)\eta(t+\tau)] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \eta(t)\eta(t+\tau)dt \tag{1.11}$$

We know that $R(\tau)$ is maximum at $\tau = 0$ and $R(0)$ is equal to the second moment about zero of the wave data $E[\eta^2(t)]$. If we assume the mean value of $\eta(t)$ is zero, $R(0)$ represents the variance of the wave record, that is $R(0)$ represents the time average of wave energy $\bar{P}$.

Let us express the average energy $\bar{P}$ in terms of wave frequency ω in radians per second. This can be done by applying the following Parseval theorem,

$$\int_{-\infty}^{\infty} \{\eta(t)\}^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\check{\eta}(\omega)|^2 d\omega \tag{1.12}$$

where $\check{\eta}(\omega)$ is the Fourier transform of $\eta(t)$. That is,

$$\begin{aligned} \check{\eta}(\omega) &= \int_{-\infty}^{\infty} \eta(t) e^{-i\omega t} dt \\ \eta(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \check{\eta}(\omega) e^{i\omega t} d\omega \end{aligned} \tag{1.13}$$

We also call $\eta(t)$ and $\check{\eta}(\omega)$ as the one-dimensional Fourier transform pair.

By applying Equation (1.12), the average wave energy $\bar{P}$ may be written as

$$\bar{P} = \lim_{T\to\infty} \frac{1}{4\pi T} \int_{-\infty}^{\infty} |\check{\eta}(\omega)|^2 d\omega \tag{1.14}$$

Then the spectral density function of random waves $\eta(t)$ is defined as follows,

$$S(\omega) = \lim_{T\to\infty} \frac{1}{2\pi T} |\check{\eta}(\omega)|^2 \tag{1.15}$$

Since the spectral density function is an even function, the time average of wave energy $\bar{P}$ can be expressed from Eqs. (1.14) and (1.15) as

$$\bar{P} = \frac{1}{2} \int_{-\infty}^{\infty} S(\omega) = \int_{0}^{\infty} S(\omega) d\omega \tag{1.16}$$

The above equation shows that the area under the spectral density function represents the average energy of random waves with respect to time. Furthermore, from the properties of the auto-correlation function, the area under the spectrum is also equal to the variance of waves $\eta(t)$.

Now let us consider two wave records $\eta_1(t)$ and $\eta_2(t)$. The cross-correlation function denoted by $R_{12}(\tau)$, is defined as

$$R_{12}(\tau) = E[\eta_1(t)\eta_2(t+\tau)] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \eta_1(t)\eta_2(t+\tau) dt \tag{1.17}$$

The cross-correlation function $R_{12}(\tau)$ is not necessarily maximum at $\tau = 0$, and unlike the auto-correlation function, $R_{12}(0)$ does not have any significant meaning.

The cross-spectral density function of two wave records $\eta_1(t)$ and $\eta_2(t)$ is defined

as follows,

$$S_{12}(\omega) = \lim_{T\to\infty} \frac{1}{2\pi T} \check{\eta}_1^*(\omega)\check{\eta}_2(\omega) \tag{1.18}$$

where $\check{\eta}_1(\omega)$ and $\check{\eta}_2(\omega)$ are conjugate functions of $\eta_1(t)$ and $\eta_2(t)$, respectively.

1.2.2 Wave-Number Spectrum

Our data was collected in different spatial locations (x, y, z). The detailed data frame will be describe in Chapter 4. To deal with this multivariate time series (wave displacement) associated with spatial locations in a stationary, homogeneous wave field, we must generate the traditional spectrum analysis to multivariate case. To be specific, we need to consider the three-dimensional Fourier transform. By introducing the coordinate vector $\boldsymbol{r}$ and wave-number vector $\boldsymbol{k}$, and letting the displacement of the sea surface be $\eta(\boldsymbol{r}, t)$ we write the three-dimensional Fourier transform pair in the following vector form,

$$\check{\eta}(\boldsymbol{k}, \omega) = \int_t \int_{\boldsymbol{r}} \eta(\boldsymbol{r}, t)\ e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}\ d\boldsymbol{r}\ dt \tag{1.19}$$

$$\eta(\boldsymbol{r}, t) = \frac{1}{(2\pi)^3} \int_\omega \int_{\boldsymbol{k}} \check{\eta}(\boldsymbol{k}, \omega)\ e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}\ d\boldsymbol{k}\ d\omega \tag{1.20}$$

where

$$\boldsymbol{r} = x\boldsymbol{i} + y\mathbf{j} \qquad d\boldsymbol{r} = dxdy$$

$$\boldsymbol{k} = k_x\boldsymbol{i} + k_y\mathbf{j} \qquad d\boldsymbol{k} = dk_xdk_y$$

Then the auto-correlation is defined as follows,

$$R(\boldsymbol{\rho}, \tau) = E[\eta(\boldsymbol{r}, t)\eta(\boldsymbol{r} + \boldsymbol{\rho}, t + \tau)]. \tag{1.21}$$

Before we could discuss in detail the extension of the Wiener-Khintchine theorem, let us first give the definition of the wave-number frequency spectral density function as follows,

$$S(\boldsymbol{k},\omega) = \lim_{T,R_1,R_2\to\infty} \frac{1}{2T\cdot 2R_1\cdot 2R_2\pi}|\check{\eta}(\boldsymbol{k},\omega)|^2 \tag{1.22}$$

The Wiener-Khintchine theorem plays an extremely significant role in the stochastic analysis of random waves. It presents the relationship between the auto-correlation function defined in the time domain and and the spectral density function defined in the frequency domain. The theorem states that for a weekly steady-state random wave, the autocorrelation function $R(\tau)$ and spectral density function $S(\omega)$ are a Fourier transform pair. That is,

$$\begin{aligned} S(\omega) &= \frac{1}{\pi}\int_{-\infty}^{\infty} R(\tau)e^{-i\omega\tau}d\tau \\ R(\tau) &= \frac{1}{2}\int_{-\infty}^{\infty} S(\omega)e^{i\omega\tau}d\omega \end{aligned} \tag{1.23}$$

Now we consider the relationship between the auto-correlation function defined in the time-space domain and the wave-number frequency spectral density function defined in the frequency-direction domain. Following the Wiener-Khintchine theorem, the similar relationships to Eq. (1.23) between the wave spectrum $S(\boldsymbol{k},\omega)$ and auto-correlation function $R(\boldsymbol{\rho},\tau)$ can be written as

$$\begin{aligned} S(\boldsymbol{k},\omega) &= \frac{1}{\pi}\int_{\tau}\int_{\boldsymbol{\rho}} R(\boldsymbol{\rho},\tau)e^{i(\boldsymbol{k}\cdot\boldsymbol{\rho}-\omega\tau)}d\boldsymbol{\rho}d\tau \\ R(\boldsymbol{\rho},\tau) &= \frac{1}{2(2\pi)^2}\int_{\omega}\int_{\boldsymbol{k}} S(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{\rho}-\omega\tau)}d\boldsymbol{k}d\omega \end{aligned} \tag{1.24}$$

We derive the wave-number frequency spectral density function in details in the

following through the Fourier transform of the auto-correlation function.

$$
\begin{aligned}
& \int_{\tau}\int_{\boldsymbol{\rho}} R(\boldsymbol{\rho},\tau)e^{i(\boldsymbol{k}\cdot\boldsymbol{\rho}-\omega\tau)}d\boldsymbol{\rho}d\tau \\
= & \int_{\tau}\int_{\boldsymbol{\rho}} E[\eta(\boldsymbol{r},t)\eta(\boldsymbol{r}+\boldsymbol{\rho},t+\tau)]e^{i(\boldsymbol{k}\cdot\boldsymbol{\rho}-\omega\tau)}d\boldsymbol{\rho}d\tau \\
= & \int_{\tau}\int_{\boldsymbol{\rho}} \{\lim_{T,R_1,R-2\to\infty}\frac{1}{8TR_1R_2}\int_{-T}^{T}\int_{-R_1}^{R_1}\int_{-R_2}^{R_2}\eta(\boldsymbol{r},t)\eta(\boldsymbol{r}+\boldsymbol{\rho},t+\tau)dtdr_1dr_2\} \\
& \times e^{i[\boldsymbol{k}\cdot(\boldsymbol{r}+\boldsymbol{\rho})]}\times e^{-i(\boldsymbol{k}\cdot\boldsymbol{r})}\times e^{-i\omega(t+\tau)}\times e^{i\omega\tau}d\boldsymbol{\rho}d\tau
\end{aligned}
$$

By changing the integration and limit, the above equation can be rewritten as

$$
\begin{aligned}
& \int_{\tau}\int_{\boldsymbol{\rho}} R(\boldsymbol{\rho},\tau)e^{i(\boldsymbol{k}\cdot\boldsymbol{\rho}-\omega\tau)}d\boldsymbol{\rho}d\tau \\
= & \lim_{T,R_1,R-2\to\infty}\frac{1}{8TR_1R_2}\int_{-T}^{T}\int_{-R_1}^{R_1}\int_{-R_2}^{R_2}\{\int_{\tau}\int_{\boldsymbol{\rho}}\eta(\boldsymbol{r}+\boldsymbol{\rho},t+\tau) \\
& e^{i[\boldsymbol{k}\cdot(\boldsymbol{r}+\boldsymbol{\rho})-\omega(t+\tau)]}d(\boldsymbol{r}+\boldsymbol{\rho})d(t+\tau)\}\times\eta(\boldsymbol{r},t)e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}dtdr_1dr_2 \\
= & \lim_{T,R_1,R-2\to\infty}\frac{1}{8TR_1R_2}\check{\eta}^{*}(\boldsymbol{k},\omega)\times\int_{-T}^{T}\int_{-R_1}^{R_1}\int_{-R_2}^{R_2}\eta(\boldsymbol{r},t)e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}dtdr_1dr_2 \\
= & \lim_{T,R_1,R-2\to\infty}\frac{1}{8TR_1R_2}\check{\eta}^{*}(\boldsymbol{k},\omega)\check{\eta}(\boldsymbol{k},\omega) \\
= & \lim_{T,R_1,R_2\to\infty}\frac{1}{2T\cdot 2R_1\cdot 2R_2}|\check{\eta}(\boldsymbol{k},\omega)|^2 \\
= & \pi S(\boldsymbol{k},\omega) \qquad (1.25)
\end{aligned}
$$

The integration of the wave-number frequency spectral density defined in Equation (1.22) yields the variance of random waves.

By integrating $S(\boldsymbol{k},\omega)$ with respect to frequency, the wave-number spectra can be defined as

$$S(\boldsymbol{k}) = \int_{-\infty}^{\infty} S(\boldsymbol{k},\omega)d\omega. \qquad (1.26)$$

By letting $\tau = 0$ and by integrating the auto-correlation function given in Equation

(1.24), we get

$$R(\boldsymbol{\rho}) = \frac{1}{(2\pi)^2} \int_{\boldsymbol{k}} S(\boldsymbol{k}) e^{-i\boldsymbol{k}\cdot\boldsymbol{\rho}} d\boldsymbol{k} \tag{1.27}$$

and its inverse yields

$$S(\boldsymbol{k}) = \int_{\boldsymbol{\rho}} R(\boldsymbol{\rho}) e^{i\boldsymbol{k}\cdot\boldsymbol{\rho}} d\boldsymbol{\rho} \tag{1.28}$$

On the other hand, if we integrate $S(\boldsymbol{k}, \omega)$ with respect to wave-number $\boldsymbol{k}$, we have the frequency spectral function that

$$S(\omega) = \int_{\boldsymbol{k}} S(\boldsymbol{k}, \omega) d\boldsymbol{k}. \tag{1.29}$$

and by letting $\boldsymbol{\rho} = 0$ and by integrating with respect to wave-number $\boldsymbol{k}$ in Equation (1.24), we have

$$R(\tau) = \frac{1}{2} \int_{\omega} S(\omega) e^{i\omega\tau} d\omega \tag{1.30}$$

and its inverse yields

$$S(\omega) = \frac{1}{\pi} \int_{\tau} R(\tau) e^{-i\omega\tau} d\tau \tag{1.31}$$

1.3 Challenging Problems

Now we are ready to introduce and discuss our project and associating statistical problem. Using ocean wave theorem and multivariate spectral analysis, we can derive the following forward model

$$C(\omega) = \int_{-\pi}^{+\pi} \mathbf{H}(\omega, \theta) D(\omega, \theta) \mathbf{H}^*(\omega, \theta) d\theta \tag{1.32}$$

where $*$denotes the complex-conjugate transpose, and $H(\omega, \theta)$ is the transfer function, defined by Isobe [19], combining various wave kinematric quantities such as pressure, velocity, acceleration and so on.

On the left-hand side of the above forward model, $C(\omega)$ denote the estimator for multivariate spectral density. The easiest and commonly used estimator is sample spectral matrix analogous to sample covariance matrix in time domain. The usual sample spectral matrix is optimal in the classical setting with large samples and fixed low dimensions. However, it performs very poorly in the high dimensional setting.

Our first goal is to provide a good estimator for multivariate spectral density no matter what the dimension is. For this good estimator, we should give a guarantee of good properties such as positive definite, smoothness as well as sparsity. We will discuss more details about the proposed estimator and corresponding algorithm in Chapter 2 and describe the theoretical analysis of the algorithm and numerical examples in Chapter 3. In Chapter 3, we compare the pros and cons of three spectral estimators, and their practical performances are also illustrated by numerical examples.

Another challenging task is to solve the forward model 1.32 for the directional spectral density $D(\omega, \theta)$. This is an inverse problem and many methods can be used to address the problem, for example, the maximum likelihood (ML), [10], [9] and [12], the extended maximum likelihood (EML), the maximum entropy method (ME), [5], [16] and [22], and Bayesian method [15]. In Chapter 4, we apply the EML method, see , [19] and [17], to solve the forward model 1.32 with our proposed estimator for multivariate spectrum.

2. NEW APPROACH TO ESTIMATE LARGE SPECTRUM

The goal of the study mentioned in Chapter 1.1 is to establish an both efficient and accurate estimator for the frequency-directional spectral distribution. Referring to the forward model (4.33), the first subproblem is to estimate the wave-number spectra effectively only with one realization of multivariate time series.

In this chapter, we show a new method to estimate wave-number spectral function for multivariate time series. Before we describe the detailed method and algorithm, we first give a brief introduction to notations and concepts of tensorial algebra in Section 2.1 which will be used throughout Chapter 2 and Chapter 3 for ease of presentation. In Section 2.2, we state and prove a necessary and sufficient condition for a complex matrix to be positive definite. The necessary and sufficient condition constrains the positive definition on the corresponding real matrices, the real matrix and imaginary matrix of the original complex matrix. In this way, we can focus on solving a real-valued estimation problem instead of a complex-valued one. In Section 2.3, we describe the motivation of the proposed method and discuss several intermediate estimators with good properties one at a time for large spectra. These intermediate estimators inspire our proposed resulting estimator with simultaneous sparsity, positive definite and smoothness. In Section 2.5, we state in detail the proposed method, derive the closed form of each step for one iteration. We also summarize the proposed estimation method as Algorithm 1 at the end of this section. To illustrate the performance of the Algorithm 1, we present a simulated example along with comparisons with three existing methods in Section 2.6.

2.1 Notation and Preliminary

Throughout Chapter 2 and Chapter 3, scalars are denoted by the lowercase letters, e.g. a, vectors are denoted by the boldface lowercase letters, e.g. $\boldsymbol{a}$ and matrices are denoted by the boldface capital letters, e.g. $\boldsymbol{A}$. Tensors are usually used to express multi-dimensional arrays The number of dimensions of a tensor is also known as the order of this tensor. In Chapter 2, we mainly address 3-order tensors. Therefore tensors are particularly referred to 3-order tensors in this work which are denoted by the calligraphic letters, e.g. $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$.

The nth element of a vector $\boldsymbol{a}$ is denoted by a_n, and the crossing element of a matrix $\boldsymbol{A}$ at the mth row and nth column is denoted by a_{mn} or for ease of distinction in proof denoted by a_{mn} or $(\boldsymbol{A})_{mn}$, $m, n = 1, \cdots, N$. In a similar way, the element of a tensor $\mathcal{A}$ throughout this work is denoted by a_{mnt} or $(\mathcal{A})_{mnt}$, $m, n = 1, \cdots, N$, $t = 1, \cdots, T$. Let $\boldsymbol{A}_{\cdot\cdot t}$ or $(\mathcal{A})_{\cdot\cdot t}$ denote the matrix abstracted from tensor $\mathcal{A}$ with fixed third index t and let $\boldsymbol{a}_{mn\cdot}$ or $(\mathcal{A})_{mm\cdot}$ denote the vector abstracted from tensor $\mathcal{A}$ with fixed first and second indices m and n.

Let $\|\cdot\|_F$ be the general Frobenius norm and $|\cdot|_1$ be the element-wise l_1-norm of all off-diagonal elements. For matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|_F$ and $|\boldsymbol{A}|_1$ are defined as,

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{m,n} (a_{mn})^2} \ .$$

and

$$|\boldsymbol{A}|_1 = \sum_{m \neq n} |a_{mn}| \ .$$

We can directly generalize the definitions of Frobenius norm and element-wise l_1-

norm for matrix and apply them to tensor $\mathcal{A}$ as follows, respectively,,

$$\|\mathcal{A}\|_F = \sqrt{\sum_t \left[\sum_{m,n} (a_{mnt})^2 \right]} \ .$$

and

$$|\mathcal{A}|_1 = \sum_t (\sum_{m \neq n} |a_{mnt}|) \ .$$

Let $\langle \cdot, \cdot \rangle$ denote the inner product. The inner product of two matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, and two tensors, $\mathcal{A}$ and $\mathcal{B}$, are respectively given by

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{m,n} a_{mn} b_{mn}$$

and

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{m,n} \left(\sum_t a_{mnt} b_{mnt} \right)$$

In Section 2.3, our goal is to simultaneously update three $2N \times 2N \times T$ tensors $\mathcal{A}$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$ in each iteration. For the simplicity of presentation, we define the following $6N \times 2N \times T$ tensor denoted by $\mathcal{U}$ by stacking up $\mathcal{A}$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$.

$$\mathcal{U} = \begin{pmatrix} \mathcal{A} \\ \boldsymbol{\Theta} \\ \boldsymbol{\Pi} \end{pmatrix} . \tag{2.1}$$

For the convenience of derivation in Chapter 3, we have to define $\mathcal{G}$-norm as $\|\cdot\|_{\mathcal{G}}$ and the corresponding inner product as $\langle \cdot, \cdot \rangle_{\mathcal{G}}$. Let $\mathcal{G}$ be a $6N \times 6N \times T$ tensor. For

each fixed ω, $(\mathcal{G})_{\cdot\cdot\omega}$ is defined as,

$$(\mathcal{G})_{\cdot\cdot\omega} = \begin{pmatrix} (\frac{1}{\mu_1} + \frac{1}{\mu_2})\boldsymbol{I}_{2N\times 2N} & 0 & 0 \\ 0 & \mu_1 \boldsymbol{I}_{2N\times 2N} & 0 \\ 0 & 0 & \mu_2 \boldsymbol{I}_{2N\times 2N} \end{pmatrix}, \quad \forall \omega = 1/T, \cdots, T/T. \quad (2.2)$$

The product of $\mathcal{G}$ and $\mathcal{U}$ is defined as,

$$(\mathcal{G}\mathcal{U})_{\cdot\cdot\omega} = (\boldsymbol{G}_{\cdot\cdot\omega})(\boldsymbol{U}_{\cdot\cdot\omega}),$$

for each fixed ω. Where $(\boldsymbol{G}_{\cdot\cdot\omega})(\boldsymbol{U}_{\cdot\cdot\omega})$ is the general product of two matrices with the form of

$$(\boldsymbol{G}_{\cdot\cdot\omega})(\boldsymbol{U}_{\cdot\cdot\omega}) = \begin{pmatrix} (\frac{1}{\mu_1} + \frac{1}{\mu_2})\boldsymbol{A}_{\cdot\cdot\omega} \\ \mu_1 \boldsymbol{\Theta}_{\cdot\cdot\omega} \\ \mu_2 \boldsymbol{\Pi}_{\cdot\cdot\omega} \end{pmatrix}, \quad \forall \omega = 1/T, \cdots, T/T.$$

Now we can define the $\mathcal{G}$-norm of tensor $\mathcal{U}$ as

$$\|\mathcal{U}\|_G = \sqrt{\langle \mathcal{U}, \mathcal{G}\mathcal{U} \rangle}\ . \quad (2.3)$$

and the corresponding inner product of $\mathcal{U}$ and $\mathcal{V}$ as

$$\langle \mathcal{U}, \mathcal{V} \rangle_{\mathcal{G}} = \langle \mathcal{U}, \mathcal{G}\mathcal{V} \rangle\ . \quad (2.4)$$

2.2 A Necessary and Sufficient Condition for a Complex Matrix to Be Positive Definite

The positive-definite property is crucial and necessary for spectral estimation both methodologically and practically. On the one hand, this is analogous to the

requirement that the covariance matrix must be positive semidefinite. On the other hand, the calculation of the inverse matrix of the estimated spectra requires positive definiteness. One example is to solve the forward model

$$C(\omega) = \int_{-\pi}^{+\pi} \mathbf{H}(\omega,\theta) D(\omega,\theta) \mathbf{H}^*(\omega,\theta) d\theta,$$

for estimating the directional spectral density $D(\omega,\theta)$ in Chapter 4. We define several necessary notations and list some useful facts here only for the presentation in Section 2.2.

Let $\boldsymbol{C}^*$ denote the conjugate transpose of $\boldsymbol{C}$. The Hermitian property can be written as $\boldsymbol{C} = \boldsymbol{C}^*$, that is, respectively, its real part $\boldsymbol{A}$ and imaginary part $\boldsymbol{B}$ satisfy

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{A}^T, \\ \boldsymbol{B} &= -\boldsymbol{B}^T. \end{aligned} \tag{2.5}$$

Let $\boldsymbol{D}$ denote the real matrix associated with $\boldsymbol{C}$, which has the form of

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{A} & -\boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{pmatrix}. \tag{2.6}$$

A necessary and sufficient condition for a complex matrix to be positive definite is given by the following Theorem 2.2.1.

Theorem 2.2.1. *If an $N \times N$ complex matrix $\boldsymbol{C}$ is a Hermitian matrix, then $\boldsymbol{C}$ is positive definite if and only if the $2N \times 2N$ real matrix $\boldsymbol{D}$ associated with $\boldsymbol{C}$ is positive definite.*

Proof. Referring to the definition of positive definite for a complex matrix, if $\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}$

is real and positive for all non-zero complex vectors $\boldsymbol{x}$, Hermitian matrix $\boldsymbol{C}$ is said to be positive definite.

We rewrite the complex vector $\boldsymbol{x}$ as $\boldsymbol{x} = \boldsymbol{x}_1 + i\boldsymbol{x}_2$, then $\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}$ can be given as its real part and its imaginary part as follows,

$$\begin{aligned} Re(\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}) &= \boldsymbol{x}_1^T\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{x}_1^T\boldsymbol{B}\boldsymbol{x}_2 + \boldsymbol{x}_2^T\boldsymbol{B}\boldsymbol{x}_1 + \boldsymbol{x}_2^T\boldsymbol{A}\boldsymbol{x}_2 \quad (2.7) \\ Im(\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}) &= \boldsymbol{x}_1^T\boldsymbol{B}\boldsymbol{x}_1 + \boldsymbol{x}_1^T\boldsymbol{A}\boldsymbol{x}_2 - \boldsymbol{x}_2^T\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{x}_2^T\boldsymbol{B}\boldsymbol{x}_2 \quad (2.8) \end{aligned}$$

Note that the fact $\boldsymbol{x}_1^T\boldsymbol{A}\boldsymbol{x}_2$ and $\boldsymbol{x}_2^T\boldsymbol{A}\boldsymbol{x}_1$ are scalars, we can get

$$\boldsymbol{x}_1^T\boldsymbol{A}\boldsymbol{x}_2 = \boldsymbol{x}_2^T\boldsymbol{A}\boldsymbol{x}_1. \quad (2.9)$$

Similarly, we have

$$\boldsymbol{x}_1^T\boldsymbol{B}\boldsymbol{x}_2 = \boldsymbol{x}_2^T\boldsymbol{B}\boldsymbol{x}_1. \quad (2.10)$$

The Hamitian property (2.5) leads to

$$\begin{aligned} \boldsymbol{x}_1^T\boldsymbol{B}\boldsymbol{x}_1 &= 0, \\ \boldsymbol{x}_2^T\boldsymbol{B}\boldsymbol{x}_2 &= 0 \quad (2.11) \end{aligned}$$

Summing Equations (2.8), (2.9) and (2.11) gives

$$Im(\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}) = 0$$

Therefore to sum up, the condition of the complex matrix $\boldsymbol{C}$ to be positive definite

is

$$Re(\boldsymbol{x}^* C \boldsymbol{x}) > 0. \tag{2.12}$$

On the other hand, let $\boldsymbol{y}$ denote the real counterpart of $\boldsymbol{x}$ by stacking the real part $\boldsymbol{x}_1$ and the imaginary part into a $2N \times 1$ vector. The condition $\boldsymbol{D}$ to be positive definite is that $\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y}$ is positive for all non-zero vectors $\boldsymbol{y}$, that is

$$\begin{aligned} \boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y} &= (\boldsymbol{x}_1^T \quad \boldsymbol{x}_2^T) \begin{pmatrix} A & -B \\ B & A \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \\ &= \boldsymbol{x}_1^T A \boldsymbol{x}_1 - \boldsymbol{x}_1^T B \boldsymbol{x}_2 + \boldsymbol{x}_2^T B \boldsymbol{x}_1 + \boldsymbol{x}_2^T A \boldsymbol{x}_2 \\ &> 0. \end{aligned} \tag{2.13}$$

which is equivalent to the condition of $\boldsymbol{C}$ to be positive by comparing Equation (2.12) and Equation (2.13). □

Using the necessary and sufficient condition of positive definition shown as Theorem 2.2.1, the problem to obtain a positive definite estimator for an $N \times N$ complex matrix $\boldsymbol{C}$ is converted to the problem to establish a positive-definite estimator for the $2N \times 2N$ real matrix $\boldsymbol{D}$ associated with $\boldsymbol{C}$.

2.3 Proposed Estimation of Large Spectral Matrix

This section is organized as follows. In Section 2.3.1, there is a brief list of literature review which also provides the background of the regularization techniques used in the estimation of covariance matrices. In this section, we also show the motivation of applying regularization techniques to estimate spectral density function. In Section 2.3.2, Section 2.3.3 and Section 2.3.4, three subproblems and the corresponding intermediate estimators are discussed in details. These intermediate estimators

directly inspire our proposed estimator which is shown in the following Section 2.4.

2.3.1 Motivation

In low dimensional setting with large sample size, the sample spectral estimator is usually optimal. However in high dimensional setting, the sample spectra performs very poorly because there usually exist many parameters to be estimated which number are much more than the number of realizations of multivariate time series. Another difficulty is that the existing optimal methods in the univariate time series much less implement to the high dimensional setting, which we have discussed in detail in Chapter 1.

In the recent literature, we note that many regularization techniques have been introduced to improve the estimation of autocovariance function, such as banding, e.g. [30] and [2], tapering, e.g. [14] and [7], and thresholding, e.g. [2], [20], [27] and [31].

On the other hand, based on fundamental spectral analysis, we have the fact that if the autocovariance function, $\gamma(l)$, of a stationary process satisfies

$$\sum_{l=-\infty}^{\infty} |\gamma(l)| < \infty.$$

then it has the following representation

$$\gamma(l) = \int_{-1/2}^{1/2} e^{2\pi i \omega h} f(\omega) d\omega, \quad h = 0, \pm 1, \pm 2, \cdots . \tag{2.14}$$

$f(\omega)$ is known as the spectral density function. According to probability theory and Equation (2.14), $\gamma(l)$ is the characteristic function of $f(\omega)$, therefore they contain the same information. To be specific, the autocovariance function expresses information in terms of lags, whereas the spectral density function expresses the same information

in terms of cycles. Therefore, it is natural to apply regularization techniques to the analogous estimation of the spectral density function.

Banding and tapering are very useful when the variables have a natural ordering and the off-diagonal elements decay to zero as they move away from the diagonal of the target matrix. Thresholding is usually used to produce consistent covariance matrix estimators when the true covariance matrix is bandable, see [1] and [8]. In this sense, thresholding is more robust than banding and tapering for most of real applications.

2.3.2 Thresholding Spectral Estimator

Thresholding technique includes a number of commonly used shrinkage procedures. The general thresholding covariance matrix estimator was proposed by Rothman, Levina and Zhu [27], denoted by $h_\lambda(z)$, which covers the hard thresholding $h_\lambda(z) = z\boldsymbol{I}_{\{|z|>\lambda\}}$, the soft thresholding $h_\lambda(z) = sign(z)(|z| - \lambda)_+$, the smoothly clipped absolute deviation thresholding [13] and the adaptive lasso thresholding [32]. The existing theoretical and empirical results show no clear favoritism to a particular thresholding rule, for example, [2], [1], [20] [27], [6], [7] and [8]. In this work, we use the soft-thresholding function because it can be formulated as the solution of a convex optimization problem.

Let $\bar{\mathcal{C}}$ denote the sample spectrum, and $\bar{\mathcal{A}} = \{a_{mnt}\}_{1\leq m,n\leq,1\leq t\leq T}$ denote the real-valued tensor of associated with $\bar{\mathcal{C}}$. Analogous to the soft-thresholding covariance matrix estimator, the soft-thresholding spectral estimator is given by

$$\mathcal{A}_{th} = \arg\min_{\mathcal{A}} \sum_{\omega} \frac{1}{2}\|\boldsymbol{A}_{\cdot\cdot\omega} - \bar{\boldsymbol{A}}_{\cdot\cdot\omega}\|_F^2 + \lambda_1 \sum_{\omega} |\boldsymbol{A}_{\cdot\cdot\omega}|_1 \quad . \tag{2.15}$$

where λ_1 is the thresholding parameter.

The reasons that we enforce the sparsity pattern on the estimator through L_1-

norm are shown as follows: Firstly, the source of energy or signal is usually limited in real applications. To detecting these limited sources with high resolution is of considerable importance in many areas. Secondly, ADCP provides seismic data with an array of receiving antennas. The four measurements in the same water depth are highly correlated, whereas the observations from different water depths are less correlated somehow, especially for one water level is far away from the other one. Though the fact that the spectral density function $s(\omega)$ is the characteristic function of the autocovariance function $r(l)$ can not directly give the conclusion of the sparsity pattern of $s(\omega)$, the spectral function $s(\omega)$ can remain to express the cross-sectional information in terms of wave numbers. Thirdly, the assumption of sparsity can massively reduce the computation complexity of matrix inversion. One example is to solve the forward model

$$C(\omega) = \int_{-\pi}^{+\pi} \mathbf{H}(\omega, \theta) D(\omega, \theta) \mathbf{H}^*(\omega, \theta) d\theta,$$

for estimating the directional spectral density $D(\omega, \theta)$ in Chapter 4, and the maximum likelihood estimator $\hat{D}(\omega, \theta)$ consists of the inverse matrix of spectra matrix, that is $C^{-1}(\omega)$.

2.3.3 Smoothed Estimator

A fundamental goal of spectral estimation is to obtain a consistent estimator. To achieve this goal, we allow each element of the spectrum with different degree of smoothness.

In this work, we use the second-order differential technique to smooth the raw spectral estimator. The smoothed resulting soft-thresholding spectral estimator can

be expressed as,

$$\begin{aligned}\tilde{\mathcal{A}}_{th} &= \arg\min_{\mathcal{A}} \sum_{\omega} \frac{1}{2} \|\boldsymbol{A}_{\cdot\cdot\omega} - \bar{\boldsymbol{A}}_{\cdot\cdot\omega}\|_F^2 + \lambda_1 \sum_{\omega} |\boldsymbol{A}_{\cdot\cdot\omega}|_1 \\ &\quad + \lambda_2 \sum_{m,n} \boldsymbol{a}_{mn\cdot}^T \boldsymbol{Q} \boldsymbol{a}_{mn\cdot\cdot} \end{aligned} \tag{2.16}$$

where λ_2 is the smoothing parameter and $\boldsymbol{Q}$ is the $2N \times 2N$ penalty matrix.

Now we state several properties of the penalty matrix $\boldsymbol{Q}$ which are used to derive the theoretical property of the proposed algorithm.

Let $\boldsymbol{B}$ denote the matrix such that $\boldsymbol{Q} = \boldsymbol{B}^T\boldsymbol{B}$. Using second-order differential technique, $\boldsymbol{B}$ has the following form of

$$\boldsymbol{B} = \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 & 1 \\ 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 \\ 1 & 0 & 0 & \cdots & 1 & -2 \end{pmatrix}.$$

We can also state that the penalty matrix $\boldsymbol{Q}$ is positive semidefinite because for all non-zero vectors $\boldsymbol{x}$, the following inequality holds,

$$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{x} = (\boldsymbol{B}\boldsymbol{x})^T(\boldsymbol{B}\boldsymbol{x}) \geq 0.$$

2.3.4 Positive Definite Estimator

As we mentioned in Section 2.2, the positive-definite estimator is usually required in the spectral estimation. However, it is hard to guarantee that neither

the soft-thresholding spectral estimator in Equation (2.15) nor the smoothed soft-thresholding spectral estimator in Equation (2.16) are always positive definite, Especially in real data analysis, both the actual soft-thresholding estimator $\mathcal{A}_{th}$ and the actual smoothed soft-thresholding estimator $\tilde{\mathcal{A}}_{th}$ can be indefinite.

To deal with the indefiniteness issue arising in Section 2.3.2 and Section 2.3.3, it is natural to revise the eigenvalues of any raw estimator $\check{\boldsymbol{A}}_{\cdot\cdot\omega}$, to guarantee that all eigenvalues are positive. One possible solution is to perform the eigen-decomposition of $\check{\boldsymbol{A}}_{\cdot\cdot\omega}$ and then project $\check{\boldsymbol{A}}_{\cdot\cdot\omega}$ into the convex cone $\{\boldsymbol{A}_{\cdot\cdot\omega} > 0\}$.

To be specific, we assume that $\check{\boldsymbol{A}}_{\cdot\cdot\omega}$ can be expressed as the eigen-decomposition

$$\check{\boldsymbol{A}}_{\cdot\cdot\omega} = \sum_{n} \gamma_{n\omega} \boldsymbol{v}_{n\omega}^T \boldsymbol{v}_{n\omega}.$$

then we can obtain a positive definite estimator $\mathcal{A}^+ = \{\boldsymbol{A}^+_{\cdot\cdot\omega}\}_\omega$ where $\boldsymbol{A}^+_{\cdot\cdot\omega}$ has the form of

$$\boldsymbol{A}^+_{\cdot\cdot\omega} = \sum_{n=1}^{N} \max\{0, \gamma_{n\omega}\} \boldsymbol{v}_{n\omega}^T \boldsymbol{v}_{n\omega}, \quad for\ each\ fixed\ \omega.$$

However, it is known that the projection of $\check{\boldsymbol{A}}_{\cdot\cdot\omega}$ may destroy the sparsity pattern and smoothness pattern. The strategy in this section can not be directly used for sparse and smoothed spectral estimator given by Equation (2.16).

2.4 Proposed Estimator

In Section 2.4, we discuss the details of the proposed method and derive the closed form solution for each subproblem. The proposed new estimator simultaneously satisfy the requirements: sparsity, positive definiteness and smoothness. In the end of Section 2.4, we summarize our method to Algorithm 1.

In order to simultaneously guarantee sparsity, positive definite and smoothness,

we introduce a new estimator for the real-valued tensor $\mathcal{A}$ associated with the spectrum $\mathcal{C}$. A natural solution is to add the positive definite constraint $\{\boldsymbol{A}_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I}\}$ for some arbitrarily small $\epsilon > 0$ to Equation (2.16). We consider the following constrained problem,

$$\begin{aligned} \hat{\mathcal{A}} &= \arg\min_{\mathcal{A}} \{ \sum_{\omega} \frac{1}{2} \|\boldsymbol{A}_{\cdot\cdot\omega} - \bar{\boldsymbol{A}}_{\cdot\cdot\omega}\|_F^2 + \lambda_1 \sum_{\omega} |\boldsymbol{A}_{\cdot\cdot\omega}|_1 \\ &\quad + \lambda_2 \sum_{m,n} \boldsymbol{a}_{mn\cdot}^T \boldsymbol{Q} \boldsymbol{a}_{mn\cdot} : \quad \boldsymbol{A}_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I} \}. \end{aligned} \tag{2.17}$$

In this way, the modified $\hat{\mathcal{A}}$ is always positive definite with sparsity pattern and smoothness.

However, due to the positive definite constraint, solving the optimization problem (2.17) is very challenging. The same problem arises when the following soft-thresholding estimator for covariance matrix is employed,

$$\check{\boldsymbol{B}} = \arg\min_{B \geq \epsilon I} \frac{1}{2} \|\boldsymbol{B} - \bar{\boldsymbol{B}}\|_F^2 + \lambda |\boldsymbol{B}|_1. \tag{2.18}$$

Rothman [26] considered to add a log-determinant barrier function ($-\tau \log\det \boldsymbol{B}$) to Equation 2.18, that is

$$\check{\boldsymbol{B}}_{ba} = \arg\min_{B \geq \epsilon I} \frac{1}{2} \|\boldsymbol{B} - \bar{\boldsymbol{B}}\|_F^2 + \lambda |\boldsymbol{B}|_1 - \tau \log\det \boldsymbol{B}. \tag{2.19}$$

and derived an iterative procedure to solve problem (2.19). However, its convergence property is unknown. Xue [31] proposed an alternating direction method to solve problem (2.18) by adding one dummy variable Θ, that is,

$$(\hat{\boldsymbol{B}}^+, \hat{\Theta}^+) = \arg\min_{B,\Theta} \{ \frac{1}{2} \|\boldsymbol{B} - \bar{\boldsymbol{B}}\|_F^2 + \lambda |\boldsymbol{B}|_1 : \boldsymbol{B} = \Theta, \Theta \geq \epsilon \boldsymbol{I} \}. \tag{2.20}$$

Then solving problem 2.18 is converted to iteratively guarantee that Θ is positive definite and $\boldsymbol{B}$ is sparse. The equality constrain $\boldsymbol{B} = \Theta$ can ultimately guarantee that the final estimator is simultaneously positive definite and sparse. Xue [31] also provided the theoretical properties of his method.

Inspired by Xue's idea, we propose a generalized alternating direction algorithm for estimating large spectral density function, that is used to solve Equation (2.17).

2.5 Generalized Alternating Direction Algorithm

In this section, we describe in details the method to solve Equation (2.17) iteratively. To be specific, we use a generalized alternating direction method to obtain a smoothed L_1-penalized spectral estimator. under the positive-definite constraint $\{\boldsymbol{A}_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I}\}$ for some arbitrarily small $\epsilon > 0$.

We first introduce new variables $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$, as well as an equality constraint as follows,

$$\begin{aligned}(\hat{\mathcal{A}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Pi}}) &= \arg\min_{\mathcal{A},\Theta,\Pi} \{\frac{1}{2}\|\mathcal{A} - \bar{\mathcal{A}}\|_F^2 + \lambda_1 \sum_{\omega} |\boldsymbol{A}_{\cdot\cdot\omega}|_1 + \lambda_2 \sum_{m,n} \boldsymbol{\pi}_{mn\cdot}^T \boldsymbol{Q} \boldsymbol{\pi}_{mn\cdot} \\ &\quad : \Theta_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I}, \mathcal{A} = \boldsymbol{\Theta}, \mathcal{A} = \boldsymbol{\Pi}\}. \end{aligned} \tag{2.21}$$

To deal with the equality constraint in (2.21), we shall minimize its augmented Lagrangian function for some given penalty parameter μ_1 and μ_2, that is,

$$\begin{aligned}L(\mathcal{A}, \boldsymbol{\Theta}, \boldsymbol{\Pi}; \boldsymbol{\Lambda}, \boldsymbol{\Delta}) &= \frac{1}{2}\|\mathcal{A} - \bar{\mathcal{A}}\|_F^2 + \lambda_1 \sum_{\omega} |\boldsymbol{A}_{\cdot\cdot\omega}|_1 + \lambda_2 \sum_{m,n} (\boldsymbol{\pi}_{mn\cdot}^T \boldsymbol{Q} \boldsymbol{\pi}_{mn\cdot}) \\ &\quad - \langle \boldsymbol{\Lambda}, \boldsymbol{\Theta} - \mathcal{A} \rangle + \frac{1}{2\mu_1}\|\boldsymbol{\Theta} - \mathcal{A}\|_F^2 \\ &\quad - \langle \boldsymbol{\Delta}, \boldsymbol{\Pi} - \mathcal{A} \rangle + \frac{1}{2\mu_2}\|\boldsymbol{\Pi} - \mathcal{A}\|_F^2 \end{aligned} \tag{2.22}$$

where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Delta}$ are Lagrange multipliers.

We firstly update $\mathcal{A}$, $\mathbf{\Theta}$ and $\mathbf{\Pi}$ by solving

$$(\mathcal{A}^{t+1}, \mathbf{\Theta}^{t+1}, \mathbf{\Pi}^{t+1}) = \arg \min_{\mathcal{A},\Pi,\Theta_{\cdot\cdot\omega} \geq \epsilon I} L(\mathcal{A}, \mathbf{\Theta}, \mathbf{\Pi}; \mathbf{\Lambda}^t, \mathbf{\Delta}^t) \tag{2.23}$$

and then update the Lagrange multipliers $\mathbf{\Lambda}^{t+1}$ and $\mathbf{\Delta}^{t+1}$ by

$$\begin{aligned} \mathbf{\Lambda}^{t+1} &= \mathbf{\Lambda}^t - \frac{1}{\mu_1}(\mathbf{\Theta}^{t+1} - \mathcal{A}^{t+1}) \\ \mathbf{\Delta}^{t+1} &= \mathbf{\Delta}^t - \frac{1}{\mu_2}(\mathbf{\Pi}^{t+1} - \mathcal{A}^{t+1}) \end{aligned}$$

For (2.23) we solve it by alternatively minimizing $L(\mathbf{\Theta}, \mathbf{\Pi}, \mathcal{A}; \mathbf{\Lambda}^t, \mathbf{\Delta}^t)$ with respect to $\mathbf{\Pi}$, $\mathbf{\Theta}$ and $\mathcal{A}$. The updating formulae for $\mathbf{\Pi}$, $\mathbf{\Theta}$, $\mathcal{A}$ as well as $\mathbf{\Lambda}$, $\mathbf{\Delta}$ are listed in the entire algorithm proceeds as follows: for $t = 0, 1, 2, \cdots$, perform the following five steps sequentially till convergence.

$$\mathbf{\Pi}\ step: \quad \mathbf{\Pi}^{t+1} = \arg\min_{\Pi} L(\mathcal{A}^t, \mathbf{\Theta}^t, \mathbf{\Pi}; \mathbf{\Lambda}^t, \mathbf{\Delta}^t) \tag{2.24}$$

$$\mathbf{\Theta}\ step: \quad \mathbf{\Theta}^{t+1} = \arg \min_{\Theta_{\cdot\cdot\omega} \geq \epsilon I} L(\mathcal{A}^t, \mathbf{\Theta}, \mathbf{\Pi}^{t+1}; \mathbf{\Lambda}^t, \mathbf{\Delta}^t) \tag{2.25}$$

$$\mathcal{A}\ step: \quad \mathcal{A}^{t+1} = \arg\min_{\Sigma} L(\mathcal{A}, \mathbf{\Theta}^{t+1}, \mathbf{\Pi}^{t+1}; \mathbf{\Lambda}^t, \mathbf{\Delta}^t) \tag{2.26}$$

$$\mathbf{\Lambda}\ step: \quad \mathbf{\Lambda}^{t+1} = \mathbf{\Lambda}^t - \frac{1}{\mu_1}(\mathbf{\Theta}^{t+1} - \mathcal{A}^{t+1}) \tag{2.27}$$

$$\mathbf{\Delta}\ step: \quad \mathbf{\Delta}^{t+1} = \mathbf{\Delta}^t - \frac{1}{\mu_2}(\mathbf{\Pi}^{t+1} - \mathcal{A}^{t+1}) \tag{2.28}$$

To further simplify the alternating direction algorithm, we derive the closed-form solutions for (2.24), (2.25) and (2.26) respectively.

Firstly, let us consider the $\mathbf{\Pi}$ *step* in Equation (2.24).

$$\begin{aligned} \mathbf{\Pi}^{t+1} &= \arg\min_{\Pi} L(\mathbf{\Theta}^t, \mathbf{\Pi}, \mathcal{A}^t; \mathbf{\Lambda}^t, \mathbf{\Delta}^t) \\ &= \arg\min_{\Pi} L(\mathbf{\Pi}, \mathcal{A}^t; \mathbf{\Delta}^t) \end{aligned} \tag{2.29}$$

By removing all items containing $\boldsymbol{\Theta}^t$ and $\boldsymbol{\Lambda}^t$, the optimal $\boldsymbol{\Pi}^{t+1}$ is equal to

$$\begin{aligned}
\boldsymbol{\Pi}^{t+1} = & \arg\min_{\Pi} \Big\{ \lambda_2 \sum_{m,n} (\boldsymbol{\pi}_{mn\cdot}^T \boldsymbol{Q} \boldsymbol{\pi}_{mn\cdot}) \\
& + \sum_{\omega} (-\langle \Delta_{\cdot\cdot\omega}^t, \Pi_{\cdot\cdot\omega} - \boldsymbol{A}_{\cdot\cdot\omega}^t \rangle + \frac{1}{2\mu_2} \|\Pi_{\cdot\cdot\omega} - \boldsymbol{A}_{\cdot\cdot\omega}^t\|_F^2) \Big\} \\
= & \arg\min_{\Pi} \Big\{ \lambda_2 \sum_{m,n} (\boldsymbol{\pi}_{mn\cdot}^T \boldsymbol{Q} \boldsymbol{\pi}_{mn\cdot}) \\
& + \frac{1}{2\mu_2} \big(\|\Pi\|_F^2 - 2\langle \Pi, \mathcal{A}^t \rangle - 2\mu_2 \langle \Pi, \Delta^t \rangle \big) \Big\}.
\end{aligned}$$

By rewriting the Frobenius norm and inner product in terms of the summation of m and n, we can obtain a unified representation for $\boldsymbol{\Pi}^{t+1}$ as follows,

$$\begin{aligned}
\boldsymbol{\Pi}^{t+1} = & \arg\min_{\Pi} \Big\{ \lambda_2 \sum_{m,n} \boldsymbol{\pi}_{mn\cdot}^T \boldsymbol{Q} \boldsymbol{\pi}_{mn\cdot} \\
& + \frac{1}{2\mu_2} \sum_{m,n} [\boldsymbol{\pi}_{mn\cdot}^T \boldsymbol{\pi}_{mn\cdot} - 2(\boldsymbol{a}_{mn\cdot}^t + \mu_2 \boldsymbol{\delta}_{mn\cdot}^t)^T \boldsymbol{\pi}_{mn\cdot}] \Big\} \\
= & \arg\min_{\Pi} \sum_{m,n} \Big\{ \lambda_2 \boldsymbol{\pi}_{mn\cdot}^T (\boldsymbol{Q} + \frac{1}{2\lambda_2\mu_2} \boldsymbol{I}) \boldsymbol{\pi}_{mn\cdot} \\
& - \frac{1}{\mu_2} (\boldsymbol{a}_{mn\cdot}^t + \mu_2 \boldsymbol{\delta}_{mn\cdot}^t)^T \boldsymbol{\pi}_{mn\cdot} \Big\}. \qquad (2.30)
\end{aligned}$$

Optimize the above equation by minimizing element-wisely, for each $m, n = 1, \cdots, N$, the optimal $\boldsymbol{\pi}_{mn\cdot}$ in the $t+1$th iteration is equal to,

$$\begin{aligned}
\boldsymbol{\pi}_{mn\cdot}^{t+1} &= \frac{1}{2\lambda_2} (\boldsymbol{Q} + \frac{1}{2\lambda_2\mu_2} \boldsymbol{I})^{-1} \frac{1}{\mu_2} (\boldsymbol{a}_{mn\cdot}^t + \mu_2 \boldsymbol{\delta}_{mn\cdot}^t) \\
&= (\boldsymbol{I} + 2\lambda_2\mu_2 \boldsymbol{Q})^{-1} (\boldsymbol{a}_{mn\cdot}^t + \mu_2 \boldsymbol{\delta}_{mn\cdot}^t). \qquad (2.31)
\end{aligned}$$

For simplicity of the computation, the updating formula for $\boldsymbol{\Pi}^{t+1}$ can be further derived as follows:

- Using the fact that $\boldsymbol{Q}$ is positive semidefinite, the eigen-decomposition of $\boldsymbol{Q}$ is given by $\boldsymbol{Q} = \boldsymbol{w}\Psi\boldsymbol{w}^T$.

- Combining with the following identity,

$$\begin{aligned}(\boldsymbol{I} + 2\lambda_2\mu_2\boldsymbol{Q})^{-1} &= \boldsymbol{w}(I + 2\lambda_2\mu_2\Psi)^{-1}\boldsymbol{w}^T \\ &= \boldsymbol{w}\ diag\{\frac{1}{1 + 2\lambda_2\mu_2\psi_{mn}}\}\boldsymbol{w}^T. \end{aligned} \tag{2.32}$$

- Obtaining the optimal $\boldsymbol{\pi}_{mn\cdot}^{t+1}$ with an easier form of

$$\boldsymbol{\pi}_{mn\cdot}^{t+1} = \boldsymbol{w}\ diag\{\frac{1}{1 + 2\lambda_2\mu_2\psi_{mn}}\}\boldsymbol{w}^T(\boldsymbol{a}_{mn\cdot}^t + \mu_2\boldsymbol{\delta}_{mn\cdot}^t). \tag{2.33}$$

Secondly, we consider the update in $\boldsymbol{\Theta}$ step, Equation (2.25). Define $(\boldsymbol{Z})_+$ as the projection of a matrix $\boldsymbol{Z}$ onto the convex cone $\{\boldsymbol{Z} \geq \epsilon\boldsymbol{I}\}$. Assume that $\boldsymbol{Z}$ has the eigen-decomposition $\sum_{n=1}^N c_n\boldsymbol{v}_n^T\boldsymbol{v}_n$, and then $(\boldsymbol{Z})_+$ can be obtained as $\sum_{n=1}^N \max(c_n, \epsilon)\boldsymbol{v}_n^T\boldsymbol{v}_n$. The $\boldsymbol{\Theta}$ step update can be analytically solved as follows:

$$\begin{aligned}\boldsymbol{\Theta}^{t+1} &= \arg\min_{\Theta_{\cdot\cdot\omega} \geq \epsilon I} L(\mathcal{A}^t, \boldsymbol{\Theta}, \boldsymbol{\Pi}^{t+1}; \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t) \\ &= \arg\min_{\Theta_{\cdot\cdot\omega} \geq \epsilon I} L(\mathcal{A}^t, \boldsymbol{\Theta}; \boldsymbol{\Lambda}^t) \end{aligned} \tag{2.34}$$

For each fixed ω, the updating formula of $\Theta_{(\ \cdot\ \cdot\omega)}^{t+1}$ can be derived as,

$$\begin{aligned}\Theta_{\cdot\cdot\omega}^{t+1} &= \arg\min_{\Theta_{\cdot\cdot\omega} \geq \epsilon I} \left\{ -\langle\Lambda_{\cdot\cdot\omega}^t, \Theta_{\cdot\cdot\omega} - \boldsymbol{A}_{\cdot\cdot\omega}^t\rangle + \frac{1}{2\mu_1} \parallel \Theta_{\cdot\cdot\omega} - \boldsymbol{A}_{\cdot\cdot\omega}^t \parallel_F^2 \right\} \\ &= \arg\min_{\Theta_{\cdot\cdot\omega} \geq \epsilon I} \left\{ \frac{1}{2\mu_1}(\|\Theta_{\cdot\cdot\omega}\|_F^2 - 2\langle\Theta_{\cdot\cdot\omega}, \boldsymbol{A}_{\cdot\cdot\omega}^t\rangle - 2\mu_1\langle\Lambda_{\cdot\cdot\omega}^t, \Theta_{\cdot\cdot\omega}\rangle) \right\} \\ &= \arg\min_{\Theta_{\cdot\cdot\omega} \geq \epsilon I} \|\Theta_{\cdot\cdot\omega} - (\boldsymbol{A}_{\cdot\cdot\omega}^t + \mu_1\Lambda_{\cdot\cdot\omega}^t)\|_F^2 \\ &= (\boldsymbol{A}_{\cdot\cdot\omega}^t + \mu_1\Lambda_{\cdot\cdot\omega}^t)_+ \end{aligned} \tag{2.35}$$

At last, we consider the closed-form solution for $\mathcal{A}$ in Equation (2.26). Define an entry-wise soft-thresholding rule for all the off-diagonal elements of a matrix $\boldsymbol{Z}$ as

$$\boldsymbol{G}(\boldsymbol{Z}, \tau) = \{g(z_{mn}, \tau)\}_{1 \le m,n \le N} \tag{2.36}$$

with

$$g(z_{mn}, \tau) = \ sign(z_{mn}) \max(|z_{mn}| - \tau, 0) I_{\{m \neq n\}} + z_{mn} I_{\{m=n\}}. \tag{2.37}$$

Then $\mathcal{A}$ step has a closed-form solution given as follows:

$$\begin{aligned}
\mathcal{A}^{t+1} = & \ \arg\min_{\mathcal{A}} \ L(\boldsymbol{\Theta}^{t+1}, \boldsymbol{\Pi}^{t+1}, \mathcal{A}; \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t) \\
= & \ \arg\min_{\mathcal{A}} \ \sum_{\omega} (\frac{1}{2} \|\boldsymbol{A}_{\cdot\cdot\omega} - \bar{\boldsymbol{A}}_{\cdot\cdot\omega}\|_F^2 + \lambda_1 |\boldsymbol{A}_{\cdot\cdot\omega}|_1) \\
& + \sum_{\omega} (-\langle \Lambda^t_{\cdot\cdot\omega}, \Theta^{t+1}_{\cdot\cdot\omega} - \boldsymbol{A}_{\cdot\cdot\omega} \rangle + \frac{1}{2\mu_1} \|\Theta^{t+1}_{\cdot\cdot\omega} - \boldsymbol{A}_{\cdot\cdot\omega}\|_F^2) \\
& + \sum_{\omega} (-\langle \Delta^t_{\cdot\cdot\omega}, \Pi^{t+1}_{\cdot\cdot\omega} - \mathcal{A}_{\cdot\cdot\omega} \rangle + \frac{1}{2\mu_2} \|\Pi^{t+1}_{\cdot\cdot\omega} - \mathcal{A}_{\cdot\cdot\omega}\|_F^2)
\end{aligned} \tag{2.38}$$

Recalculating and simplifying the above equation, we can obtain

$$\begin{aligned}
& \mathcal{A}^{t+1} \\
= & \ \arg\min_{\mathcal{A}} \ \sum_{\omega} \Bigg(\frac{1}{2} \|\boldsymbol{A}_{\cdot\cdot\omega} - \bar{\boldsymbol{A}}_{\cdot\cdot\omega}\|_F^2 + \lambda_1 |\boldsymbol{A}_{\cdot\cdot\omega}|_1 \\
& + \frac{1}{2\mu_1} \|\Theta^{t+1}_{\cdot\cdot\omega} - \boldsymbol{A}_{\cdot\cdot\omega}\|_F^2 + \langle \Lambda^t_{\cdot\cdot\omega}, \boldsymbol{A}_{\cdot\cdot\omega} \rangle \\
& + \frac{1}{2\mu_2} \|\Pi^{t+1}_{\cdot\cdot\omega} - \boldsymbol{A}_{\cdot\cdot\omega}\|_F^2 + \langle \Delta^t_{\cdot\cdot\omega}, \boldsymbol{A}_{\cdot\cdot\omega} \rangle \Bigg) \\
= & \ \arg\min_{\mathcal{A}} \ \sum_{\omega} \Bigg(\frac{1}{2} \|\boldsymbol{A}_{\cdot\cdot\omega} - \frac{\mu_1\mu_2(\bar{\boldsymbol{A}} - \Lambda^t - \Delta^t)_{\cdot\cdot\omega} + \mu_1 \Pi^{t+1}_{\cdot\cdot\omega} + \mu_2 \Theta^{t+1}_{\cdot\cdot\omega}}{\mu_1\mu_2 + \mu_1 + \mu_2} \|_F^2 \\
& + \frac{\lambda_1 \mu_1 \mu_2}{\mu_1\mu_2 + \mu_1 + \mu_2} |\boldsymbol{A}_{\cdot\cdot\omega}|_1 \Bigg).
\end{aligned} \tag{2.39}$$

For each fixed ω, the optimal solution of Equation (2.26) is given by

$$\begin{aligned} & \boldsymbol{A}_{\cdot\cdot\omega}^{t+1} \\ = & \frac{1}{\mu_1\mu_2+\mu_1+\mu_2} \times G\bigg(\mu_1\mu_2(\bar{\boldsymbol{A}}-\Lambda^t-\Delta^t)_{\cdot\cdot\omega}+\mu_1\Pi_{\cdot\cdot\omega}^{t+1}+\mu_2\Theta_{\cdot\cdot\omega}^{t+1},\lambda_1\mu_1\mu_2\bigg) \\ = & \frac{1}{1+\frac{1}{\mu_1}+\frac{1}{\mu_2}} \times G\bigg((\bar{\boldsymbol{A}}-\Lambda^t-\Delta^t)_{\cdot\cdot\omega}+\frac{1}{\mu_1}\Theta_{\cdot\cdot\omega}^{t+1}+\frac{1}{\mu_2}\Pi_{\cdot\cdot\omega}^{t+1},\lambda_1\bigg). \end{aligned} \tag{2.40}$$

So far, we derive all the closed-form solutions for Equations (2.24)-(2.28). The following Algorithm 1 shows the complete details of the proposed generalized alternating direction method for solving problem (2.23).

Algorithm 1 The alternating direction method for proposed estimator.

1: Input: μ_1, μ_2, $\mathcal{A}^0$, $\mathbf{\Lambda}^0$ and $\mathbf{\Delta}^0$.

2: Iterative alternating direction augmented Lagrangian step: for the tth iteration

2.1 In $\mathbf{\Pi}$ step, update $\mathbf{\Pi}^{t+1}$. For each fixed m, n, solve

$$\boldsymbol{\pi}_{mn\cdot}^{t+1} = \boldsymbol{w}\ diag\{\frac{1}{1+2\lambda_2\mu_2\psi_{mn}}\}\boldsymbol{w}^T(\boldsymbol{a}_{mn\cdot}^t+\mu_2\boldsymbol{\delta}_{mn\cdot}^t) \quad .$$

2.2 In $\mathbf{\Theta}$ step, update $\mathbf{\Theta}^{t+1}$. For each fixed ω, solve

$$\Theta_{\cdot\cdot\omega}^{t+1} = (\boldsymbol{A}_{\cdot\cdot\omega}^t+\mu_1\Lambda_{\cdot\cdot\omega}^t)_+ \quad .$$

2.3 In $\mathcal{A}$ step, update $\mathcal{A}^{t+1}$. For each fixed ω, solve

$$\boldsymbol{A}_{\cdot\cdot\omega}^{t+1} = \frac{1}{1+\frac{1}{\mu_1}+\frac{1}{\mu_2}} \times G\bigg((\bar{\boldsymbol{A}}-\Lambda^t-\Delta^t)_{\cdot\cdot\omega}+\frac{1}{\mu_1}\Theta_{\cdot\cdot\omega}^{t+1}+\frac{1}{\mu_2}\Pi_{\cdot\cdot\omega}^{t+1},\lambda_1\bigg) \quad .$$

2.4 Update $\mathbf{\Lambda}^{t+1} = \mathbf{\Lambda}^t - \frac{1}{\mu_1}(\mathbf{\Theta}^{t+1}-\mathcal{A}^{t+1})$;

2.5 Update $\mathbf{\Delta}^{t+1} = \mathbf{\Delta}^t - \frac{1}{\mu_2}(\mathbf{\Pi}^{t+1}-\mathcal{A}^{t+1})$.

3: Repeat the above cycle till convergence.

The selection of the starting value for $\mathcal{A}^0$, $\mathbf{\Theta}^0$ and $\mathbf{\Pi}^0$ is much more flexible

than other methods. In our numerical experiments, we use the sample estimator for multivariate spectral density as the initial value and We set the starting points of $\mathbf{\Lambda}$ and $\mathbf{\Delta}$ as zero tensors.

Unlike λ_1 and λ_2, μ_1 and μ_2 have no influence on the final spectral estimator which can be guaranteed by the proof of Theorem 3.2.1. In our implementation in the next section, we set $\mu_1 = \mu_2$ for simplicity.

2.6 Numerical Examples

In this section, we compare the proposed alternating direction method with three existing estimators of wave-number spectral function for multivariate time series.

2.6.1 Compared Estimators

The first compared method is the sample spectral estimator. Using Fourier transform technique, the raw multivariate time series data $\{\boldsymbol{x}(t)\}$ can be transferred from time domain to frequency domain. Let $\{\boldsymbol{u}(\omega_k)\}$ denote the transferred data. The sample spectrum for frequency ω_1 is defined as

$$\bar{\boldsymbol{C}}(\omega_1) = \begin{pmatrix} u_1(\omega_1) \\ u_2(\omega_1) \\ \vdots \\ u_N(\omega_1) \end{pmatrix} \times \left(u_1^*(\omega_1), u_2^*(\omega_1), \cdots, u_N^*(\omega_1) \right)$$

Since the rank of $\bar{\boldsymbol{C}}(\omega_1)$ is 1. and combining the fact that there is only one realization of the multivariate time series, we know that sample spectral estimator is singular. Figure (2.1) illustrates the reason why sample spectral estimator performs very poorly in high dimensional setting.

The second compared method, banding method, comes from one of the regularization techniques. Wu (2003) [30] and Bickel (2008) [2] proposed banding estimators

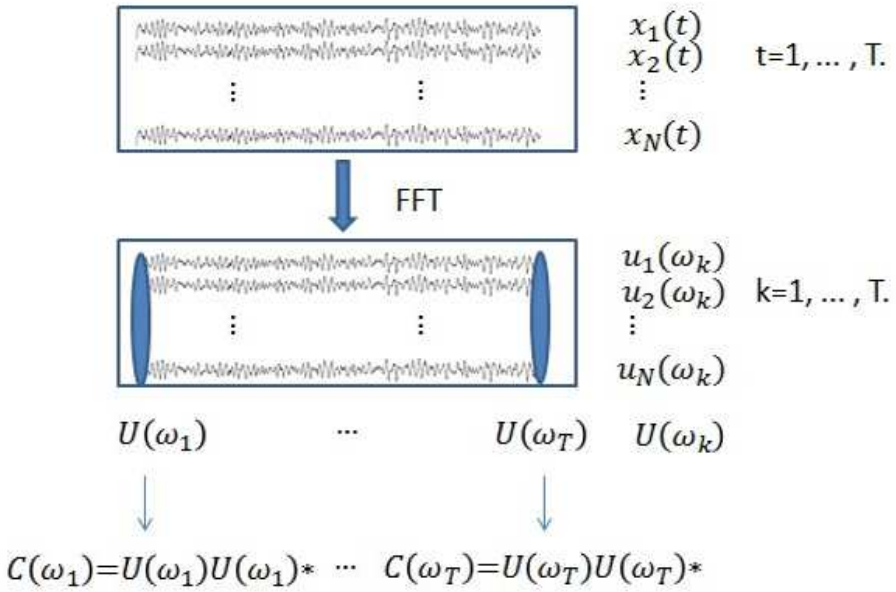

Figure 2.1: Illustration of the singularity of sample spectra.

for autocovariance function. The analogous estimation of spectra is known as traditional kernel smoothing methods, [3] and [4]. that is to ensure the wave-number spectral estimate is positive definite, the same bandwidth is used in smoothing all the elements. Based on the sample spectral estimator, the banding estimator for wave-number spectra can be similarly defined as,

$$\boldsymbol{C}_b(\omega_1) = \sum_{k=1}^{B_1} \bar{\boldsymbol{C}}(\omega_k).$$

Smoothing all elements of the wave-number spectra with the same bandwidth in frequency is a very rigid constraint. In many situations, in order to establish the optimal estimators, we must allow different entries of the wave-number spectra have different smoothness. Another difficulty is the selection of bandwidth and the number

of bins due to the dependence structure of time series data.

The third compared method is proposed by Dai & Guo [11] for estimating a positive definite multivariate spectra. Dai & Guo's method is actually an extension of tapering regularization method with smoothed Cholesky decomposition components. The estimation procedure consists of the following four steps:

(1) Obtain a positive definite and asymptotically unbiased spectral estimator using the multitaper spectral estimator, [29];

(2) Perform the Cholesky decomposition on the raw estimator calculated in (1);

(3) Smooth each of the Cholesky decomposition components with its own smoothing parameter to allow optimal smoothing for all components;

(4) Reconstruct the spectral estimator from the smoothed Cholesky decomposition components

The major merit of Dai & Guo's method is that the final spectral estimator produced by the above four steps is consistent, positive definite and smooth in frequency. However, the estimator proposed by Dai (2004) highly depends on the raw estimator which must be positive definite. Another disadvantage of Dai's method is that the substep, the Cholesky decomposition, is time-consuming, therefore the four-step procedure is difficult to implement in the high dimensional setting.

2.6.2 Simulated Data

In this section, we illustrate that the estimator produced by Algorithm 1 is not only consistent and positive definite, but has the flexibility to select the initial value for the raw estimator. The proposed estimator allows different components of the spectral matrix with their own smoothing parameters. Here we also force the proposed estimator to have a sparse pattern, because in Chapter 4, we will solve the

forward model (1.32) by inverting the resulting estimator for multivariate spectral matrix, and a sparse estimator will efficiently reduce the inverting procedure.

We simulate a 10-dimensional time series with 120 observations, that is $T = 120$ and $N = 10$. The selection of its spectrum $\boldsymbol{S}(\omega)$ is given by its complex Cholesky decomposition, denoted by $\boldsymbol{R}(\omega)$, as follows,

- The diagonal elements are defined as,

$$\begin{aligned}
R_{11}(\omega) &= 0.5\cos(2\pi\omega) + 1.5 \\
R_{22}(\omega) &= 0.5\cos(2\pi\omega) + 1.5 \\
R_{33}(\omega) &= 0.4\cos(4\pi\omega) + 1.2 \\
R_{44}(\omega) &= 0.4\cos(4\pi\omega) + 1.2 \\
R_{55}(\omega) &= 0.5\cos(6\pi\omega) + 1.5 \\
R_{66}(\omega) &= 0.5\cos(6\pi\omega) + 1.5 \\
R_{77}(\omega) &= 0.4\cos(8\pi\omega) + 1.2 \\
R_{88}(\omega) &= 0.4\cos(8\pi\omega) + 1.2 \\
R_{99}(\omega) &= 0.5\cos(10\pi\omega) + 1.5 \\
R_{10,10}(\omega) &= 0.5\cos(10\pi\omega) + 1.5
\end{aligned}$$

- The first off-diagonal elements are defined as,

$$\begin{aligned}
R_{21}(\omega) &= 0.9\cos(4\pi\omega) + 1.2i\sin(2\pi\omega) \\
R_{32}(\omega) &= 0.3\cos(4\pi\omega) + 0.4i\sin(2\pi\omega) \\
R_{43}(\omega) &= 0.9\cos(8\pi\omega) + 1.2i\sin(2\pi\omega) \\
R_{54}(\omega) &= 0.3\cos(8\pi\omega) + 0.4i\sin(2\pi\omega) \\
R_{65}(\omega) &= 0.9\cos(12\pi\omega) + 1.2i\sin(4\pi\omega) \\
R_{76}(\omega) &= 0.3\cos(12\pi\omega) + 0.4i\sin(4\pi\omega) \\
R_{87}(\omega) &= 0.9\cos(16\pi\omega) + 1.2i\sin(6\pi\omega) \\
R_{98}(\omega) &= 0.3\cos(16\pi\omega) + 0.4i\sin(6\pi\omega) \\
R_{10,9}(\omega) &= 0.9\cos(20\pi\omega) + 1.2i\sin(8\pi\omega).
\end{aligned}$$

- The other off-diagonal ones are defined as

$$R_{mn}(\omega) = 0.$$

From $R(\omega)$, we generate the time series

$$\boldsymbol{X}_t = \begin{pmatrix} \boldsymbol{x}_1(t) \\ \vdots \\ \boldsymbol{x}_{10}(t) \end{pmatrix} = \sum_{k=1}^{120} \boldsymbol{R}(\omega_k) e^{2\pi i \omega_k t} \boldsymbol{Z}(k),$$

where $\omega_k = k/120$ for $k = 1, \cdots, 120$. For $\omega_k = 0, 0.5$ and 1, $\boldsymbol{Z}(\omega_k)$ is generated from a 10-dimensional real normal with zero mean and covariance matrix $(1/120)\boldsymbol{I}_1 0$. For $\omega_k \neq 0, 0.5$ and 1, $\boldsymbol{Z}(\omega_k)$ is generated from a 10-dimensional complex normal with zero mean and covariance matrix $(1/120)\boldsymbol{I}_1 0$,

2.6.3 Numerical Results

We compare the performance of our method, Guo & Dai's smoothed tapering method, the banding method and the sample spectra.

First, we compare the performance of four estimators. For our proposed smoothed soft thresholding estimator, the thresholding parameter and the smoothing estimator were chosen over 16 thresholding parameters $\lambda_1 = \{100, 110, \cdots, 250\}$ and over 31 smoothing parameters $\lambda_2 = \{0.5, 0.6, \cdots, 3.5\}$ by twofold cross-validation ([21]).

The estimation performance is measured by the total losses under the Frobenius norm, that is,

$$totalerror = \|\hat{\mathcal{A}} - \mathcal{A}\|_F^2.$$

Moreover, we compare the percentage of positive definiteness to check the positive definiteness, which is measured by the positive rate of spectra for all frequencies. The selection performance is evaluated by both the sensitivity defined as,

$$\frac{\#\{(m,n,t) : \hat{\mathcal{A}}_{mnt} = 0 \& \mathcal{A}_{mnt} = 0\}}{\#\{(m,n,t) : \mathcal{A}_{mnt} = 0\}} \tag{2.41}$$

and the specificity defined as,

$$\frac{\#\{(m,n,t) : \hat{\mathcal{A}}_{mnt} \neq 0 \& \mathcal{A}_{mnt} \neq 0\}}{\#\{(m,n,t) : \mathcal{A}_{mnt} \neq 0\}} \tag{2.42}$$

The bigger sensitivity is seen the better, and it's the same for specificity. However, the sensitivity and the specificity have a trade-off relationship that the increase of one yields the decrease of the other. Therefore, we use a better criterion, the summation of specificity and sensitivity, to evaluate the sparsity performance of four estimators.

Comparison is based on 100 replications. And summing up the above interested statistics as follows:

- estimation error.
- positive definiteness rate.
- sensitivity.
- specificity.
- the summation of sensitivity and specificity.

Table 2.1, Table 2.2 and Table 2.3 provide numerical summaries for the results of sample spectra, smoothed tapering method, banding method and the proposed method. For each method, we reported the evaluation criterion for total errors, estimation, selection and positive definiteness.

	Error
Sample Spectra	164,688(1,861)
ET Method	15,208(204)
Banding Method	15,743(152)
Our Method	6,320(58)

Table 2.1: The comparison of the estimation performances of four methods: sample spectra, extended tapering method, banding method and our proposed method.

	Positive Definite
Sample Spectra	0%
ET Method	100%
Banding Method	58%
Our Method	100%

Table 2.2: The comparison of the positive definiteness rates of four methods: sample spectra, extended tapering method, banding method and our proposed method.

	Sensitivity	Specificity	Sensitivity+Specificity
Sample Spectrum	100%	0%	1.00
Guo Dai's Method	100%	0%	1.00
Banding	100%	0%	1.00
Our Method	68(1.5)%	44(1.3)%	1.12(0.0035)

Table 2.3: The comparison of the sparsity rates of four methods: sample spectra, extended tapering method, banding method and our proposed method.

At the end, we illustrate the smoothing performance by comparing Figure 2.2, Figure 2.3 Figure 2.4, Figure 2.5 and Figure 2.6.

Figure 2.2, Figure 2.3 show the comparisons of the sample spectra, extended tapering method, banding method and our method (black line) with the true spectrum (red line) respectively.

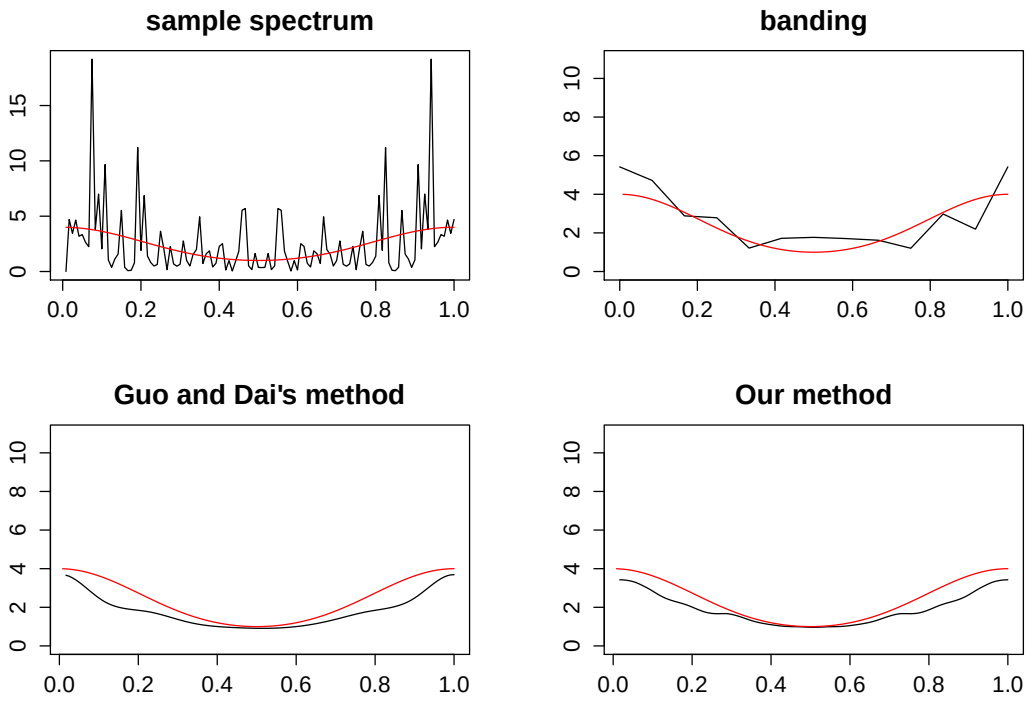

Figure 2.2: The comparison of four spectral estimators: the sample spectrum, extended tapering method, banding method and our method (black line) with the true spectrum (red line).

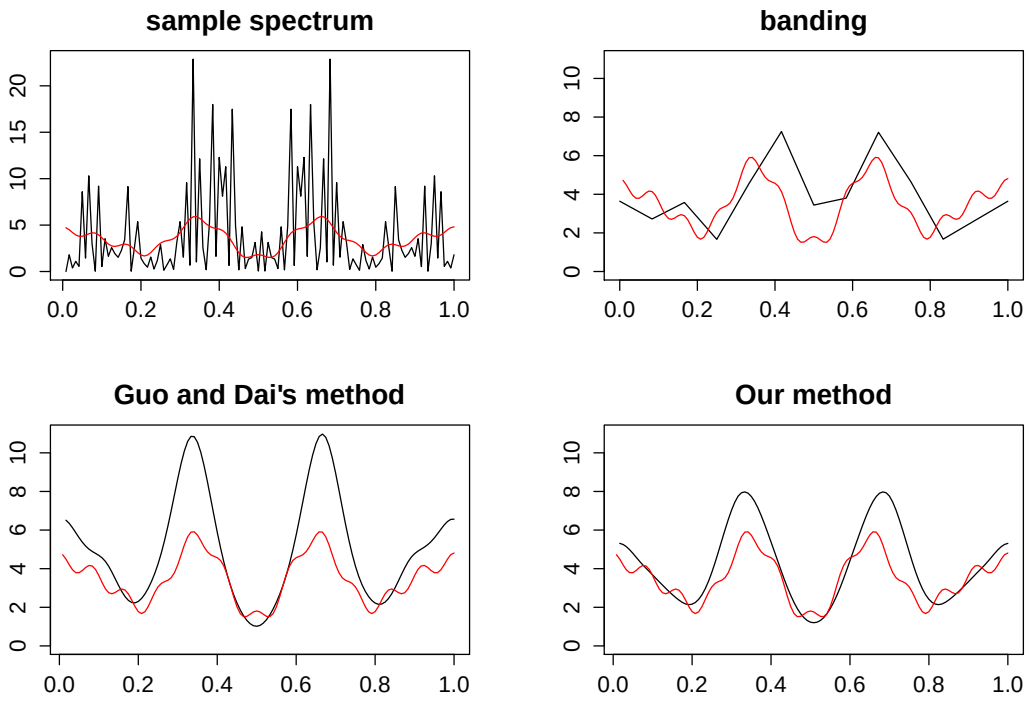

Figure 2.3: The comparison of four cross-spectral estimators: the sample cross-spectrum, extended tapering method, banding method and our method (black line) with the true spectrum (red line).

Figure 2.4, Figure 2.5 and Figure 2.6 show the comparisons of the sample spectra, extended tapering method, banding method and our method (black line) with the true spectrum (red line) with 95% percentage confidence interval respectively.

The numerical experiments in this section illustrate that our proposed wavenumber spectral estimator is more accurate than the smoothed tapering estimator proposed by Dai and Guo (2004), especially for the estimation of boundary spectra. Our method also significantly outperform the other two estimators while considering the requirements of positive definite and smoothness. Moreover, the sparsity pattern of the proposed estimator can massively reduce the computation complexity for further study.

In the next chapter, we will analyze in detail the convergence property of Algorithm 1 theoretically.

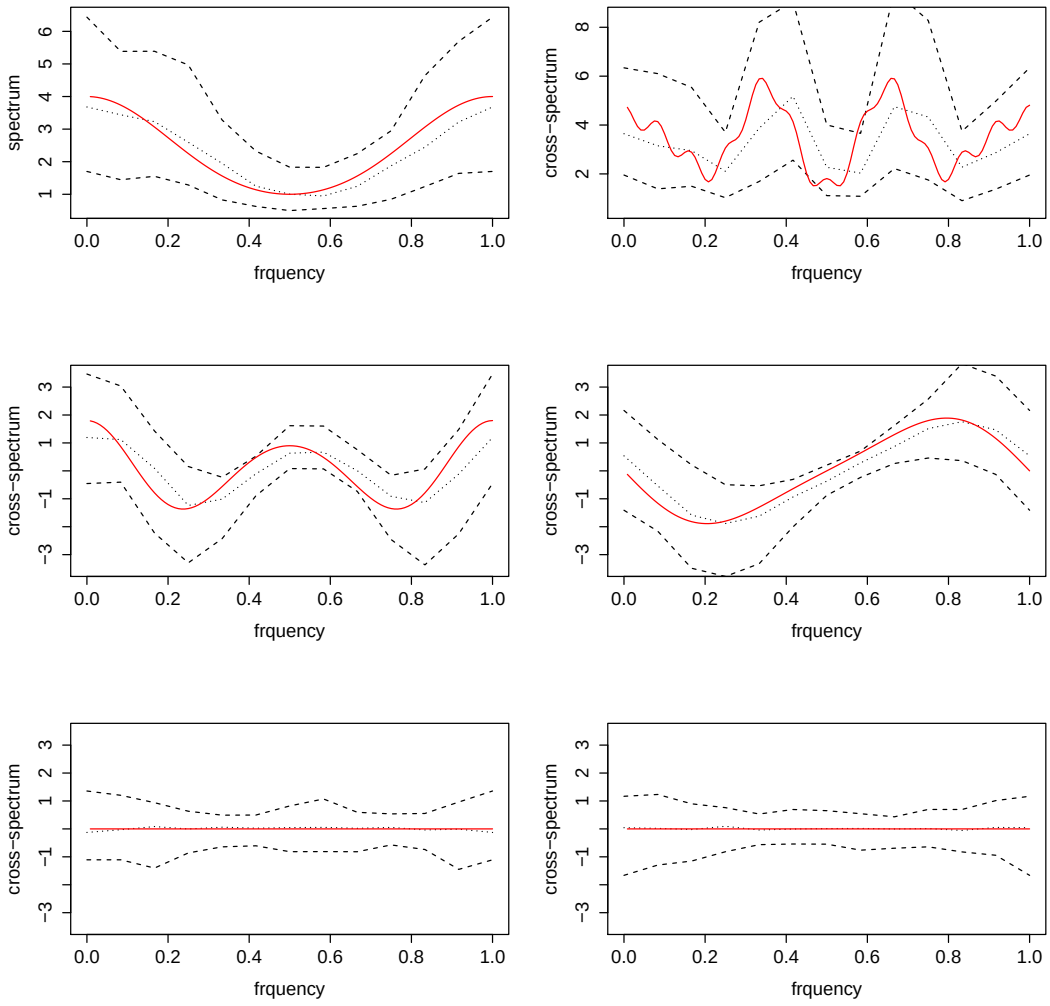

Figure 2.4: The comparison of two estimators with 95% percentage confidence interval: banding estimator (black line) and true spectrum (red line).

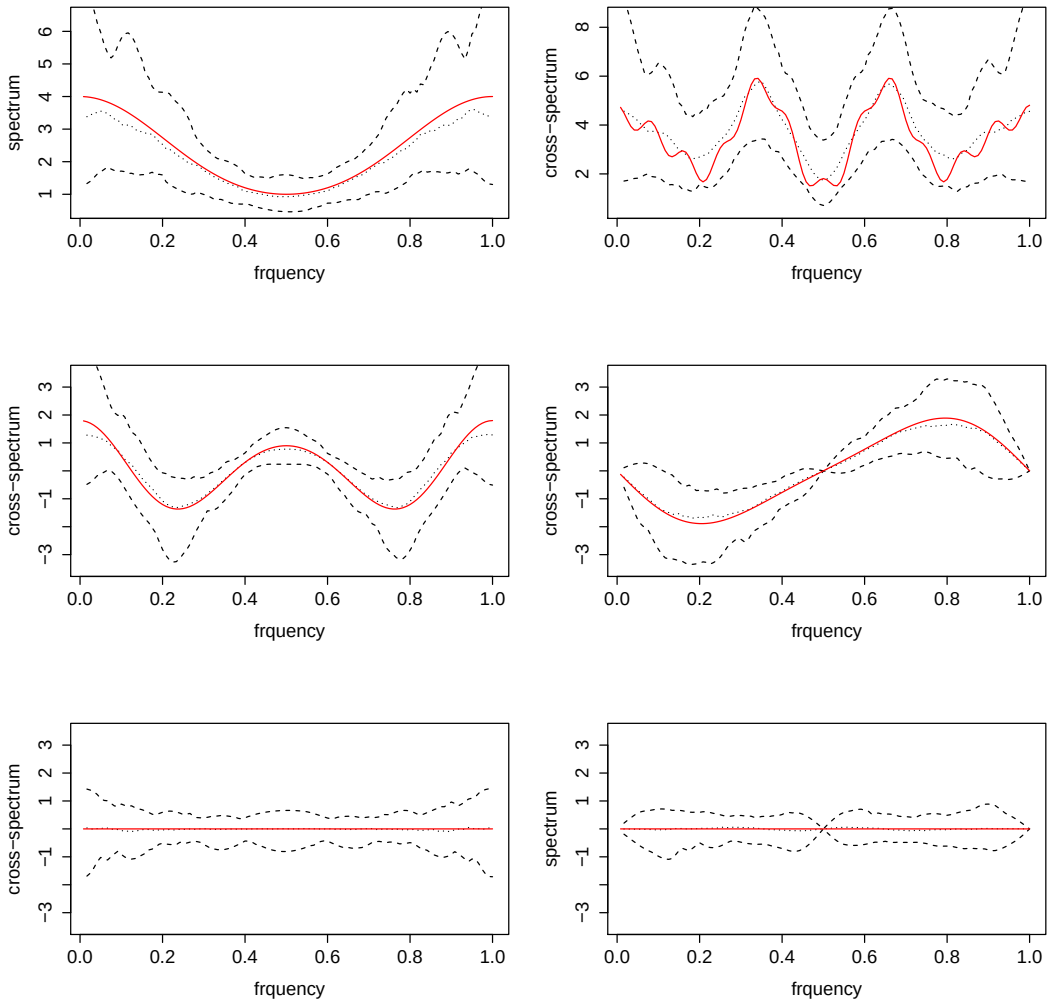

Figure 2.5: The comparison of two estimators with 95% percentage confidence interval: extended tapering estimator (black line) and true spectrum (red line).

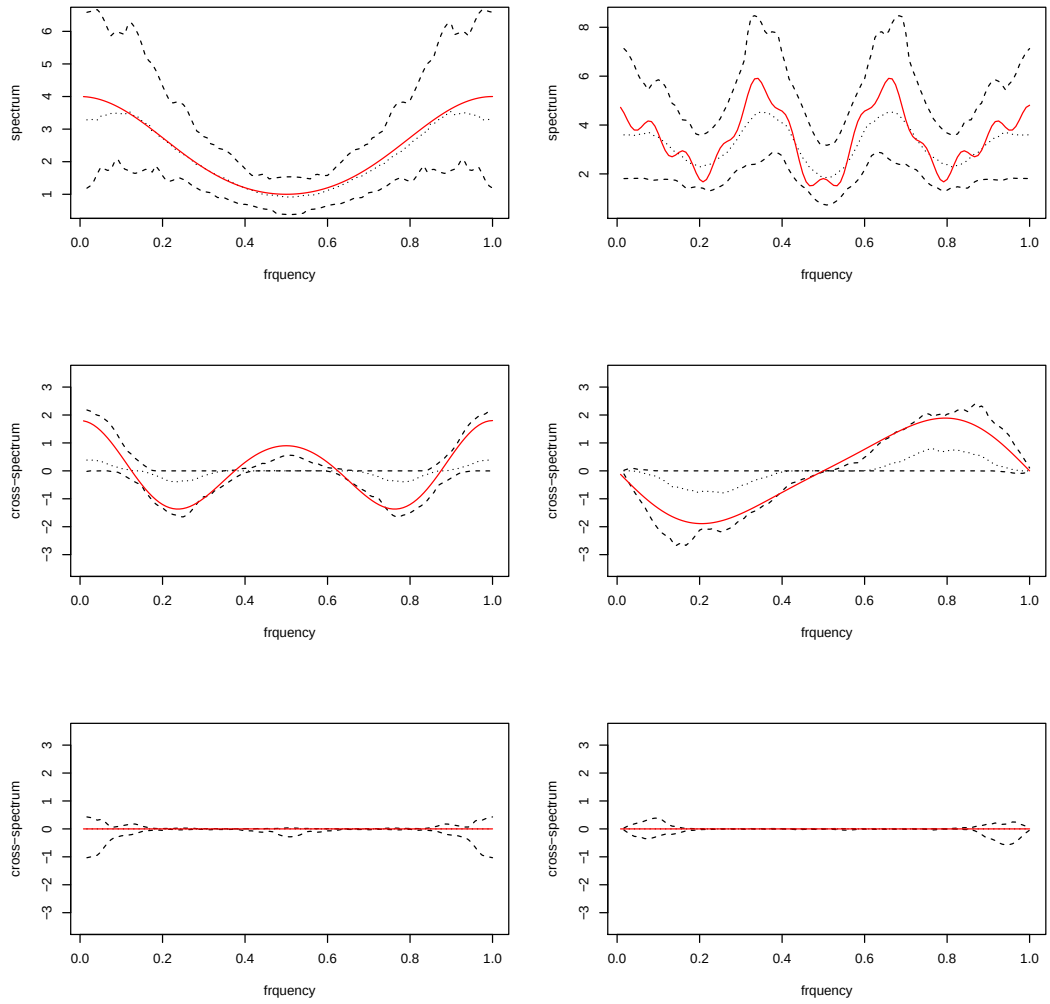

Figure 2.6: The comparison of two estimators with 95% percentage confidence interval: proposed estimator (black line) and true spectrum (red line).

3. THEORETICAL ANALYSIS OF ALGORITHM 1

In this chapter, we prove that the sequence $(\mathcal{A}^t, \mathbf{\Theta}^t, \mathbf{\Pi}^t, \mathbf{\Lambda}^t, \mathbf{\Delta}^t)$ produced by Algorithm 1 from any starting point converges to an optimal solution of problem (2.17). To complete the proof of this main theorem, we firstly describe and prove two lemmas which is necessary and helpful to prove the main resulting theorem.

3.1 Convergence Analysis of the Algorithm

In this section, we prove that the sequence $(\mathcal{A}^t, \mathbf{\Theta}^t, \mathbf{\Pi}^t, \mathbf{\Lambda}^t, \mathbf{\Delta}^t)$ produced by the generalized alternating direction algorithm converges to an optimal solution $(\hat{\mathcal{A}}, \hat{\mathbf{\Theta}}, \hat{\mathbf{\Pi}}, \hat{\mathbf{\Lambda}}, \hat{\mathbf{\Delta}})$ where $(\hat{\mathcal{A}}, \hat{\mathbf{\Theta}}, \hat{\mathbf{\Pi}})$ is an optimal solution of problem (2.17), $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Delta}}$ are the optimal dual variable.

Before we give the main theorem about the global convergence of Algorithm 1, we need the following lemmas.

Lemma 3.1.1. *If* $(\hat{\mathcal{A}}, \hat{\mathbf{\Theta}}, \hat{\mathbf{\Pi}}, \hat{\mathbf{\Lambda}}, \hat{\mathbf{\Delta}})$ *is optimal to Equation (2.21), the followings hold*

$$\begin{aligned} \frac{1}{({}^1\lambda_\omega)}(-\hat{\mathbf{\Lambda}} - \hat{\mathbf{\Delta}} - \hat{\mathcal{A}} + \bar{\mathcal{A}}) \quad &\in \quad \partial|(\hat{\mathcal{A}})_{mn\omega}|, \\ &\forall \quad m, n = 1, \cdots, N \text{ and } m \neq n, \end{aligned} \tag{3.1}$$

$$(\mathcal{A} - \bar{\mathcal{A}} + \hat{\mathbf{\Lambda}} + \hat{\mathbf{\Delta}})_{mm\omega} = 0, \forall m = 1, \cdots, N, \tag{3.2}$$

$$\hat{\mathbf{\Theta}} = \hat{\mathcal{A}}, \tag{3.3}$$

$$\hat{\mathbf{\Theta}}_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I}, \tag{3.4}$$

$$\langle \hat{\mathbf{\Lambda}}, \mathbf{\Theta} - \hat{\mathbf{\Theta}} \rangle \leq 0, \quad \forall \{ \mathbf{\Theta} : \ \mathbf{\Theta}_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I} \}, \tag{3.5}$$

$$\hat{\mathbf{\Pi}} = \hat{\mathcal{A}}, \tag{3.6}$$

and

$$2({}^2\lambda_{mn})(\boldsymbol{Q}\hat{\boldsymbol{\pi}}_{mn\cdot})_\omega = (\hat{\mathbf{\Delta}})_{mn\omega}. \tag{3.7}$$

Proof of Lemma 1. Since $(\hat{\mathcal{A}}, \hat{\mathbf{\Theta}}, \hat{\mathbf{\Pi}}, \hat{\mathbf{\Lambda}}, \hat{\mathbf{\Delta}})$ is optimal to Equation (2.23), the equality constrains Equation (3.3) and Equation (3.6) and the inequality constrain Equation (3.4) are automatically satisfied, that is

$$\begin{aligned} \hat{\mathbf{\Theta}} &= \hat{\mathcal{A}}, \\ \hat{\mathbf{\Theta}}_{\cdot\cdot\omega} &\geq \epsilon \boldsymbol{I}, \quad \forall \omega, \\ \hat{\mathbf{\Pi}} &= \hat{\mathcal{A}}, \end{aligned}$$

The partial derivative of the augmented Lagrangian function (2.22) with respect to $(\mathcal{A})_{mn\omega}, m \neq n$ and $(\mathcal{A})_{mm\omega}$ are respectively given by,

$$\begin{aligned} \frac{\partial}{\partial(\mathcal{A})_{mn\omega}} L(\mathcal{A}, \mathbf{\Theta}, \mathbf{\Pi}; \mathbf{\Lambda}, \mathbf{\Delta}) &= (\mathcal{A} - \bar{\mathcal{A}})_{mn\omega} + ({}^1\lambda_\omega) \frac{\partial |\boldsymbol{A}_{\cdot\cdot\omega}|_1}{\partial(\mathcal{A})_{mn\omega}} \\ &\quad + (\mathbf{\Lambda})_{mn\omega} + \frac{1}{\mu_1}(\mathcal{A} - \mathbf{\Theta})_{mn\omega} \\ &\quad + (\mathbf{\Delta})_{mn\omega} + \frac{1}{\mu_2}(\mathcal{A} - \mathbf{\Pi})_{mn\omega} \end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial(\mathcal{A})_{mm\omega}}L(\mathcal{A},\mathbf{\Theta},\mathbf{\Pi};\mathbf{\Lambda},\mathbf{\Delta}) &= (\mathcal{A}-\bar{\mathcal{A}})_{mm\omega} \\ &\quad +(\mathbf{\Lambda})_{mm\omega}+\frac{1}{\mu_1}(\mathcal{A}-\mathbf{\Theta})_{mm\omega} \\ &\quad +(\mathbf{\Delta})_{mm\omega}+\frac{1}{\mu_2}(\mathcal{A}-\mathbf{\Pi})_{mm\omega}.\end{aligned}$$

Using the fact that $(\hat{\mathcal{A}},\hat{\mathbf{\Theta}},\hat{\mathbf{\Pi}},\hat{\mathbf{\Lambda}},\hat{\mathbf{\Delta}})$ is an optimal solution to problem (2.17), the optimality conditions of $\mathcal{A}$ can be, respectively, given by

$$\begin{aligned}0 \;\in\; & (\hat{\mathcal{A}}-\bar{\mathcal{A}}+\frac{1}{\mu_1}(\hat{\mathcal{A}}-\hat{\mathbf{\Theta}})+\frac{1}{\mu_2}(\hat{\mathcal{A}}-\hat{\mathbf{\Pi}})+\hat{\mathbf{\Lambda}}+\hat{\mathbf{\Delta}})_{mn\omega} \\ & +({}^1\lambda_\omega)\frac{\partial|\boldsymbol{A}_{\cdot\cdot\omega}|_1}{\partial(\mathcal{A})_{mn\omega}}|_{\hat{\mathcal{A}}}\end{aligned} \tag{3.8}$$

and

$$0=(\hat{\mathcal{A}}-\bar{\mathcal{A}}+\frac{1}{\mu_1}(\hat{\mathcal{A}}-\hat{\mathbf{\Theta}})+\frac{1}{\mu_2}(\hat{\mathcal{A}}-\hat{\mathbf{\Pi}})+\hat{\mathbf{\Lambda}}+\hat{\mathbf{\Delta}})_{mm\omega}. \tag{3.9}$$

By using Equation (3.3) and Equation (3.6), Equation (3.8) and Equation (3.9) can be, respectively, rewritten as

$$\frac{1}{({}^1\lambda_\omega)}(-\hat{\mathbf{\Lambda}}-\hat{\mathbf{\Delta}}-\hat{\mathcal{A}}+\bar{\mathcal{A}})\in\partial|(\hat{\mathcal{A}})_{mn\omega}|.$$

and

$$(\hat{\mathcal{A}}-\bar{\mathcal{A}}+\hat{\mathbf{\Lambda}}+\hat{\mathbf{\Delta}})_{mm\omega}=0.$$

The fact that $(\hat{\mathcal{A}},\hat{\mathbf{\Theta}},\hat{\mathbf{\Pi}},\hat{\mathbf{\Lambda}},\hat{\mathbf{\Delta}})$ is optimal also leads to the following relationship

between $\boldsymbol{\Theta}$ and $\hat{\boldsymbol{\Theta}}$,

$$L(\hat{\mathcal{A}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Pi}}; \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Delta}}) \leq L(\hat{\mathcal{A}}, \boldsymbol{\Theta}, \hat{\boldsymbol{\Pi}}, ; \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Delta}}), \quad \forall \{\boldsymbol{\Theta} : \ \boldsymbol{\Theta}_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I}\}. \tag{3.10}$$

Taylor's theorem states that an expansion of $L(\hat{\mathcal{A}}, \boldsymbol{\Theta}, \hat{\boldsymbol{\Pi}}; \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Delta}})$ about a point $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}$ can be represented by

$$\begin{aligned} & L(\hat{\mathcal{A}}, \boldsymbol{\Theta}, \hat{\boldsymbol{\Pi}}; \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Delta}}) \\ = \ & L(\hat{\mathcal{A}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Pi}}; \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Delta}}) + \langle \frac{\partial L}{\partial \boldsymbol{\Theta}}|_{\hat{\boldsymbol{\Theta}}}, \boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}} \rangle + O(\|\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}}\|_F^2) \\ \approx \ & L(\hat{\mathcal{A}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Pi}}; \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Delta}}) + \langle -\hat{\boldsymbol{\Lambda}} + \frac{1}{\mu_1}(\hat{\boldsymbol{\Theta}} - \hat{\mathcal{A}}), \boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}} \rangle. \end{aligned} \tag{3.11}$$

Now by using (3.3), Equation (3.11) can be rewritten as

$$L(\hat{\mathcal{A}}, \boldsymbol{\Theta}, \hat{\boldsymbol{\Pi}}; \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Delta}}) \approx L(\hat{\mathcal{A}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Pi}}; \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Delta}}) + \langle -\hat{\boldsymbol{\Lambda}}, \boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}} \rangle. \tag{3.12}$$

The summation of Equation (3.10) and Equation (3.12) yields,

$$\langle \hat{\boldsymbol{\Lambda}}, \boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}} \rangle \leq 0, \quad \forall \{\boldsymbol{\Theta} : \ \boldsymbol{\Theta}_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I}\}.$$

Now let us consider the partial derivative of $L(\mathcal{A}, \boldsymbol{\Theta}, \boldsymbol{\Pi}; \boldsymbol{\Lambda}, \boldsymbol{\Delta})$ with respect to $(\boldsymbol{\Pi})_{mn\omega}$, that is

$$\frac{\partial}{\partial (\boldsymbol{\Pi})_{mn\omega}} L(\mathcal{A}, \boldsymbol{\Theta}, \boldsymbol{\Pi}; \boldsymbol{\Lambda}, \boldsymbol{\Delta}) = 2({}^2\lambda_{mn})(\boldsymbol{Q}\boldsymbol{\pi}_{mn\cdot})_\omega - (\boldsymbol{\Delta})_{mn\omega} + \frac{1}{\mu_2}(\boldsymbol{\Pi} - \mathcal{A})_{mn\omega} \tag{3.13}$$

the optimal solution $(\hat{\mathcal{A}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Pi}}, \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Delta}})$ to Equation (2.23) implies

$$2({}^2\lambda_{mn})(\boldsymbol{Q}\hat{\boldsymbol{\pi}}_{mn\cdot})_\omega - (\hat{\boldsymbol{\Delta}})_{mn\omega} + \frac{1}{\mu_2}(\hat{\boldsymbol{\Pi}} - \hat{\mathcal{A}})_{mn\omega} = 0 \tag{3.14}$$

combining Equation (3.14) with Equation (3.6), we can get that

$$2({}^2\lambda_{mn})(\boldsymbol{Q}\hat{\boldsymbol{\pi}}_{mn\cdot})_\omega = (\hat{\boldsymbol{\Delta}})_{mn\omega}$$

□

Now directly using Lemma 1, we are ready to state the following Lemma 2 and prove this lemma.

Lemma 3.1.2. *Assume that* $(\hat{\mathcal{A}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Pi}})$ *is an optimal solution of problem (2.17), and* $\hat{\boldsymbol{\Lambda}}$ *and* $\hat{\boldsymbol{\Delta}}$ *are the corresponding optimal dual variable associated with the equality constraints* $\mathcal{A} = \boldsymbol{\Theta}$ *and* $\mathcal{A} = \boldsymbol{\Pi}$*. Then the sequence* $\{(\mathcal{A}^t, \boldsymbol{\Theta}^t, \boldsymbol{\Pi}^t, \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t)\}_t$ *produced by Algorithm 1 satisfies*

$$\|\mathcal{U}^t - \hat{\mathcal{U}}\|_G^2 - \|\mathcal{U}^{t+1} - \hat{\mathcal{U}}\|_G^2 \geq \|\mathcal{U}^t - \mathcal{U}^{t+1}\|_G^2, \tag{3.15}$$

where $\hat{\mathcal{U}}$ *and* $\mathcal{U}^t$ *are* $6N \times 2N \times T$ *tensors, respectively defined as* $\hat{\mathcal{U}} = (\hat{\mathcal{A}}, \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Delta}})^T$ *and* $\mathcal{U}^t = (\mathcal{A}^t, \boldsymbol{\Theta}^t, \boldsymbol{\Delta}^t)^T$*.*

Proof of lemma 2. The first subproblem in Algorithm 1 is to update $\boldsymbol{\Pi}$ in the tth iteration, that is,

$$L(\mathcal{A}^t, \boldsymbol{\Theta}^t, \boldsymbol{\Pi}^{t+1}; \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t) = \min_{\Pi} L(\mathcal{A}^t, \boldsymbol{\Theta}^t, \boldsymbol{\Pi}; \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t). \tag{3.16}$$

Using the Taylor expansion of $(\mathcal{A}^t, \boldsymbol{\Theta}^t, \boldsymbol{\Pi}^{t+1}; \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t)$ about the point $\boldsymbol{\Pi} = \boldsymbol{\Pi}^{t+1}$, the optimality condition for Equation (3.16) is equivalent to

$$2({}^2\lambda_{mn})(\boldsymbol{Q}\boldsymbol{\pi}_{mn\cdot}^{t+1})_\omega - (\boldsymbol{\Delta}^t)_{mn\omega} + \frac{1}{\mu_2}(\boldsymbol{\Pi}^{t+1} - \mathcal{A}^t)_{mn\omega} = 0. \tag{3.17}$$

Now using the updating formula for $\boldsymbol{\Delta}^t$, that is,

$$\boldsymbol{\Delta}^{t+1} = \boldsymbol{\Delta}^t - \frac{1}{\mu_2}(\boldsymbol{\Pi}^{t+1} - \mathcal{A}^{t+1}). \quad Eq.(2.28)$$

Equation (3.17) can be rewritten as

$$2({}^2\lambda_{mn})(\boldsymbol{Q}\boldsymbol{\pi}_{mn\cdot}^{t+1})_\omega = (\boldsymbol{\Delta}^{t+1} - \frac{1}{\mu_2}(\mathcal{A}^{t+1} - \mathcal{A}^t))_{mn\omega} \quad . \tag{3.18}$$

Using the fact that the penalty matrix $\boldsymbol{Q}$ is positive definite, the combination of Equation (3.7) and Equation (3.18) implies that

$$\langle(\boldsymbol{\Delta}^{t+1} - \hat{\boldsymbol{\Delta}}_+^+ \frac{1}{\mu_2}(\mathcal{A}^t - \mathcal{A}^{t+1}), \boldsymbol{\Pi}^{t+1} - \hat{\boldsymbol{\Pi}}^+\rangle \geq 0. \tag{3.19}$$

Secondly, we derive the optimality condition for the second subproblem in Algorithm 1, that is, to update $\boldsymbol{\Theta}$ in the tth iteration, that is,

$$L(\mathcal{A}^t, \boldsymbol{\Theta}^{t+1}, \boldsymbol{\Pi}^{t+1}; \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t) = \min_{\{\Theta:\Theta_{\cdot\cdot\omega} \geq \epsilon I\}} L(\mathcal{A}^t, \boldsymbol{\Theta}, \boldsymbol{\Pi}^{t+1}; \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t). \tag{3.20}$$

Using the Taylor expansion of $L(\mathcal{A}^t, \boldsymbol{\Theta}, \boldsymbol{\Pi}^{t+1}; \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t)$ about the point $\boldsymbol{\Theta} = \boldsymbol{\Theta}^{t+1}$, the optimality condition for minimizing Equation (3.20) at the point $\boldsymbol{\Theta}^{t+1}$ must satisfy the following inequality

$$\langle -\boldsymbol{\Lambda}^t + \frac{1}{\mu_1}(\boldsymbol{\Theta}^{t+1} - \mathcal{A}^t), \boldsymbol{\Theta} - \boldsymbol{\Theta}^{t+1}\rangle \geq 0, \ \forall\{\boldsymbol{\Theta} : \boldsymbol{\Theta}_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I}\}. \tag{3.21}$$

By plugging in the updating formula for $\boldsymbol{\Lambda}^t$, given by Equation (2.27), Equation

(3.21) can be rewritten as

$$\langle \mathbf{\Lambda}^{t+1} - \frac{1}{\mu_1}(\mathcal{A}^{t+1} - \mathcal{A}^t), \mathbf{\Theta} - \mathbf{\Theta}^{t+1} \rangle \leq 0, \ \forall \{\mathbf{\Theta} : \mathbf{\Theta}_{..\omega} \geq \epsilon \boldsymbol{I}\}. \tag{3.22}$$

Now letting $\mathbf{\Theta} = \mathbf{\Theta}^{t+1}$ in Equation (3.5) and $\mathbf{\Theta} = \hat{\mathbf{\Theta}}^{+}$ in Equation (3.21), (3.5) and (3.21) can be, respectively, represented as

$$\langle \hat{\mathbf{\Lambda}}^{+}, \mathbf{\Theta}^{t+1} - \hat{\mathbf{\Theta}}^{+} \rangle \leq 0. \tag{3.23}$$

and

$$\langle \mathbf{\Lambda}^{t+1} - \frac{1}{\mu_1}(\mathcal{A}^{t+1} - \mathcal{A}^t), \hat{\mathbf{\Theta}}^{+} - \mathbf{\Theta}^{t+1} \rangle \leq 0. \tag{3.24}$$

The summation of (3.23) and (3.24) leads to the following inequality,

$$\langle (\mathbf{\Lambda}^{t+1} - \hat{\mathbf{\Lambda}}^{+}) + \frac{1}{\mu_1}(\mathcal{A}^t - \mathcal{A}^{t+1}), \mathbf{\Theta}^{t+1} - \hat{\mathbf{\Theta}}^{+} \rangle \geq 0. \tag{3.25}$$

Thirdly, the optimality conditions for the third subproblem in Algorithm 1, that is, the optimal minimizer of $L(\mathcal{A}, \mathbf{\Theta}^{t+1}, \mathbf{\Pi}^{t+1}; \mathbf{\Lambda}^t, \mathbf{\Delta}^t)$ in the tth iteration

$$\mathcal{A}^{t+1} = \arg\min_{\mathcal{A}} L(\mathcal{A}, \mathbf{\Theta}^{t+1}, \mathbf{\Pi}^{t+1}; \mathbf{\Lambda}^t, \mathbf{\Delta}^t). \tag{3.26}$$

Using the Taylor expansion of $L(\mathcal{A}, \mathbf{\Theta}^{t+1}, \mathbf{\Pi}^{t+1}; \mathbf{\Lambda}^t, \mathbf{\Delta}^t)$ about the point $\mathcal{A} = \mathcal{A}^{t+1}$, Equation (3.26) is equivalent to the following conditions for the off-diagonal elements

and the diagonal elements, respectively,

$$\begin{aligned} 0 \quad \in \quad & (\mathcal{A}^{t+1} - \bar{\mathcal{A}})_{mn\omega} + ({}^1\lambda_\omega)\partial|(\mathcal{A}^{t+1})_{mn\omega}| + (\mathbf{\Lambda}^t)_{mn\omega} + (\mathbf{\Delta}^t)_{mn\omega} \\ & + \frac{1}{\mu_1}(\mathcal{A}^{t+1} - \mathbf{\Theta}^{t+1})_{mn\omega} + \frac{1}{\mu_2}(\mathcal{A}^{t+1} - \mathbf{\Pi}^{t+1})_{mn\omega}, \\ & \forall m, n = 1, \cdots, N, m \neq n \; and \; \omega = 1/T, \cdots, 1. \end{aligned} \tag{3.27}$$

and

$$\begin{aligned} 0 \quad = \quad & ((\mathcal{A}^{t+1} - \bar{\mathcal{A}}) + \mathbf{\Lambda}^t + \mathbf{\Delta}^t + \frac{1}{\mu_1}(\mathcal{A}^{t+1} - \mathbf{\Theta}^{t+1}) + \frac{1}{\mu_2}(\mathcal{A}^{t+1} - \mathbf{\Pi}^{t+1}))_{mm\omega} \\ & \forall m, n = 1, \cdots, N; \omega = 1/T, \cdots, 1. \end{aligned} \tag{3.28}$$

Now plugging in the updating formulae for $\mathbf{\Lambda}^t$ and $\mathbf{\Delta}^t$, that is, Equation (2.27) and Equation (2.28), Equation (3.27) and Equation (3.28) can be, respectively, rewritten as

$$\frac{1}{({}^1\lambda_\omega)}(-(\mathcal{A}^{t+1} - \bar{\mathcal{A}}) - \mathbf{\Lambda}^{t+1} - \mathbf{\Delta}^{t+1})_{mn\omega} \in \partial|(\mathcal{A}^{t+1})_{mn\omega}| \tag{3.29}$$

and

$$(-(\mathcal{A}^{t+1} - \bar{\mathcal{A}}) - \mathbf{\Lambda}^{t+1} - \mathbf{\Delta}^{t+1})_{mm\omega} = 0. \tag{3.30}$$

Using the fact that the partial differential function $\partial|\cdot|$ is monotone, Equation (3.1) and Equation (3.29) yield that

$$\langle -(\mathbf{\Lambda}^{t+1} - \hat{\mathbf{\Lambda}}^+) - (\mathbf{\Delta}^{t+1} - \hat{\mathbf{\Delta}}^+) - (\mathcal{A}^{t+1} - \hat{\mathcal{A}}^+), \mathcal{A}^{t+1} - \hat{\mathcal{A}}^+ \rangle \geq 0. \tag{3.31}$$

And the summation of Equation (3.2) and Equation (3.30) leads to the following

equality that

$$\langle -(\mathbf{\Lambda}^{t+1} - \hat{\mathbf{\Lambda}}^{+}) - (\mathbf{\Delta}^{t+1} - \hat{\mathbf{\Delta}}^{+}) - (\mathcal{A}^{t+1} - \hat{\mathcal{A}}^{+}), \mathcal{A}^{t+1} - \hat{\mathcal{A}}^{+} \rangle = 0. \tag{3.32}$$

In fact, Equation (3.32) is a special case of Equation (3.31), therefore Equations (3.32) and (3.31) can unite into Equation (3.31).

Summing Equation (3.19), Equation (3.25) and Equation (3.31) gives that

$$\begin{aligned} & \|\mathcal{A}^{t+1} - \hat{\mathcal{A}}^{+}\|_F^2 + \langle \mathbf{\Lambda}^{t+1} - \hat{\mathbf{\Lambda}}^{+}, \mathcal{A}^{t+1} - \hat{\mathcal{A}}^{+} \rangle + \langle \mathbf{\Delta}^{t+1} - \hat{\mathbf{\Delta}}^{+}, \mathcal{A}^{t+1} - \hat{\mathcal{A}}^{+} \rangle \\ \leq \; & \langle \mathbf{\Lambda}^{t+1} - \hat{\mathbf{\Lambda}}^{+}, \mathbf{\Theta}^{t+1} - \hat{\mathbf{\Theta}}^{+} \rangle + \frac{1}{\mu_1} \langle \mathcal{A}^{t} - \mathcal{A}^{t+1}, \mathbf{\Theta}^{t+1} - \hat{\mathbf{\Theta}}^{+} \rangle \\ & + \langle \mathbf{\Delta}^{t+1} - \hat{\mathbf{\Delta}}^{+}, \mathbf{\Pi}^{t+1} - \hat{\mathbf{\Pi}}^{+} \rangle + \frac{1}{\mu_2} \langle \mathcal{A}^{t} - \mathcal{A}^{t+1}, \mathbf{\Pi}^{t+1} - \hat{\mathbf{\Pi}}^{+} \rangle. \end{aligned} \tag{3.33}$$

Now by respectively rewriting the updating formulae (2.27) and (2.28) as

$$\mathbf{\Theta}^{t+1} = \mu_1 (\mathbf{\Lambda}^{t} - \mathbf{\Lambda}^{t+1}) + \mathcal{A}^{t+1}. \tag{3.34}$$

and

$$\mathbf{\Delta}^{t+1} = \mu_2 (\mathbf{\Delta}^{t} - \mathbf{\Delta}^{t+1}) + \mathcal{A}^{t+1}. \tag{3.35}$$

as well as using Equations (3.34), Equation (3.35) and the equality constrains $\hat{\mathbf{\Theta}}^{+} =$

$\hat{\mathcal{A}}^+$ and $\hat{\mathbf{\Pi}}^+ = \hat{\mathcal{A}}^+$, Equation (3.33) can be represented as

$$\begin{aligned}
& \|\mathcal{A}^{t+1} - \hat{\mathcal{A}}^+\|_F^2 \\
\leq\ & -\langle \mathbf{\Lambda}^{t+1} - \hat{\mathbf{\Lambda}}^+, \mathcal{A}^{t+1} - \hat{\mathcal{A}}^+ \rangle - \langle \mathbf{\Delta}^{t+1} - \hat{\mathbf{\Delta}}^+, \mathcal{A}^{t+1} - \hat{\mathcal{A}}^+ \rangle \\
& + \langle \mathbf{\Lambda}^{t+1} - \hat{\mathbf{\Lambda}}^+, \mu_1(\mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1}) + \mathcal{A}^{t+1} - \hat{\mathcal{A}}^+ \rangle \\
& + \frac{1}{\mu_1} \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mu_1(\mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1}) + \mathcal{A}^{t+1} - \hat{\mathcal{A}}^+ \rangle \\
& + \langle \mathbf{\Delta}^{t+1} - \hat{\mathbf{\Delta}}^+, \mu_2(\mathbf{\Delta}^t - \mathbf{\Delta}^{t+1}) + \mathcal{A}^{t+1} - \hat{\mathcal{A}}^+ \rangle \\
& + \frac{1}{\mu_2} \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mu_2(\mathbf{\Delta}^t - \mathbf{\Delta}^{t+1}) + \mathcal{A}^{t+1} - \hat{\mathcal{A}}^+ \rangle. \qquad (3.36)
\end{aligned}$$

Then using algebraic derivation, we can simplify Equation (3.36) as

$$\begin{aligned}
& \|\mathcal{A}^{t+1} - \hat{\mathcal{A}}^+\|_F^2 \\
\leq\ & \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1} \rangle + \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1} \rangle \\
& + \mu_1 \langle \mathbf{\Lambda}^{t+1} - \hat{\mathbf{\Lambda}}^+, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1} \rangle + \mu_2 \langle \mathbf{\Delta}^{t+1} - \hat{\mathbf{\Delta}}^+, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1} \rangle \\
& + (\frac{1}{\mu_1} + \frac{1}{\mu_2}) \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathcal{A}^{t+1} - \hat{\mathcal{A}}^+ \rangle. \qquad (3.37)
\end{aligned}$$

By rewriting $\hat{\mathbf{\Lambda}}^+ - \mathbf{\Lambda}^{t+1}$, $\hat{\mathbf{\Delta}}^+ - \mathbf{\Delta}^{t+1}$ and $\hat{\mathcal{A}}^+ - \mathcal{A}^{t+1}$ as

$$\hat{\mathbf{\Lambda}}^+ - \mathbf{\Lambda}^{t+1} = (\hat{\mathbf{\Lambda}}^+ - \mathbf{\Lambda}^t) + (\mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1}) \qquad (3.38)$$

$$\hat{\mathbf{\Delta}}^+ - \mathbf{\Delta}^{t+1} = (\hat{\mathbf{\Delta}}^+ - \mathbf{\Delta}^t) + (\mathbf{\Delta}^t - \mathbf{\Delta}^{t+1}) \qquad (3.39)$$

$$\hat{\mathcal{A}}^+ - \mathcal{A}^{t+1} = (\hat{\mathcal{A}}^+ - \mathcal{A}^t) + (\mathcal{A}^t - \mathcal{A}^{t+1}). \qquad (3.40)$$

we can plug Equations (3.38)-(3.40) into Equation (3.37), and then derive the following representation referring to the definition of the Frobenius norm of tensors in

Section 2.1,

$$
\begin{aligned}
& \|\mathcal{A}^{t+1} - \hat{\mathcal{A}}^{+}\|_F^2 \\
\leq\ & \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1} \rangle + \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1} \rangle \\
& + \mu_1 \langle \mathbf{\Lambda}^t - \hat{\mathbf{\Lambda}}^{+}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1} \rangle - \mu_1 \|\mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1}\|_F^2 \\
& + \mu_2 \langle \mathbf{\Delta}^t - \hat{\mathbf{\Delta}}^{+}, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1} \rangle - \mu_2 \|\mathbf{\Delta}^t - \mathbf{\Delta}^{t+1}\|_F^2 \\
& + (\frac{1}{\mu_1} + \frac{1}{\mu_2}) \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathcal{A}^t - \hat{\mathcal{A}}^{+} \rangle - (\frac{1}{\mu_1} + \frac{1}{\mu_2}) \|\mathcal{A}^t - \mathcal{A}^{t+1}\|_F^2. \qquad (3.41)
\end{aligned}
$$

rearranging the right-hand side of Equation (3.41), we can get that

$$
\begin{aligned}
& \mu_1 \|\mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1}\|_F^2 + \mu_2 \|\mathbf{\Delta}^t - \mathbf{\Delta}^{t+1}\|_F^2 + (\frac{1}{\mu_1} + \frac{1}{\mu_2}) \|\mathcal{A}^t - \mathcal{A}^{t+1}\|_F^2 \\
& + \|\mathcal{A}^{t+1} - \hat{\mathcal{A}}^{+}\|_F^2 - \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1} \rangle - \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1} \rangle \\
\leq\ & \mu_1 \langle \mathbf{\Lambda}^t - \hat{\mathbf{\Lambda}}^{+}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1} \rangle + \mu_2 \langle \mathbf{\Delta}^t - \hat{\mathbf{\Delta}}^{+}, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1} \rangle \\
& + (\frac{1}{\mu_1} + \frac{1}{\mu_2}) \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathcal{A}^t - \hat{\mathcal{A}}^{+} \rangle. \qquad (3.42)
\end{aligned}
$$

Using the notation of $\mathcal{U}^{t+1}$ and $\hat{\mathcal{U}}$, Equation (3.42) can be rewritten as

$$
\begin{aligned}
& \|\mathcal{U}^t - \mathcal{U}^{t+1}\|_G^2 + \|\mathcal{A}^{t+1} - \hat{\mathcal{A}}\|_F^2 \\
& - \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1} \rangle - \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1} \rangle \\
\leq\ & \langle \mathcal{U}^t - \hat{\mathcal{U}}, \mathcal{U}^t - \mathcal{U}^{t+1} \rangle_G. \qquad (3.43)
\end{aligned}
$$

It is known that for the $\mathcal{G}$ norm $\|\cdot\|_G$ the following identity holds,

$$
\|\mathcal{U}^{t+1} - \hat{\mathcal{U}}\|_G^2 = \|\mathcal{U}^{t+1} - \mathcal{U}^t\|_G^2 - 2\langle \mathcal{U}^t - \hat{\mathcal{U}}, \mathcal{U}^t - \mathcal{U}^{t+1} \rangle)_G + \|\mathcal{U}^t - \hat{\mathcal{U}}\|_G^2. \qquad (3.44)
$$

Using Equation (3.44), Equation (3.43) can be rewritten at

$$\begin{aligned}
& \|\mathcal{U}^t - \hat{\mathcal{U}}\|_G^2 - \|\mathcal{U}^{t+1} - \hat{\mathcal{U}}\|_G^2 \\
= \; & 2\langle \mathcal{U}^t - \hat{\mathcal{U}}, \mathcal{U}^t - \mathcal{U}^{t+1}\rangle)_G - \|\mathcal{U}^{t+1} - \mathcal{U}^t\|_G^2 \\
\geq \; & 2\|\mathcal{U}^t - \mathcal{U}^{t+1}\|_G^2 + 2\|\mathcal{A}^{t+1} - \hat{\mathcal{A}}\|_F^2 - \|\mathcal{U}^{t+1} - \mathcal{U}^t\|_G^2 \\
& -2\langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1}\rangle - 2\langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1}\rangle \\
= \; & \|\mathcal{U}^t - \mathcal{U}^{t+1}\|_G^2 + 2\|\mathcal{A}^{t+1} - \hat{\mathcal{A}}\|_F^2 \\
& -2\langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1}\rangle - 2\langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1}\rangle. \qquad (3.45)
\end{aligned}$$

At last, we will prove that

$$\|\mathcal{A}^{t+1} - \hat{\mathcal{A}}\|_F^2 - \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1}\rangle - \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1}\rangle \geq 0. \qquad (3.46)$$

Then the summation of Equation (3.45) and Equation (3.46) implies

$$\|\mathcal{U}^t - \hat{\mathcal{U}}\|_G^2 - \|\mathcal{U}^{t+1} - \hat{\mathcal{U}}\|_G^2 \geq \|\mathcal{U}^t - \mathcal{U}^{t+1}\|_G^2.$$

Using Equation (3.29) and Equation (3.30) for t instead of $t+1$, they can be rewritten respectively as,

$$\frac{1}{({}^1\lambda_\omega)}(-(\mathcal{A}^t - \bar{\mathcal{A}}) - \mathbf{\Lambda}^t - \mathbf{\Delta}^t)_{mn\omega} \in \partial|(\mathcal{A}^t)_{mn\omega}| \qquad (3.47)$$

and

$$(-(\mathcal{A}^t - \bar{\mathcal{A}}) - \mathbf{\Lambda}^t - \mathbf{\Delta}^t)_{mm\omega} = 0. \qquad (3.48)$$

The combination of Equations (3.29), (3.30), (3.47) and (3.48) and using the fact

that the partial differential function $\partial|\cdot|$ is a monotone function yields the following inequality

$$\langle -(\mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1}) - (\mathbf{\Delta}^t - \mathbf{\Delta}^{t+1}) - (\mathcal{A}^t - \mathcal{A}^{t+1}), \mathcal{A}^t - \mathcal{A}^{t+1} \rangle \geq 0. \tag{3.49}$$

Using the Frobenius norm $\|\cdot\|_F$, Equation (3.49) can be represented as

$$-\langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1} + \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1} \rangle \geq \|\mathcal{A}^t - \mathcal{A}^{t+1}\|_F^2.$$

Equation (3.50) implies that

$$\begin{aligned} & \|\mathcal{A}^{t+1} - \hat{\mathcal{A}}\|_F^2 - \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1} \rangle - \langle \mathcal{A}^t - \mathcal{A}^{t+1}, \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1} \rangle \\ \geq & \|\mathcal{A}^{t+1} - \hat{\mathcal{A}}\|_F^2 + \|\mathcal{A}^t - \mathcal{A}^{t+1}\|_F^2 \\ \geq & 0. \end{aligned}$$

□

3.2 Convergence of Algorithm 1

Now we are ready to give the main convergence result of Algorithm 1 as follows.

Theorem 3.2.1. *The sequence* $(\mathcal{A}^t, \mathbf{\Theta}^t, \mathbf{\Pi}^t, \mathbf{\Lambda}^t, \mathbf{\Delta}^t)$ *produced by Algorithm 1 from any starting point converges to an optimal solution of problem (2.17).*

Proof of Theorem 3.2.1. From Lemma 2, we can easily derive that

(1) $\|\mathcal{U}^t - \mathcal{U}^{t+1}\|_G \to 0$;

(2) $\{\mathcal{U}^t\}$ lies in a compact region;

(3) $\|\mathcal{U}^t - \hat{\mathcal{U}}\|_G^2$ is monotonically non-increasing and thus converges.

Using (1) and the notation of $\mathcal{U}^t$ and $\mathcal{U}^{t+1}$, we can directly get that

$$\begin{aligned} \mathcal{A}^t - \mathcal{A}^{t+1} &\rightarrow 0, \quad (3.50) \\ \mathbf{\Lambda}^t - \mathbf{\Lambda}^{t+1} &\rightarrow 0, \quad (3.51) \\ \mathbf{\Delta}^t - \mathbf{\Delta}^{t+1} &\rightarrow 0. \quad (3.52) \end{aligned}$$

Then using the updating formulae for $\mathbf{\Lambda}$ and $\mathbf{\Delta}$ in Algorithm 1, that is Equation (2.27) and Equation (2.28), we can obtain

$$\begin{aligned} \mathbf{\Theta}^t - \mathbf{\Theta}^{t+1} &\rightarrow 0, \\ \mathbf{\Theta}^t - \mathcal{A}^t &\rightarrow 0. \quad (3.53) \end{aligned}$$

and

$$\begin{aligned} \mathbf{\Pi}^t - \mathbf{\Pi}^{t+1} &\rightarrow 0, \\ \mathbf{\Pi}^t - \mathcal{A}^t &\rightarrow 0. \end{aligned}$$

From (2), we obtain that $\mathcal{U}^t$ contains a subsequence $\mathcal{U}^{t_l}$ that converges to $\mathcal{U}^*$, that is,

$$\begin{aligned} \mathcal{A}^{t_l} &\rightarrow \mathcal{A}^*, \\ \mathbf{\Lambda}^{t_l} &\rightarrow \mathbf{\Lambda}^*, \\ \mathbf{\Delta}^{t_l} &\rightarrow \mathbf{\Delta}^*. \end{aligned}$$

From Equation (3.53), we get that

$$\boldsymbol{\Theta}^{t_l} \rightarrow \boldsymbol{\Theta}^* = \mathcal{A}^*. \tag{3.54}$$

Similarly, from Equation (3.54), we obtain that

$$\boldsymbol{\Pi}^{t_l} \rightarrow \boldsymbol{\Pi}^* = \mathcal{A}^*. \tag{3.55}$$

Therefore, $(\mathcal{A}^*, \boldsymbol{\Theta}^*, \boldsymbol{\Pi}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Delta}^*)$ is a limit point of $\{(\mathcal{A}^t, \boldsymbol{\Theta}^t, \boldsymbol{\Pi}^t, \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t)\}$.

Note that using Equation (3.29) and Equation (3.30) for $*$ instead of $t+1$, they can be respectively represented as,

$$\frac{1}{({}^1\lambda_\omega)}(-(\mathcal{A}^* - \bar{\mathcal{A}}) - \boldsymbol{\Lambda}^* - \boldsymbol{\Delta}^*)_{mn\omega} \in \partial|(\mathcal{A}^*)_{mn\omega}| \tag{3.56}$$

and

$$(-(\mathcal{A}^* - \bar{\mathcal{A}}) - \boldsymbol{\Lambda}^* - \boldsymbol{\Delta}^*)_{mm\omega} = 0. \tag{3.57}$$

Using Equation (3.22) for t_l instead of t, we have

$$\langle \boldsymbol{\Lambda}^{t_l+1} - \frac{1}{\mu_1}(\mathcal{A}^{t_l+1} - \mathcal{A}^{t_l}), \boldsymbol{\Theta} - \boldsymbol{\Theta}^{t_l+1} \rangle \leq 0, \quad \forall \{\boldsymbol{\Theta} : \boldsymbol{\Theta}_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I}\}. \tag{3.58}$$

By letting Equation (3.58) go to the limit, we can get that

$$\langle \boldsymbol{\Lambda}^*, \boldsymbol{\Theta} - \boldsymbol{\Theta}^* \rangle \leq 0, \quad \forall \{\boldsymbol{\Theta} : \boldsymbol{\Theta}_{\cdot\cdot\omega} \geq \epsilon \boldsymbol{I}\}. \tag{3.59}$$

Using Equation

$$\langle \boldsymbol{\Lambda}^*, \boldsymbol{\Theta} - \boldsymbol{\Theta}^* \rangle \leq 0, \quad \forall \{\boldsymbol{\Theta} : \boldsymbol{\Theta}_{\cdot\omega} \geq \epsilon \boldsymbol{I}\}. \tag{3.60}$$

Equations (3.56), (3.57), (3.59) and (3.60) together with the equalities $\boldsymbol{\Theta}^* = \mathcal{A}^*$ and $\boldsymbol{\Delta}^* = \mathcal{A}^*$ mean that $(\mathcal{A}^*, \boldsymbol{\Theta}^*, \boldsymbol{\Pi}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Delta}^*)$ is an optimal solution to problem (2.17). Therefore, we show that any limit point of $\{(\mathcal{A}^t, \boldsymbol{\Theta}^t, \boldsymbol{\Pi}^t, \boldsymbol{\Lambda}^t, \boldsymbol{\Delta}^t)\}_t$ is an optimal solution to problem (2.17). □

4. ADCP DATA AND FORWARD MODEL

4.1 Introduction

ADCP is used to measure both waves and currents that can provide subsurface velocity data beneath ocean waves. Our purpose is to use subsurface velocity data to estimate the frequency-directional distribution of ocean waves. The typical ADCP has 4-direction beams and m-sensor on each beam, see Fig.(4.1), where m depends on the width of sampling interval and for our data $m = 25$. There are 4 sensors in one horizontal plane which we call as layer. Our observations are time series

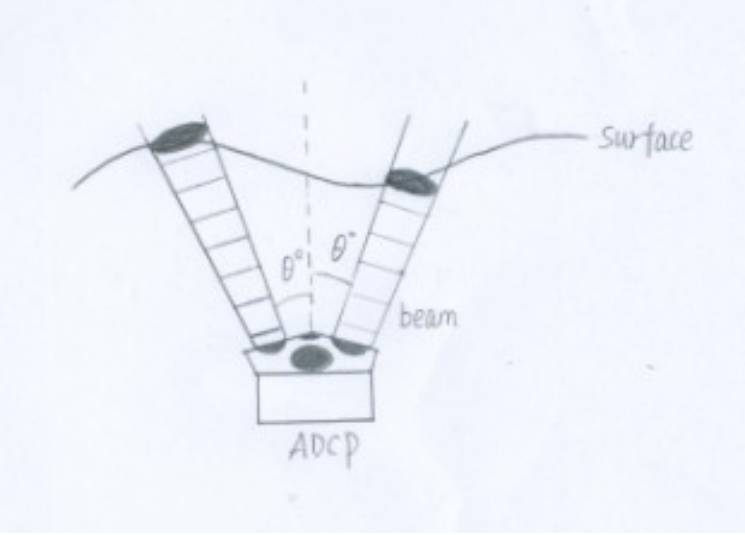

Figure 4.1: Single bottom-mounted acoustic doppler current profiler.

data of subsurface velocities beneath ocean waves at a number of spatial locations ($4m$). Hence the set of sensors constitutes a spatial array, see Fig.(4.2), although it is extremely sparse, and has a highly non-uniform lag distribution. Also, since the

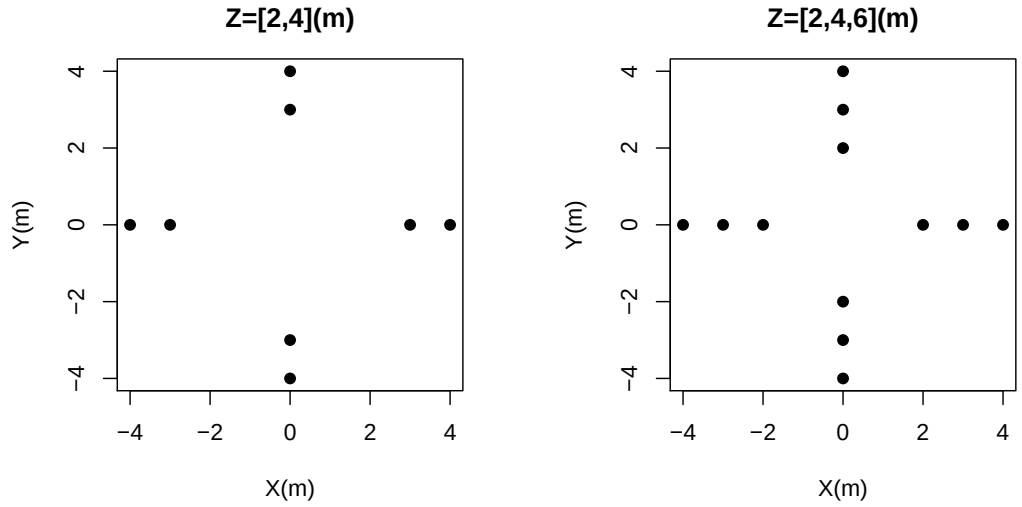

Figure 4.2: The sensor array of single bottom-mounted acoustic doppler current profiler.

sensors are located at different depths, the signal-to-noise ratio varies from sensor to sensor.

Now let us see the data frame of interest. The observations we obtain are subsurface velocities beneath ocean waves. The location of the n^{th} sensor is represented as a 3-dimension vector, $(\boldsymbol{x}_n, z_n)$, where $\boldsymbol{x}_n = (x_1^{(n)}, x_2^{(n)})$, which is a 2-dimension vector in horizontal plane, and z_n denotes the vertical distance between ADCP and the n^{th} sensor.

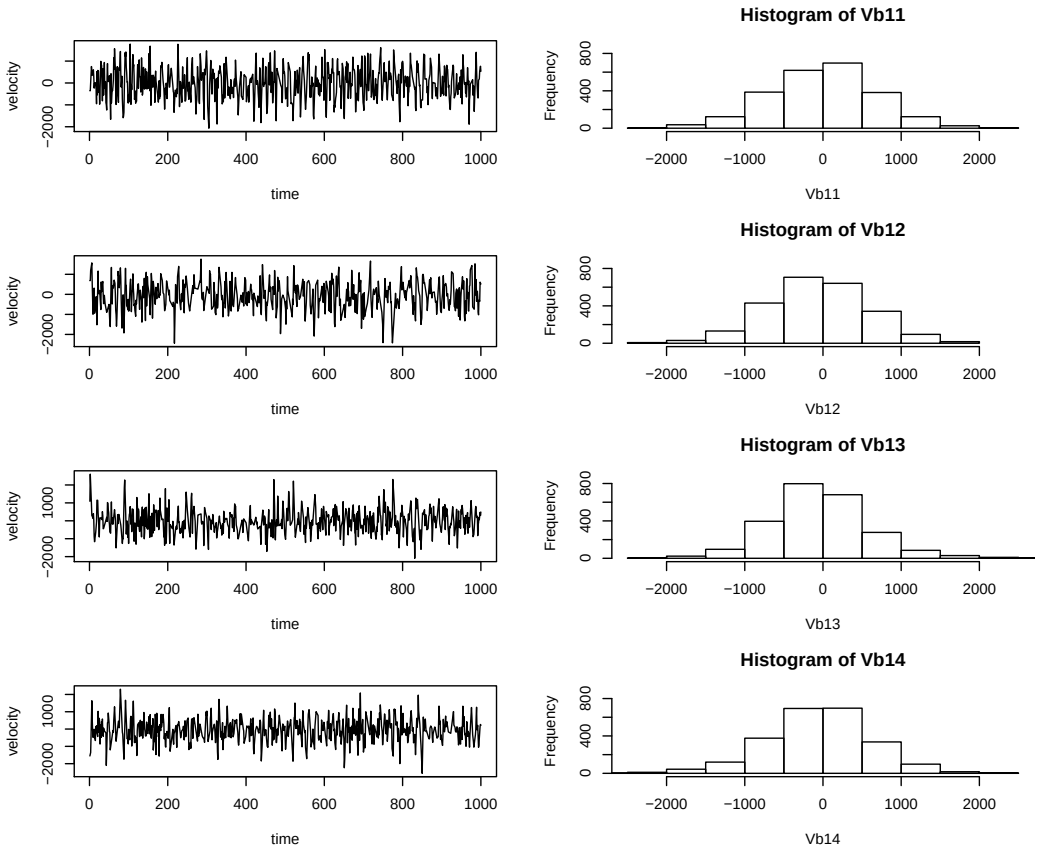

Figure 4.3: The comparison of 4 profiles within the same depth in shallow water.

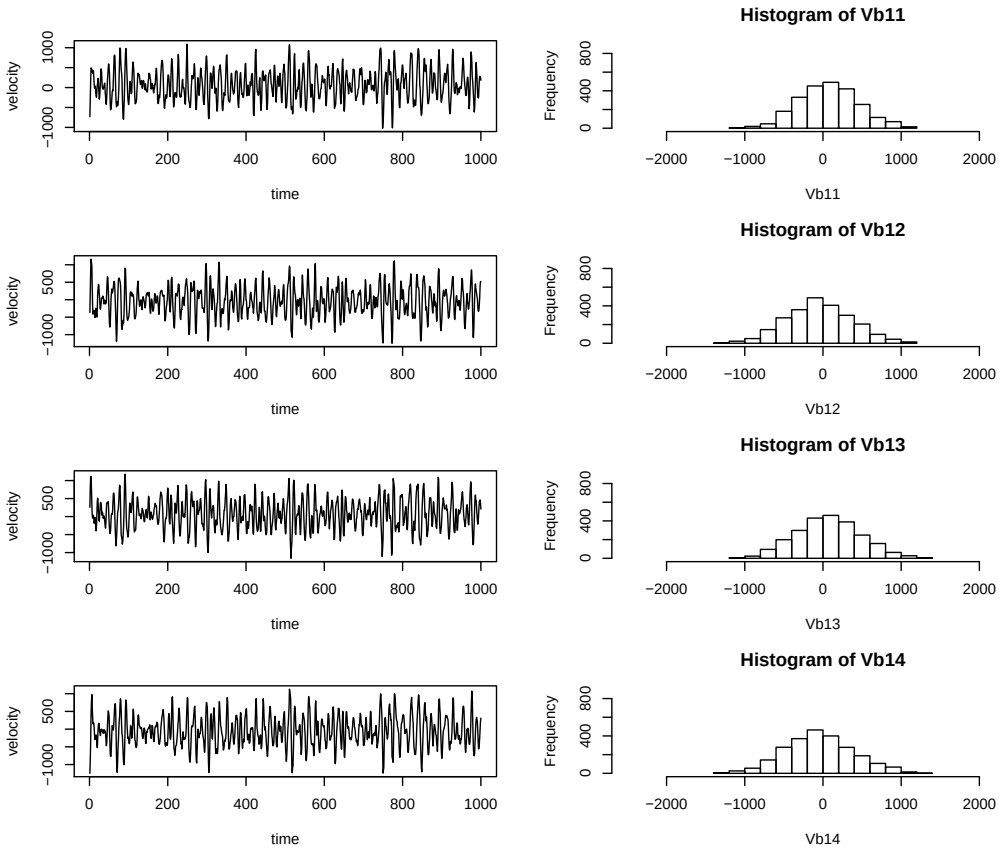

Figure 4.4: The comparison of 4 profiles within the same depth in middle depth water.

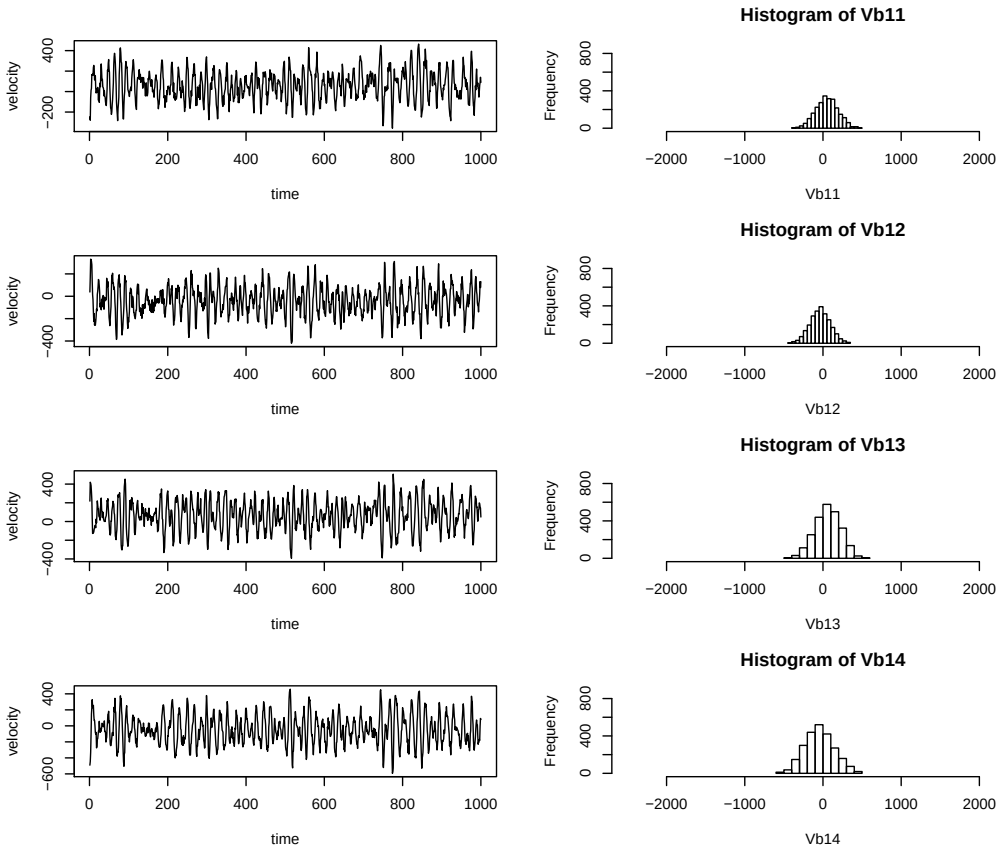

Figure 4.5: The comparison of 4 profiles within the same depth in deep water.

For the n^{th} sensor located at $(\boldsymbol{x}_n, z_n)$, the response is a time series, $\xi(\boldsymbol{x}_n, t; z_n)$ where t is the sampling time and $t = 1, \ldots, N$. However the goal in Oceanography is to find the surface displacement of the waves at a fixed location in horizontal plane, denoted by $\eta(\boldsymbol{x}_n, t)$.

The relation between $\xi(\boldsymbol{x}_n, z_n, t)$ and $\eta(\boldsymbol{x}_n, t)$ is represented by the spectral transfer function $H(\omega, \theta, \boldsymbol{x}_n, z_n)$ which models the trend that how velocities vary from the n^{th} sensor to a plane wave of radian frequency ω traveling in the direction θ. The column vector $\mathbf{H}(\omega, \theta)$ has entries $H(\omega, \theta, \mathbf{x}_n, z_n)$. Figure 4.3, Figure 4.4 and Figure

4.5 show the observations of wave profile in three different water depth.

We want to infer the frequency-directional characteristics of the surface displacement $\zeta(\mathbf{x}, t)$ from measurements of a linearly-related interior field, $u(\mathbf{x}, z, t)$, taken at various sensor locations $(\mathbf{x}_n, z_n)$. This idea arises the forward model as follows: denoting the cross-spectral matrix between the various sensors by $C(\omega)$ and the frequency-directional distribution of the surface waves by the scalar function $D(\omega, \theta)$, then the forward relation is

$$C(\omega) = \int_{-\pi}^{+\pi} \mathbf{H}(\omega, \theta) D(\omega, \theta) \mathbf{H}^*(\omega, \theta) d\theta \tag{4.1}$$

where $H^*(\omega, \theta)$ denotes the complex-conjugate transpose function of $H(\omega, \theta)$.

4.2 Characteristic of Surface Ocean Waves

The purpose of analysis of ADCP data is to study the characteristics of surface ocean wave, such as height and direction. However, the transfer function between the surface displacement $\eta(\mathbf{x}, t)$ and the interior velocity $\xi(\mathbf{x}, z, t)$ is defined on frequency ω and direction θ. Let us first see the definition of the transfer function defined in wave-number spectral space in the following.

For surface displacement $\eta(\boldsymbol{x}, t)$, we can write its three-dimensional Fourier transform $\check{\eta}(\boldsymbol{k}, \omega)$ as

$$\check{\eta}(\boldsymbol{k}, \omega) = \int_{\boldsymbol{x}} \int_t e^{i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)} \eta(\boldsymbol{x}, t) dt d\boldsymbol{x} \tag{4.2}$$

Inversely,

$$\begin{aligned} \eta(\boldsymbol{x}, t) &= \frac{1}{(2\pi)^3} \int_{\boldsymbol{k}} \int_\omega e^{-i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)} \check{\eta}(\boldsymbol{k}, \omega) d\omega d\boldsymbol{k} \\ &= \int_{\boldsymbol{k}} \int_\omega e^{-i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)} dA(\boldsymbol{k}, \omega) \end{aligned} \tag{4.3}$$

where

$$E[dA(\boldsymbol{k},\omega)dA^*(\boldsymbol{k}',\omega')] = \begin{cases} S(\boldsymbol{k},\omega)d\boldsymbol{k}d\omega & \text{if } \boldsymbol{k}=\boldsymbol{k}' \text{and } \omega=\omega' \\ 0 & \text{Otherwise} \end{cases} \tag{4.4}$$

The wave number vector $\boldsymbol{k}$ can be expressed in terms of wave energy propagation direction θ, and frequency ω, through the dispersion equation

$$\omega^2 = kg\tanh(kh); \tag{4.5}$$

hence, $S(\boldsymbol{k},\omega)$ represents a directional wave spectrum.

Now let us consider the wave profiles measured at two locations $\boldsymbol{x}$ and $\boldsymbol{y}$ separated by a distance $\boldsymbol{r}$, that is $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{r}$. Similarly to Eq.(4.3), we can rewrite the observation $\eta(\boldsymbol{y}, t+\tau)$ as a function of its inverse Fourier transform $\check{\eta}(\boldsymbol{k}',\omega')$

$$\begin{aligned} \eta(\boldsymbol{y},t+\tau) &= \frac{1}{(2\pi)^3}\int_{\boldsymbol{k}'}\int_{\omega'} e^{-i[\boldsymbol{k}'\cdot\boldsymbol{y}-\omega'(t+\tau)]}\check{\eta}(\boldsymbol{k}',\omega')d\omega' d\boldsymbol{k}' \\ &= \int_{\boldsymbol{k}'}\int_{\omega'} e^{-i[\boldsymbol{k}'\cdot\boldsymbol{y}-\omega'(t+\tau)]}dA(\boldsymbol{k}',\omega') \end{aligned} \tag{4.6}$$

Hence, we can derive the auto-correlation function of the process $\{\eta(\boldsymbol{x},t)\}$ at two locations separated by a distance $\boldsymbol{r}$ and at two time points separated by a delay τ has the following form,

$$\begin{aligned} R(\boldsymbol{r},\tau) &= E[\eta(\boldsymbol{x},t)\eta^*(\boldsymbol{y},t+\tau)] \\ &= E\{\int_{\boldsymbol{k}}\int_{\omega} e^{-i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}dA(\boldsymbol{k},\omega)\int_{\boldsymbol{k}'}\int_{\omega'} e^{i[\boldsymbol{k}'\cdot\boldsymbol{y}-\omega'(t+\tau)]}dA^*(\boldsymbol{k}',\omega')\} \\ &= E\{\int_{\boldsymbol{k}}\int_{\boldsymbol{k}'}\int_{\omega}\int_{\omega'} e^{-i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)+i[\boldsymbol{k}'\cdot\boldsymbol{y}-\omega'(t+\tau)]}dA(\boldsymbol{k},\omega)dA^*(\boldsymbol{k}',\omega')\} \\ &= \int_{\boldsymbol{k}}\int_{\omega} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega\tau)}S(\boldsymbol{k},\omega)d\omega d\boldsymbol{k} \end{aligned} \tag{4.7}$$

The above equation (4.7) implies that the transformation between auto-correlation function of the process $\{\eta(\boldsymbol{x}, t)\}$, and directional wave spectrum $S(\boldsymbol{k}, \omega)$ is the Fourier transform pair.

On the other hand, we can rewrite the above auto-correlation function as a function of a one-dimensional Fourier transform

$$\begin{aligned} R(\boldsymbol{r}, \tau) &= \int_{\boldsymbol{k}} \int_{\omega} e^{i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega\tau)} S(\boldsymbol{k}, \omega) d\omega d\boldsymbol{k} \\ &= \int_{\omega} [\int_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} S(\boldsymbol{k}, \omega) d\boldsymbol{k}] e^{-i\omega\tau} d\omega \end{aligned} \tag{4.8}$$

We denote this one-dimensional Fourier transform as $C(\boldsymbol{r}, \omega)$, which is also the cross-spectral density function between wave profiles measured at two locations separated by a distance $\boldsymbol{r}$, that is,

$$C(\boldsymbol{r}, \omega) = \int_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} S(\boldsymbol{k}, \omega) d\boldsymbol{k} \tag{4.9}$$

The above equation (4.9) implies that if cross-spectra are available for an infinite number of $\boldsymbol{r}$, the wave directional spectra $S(\boldsymbol{k}, \omega)$ can be evaluated by the inverse Fourier transform of $C(\boldsymbol{r}, \omega)$.

Comparing Eq.(4.7) and Eq.(4.9), we can obtain the cross-spectral density function $C(\boldsymbol{r}, \omega)$ is the inverse Flourier transform of the auto-correlation function $R(\boldsymbol{r}, \tau)$.

4.3 Characteristic of Subsurface Ocean Waves

So far, we only discuss the characteristics of surface ocean waves. However, we actually measure all observations beneath the surface of ocean. In order to obtain the relationship similar to that given in Eq.(4.9) for subsurface velocity data, we need to involve the transfer function which is proposed by Isobe et al.(1984) [18] and

defined as the form,

$$H(\boldsymbol{k},\omega) = G(k,\omega)(\cos\theta)^{\alpha}(\sin\theta)^{\beta} \tag{4.10}$$

Here, $G(k,\omega)$ represents a function showing a linear relationship between the kinematic quantity and surface wave profile. For example, the velocity in the x-direction is given by letting $G(k,\omega) = \omega\cosh kz/\sinh kh$, $\alpha = 1$ and $\beta = 0$, where h is the water depth and z is the relative height of sensor from the bottom of ADCP.

Now let us consider the three-dimensional Fourier transform of interior velocity $\xi(\boldsymbol{x},t)$ and also derive the relationship similar to that given in Eq.(4.9) for these subsurface observations.

The forward Fourier transform for interior velocity and its inverse form express as

$$\begin{aligned} \check{\xi}(\boldsymbol{k},\omega) &= \int_t\int_{\boldsymbol{x}} \xi(\boldsymbol{x},t)\ e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}d\boldsymbol{x}dt \\ \xi(\boldsymbol{x},t) &= \frac{1}{(2\pi)^3}\int_{\omega}\int_{\boldsymbol{k}} \check{\xi}(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}d\boldsymbol{k}d\omega \end{aligned} \tag{4.11}$$

Now by introducing the transfer function, we can formulate the relationship between two transformed quantities $\check{\xi}(\boldsymbol{k},\omega)$ and $\check{\eta}(\boldsymbol{k},\omega)$ as,

$$\check{\xi}(\boldsymbol{k},\omega) = H(\boldsymbol{k},\omega)\cdot\check{\eta}(\boldsymbol{k},\omega). \tag{4.12}$$

Hence, the wave kinematic quantity $\xi(\boldsymbol{x},t)$ may be rewritten as

$$\begin{aligned} \xi(\boldsymbol{x},t) &= \frac{1}{(2\pi)^3}\int_{\omega}\int_{\boldsymbol{k}} H(\boldsymbol{k},\omega)\check{\eta}(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}\ d\boldsymbol{k}\ d\omega \\ &= \int_{\omega}\int_{\boldsymbol{k}} H(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}\ dA(\boldsymbol{k},\omega) \end{aligned} \tag{4.13}$$

From so on, let us consider two records of interior velocity at two locations $\boldsymbol{x}_m$

and $\boldsymbol{x}_n$ separated by a distance $\boldsymbol{r}$ and at two time points t and $t+\tau$. We denote these records as $\xi(\mathbf{x}_m, t)$ and $\xi(\mathbf{x}_n, t+\tau)$ respectively. They have similar forms to Eq.(4.13) as follows,

$$\begin{aligned}\xi(\boldsymbol{x}_m, t) &= \frac{1}{(2\pi)^3}\int_{\omega}\int_{\boldsymbol{k}} H_m(\boldsymbol{k},\omega)\breve{\eta}(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{x}_m-\omega t)}\, d\boldsymbol{k}\, d\omega \\ &= \int_{\omega}\int_{\boldsymbol{k}} H_m(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{x}_m-\omega t)}\, dA(\boldsymbol{k},\omega) \end{aligned} \tag{4.14}$$

and

$$\begin{aligned}\xi(\boldsymbol{x}_n, t+\tau) &= \frac{1}{(2\pi)^3}\int_{\omega'}\int_{\boldsymbol{k}'} H_n(\boldsymbol{k}',\omega')\breve{\eta}(\boldsymbol{k}',\omega')e^{-i[\boldsymbol{k}'\cdot\boldsymbol{x}_n-\omega'(t+\tau)]}\, d\boldsymbol{k}'\, d\omega' \\ &= \int_{\omega}\int_{\boldsymbol{k}} H_n(\boldsymbol{k},\omega)e^{-i[\boldsymbol{k}'\cdot\boldsymbol{x}_n-\omega'(t+\tau)]}\, dA(\boldsymbol{k}',\omega') \end{aligned} \tag{4.15}$$

By comparing Eq.(4.14), Eq.(4.15) and Eq.(4.6) as well as using a similar derivation to Eq.(4.7), we can obtain the following relationship,

$$\begin{aligned}& E[\xi(\boldsymbol{x}_m, t)\xi^*(\boldsymbol{x}_n, t+\tau)] \\ =\ & \int_{\omega}\int_{\boldsymbol{k}} H_m(\boldsymbol{k},\omega)H_n^*(\boldsymbol{k},\omega)e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega\tau)}S(\boldsymbol{k},\omega)d\boldsymbol{k}d\omega \end{aligned} \tag{4.16}$$

Proposition 4.3.0.1. *The relationship between the wave-number frequency spectrum of surface wave, $S_{\eta\eta}(\boldsymbol{k},\omega)$ and the cross-correlation function of two inner wave records at locations $\boldsymbol{x}_m$ and $\boldsymbol{x}_n$, $R_{mn}(\boldsymbol{r},\tau)$ can be written as*

$$R_{mn}(\boldsymbol{r},\tau) = \frac{1}{2(2\pi)^2}\int_{\boldsymbol{k}}\int_{\omega} H_m^*(\boldsymbol{k},\omega)S_{\eta\eta}(\boldsymbol{k},\omega)H_n(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega\tau)}d\omega d\boldsymbol{k} \tag{4.17}$$

Proof. The proof of the above proposition is directly following Wiener-Khintchine theorem which is derived as follows:

The cross-correlation function of two inner wave profiles $\xi_m(\mathbf{x}_m, t)$ and $\xi_n(\mathbf{x}_n, t+\tau)$, denoted by $R_{mn}(\boldsymbol{r}, \tau)$, is defined as

$$R_{mn}(\boldsymbol{r}, \tau) = E[\xi_m^*(\boldsymbol{x}_m, t) \cdot \xi_n(\boldsymbol{x}_n, t+\tau)] \tag{4.18}$$

Hence,

$$\begin{aligned} & R_{mn}(\boldsymbol{r}, \tau) \\ \overset{a.s.}{=} & \lim_{R_1, R_2 \to \infty} \frac{1}{4R_1R_2} \lim_{T \to \infty} \frac{1}{2T} \int_D \int_{-T}^{T} \xi_m^*(\boldsymbol{x}_m, t) \xi_n(\boldsymbol{x}_n, t+\tau) dt d\boldsymbol{x}_m \\ = & \lim_{R_1, R_2 \to \infty} \frac{1}{4R_1R_2} \lim_{T \to \infty} \frac{1}{2T} \int_{\boldsymbol{x}_m} \int_t \xi_m^*(\boldsymbol{x}_m, t) \xi_n(\boldsymbol{x}_m + \boldsymbol{r}, t+\tau) dt d\boldsymbol{x}_m \end{aligned} \tag{4.19}$$

where $D := [-R_1, R_1] \times [-R_2, R_2]$.

Using Eq. (4.14) and (4.15), we can re-write $\xi_m^*(\boldsymbol{x}_m, t)\xi_n(\boldsymbol{x}_m + \boldsymbol{r}, t+\tau)$ as

$$\begin{aligned} & \xi_m^*(\mathbf{x}_m, t)\xi_n(\mathbf{x}_m + \boldsymbol{r}, t+\tau) \\ = & \{\frac{1}{(2\pi)^3} \int_{\boldsymbol{k}} \int_{\omega} H_m^*(\boldsymbol{k}, \omega) \check{\eta}^*(\boldsymbol{k}, \omega) e^{i(\boldsymbol{k} \cdot \mathbf{x}_m - \omega t)} d\boldsymbol{k} d\omega\} \times \\ & \{\frac{1}{(2\pi)^3} \int_{\boldsymbol{k}'} \int_{\omega'} H_n(\boldsymbol{k}', \omega') \check{\eta}(\boldsymbol{k}', \omega') e^{-i[\boldsymbol{k}' \cdot (\mathbf{x}_m + \boldsymbol{r}) - \omega'(t+\tau)]} d\boldsymbol{k}' d\omega'\} \\ = & \frac{1}{(2\pi)^6} \int_{\boldsymbol{k}} \int_{\omega} [H_m^*(\boldsymbol{k}, \omega) \check{\eta}^*(\boldsymbol{k}, \omega)] \times \{\int_{\boldsymbol{k}'} \int_{\omega'} [H_n(\boldsymbol{k}', \omega') \check{\eta}(\boldsymbol{k}', \omega') e^{-i(\boldsymbol{k}' \cdot \boldsymbol{r} - \omega'\tau)}] \\ & \times [e^{i(\boldsymbol{k} - \boldsymbol{k}') \cdot \boldsymbol{x}_m} \times e^{-i(\omega - \omega')t}] d\omega' d\boldsymbol{k}'\} d\omega d\boldsymbol{k} \end{aligned} \tag{4.20}$$

We know that δ-function with one argument ω has the following integral representation,

$$\delta(\omega) = \frac{1}{2\pi} \int_t e^{\pm i\omega t} dt \tag{4.21}$$

and property,

$$\int_{\omega'} F(\omega') \delta(\omega - \omega') d\omega' = F(\omega) \tag{4.22}$$

The key step in the following derivation is the use of both the generalized integral representation of the δ-function with three argument $(\boldsymbol{k}, \omega)$ which is written as,

$$\delta(\boldsymbol{k}, \omega) = \frac{1}{(2\pi)^3} \int_{\boldsymbol{r}} \int_t e^{\pm i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)} dt d\boldsymbol{r} \tag{4.23}$$

and the generalized property of three-argument δ-function,

$$\int_{\boldsymbol{k}', \omega'} F(\boldsymbol{k}', \omega') \delta(\boldsymbol{k} - \boldsymbol{k}', \omega - \omega') d\omega' d\boldsymbol{k}' = F(\boldsymbol{k}, \omega) \tag{4.24}$$

Now plugging Eq. (4.20) into Eq. (4.19), we obtain

$$\begin{aligned}
& R_{mn}(\boldsymbol{r}, \tau) \\
= & \lim_{R_1, R_2 \to \infty} \frac{1}{4R_1R_2} \lim_{T\to\infty} \frac{1}{2T} \int_{\boldsymbol{x}_m} \int_t \xi_m^*(\boldsymbol{x}_m, t) \xi_n(\boldsymbol{x}_n, t+\tau) dt d\boldsymbol{x}_m \\
= & \lim_{R_1, R_2 \to \infty} \frac{1}{4R_1R_2} \lim_{T\to\infty} \frac{1}{2T} \int_{\boldsymbol{x}_m} \int_t \{ \frac{1}{(2\pi)^6} \int_{\boldsymbol{k}} \int_\omega [H_m^*(\boldsymbol{k}, \omega) \check{\eta}^*(\boldsymbol{k}, \omega)] \times \\
& \int_{\boldsymbol{k}'} \int_{\omega'} [H_n(\boldsymbol{k}', \omega') \check{\eta}(\boldsymbol{k}', \omega') e^{-i(\boldsymbol{k}'\cdot\boldsymbol{r} - \omega'\tau)}] \times \\
& [e^{i(\boldsymbol{k} - \boldsymbol{k}')\cdot\boldsymbol{x}_m} \times e^{-i(\omega - \omega')t}] d\omega' d\boldsymbol{k}' d\omega d\boldsymbol{k} \} dt d\boldsymbol{x}_m
\end{aligned}$$

Changing the integration order, the above equation can be rewritten as

$$\begin{aligned}
& R_{mn}(\boldsymbol{r}, \tau) \\
= & \lim_{R_1, R_2 \to \infty} \frac{1}{4R_1R_2} \lim_{T\to\infty} \frac{1}{2T} \frac{1}{(2\pi)^6} \int_{\boldsymbol{k}} \int_\omega [H_m^*(\boldsymbol{k}, \omega) \check{\eta}^*(\boldsymbol{k}, \omega)] \times \\
& \{ \int_{\boldsymbol{k}'} \int_{\omega'} [H_n(\boldsymbol{k}', \omega') \check{\eta}(\boldsymbol{k}', \omega') e^{-i(\boldsymbol{k}'\cdot\boldsymbol{r} - \omega'\tau)}] \times \\
& [\int_{\boldsymbol{x}_m} \int_t e^{i(\boldsymbol{k} - \boldsymbol{k}')\cdot\boldsymbol{x}_m} \times e^{-i(\omega - \omega')t} dt d\boldsymbol{x}_m] d\omega' d\boldsymbol{k}' \} d\omega d\boldsymbol{k}
\end{aligned}$$

Now referring to the definition of δ function and using its properties, we can derive

the followings

$$
\begin{aligned}
& R_{mn}(\boldsymbol{r},\tau) \\
= & \lim_{R_1,R_2\to\infty}\frac{1}{4R_1R_2}\lim_{T\to\infty}\frac{1}{2T}\frac{1}{(2\pi)^6}\int_{\boldsymbol{k}}\int_{\omega}[H_m^*(\boldsymbol{k},\omega)\check{\eta}^*(\boldsymbol{k},\omega)]\times \\
& \{\int_{\boldsymbol{k}'}\int_{\omega'}[H_n(\boldsymbol{k}',\omega')\check{\eta}(\boldsymbol{k}',\omega')e^{-i(\boldsymbol{k}'\cdot\boldsymbol{r}-\omega'\tau)}]\times \\
& [(2\pi)^3\delta(\boldsymbol{k}-\boldsymbol{k}',\omega-\omega')]d\omega' d\boldsymbol{k}'\}d\omega d\boldsymbol{k} \\
= & \lim_{R_1,R_2\to\infty}\frac{1}{4R_1R_2}\lim_{T\to\infty}\frac{1}{2T}\frac{1}{(2\pi)^3}\int_{\boldsymbol{k}}\int_{\omega}[H_m^*(\boldsymbol{k},\omega)\check{\eta}^*(\boldsymbol{k},\omega)]\times \\
& \{H_n(\boldsymbol{k},\omega)\check{\eta}(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega\tau)}\}d\omega d\boldsymbol{k}
\end{aligned}
$$

Changing the order of limit and integration, $R_{mn}(\boldsymbol{r},\tau)$ can be simplified as the representation of

$$
\begin{aligned}
& R_{mn}(\boldsymbol{r},\tau) \\
= & \frac{1}{2(2\pi)^2}\int_{\boldsymbol{k}}\int_{\omega}H_m^*(\boldsymbol{k},\omega)\{\lim_{R_1,R_2\to\infty}\frac{1}{4R_1R_2}\lim_{T\to\infty}\frac{1}{2T\pi}[\check{\eta}^*(\boldsymbol{k},\omega)\check{\eta}(\boldsymbol{k},\omega)]\}\times \\
& H_n(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega\tau)}d\omega d\boldsymbol{k} \\
= & \frac{1}{2(2\pi)^2}\int_{\boldsymbol{k}}\int_{\omega}H_m^*(\boldsymbol{k},\omega)S_{\eta\eta}(\boldsymbol{k},\omega)H_n(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega\tau)}d\omega d\boldsymbol{k} \qquad (4.25)
\end{aligned}
$$

The last derivation comes about because of the definition of wave-number frequency spectrum refereed to Eq. (1.22) and Eq. (4.4). □

4.3.1 The Forward Model

Now we can derive the forward model as follows:

By re-writing Eq. (4.17) as

$$\begin{aligned} & 2(2\pi)R_{mn}(\boldsymbol{r},\tau) \\ = \; & \frac{1}{2\pi}\int_{\boldsymbol{k}}\int_{\omega} H_m^*(\boldsymbol{k},\omega)S_{\eta\eta}(\boldsymbol{k},\omega)H_n(\boldsymbol{k},\omega)e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega\tau)}d\omega d\boldsymbol{k} \\ = \; & \frac{1}{2\pi}\int_{\omega}\{\int_{\boldsymbol{k}} H_m^*(\boldsymbol{k},\omega)S_{\eta\eta}(\boldsymbol{k},\omega)H_n(\boldsymbol{k},\omega)e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}d\boldsymbol{k}\}e^{i\omega\tau}d\omega \end{aligned} \tag{4.26}$$

we could write the inverse Fourier transform of Eq. (4.26) as

$$\int_{\boldsymbol{k}} H_m^*(\boldsymbol{k},\omega)S_{\eta\eta}(\boldsymbol{k},\omega)H_n(\boldsymbol{k},\omega)e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}d\boldsymbol{k} = \int_{\tau} 2(2\pi)R_{mn}(\boldsymbol{r},\tau)\cdot e^{-i\omega\tau}d\tau \tag{4.27}$$

that is,

$$\frac{1}{2(2\pi)}\int_{\boldsymbol{k}} H_m^*(\boldsymbol{k},\omega)S_{\eta\eta}(\boldsymbol{k},\omega)H_n(\boldsymbol{k},\omega)e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}d\boldsymbol{k} = \int_{\tau} R_{mn}(\boldsymbol{r},\tau)\cdot e^{-i\omega\tau}d\tau \tag{4.28}$$

According to the definition of expectation and the property of stationary processes, we obtain

$$\begin{aligned} & \int_{\tau} R_{mn}(\boldsymbol{r},\tau)\cdot e^{-i\omega\tau}d\tau \\ = \; & \int_{\tau} E[\xi_m^*(\boldsymbol{x}_m,t)\xi_n(\boldsymbol{x}_m+\boldsymbol{r},t+\tau)]\cdot e^{-i\omega\tau}d\tau \\ \overset{a.s.}{=} \; & \int_{\tau} \lim_{R_1,R_2\to\infty}\frac{1}{4R_1R_2}\lim_{T\to\infty}\frac{1}{2T} \\ & \left[\int_D\int_{-T}^{T}\xi_m^*(\boldsymbol{x}_m,t)\xi_n(\boldsymbol{x}_m+\boldsymbol{r},t+\tau)dtd\boldsymbol{x}_m\right]\times e^{-i\omega(t+\tau)}e^{i\omega t}d\tau \end{aligned} \tag{4.29}$$

By changing the order of limit and integration, the above equation can be rewritten as,

$$\begin{aligned}
&\int_{\tau} R_{mn}(\boldsymbol{r},\tau)\cdot e^{-i\omega\tau}d\tau \\
=\ & \lim_{R_1,R_2\to\infty}\frac{1}{4R_1R_2}\lim_{T\to\infty}\frac{1}{2T}\int_{\boldsymbol{x}_m} \\
& \left\{\int_t \xi_m^*(\boldsymbol{x}_m,t)e^{i\omega t}[\int_{\tau}\xi_n(\boldsymbol{x}_m+\boldsymbol{r},t+\tau)e^{-i\omega(t+\tau)}d\tau]dt\right\}d\boldsymbol{x}_m \\
=\ & \lim_{R_1,R_2\to\infty}\frac{1}{4R_1R_2}\lim_{T\to\infty}\frac{1}{2T}\int_{\boldsymbol{x}_m}\{\int_t \xi_m^*(\boldsymbol{x}_m,t)e^{i\omega t}\}\check{\xi}_n^{(2)}(\boldsymbol{x}_m+\boldsymbol{r},\omega)dtd\boldsymbol{x}_m \\
=\ & \int_{\boldsymbol{x}_m}\{\check{\xi}_m^{(2)*}(\boldsymbol{x}_m,\omega)\}\check{\xi}_n^{(2)}(\boldsymbol{x}_m+\boldsymbol{r},\omega)d\boldsymbol{x}_m \qquad (4.30)
\end{aligned}$$

where $\check{\xi}^{(2)}(\boldsymbol{x},\omega)$ denote the forward Fourier transform of $\xi(\boldsymbol{x},t)$ by fixing the first argument $\boldsymbol{x}$.

Let us define the cross-spectral density function between wave profiles measured at two locations $\boldsymbol{x}_m$ and $\boldsymbol{x}_n$ separated by a distance vector $\boldsymbol{r}=\boldsymbol{x}_n-\boldsymbol{x}_m$ as,

$$C_{mn}(\omega)=\int_{\boldsymbol{x}_m}\{\check{\xi}_m^{(2)*}(\boldsymbol{x}_m,\omega)\}\check{\xi}_n^{(2)}(\boldsymbol{x}_m+\boldsymbol{r},\omega)d\boldsymbol{x}_m \qquad (4.31)$$

Then we obtain the forward model,

$$C_{mn}(\omega)=\frac{1}{2(2\pi)}\int_{\boldsymbol{k}} H_m(\boldsymbol{k},\omega)H_n^*(\boldsymbol{k},\omega)e^{-i\boldsymbol{k}\cdot(\boldsymbol{x}_n-\boldsymbol{x}_m)}S(\boldsymbol{k},\omega)d\boldsymbol{k} \qquad (4.32)$$

4.4 Maximum Likelihood Method

In this section, we will apply the maximum likelihood technique to derive an estimated wave-number frequency spectrum based on the forward model, that is,

$$C_{mn}(\omega)=\frac{1}{2(2\pi)}\int_{\boldsymbol{k}} H_m(\boldsymbol{k},\omega)H_n^*(\boldsymbol{k},\omega)e^{-i\boldsymbol{k}\cdot(\boldsymbol{x}_n-\boldsymbol{x}_m)}S(\boldsymbol{k},\omega)d\boldsymbol{k} \qquad (4.33)$$

According to the above forward model, the wave directional spectra $S(\boldsymbol{k}, \omega)$ and the cross-spectra between two quantities ξ_m and ξ_n, denoted by $C_{mn}(\omega)$, at different location $\boldsymbol{x}_m$ and $\boldsymbol{x}_n$, respectively, are linearly related. Hence, the desired wave-number frequency spectrum $\hat{S}(\boldsymbol{k}, \omega)$ is presented as a linear combination of cross-spectra $C_{mn}(\omega)$. We may write

$$\hat{S}(\boldsymbol{k}, \omega) = \sum_{m,n} \alpha_{mn}(\boldsymbol{k}) C_{mn}(\omega). \tag{4.34}$$

$\alpha_{mn}(\boldsymbol{k})$ and $\alpha_{nm}(\boldsymbol{k})$ are complex conjugates so that $\hat{S}(\boldsymbol{k}, \omega)$ is a real-valued function which is the same as the directional spectra $S(\boldsymbol{k}, \omega)$. Further, we assume that the coefficients $\alpha_{mn}(\boldsymbol{k})$ can be expressed as a factorial form,

$$\alpha_{mn}(\boldsymbol{k}) = \gamma_m(\boldsymbol{k}) \gamma_n^*(\boldsymbol{k}) \tag{4.35}$$

and Eq. (4.34) can be written as

$$\hat{S}(\boldsymbol{k}, \omega) = \sum_{m,n} \gamma_m(\boldsymbol{k}) C_{mn}(\omega) \gamma_n^*(\boldsymbol{k}). \tag{4.36}$$

By plugging Eq.(4.33) into Eq.(4.36), the estimated wave-number frequency spectra becomes

$$\hat{S}(\boldsymbol{k}, \omega) = \int_{\boldsymbol{k}'} \varpi(\boldsymbol{k}, \boldsymbol{k}') S(\boldsymbol{k}', \omega) d\boldsymbol{k}' \tag{4.37}$$

where

$$\varpi(\boldsymbol{k}, \boldsymbol{k}') = \frac{1}{2(2\pi)} \sum_{m,n} \gamma_m(\boldsymbol{k}) H_m(\boldsymbol{k}', \omega) e^{i\boldsymbol{k}' \cdot \boldsymbol{x}_m} e^{-i\boldsymbol{k}' \cdot \boldsymbol{x}_n} H_n^*(\boldsymbol{k}', \omega) \gamma_n^*(\boldsymbol{k}) \tag{4.38}$$

Furthermore, for convenience, $\varpi(\boldsymbol{k}, \boldsymbol{k}')$ may be written as

$$\varpi(\boldsymbol{k}, \boldsymbol{k}') = \sum_{m,n} \gamma_m(\boldsymbol{k}) T_{m.n}(\boldsymbol{k}', \omega) \gamma_n^*(\boldsymbol{k}) \tag{4.39}$$

where

$$T_{mn}(\boldsymbol{k}', \omega) = \frac{1}{2(2\pi)} H_m(\boldsymbol{k}', \omega) e^{i\boldsymbol{k}'\cdot\boldsymbol{x}_m} e^{-i\boldsymbol{k}'\cdot\boldsymbol{x}_n} H_n^*(\boldsymbol{k}', \omega) \tag{4.40}$$

In Eq.(4.37), $S(\boldsymbol{k}', \omega)$ is the unknown true spectrum, $\hat{S}(\boldsymbol{k}, \omega)$ is the estimated spectrum, and $\varpi(\boldsymbol{k}, \boldsymbol{k}')$ is known as the wave-number window function. The smaller the window function, the higher the resolution of $\hat{S}(\boldsymbol{k}, \omega)$. The best estimator can be achieved when $\varpi(\boldsymbol{k}, \boldsymbol{k}')$ approximately equals to the δ-function centered at $\boldsymbol{k}'$.

However, the δ function is a function defined on the real line which is zero everywhere except at the center $\boldsymbol{k}'$ where it is infinite,

$$\delta(\boldsymbol{k}; \boldsymbol{k}') = \begin{cases} +\infty & \text{if } \boldsymbol{k} = \boldsymbol{k}' \\ 0 & \text{otherwise.} \end{cases} \tag{4.41}$$

and which is also constrained to satisfy the identity,

$$\int_{\boldsymbol{k}} \varpi(\boldsymbol{k}; \boldsymbol{k}') d\boldsymbol{k} = 1 \tag{4.42}$$

If we remove the constraint $\int_{\boldsymbol{k}} \varpi(\boldsymbol{k}; \boldsymbol{k}') d\boldsymbol{k} = 1$, the estimation $\hat{S}(\boldsymbol{k}, \omega)$ is equivalent to the true energy spreading spectrum $S(\boldsymbol{k}', \omega)$ up to a multiplier. And if we also let the wave-number window function be 1 at $\boldsymbol{k}'$, a windows function which looks most like a Dirac δ function will minimize $\hat{S}(\boldsymbol{k}, \omega)$ given in Eq.(4.37).

Now the problem was converted to minimize $\hat{S}(\boldsymbol{k}, \omega)$ under the following condi-

tion,

$$\varpi(\boldsymbol{k},\boldsymbol{k}) = \sum_{m,n} \gamma_m(\boldsymbol{k}) T_{m.n}(\boldsymbol{k},\omega)\gamma_n^*(\boldsymbol{k}) = 1 \tag{4.43}$$

This, in turn, may be likened to the problem of maximizing the following quantity:

$$\frac{\varpi(\boldsymbol{k},\boldsymbol{k})}{\hat{S}(\boldsymbol{k},\omega)} = \frac{\sum_{m,n} \gamma_m(\boldsymbol{k}) T_{m.n}(\boldsymbol{k},\omega)\gamma_n^*(\boldsymbol{k})}{\sum_{m,n} \gamma_m(\boldsymbol{k}) C_{m.n}(\omega)\gamma_n^*(\boldsymbol{k})} \tag{4.44}$$

This is also equivalent to finding the maximum eigenvalue λ which satisfies the following relationship for the given matrices $T_{m.n}(\boldsymbol{k})$ and $C_{m.n}(\omega)$:

$$\sum_n T_{mn}(\boldsymbol{k},\omega)\gamma_n^* = \lambda \sum_n C_{mn}(\omega)\gamma_n^* \tag{4.45}$$

and hence, we have

$$\sum_m \sum_n C_{l.m}^{-1}(\omega) T_{mn}(\boldsymbol{k},\omega)\gamma_n^* = \lambda\gamma_l^* \tag{4.46}$$

where $C_{lm}^{-1}(\omega)$ is the inverse matrix of $C_{lm}(\omega)$.

Thus, from Eq.(4.43), the estimated spectra is inversely proportional to the maximum eigenvalue λ_{max}. That is,

$$\hat{S}(\boldsymbol{k},\omega) \propto \frac{1}{\lambda_{max}} \tag{4.47}$$

where λ_{max} can be obtained as

$$\begin{aligned} \lambda_{max} &= \sum_m \sum_n C_{m.n}^{-1}(\omega) T_{mn}(\boldsymbol{k},\omega)\gamma_n^* \\ &= \sum_m \sum_n H_m^*(\boldsymbol{k},\omega) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_m} C_{m.n}^{-1}(\omega) e^{i\boldsymbol{k}\cdot\boldsymbol{x}_n} H_n(\boldsymbol{k},\omega) \end{aligned} \tag{4.48}$$

By applying the relationship for all frequencies, the estimated energy distribution

as a function of wave number and frequency ω is

$$\hat{S}(\boldsymbol{k},\omega) = \frac{c_0}{\sum_m \sum_n H_m^*(\boldsymbol{k},\omega) C_{m.n}^{-1}(\omega) H_n(\boldsymbol{k},\omega)} e^{i\boldsymbol{k}\cdot(\boldsymbol{x}_n - \boldsymbol{x}_m)} \tag{4.49}$$

where c_0 is a constant determined such that the integration of $\hat{S}(\boldsymbol{k},\omega)$ with respect to $\boldsymbol{k}$ yields a point spectrum. That is

$$S(\omega) = \int_{\boldsymbol{k}} S(\boldsymbol{k},\omega) d\boldsymbol{k} \tag{4.50}$$

In Chapter 2 and Chapter 3, we propose a new approach to estimate wave-number spectra which is the left-hand side of the forward model. The solution for forward model shown in Equation (4.49) requires to calculate the inverse of the estimated wave-number spectra. In the next section, we show how to convert a complex matrix inversion to a real matrix inversion

4.5 New Approach to Complex Matrix Inversion

Matrix inversion is required in many signal processing systems. For example, in mobile telecommunication systems the estimation of parameters for equalizers involves matrix inversion [25]. Another interesting example is provided by frequency-wavenumber spectrum analysis [9], where the inverse of a spectral matrix is used to construct the high-resolution estimate for the frequency-wavenumber spectrum.

We confront the first difficulty while dealing with multivariate spectral estimation is how to efficiently inverse large complex matrices. On the one hand, there are many very efficient implementation of real matrix inversion already provided in software. On the other hand, however, no implementation of the matrix-version algorithm for complex matrices is provided in R Library. The question arises as to the rationale for the inverse of a complex matrix using real matrix inversion.

4.5.1 The Problem of Complex Matrix Inversion

Let C denote a $N \times N$ complex matrix. Let A and B denote real matrices which are the real portion and the imaginary portion of C, respectively, that is, $C = A+iB$. The goal is to find the inverse matrix of C denoted by C^{-1}. This task is equivalent to solving the following system of complex equations for real matrices X and Y,

$$(A + iB) \cdot (X + iY) = I_N. \tag{4.51}$$

we can rewrite the above complex equations as,

$$(AX - BY) + i(BX + AY) = I_N + i0_N. \tag{4.52}$$

which leads to the following set of real equations:

$$\begin{aligned} AX - BY &= I_N \\ BX + AY &= 0_N. \end{aligned} \tag{4.53}$$

Using a real matrix inversion, this problem can be written as the following formula,

$$\begin{pmatrix} A & -B \\ B & A \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} I_N \\ 0_N \end{pmatrix}. \tag{4.54}$$

The solution of the above real equations can be represented as the form of a $2N \times 2N$

matrix inversion,

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} A & -B \\ B & A \end{pmatrix}^{-1} \begin{pmatrix} I_N \\ 0_N \end{pmatrix}. \tag{4.55}$$

Consequently, the inverse of complex matrix C is given by

$$\begin{aligned} C^{-1} &= (I_N,\ iI_N) \cdot \begin{pmatrix} X \\ Y \end{pmatrix} \\ &= (I_N,\ iI_N) \cdot \begin{pmatrix} A & -B \\ B & A \end{pmatrix}^{-1} \begin{pmatrix} I_N \\ 0_N \end{pmatrix}. \end{aligned} \tag{4.56}$$

The proposed conversion from $N \times N$ complex matrix inversion to $2N \times 2N$ real matrix inversion works for both pseudo-inverse and inverse matrices. It is also possible to directly use the highly optimized real matrix inversion algorithm in many software.

4.5.2 Reducing the Complex Matrix Inversion

In this section, we continue to improve the inversion of complex matrix C by reducing $2N \times 2N$ real matrix inversion to $N \times N$ real matrix inversion. Let D be the following $2N \times 2N$ real matrix,

$$D = \begin{pmatrix} A & -B \\ B & A \end{pmatrix}$$

.

First we split D^{-1} in four $N \times N$ submatrices. We start with the derivation of

the form of each $X_{ij}, i, j = 1, 2$,

$$D^{-1} = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}. \tag{4.57}$$

Basic algebra calculations lead to the following set of equations,

$$\begin{aligned} X_{12} &= A^{-1}B(BA^{-1}B + A)^{-1} \\ X_{21} &= -(BA^{-1}B + A)^{-1}BA^{-1} \\ X_{11} &= A^{-1} + A^{-1}BX_{21} \\ X_{22} &= (BA^{-1}B + A)^{-1}. \end{aligned} \tag{4.58}$$

We can also show the following facts with more complicated matrix algebra,

$$\begin{aligned} X_{11} &= X_{22} \\ X_{12} &= -X_{21}. \end{aligned} \tag{4.59}$$

Let X_0 denote $A^{-1}B$. Now we can obtain the more explicit formulas for X_{ij}s,

$$\begin{aligned} X_{11} = \quad X_{22} \quad &= (BX_0 + A)^{-1} \\ X_{12} = -X_{21} \quad &= (B + AX_0^{-1})^{-1}. \end{aligned} \tag{4.60}$$

Conclusions as a result, the resulting inverse matrix of the original $N \times N$ complex matrix $A + iB$ is $(BA^{-1}B + A)^{-1} - i(B + AB^{-1}A)^{-1}$, which is calculated through two real matrix multiplications and four real matrix inversions.

4.6 Application to ADCP Data

In this section, we apply our propose method as well as other three methods to estimate the wave-number spectra and implement these estimators to yield the directional spectral distribution using Equation (4.49) respectively.

The partial simulated results of the wave-number spectral estimators are shown in Figure 4.6, Figure 4.7, Figure 4.8 and Figure 4.9.

Figure 4.10 shows the partial simulated results of the wave-number spectral estimator used our proposed method.

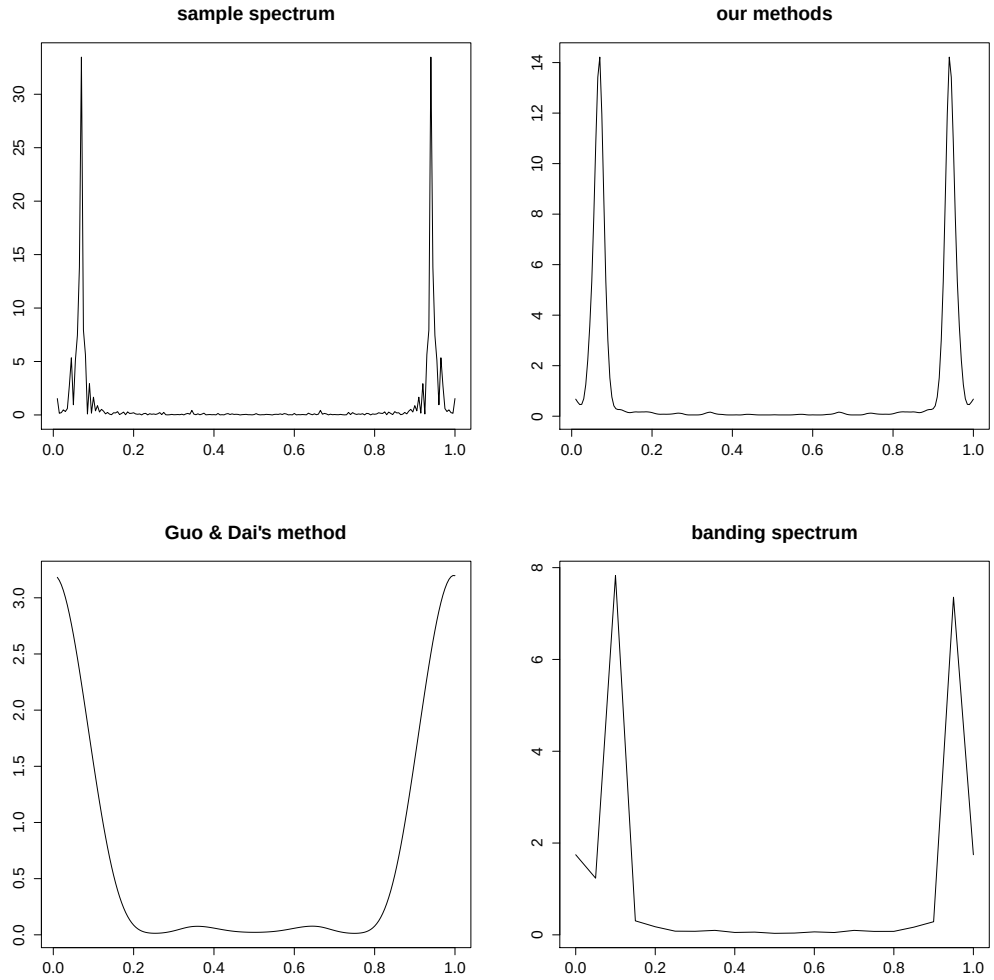

Figure 4.6: The comparison of four methods for selected spectra in shallow water.

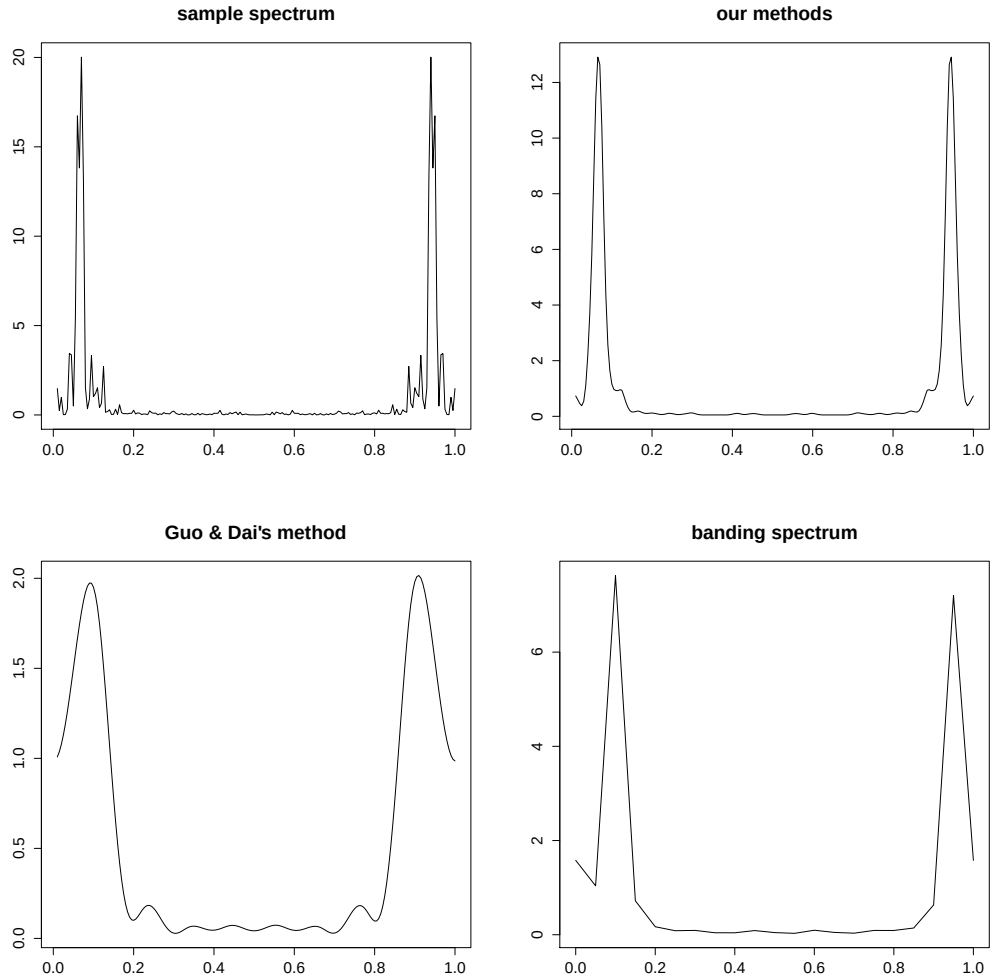

Figure 4.7: The comparison of four methods for selected spectra in deep water.

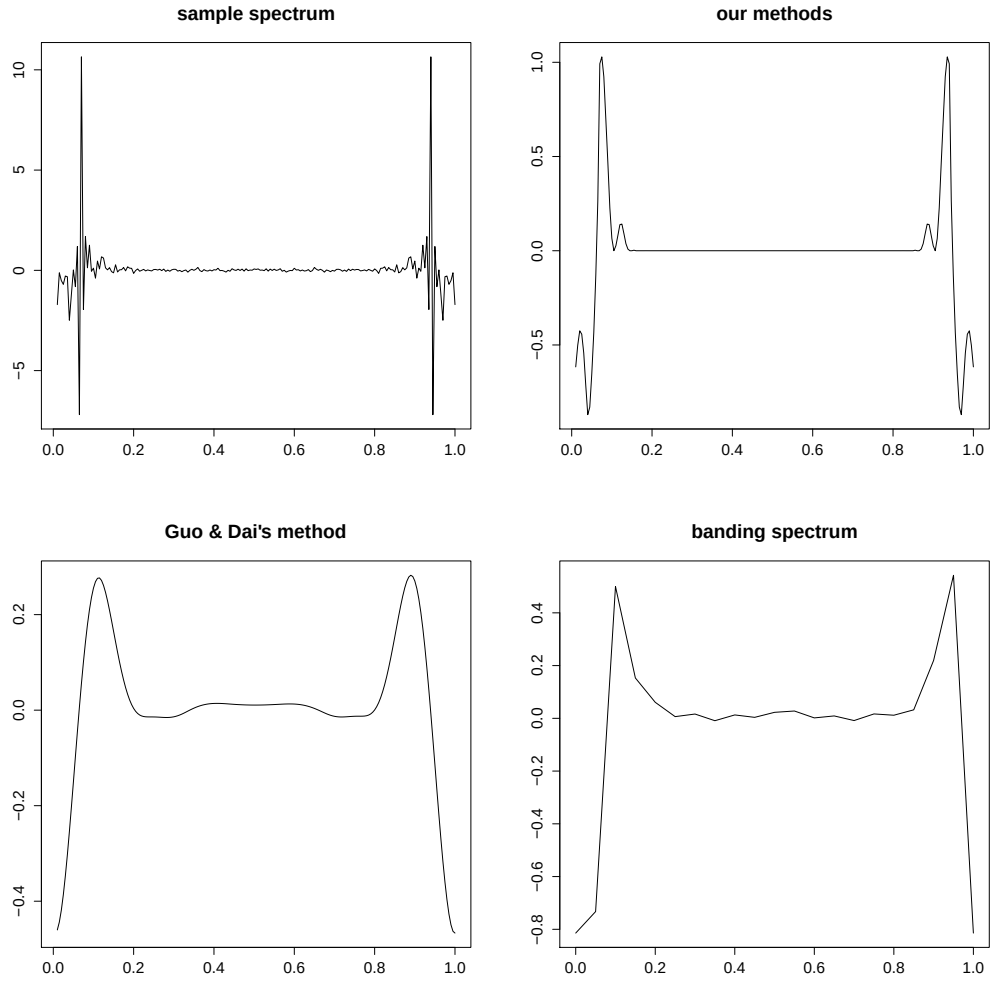

Figure 4.8: The comparison of four methods for the real part of selected cross-spectra in deep water.

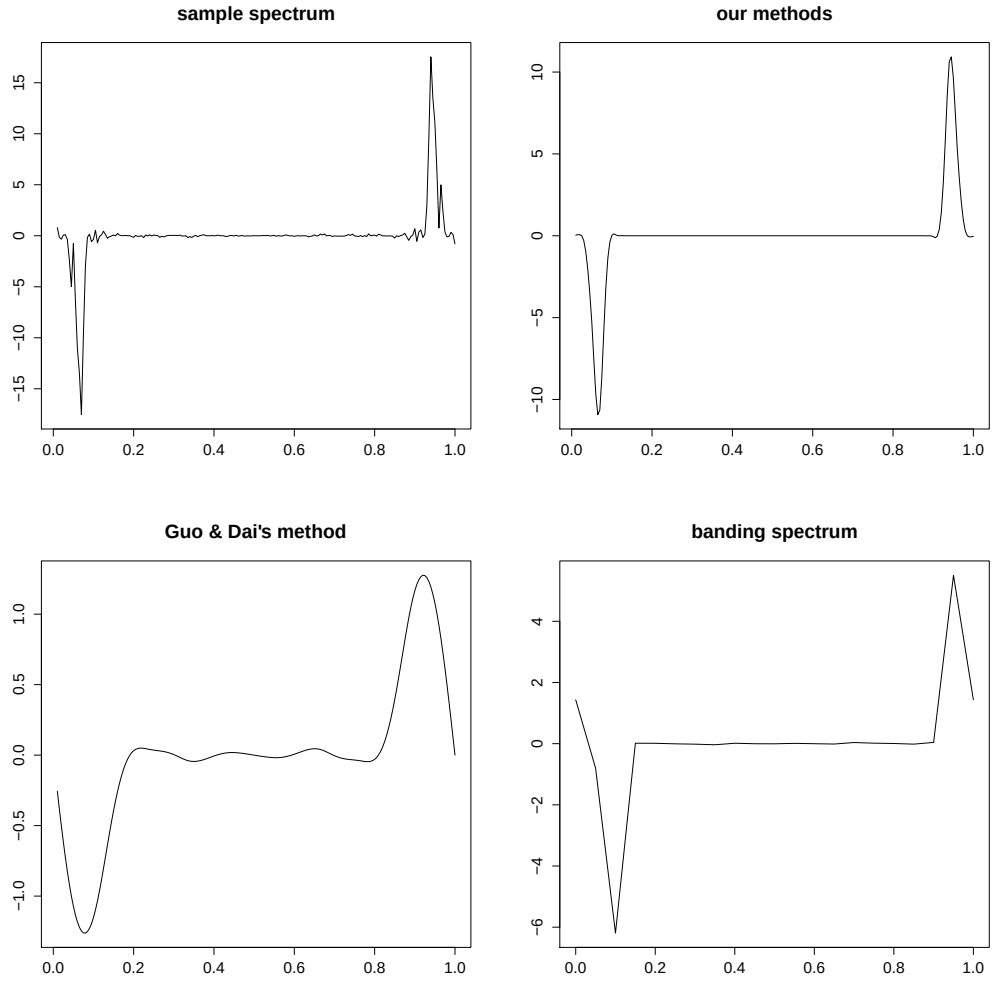

Figure 4.9: The comparison of four methods for the imaginary part of selected cross-spectra in deep water.

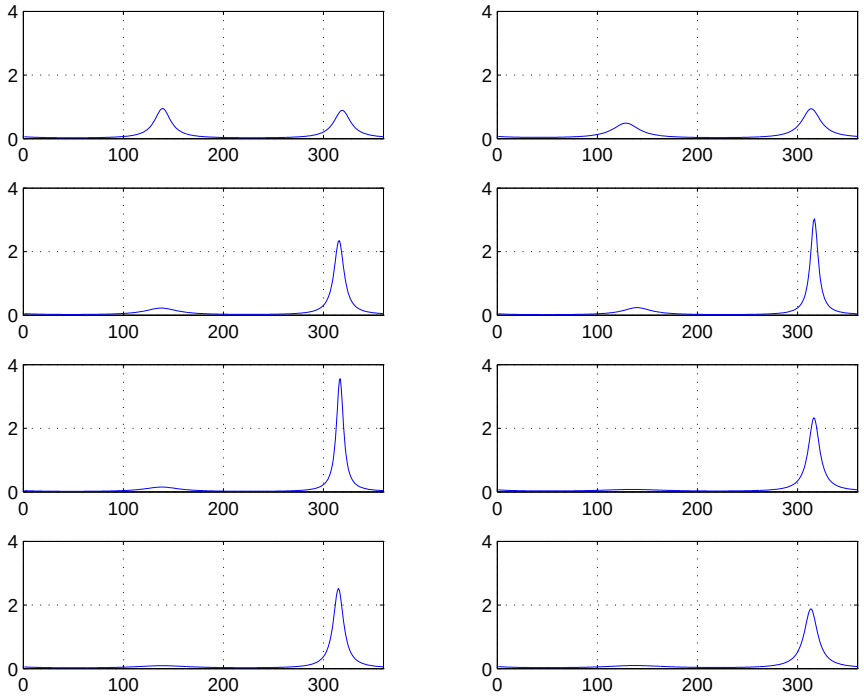

Figure 4.10: The comparison of four methods for the imaginary part of selected cross-spectra in deep water.

5. CONCLUSION AND DISCUSSION

In Chapter 2, we propose a new approach to regularize real/complex 3-mode tensor with positive definite, sparsity and smoothness. This proposed approach can be directly used to estimate wave-number spectra in oceanography, meteorology and other signal processing. We regard the estimation of spectra as a convex optimization problem. By introducing two dual variables, we can alternatively minimizing the target Lagrangian function with respect to different intermediate variables in order to fulfill the requirements: positive definite, sparsity and smoothness respectively. The convergence of Algorithm 1 leads to that the resulting estimator simultaneously satisfies positive definite, sparsity and smoothness.

Compared to the extended tapering approaches of Dai and Guo (2004), Rosen and Stoffer (2007), a significant advantage of the SSP approach is that it enforces a sparsity pattern on the resulting estimator meanwhile keeps the requirements of smoothness and positive definite. While the extended tapering approaches must be required to use a raw positive definite estimator as start point so that the resulting estimator can remain the positive definiteness. Besides, due to allow optimal smoothing for each element of the spectral matrix, the sparsity of the resulting estimator is very poor. This may exceedingly increase the computation complexity for further study, such as matrix inversion. Another advantage comes from the view of spectral analysis and energy representation that the SSP approach can denoise the sources of energy by removing all elements associated with white noise. At the same time, the SSP approach may enlarge the signal of interest somehow. Compared the selection of the starting points, the SSP approach is much more flexible to set the starting points than the extended tapering approaches, because the alternating directive algorithm

can automatically guarantee the requirement of positive definite no matter what the properties of the starting tensor could have.

In Chapter 3, we provide the theoretical analysis of the convergence of proposed alternating directive algorithm. The details of the corresponding proofs of lemmas and the main theory are given in this chapter. The numerical examples in the end of Chapter 2 illustrate the above conclusion.

We continue to analyze ADCP data and estimate the frequency-directional distribution of waves by applying our proposed SSP estimator. We use the extended maximum likelihood method to solve the forward model for the target distribution. The resulting estimator can express more accurate characteristic of the energy source of the ocean waves.

REFERENCES

[1] P.J. Bickel and E. Levina. Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604, 2008.

[2] P.J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227, 2008.

[3] D.R. Brillinger. *Time series: data analysis and theory*, volume 36 of *Classics in Applied Mathematics*. Siam, 2001.

[4] P.J. Brockwell and R.A. Davis. *Time series: theory and methods*. Springer, 2009.

[5] J.P. Burg. *Maximum entropy spectral analysis.* PhD thesis, Standford University, 1975.

[6] T. Cai and W. Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684, 2011.

[7] T. Cai, C-H Zhang, and H.H. Zhou. Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38(4):2118–2144, 2010.

[8] T. Cai and H. Zhou. Minimax estimation of large covariance matrices under l1 norm. *Statistica Sinica*, 22(4):1319–1378, 2012.

[9] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969.

[10] J. Capon, R.J. Greenfield, and R.J. Kolker. Multidimensional maximum-likelihood processing of a large aperture seismic array. *Proceedings of the IEEE*, 55(2):192–211, 1967.

[11] M. Dai and W. Guo. Multivariate spectral analysis using cholesky decomposition. *Biometrika*, 91(3):629–643, 2004.

[12] R.E. Davis and L.A. Regier. Methods for estimating directional wave spectra from multi-element arrays. *Journal of Marine Research*, 35(3):453–478, 1977.

[13] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

[14] R. Furrer and T. Bengtsson. Estimation of high-dimensional prior and posterior covariance matrices in kalman filter variants. *Journal of Multivariate Analysis*, 98(2):227–255, 2007.

[15] N. Hashimoto and K. Konbune. Directional spectrum estimation from a bayesian approach. In B.L. Edge, editor, *Coastal Engineering 1988 Proceedings*, pages 62–76, Torremolinos, Spain, 1988. American Society of Civil Engineers.

[16] N. Hashimoto, T. Nagai, and T. Asai. Extension of the maximum entropy principle method for directional wave spectrum estimation. In B.L. Edge, editor, *Coastal Engineering 1994 Proceedings*, pages 232–246, Kobe, Japan, 1994. American Society of Civil Engineers.

[17] M. Isobe. On joint distribution of wave heights and directions. In B.L. Edge, editor, *Coastal Engineering 1988 Proceedings*, pages 524–538, Torremolinos, Spain, 1988. American Society of Civil Engineers.

[18] M. Isobe and K. Kondo. Method for estimating directional wave spectrum in incident and reflected wave field. In B.L. Edge, editor, *Coastal Engineering 1984 Proceedings*, pages 467–483, Houston, Texas, 1984. American Society of Civil Engineers.

[19] M. Isobe, K. Kondo, and K. Horikawa. Extension of mlm for estimating directional wave spectrum. In *Symposium on Description and Modelling of Directional Seas*, pages 1–15, 1984.

[20] N.E. Karoui. Operator norm consistent estimation of large-dimensional sparse covariance matrices. *The Annals of Statistics*, 36(6):2717–2756, 2008.

[21] G.P. Nason. Wavelet shrinkage using cross-validation. *Journal of the Royal Statistical Society, Series B*, 58(2):463–479, 1996.

[22] O. Nwogu. Maximum entropy estimation of directional wave spectra from an array of wave probes. *Applied Ocean Research*, 11(4):176–182, 1989.

[23] W.J. Pierson Jr. Wind generated gravity waves. In H.E. Landsberg, editor, *Advances in geophysics*, volume 2, pages 93–178, 1955.

[24] W.J. Pierson Jr, G. Neumann, and R.W. James. Practical methods for observing and forecasting ocean waves by means of wave spectra and statistics. Publication No. 603, U.S. Navy Hydrographic Office, 1955.

[25] R. Prasad, W. Mohr, and W. Konhauser. *Third generation mobile communication systems*. Artech House, Inc., 2000.

[26] A.J. Rothman. Positive definite estimators of large covariance matrices. *Biometrika*, 99(3):733–740, 2012.

[27] A.J. Rothman, E. Levina, and J. Zhu. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104(485):177–186, 2009.

[28] M. St Dinis and W.J. Pierson Jr. On the motions of ships in confused seas. *Transactions, Society of Naval Architects and Marine Engineers*, 61:260–357, 1953.

[29] D.J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.

[30] W.B. Wu and M. Pourahmadi. Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, 90(4):831–844, 2003.

[31] L. Xue, S. Ma, and H. Zou. Positive-definite â1-penalized estimation of large covariance matrices. *Journal of the American Statistical Association*, 107(500):1480–1491, 2012.

[32] H. Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.