ARRAY PROCESSING
BASED ON
TIME-FREQUENCY ANALYSIS AND
HIGHER-ORDER STATISTICS

SUWANDI RUSLI LIE

NATIONAL UNIVERSITY OF SINGAPORE

2007

ARRAY PROCESSING BASED ON TIME-FREQUENCY ANALYSIS AND HIGHER-ORDER STATISTICS

SUWANDI RUSLI LIE

(*B.S.E.E. and M.S.E.E., University of Wisconsin - Madison, U.S.A.*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2007

Acknowledgment

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. A. Rahim Leyman, for his guidance, support, and his patience during the period of study. The fact that great freedom and patience were given has enabled me to drive the research in the direction of my own interests and hence enjoy the process of intellectual discovery. He had also generated many ideas for me to explore. Some of those ideas have been realized and came into this thesis, especially in the areas of higher-order statistics and time-frequency analysis. Many thanks also to my co-supervisor, Dr. Chew Yong Huat, for his helps and his care during my long journey towards this doctoral degree.

I also would like to thank all my colleagues, Fang Jun, Chen Xi, and Weiying for their friendship and helpful discussions. Special thanks also for my friend Teh Keng Ho, whom I met the first time in CWC and who often challenged me with his analysis in networking problem. Many thanks also for all my friends, such as Esther, Sofi, Stephanus, Mingkun, and Victor who have given me so many helps and encouragements during my study. I would like to acknowledge Agency for Science, Technology, and Research (A*STAR) and National University of Singapore for their generous financial support and Institute for Infocomm Research (I^2R) for their facilities.

Finally, profound thanks should also be given to my beloved, Jules. I am deeply indebted to her for her untiring support and encouragement, especially during the hardest time of the journey. Of course, a deep thanks to my parents, who have supported throughout my life, with constant love, wisdom, and encouragement. Last and most importantly, a wholehearted thanks to my Lord and Saviour for the love and care that see me throughout this journey.

Contents

Summary

In this thesis, we first explain the motivations behind this work and listed the type the array processing problems, which will be dealt with. Mathematical background and preliminary concepts, which are useful to this work, are reviewed in Chapter 2. In Chapter 3, two algorithms for parameter estimation of wideband LFM array signals are devised. Parameters of interest are the DOAs, initial frequencies and frequency rates. The new algorithm that uses least squares method is presented, and is extended to another algorithm by using total least squares method. In Chapter 4, a parameter estimation algorithm for the general PPS, in which LFM signal is a subclass of it, is devised. The estimation parameters are the highest-order frequency parameters and DOA. Spatial Higher-order Instantaneous Moment (SHIM) and its property are introduced and a search-free algorithm is devised. In Chapter 5, a non-parametric estimation algorithm for time-frequency signals, which is even a wider class of signals than PPS, is devised. The primary interest is to recover each of the original signals when the channel is non-invertible (resulting from the underdetermined condition of more inputs than outputs). Properties of Spatial Time-Frequency Distributions (STFDs) are discussed. Following that, the algorithm is outlined and proposed. In Chapter 6, two parametric estimation algorithms for random signals in the presence of unknown Gaussian noise are proposed. The first one is a fourth-order-statistics (FOS) -based

algorithm. The second one is a mixed-order-statistics-based algorithm, which is extended from the first algorithm. The well-known root-multiple signal classification (Root-MUSIC) algorithm is incorporated in the proposed algorithms. Finally, Chapter 7 summarizes the main contributions of the dissertation and provides the future research direction.

List of Tables

List of Figures

Abbreviations

AF	Ambiguity Function
BSS	Blind Source Separation
CBF	Chirp Beam-former
CP(s)	Cross Point(s)
CRB	Cramer-Rao Bound
CT(s)	Cross-term(s)
DOA(s)	Direction of Arrival(s)
DPT	Discrete Polynomial Transform
DTFT	Discrete Time Fourier Transform
Eqn(s).	Equation(s)
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Techniques
EVD	Eigen-Value Decomposition
FT	Fourier Transform
FFT	Fast Fourier Transform
FOS	Fourth-Order Statistics

FMCW	Frequency Modulated Continuous Wave
HAF	Higher-order Ambiguity Function
HIM	Higher-order Instantaneous Moment
HO	Higher-Order
HOS	Higher-Order Statistics
i.i.d.	Independent and Identically Distributed
JAFE	Joint Angle and Frequency Estimation
LFM	Linear Frequency Modulated
LS	Least Squares
MAP(s)	Multiple Auto Point(s)
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MUSIC	MUltiple SIgnal Classification
MWV	Modified Wigner-Ville
NMSE	Normalized Mean Square Error
pdf	Probability Density Function
PPS	Polynomial Phased Signal
SAP(s)	Single Auto Point(s)
SAR	Synthetic Aperture Radar

SAS	Synthetic Aperture Sonar
SHIM	Spatial Higher-order Instantaneous Moment
SIR	Signal-to-Interference Ratio
SNR	Signal-to-Noise Ratio
SOS	Second-Order Statistics
STFD(s)	Spatial Time-Frequency Distribution(s)
SVD	Single Value Decomposition
TF	Time-Frequency
TFD(s)	Time-Frequency Distribution(s)
TLS	Total Least Squares
UBSS	Underdetermined Blind Source Separation
WVD	Wigner-Ville Distribution

Notations

$(\cdot)^T$	Matrix transpose
$(\cdot)^*$	Complex conjugate
$(\cdot)^H$	Hermitian transpose
$(\cdot)^{-1}$	Generalized inverse
$(\cdot)^\dagger$	Moore-Penrose pseudo-inverse
$E[\cdot]$, $E\{\cdot\}$	Statistical expectation
$\|\cdot\|$, $\|\cdot\|_2$	2-norm
$\|\cdot\|_{\mathrm{F}}$	Frobenius norm
$\otimes$	Kronecker product
$\odot$	Khatri-Rao product, which is column-wise Kronecker product $\mathbf{A} \odot \mathbf{B} = [\mathbf{a_1} \otimes \mathbf{b_1}, \quad \mathbf{a_2} \otimes \mathbf{b_2}, \quad \cdots]$
$\circ$	Element-wise (Schur-Hadamard) matrix product
$[\mathbf{A}]_{l,k}$ or $(\mathbf{A})_{l,k}$	The element of matrix $\mathbf{A}$ in row l-th and column k-th
$[\mathbf{v}]_l$ or $(\mathbf{v})_l$	The l-th element of vector $\mathbf{v}$
rank($\mathbf{A}$)	The rank of matrix $\mathbf{A}$
span($\mathbf{A}$)	The column span of matrix $\mathbf{A}$
tr($\mathbf{A}$), trace($\mathbf{A}$)	Trace or sum of diagonal elements of matrix $\mathbf{A}$
diag($\mathbf{v}$)	Diagonal matrix formed by elements of vector $\mathbf{v}$

$\text{diag}(\mathbf{A})$	Column vector formed by diagonal elements of matrix $\mathbf{A}$
$\text{vec}(\mathbf{A})$	The column vector obtained from matrix $\mathbf{A}$ by stacking the column vectors of $\mathbf{A}$ from left to right.
$\mathbb{C}^n$	The set of $n \times 1$ column vectors with complex entries
$\mathbb{C}^{n \times m}$	The set of $n \times m$ matrices with complex entries
$\mathbf{I}$	The identity matrix
$\mathbf{P_A}$	The projection matrix onto space $span(\mathbf{A})$
$\Re\{c\}$	The real component of c
$\Im\{c\}$	The imaginary component of c
$\angle c$	The complex phase/angle of c

Chapter 1

Introduction

1.1 Background

Generally, this thesis focused on the parametric and non-parametric estimation of signals in array systems. The parameters to be estimated include DOA and the frequency parameters of signals. The most classical frequency parameter estimation is the signal spectral estimation, which is still of interests in many applications. In addition to that, research scope on spectral estimation has been broadened over the last decades, not only just applying to sinusoidal signals but also applying to wider class of signals which are more suitable in the real world settings. In the following subsection, we will introduce polynomial phase signals (PPS), which is the class of signals that this thesis is focused in. Thereafter, three non-classical array processing problems which will be studied from Chapter 2 onward are introduced.

1.1.1 Polynomial Phase Signals

Most of the research focused in spectral estimation of sinusoid signals. This class of signals consists of signals with their phases being a linear function of time, or equivalently, their (instantaneous) frequencies are constant. Estimation of the frequency of this class of signals has been well investigated. A more general class of signals consists of PPS where, as its name implies, its phase, $\phi(t)$, is a polynomial function of time (see Eqn. (1.1)). Furthermore, this class of signals also has its frequency varies as a polynomial function of time, because its angular instantaneous frequency, $\phi'(t)$, is just the derivative of the phase with respect to time.

$$\begin{cases} s(t) & \triangleq Ae^{j\phi(t)} \\ \phi(t) & \triangleq \sum_{i=0}^{K} a_i t^i \\ \phi'(t) & = \frac{d\phi(t)}{dt} = \sum_{i=0}^{K} i a_i t^{i-1} \end{cases} \tag{1.1}$$

A very common example in this class of signals is the linear chirp signal, where the phase is a quadratic function of time ($K = 2$ in Eqn. (1.1)). Thus, the frequency of this chirp signal is a linear function of time and hence it is also referred as LFM signal.

Polynomial phase signals occur in natural phenomena, e.g., gravity waves [1] and seismography. Bats' sonar-like maneuver and their way of navigating relying on chirp (second-order PPS) are of interest to researchers for a long time. Aside

from that, applications of chirp signals have also been reported in radar [2] and sonar [3].

The thesis also looks into multi-component PPS signal, which is defined as

$$r(t) = \sum_{i=1}^{K} s_i(t)$$

where each $s_i(t)$ is of the form of Eqn. (1.1) with its own set of frequency parameters.

1.1.2 Radar Applications

Generally, radar can be classified as two major groups, i.e. pulse radar and FMCW radar. A pulse radar transmits the pulse wave such that when the reflected wave received by radar, the propagation time can be measured from the duration from the moment the pulse is transmitted to the moment the the reflected pulse is received. On the other hand, a FMCW radar does not transmit a short pulse signal but transmits continuous signal. This radar changes the frequency of the sinusoid signal linearly as a sawtooth function within a frequency band. To extract the propagation time, the received signal, and transmitted signal are multiplied and passed through a low pass filter. The output signal after passing through a filter will be a single sinusoid with frequency Δf directly proportional to the distance from the target (see Fig. 1.1). This operation together with Fourier transform

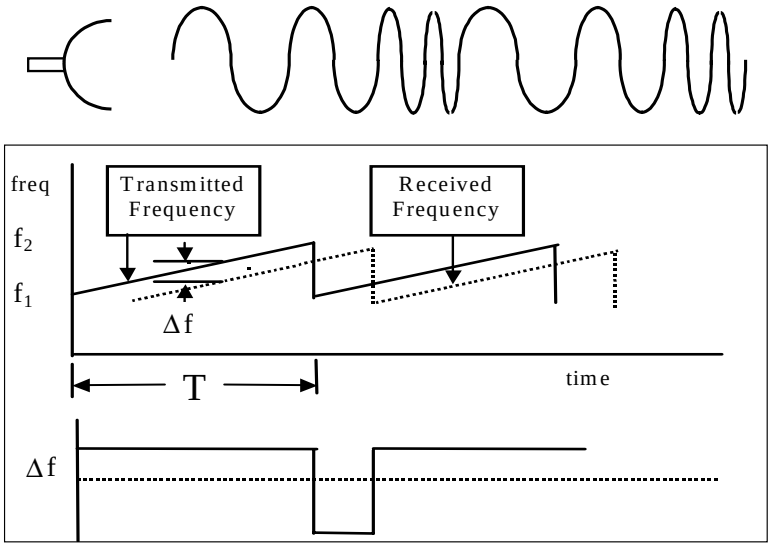

Figure 1.1: The FMCW radar transmitted (solid) and received signal frequency (dashed). The region where the Δf is valid is in region T

(FT) for frequency analysis is actually called ambiguity function (AF); we will generalize AF to higher-order ambiguity function (HAF) in the following chapters. Mathematically, this AF operation in the complex form is written in the form

$$Af(k) \triangleq FT\{s(\Delta n)r^*(\Delta n)\} \tag{1.2}$$

where $s(\Delta n)$ are the samples of the current transmitted signal and $r(\Delta n)$ the samples of the reflected/received signal. The more general form of AF is defined as

$$Af(\gamma, \tau) \triangleq \int x(t-\tau)x^*(t+\tau)e^{-jt\gamma}dt \tag{1.3}$$

where $x(t)$ is the signal or data for the analysis, τ is delay parameter, and γ is a dummy variable.

If there is only one FMCW radar operating in a certain frequency band, the radar is capable of detecting multiple objects and estimating their relative positions from radar. However, in the case of multiple transmitting radars operating in the same frequency band, such as in anti-collision warning system of automobiles, each radar will create interference burying the signal reflected from the targets. This is critical as it could create collisions on the road.

In order to understand this vividly, suppose that there are one main radar, one interference radar, and one object. The signals transmitted by the main radar and the interference radar during period T are $s_o(t) \triangleq A_o e^{j\omega_o t+\nu_o t^2}$ and $s_i(t) \triangleq A_i e^{j\omega_i t+\nu_i t^2}$, respectively. Assuming also that the signal scattered by the object to main radar is only the signal transmitted from main radar, then the noise-free received signal by the main radar is $r(t) = s_o(t-\tau) + s_i(t)$, where, without loss of generality, the delay time for $s_i(t)$ to reach the receiver has been ignored. The result from the radar ambiguity function would be the FT of the following $y(t)$,

$$y(t) = A_1 exp\{j(2\nu_o\tau t + \omega_o\tau - \nu_o\tau^2)\} + A_2 exp\{j((\omega_o - \omega_i)t + (\nu_o - \nu_i)t^2)\}$$

where A_1 and A_2 contain the attenuated amplitudes of A_o^2 and $A_o A_i$. The second term of $y(t)$ will not appear if there is no interference radar. The second term is a chirp signal, which will bury the signal of interest if its received amplitude is large, because the chirp component has energy spreads over the entire frequency band of interest. Hence, suppression of this chirp component would be important. This

could be done by estimating the frequency rate and removing the second term through filtering (advert to Chapter 3).

Another example is in the application of Doppler radar, where the relative velocity of the object toward or away from the radar is proportional to the Doppler frequency shift of the object. Furthermore, if the object is accelerating radially then the radial acceleration is proportional to the Doppler frequency sweep rate, i.e., frequency rate. Hence, estimation of frequency rate is essential to extract the acceleration of the object. Therefore, the knowledge of initial frequencies and frequency rates will give the knowledge of the distance of the objects from the radar, the radial acceleration, and the radial velocity of the object. Consequently, estimation of these parameters, or in general the parameters of PPS, would be essential for various radar applications.

1.1.3 Array Processing

Basically, all of the problems covered by this thesis are in the area of array processing, which can also be treated as multiple-input and multiple-output (MIMO) problems. From practical standpoint, the setting can be interpreted as multiple-antenna base station receiving signals from multiple users, or the antenna array of radar receiving signals reflected from multiple targets. There are many more problems can be interpreted from this array processing setting. Figure 1.2 summarizes the general model considered in this thesis.

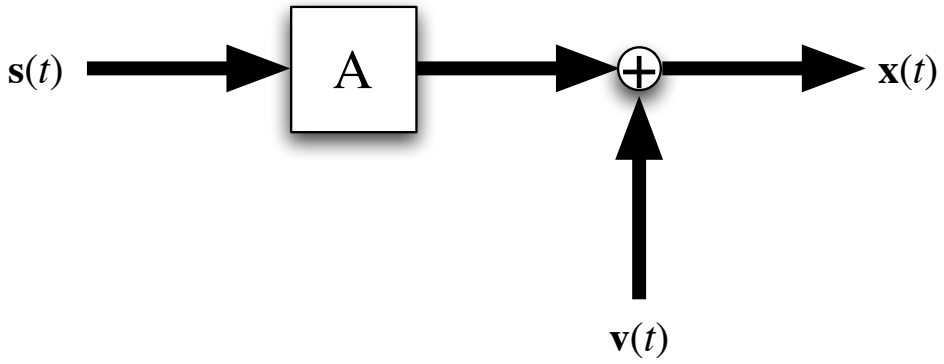

Figure 1.2: The Channel Input-Output Model

Classically, sources are assumed to be narrowband, such that the channel (mixing matrix or array manifold) $\mathbf{A}$ is undergone flat fading. The channel is further assumed to be unchanged within the estimation period. The noise is assumed to be spatially and temporally white Gaussian noise. The literature survey of the classical parametric DOA estimation methods could be found, for example in [4]. In this thesis, the channel is assumed to be non-convolutive. Furthermore, there are three different types of non-classical problems that are under consideration, and are illustrated in Fig. 1.3 (b), (c) and (d).

In the first type is the array processing problem shown in Fig. 1.3 (b) the channel, $\mathbf{A}(\boldsymbol{\theta}, t)$, and the multiple input or transmitted signals, $\mathbf{s}_{\boldsymbol{\theta}}(t)$, are modeled to be function of parameters, $\boldsymbol{\theta}$. The objective in parametric array processing is to estimate these parameters. In this thesis, the parameters, $\boldsymbol{\theta}$, include DOAs, frequencies, frequency rates, and other frequency-related parameters of the sources. If the interest is to recover these signals, they could be constructed by estimating these parameters. Alternatively, the estimated parameters could also be used in

beamforming. The main difference from the classical parametric array processing, owing to the wideband LFM or PPS case, is that the channel $\mathbf{A}(\theta, t)$ is also function of time, which is theoretically a very challenging problem to deal with compared to the classical case in Fig. 1.3 (a). This is because the estimation of the signal covariance matrix is difficult, attributable to non-ergodicity of the observed signal covariance. The noise $\mathbf{v}(t)$ is assumed to be spatially and temporally white Gaussian noise.

The second type is the parametric array processing of random sources with possibly correlated noise (Fig. 1.3 (c)). However, here the sources, $\mathbf{s}(t)$, are narrowband random signals, which are not parametrically modeled and only the channel, $\mathbf{A}(\boldsymbol{\theta})$ is assumed to be a function of parameters, which are DOAs. Here, the objective is to estimate DOAs. The main difference from the classical parametric array processing is that the noise, $\mathbf{v}(t)$, is not restricted to spatially and temporally white Gaussian noise. In fact, in many applications, the noise is not always white spatially and temporally. If one is interested in restoring the original signals, it could be done by solving the least squares problem (by using pseudo-inverse of the channel, $\hat{\mathbf{s}}(t) = \mathbf{A}^{\dagger}(\hat{\boldsymbol{\theta}})\mathbf{x}(t)$), because the channel, $\mathbf{A}(\boldsymbol{\theta})$, is independent of time and has a known structure. Again, one could use the estimated DOAs for beamforming applications if the interest is not to restore the original source signals.

The third type is the non-parametric estimation of time-frequency signals or non-stationary signals (Fig. 1.3 (d)). The signals $\mathbf{s}(t)$ is assumed to be narrow-

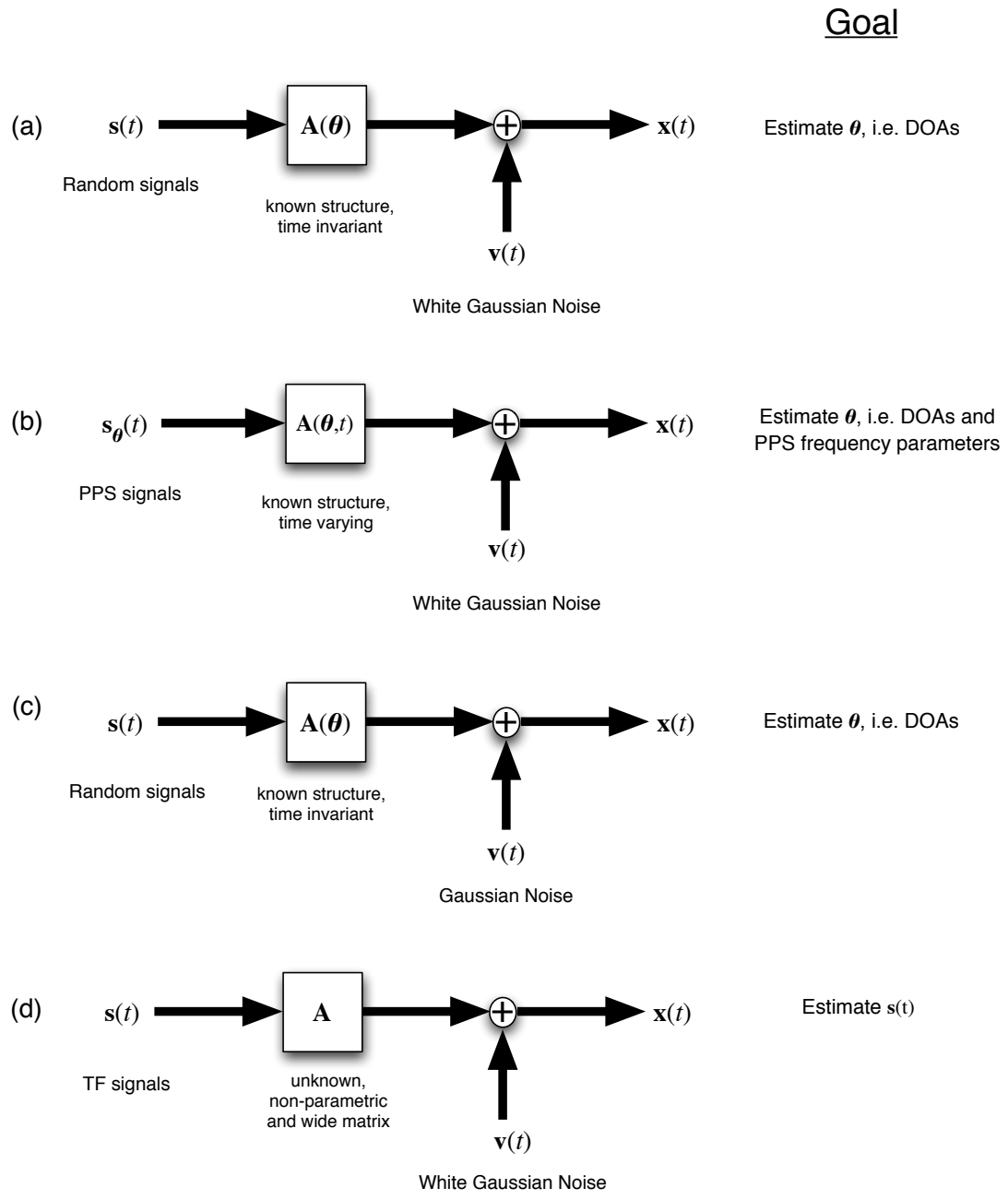

Figure 1.3: (a). Classical parametric array processing, (b). First case: PPS array processing, (c). Second case: array processing in presence of unknown zero-mean Gaussian noise, (d) Third case: non-parametric (blind) array processing

band, and the channel, $\mathbf{A}$, is assumed to be independent of time and parameters, however, it is unknown. Besides that, it is assumed that there are more sources

than receiving antennas, and hence, the channel can be represented by a wide matrix. Practically, this could happen when there are more users transmitting than base station's antennas in a single cell. In this condition, even if $\mathbf{A}$ is known, one cannot obtain $\mathbf{s}(t)$ directly by solving least squares problem described briefly in the second type of array processing. Here, the objective is to obtain $\mathbf{s}(t)$, which is unknown, but each signal is assumed to have a distinct time-frequency signature. This type of non-parametric estimation where the channel is unknown is called blind source separation (BSS) and its literature surveys could be found in [5]. The BSS problem that assume more sources than sensors is called underdetermined BSS (UBSS).

1.2 Organization of the Thesis and Contributions

The organization of the thesis is as follows: In Chapter 2, mathematical background and preliminary concepts are covered. Time-frequency distributions, which are the core for analyzing the non-stationary signals, are discussed. The quadratic time-frequency distributions and some of their properties are briefly discussed. Higher-order statistics employed in this thesis, such as cumulants and moments, are also explained. The signal models that were explained in the previous section will be elaborated in detail in Chapter 2. Following that, some of subspace-based DOA estimation techniques are reviewed.

In Chapter 3, two algorithms for parameter estimation of wideband LFM array signals for the first type of array processing setting are devised. Parameters of interest are the DOAs, initial frequencies and frequency rates. Initially, a review of the existing algorithms, as well as, their strengths and weaknesses, are presented. Following that, the mathematical model is reviewed for the LFM signals in order to demonstrate the idea behind the proposed algorithm. The first algorithm that uses least squares method is presented, and is extended to the second algorithm by using total least squares method. Simulation results are presented and comparison with an existing algorithm is made. Finally, Cramer-Rao Bound (CRB) and the performance analysis are derived.

Most of the materials in Chapter 3 have been published in

- S. Lie, A. R. Leyman and Y. H. Chew, "Parameter estimation of wideband chirp signals in sensor arrays through DPT," in *Proc. 37th Asilomar Conf. on Sign., Syst. and Comp.*, Pacific Grove, CA, Nov. 2003.

- S. Lie, A. R. Leyman and Y. H. Chew, "Wideband chirp parameter estimation in sensor arrays through DPT," *IEE Electronic Letters*, vol. 39, no. 23, pp. 1633-1634, Nov. 2003.

In Chapter 4, a parameter estimation algorithm for the class of PPS, in which LFM signal is a subclass of, is devised. The estimation parameters are the highest-order frequency parameter and DOA. In the case of LFM signal, and quadrature

FM signal, the highest-order frequency parameters are frequency rate for LFM signal, and frequency acceleration for quadrature FM signal. Spatial Higher-order Instantaneous Moment (SHIM) and its property are introduced in Chapter 4. Furthermore, a review on the joint angle-frequency estimation algorithms is also presented. The proposed algorithm is devised using SHIM. Thereafter, a brief analysis and the identifiability condition are discussed. Finally, results are presented and comparison with Maximum Likelihood (ML) estimation is demonstrated.

Most of the materials in Chapter 4 have been published in

- S. Lie, A. R. Leyman and Y. H. Chew, "Wideband polynomial-phase parameter estimation in sensor array," in *Proc. of the 3rd IEEE International Symposium on Sign. Proc. and Info. Tech.*, Darmstadt, Germany, Dec 2003.

In Chapter 5, a non-parametric estimation algorithm for time-frequency signals, even wider class than PPS, is devised. The primary interest is to recover each of the original signals even if the channel is unknown and non-invertible (resulting from the underdetermined condition of more inputs than outputs). This chapter starts with a brief review of existing algorithms and introduction to the problem. Properties of Spatial Time-Frequency Distributions (STFDs) are discussed. Following that, the algorithm is outlined and proposed. A new property of the existing subspace separation method is discussed and employed in the proposed algorithm. Simulation results are presented to show its effectiveness. The results are also compared with the existing algorithm.

Most of the materials in Chapter 5 have been published in

- S. Lie, A. R. Leyman and Y. H. Chew, "Underdetermined source separation for non-stationary signal," *The 32nd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Hawaii, USA, April 2007.

In Chapter 6, two parametric estimation algorithms for random signals in the presence of unknown Gaussian noise are proposed. Introduction and review of the existing algorithms are discussed. A fourth-order-statistics (FOS) -based algorithm is devised and it is extended to mixed-order-statistics-based algorithm. Simulation results are demonstrated and compared to an existing fourth-order (FO) algorithm and an existing second-order (SO) algorithm. Root-multiple signal classification (Root-MUSIC) algorithm is incorporated in the proposed algorithms. Thereafter, we end the chapter with a short discussion.

Most of the materials in Chapter 6 have been published in

- S. Lie, A. R. Leyman and Y. H. Chew, "Fourth-order and weighted mixed order direction of arrival estimators," *IEEE Signal Processing Letters*, vol. 13, no.11, Nov 2006.

Finally, Chapter 7 summarizes the main contributions of the dissertation. The directions of the future research are discussed.

Chapter 2

Mathematical Preliminaries

In this chapter, we review some of the background and mathematical theories, which will be used in the thesis. The scopes to be covered include time-frequency distributions (TFDs), cumulants, moments, and subspace-based direction-of-arrival estimation methods. Readers are assumed to have some basic understanding in parameter estimation theory and time-frequency analysis, hence the review on these topics is only minimally elaborated.

2.1 Time-Frequency Distributions

In this section we define the TFD of a signal. The reason why the TFD of a signal is important is because most signals encountered in many real-life situations are not necessarily stationary, e.g., speech, music, and PPS. A signal is said to be non-stationary if its intrinsic characteristics vary with time [6]. For example, in speech and music, we could clearly hear the variations of frequencies or notes over

time. If we apply FT to this type of signal, we can only observe the frequency content of the signal. To observe vividly, see the following illustration in Fig. 2.1. In the illustration, we show how the frequency of a signal with time-varying frequency changes with time in the $\omega - t$ plane. Applying FT to the signal only gives the three frequencies shown along the ω-axis. Hence, FT does not allow one to observe how the frequencies vary in time. Therefore, the FT is not suitable to analyze the non-stationary signals. The preferred method to analyze this type of signals is to use a description of the signal that involves both time and frequency. This method is called time-frequency (TF) analysis, which maps a signal (i.e., a one-dimensional function of time) onto an image (i.e., a two-dimensional function of time and frequency) that displays the spectral components of the signal as a function of time (see the illustration on the box in Fig. 2.1). Conceptually, one may think of this mapping as a time-varying spectral representation of the signal, analogous to musical score.

2.1.1 Definitions

In general, let us define this time-varying spectral representation as $P(t, \omega)$. The definition of the moments or global averages of time and frequency are

$$\begin{aligned} \langle \omega \rangle &\triangleq \int\int \frac{1}{|s(t)|^2} \omega P(t,\omega) dt d\omega \\ \langle t \rangle &\triangleq \int\int \frac{1}{|S(\omega)|^2} t P(t,\omega) dt d\omega \end{aligned} \tag{2.1}$$

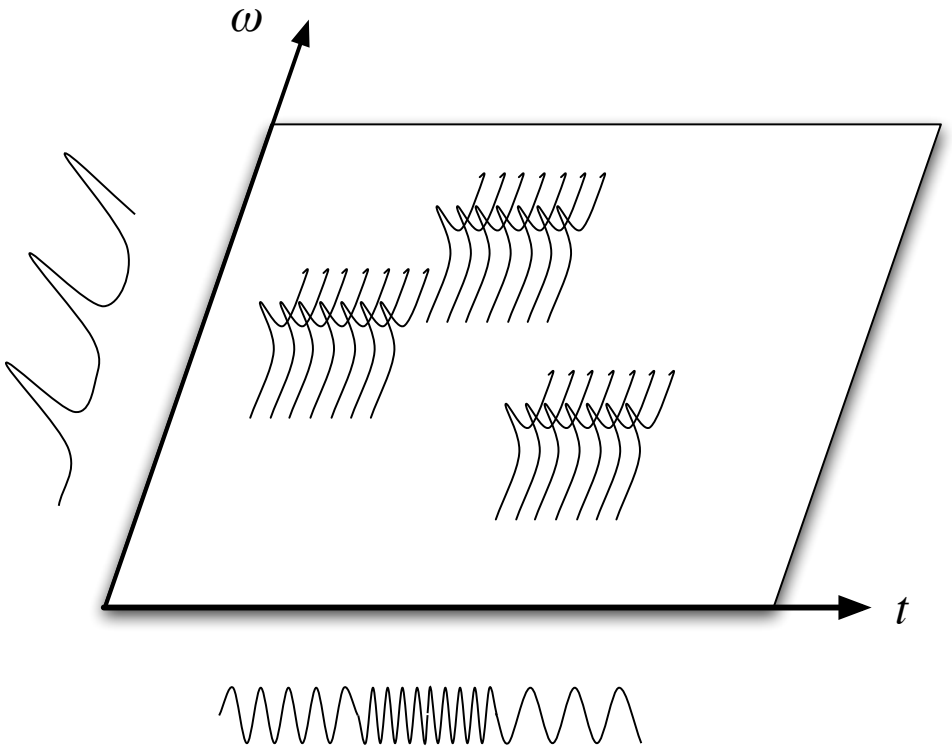

Figure 2.1: Signal with varying frequencies over time

where $s(t)$ and $S(\omega)$ are the signal and its Fourier-tranformed pair. The definition of the conditional moments or local averages of time and frequency are

$$\begin{aligned} \langle \omega \rangle(t) &\triangleq \frac{1}{|s(t)|^2} \int \omega P(t,\omega) d\omega \\ \langle t \rangle(\omega) &\triangleq \frac{1}{|S(\omega)|^2} \int t P(t,\omega) dt \end{aligned} \tag{2.2}$$

Moments of the function of time and frequency, are defined as follows

$$\begin{aligned} \langle g(\omega) \rangle(t) &\triangleq \frac{1}{|s(t)|^2} \int g(\omega) P(t,\omega) d\omega \\ \langle h(t) \rangle(\omega) &\triangleq \frac{1}{|S(\omega)|^2} \int h(t) P(t,\omega) dt \end{aligned} \tag{2.3}$$

where $g(.)$ and $h(.)$ are continuous functions. With these definitions, we have the conditional spread in time and frequency defined as,

$$\begin{aligned} \sigma^2_{t|\omega} &\triangleq \langle t^2 \rangle(\omega) - \langle t \rangle^2(\omega) \\ \sigma^2_{\omega|t} &\triangleq \langle \omega^2 \rangle(t) - \langle \omega \rangle^2(t) \end{aligned} \tag{2.4}$$

Assuming that the signal has a model as follows: $s(t) = A(t)\exp(j\phi(t))$, which is the typical model of speech and communication signals, where $A(t)$ is the slow time-varying amplitude and $\phi(t)$ is the time-varying phase. The instantaneous angular frequency is then defined as $\phi'(t) \triangleq \frac{\partial \phi(t)}{\partial t}$.

2.1.2 Types of TFD

Generally, TFDs could be classified into two classes. One class is known as the linear TFD, such as spectrogram of windowed FT and scalogram of wavelets transform. It is called linear because the operator applied to the sum of signals is equal to the sum of the operators applied to each of the signals. The operator in this case could be windowed FT or wavelets transform. In the next section, we will briefly describe the windowed FT because it is related to the well-established FT. The other class of distribution is called quadratic distribution, such as Cohen's Class distribution and Wigner-Ville distribution (WVD). It is called quadratic because the operator applied to, e.g., sum of two signals, will lead to sum of the operators

applied to each of the signals plus the operators applied to the product of the two signals. We will express this quadratic property mathematically when defining the Cohen's Class distributions.

2.1.3 Windowed Fourier Transform

Windowed Fourier transform, or Short-time Fourier transform, of the signal $f(t)$ is defined as follows

$$F(t,\omega) \triangleq \int_{-\infty}^{\infty} f(u)g(u-t)e^{-j\omega u}du \tag{2.5}$$

with a symmetric window $g(t) = g(-t)$ which is also normalized[§], i.e., $\|g\|_{L^2} = 1$. The multiplication window $g(u-t) = \delta(u-t)$ localizes the Fourier integral in the neighborhood of $t = u$. Hence, the window determines the TF support or the resolution of the transform, which is independent of the signal $f(t)$ and location of signal in TF plane. In the following examples, we will illustrate this property.

Example 1. Suppose the signal is $f(t) = e^{j\omega_0 t}$ then its windowed FT is

$$F(t,\omega) = G(\omega-\omega_0)e^{-jt(\omega-\omega_0)} \tag{2.6}$$

For a given time instant, its energy is spread over interval $[\omega_0 - \sigma_{\omega|t}/2, \omega_0 + \sigma_{\omega|t}/2]$. Here $G(\omega)$ is the FT of $g(t)$, and $\sigma^2_{\omega|t}$ is the conditional frequency spread of $G(\omega)$.

[§] L^2-norm: $\|g\|^2_{L^2} = \int |g(t)|^2 dt$

Example 2. The windowed FT of a Dirac $f(t) = \delta(t - t_0)$ is given by

$$F(t, \omega) = g(t_0 - t)e^{-jft_0} \tag{2.7}$$

Hence, for a given frequency, its energy is spread over the time interval $[t_0 - \sigma_{t|\omega}/2, t_0 + \sigma_{t|\omega}/2]$. Here $\sigma^2_{t|\omega}$ is the conditional time spread of $g(t)$.

Example 3. Consider a chirp (LFM) with a Gaussian envelope and a Gaussian window,

$$s(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\beta t^2/2 + j\omega_0 t} \quad \text{and} \quad g(t) = (a/\pi)^{1/4} e^{-at^2/2} \tag{2.8}$$

where $\frac{1}{2\alpha}$ and $\frac{1}{2a}$ are the variances of Gaussian distributions. The spectrogram, which is defined as

$$S(t, \omega) \triangleq \|F(t, \omega)\|^2, \tag{2.9}$$

is given by [7]

$$S(t, \omega) = \frac{P(t)}{\sqrt{2\pi\sigma^2_{\omega|t}}} \exp\left[-\frac{(\omega - \langle\omega\rangle(t))^2}{2\sigma^2_{\omega|t}}\right] \tag{2.10}$$

$$= \frac{P(\omega)}{\sqrt{2\pi\sigma^2_{t|\omega}}} \exp\left[-\frac{(t - \langle t\rangle(\omega))^2}{2\sigma^2_{t|\omega}}\right] \tag{2.11}$$

where

$$P(t) = \sqrt{\frac{a\alpha}{\pi(\alpha+a)}}\exp\left[-\frac{a\alpha}{\alpha+a}t^2\right] \tag{2.12}$$

$$P(\omega) = \sqrt{\frac{a\alpha/\pi}{\alpha a^2+a(\alpha^2+\beta^2)}}\exp\left[-\frac{a\alpha}{\alpha a^2+a(\alpha^2+\beta^2)}\omega^2\right] \tag{2.13}$$

and

$$\langle\omega\rangle(t) = \frac{a}{a+\alpha}\beta t+\omega_0 \tag{2.14}$$

$$\sigma^2_{\omega|t} = \frac{1}{2}(\alpha+a)+\frac{1}{2}\frac{\beta}{\alpha+a} \tag{2.15}$$

$$\langle t\rangle(\omega) = \frac{a\beta}{\alpha a^2+a(\alpha^2+\beta^2)}\omega \tag{2.16}$$

$$\sigma^2_{t|\omega} = \frac{1}{2}\frac{(\alpha+a)^2+\beta^2}{\alpha a^2+a(\alpha^2+\beta^2)} \tag{2.17}$$

The concentration of energy for an instantaneous time is along the estimated instantaneous frequency, $\langle\omega\rangle(t)$, given by Eqn. (2.14). Similarly, for a given frequency, its concentration of energy is along estimated group delay, $\langle t\rangle(\omega)$, given by Eqn. (2.16). As the window gets narrow, i.e. $a \to \infty$, the estimate of instantaneous frequency approaches $\beta t+\omega_0$. However, with this limit, the estimate of group delay approaches zero. This is understandable, because as $a \to \infty$ we have a flat window in frequency domain, which corresponding to the case where there is no windowing.

Conversely, if we want to focus on temporal properties for a given frequency,

we must take $a \to 0$. In this case $\langle t \rangle(\omega) \to \beta/(\alpha^2 + \beta^2)$, which gives the correct group delay. Thus, we conclude that in spectrogram, $\langle \omega \rangle(t)$ does not always give the actual instantaneous frequency and group delay. They are dependent on the window function chosen.

2.1.4 Cohen's Class Distribution

The Cohen's class of quadratic distribution [8] is defined as follows

$$P(t,\omega) \triangleq \frac{1}{4\pi} \int\int\int e^{-j(\theta t+\tau\omega-\theta u)} \kappa(\theta,\tau) s^*(u-\frac{\tau}{2}) s(u+\frac{\tau}{2})\, du\, d\tau\, d\theta \qquad (2.18)$$

where τ, θ and u are the dummy variables, and $\kappa(\theta,\tau)$ is the kernel. The choices of kernels and their properties could be found in [7, 9]. When the kernel is equal to one, i.e., $\kappa(\theta,\tau) = 1$, Eqn. (2.18) gives the WVD, which is given by

$$\mathrm{WV}_s(t,\omega) \triangleq \frac{1}{2\pi} \int s^*(t-\tau/2) s(t+\tau/2) e^{-j\omega\tau} d\tau \qquad (2.19)$$

This can be shown by noting that $\int e^{-j\theta(t-u)} d\theta = 2\pi\delta(t-u)$. Meanwhile, we will consider only WVD in order to observe its properties, which are also inherited by some members of the Cohen's class distributions. The WVD has many properties [7, 10]. Only three of them are shown here, because these properties distinguish WVD and some of the Cohen's class members from the windowed FT.

Property 1: WVD Time-Frequency Support

Time-frequency support means that the distribution is zero whenever signal or spectrum is zero. Suppose that a signal $s(t)$ has energy that is concentrated around (t_0, ω_0), then the WVD of $s(t)$ also has its energy centered at (t_0, ω_0), with equal time and frequency support (spread of $s(t)$ and its spectrum). This is illustrated by the following proposition [10]:

Proposition 1. *If the support of $s(t)$ is $[t_0 - T/2, t_0 + T/2]$, then for all ω the support in t of $\mathrm{WV}_s(t,\omega)$ is included in this interval. If the support of signal spectrum, $S(\omega)$ is $[\omega_0 - \Delta/2, \omega_0 + \Delta/2]$, then for all t the support in ω of $\mathrm{WV}_s(t,\omega)$ is included in this interval. (see Mallat's [10] for a proof)*

Hence, the direct consequences of the above proposition are:

$$\text{If} \quad s(t) = \delta(t - t_0) \implies \mathrm{WV}_s(t,\omega) = \delta(t - t_0)$$

$$\text{If} \quad s(t) = exp(j\omega_0 t) \implies \mathrm{WV}_s(t,\omega) = \frac{1}{2\pi}\delta(\omega - \omega_0)$$

This implies that WVD does not spread the time or frequency support of Dirac or sinusoid functions, unlike windowed FT (c.f. *Examples 1 and 2*)

Property 2: The Local Average Frequency of WVD at t_0 gives $\phi'(t_0)$

The instantaneous frequency of FM signals at any fixed time instant t_0 is just the local average of the frequency computed relative to the WVD. A frequency modulated signal is defined as $s(t) = a(t)exp(j\phi(t))$, where $a(t)$ is the slow time-varying amplitude and $\phi'(t)$ is the frequency modulation or the instantaneous frequency. Therefore, for a fixed t_0, the mass or energy of WVD is concentrated around the instantaneous frequency. Formally, this property is given by the following proposition [10].

Proposition 2. *If $s(t) = a(t)exp(j\phi(t))$, then*

$$\phi'(t) = \frac{\int \omega \mathrm{WV}_s(t,\omega)d\omega}{|s(t)|^2} \tag{2.20}$$

(see Mallat's [10] for the proof)

For example, from the same chirp signal with Gaussian envelope in *Example 3*, we obtain its WVD as,

$$WV(t,\omega) = \frac{1}{\pi} e^{-\alpha t^2 - (\omega - \beta t - \omega_0)^2)/\alpha} \tag{2.21}$$

From Eqn. (2.21), the energy is concentrated at $\omega = \beta t + \omega_0$ which is the instantaneous frequency of the signal. This is not the case in *Example 3* when windowed FT is used. However, in *Example 1*, the energy is concentrated along the instan-

taneous frequency. We can conclude that in windowed FT, local average of the distribution does not guarantee to give instantaneous frequency (this also applies to the group delay, which is not discussed here).

Properties 1 and 2 are the main advantages of some of the Cohen's class members, including the WVD, over the windowed FT. In the following, we will show the drawback of quadratic distributions.

Property 3: The Interferences or Cross-Term

Interference or cross-term is quite a severe issue because it could become nuisance in some applications, for example, one might detects more signals are present than there are, because the interference appears like signal after being processed by the quadratic TFD. This topic has generated many interests from many people [8, 11–15]. All of them try various ways to suppress the interference or the cross-terms. However, these cross-terms could be useful if one knows how to exploit them, which will be shown in Chapter 5. The cross-terms arise when the WVD is applied to a composite signal, which is formed by the linear combination of two or more distinct signals. For example, if $s(t) = g(t) + h(t)$ then the WVD will produce two cross-terms, i.e., $\mathrm{WV}_{g,h}(t,\omega) + \mathrm{WV}_{h,g}(t,\omega)$, mathematically,

$$\begin{aligned}\mathrm{WV}_s(t,\omega) &= \mathrm{WV}_g(t,\omega) + \mathrm{WV}_{g,h}(t,\omega) + \mathrm{WV}_{h,g}(t,\omega) + \mathrm{WV}_h(t,\omega) \\ &= \mathrm{WV}_g(t,\omega) + 2\Re\{\mathrm{WV}_{g,h}(t,\omega)\} + \mathrm{WV}_h(t,\omega)\end{aligned} \tag{2.22}$$

where the cross-term is defined as

$$\mathrm{WV}_{h,g}(t,\omega) \triangleq \int h(t+\tau/2)g^*(t-\tau/2)e^{-j\omega\tau}d\tau \tag{2.23}$$

2.1.5 Ambiguity Function

Ambiguity Function (AF) is defined as follows,

$$\mathrm{AF}(\theta,\tau) \triangleq \int s^*(u-\tau/2)s(u+\tau/2)e^{j\theta u}\,du \tag{2.24}$$

where u is a dummy variable, and τ is the delay variable. Ambiguity Function is related to Cohen's class distribution in Eqn. (2.18) as follows,

$$P(t,\omega) = \frac{1}{4\pi^2}\int\int M(\theta,\tau)e^{-j\theta t - j\tau\omega}\,d\theta\,d\tau \tag{2.25}$$

where

$$M(\theta,\tau) = \kappa(\theta,\tau)\mathrm{AF}(\theta,\tau) \tag{2.26}$$

From Eqn. (2.25), $P(t,\omega)$ is related to $M(\theta,\tau)$ through 2-dimensional (2-D) FT. Hence, the multiplication of kernel in the ambiguity domain ($\theta-\tau$ domain) in Eqn. (2.26) with appropriate low-pass masking will lead to cross-terms smoothing in TF domain ($t-\omega$ domain), e.g. by Choi-Williams kernel $\kappa(\theta,\tau) = e^{-\sigma^2\tau^2\theta^2}$. The idea of smoothing is based on the fact that cross-terms are oscillatory in TF

domain (see [10]).

2.1.6 Higher-order Ambiguity Function (HAF)

The higher-order variant of AF is called higher-order ambiguity function (HAF) [16, 17], and its discrete-time implementation is called discrete polynomial transform (DPT) [18] . The DPT is defined as follows,

$$\mathrm{DPT}_K[s(n),\tau] \triangleq \mathrm{DTFT}\{DP_K[s(n),\tau]\} \tag{2.27}$$

where DTFT denotes the discrete-time Fourier transform and $DP_K[\cdot]$ is the higher-order instantaneous (HIM) operator. The HIM operator [19] is defined as follows,

$$\begin{aligned} DP_1[s(n),\tau] &\triangleq s(n) \\ DP_2[s(n),\tau] &\triangleq s(n)s^*(n-\tau) \\ &\vdots \\ DP_K[s(n),\tau] &\triangleq DP_2[DP_{K-1}[s(n),\tau],\tau] \end{aligned} \tag{2.28}$$

and τ is an arbitrary positive integer parameter which could be chosen according to [18]. The main usage of DPT or HAF is in PPS applications as we will see in Chapters 3 and 4.

2.2 Moments and Cumulants

In this section, we review the definitions of statistical moments and cumulants, because we will use them throughout the thesis. When random variables are non-Gaussian, the first two moments are not sufficient to define their probability density functions or cumulative distribution functions (pdf or cdf). Consequently, higher-order statistics (HOS), can reveal other information about them than just second-order statistics (SOS) can provide. Ideally, the entire pdf is needed to characterize a non-Gaussian random variable, however, in practice this may not be always available. Under these circumstances, the pdf may be characterized by its moments or cumulants. It should be noted that some distributions do not possess finite moments of all orders. Fortunately, for sufficiently large number of the sources of randomness and under certain conditions stated in page 538 of [20], the cdf of the random variable could be approximated by Edgeworth's expansion involving HOS to up to fourth-order [21]. In fact, if the number of the sources of randomness is n, then the error of approximation is in the order of less than $1/n$. Thus, with just up to fourth-order statistics, one could approximate the cdf very well. In practice, the approximations often turn out to be remarkably good even when only the first three or four moments are used [22].

2.2.1 Definitions and Properties

We focus our presentation on real random variables. For the complex case, readers can refer to [22, 23] and the references therein. Let $\{y(n)\}, n = 0, \pm 1, \pm 2, \pm 3, \ldots$ be a random process, stationary up to order q; then, the p^{th}-order moment ($p \leq q$), $M_{p,y}(\tau_1, \tau_2, \ldots, \tau_{p-1})$, is defined as the joint p^{th}-order moment of the random variables, $y(n), y(n+\tau_1), \ldots, y(n+\tau_{p-1})$. Because of the assumed stationarity, the p^{th}-order moment is a function only of the $(p-1)$ lags, $\{\tau_i\}_{i=1}^{p-1}$. We now write the moment of a stationary random process as

$$\begin{aligned} M_{p,y}(\tau_1, \tau_2, \ldots, \tau_{p-1}) &\stackrel{\triangle}{=} Mom[y(n), y(n+\tau_1), \ldots, y(n+\tau_{p-1})] \\ &= E[y(n)y(n+\tau_1)\ldots y(n+\tau_{p-1})] \end{aligned} \tag{2.29}$$

where $E[\cdot]$ is the statistical expectation operator. The p^{th}-order cumulant exists, if all absolute moments of q^{th}-orders $q \leq p$ exist (and are bounded). Similarly, all the p^{th}-order cumulants of $\{y(n)\}$ are $(p-1)$-dimensional functions, which can be written in the form

$$C_{py}(\tau_1, \tau_2, \ldots, \tau_{p-1}) \stackrel{\triangle}{=} \text{Cum}[y(n), y(n+\tau_1), \ldots, y(n+\tau_{p-1})]. \tag{2.30}$$

The general relationship between moments and cumulants of any order can be found in [24]. Cumulants of orders greater than one are invariant to shift of mean. We will assume that the processes of interest are all with zero-mean.

Hence, the second-order moment (autocorrelation) of the zero-mean random process $\{y(n)\}$ is defined as

$$M_{2,y}(\tau) \triangleq E[y(n)y(n+\tau)]. \tag{2.31}$$

In this case, the second-order cumulants $C_{2,y}(\tau)$ are the same as $M_{2,y}(\tau)$, i.e. $C_{2,y}(\tau) = M_{2,y}(\tau)\ \forall\tau$. The third-order moment is defined as

$$M_{3,y}(\tau_1, \tau_2) \triangleq E[y(n)y(n+\tau_1)y(n+\tau_2)] \tag{2.32}$$

and again $C_{3,y}(\tau_1, \tau_2) = M_{3,y}(\tau_1, \tau_2)\ \forall\tau_1, \tau_2$, where $C_{3,y}(\tau_1, \tau_2)$ is the third-order cumulant. The fourth-order moment is defined as

$$M_{4,y}(\tau_1, \tau_2, \tau_3) \triangleq E[y(n)y(n+\tau_1)y(n+\tau_2)y(n+\tau_3)] \tag{2.33}$$

and the fourth-order cumulant is

$$\begin{aligned} C_{4,y}(\tau_1, \tau_2, \tau_3) = {} & M_{4,y}(\tau_1, \tau_2, \tau_3) \\ & -C_{2,y}(\tau_1)C_{2,y}(\tau_2 - \tau_3) - C_{2,y}(\tau_2)C_{2,y}(\tau_3 - \tau_1) \\ & -C_{2,y}(\tau_3)C_{2,y}(\tau_1 - \tau_2). \end{aligned} \tag{2.34}$$

As we have seen above, although the second- and third-order cumulants (of zero-mean processes) are identical with the autocorrelation and the third-order mo-

ments, respectively, the fourth-order moments are different from the fourth-order cumulants. The third- and higher-order cumulants of Gaussian processes are zero (proof can be found in Appendix A). Since cumulants of order $p > 2$ of a Gaussian process are zero, the cumulants provide a quantitative measure of its deviation from Gaussianity.

The properties of moments and cumulants can be summarized as follows (the proof could be found in [23]):

P1 If $\lambda_i, i = 1, \ldots, p$ are constants, and $y_i, i = 1, \ldots, p$ are random variables, then

$$\mathrm{Mom}(\lambda_1 y_1, \ldots, \lambda_p y_p) = (\prod_{i=1}^{p} \lambda_i)\mathrm{Mom}(y_1, \ldots, y_p)$$

and

$$\mathrm{Cum}(\lambda_1 y_1, \ldots, \lambda_p y_p) = (\prod_{i=1}^{p} \lambda_i)\mathrm{Cum}(y_1, \ldots, y_p).$$

P2 Moments and cumulants are symmetric functions in their arguments, i.e.

$$\mathrm{Mom}(y_1, \ldots, y_p) = \mathrm{Mom}(y_{j+1}, \ldots, y_p, y_1, \ldots, y_j)$$

and

$$\text{Cum}(y_1, \ldots, y_p) = \text{Cum}(y_{j+1}, \ldots, y_p, y_1, \ldots, y_j).$$

P3 If the random variables $\{y_i\}_{i=1}^p$ are independent of the random variables $\{z_i\}_{i=1}^p$, then

$$\text{Cum}(y_1 + z_1, \ldots, y_p + z_p) = \text{Cum}(y_1, \ldots, y_p) + \text{Cum}(z_1, \ldots, z_p)$$

whereas in general

$$\begin{aligned}\text{Mom}(y_1 + z_1, \ldots, y_p + z_p) &\triangleq E[(y_1 + z_1)(y_2 + z_2) \cdots (y_p + z_p)] \\ &\neq \text{Mom}(y_1, \ldots, y_p) + \text{Mom}(z_1, \ldots, z_p).\end{aligned}$$

However, for random variables $\{z_1, y_1, y_2, \ldots, y_p\}$, we have

$$\text{Cum}(y_1 + z_1, y_2, \ldots, y_p) = \text{Cum}(y_1, y_2, \ldots, y_p) + \text{Cum}(z_1, y_2, \ldots, y_p)$$

and

$$\text{Mom}(y_1 + z_1, y_2, \ldots, y_p) = \text{Mom}(y_1, y_2, \ldots, y_p) + \text{Mom}(z_1, y_2, \ldots, y_p).$$

P4 If a subset of the random variables $\{y_i\}_{i=1}^{p}$ is independent from the rest, then

$$\text{Cum}(y_1, y_2, \ldots, y_p) = 0$$

whereas in general

$$\text{Mom}(y_1, y_2, \ldots, y_p) \neq 0.$$

2.2.2 Ergodicity and Moments

Ergodicity deals with the relationship between statistical averages and sample averages. A process $\{y(n)\}$ is ergodic in the most general form if, with probability one, all of its moments can be determined from a single realization [25]. In other words, the expected value $E[\cdot]$ (or ensemble averages) can be replaced by time averages, i.e.,

$$\begin{aligned} E[y(n)y(n+\tau_1)\cdots y(n+\tau_{p-1})] &= \langle y(n)y(n+\tau_1)\cdots y(n+\tau_{p-1})\rangle \\ &= \lim_{T\to\infty} \frac{1}{2T+1} \sum_{n=-T}^{+T} y(n)y(n+\tau_1)\cdots y(n+\tau_{p-1}) \end{aligned} \tag{2.35}$$

where $\langle\cdot\rangle$ is the time-average operator, which has the same properties as the ensemble average operation $E[\cdot]$ if the process is ergodic.

We see from Eqn. (2.35) that time-averages of higher-order moments are func-

tions of infinitely many random variables and, therefore, can be viewed as random variables themselves. Ergodicity implies that the time averages of all possible sample sequences are equal to the same constant which, in turn, equals the ensemble average. Throughout this thesis we assume that if the process is ergodic, then Eqn. (2.35) holds for all orders up to p.

In practice, when we are given a finite length single realization of an ergodic process, i.e., $\{y(n)\}, n = -T, \ldots, +T$, we cannot compute the limits of Eqn. (2.35) but the estimates are

$$\langle y(n)y(n+\tau_1)\cdots y(n+\tau_{p-1})\rangle \approx \frac{1}{2T+1}\sum_{n=-T}^{+T} y(n)y(n+\tau_1)\cdots y(n+\tau_{p-1}) \quad (2.36)$$

The estimation of higher-order moments of a stochastic process can be found in detail in [21–23, 25]. In Chapter 6, we explicitly use HOS and the mixture of HOS-SOS. The main reason HOS is used because no matter if the Gaussian noise is colored or not, higher-order cumulants of the noise is zero. We discuss more on the use of HOS in Chapter 6.

2.3 Array Processing

The objective of array processing is to extract as much information as possible from the signals impinged on an antenna array or to recover each of the source signals that have been transmitted. Information of interest includes e.g., DOAs,

source frequencies and the number of sources. Basically, there are two types of array processing models.

The first type is non-parametric array processing, which does not assume any parameter-dependent structure on the signal and the array manifold. An example is given in Fig. 1.3 (d). Normally, the non-parametric array manifold is classified into convolutive and non-convolutive [5]. In this thesis, we limit the subject to non-convolutive or linear array manifold. Mathematically, the received signal, $\mathbf{x}(n)$, is described as

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{v}(n) \tag{2.37}$$

where $\mathbf{A}$ is the $M \times K$ linear array manifold matrix, $\mathbf{s}(n)$ is the $K \times 1$ vector consists of K source signals and $\mathbf{v}(n)$ is the $M \times 1$ vector of the complex noise which has zero mean. If the array processing is to recover each source signal without prior knowledge of the channel, this is known as blind beamforming or BSS [26,27]. If the objective is to estimate the array manifold, then it is called blind identification. When the array manifold matrix is long or square, solving blind identification problem is also directly solving BSS, through the use of pseudo-inverse as follows,

$$\hat{\mathbf{s}}(n) = \hat{\mathbf{A}}^{\dagger}\mathbf{x}(n) \tag{2.38}$$

As we will see in Chapter 5, the array manifold could be wide matrix. In this case, estimating array manifold (solving blind identification) does not solve BSS,

because there are many solutions for $\hat{\mathbf{s}}(n)$ that satisfy Eqn. (2.37) (assuming no noise), and pseudo-inverse only solves for the minimum-normed $\hat{\mathbf{s}}(n)$, which is not necessary anywhere close to actual $\mathbf{s}(n)$. We will see in detail for this type of BSS method in Chapter 5.

The second type is the parametric array processing, which assume that either the array manifold or both the array manifold and signals have parameter-dependent structure. The structure is known to the receiver, but the parameters are unknown. Recently, parametric array processing algorithms have been developed for various kind of problems [4, 28–32]. The most classical array processing problems involve estimation of DOA and number of sources in the presence of spatially white Gaussian noise [4, 33–39]. Hence, in the following subsections, we will only review on the classical DOA estimation algorithms, especially the subspace-based algorithms, because they are known for their low computational complexity and yet superb performance.

2.3.1 Parametric Signal Model

In this section, we will briefly review the techniques of classical subspace DOA estimation techniques. These techniques are indirectly incorporated or modified to suit to applications to be presented later. The signal received by M-sensor array

is typically modeled as follows,

$$\mathbf{x}(n) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(n) + \mathbf{v}(n) \qquad n = 0, 1, \ldots, N-1 \tag{2.39}$$

where N is the number of snapshots, $\mathbf{A}(\boldsymbol{\theta})$ is the $M \times K$ array manifold matrix, $\mathbf{s}(n)$ is the $K \times 1$ vector of source signals and $\mathbf{v}(n)$ is the $M \times 1$ vector of circularly complex Gaussian noise having zero mean, and its elements are temporally and spatially white with autocorrelation $\mathbf{R}_{\mathbf{v}} = \sigma_n^2\mathbf{I}$. The vector of source signals, $\mathbf{s}(n)$, is unknown to receiver.

The array manifold matrix $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\boldsymbol{\theta}_1), \ldots, \mathbf{a}(\boldsymbol{\theta}_K)]$ consists of the steering vector of each source $\mathbf{a}(\boldsymbol{\theta}_i)$. Assuming the sources are far away from the sensors such that the resulting waves are plane waves, then the array steering vector of the i^{th} source has the following structure

$$\mathbf{a}(\boldsymbol{\theta}_i) \triangleq \left[1, e^{j2\pi\tau_2(\boldsymbol{\theta}_i)}, \ldots, e^{j2\pi\tau_{M-1}(\boldsymbol{\theta}_i)}\right]^T, \tag{2.40}$$

where the propagation delay between the first and the m^{th} sensor is given as

$$\tau_m(\boldsymbol{\theta}_i) = \frac{\boldsymbol{\upsilon}^T(\boldsymbol{\theta}_i)\mathbf{d}_m}{c}, \qquad m = 1, \ldots, M. \tag{2.41}$$

In Eqn. (2.41),

- $\boldsymbol{\theta}_i = [\phi_i,\ \psi_i]^T$, where ϕ_i is the elevation and ψ_i is the azimuth with respect

to the coordinate system.

- $\boldsymbol{\upsilon}(\boldsymbol{\theta}_i) = [\cos\phi_i \sin\psi_i,\ \sin\phi_i \sin\psi_i,\ \cos\psi_i]^T$ is the unit vector in the direction of $\boldsymbol{\theta}_i$.

- $\mathbf{d}_m = [\alpha_m,\ \beta_m,\ \gamma_m]^T$ is the coordinate of the physical antenna sensor relative to sensor 1, which is set as the origin of the coordinate. The values of the coordinate are normalized by the carrier wavelength.

- c is the speed of propagation of carrier in the medium, e.g. propagation of electromagnetic waves in the air is $3 \cdot 10^8 \frac{m}{s}$ and propagation of sound in the water is $1500 \frac{m}{s}$.

In Fig. 2.2, we illustrate the plane wave from source i impinging on the antenna array.

For uniform linear array (ULA) and if the signal is from far-field, the array steering vector of the i^{th} source simplifies to the following structure

$$\mathbf{a}(\theta_i) \triangleq \left[1, e^{j2\pi\tau_2(\theta_i)}, \ldots, e^{j2\pi\tau_{M-1}(\theta_i)}\right]^T, \tag{2.42}$$

where the propagation delay between the first and the m^{th} sensors is given as

$$\tau_m(\theta) = (m-1)\frac{d\sin\theta}{c}, \qquad m = 1, \ldots, M \tag{2.43}$$

where d is the physical antenna spacing normalized by the carrier wavelength (see

Fig. 2.3). Note that for the estimated DOA, $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, to be unambiguous, the $d < \lambda/2$ condition must be satisfied, which is analogous to Nyquist sampling theorem for time series.

For convenience, we construct a data matrix, $\mathbf{X}$, by stacking the data vectors*,

*Data matrix (vector) and observed/received signal matrix (vector) are used interchangeably

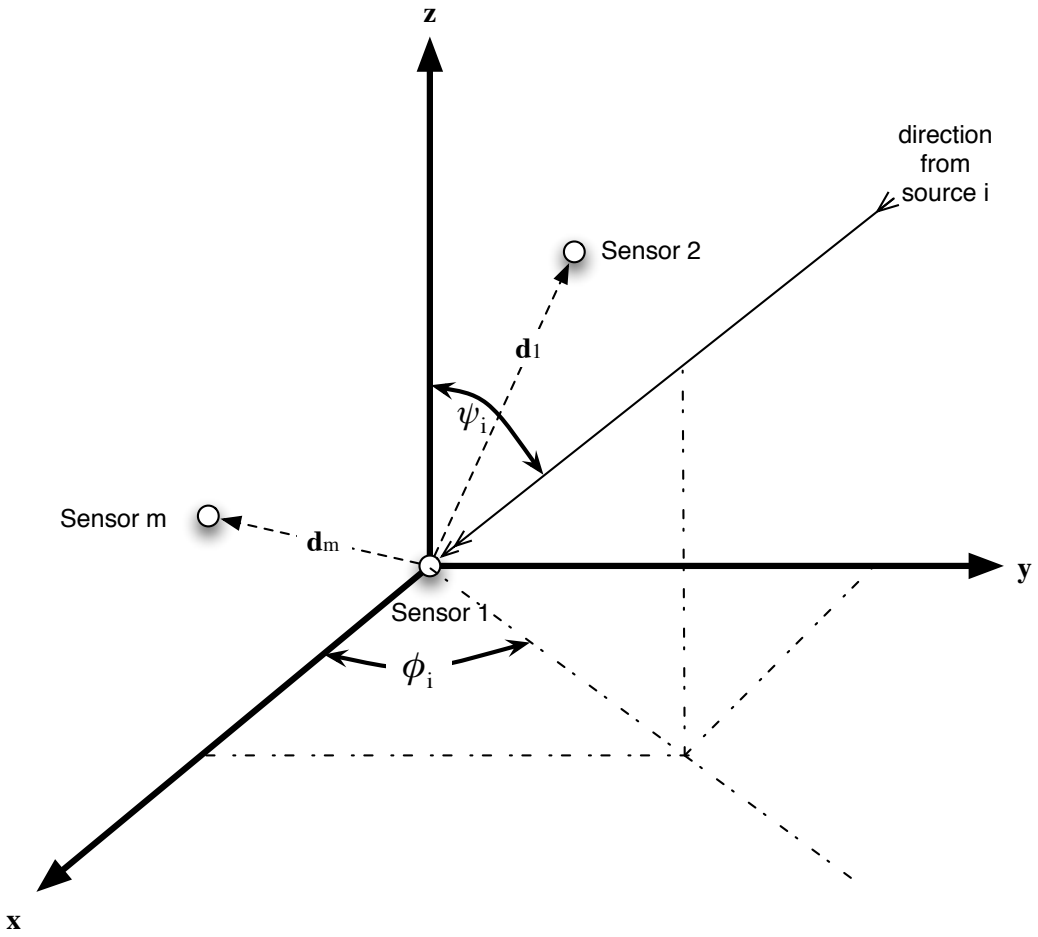

Figure 2.2: Plane wave impinging from (ϕ_i, ψ_i) direction to antenna array

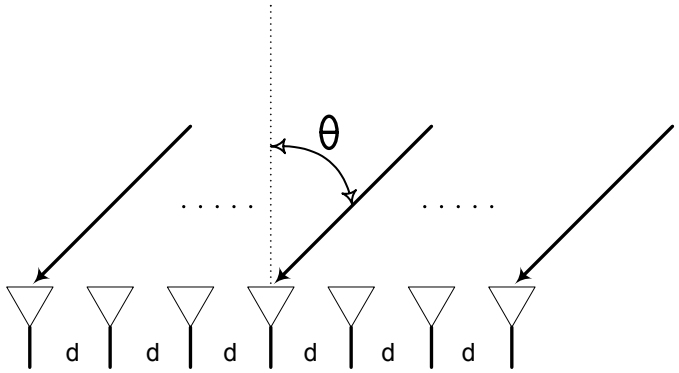

Figure 2.3: Plane wave impinging from θ direction to ULA with d element interspacing

$\mathbf{x}(n)$, as follows,

$$\begin{aligned} \mathbf{X} &= [\mathbf{x}(0), \ldots, \mathbf{x}(N-1)] \\ &= \mathbf{A}(\boldsymbol{\theta})\mathbf{S} + \mathbf{V} \end{aligned} \tag{2.44}$$

where $\mathbf{S} = [\mathbf{s}(0), \ldots, \mathbf{s}(N-1)]$ and $\mathbf{V} = [\mathbf{v}(0), \ldots, \mathbf{v}(N-1)]$ are the source signals and noise matrices resulted from stacking, respectively. From now on for simplicity, we will consider only ULA. The weighted subspace fitting and MUSIC algorithm could be used for any array settings. However, root-MUSIC algorithm is limited to ULA, while ESPRIT algorithm is required to have a pair identical subarrays.

2.3.2 Review of Weighted Subspace Fitting Algorithm

The first algorithm we will contemplate on is the Weighted Subspace Fitting algorithm [33, 34]. It is basically based on geometrical interpretation of the data, i.e. the signal subspace. The subspace spanned by column vectors of $\mathbf{A}(\boldsymbol{\theta})$ is termed

as the signal subspace. In the absence of noise, the subspace spanned by $\mathbf{X}$ is the same signal subspace spanned by column vectors of $\mathbf{A}(\boldsymbol{\theta})$. Suppose that the true covariance matrix of $\mathbf{x}(n)$ is available, i.e. $\mathbf{R_x} \triangleq E[\mathbf{x}(n)\mathbf{x}(n)^H]$, then its eigenvalue decomposition (EVD) is expressed as

$$\mathbf{R_x} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H = \mathbf{U_S}\boldsymbol{\Lambda_S}\mathbf{U_S}^H + \sigma_n^2\mathbf{U_N}\mathbf{U_N}^H \tag{2.45}$$

which reveals that a basis for signal subspace is formed by the columns of the $M \times K$ matrix $\mathbf{U_S}$. Since the noise is assumed to be white, uncorrelated with the source signals, and of equal variance, then $\mathbf{U_N}$ is the matrix of eigenvectors correspond to the $M - K$ smallest eigenvalues, which is all equal to σ_n^2. These eigenvectors (columns of $\mathbf{U_N}$) serve as the bases for noise subspace, which is orthogonal to the signal subspace. The orthogonality between signal and noise subspace is exploited not only in the weighted subspace algorithm, but also in other subspace algorithms [33–38]. Since the true covariance matrix is not accessible in practice, one can use the consistent estimate of the covariance matrix which is the sample covariance matrix $\hat{\mathbf{R}}_\mathbf{x} \triangleq \mathbf{X}\mathbf{X}^H/N$. Thus, its EVD is given by

$$\hat{\mathbf{R}}_\mathbf{x} = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^H = \hat{\mathbf{U}}_\mathbf{S}\hat{\boldsymbol{\Lambda}}_\mathbf{S}\hat{\mathbf{U}}_\mathbf{S}^H + \hat{\mathbf{U}}_\mathbf{N}\hat{\boldsymbol{\Lambda}}_\mathbf{N}\hat{\mathbf{U}}_\mathbf{N}^H \tag{2.46}$$

and the estimate for the signal subspace is $\hat{\mathbf{U}}_\mathbf{S}$. As the name subspace fitting implies, the method is to fit subspace spanned by the columns of $\mathbf{X}$ to that of

span$\{\mathbf{A}(\boldsymbol{\theta})\}$. Since span$\{\mathbf{U_S}\}$ = span$\{\mathbf{A}(\boldsymbol{\theta})\}$, then $\mathbf{U_S} = \mathbf{A}(\boldsymbol{\theta})\mathbf{T}$ where $\mathbf{T}$ is a square and non-singular matrix of size $K \times K$. Finally, we can formulate the subspace fitting estimation as follows

$$[\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}] = \arg\min_{\boldsymbol{\theta},\mathbf{T}} \|\hat{\mathbf{U}}_{\mathbf{S}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{T}\|_F \tag{2.47}$$

However since each column of $\hat{\mathbf{U}}_{\mathbf{S}}$ is perturbed differently, a modified version of the above formulation with weighting will perform better [33, 34]. The modified version is given by

$$[\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}] = \arg\min_{\boldsymbol{\theta},\mathbf{T}} \|\hat{\mathbf{U}}_{\mathbf{S}}\mathbf{W}^{\frac{1}{2}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{T}\|_F \tag{2.48}$$

where $\mathbf{W}$ is the weighting matrix, which is a diagonal and positive definite matrix. Solving for $\mathbf{T}$, we obtain $\mathbf{T} = \mathbf{A}(\boldsymbol{\theta})^{\dagger}\hat{\mathbf{U}}_{\mathbf{S}}\mathbf{W}^{\frac{1}{2}}$. After the substitution of $\mathbf{T}$ into Eqn. (2.48), we obtain

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= \arg\min_{\boldsymbol{\theta}} \|\hat{\mathbf{U}}_{\mathbf{S}}\mathbf{W}^{\frac{1}{2}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta})^{\dagger}\hat{\mathbf{U}}_{\mathbf{S}}\mathbf{W}^{\frac{1}{2}}\|_F \\
&= \arg\min_{\boldsymbol{\theta}} \|(\mathbf{I} - \mathbf{A}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta})^{\dagger})\hat{\mathbf{U}}_{\mathbf{S}}\mathbf{W}^{\frac{1}{2}}\|_F \\
&= \arg\min_{\boldsymbol{\theta}} \|\mathbf{P}^{\perp}_{\mathbf{A}(\boldsymbol{\theta})}\hat{\mathbf{U}}_{\mathbf{S}}\mathbf{W}^{\frac{1}{2}}\|_F \\
&= \arg\min_{\boldsymbol{\theta}} \mathrm{tr}\{\mathbf{P}^{\perp}_{\mathbf{A}(\boldsymbol{\theta})}\hat{\mathbf{U}}_{\mathbf{S}}\mathbf{W}\hat{\mathbf{U}}_{\mathbf{S}}{}^{H}\}
\end{aligned} \tag{2.49}$$

It has been shown in [33,34,39], that the optimal weighting which gives the lowest asymptotic error variance is $\mathbf{W}_o = (\hat{\mathbf{\Lambda}}_\mathbf{S} - \hat{\sigma}_n^2\mathbf{I})^2\hat{\mathbf{\Lambda}}_\mathbf{S}^{-1}$, where $\hat{\sigma}_n^2$ is any consistent estimate of the noise variance, e.g., average of the diagonal elements of $\hat{\mathbf{\Lambda}}_\mathbf{N}$. To implement this algorithm, one would require a multidimensional search algorithm such as Gauss-Newton, which is fast but requires a relatively accurate initial guess. Therefore, this algorithm suffers from computational complexity (because of multidimensional search) and from the possibility to be trapped at a local minimum point, which leads to wrong estimates. One way to reduce occurrence of being trapped in local minimum is by using Genetic Algorithms [40] or Simplex algorithm [41], however it will not completely remove the possibility of being trapped. In the next subsection, we introduce a DOA estimation method which uses only one-dimensional search. It is computationally inexpensive and one could even plot the cost function and locate the smallest minima which correspond to the estimate.

2.3.3 Review of MUSIC Algorithm

MUltiple SIgnal Classification algorithm (MUSIC) was introduced in [35,42]. MUSIC is similar to the weighted subspace fitting algorithm, except that it involves only one dimensional search. The basic idea is that the steering array in the directions of true DOAs is perpendicular to noise subspace, provided that access to

true noise subspace is available. Mathematically,

$$\mathbf{a}(\theta_i)^H \mathbf{U_N} = \mathbf{0} \qquad i = 1, \ldots, K \tag{2.50}$$

where $\mathbf{0}$ is the zero vector. MUSIC spectrum uses its inverse, as follow

$$P_{MUSIC}(\theta) = \frac{1}{\|\mathbf{a}(\theta)^H \mathbf{U_N}\|^2} \tag{2.51}$$

so that if $\theta = \theta_i$, the so called MUSIC spectrum, gives a very sharp peak for each impinging DOA. There is also a search-free and improved version of MUSIC, which is called root-MUSIC and was proposed in [36]. Root-MUSIC basically solves the denominator in Eqn(2.51) for the K complex roots which are closest to unit circle. With the elements of the steering vector set as $e^{j2\pi\tau_1(\theta)} = z$, the ULA steering vector is $\mathbf{a}(z) = [z^0, z^1, \cdots, z^{M-1}]^T$. Thus, denominator of Eqn. (2.51) is a polynomial of z and the phases of its K roots closest to the unit circle on a complex plane give the estimates $\{\tau(\theta_i)\}_{i=1}^K$, which eventually lead to the estimates of DOAs. This algorithm will be used in Chapter 6 to perform DOA estimation together with HOS and mixture of HOS-SOS.

2.3.4 Review of ESPRIT Algorithm

The Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) was introduced in [38]. It was originally meant for two identical sets of

sensor arrays with any geometry. Its variant, which we will discuss here, applies to sensor arrays for which one could choose two subsets of the sensor arrays such that they are identical. In harmonic retrieval of time series data, this variant from ESPRIT is called matrix pencil method [37].

Assuming ULA settings, then the array manifold matrix would be of the form

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & \cdots & 1 \\ \psi_1 & \cdots & \psi_K \\ \vdots & \ddots & \vdots \\ \psi_1^{M-1} & \cdots & \psi_K^{M-1} \end{bmatrix} \tag{2.52}$$

where $\psi_i = e^{j2\pi\tau_1(\theta_i)}$. In fact, multiple identical subarrays could be chosen from this ULA setting, such as choosing the first $M-1$ sensors and the last $M-1$ sensors. Assuming no noise for the time being, the response or the data matrix observed from the first and the second subarrays could be shown as follows

$$\begin{aligned} \mathbf{X}_1 &= \tilde{\mathbf{A}}\mathbf{S} \\ \mathbf{X}_2 &= \tilde{\mathbf{A}}\boldsymbol{\Psi}\mathbf{S} \end{aligned} \tag{2.53}$$

where $\boldsymbol{\Psi} = \text{diag}\{\psi_1, \cdots, \psi_K\}$ is a diagonal matrix, and $\tilde{\mathbf{A}}$ is the array manifold matrix $\mathbf{A}(\boldsymbol{\theta})$ with the last row deleted. This property is called shift invariance, because both the data matrix observed by the first subarray (first $(M-1)$ sensors)

and the shifted subarray (the last $(M-1)$ sensors) have the same invariant subspace span$\{\tilde{\mathbf{A}}\}$.

In ESPRIT, the received signal $\mathbf{X}$ is low-rank approximated at initial step upon successive estimation of the number of sources, $\hat{K}$. Suppose that the singular value decomposition of data matrix $\mathbf{X}$ is given by

$$\mathbf{X} = \mathbf{P}\boldsymbol{\Sigma}\mathbf{Q}^H \tag{2.54}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are $(M \times M)$-matrix consist of the left singular vectors and $(N \times N)$-matrix consist of the right singular vectors, respectively. The $(M \times N)$-matrix, $\boldsymbol{\Sigma}$, contains the singular values in its diagonal. Let the $(\hat{K} \times \hat{K})$-matrix, $\hat{\boldsymbol{\Sigma}}$, contains only $\hat{K}$-largest singular values in its diagonal, while $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$ are $(M \times \hat{K})$-matrix and $(N \times \hat{K})$-matrix consist of the left and right singular vectors corresponding to the $\hat{K}$-largest singular values, respectively. The low rank approximation of data in Eqn. (2.44) is then given by

$$\mathbf{X} \approx \hat{\mathbf{X}} \triangleq \hat{\mathbf{P}}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{Q}}^H. \tag{2.55}$$

Subsequently, the span of signal subspace is equal to the span of array manifold, i.e., span$\{\mathbf{A}(\boldsymbol{\theta})\}$ = span$\{\hat{\mathbf{P}}\}$. Thus, the array manifold matrix is related to signal subspace by

$$\hat{\mathbf{P}} = \mathbf{A}(\boldsymbol{\theta})\mathbf{T} \tag{2.56}$$

where $\mathbf{T}$ is an invertible square matrix of size $(\hat{K} \times \hat{K})$. By forming two subarrays as described for Eqn. (2.53), its effect to the signal subspace is given by

$$\begin{aligned} \hat{\mathbf{P}}_1 &= \tilde{\mathbf{A}}\mathbf{T} \\ \hat{\mathbf{P}}_2 &= \tilde{\mathbf{A}}\mathbf{\Psi}\mathbf{T} \end{aligned} \tag{2.57}$$

where $\hat{\mathbf{P}}_1$ and $\hat{\mathbf{P}}_2$ are matrices derived from $\hat{\mathbf{P}}$ with the first and the last row deleted, respectively. These matrices correspond to subspaces of the subarray 1 (formed by first $(M-1)$ sensors) and subarray 2 (formed by the last $(M-1)$ sensors), respectively. Thus, by combining this pair of equations, we have

$$\begin{aligned} \hat{\mathbf{P}}_1^{\dagger}\hat{\mathbf{P}}_2 &= \mathbf{T}^{-1}\tilde{\mathbf{A}}^{\dagger}\tilde{\mathbf{A}}\mathbf{\Psi}\mathbf{T} \\ &= \mathbf{T}^{-1}\mathbf{\Psi}\mathbf{T} \end{aligned} \tag{2.58}$$

We observe that Eqn. (2.58) shows a non-symmetric EVD. Thus, the diagonal matrix $\mathbf{\Psi}$ is the eigenvalue matrix of $\hat{\mathbf{P}}_1^{\dagger}\hat{\mathbf{P}}_2$, which means the estimate of ψ_i is the eigenvalue. Finally, the estimate of DOAs can be obtained from the phase of ψ_i. More complete background estimation theory and array processing can be found in [4, 43–45].

Chapter 3

Wideband LFM Array Parameters Estimation Method

3.1 Background

This chapter discusses on the parameter estimation of several wideband PPS sources in sensor arrays. PPS is a more accurate model for signals having continuous instantaneous phase/frequency over a finite extent time interval. According to Weierstrass' approximation theorem [46], the instantaneous phase/frequency can be well approximated by a polynomial within a finite observation interval. The estimation of PPS parameters is an important problem because they arise in diverse practical communication applications. For example, in SAR, SAS, inverse SAR, inverse SAS, Doppler radar, and sonar imaging, the returned signals are continuous frequency modulated signals [2, 47–50]. Aside from that, LFM signals are widely used in pulse-compression radar and sonar, particularly the ones using chirp pulse, i.e. FMCW radar [2, 47].

There has been much attention put into both the constant amplitude and slow time-varying PPS signals [18, 49–56]. However, most of these efforts are with one receiver sensor. There have been growing interests in estimating the PPS parameters and DOAs in sensor array systems [31,32,57–61]. Some authors solved this problem by using the narrowband assumption and by assuming the array manifold (mixing matrix or spatial signature) is unchanged over the observation period. Particularly, in [32, 57], STFD has been introduced and used for DOA estimation of narrowband FM sources. However, this requires a few time-frequency points which correspond to the sources' time-frequency signatures. Several exact and approximate ML algorithms have been proposed in [31]. Several authors have attempted to extend these narrowband assumption to the case of wideband FM sources [58–60,62]. In [58], LFM source parameters are estimated by the subspace method, however, the approach is limited by the need for all sources to have the same central frequency. In [59, 60] the wideband extension of STFD is severely restricted by the short sliding data window size. In [62], an iterative approach is proposed, but it could lead to strongly biased DOA estimates [59], and convergence is not guaranteed. In [61], DOAs of LFM sources are estimated more efficiently, however, all of the algorithms proposed require the knowledge on the frequency rate of all sources. In [63], the wideband technique, also known as the chirp beamformer (CBF), which is free from the restrictions in the techniques mentioned above, was devised to extract the parameters of interest through making a 3-D search. The CBF is exactly equivalent to the ML estimation technique for a single PPS source,

however, the authors in [63] directly extended it to multiple sources scenario. Since this method is ML based, which requires to perform 3-D search, the computational complexity is very high.

It is also worthwhile to note that, the conventional wideband array processing techniques such as coherent-subspace MUSIC [64,65] and wideband ML approaches [66], which are meant for general wideband signals, are underperformed when dealing with PPS. This is because these algorithms do not take advantage of the PPS structure. As reported in [59], these methods also suffer severely from limited window length.

In this chapter, we describe a simple parameter estimation technique for wideband chirp signals. We exploit the distinctiveness of the initial frequencies, and the invariance of the frequency rates observed from different sensors. The proposed technique is effective and computationally inexpensive. Simulation results and comparative results are included to validate the proposed algorithm.

In the next chapter, we describe a subspace-based parameter estimation method for wideband PPS source. The new operator named as SHIM is introduced and used to transform a PPS received by array sensors into a classical narrowband sinusoid estimation problem. One could estimate the highest-order frequency parameter and DOA separately, but we use a subspace algorithm [28] to jointly estimate both parameters.

3.2 Parametric PPS Models

There are many models can be used to describe PPS impinged on sensor arrays, depending on the type of array structures. In this thesis, only the ULA structure is considered, and it can be extended to a rectangular array or any other ULA based structure. For the PPS model, there are constant amplitude PPS, slow time-varying amplitude PPS, and fast time-varying amplitude PPS. Again, in this thesis, we consider only the constant amplitude PPS. The algorithms for the constant amplitude PPS could be extended with some degradations to slow time-varying amplitude PPS [19]. For the array manifold model, we consider the non-stationary array manifold. The array manifold can be either (quasi-)stationary if the PPS impinged on the sensor array are narrowband signals, or non-stationary if the PPS impinged on the sensor array are wideband signals. Of course, the non-stationary (wideband) model is more generalized than the stationary (narrowband) model, and that is the reason we consider the non-stationary model.

The signal model for a constant amplitude PPS from L sources arriving at an ULA composed of M sensors can be delineated as:

$$\mathbf{x}(n) = \mathbf{A}(n)\mathbf{s}(n) + \mathbf{v}(n) \qquad n = 0, 1, \ldots N \tag{3.1}$$

where N is the number of snapshots, $\mathbf{A}(n)$ is the $M \times L$ array manifold matrix, $\mathbf{s}(n)$ is the $L \times 1$ vector of PPS, and $\mathbf{v}(n)$ is the $M \times 1$ vector of complex cir-

cularly Gaussian noise that is of zero mean, temporally and spatially white with autocorrelation $\mathbf{R_v} \triangleq \sigma_n^2\mathbf{I}$. The vector of PPS is modeled as

$$\mathbf{s}(n) \triangleq \left[s_1(n), s_2(n), \ldots, s_L(n)\right]^T \tag{3.2}$$

where the PPS of order K of i^{th} source is modeled as

$$s_i(n) = A_i e^{j\alpha_i} \exp\left(j \sum_{k=1}^{K} a_{i,k}(n\Delta)^k\right) \tag{3.3}$$

In Eqn. (3.3), A_i is the unknown constant amplitude of the i^{th} source, $a_{i,k} = 2\pi f_{i,k}$ is the k^{th}-order angular frequency coefficient of the i^{th} source, α_i is the unknown phase and Δ is the sampling interval. The k^{th}-order frequency coefficient of the i^{th} source is $f_{i,k}$ and has the unit of Hz/s^k.

The array manifold matrix is $\mathbf{A}(n) = [\mathbf{a}(\theta_1, n), \ldots, \mathbf{a}(\theta_L, n)]$. The array steering vector of the i^{th} source is modeled as

$$\mathbf{a}(\theta_i, n) \triangleq \left[1, e^{j\varphi_i'(n)\psi_i}, \ldots, e^{j\varphi_i'(n)\psi_i(M-1)}\right]^T, \tag{3.4}$$

where the instantaneous frequency is given by $\varphi_i'(n) \triangleq \sum_{k=0}^{K-1}(k+1)a_{i,k+1}(n\Delta)^k$, which is assumed to be constant during the time interval for the signal to propagate across the array aperture. The DOA, θ_i, which lies inside the range from $-\frac{\pi}{2}$ to

$+\frac{\pi}{2}$, is related to ψ_i by

$$\psi_i = \frac{d}{c}\sin(\theta_i), \tag{3.5}$$

where d is the spacing between adjacent sensors of the ULA and c is the propagation speed in the medium (see Fig. 2.3 for the ULA figure).

3.3 Review of Chirp Beamformer

Since $\mathbf{v}(n) \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I})$, then $\mathbf{x}(n) \sim \mathcal{N}(\mathbf{A}(n)\mathbf{s}(n), \sigma_n^2\mathbf{I})$. Given the distribution of $\mathbf{x}(n)$, the optimal estimates could be obtained from ML method. The estimates of ML method are obtained by maximizing the logarithm of the probability density function of the received signals with respect to all the unknowns, including the unknown parameters of interest, and the unknown nuisance parameters. The log likelihood function is given by,

$$\mathcal{L}(\boldsymbol{\psi}) = -\sum_{n=0}^{N-1} \|\mathbf{x}(n) - \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n)\mathbf{G}(\boldsymbol{\varpi}, n)\boldsymbol{\beta}\|^2 \tag{3.6}$$

where the new definitions are $\boldsymbol{\theta} \triangleq [\theta_1, \cdots, \theta_L]^T$, $\boldsymbol{\varpi} \triangleq [\boldsymbol{\varpi}_1^T, \cdots, \boldsymbol{\varpi}_L^T]^T$, $\mathbf{G}(\boldsymbol{\varpi}, n) \triangleq \text{diag}\{g(\boldsymbol{\varpi}_1, n), \cdots, g(\boldsymbol{\varpi}_L, n)\}$, $\boldsymbol{\beta} \triangleq [A_1 e^{j\alpha_1}, \cdots, A_L e^{j\alpha_L}]^T$ and $\boldsymbol{\psi} \triangleq [\boldsymbol{\theta}^T, \boldsymbol{\varpi}^T, \boldsymbol{\beta}^T]^T$. The array manifold matrix $\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n)$ is the same as[††] $\mathbf{A}(n)$ in Eqn. (3.4). We have also defined $\boldsymbol{\varpi}_i \triangleq [a_{i,1}, \cdots, a_{i,K}]^T$, and $g(\boldsymbol{\varpi}_i, n) \triangleq \exp\left(j\sum_{k=1}^{K} a_{i,k}(n\Delta)^k\right)$. All these new definitions are used only in this section in order to simplify the

[††]Here we explicitly denote its dependency on $\boldsymbol{\theta}$ and $\boldsymbol{\varpi}$ for derivation purpose

derivation and to show the dependency of the functions on parameters of interest.

Expanding Eqn. (3.6) gives,

$$\begin{aligned}\mathcal{L}(\boldsymbol{\psi}) &= -\boldsymbol{\beta}^H \left(\sum_{n=0}^{N-1} \mathbf{G}(\boldsymbol{\varpi}, n)^H \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n)^H \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n) \mathbf{G}(\boldsymbol{\varpi}, n) \right) \boldsymbol{\beta} \\ &\quad + 2\Re \left(\sum_{n=0}^{N-1} \mathbf{x}(n)^H \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n) \mathbf{G}(\boldsymbol{\varpi}, n) \boldsymbol{\beta} \right) - \sum_{n=0}^{N-1} \mathbf{x}(n)^H \mathbf{x}(n). \quad (3.7)\end{aligned}$$

Maximizing Eqn. (3.7) above with respect to $\boldsymbol{\beta}$ yields

$$\begin{aligned}\hat{\boldsymbol{\beta}} = &\left(\sum_{n=0}^{N-1} \mathbf{G}(\boldsymbol{\varpi}, n)^H \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n)^H \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n) \mathbf{G}(\boldsymbol{\varpi}, n) \right)^{-1} \\ &\cdot \left(\sum_{n=0}^{N-1} \mathbf{G}(\boldsymbol{\varpi}, n)^H \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n)^H \mathbf{x}(n) \right). \quad (3.8)\end{aligned}$$

This value of $\hat{\boldsymbol{\beta}}$ is the optimum estimate of $\boldsymbol{\beta}$ given that the other parameters are known. Substituting Eqn. (3.8) into Eqn. (3.6) gives the simplified likelihood function

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\varpi}) = &\left(\sum_{n=0}^{N-1} \mathbf{x}(n)^H \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n) \mathbf{G}(\boldsymbol{\varpi}, n) \right) \cdot \\ &\left(\sum_{n=0}^{N-1} \mathbf{G}(\boldsymbol{\varpi}, n)^H \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n)^H \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n) \mathbf{G}(\boldsymbol{\varpi}, n) \right)^{-1} \cdot \\ &\left(\sum_{n=0}^{N-1} \mathbf{G}(\boldsymbol{\varpi}, n)^H \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n)^H \mathbf{x}(n) \right) - \sum_{n=0}^{N-1} \mathbf{x}(n)^H \mathbf{x}(n) \quad (3.9)\end{aligned}$$

which is independent of $\boldsymbol{\beta}$. Furthermore, by removing the last term in the last

equation, it can be simplified to

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta},\boldsymbol{\varpi}) &= \left(\sum_{n=0}^{N-1}\mathbf{x}(n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathbf{G}(\boldsymbol{\varpi},n)\right). \\ &\quad \left(\sum_{n=0}^{N-1}\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathbf{G}(\boldsymbol{\varpi},n)\right)^{-1}. \\ &\quad \left(\sum_{n=0}^{N-1}\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\mathbf{x}(n)\right) \end{aligned} \tag{3.10}$$

Mathematically, ML estimation for multiple wideband polynomial phase signals is given by

$$[\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\varpi}}] = \arg\max_{\boldsymbol{\theta},\boldsymbol{\varpi}} \mathcal{L}(\boldsymbol{\theta},\boldsymbol{\varpi}) \tag{3.11}$$

However, due to significant number of unknown parameters involved when maximizing Eqn. (3.11), a very large dimensional search is required. Hence, authors of [63] proposed to use the log likelihood function in Eqn. (3.10) with the assumption that only single source is present. By assuming only source i is present, Eqn. (3.10) simplifies to

$$\mathcal{L}(\theta_i,\boldsymbol{\varpi}_i) = \frac{1}{NM}\left|\sum_{n=0}^{N-1}\mathbf{x}(n)^H\mathbf{a}(\theta_i,\boldsymbol{\varpi}_i)g(\boldsymbol{\varpi}_i,n)\right|^2 \tag{3.12}$$

Since the chirp beamformer (CBF) is meant only for chirp (LFM) signal, Eqn.

(3.12) is simplified further to[‡]

$$\mathcal{L}(\theta, a_1, a_2) = \frac{1}{NM} \left| \sum_{n=0}^{N-1} \mathbf{x}(n)^H \mathbf{a}(\theta, a_1, a_2) e^{ja_1\Delta n + ja_2(\Delta n)^2} \right|^2 \tag{3.13}$$

Equation (3.13) is the CBF cost function and has only three parameters involved in the search/maximization. Although in the derivation, single source is assumed, the CBF could be applied to multiple sources case (see [63]).

3.4 The Proposed Algorithms

3.4.1 Algorithm Utilizing (Weighted) Least Squares

We begin by defining, $\mathbf{y}(n) = \mathbf{A}(n)\mathbf{s}(n)$, which is the noise-free signal components of Eqn. (3.1). Since the signals in this chapter are second-order PPS, for convenience the notation of frequency rates and initial frequencies are redefined as $a_{i,2} \triangleq b_i$ and $a_{i,1} \triangleq a_i$, respectively (c.f. Eqn. (3.3) for the notation of frequency parameters in PPS). Hence, it can be easily seen that

$$\begin{aligned} \mathbf{y}(n) &= [\mathbf{a}(\theta_1, n), \ldots, \mathbf{a}(\theta_L, n)] \begin{bmatrix} A_1 e^{j(\alpha_1 + a_1\Delta n + b_1(\Delta n)^2)} \\ \vdots \\ A_L e^{j(\alpha_L + a_L\Delta n + b_L(\Delta n)^2)} \end{bmatrix} \\ &= \mathbf{y}_1(n) + \cdots + \mathbf{y}_L(n) \end{aligned} \tag{3.14}$$

[‡]subscript i has been dropped because CBF is an ML estimation technique for single source

where we have $\mathbf{y}_i(n) = A_i e^{j(\alpha_i + a_i\Delta n + b_i(\Delta n)^2)}\mathbf{a}(\theta_i, n)$. Rearranging the latter equation, we obtain the following expression

$$\mathbf{y}_i(n) = A_i e^{j\alpha_i} \begin{bmatrix} e^{j(a_i\Delta n + b_i(\Delta n)^2)} \\ e^{j\psi_i a_i} e^{j[(a_i + 2b_i\psi_i)\Delta n + b_i(\Delta n)^2]} \\ \vdots \\ e^{j(M-1)\psi_i a_i} e^{j[(a_i + 2b_i\psi_i(M-1))\Delta n + b_i(\Delta n)^2]} \end{bmatrix} \tag{3.15}$$

To vividly observe the main thrust of the proposed algorithm, we define the following,

Definition 1. *Observing from the m^{th} sensor, the frequency rate, the initial frequency and the phase of the i^{th} source are respectively given by:*

i. $\nu_{m,i} \triangleq b_i$, *as the frequency rate at sensor* m

ii. $\omega_{m,i} \triangleq (a_i + 2b_i\psi_i m)$, *as the initial frequency at sensor* m

iii. $\phi_{m,i} \triangleq \alpha_i + \psi_i a_i m$, *as the phase at sensor* m.

Rewriting Eqn. (3.15) by using the definitions given above, we obtain

$$\mathbf{y}_i(n) = A_i \begin{bmatrix} e^{j[\phi_{0,i} + \omega_{0,i}\Delta n + \nu_{0,i}(\Delta n)^2]} \\ \vdots \\ e^{j[\phi_{M-1,i} + \omega_{M-1,i}\Delta n + \nu_{M-1,i}(\Delta n)^2]} \end{bmatrix} \tag{3.16}$$

Note that from (i), (ii),and (iii) given in Definition 1, the frequency rates are con-

stant from one sensor to another, while the initial frequencies and phases vary linearly from one sensor to the adjacent one. Given the above definition, we commence by processing the output of each array element data individually to obtain the estimates of $\{\hat{\omega}_{m,i},\ \hat{\nu}_{m,i},\ \hat{\phi}_{m,i}\}$, for $i = 1, 2, \ldots, L$ and $m = 0, 1, \ldots, M-1$. This can be done by appealing to a modified version of multi-component Discrete Polynomial phase Transform (MC-DPT) method [67], which exploits the sensor array structure. We further process $\{\hat{\omega}_{m,i},\ \hat{\nu}_{m,i}\}$ to retrieve the values of $\{\hat{a}_i,\ \hat{b}_i,\ \hat{\psi}_i\}$ for $i = 1, 2, \ldots L$.

To obtain the estimate of $\hat{b}_i$, we propose the following averaging

$$\hat{b}_i = \frac{1}{M} \sum_{m=0}^{M-1} \widehat{\nu_{m,i}} \tag{3.17}$$

which will result in an improved estimate of $\hat{b}_i$ because of multiple sensors. To extract a_i from $\hat{\omega}_{m,i}$, we utilize the relationship observed from (ii). Stacking (ii) over $m = 0, \ldots, M-1$,

$$\boldsymbol{\omega}_i = \boldsymbol{\mathcal{A}}_i \boldsymbol{\gamma}_i \tag{3.18}$$

where $\boldsymbol{\mathcal{A}}_i \triangleq [\mathbf{1}, 2b_i\boldsymbol{\zeta}]$, $\boldsymbol{\zeta} \triangleq [0, 1, \ldots, M-1]^T$, $\mathbf{1} = [1, 1, \ldots, 1]^T$ with size of $M \times 1$ and $\boldsymbol{\gamma}_i \triangleq [a_i, \psi_i]^T$. By substituting the $\boldsymbol{\omega}_i$ and b_i with their estimates, $\hat{\boldsymbol{\omega}}_i$ and $\hat{b}_i$, we obtain estimate of $\boldsymbol{\gamma}_i$, as

$$\hat{\boldsymbol{\gamma}}_i = \hat{\boldsymbol{\mathcal{A}}}_i^{\dagger} \hat{\boldsymbol{\omega}}_i \tag{3.19}$$

where $\hat{\boldsymbol{\mathcal{A}}}_i^{\dagger}$ is the pseudo-inverse of $\hat{\boldsymbol{\mathcal{A}}}_i = [\mathbf{1}, 2\hat{b}_i\boldsymbol{\zeta}]$.

From Eqn. (3.19), one could directly obtain the estimate of ψ_i from the second component of $\hat{\boldsymbol{\gamma}}_i$. However, the estimate of ψ_i obtained this way is of poor accuracy as we will show this in the following. Assume $\hat{\boldsymbol{\omega}}_i = \boldsymbol{\omega}_i + \boldsymbol{\delta\omega}_i$, where $\boldsymbol{\delta\omega}_i$ is the estimation error vector or perturbation vector, and substitute into Eqn. (3.19), it yields

$$\hat{\boldsymbol{\gamma}}_i = \hat{\boldsymbol{\mathcal{A}}}_i^{\dagger}\boldsymbol{\mathcal{A}}_i\boldsymbol{\gamma}_i + \hat{\boldsymbol{\mathcal{A}}}_i^{\dagger}\boldsymbol{\delta\omega}_i \tag{3.20}$$

Since $\hat{\boldsymbol{\mathcal{A}}}_i$ is a $M \times 2$-matrix, then its pseudo-inverse is given by $\hat{\boldsymbol{\mathcal{A}}}_i^{\dagger} = (\hat{\boldsymbol{\mathcal{A}}}_i^T\hat{\boldsymbol{\mathcal{A}}}_i)^{-1}\hat{\boldsymbol{\mathcal{A}}}_i^T$. We next show how the inverse 2×2-matrix of $(\hat{\boldsymbol{\mathcal{A}}}_i^T\hat{\boldsymbol{\mathcal{A}}}_i)$ could be derived analytically.

$$\hat{\boldsymbol{\mathcal{A}}}_i^T\hat{\boldsymbol{\mathcal{A}}}_i = \begin{bmatrix} \mathbf{1}^T \\ 2\hat{b}_i\boldsymbol{\zeta}^T \end{bmatrix} \begin{bmatrix} \mathbf{1} & 2\hat{b}_i\boldsymbol{\zeta} \end{bmatrix} = \begin{bmatrix} \mathbf{1}^T\mathbf{1} & 2\hat{b}_i\mathbf{1}^T\boldsymbol{\zeta} \\ 2\hat{b}_i\boldsymbol{\zeta}^T\mathbf{1} & 4\hat{b}_i^2\boldsymbol{\zeta}^T\boldsymbol{\zeta} \end{bmatrix} \tag{3.21}$$

$$= \begin{bmatrix} M & \hat{b}_i M(M-1) \\ \hat{b}_i M(M-1) & \frac{2}{3}\hat{b}_i^2(M-1)M(2M-1) \end{bmatrix} \tag{3.22}$$

Here we make use of the fact that

$$\begin{aligned} \mathbf{1}^T\boldsymbol{\zeta} &= \sum_{l=0}^{M-1} l = (M-1)M/2 \\ \boldsymbol{\zeta}^T\boldsymbol{\zeta} &= \sum_{l=0}^{M-1} l^2 = \frac{1}{6}(M-1)M(2M-1) \end{aligned} \tag{3.23}$$

Hence, by evaluating the inverse of Eqn. (3.22), the pseudo-inverse of $\hat{\boldsymbol{\mathcal{A}}}_i$ is then

given as

$$\begin{aligned}\hat{\boldsymbol{\mathcal{A}}}_i^{\dagger} &= (\hat{\boldsymbol{\mathcal{A}}}_i^T\hat{\boldsymbol{\mathcal{A}}}_i)^{-1}\hat{\boldsymbol{\mathcal{A}}}_i^T \\ &= \frac{1}{\varrho}\begin{bmatrix} \frac{2}{3}\hat{b}_i^2(M-1)M(2M-1) & -\hat{b}_iM(M-1) \\ -\hat{b}_iM(M-1) & M \end{bmatrix}\begin{bmatrix} \mathbf{1}^T \\ 2\hat{b}_i\boldsymbol{\zeta}^T \end{bmatrix} \end{aligned} \tag{3.24}$$

where the determinant of matrix $\hat{\boldsymbol{\mathcal{A}}}_i^T\hat{\boldsymbol{\mathcal{A}}}_i$ is $\varrho \triangleq \frac{1}{3}\hat{b}_i^2M^2(M^2-1)$. Meanwhile, we will consider the first term on the right hand side of Eqn. (3.20) only. This term involves $\hat{\boldsymbol{\mathcal{A}}}_i^{\dagger}\boldsymbol{\mathcal{A}}_i$, which could be simplified to

$$\hat{\boldsymbol{\mathcal{A}}}_i^{\dagger}\boldsymbol{\mathcal{A}}_i = (\hat{\boldsymbol{\mathcal{A}}}_i^T\hat{\boldsymbol{\mathcal{A}}}_i)^{-1}(\hat{\boldsymbol{\mathcal{A}}}_i^T\boldsymbol{\mathcal{A}}_i) \tag{3.25}$$

$$= (\hat{\boldsymbol{\mathcal{A}}}_i^T\hat{\boldsymbol{\mathcal{A}}}_i)^{-1}\begin{bmatrix} M & b_iM(M-1) \\ \hat{b}_iM(M-1) & \frac{2}{3}\hat{b}_ib_i(M-1)M(2M-1) \end{bmatrix} \tag{3.26}$$

$$= \frac{1}{\varrho}\begin{bmatrix} \frac{1}{3}\hat{b}_i^2M^2(M^2-1) & 0 \\ 0 & \frac{1}{3}\hat{b}_ib_iM^2(M^2-1) \end{bmatrix} \tag{3.27}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & b_i/\hat{b}_i \end{bmatrix} \tag{3.28}$$

By substituting of Eqn. (3.28) into Eqn. (3.20), we have the following,

$$\begin{bmatrix} \hat{a}_i \\ \hat{\psi}_i \end{bmatrix} = \hat{\boldsymbol{\gamma}}_i = \begin{bmatrix} a_i \\ \psi_ib_i/\hat{b}_i \end{bmatrix} + \hat{\boldsymbol{\mathcal{A}}}_i^{\dagger}\boldsymbol{\delta\omega}_i \tag{3.29}$$

In Appendix C (see Eqn. (C.26)), we show that the second term of the right hand

side of Eqn. (3.29) is of zero mean. Therefore, the first element of $\hat{\boldsymbol{\gamma}}_i$ will give the unbiased estimate of a_i. On the contrary, the second element $\hat{\boldsymbol{\gamma}}_i$ gives an estimate of ψ_i, which depends on the value of b_i, and hence does not necessarily unbiased. By the first-order perturbation analysis of $\psi_i b_i/\hat{b}_i$, we can analyze the bias of the estimate, as follows

$$\begin{aligned} \hat{\psi}_i &= \frac{1}{b_i + \delta b_i} b_i \psi_i \\ &\approx \psi_i - \psi_i \frac{\delta b_i}{b_i} \end{aligned} \tag{3.30}$$

where δb_i are perturbations on b_i. From Eqn. (3.30), it can be seen that if the value of b_i is so small such that $\frac{\delta b_i}{b_i}$ is of order of 10^{-1} or more, then the error will be of the same order times ψ_i, which is rather large in this case because the conversion back to $\hat{\theta}_i = \sin^{-1}(\frac{c}{d}\hat{\psi}_i)$ is very sensitive to error. To overcome this problem when estimating ψ_i, we use $\widehat{\nu}_{m,i}$, $\widehat{\omega}_{m,i}$ and $\widehat{\phi}_{m,i}$ to reconstruct the non-parametric estimate of single source data, $\hat{\mathbf{x}}_i(n)$, as follows

$$\hat{\mathbf{x}}_i(n) = \begin{bmatrix} e^{j[\hat{\phi}_{0,i} + \hat{\omega}_{0,i}\Delta n + \hat{\nu}_{0,i}(\Delta n)^2]} \\ \vdots \\ e^{j[\hat{\phi}_{M-1,i} + \hat{\omega}_{M-1,i}\Delta n + \hat{\nu}_{M-1,i}(\Delta n)^2]} \end{bmatrix}. \tag{3.31}$$

Following that, to estimate the DOAs, we can now appeal to CBF technique [63], which is the compressed likelihood function of ML estimation method for one source chirp signal. The idea is to search for the DOA of the i^{th} source through

CBF while assuming the other unknown parameters are equal to the estimated parameters. Mathematically, this is done by performing one-dimensional search on the CBF cost function (Eqn. (3.13)) with $\hat{\mathbf{x}}_i(n)$, $\hat{a}_i$, and $\hat{b}_i$ have been substituted in, i.e.,

$$\hat{\psi}_i = \arg\max_{\psi_i} \frac{1}{N^2} \left| \sum_{n=0}^{N-1} \hat{\mathbf{x}}_i^H(n) \hat{\mathbf{a}}(\theta_i, n) e^{j(\hat{a}_i \Delta n + \hat{b}_i (\Delta n)^2)} \right|^2 \tag{3.32}$$

where $\hat{\mathbf{a}}(\theta_i, n) = [1, e^{j(\hat{a}_i + 2\hat{b}_i \Delta n)\psi_i}, \ldots, e^{j(\hat{a}_i + 2\hat{b}_i \Delta n)\psi_i (M-1)}]^T$. The original CBF is a 3-D search algorithm for a_i, b_i and ψ_i applied to multiple chirp signals. On the contrary, here is a 1-D search algorithm for DOA applied to single chirp signal.

Application of MC-DPT

In this chapter, the MC-DPT is used with some modifications to obtain $\{\hat{\omega}_{m,i}$, $\hat{\nu}_{m,i}$, $\hat{\phi}_{m,i}\}$ of each sensor. The reasons that MC-DPT is used are the close-to-CRB performance, the capability to estimate parameters of multiple chirp signals, the low computation complexity and the capability to provide automatically the non-parametric estimates of single source data. The last listed reason above will be demonstrated in the proposed algorithm steps because the non-parametric estimate is by the product of the MC-DPT steps. There are methods other than MC-DPT, such as Radon transform-based technique [68] and the product of higher-order ambiguity function-based technique [49], which can be used to estimate multiple LFM signal parameters. However, both techniques aiming at estimating the frequency rates only, but not the frequencies. In order to estimate the frequencies, one might

have to resort to MC-DPT steps. That is the reason we prefer MC-DPT over these two existing techniques. Additionally, both techniques are more computationally complex compared to MC-DPT.

The modifications that could be made to improve the estimates are as follows:

1. In the first estimation stage of original MC-DPT [67], to remove the i^{th} chirp of the signal at each sensor's measurement, the data are multiplied with $exp\{-j\hat{b}_i(n\Delta)^2\}$, where $\hat{b}_i$ is obtained from Eqn. (3.17).

2. In the second estimation stage, the non-parametric estimate of i^{th} component, $\hat{\mathbf{x}}_i(n)$, is obtained through following steps:

   (a) The data are modulated with negative of estimated initial frequencies and rate of the k^{th} source, as

   $$\tilde{\tilde{\mathbf{x}}}_k(n) \triangleq \mathbf{x}(n) \exp\{-j\hat{b}_k(n\Delta)^2\} \circ e^{-j\hat{\boldsymbol{\omega}}_k n\Delta} \tag{3.33}$$

   where $\circ$ is the Schur-Hadamard (element-wise) matrix product. This step is used to shift the k^{th} component to dc.

   (b) The k^{th} source is filtered out by its averaged data as follows,

   $$\overline{\mathbf{X}} = \tilde{\tilde{\mathbf{X}}}_k - \frac{1}{N}\left(\sum_{n=0}^{N-1} \tilde{\tilde{\mathbf{x}}}_k(n)\right)\mathbf{1}^T \tag{3.34}$$

   where $\overline{\mathbf{X}} \triangleq [\overline{\mathbf{x}}(0), \ldots, \overline{\mathbf{x}}(N-1)]$ and $\tilde{\tilde{\mathbf{X}}}_k \triangleq [\tilde{\tilde{\mathbf{x}}}_k(0), \ldots, \tilde{\tilde{\mathbf{x}}}_k(N-1)]$.

(c) The non-parametric estimate of the signal with k^{th} source removed is retrieved by

$$\hat{\mathbf{x}}_i(n) \triangleq \overline{\mathbf{x}}(n) e^{j\hat{b}_k(n\Delta)^2} \circ e^{j\hat{\boldsymbol{\omega}}_k n\Delta} \tag{3.35}$$

(d) Repeating these steps (a)–(c) for all $\{k : k \neq i, 1 \leq k \leq L\}$ and replacing $\mathbf{x}(n)$ in Eqn. (3.33) with $\hat{\mathbf{x}}_i(n)$, we will get the data which contain only the i^{th} component.

3. Following that, root-MUSIC is used instead of fast FT (FFT) to extract finer estimate of frequencies and frequency rates.

Now, we summarize the algorithm as follows:

1. To obtain $\{\hat{\boldsymbol{\omega}}_1, \ldots, \hat{\boldsymbol{\omega}}_L\}$, $\{\hat{b}_1, \ldots, \hat{b}_L\}$, and $\{\hat{\mathbf{x}}_1(n), \ldots, \hat{\mathbf{x}}_L(n)\}$ for $n = 0$ to $N-1$, we perform the modified MC-DPT on the data, $\mathbf{x}(n)$.

2. Compute Eqn. (3.19) for each $i = 1$ to L, then $\hat{a}_i = [\hat{\boldsymbol{\gamma}}_i]_{1,1}$, which is the first element of $\hat{\boldsymbol{\gamma}}$.

3. To get $\hat{\psi}_i$, perform the 1-D search for each $i = 1, \ldots, L$ using Eqn. (3.32). Subsequently, the estimate of DOAs, $\hat{\theta}_i$ are found by $\sin^{-1}(\frac{c}{d}\hat{\psi}_i)$.

It should however be noted that like most of the other harmonic retrieval estimators, the proposed estimator given in Eqn. (3.17) and (3.19) is formulated particularly based on spatially white and equal sensor gain/attenuation assumptions. It is possible to generalize the least squares (LS) approach in Eqns. (3.17) and

(3.19) by using the weighted LS approach in order to cope with different sensors gain/attenuation and noise variances. In the weighted LS case, Eqn. (3.17) and (3.19) should be replaced by $\sum_m w_m \widehat{\nu_m} / \sum_{m'} w_{m'}$ and $(\hat{\boldsymbol{A}}^T \mathbf{W}^{-1} \hat{\boldsymbol{A}})^{-1} \hat{\boldsymbol{A}} \mathbf{W}^{-1} \boldsymbol{\omega}$ respectively, where $\mathbf{W} = diag[\kappa_0, .., \kappa_{M-1}]$, and $\{\kappa_m\}, \{w_m\}$ are the weights. Of course, perfect knowledge of these heteroscedastic model's relative variance weights, $\{w_m\}$ and $\{\kappa_m\}$, are needed or to be estimated. One way to do the estimation is to measure the noise variance when no signal is transmitted.

3.4.2 Algorithm Utilizing TLS - LS

Further improvement could be made in the estimation of a_i and possibly ψ_i. This could be achieved by noting that the structure of $[\hat{\boldsymbol{A}}_i, \hat{\boldsymbol{\omega}}_i]$ is error-free in the first column, and contains errors in the second and third column. This leads to a solution using mixed total least squares - least squares (TLS-LS) approach as presented in [69]. The initial step of solving this problem, similar to the total least squares (TLS) approach, is to find a new estimate of $[\hat{\boldsymbol{A}}, \hat{\boldsymbol{\omega}}]$ such[§] that it has a rank of two and the first error-free column of $\hat{\boldsymbol{A}}$ is unchanged. The reason to have a rank of two is that under noiseless condition, span($\boldsymbol{\omega}$) = span($\boldsymbol{A}$) due to Eqn. (3.18).

[§]The subscript i has been suppressed for notational convenience in this chapter. The algorithm in this chapter are to be applied to each source individually.

Hence, mathematically it can be stated as

$$\min_{\substack{\boldsymbol{\alpha},\boldsymbol{\omega} \\ rank\{[\mathbf{1},\boldsymbol{\alpha},\boldsymbol{\omega}]\}=2}} \|[\mathbf{1},\boldsymbol{\alpha},\boldsymbol{\omega}] - [\mathbf{1},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\omega}}]\| \tag{3.36}$$

under unitarily invariant norm. In Eqn. (3.36), $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\omega}}$ are our estimates obtained previously using MC-DPT. Furthermore, $\hat{\boldsymbol{\alpha}}$ is[‡] the second column of $\hat{\boldsymbol{\mathcal{A}}}$. The solution to Eqn. (3.36) is obtained by directly applying the result presented in [70], and is given as follows,

$$[\boldsymbol{\alpha}_o,\boldsymbol{\omega}_o] = \boldsymbol{q}_1\boldsymbol{q}_1^T[\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\omega}}] + \boldsymbol{Q}_2\mathcal{H}_1\{\boldsymbol{Q}_2^T[\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\omega}}]\} \tag{3.37}$$

where $[\boldsymbol{q}_1,\boldsymbol{Q}_2] = \boldsymbol{Q}$ denotes the Q-part (unitary matrix) of the full QR factorization of matrix $[\mathbf{1},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\omega}}]$ and $\mathcal{H}_r\{\mathbf{M}\}$ is an operator that gives the low rank-r matrix approximation to $\mathbf{M}$, which is obtained by performing SVD to $\mathbf{M}$.

The following step to solve this TLS-LS problem is to find $\boldsymbol{\gamma}$ which can satisfy

[‡]The variable $\boldsymbol{\alpha}$ is different from α_i, which is the complex phase of source i defined in Eqn. (3.3)

$[\mathbf{1}, \boldsymbol{\alpha}_o]\boldsymbol{\gamma} = \boldsymbol{\omega}_o$. Equivalently, the solution could be obtained by solving

$$\begin{aligned}
\mathbf{0} &= [\mathbf{1}, \boldsymbol{\alpha}_o, \boldsymbol{\omega}_o][\boldsymbol{\gamma}^T, -1]^T \\
\mathbf{0} &= \boldsymbol{Q}^T[\mathbf{1}, \boldsymbol{\alpha}_o, \boldsymbol{\omega}_o] \begin{bmatrix} \boldsymbol{\gamma} \\ -1 \end{bmatrix} \\
&= \underbrace{\begin{bmatrix} r_{11} & \boldsymbol{q}_1^T[\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\omega}}] \\ \mathbf{0} & \mathcal{H}_1\{\boldsymbol{Q}_2^T[\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\omega}}]\} \end{bmatrix}}_{\mathbf{R}} \begin{bmatrix} a \\ \psi \\ -1 \end{bmatrix}
\end{aligned} \tag{3.38}$$

where r_{11} is just the $(1,1)^{\text{th}}$ element of $\mathbf{R}$. This equation could be decomposed into

$$\mathcal{H}_1\{\boldsymbol{Q}_2^T[\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\omega}}]\}[\psi, -1]^T = \mathbf{0} \tag{3.39}$$

and

$$r_{11}a + \boldsymbol{q}_1^T[\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\omega}}][\psi, -1]^T = \mathbf{0} \tag{3.40}$$

The solution for ψ can be obtained by using the right singular vector corresponding to the second largest singular value of $\boldsymbol{Q}_2^T[\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\omega}}]$ and normalizing the last element of the singular vector to -1. The reason to use the right singular vector corresponding to the second largest singular value is that $\mathcal{H}_1\{.\}$ nulls the second largest singular value, thus the corresponding right singular vector belongs to the null space. Furthermore, by using the second largest singular value, one does not require an additional SVD, which will result in additional computation complexity.

Finally, the Eqn. (3.40) could be solved for a, given that the estimate ψ is known. Mathematically, it can be summarized as follows,

$$\psi_o = -\frac{v_{21}}{v_{22}} \tag{3.41}$$

$$a_o = -\frac{1}{r_{11}}\boldsymbol{q}_1^T \begin{bmatrix} \hat{\boldsymbol{\alpha}} & \hat{\boldsymbol{\omega}} \end{bmatrix} \begin{bmatrix} \psi_o \\ -1 \end{bmatrix} \tag{3.42}$$

where $\mathbf{v_2} \triangleq [v_{21}^T, v_{22}^T]^T$ is the right singular vector corresponding to the second largest singular value.

3.5 Results and Discussion

Frequency spreading is very common in sonar due to non-stationarity of the water [72]. Therefore, in the subsequent examples, we will use the sonar settings to demonstrate the advantage of the proposed algorithms. Simulation results were plotted in Fig. 3.1 – Fig. 3.3, with mean square error (MSE) estimates versus SNR. The MSE for each SNR was obtained through 500 Monte Carlo runs. The number of samples is $N = 256$, with sampling interval $\Delta = 0.001$. The number of ULA sensors is $M = 10$, with inter-element spacing of $d = 1.5m$. The speed of propagation is assumed to be $c = 1500\frac{m}{s}$, which is a sonar propagation application. The first signal has initial frequency, frequency rate and direction of arrival, which are 420Hz, $-50\frac{\text{Hz}}{\text{s}}$ and 10^o correspondingly. The second signal parameters are 300Hz, $20\frac{\text{Hz}}{\text{s}}$ and 20^o, with the signal power of 0.8 relative to the first signal.

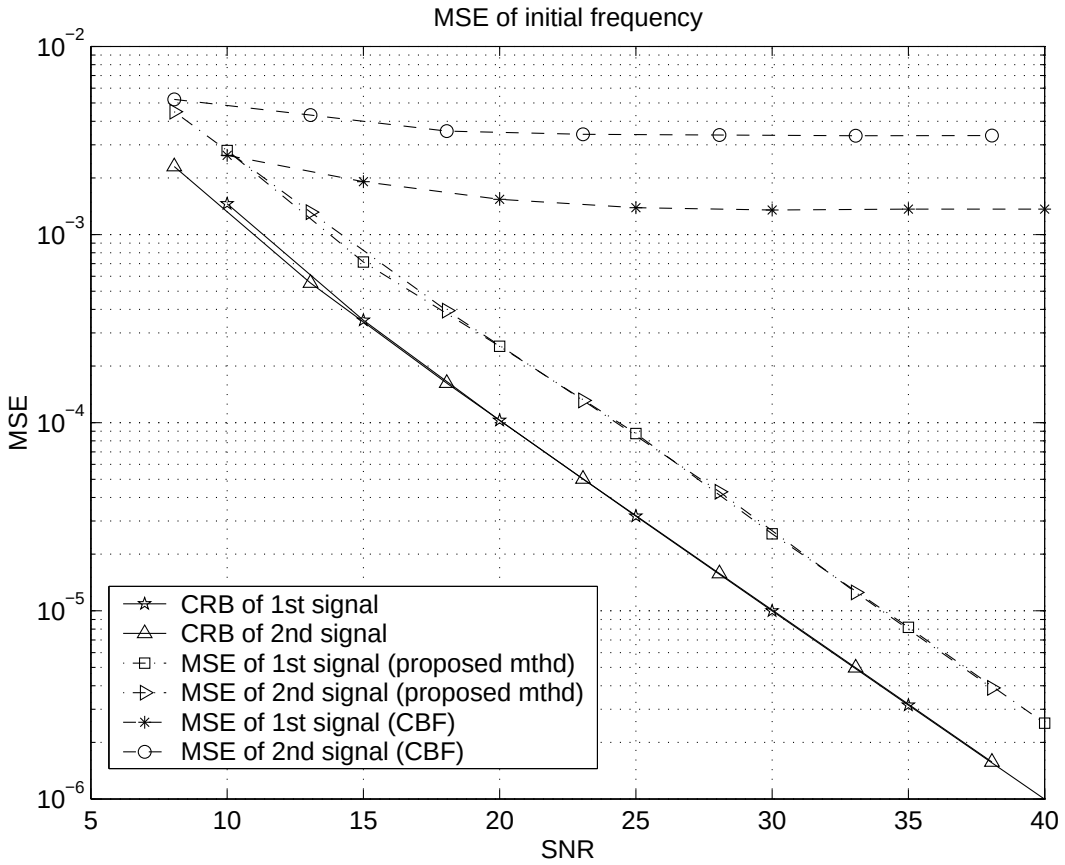

Figure 3.1: Comparison of MSE of f_1 $(Hz)^2$ vs. SNR(dB) among CBF, proposed LS-based algorithm and CRB

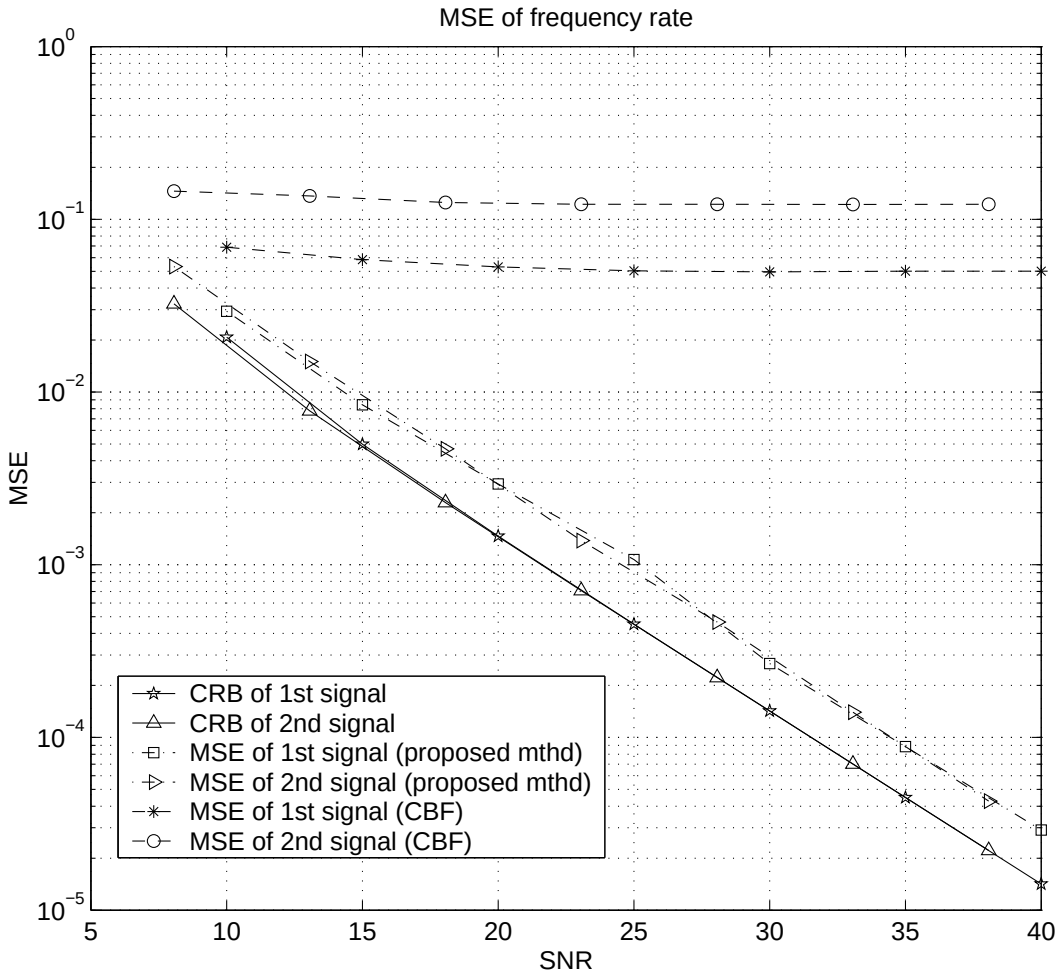

Figure 3.2: Comparison of MSE of f_2 $(Hz/s)^2$ vs. SNR(dB) among CBF, proposed LS-based algorithm and CRB

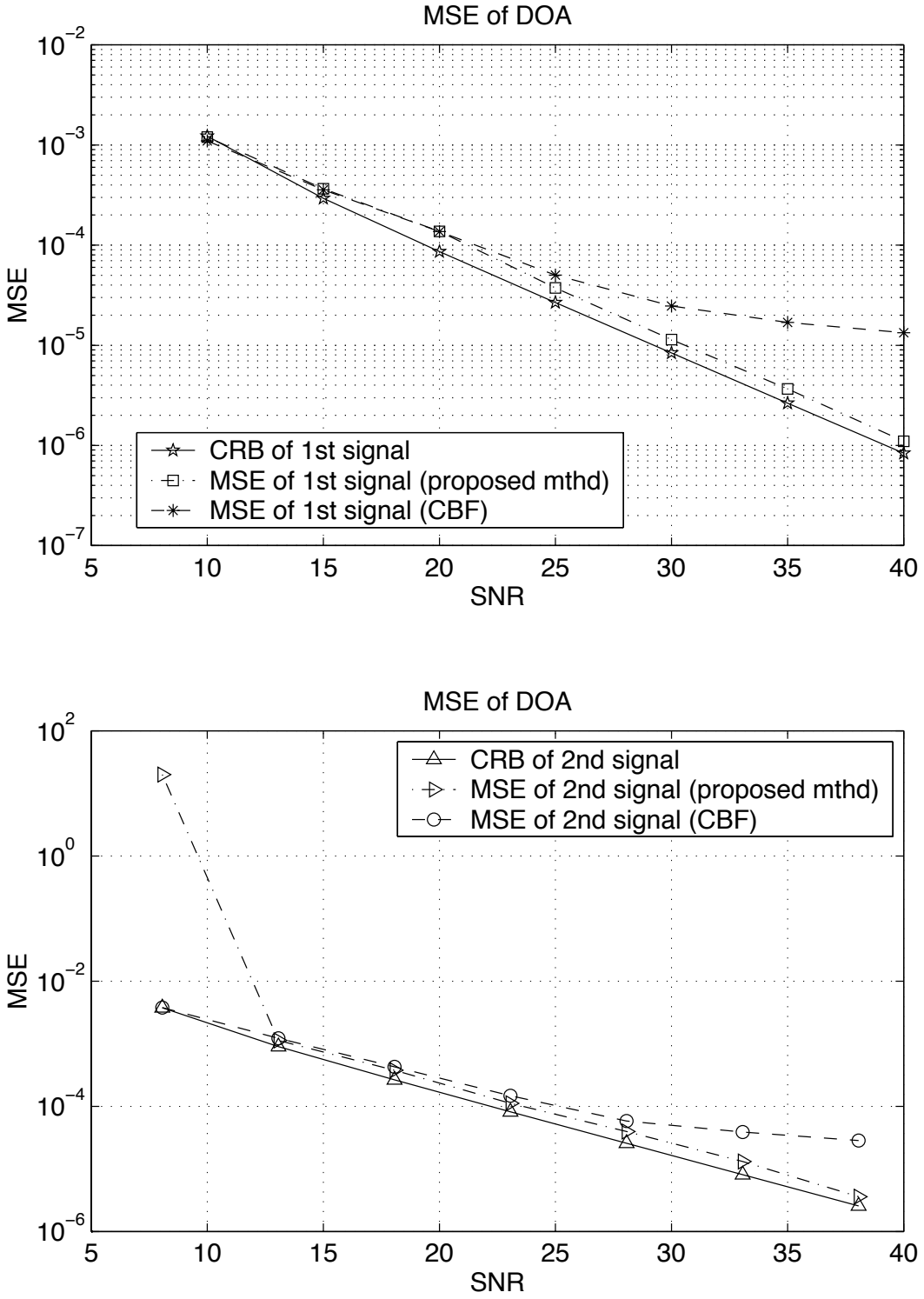

Figure 3.3: Comparison of MSE of θ $(^o)^2$ vs. SNR(dB) among CBF, proposed LS-based algorithm and CRB

For comparative purposes, CBF is implemented with two simplex search algorithms initiated at the true parameters to prevent it from being trapped at the local maxima. In addition to that CRB is also computed, and the derivation can be found in Appendix B.

All the three figures show that when the SNR is above 15 dB, the performance of the proposed algorithm remains close to CRB. On the other hand, the performance of CBF does not improve much as SNR increases, due to the bias of the estimates which does not vanish when SNR increases [63]. The MSE of an estimate is given by the variance of the estimate plus the square of the bias. By first-order perturbation analysis, the variance of the estimates diminishes when the noise variance diminishes or when SNR increases. Mathematically, it can be expressed as

$$\text{MSE(SNR)} = \text{Variance(SNR)} + \text{Bias}^2$$

Hence, the reason that CBF performance does not improve when SNR increases is because its bias is constant with respect to SNR.

The close-to-CRB performance of the proposed technique in frequency and frequency rate estimation is inherited from the close-to-CRB performance of MC-DPT algorithm. In fact, the detailed analysis given in Appendix C shows that the proposed algorithm is unbiased. The close-to-CRB performance in estimation of DOAs is because the proposed method has the capability to isolate the interferences due to other sources. This is because the proposed algorithm uses

non-parametric estimates of each source's data in CBF cost function for DOA estimation (Eqn. (3.32)). On the other hand, the original CBF [63] uses the unprocessed data from multiple signal sources in its cost function (Eqn. (3.13)). It can be verified through analysis that the estimates of DOA by the proposed method is unbiased under high SNR assumption. The details can be found in Appendix D.

From computation complexity perspective, the proposed method uses multiple 1-D searches if one incorporated FFT in the estimation of frequencies and frequency rates, and one 1-D search if one incorporated root-MUSIC instead of FFT. Hence, the proposed method is less complex than CBF which uses a 3-D search. Assuming that all of the 1-D searches are performed on N grid points, then the computation complexity of the proposed method is of order $\mathcal{O}(2K^2MN + 2KMN\log_2 N + N^2M)$ had one incorporated FFT, and of order $\mathcal{O}(2K^2MN + 6K^3MN + N^2M)$ had one incorporated root-MUSIC with window length of $2K$ (default window length in MATLAB). On the other hand, assuming that the 3-D search is performed on the $N \times N \times N$ three-dimensional grid points, then the computational complexity of CBF method is of order $\mathcal{O}(N^4M)$, which is significantly larger than the proposed method. Alternatively, one could use simplex or genetic algorithm to reduce number of searches in CBF method, however, without proper initial iteration points these algorithms will easily converge to one of the many spurious local maximum points and lead to incorrect estimation.

The only disadvantage of the proposed algorithm is in estimating multiple

sources that have the same frequency rates and/or frequencies. This limitation is inherited from MC-DPT algorithm which is incorporated in our proposed method.

The above mentioned results are for the LS method. As for the TLS-LS method, we did a simulation using the same parameters as in LS method. It could be observed from Fig. 3.4–Fig. 3.6 that the performance of TLS-LS-based method performs better when SNR is small. The reason is that TLS-LS, when computing the solution, takes into account the model error whilst LS does not.

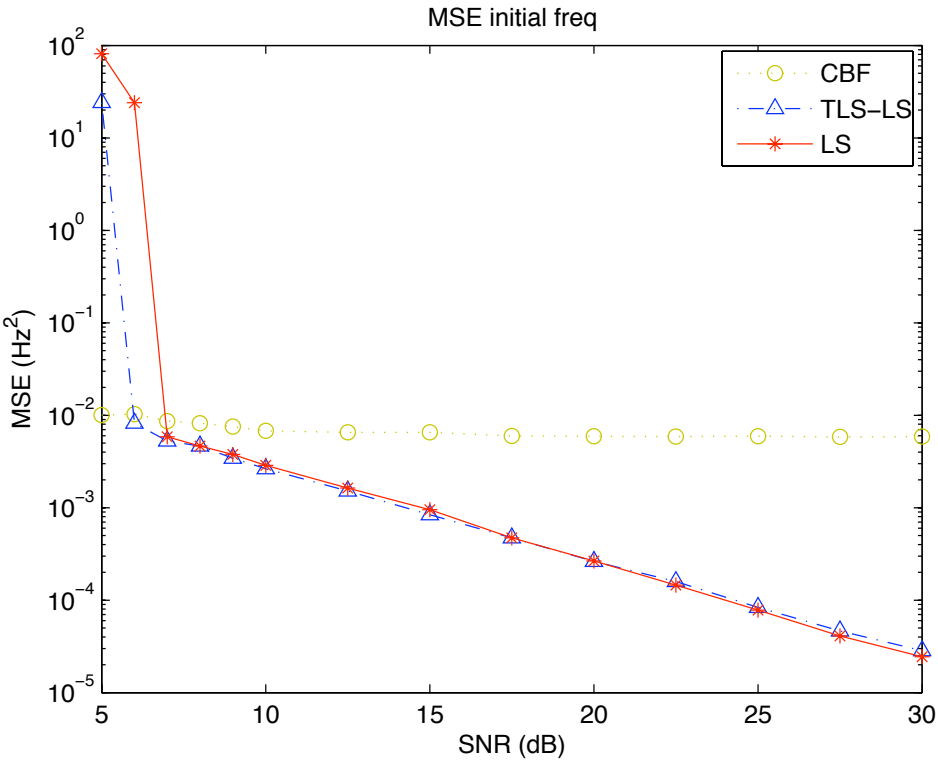

Figure 3.4: Comparison of MSE of f_1 $(Hz)^2$ among CBF, proposed LS-based and TLS-LS based algorithms

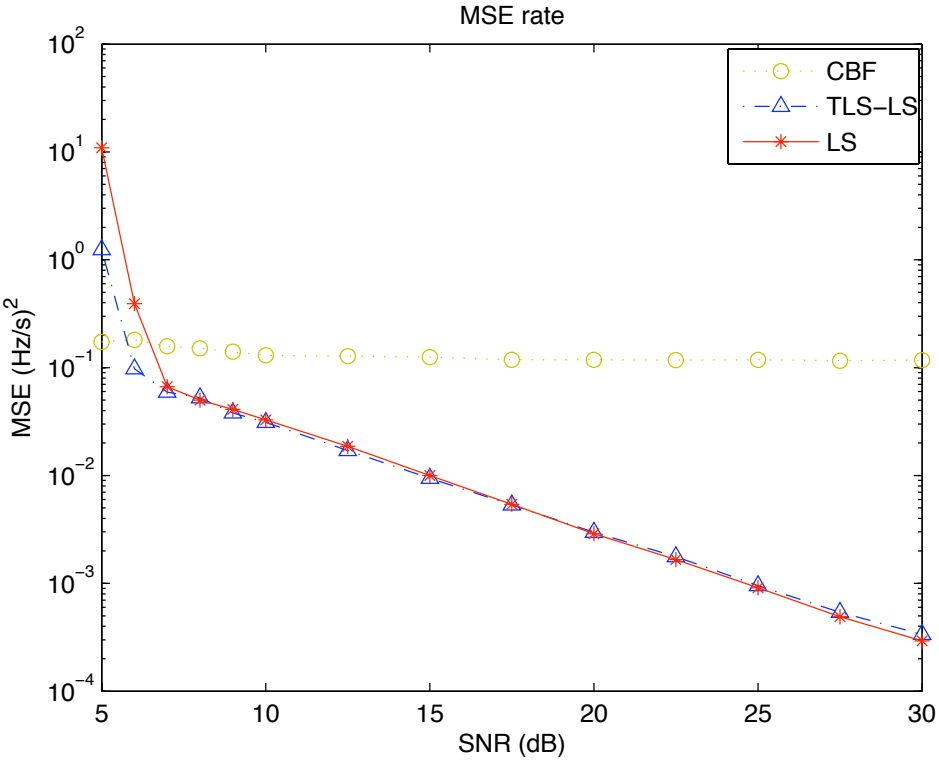

Figure 3.5: Comparison of MSE of f_2 $(Hz/s)^2$ among CBF, proposed LS-based and TLS-LS based algorithms

3.6 Summary

In this chapter, we have demonstrated a new wideband array processing technique based on LS estimation, and an improved technique based on TLS-LS estimation, for the estimation of multiple LFM signal parameters. The proposed method is compared to recently developed CBF algorithm. The proposed method is better in terms of accuracy, i.e. lower MSE and lower computational cost. However, the proposed algorithm in this chapter can only apply to second-order PPS. In the following chapter, we will devise a new algorithm that address PPS in general.

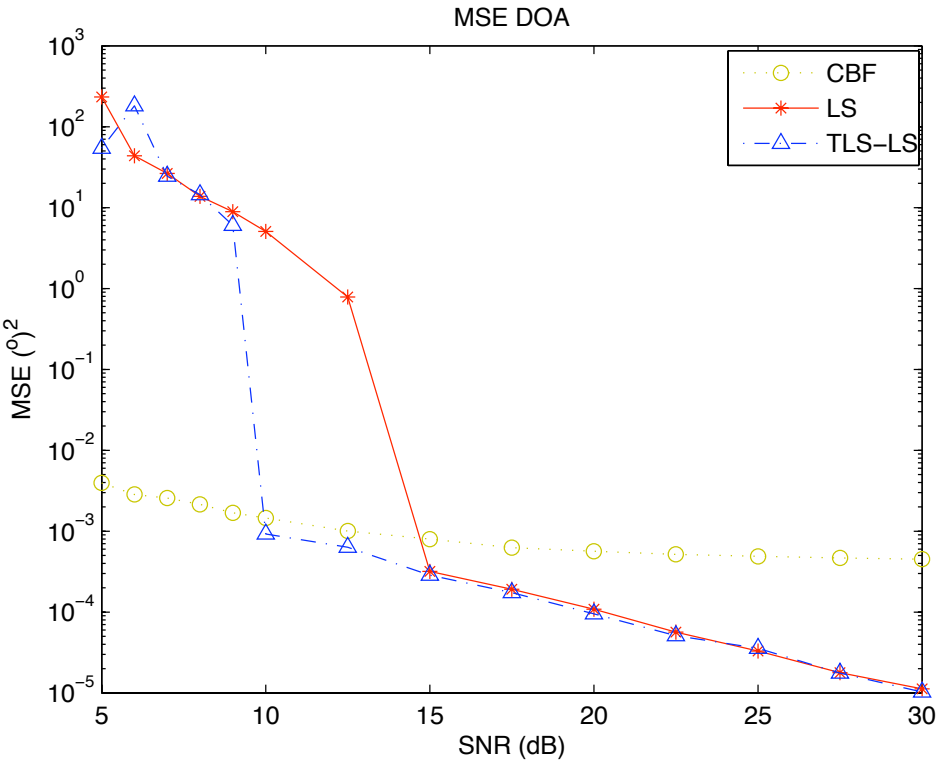

Figure 3.6: Comparison of MSE of DOA $(^o)^2$ among CBF, proposed LS-based and TLS-LS based algorithms

Chapter 4

Joint Estimation Method for Wideband PPS Impinged on Array Sensors

4.1 Introduction

After dealing with LFM signals in the previous chapter, we look into a more general class of PPS in this chapter. Herein, we estimate the parameters of a PPS impinged on array sensors. In this chapter, we describe a simple and search-free parameter estimation technique of wideband PPS, primarily to estimate its highest-order phase/frequency coefficient and DOA, which are the two main parameters of interest in many applications. The other frequency-related parameters of PPS are all dependent on the unknown propagation delay between the transmitter and the receiver and will not be estimated [18]. In the proposed algorithm in this chapter, we exploit the dual shift-invariance property in SHIM, which is a spatial variant of HIM [19]. After performing SHIM, the output are post-processed with ESPRIT-based joint angle-frequency estimation (JAFE) algorithm [28]. JAFE is originally

intended for narrowband sinusoidal signals impinging on an array of sensors. The proposed technique in this chapter is computationally inexpensive, because it is search-free. Simulation results are presented and comparisons are made to validate the strength of the proposed algorithm.

4.2 Single-Component PPS Model and SHIM

In Section 3.2 of the previous chapter, we have defined the general model of multiple wideband PPS impinging on a ULA in Eqn. (3.1). Assuming no noise for the time being, the wideband signal model for a single PPS impinging on the m^{th} sensor of a ULA is therefore given by

$$[\mathbf{z}(n)]_m \triangleq A \exp\left\{ j \left[a_K (n\Delta)^K + \sum_{k=0}^{K-1} (a_k + (k+1) a_{k+1} \psi(m-1)) (n\Delta)^k \right] \right\} \quad (4.1)$$

where $[\mathbf{z}(n)]_m$ denotes the m^{th} component of $\mathbf{z}(n)$.

Our proposed method is motivated by the works reported in Chapter 3 on second-order PPS, and particularly from Definition 1. Specifically, the first and second definitions imply that the highest-order frequency parameters are invariant from one sensor to another, and the electrical angle (ψ) appears in the second highest order frequency parameter varies linearly from one sensor to another in ULA setting, respectively. To exploit these properties, we introduce the spatial higher-order instantaneous moment (SHIM).

Definition 2. *Suppose* $\mathbf{w}(n)$ *is the time-sampled data vector, then the operator* $\mathbf{DP}_K[\cdot,\tau]$ *is defined as the SHIM operator of order* K *with delay* τ *applied to* $\mathbf{w}(n)$*, mathematically,*

$$\mathbf{DP}_K[\mathbf{w}(n),\tau] \triangleq \Big[DP_K[(\mathbf{w}(n))_1,\tau],\ldots,DP_K[(\mathbf{w}(n))_M,\tau]\Big]^T, \tag{4.2}$$

for $N_i \leq n \leq N-1$. $DP_K[\cdot]$ *is the (HIM) operator defined in [19] i.e.,*

$$\begin{aligned} DP_1[s(n),\tau] &\triangleq s(n) \\ DP_2[s(n),\tau] &\triangleq s(n)s^*(n-\tau) \\ &\vdots \\ DP_K[s(n),\tau] &\triangleq DP_2[DP_{K-1}[s(n),\tau],\tau]. \end{aligned} \tag{4.3}$$

*The initial sample is defined** *as* $N_i \triangleq (K-1)\tau$*; and* τ *is an arbitrary positive integer less than the number of samples,* N*, and it could be chosen according to [18].*

4.3 Proposed Algorithm

Here we will see how the newly introduced SHIM operator can be applied to a wideband PPS impinged on ULA. Because the SHIM operator basically operates at each sensor independently, all the temporal properties of $DP_K[\cdot]$ hold. Hence,

*It is called initial sample because for $n' < N_i$ the $DP_K[s(n'),\tau]$ could not be constructed (see Eqn. (4.3))

the following theorem from [18], can be applied to signal measured from every sensor.

Theorem 1. *Let $s(n)$ be a polynomial-phase signal of order K, such that, $s(n) = e^{j\alpha}\exp\left(j\sum_{k=1}^{K} a_k(n\Delta)^k\right)$ (c.f. Eqn. (3.3)), for $0 \le n \le N-1$, where a_m is the m^{th}-order frequency parameter, which takes a real value. Then for all positive integers τ*

$$DP_K[s(n),\tau] = \exp\{j(\phi_0\Delta n + \gamma_0)\} \tag{4.4}$$

for $N_i \le n \le N-1$, where ϕ_0 and γ_0 are the sinusoidal frequency and phase in DPT domain, and they are given as follows

$$\phi_0 = K!(\tau\Delta)^{K-1}a_K \tag{4.5}$$

$$\gamma_0 = (K-1)!(\tau\Delta)^{K-1}a_{K-1} - 0.5(K-1)K!(\tau\Delta)^K a_K \tag{4.6}$$

Therefore, by Definition 2 and by applying Theorem 1 to Eqn. (4.1), we obtain the following proposition.

Proposition 3. *If $\mathbf{z}(n)$ is the noise-free received wideband PPS of order K at the antenna array, then the result when applying the K^{th}-order SHIM to this signal is given by*

$$\mathbf{y}(n,\tau) = \mathbf{DP}_K[\mathbf{z}(n),\tau], \qquad n = N_i,\ldots,N-1 \tag{4.7}$$

$$= \mathbf{a}(\mu)Be^{j(\phi n+\gamma)} \tag{4.8}$$

where

$$\begin{aligned}
\mathbf{a}(\mu) &= [1, e^{j\mu}, \ldots, e^{j\mu(M-1)}]^T, && (4.9)\\
[\mathbf{y}(n,\tau)]_m &= B\exp\{j(\phi n + \mu m + \gamma)\}, && (4.10)\\
\phi &= K!\tau^{K-1}\Delta^K a_K, && (4.11)\\
\mu &= \phi\psi/\Delta, && (4.12)\\
\gamma &= (K-1)!(\tau\Delta)^{K-1}a_{K-1} - 0.5(K-1)K!(\tau\Delta)^K a_K, && (4.13)\\
B &= A^{2^{K-1}}. && (4.14)
\end{aligned}$$

After SHIM operation, we can obtain ϕ and γ, which are the frequency and phase of the signal, respectively, and the vector $\mathbf{a}(\mu)$, which is the narrowband steering array vector with electrical angle, μ. Therefore, the original data vectors of wideband PPS impinged on array are transformed into data vectors of narrowband sinusoidal impinged on array, which are given in Eqn. (4.8). Applying SHIM has dual advantages. Firstly, wideband steering array is transformed into narrowband steering array. Secondly, the PPS is transformed into sinusoidal signal, which can readily use the existing classical array processing algorithms or harmonic retrieval algorithms. Hereafter, the notation of τ in $\mathbf{y}(n,\tau)$ will be suppressed for simplification purpose. By exploiting Proposition 3 and by stacking together all

the $N - N_i$ samples, we get the following expression

$$\begin{aligned} \widetilde{\mathbf{Y}} &= [\mathbf{y}(N_i), \mathbf{y}(N_i + 1), \ldots, \mathbf{y}(N-1)] \\ &= Be^{j\gamma}\mathbf{a}(\mu)\left[e^{j\phi N_i}, e^{j\phi(N_i+1)}, \ldots, e^{j\phi(N-1)}\right] \end{aligned} \tag{4.15}$$

which possesses the shift-invariant property in two directions, i.e. along column direction and row direction (see Chapter 2 for the shift-invariant property). The $(M \times (N - N_i))$-matrix in Eqn. (4.15) is shift-invariant in column direction because the column subspace of the matrix formed by deleting the first row is the same as the column subspace of the matrix formed by deleting the last row. Similarly, the $(M \times (N - N_i))$-matrix in Eqn. (4.15) is shift-invariant in row direction because the row subspace of the matrix formed by deleting the first column is the same as the row subspace of the matrix formed by deleting the last column.

By constructing p-factor temporal smoothed data matrix as follow,

$$\begin{aligned}
\mathbf{Y}_p &= \begin{bmatrix} \mathbf{y}(N_i) & \mathbf{y}(N_i+1) & \cdots & \mathbf{y}(N-p) \\ \mathbf{y}(N_i+1) & \mathbf{y}(N_i+2) & \cdots & \mathbf{y}(N-p+1) \\ \vdots & \vdots & & \vdots \\ \mathbf{y}(N_i+p-1) & \mathbf{y}(N_i+p) & \cdots & \mathbf{y}(N-1) \end{bmatrix} \quad \in \mathbb{C}^{pM\times(N-p-N_i+1)} \\
&= Be^{j(\phi N_i+\gamma)} \begin{bmatrix} \mathbf{a}(\mu) & \mathbf{a}(\mu)e^{j\phi} & \cdots & \mathbf{a}(\mu)e^{j(N-N_i-p)\phi} \\ \mathbf{a}(\mu)e^{j\phi} & \mathbf{a}(\mu)e^{j2\phi} & \cdots & \mathbf{a}(\mu)e^{j(N-N_i-p+1)\phi} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}(\mu)e^{j(p-1)\phi} & \mathbf{a}(\mu)e^{jp\phi} & \cdots & \mathbf{a}(\mu)e^{j(N-N_i-1)\phi} \end{bmatrix}
\end{aligned} \tag{4.16}$$

we obtain a more general form of dual-shift invariant to appear in the column direction by deleting or selecting four distinct sets of rows, which will be shown next. The process of selecting certain rows of a matrix could be performed by left multiplication of the matrix with the selection matrix. The selection matrices, which will be used to to form the two pairs of shift-invariance matrices, are

$$\begin{aligned}
\mathbf{J}_{x,\phi} &= [\mathbf{I}_{p-1} \;\; \mathbf{0}_1] \otimes \mathbf{I}_M \\
\mathbf{J}_{y,\phi} &= [\mathbf{0}_1 \;\; \mathbf{I}_{p-1}] \otimes \mathbf{I}_M \\
\mathbf{J}_{x,\mu} &= \mathbf{I}_p \otimes [\mathbf{I}_{M-1} \;\; \mathbf{0}_1] \\
\mathbf{J}_{y,\mu} &= \mathbf{I}_p \otimes [\mathbf{0}_1 \;\; \mathbf{I}_{M-1}]
\end{aligned} \tag{4.17}$$

where $\mathbf{0}_1$ is a column vector containing zeros. A pair of selection matrices, $\mathbf{J}_{x,\phi}$ and $\mathbf{J}_{y,\phi}$, select the 1^{st} to $((p-1)M)^{\text{th}}$ rows, and the $(M+1)^{\text{th}}$ to $(pM)^{\text{th}}$ rows of the matrix they applied to, respectively. The other pair of selection matrices, $\mathbf{J}_{x,\mu}$ and $\mathbf{J}_{y,\mu}$, select the $(IM+1)^{\text{th}}$ to $(IM+M-1)^{\text{th}}$ rows, and the $(IM+2)^{\text{th}}$ to $(IM+M)^{\text{th}}$ rows, for $I=0,\ldots,p-1$, of the matrix they applied to, respectively. Therefore, applying the selection matrices to Eqn. (4.16) gives two pairs of shift-invariant matrices, given by

$$\begin{cases} \mathbf{J}_{x,\phi}\mathbf{Y}_p &= \mathbf{a}' B e^{j(\phi N_i+\gamma)}\mathbf{a}'''^T \\ \mathbf{J}_{y,\phi}\mathbf{Y}_p &= \mathbf{a}' B e^{j(\phi N_i+\gamma+\phi)}\mathbf{a}'''^T \end{cases},$$

$$\begin{cases} \mathbf{J}_{x,\mu}\mathbf{Y}_p &= \mathbf{a}'' B e^{j(\phi N_i+\gamma)}\mathbf{a}'''^T \\ \mathbf{J}_{y,\mu}\mathbf{Y}_p &= \mathbf{a}'' B e^{j(\phi N_i+\gamma+\mu)}\mathbf{a}'''^T \end{cases}$$

where

$$\mathbf{a}' \triangleq [\mathbf{a}(\mu)^T, \mathbf{a}(\mu)^T e^{j\phi}, \ldots, \mathbf{a}(\mu)^T e^{j(p-2)\phi}]^T, \tag{4.18}$$

$$\mathbf{a}'' \triangleq [\tilde{\mathbf{a}}(\mu)^T, \tilde{\mathbf{a}}(\mu)^T e^{j\phi}, \ldots, \tilde{\mathbf{a}}(\mu)^T e^{j(p-1)\phi}]^T, \tag{4.19}$$

$$\mathbf{a}''' \triangleq [1, e^{j\phi}, \ldots, e^{j(N-N_i-p)\phi}]^T, \tag{4.20}$$

$$\tilde{\mathbf{a}}(\mu) \triangleq [1, e^{j\mu}, \ldots, e^{j(M-2)\mu}]^T. \tag{4.21}$$

However, since in practice the matrix $\mathbf{Y}_p$ contains noise, it becomes full-rank

rather than rank-one (or low-rank in general), it is preferably to perform low-rank approximation using singular value decomposition (SVD) of $\mathbf{Y}_p$, i.e.,

$$\mathbf{Y}_p = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{4.22}$$

where its rank-one approximation is given by

$$\hat{\mathbf{Y}}_p = \mathbf{u}s_1\boldsymbol{\nu}^H \tag{4.23}$$

Here s_1 is the largest singular value, $\mathbf{u}$ and $\boldsymbol{\nu}$ are the corresponding left- and right-singular vectors. In the case of no noise, $\mathbf{u} = [\mathbf{a}(\mu)^T, \mathbf{a}(\mu)^T e^{j\phi}, \ldots, \mathbf{a}(\mu)^T e^{j(p-1)\phi}]^T t$, where t is a complex scaling constant. Therefore, the application of selection matrices onto the left-singular vector corresponding to the largest singular value leads to the following sets of equations,

$$\begin{cases} \mathbf{J}_{x\phi}\mathbf{u} &= \mathbf{a}'t \\ \mathbf{J}_{y\phi}\mathbf{u} &= \mathbf{a}'\lambda_\phi t = \mathbf{a}'e^{j\phi}t \end{cases} \qquad \begin{cases} \mathbf{J}_{x\mu}\mathbf{u} &= \mathbf{a}''t \\ \mathbf{J}_{y\mu}\mathbf{u} &= \mathbf{a}''\lambda_\mu t = \mathbf{a}''e^{j\mu}t \end{cases} \tag{4.24}$$

where $\lambda_\phi \triangleq e^{j\phi}$ and $\lambda_\mu \triangleq e^{j\mu}$. Let us define $\mathbf{u}_{x\phi} \triangleq \mathbf{J}_{x\phi}\mathbf{u}$, $\mathbf{u}_{y\phi} \triangleq \mathbf{J}_{y\phi}\mathbf{u}$, $\mathbf{u}_{x\mu} \triangleq \mathbf{J}_{x\mu}\mathbf{u}$ and $\mathbf{u}_{y\mu} \triangleq \mathbf{J}_{y\mu}\mathbf{u}$, by using these definitions and Eqn. (4.24), we have the following equations,

$$\mathbf{u}_{x\phi}^\dagger \mathbf{u}_{y\phi} = \lambda_\phi = e^{j\phi} \qquad \mathbf{u}_{x\mu}^\dagger \mathbf{u}_{x\mu} = \lambda_\mu = e^{j\mu} \tag{4.25}$$

Therefore, we can estimate ϕ and μ by $\hat{\phi} = \angle\lambda_\phi$ and $\hat{\mu} = \angle\lambda_\mu$ after we perform three steps: initially transform the PPS by SHIM, secondly perform a rank-one approximation through SVD of its p-temporally smoothened data matrix, and finally apply the proper selection matrices (Eqn. (4.17)) to its left-singular vector corresponding to the largest singular value. Using $\hat{\phi}$ and $\hat{\mu}$, the estimate of the K^{th}-order frequency parameter and DOA can be extracted by the following relations,

$$\hat{a}_K = \frac{\hat{\phi}}{K!\tau^{K-1}\Delta^K} \tag{4.26}$$

$$\hat{\theta} = \arcsin\left(\frac{c\Delta\hat{\mu}}{d\hat{\phi}}\right) \tag{4.27}$$

where we have made used of Eqn. (3.5), Eqn. (4.11), and Eqn. (4.11).

For better estimation of ϕ and μ, spatiotemporal smoothing and forward-backward averaging could be performed instead of just using temporal smoothing [28]. Spatiotemporal smoothed data matrix is constructed as

$$\mathbf{Y}_{p,L'} = [\mathbf{J}_1\mathbf{Y}_p, \mathbf{J}_2\mathbf{Y}_p, \ldots, \mathbf{J}_{L'}\mathbf{Y}_p] \qquad \in \mathbb{C}^{p(M-L'+1)\times L'(N-p+1)} \tag{4.28}$$

where the selection matrix, $\mathbf{J}_l \in \mathbb{R}^{p(M-L'+1)\times pM}$, selects a number of rows from the data matrix $\mathbf{Y}_p$ that corresponds to the l^{th} subarray. The first subarray in our ULA model, basically consists of data received from 1^{st} sensor to $(M-L'+1)^{\text{th}}$ sensor. The data for l^{th} subarray, would be data received from l^{th} sensor to $(M-L'+l)^{\text{th}}$ sensor.

Furthermore, whitening could be incorporated to improve the performance at the cost of additional computational complexity. However, as shown in [28], there is only slight improvement in estimating frequency parameter at low signal-to-noise ratio (SNR). Hence, in Section 4.5, we will assume whitening is incorporated to simplify the results reported in [28].

However, in practice, the estimation of λ_ϕ and λ_μ by using Eqn. (4.25) may not be accurate, because the SHIM-operated data is not noise free and, hence, $\mathbf{u}_{x\phi}$ and $\mathbf{u}_{y\phi}$ might not be sharing common subspace. Similarly, $\mathbf{u}_{x\mu}$ and $\mathbf{u}_{y\mu}$ might not be sharing common subspace either. Thus, it is recommended to use JAFE method [28], which will be reviewed in the next section.

4.4 Review of Joint Angle Frequency Method

In this section, we review the JAFE algorithm. In this algorithm [28], *super-generalized Schur decomposition* is used. It basically solves generalized Schur problem for more than two matrices that share a common subspace. In [73], the super-generalized Schur decomposition is introduced and is used in the analytical constant modulus algorithm to blindly separate different source signals. In [74], the super-generalized Schur decomposition was applied to jointly estimate DOAs and delays. In [28], the decomposition is applied to jointly estimate DOAs and frequency of signals. The material presented in this section outlines the procedure

to convert the problem in estimating the DOAs and the frequencies of the sources into a form which can be solved by the super generalized Schur decomposition algorithm.

When estimating DOAs and frequencies, the observed L source signals from the M sensors are given by

$$\mathbf{y}(n,\tau) = \mathbf{A}(\boldsymbol{\psi})\mathbf{s}(n) \tag{4.29}$$

where $\mathbf{A}(\boldsymbol{\psi}) = [\mathbf{a}(\psi_1), \ldots, \mathbf{a}(\psi_L)]$ is the array manifold matrix, $\mathbf{s}(n) = [B_1 e^{j(\phi_1 n+\gamma_1)}, \ldots, B_L e^{j(\phi_L n+\gamma_L)}]^T$ are the sinusoidal signal sources, and $\mathbf{a}(\psi_i) = [1, e^{j\psi_i}, \ldots, e^{j\psi_i(M-1)}]^T$ is the steering array. The objective here is to estimate ψ_i and ϕ_i. The setup and the objective are the same for our single source model given in Eqn. (4.8). We have seen that for the single source case in Eqn.(4.16), the p-factor temporally smoothed data matrix, as well as spatiotemporally smoothed data matrix, possesses two shift-invariant properties in the column direction. By left-multiplicating the selection matrices in Eqn. (4.17) to the p-factor temporal smoothed data, we get two pairs of shift-invariant matrices (Eqn. (4.18)). By left-multiplicating the selection matrices to left-singular vector of the signal, we obtain Eqn. (4.24). Similarly, for the case of multiple sources, after forming p-temporally smoothed data matrix, $\mathbf{Y}_p$, we find the rank-L approximation of

$\mathbf{Y}_p$. Suppose that its SVD is $\mathbf{Y}_p = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, then

$$\hat{\mathbf{Y}}_p = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^H \tag{4.30}$$

is the low-rank-L approximation of the data matrix. Here, $\hat{\mathbf{\Sigma}}$ is a $(L \times L)$-diagonal matrix with up to L largest singular values of $\mathbf{Y}_p$ in its diagonal; and the matrices $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ contain the left- and right-singular vectors corresponding to the L-largest singular values.

In the case of no noise, $\hat{\mathbf{U}} = \mathcal{A}(\boldsymbol{\phi}, \boldsymbol{\psi})\mathbf{T}$ for a unique non-singular matrix $\mathbf{T}$ and $\mathcal{A}(\boldsymbol{\phi}, \boldsymbol{\psi}) \triangleq [\breve{\mathbf{a}}(\phi_1, \psi_1), \ldots, \breve{\mathbf{a}}(\phi_L, \psi_L)]$, where

$$\breve{\mathbf{a}}(\phi_i, \psi_i) \triangleq [\mathbf{a}(\psi_i)^T, \mathbf{a}(\psi_i)^T e^{j\phi_i}, \ldots, \mathbf{a}(\psi_i)^T e^{j(p-1)\phi_i}]^T \tag{4.31}$$

$$\mathbf{a}(\psi_i) \triangleq [1, e^{j\psi_i}, \ldots, e^{j(M-1)\psi_i}]^T \tag{4.32}$$

Similar to the single source scenario given in Eqn. (4.24), we have the following two pairs of shift-invariant matrices,

$$\begin{cases} \mathbf{U}_{x\phi} \triangleq \mathbf{J}_{x\phi}\hat{\mathbf{U}} = \mathbf{A}_{x\phi}\mathbf{T} \\ \mathbf{U}_{y\phi} \triangleq \mathbf{J}_{y\phi}\hat{\mathbf{U}} = \mathbf{A}_{x\phi}\mathbf{\Phi}\mathbf{T} \end{cases} \quad \begin{cases} \mathbf{U}_{x\psi} \triangleq \mathbf{J}_{x\psi}\hat{\mathbf{U}} = \mathbf{A}_{x\psi}\mathbf{T} \\ \mathbf{U}_{y\psi} \triangleq \mathbf{J}_{y\psi}\hat{\mathbf{U}} = \mathbf{A}_{x\psi}\mathbf{\Psi}\mathbf{T} \end{cases} \tag{4.33}$$

where $\mathbf{\Phi} \triangleq \text{diag}\{\lambda_{\phi_1}, \ldots, \lambda_{\phi_L}\}$ and $\mathbf{\Psi} \triangleq \text{diag}\{\lambda_{\psi_1}, \ldots, \lambda_{\psi_L}\}$ are the complex diagonal matrices. In Eqn. (4.33), we have $\mathbf{A}_{x\phi} \triangleq [\mathbf{a}'(\phi_1, \psi_1), \ldots, \mathbf{a}'(\phi_L, \psi_L)]$ and $\mathbf{A}_{x\psi} \triangleq [\mathbf{a}''(\phi_1, \psi_1), \ldots, \mathbf{a}''(\phi_L, \psi_L)]$, where

$$\mathbf{a}'(\phi_i, \psi_i) \triangleq [\mathbf{a}(\psi_i)^T, \mathbf{a}(\psi_i)^T e^{j\phi_i}, \ldots, \mathbf{a}(\psi_i)^T e^{j(p-2)\phi_i}]^T \tag{4.34}$$

$$\mathbf{a}''(\phi_i, \psi_i) \triangleq [\tilde{\mathbf{a}}(\psi_i)^T, \tilde{\mathbf{a}}(\psi_i)^T e^{j\phi_i}, \ldots, \tilde{\mathbf{a}}(\psi_i)^T e^{j(p-1)\phi_i}]^T \tag{4.35}$$

$$\tilde{\mathbf{a}}(\psi_i) \triangleq [1, e^{j\psi_i}, \ldots, e^{j(M-2)\psi_i}]^T \tag{4.36}$$

In fact, Eqn. (4.33) gives two matrix pencil problems, each of them could be solved independently by using generalized Schur decomposition. However, in practice, the data is noisy and, hence, $\hat{\mathbf{U}} \approx \mathcal{A}(\boldsymbol{\phi}, \boldsymbol{\psi})\mathbf{T}$. As such, $\mathbf{U}_{x\phi}$ and $\mathbf{U}_{y\phi}$ might not be sharing a common subspace, and also $\mathbf{U}_{x\mu}$ and $\mathbf{U}_{y\mu}$ might not be sharing a common subspace. The estimated $\mathbf{T}$ also might not be the same in both matrix pencil problems, therefore, solving them independently with generalized Schur decomposition is not desireable. In order to solve this set of equations, the two pairs of matrices in Eqn. (4.33) should be reformulated so that the two matrix pencil problems share a common subspace and a common unknown $\mathbf{T}$. The super-generalized Schur decomposition algorithm can be used to find the eigenvalues of the modified problem. Note that the requirement to have a common $\mathbf{T}$ is to provide automatic pairing of the eigenvalues that belong to the same source.

Firstly, we will form a new set of four data matrices which ensure that there

are two common subspaces. Let us use span{**M**} to denote the space spanned by column vectors of **M**, then by considering the following facts from Eqn. (4.33),

$$\begin{aligned} \text{span}\{\mathbf{U}_{x\phi}\} &= \text{span}\{\mathbf{U}_{y\phi}\} = \text{span}\{\mathbf{A}_{x\phi}\} \\ \text{span}\{\mathbf{U}_{x\psi}\} &= \text{span}\{\mathbf{U}_{y\psi}\} = \text{span}\{\mathbf{A}_{x\psi}\} \\ \text{rank}\{\mathbf{U}_{x\phi}\} &= \text{rank}\{\mathbf{U}_{y\phi}\} = \text{rank}\{\mathbf{U}_{x\psi}\} = \text{rank}\{\mathbf{U}_{y\psi}\} = k \end{aligned} \tag{4.37}$$

which implied

$$\begin{aligned} \text{span}\{[\mathbf{U}_{x\phi}\,\mathbf{U}_{y\phi}]\} &= \text{span}\{\mathbf{A}_{x\phi}\} \\ \text{span}\{[\mathbf{U}_{x\psi}\,\mathbf{U}_{y\psi}]\} &= \text{span}\{\mathbf{A}_{x\psi}\} \\ \text{rank}\{[\mathbf{U}_{x\phi}\,\mathbf{U}_{y\phi}]\} &= \text{rank}\{[\mathbf{U}_{x\psi}\,\mathbf{U}_{y\psi}]\} = k \end{aligned} \tag{4.38}$$

Following that, let us consider the QR decompositions of the matrices as follows

$$[\mathbf{U}_{x\phi}\,\mathbf{U}_{y\phi}] = \mathbf{A}_{x\phi}\begin{bmatrix} \mathbf{T} & \mathbf{\Phi T} \end{bmatrix} = \mathbf{Q}_{\phi}\begin{bmatrix} \mathbf{E}_{x\phi} & \mathbf{E}_{y\phi} \\ \mathbf{0} & * \end{bmatrix} \tag{4.39}$$

$$[\mathbf{U}_{x\psi}\,\mathbf{U}_{y\psi}] = \mathbf{A}_{x\psi}\begin{bmatrix} \mathbf{T} & \mathbf{\Psi T} \end{bmatrix} = \mathbf{Q}_{\psi}\begin{bmatrix} \mathbf{E}_{x\psi} & \mathbf{E}_{y\psi} \\ \mathbf{0} & * \end{bmatrix} \tag{4.40}$$

where $*$ denotes some upper-triangular matrices. Let $\tilde{\mathbf{Q}}_{\phi} = [\mathbf{q}_{\phi 1} \ldots \mathbf{q}_{\phi L}]$, $\tilde{\mathbf{Q}}_{\psi} = [\mathbf{q}_{\psi 1} \ldots \mathbf{q}_{\psi L}]$ denote the first L-column vectors of $\mathbf{Q}_{\phi}$ and $\mathbf{Q}_{\psi}$, respectively, and $\tilde{\tilde{\mathbf{Q}}}_{\phi}$, $\tilde{\tilde{\mathbf{Q}}}_{\psi}$ denote the last L-column vectors of $\mathbf{Q}_{\phi}$ and $\mathbf{Q}_{\psi}$, respectively. From Eqn.

(4.38) and the QR decomposition properties, we obtain

$$\begin{aligned} \text{span}\{\tilde{\mathbf{Q}}_\phi\} &= \text{span}\{\mathbf{A}_{x\phi}\} \\ \text{span}\{\tilde{\mathbf{Q}}_\psi\} &= \text{span}\{\mathbf{A}_{x\psi}\} \end{aligned} \tag{4.41}$$

Therefore, by rearranging Eqns.(4.39) and (4.40), we get

$$\mathbf{Q}_\phi{}^H[\mathbf{U}_{x\phi}\,\mathbf{U}_{y\phi}] = \begin{pmatrix} \mathbf{E}_{x\phi} & \mathbf{E}_{y\phi} \\ \mathbf{0} & * \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{Q}}_\phi^H\mathbf{A}_{x\phi}\mathbf{T} & \tilde{\mathbf{Q}}_\phi^H\mathbf{A}_{x\phi}\mathbf{\Phi T} \\ \mathbf{0} & \tilde{\tilde{\mathbf{Q}}}_\phi{}^H\mathbf{A}_{x\phi}\mathbf{\Phi T} \end{pmatrix} \tag{4.42}$$

$$\mathbf{Q}_\psi{}^H[\mathbf{U}_{x\psi}\,\mathbf{U}_{y\psi}] = \begin{pmatrix} \mathbf{E}_{x\psi} & \mathbf{E}_{y\psi} \\ \mathbf{0} & * \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{Q}}_\psi^H\mathbf{A}_{x\psi}\mathbf{T} & \tilde{\mathbf{Q}}_\psi^H\mathbf{A}_{x\psi}\mathbf{\Psi T} \\ \mathbf{0} & \tilde{\tilde{\mathbf{Q}}}_\psi{}^H\mathbf{A}_{x\psi}\mathbf{\Psi T} \end{pmatrix} \tag{4.43}$$

where $\tilde{\tilde{\mathbf{Q}}}_\phi{}^H\mathbf{A}_{x\phi}\mathbf{\Phi T}$ and $\tilde{\tilde{\mathbf{Q}}}_\psi{}^H\mathbf{A}_{x\psi}\mathbf{\Psi T}$ are zero matrices if the data are noise free because $\tilde{\tilde{\mathbf{Q}}}_\phi$ and $\tilde{\tilde{\mathbf{Q}}}_\psi$ span the orthogonal spaces of $\text{span}\{\tilde{\mathbf{Q}}_\phi\}$ and $\text{span}\{\tilde{\mathbf{Q}}_\psi\}$. The QR decomposition step has implicitly formed two new matrices pairs, i.e., ($\mathbf{E}_{x\phi}$, $\mathbf{E}_{y\phi}$) and ($\mathbf{E}_{x\psi}$, $\mathbf{E}_{y\psi}$) which lie on $\text{span}\{\tilde{\mathbf{Q}}_\phi^H\mathbf{A}_{x\phi}\}$ and $\text{span}\{\tilde{\mathbf{Q}}_\psi^H\mathbf{A}_{x\psi}\}$, respectively, from the original matrices pairs, $\mathbf{U}_{x\phi}$, $\mathbf{U}_{y\phi}$, $\mathbf{U}_{x\psi}$ and $\mathbf{U}_{y\psi}$, which are not necessary lying on any common spaces due to the presence of noise. This idea is similar to TLS approximation, where a new set of matrices with common subspace is constructed from the original set of data matrices which do not have any common subspace (see [69]).

Secondly, we will use the newly derived matrices, i.e., ($\mathbf{E}_{x\phi}$, $\mathbf{E}_{y\phi}$) and ($\mathbf{E}_{x\psi}$,

$\mathbf{E}_{y\psi}$), to establish two new matrix pencil problems. Note that $\tilde{\mathbf{Q}}_{\phi}^{H}\mathbf{A}_{x\phi}$ and $\tilde{\mathbf{Q}}_{\psi}^{H}\mathbf{A}_{x\psi}$ are non-singular because $\tilde{\mathbf{Q}}_{\phi}$, $\mathbf{A}_{x\phi}$, $\tilde{\mathbf{Q}}_{\psi}$ and $\mathbf{A}_{x\psi}$ are all full column rank matrices. Hence, by retaining only two sets of equations in the following

$$\begin{cases} \mathbf{E}_{x\phi} = & \tilde{\mathbf{Q}}_{\phi}^{H}\mathbf{A}_{x\phi}\mathbf{T} \\ \mathbf{E}_{y\phi} = & \tilde{\mathbf{Q}}_{\phi}^{H}\mathbf{A}_{x\phi}\mathbf{\Phi}\mathbf{T} \end{cases} \qquad \begin{cases} \mathbf{E}_{x\psi} = & \tilde{\mathbf{Q}}_{\psi}^{H}\mathbf{A}_{x\psi}\mathbf{T} \\ \mathbf{E}_{y\phi} = & \tilde{\mathbf{Q}}_{\psi}^{H}\mathbf{A}_{x\psi}\mathbf{\Psi}\mathbf{T} \end{cases} \tag{4.44}$$

we established two sets of matrix pencil problems, and within each set has common subspace being enforced. Subsequently, we can form two eigenvalue decomposition equations by using Eqn. (4.44), i.e.,

$$\begin{cases} \mathbf{E}_{x\phi}^{-1}\mathbf{E}_{y\phi} & = \mathbf{T}^{-1}\mathbf{\Phi}\mathbf{T} \\ \mathbf{E}_{x\psi}^{-1}\mathbf{E}_{y\psi} & = \mathbf{T}^{-1}\mathbf{\Psi}\mathbf{T} \end{cases} \tag{4.45}$$

and both equations have a common matrix $\mathbf{T}$. This common matrix $\mathbf{T}$ forces these equations to have common eigenvectors and as a result, it gives automatic pairing of the eigenvalues belonging to the same source. However, solving these two eigenvalue decomposition problems in Eqn. (4.45) are numerically unstable because they involve matrix inversions in the left-hand side of Eqn. (4.45). Therefore, it is more desirable to solve the original version of the problem in the matrix pencil

form (Eqn. (4.44)) given by

$$\begin{cases} \mathbf{E}_{y\phi}\mathbf{q} &= \lambda_{\phi}\mathbf{E}_{x\phi}\mathbf{q} \\ \mathbf{E}_{y\psi}\mathbf{q} &= \lambda_{\psi}\mathbf{E}_{x\psi}\mathbf{q} \end{cases} \tag{4.46}$$

The super-generalized Schur decomposition algorithm can then be used to maintain automatic pairing of eigenvalues λ_{ϕ_i}, λ_{ψ_i}, which will be shown later.

Finally, we will rearrange Eqn. (4.44) into the QZ-structures. In other words, we transform the problem Eqn. (4.46) into a generalized Schur decomposition problem. Let us perform the full QR decompositions as follows

$$\mathbf{T}^{-1} = \mathbf{Z}^{H}\mathbf{R}_{T}^{-1} \tag{4.47}$$

$$\tilde{\mathbf{Q}}_{\phi}^{H}\mathbf{A}_{x\phi} = \mathbf{Q}'^{H}_{\phi}\mathbf{R}_{\phi} \tag{4.48}$$

$$\tilde{\mathbf{Q}}_{\psi}^{H}\mathbf{A}_{x\psi} = \mathbf{Q}'^{H}_{\psi}\mathbf{R}_{\psi} \tag{4.49}$$

where $\mathbf{Z}^{H}$, $\mathbf{Q}'^{H}_{\phi}$ and $\mathbf{Q}'^{H}_{\psi}$ are unitary matrices (the Q-part of QR decomposition), and $\mathbf{R}_{T}^{-1}$, $\mathbf{R}_{\phi}$ and $\mathbf{R}_{\psi}$ are square upper-triangular matrices (the R-part of QR decomposition). These upper-triangular matrices have non-zero diagonal elements because of the non-singularity of left-hand sides of the above equations. Thus, we

have the following set of matrices

$$\begin{cases} \mathbf{E}_{x\phi} &= \mathbf{Q}'^{H}_{\phi}\mathbf{R}_{\phi}\mathbf{R}_{T}\mathbf{Z} \triangleq \mathbf{Q}'^{H}_{\phi}\mathbf{R}_{x\phi}\mathbf{Z} \\ \mathbf{E}_{y\phi} &= \mathbf{Q}'^{H}_{\phi}\mathbf{R}_{\phi}\mathbf{\Phi}\mathbf{R}_{T}\mathbf{Z} \triangleq \mathbf{Q}'^{H}_{\phi}\mathbf{R}_{y\phi}\mathbf{Z} \end{cases}$$
$$\begin{cases} \mathbf{E}_{x\psi} &= \mathbf{Q}'^{H}_{\psi}\mathbf{R}_{\psi}\mathbf{R}_{T}\mathbf{Z} \triangleq \mathbf{Q}'^{H}_{\psi}\mathbf{R}_{x\psi}\mathbf{Z} \\ \mathbf{E}_{y\psi} &= \mathbf{Q}'^{H}_{\psi}\mathbf{R}_{\psi}\mathbf{\Psi}\mathbf{R}_{T}\mathbf{Z} \triangleq \mathbf{Q}'^{H}_{\psi}\mathbf{R}_{y\psi}\mathbf{Z} \end{cases} \tag{4.50}$$

where $\mathbf{R}_{x\phi}$, $\mathbf{R}_{y\phi}$, $\mathbf{R}_{x\psi}$ and $\mathbf{R}_{y\psi}$ are also square upper-triangular matrices with non-zero diagonal elements. The set of equations in (4.50) is basically of the form of two sets of generalized Schur decomposition [70] with a common factor $\mathbf{Z}$ which is derived from the common matrix $\mathbf{T}$. Hence the generalized eigenvalues are the ratio of the diagonals i.e. $(\lambda_{\phi i}, \lambda_{\psi i}) = ([\mathbf{R}_{y\phi}]_{ii}/[\mathbf{R}_{x\phi}]_{ii}, [\mathbf{R}_{y\psi}]_{ii}/[\mathbf{R}_{x\psi}]_{ii}$). Automatic pairing between the eigenvalues is ensured because the eigenvalues are linked by the common matrix $\mathbf{Z}$ defined in Eqn. (4.50). Given that there are more than one pair of matrices $\mathbf{E}$'s in Eqn. (4.50), the super generalized Schur decomposition algorithm can be used to solve for all the $\mathbf{Q}$'s, $\mathbf{R}$'s and $\mathbf{Z}$ (see [73] for the algorithm steps). Subsequently the estimates of $\mathbf{\Phi}$ and $\mathbf{\Psi}$, are given by

$$\begin{aligned} \hat{\mathbf{\Phi}} &= \text{diag}\{\mathbf{R}_{y\phi}\}\text{diag}\{\mathbf{R}_{x\phi}\}^{-1} \\ \hat{\mathbf{\Psi}} &= \text{diag}\{\mathbf{R}_{y\psi}\}\text{diag}\{\mathbf{R}_{x\psi}\}^{-1} \end{aligned} \tag{4.51}$$

with the i^{th} diagonal element of $\hat{\mathbf{\Phi}}$ being paired with i^{th} diagonal element of $\hat{\mathbf{\Psi}}$.

In summary, in this section, we have given the steps how DOAs and frequencies can be estimated simultaneously. It is done by forming the low-rank approximated data matrix in Eqn. (4.30) into Eqn. (4.50) which has four QZ-structured matrices with two common **Q** matrices and one common **Z** matrix. These two pairs of QZ structures could be solved simultaneously by super generalized Schur decomposition algorithm or also called extended QZ algorithm [73]. In fact, the super generalized Schur algorithm could solve for any number of QZ-structured matrices with one or many common factors of **Q**'s or **Z**. However, the details on the super generalized Schur algorithm will not be described in this thesis but it can be found in [73].

4.5 Analysis and Identifiability Condition

The performance of the JAFE algorithm under zero-mean Gaussian noise or error perturbation has been studied in [28]. However, in our proposed algorithm, the noisy data $\mathbf{x}(n)$ in Eqn. (3.1), which contains zero-mean Gaussian noise, is first operated with SHIM. The noisy SHIM-operated data, $\mathbf{DP}_K[\mathbf{x}(n), \tau]$, is then used in JAFE algorithm to perform estimation. Therefore, to study the performance in DOA and frequency estimation in the first step, we will analyze the perturbation error of the SHIM-operated data. The second step is to analyze the JAFE algorithm's performance when estimating the phases, μ and ϕ, using the SHIM-operated data. Here, we will simply recall the results in [28] and apply to our

algorithm. The third step is to relate the performance in estimating phases to performance in estimating of DOA and frequency parameters. Subsequently, we will show the identifiability conditions on Δ, τ, a_K and d/c, which have to be met for JAFE algorithm to estimate unambiguously. Finally, the optimal choice of Δ that minimizes the MSE is derived.

4.5.1 The Statistics of $\delta\mathbf{y}(n)$

Since SHIM is a non-linear operator, there will be many cross-terms generated between signal and noise resulted from its application to noisy signal $\mathbf{x}(n)$ of Eqn. (3.1). Hence, the easiest way to access the statistics of the perturbation error or the noise in the transformed domain, $\delta\mathbf{y}(n)$, is to deal with the difference between SHIM of $\mathbf{x}(n)$ (the noisy signal data) and SHIM of $\mathbf{z}(n)$ (the noiseless data signal of Eqn. (4.1)). Mathematically, $\delta\mathbf{y}(n)$ can be obtained by

$$\delta\mathbf{y}(n) = \mathbf{DP}_K[\mathbf{x}(n), \tau] - \mathbf{DP}_K[\mathbf{z}(n), \tau] \tag{4.52}$$

Let us recall from [18] the perturbation error of the HIM-operated data in the single antenna scenario, which we label it m^{th} antenna, as follows,

$$\begin{aligned}
[\delta\mathbf{y}(n)]_m &= \mathrm{DP}_K[[\mathbf{x}(n)]_m,\tau] - \mathrm{DP}_K[[\mathbf{z}(n)]_m,\tau] \\
&= \prod_{q=0}^{K-1}\left[\left([\mathbf{x}^{\{q\}}(n-q\tau)]_m\right)^{\binom{K-1}{q}} - \left([\mathbf{z}^{\{q\}}(n-q\tau)]_m\right)^{\binom{K-1}{q}}\right] \\
&= \eta_{(m)}(n)\mathrm{DP}_K[[\mathbf{z}(n)]_m,\tau] \\
&= Be^{j(\phi n+\gamma)}\eta_{(m)}(n)[\mathbf{a}(\mu)]_m
\end{aligned} \tag{4.53}$$

where

$$\begin{aligned}
\eta_{(m)}(n) &\triangleq \prod_{q=0}^{K-1}\left[1+\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{[\mathbf{z}^{\{q\}}(n-q\tau)]_m}\right]^{\binom{K-1}{q}} - 1 \\
&= \prod_{q=0}^{K-1}\left[1+\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{s^{\{q\}}(n-q\tau)[\mathbf{a}^{\{q\}}(\theta,n-q\tau)]_m}\right]^{\binom{K-1}{q}} - 1 \\
&= \prod_{q=0}^{K-1}\left[\sum_{i=0}^{\binom{K-1}{q}}\binom{\binom{K-1}{q}}{i}\left(\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{s^{\{q\}}(n-q\tau)[\mathbf{a}^{\{q\}}(\theta,n-q\tau)]_m}\right)^i\right] - 1,
\end{aligned} \tag{4.54}$$

$s(n)$ is the PPS given in Eqn. (3.3), $\mathbf{a}(\theta,n)$ is the wideband steering array given in Eqn. (3.4), and $[\mathbf{v}(n)]_m$ is the white Gaussian noise observed in sensor m given in Eqn. (3.1). We have also used the following notations: the even/odd-conjugation notation is given as

$$s^{\{q\}}(n) = \begin{cases} s(n), & q \quad \text{even} \\ s^*(n), & q \quad \text{odd} \end{cases} \tag{4.55}$$

and the binomial coefficient is given as $\binom{p}{q} = \frac{p!}{q!(p-q)!}$. Let us consider $\tau = N/K$, which is the value that DPT can achieve the lowest MSE in estimating a_K. With this optimal value of τ, we have the following statistics of $\eta_m(n)$:

$$E\{\eta_m(n)\} = 0 \tag{4.56}$$

$$E\{]\eta_m(n)]^2\} = 0 \tag{4.57}$$

$$E\{\eta_m(n)\eta_m^*(n)\} = \mathcal{K}_{a_K}(K, \text{SNR}) \tag{4.58}$$

(see Eqns.(C.9), (C.12) and (C.13)), where

$$\mathcal{K}_{a_K}(K, \text{SNR}) \triangleq \left[\prod_{q=0}^{K-1} \sum_{i=0}^{\binom{K-1}{q}} \binom{\binom{K-1}{q}}{i}^2 i! \left(\frac{1}{\text{SNR}}\right)^i\right] - 1 \tag{4.59}$$

and $SNR =\triangleq A^2/\sigma_n^2$.

By stacking the perturbation errors in Eqn. (4.53) into vector $\delta\mathbf{y}(n)$, we obtain

$$\delta\mathbf{y}(n) = Be^{j(\phi n+\gamma)}(\mathbf{a}(\mu) \circ \boldsymbol{\eta}(n)) \tag{4.60}$$

where $\circ$ denotes the Schur-Hadamard (element-wise) matrix product and $\boldsymbol{\eta}(n)) = [\eta_0(n), \ldots, \eta_{M-1}(n)]^T$. By using Eqn. (4.56), we have

$$E\{\delta\mathbf{y}(n)\} = Be^{j(\phi n+\gamma)}(\mathbf{a}(\mu) \circ E\{\boldsymbol{\eta}(n)\}) = \mathbf{0} \tag{4.61}$$

Furthermore, because the noise is also assumed to be spatially white, its autocorrelation matrix is given by

$$\begin{aligned} \mathbf{R_v} &\triangleq E\{\delta\mathbf{y}(n)\delta\mathbf{y}(n)^H\} = B^2 E\{\boldsymbol{\eta}(n)\boldsymbol{\eta}(n)^H\} \\ &= B^2 E\{\eta_m(n)\eta_m^*(n)\}\mathbf{I} = B^2\mathcal{K}_{a_K}(K,\mathrm{SNR})\mathbf{I} \end{aligned} \tag{4.62}$$

where $\mathbf{I}$ is the identity matrix.

4.5.2 Performance of JAFE in our Proposed Algorithm

With the first and second-order statistics of the noise in the SHIM domain, which is also zero mean and spatially white as well as temporally white (assuming $\tau = N/K$), the performance analysis of JAFE is can be applied to noisy version of $\mathbf{y}(n)$. We now recall the results for whitened and spatially-temporally smoothed data matrix from [28], which are

$$E\{|\delta\mu|^2\} = \frac{\sigma_{n'}^2}{\xi}\frac{2}{p(N'-p+1)}\frac{1}{L'(M-L')^2} \tag{4.63}$$

$$E\{|\delta\phi|^2\} = \frac{\sigma_{n'}^2}{\xi}\frac{1}{(p-1)^2(N'-p+1)}\frac{1}{L'(M-L'+1)} \tag{4.64}$$

where L' and p are the spatial and temporal smoothing factors, respectively. The number of antennas is M, and N' is the effective number of samples in the trans-

formed domain. Effectively, the number of samples, N', is given as

$$N' = N - N_i = N - (K-1)\tau \tag{4.65}$$

where, K is the order of PPS. The effective SNR is defined as $\xi/\sigma_{n'}^2$, where

$$\sigma_{n'}^2 = A^{2^{2K-2}}\mathcal{K}(K, SNR), \tag{4.66}$$

$$\frac{1}{\xi} = \frac{\sigma_Y^2}{|t|^2(\sigma_Y^2 - \sigma_{n'}^2)}, \tag{4.67}$$

and t is the complex scaling factor and $\sigma_Y^2 = |s_1|^2$ is the largest eigenvalue of the covariance matrix of the spatiotemporally smoothed data matrix, $\mathbf{Y}_{p,L'}$, which is noisy and whitened. The largest singular value, s_1 and t are the same if the data matrix used is noise-free and not spatiotemporally smoothed.

4.5.3 The Performance Analysis of θ and a_K

It is noteworthy to see that in the angle-frequency estimation model (JAFE problem) given in [28], DOA and frequency of interest are related to μ and ϕ, respectively, and independent of each other. While on the other hand, in the wideband PPS problem, as seen in Eqns.(4.26) and (4.27), the DOA is related to both μ and ϕ, and K^{th}-order frequency parameter is only related to ϕ. This would stifle an accurate DOA estimation. To observe the effect of the parameters on estimation accuracy, the first-order perturbation approximation of $\hat{\psi}$, derived by first-order

Taylor approximation for Eqn. (4.12) (a function of two variables, ϕ and μ), is given by

$$\delta\psi \approx \frac{\delta\mu\Delta}{\phi} - \frac{\mu\Delta\delta\phi}{\phi^2} \tag{4.68}$$

Thus, by using Eqn. (4.68), we obtain the MSE as follows,

$$\begin{aligned} E(\delta\psi^2) &\approx \frac{\Delta^2}{\phi^2}E(\delta\mu^2) + \frac{\mu^2\Delta^2}{\phi^4}E(\delta\phi^2) \\ &= \frac{\Delta^2}{\phi^2}E(\delta\mu^2) + \frac{\psi^2}{\phi^2}E(\delta\phi^2) \end{aligned} \tag{4.69}$$

We have used Eqn. (4.12) to get Eqn. (4.69). Since $\psi = \frac{d}{c}\sin\theta$, we have $\delta\theta \approx \frac{c\delta\psi}{d\cos\theta}$, and by substituting Eqn. (4.11) into Eqn. (4.69), we obtain

$$E(\delta\theta^2) \approx \frac{(\Delta c/d)^2 E(\delta\mu^2) + (\sin\theta)^2 E(\delta\phi^2)}{(K!\tau^{K-1}\Delta^K a_K \cos\theta)^2} \tag{4.70}$$

Similarly, the MSE of $\hat{a}_K$ can be obtained straightforwardly from Eqn. (4.11) as follows,

$$E(\delta a_K^2) \approx \frac{E(\delta\phi^2)}{(K!\tau^{K-1}\Delta^K)^2} \tag{4.71}$$

From Eqns.(4.69) and (4.70), it can be observed that the MSE of $\hat{\psi}$ and $\hat{\theta}$ are dependent upon the MSE of $\hat{\phi}$. By increasing $\Delta c/d$, we will reduce the effect of estimation error, $\delta\phi^2$. Furthermore, if we increase Δ, we are relaxing the hardware requirement from sampling fast, and concurrently reducing the MSE of DOA and a_K.

It should also be noted that the performance of DOA relies on the actual parameter a_K. The larger a_K is, which roughly corresponds to larger bandwidth, the better the performance of DOA. As for smaller a_K, one might still be able to compensate the performance by increasing Δ, but it is not possible for very small a_K because Δ is upper bounded as we will see in Subsection 4.5.4. In this case, we apply the following steps:

1. Estimate a_K by applying the harmonic estimator to $DP_K[[\mathbf{x}(n)]_m, \tau]$

2. By using $\hat{a}_K$ estimated from the previous step, a non-parametric estimate of the PPS data of order $(K-1)$, $[\tilde{\mathbf{x}}(n)]_m$, can be constructed

$$[\tilde{\mathbf{x}}(n)]_m = [\mathbf{x}(n)]_m e^{-j\hat{a}_K(\Delta n)^K} \quad \text{for } n = 0, \ldots, N-1 \tag{4.72}$$

3. Apply SHIM of order $K-1$ to the non-parametric estimate of the $(K-1)^{\text{th}}$-order PPS data, we have

$$\mathbf{y}(n, \tau) = \mathbf{DP}_{K-1}[\tilde{\mathbf{x}}(n), \tau], \quad n = (K-2)\tau, \ldots, N-1 \tag{4.73}$$

4. Apply JAFE algorithm to Eqn. (4.73) and extract the DOA estimate by using Eqn. (4.27).

For completeness, the theoretical performance of the algorithm is given by

$$E(\delta\theta^2) \approx \frac{(\Delta c/d)^2 \frac{2}{p(M-L')^2} + (\sin\theta)^2 \frac{1}{(p-1)^2(M-L'+1)}}{(K!\tau^{K-1}\Delta^K a_K cos\theta)^2 (N'-p+1)L'} \left(\frac{\sigma_{n'}^2}{\xi}\right) \tag{4.74}$$

and,

$$E(\delta a_K^2) \approx \frac{1}{(K!\tau^{K-1}\Delta^K(p-1))^2(N'-p+1)L'(M-L'+1)} \left(\frac{\sigma_{n'}^2}{\xi}\right) \tag{4.75}$$

which are obtained by substituting Eqn. (4.63) and (4.64) into Eqn. (4.70) and (4.71), respectively.

4.5.4 The Identifiability Condition

Before proceeding to the choice of optimal parameters, the conditions or the bounds for these parameters to be estimated unambiguously are analyzed. The conditions to estimate DOA and a_K unambiguously are $|\phi| < \pi$ and $|\mu| < \pi$. If one of these conditions is not satisfied, e.g. when $\phi = \upsilon\pi + \epsilon$, where $-\pi < \epsilon < \pi$ is a real-valued variable and υ is any integer, then ϕ is estimated as ϵ, assuming perfect estimation. This is called ambiguous estimation because for any value of υ, ϕ's are estimated as ϵ. These conditions for DOA and a_k, could be rewritten

using Eqn. (4.11) and (4.12) as,

$$\tau^{K-1}\Delta^{K-1}K! < \frac{\pi}{|a_K|\Delta} \tag{4.76}$$

$$\tau^{K-1}\Delta^{K-1}K! < \frac{\pi}{|a_K|(d/c)} \tag{4.77}$$

These inequalities could be combined into one inequality,

$$\tau^{K-1}\Delta^{K-1}K! < \frac{\pi}{|a_K|}\min(\frac{c}{d}, \frac{1}{\Delta}) \tag{4.78}$$

This inequality upper bounds the choices of the parameter Δ one can use, if all the other parameters are held constant. Practically, the choice of d (antenna array spacing) could not be varied on the fly because the hardware is fixed. The choice of τ and Δ could be varied on the fly simply by regenerating the data matrix of SHIM and by down-sampling assuming that the data have been densely sampled and collected. The choice of optimum τ suggested by [18] is independent of Δ, hence it is selected first. In the next subsection, together with the upper bound of Δ, the optimum choice of Δ is derived.

E. The Optimum Δ

To quantify the performance when jointly estimating the two parameters, the geometric mean between the MSE of θ and a_K is computed. The geometric mean is used because it alleviates the need of proper scaling factor. On the other hand,

proper scaling factor is needed if the arithmetic mean had been used. Thus the joint performance measure is expressed as follows,

$$\bar{\mathcal{C}}_{\theta,a_K}(\Delta) \triangleq \sqrt{E\{\delta\theta^2\}E\{\delta a_K^2\}} \tag{4.79}$$

$$= \frac{\sqrt{((\Delta c/d)^2 E(\delta\mu^2) + (\sin\theta)^2 E(\delta\phi^2))E(\delta\phi^2)}}{(K!\tau^{K-1}\Delta^K)^2|a_K\cos\theta|} \tag{4.80}$$

It is clear from Eqn. (4.78) that Δ is upper bounded by

$$\Delta < \min\left(\left(\frac{\pi}{|a_K|\tau^{K-1}K!}\right)^{\frac{1}{K}}, \left(\frac{\pi}{|a_K|\frac{d}{c}\tau^{K-1}K!}\right)^{\frac{1}{K-1}}\right) \tag{4.81}$$

Since Δ is upper bounded by Eqn. (4.81) and lower bounded by 0, the optimum choice of Δ, such that Eqn. (4.80) is globally minimized, could be either within the boundary or at the boundary points. If it is within the boundary points, then the necessary condition for it to be minimum is that its derivative should equal to 0. From Eqn. (4.80), it would be better to use logarithmic cost function to find its derivative. Taking the derivative of Eqn. (4.80) with respect to Δ give

$$\begin{aligned}\frac{\partial}{\partial\Delta}\log(\bar{\mathcal{C}}_{\theta,a_K}(\Delta)) &= \frac{\partial}{\partial\Delta}\Big(\frac{1}{2}\log((\Delta c/d)^2 E(\delta\mu^2) + (\sin\theta)^2 E(\delta\phi^2)) \\ &\quad -2K\log(\Delta) + constants\Big) \\ &= \frac{\Delta(c/d)^2 E(\delta\mu^2)}{(\Delta c/d)^2 E(\delta\mu^2) + (\sin\theta)^2 E(\delta\phi^2)} - \frac{2K}{\Delta} \\ &= \frac{(1-2K)(\Delta c/d)^2 E(\delta\mu^2) - 2K(\sin\theta)^2 E(\delta\phi^2)}{((\Delta c/d)^2 E(\delta\mu^2) + (\sin\theta)^2 E(\delta\phi^2))\Delta}\end{aligned} \tag{4.82}$$

However the numerator is always negative for $K > 1$ and the denominator is always positive. Because logarithm is a monotonically increasing function, Eqn. (4.80) monotonically decreases as Δ increases. However the identifiability conditions in Eqn. (4.78) restrict Δ to be strictly less than its upper bound. Hence, we conclude that there is no optimal Δ. The more Δ is closer to its upper bound, the better the performance of the algorithm.

4.6 Results and Discussion

4.6.1 Simulation Examples

In the following, we apply the proposed algorithm to sonar. Frequency spreading is very common in sonar because signals travelled under the water are non-stationary [72]. In these simulation examples, we consider a ten-element ULA with inter-element spacing of $d = 1.5m$ and the propagation speed of sonar in water $c = 1500m/s$. In Fig. 4.1, root mean square error (RMSE) of a LFM signal, i.e. $K = 2$, is computed through 100 Monte-Carlo runs. The signal parameters[†] are $f_0 = 0$, $f_1 = 420Hz$, $f_2 = 100\frac{Hz}{s}$, $\theta = 40^o$ with sampling interval $\Delta = 4 \times 10^{-3}s$, number of samples, $N = 256$, and $\tau = 128$. In addition to that, the spatiotemporal smoothing of factor $m = 64$ and $L' = 3$ together with forward-backward data extension have been incorporated in the algorithm. In Fig. 4.1, the RMSE of f_2 and DOA are plotted against the SNR. For performance comparison, the results of the ML

[†]here we use the frequency parameters instead of angular frequency parameters which are related by $a_i = 2\pi f_i$

method[‡] and the simulation results of the proposed method, as well as theoretical results of the proposed method, are shown in Fig. 4.1. The performance of the ML method is included just to show how close the performance of the proposed method to the performance of the ML method, although ML method is not a fair estimation method to be compared with. It is not fair for performance comparison because ML method estimates the three parameters, i.e., f_2, DOA and f_1, where f_1 is an unknown nuisance parameter [44], while the proposed method does not estimate f_1. By estimating nuisance parameter, ML improves the performance in estimating the parameters of interests. Therefore, the proposed method is performing much poorer than ML method as seen on right side of Fig. 4.1. Hence, in the figures following Fig. 4.1, we will not include the performance of the ML method, because in the estimation of PPS of order three, there are two unknown nuisance parameters (f_2 and f_1), which will cause the performance of the proposed method become even worse than the ML method.

In Fig. 4.2, the RMSE of both f_2 and θ are plotted against Δ, while SNR is held at 30dB, to show the dependency of RMSEs toward the choice of Δ. Note that the simulation for $\Delta \geqslant 4 \cdot 10^{-3}$ is not shown because it violates the ambiguity condition in Eqn. (4.78), and hence will result in significant RMSE. It can be clearly seen that the simulation results plotted are very close to the theoretical results.

In the following simulation, we show that the algorithm works well for PPS of

[‡]The ML method for single component PPS of order 2 is actually CBF method (see Section CBFsection).

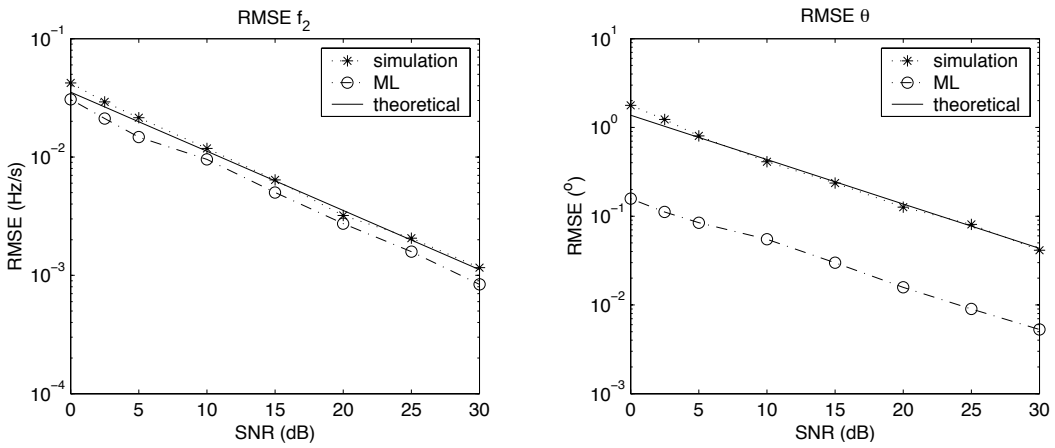

Figure 4.1: Comparison of simulation results between the ML and the proposed method. RMSE of f_2 (Hz/s) and DOA (^{o}) as function of SNR are in (a) and (b) respectively

order $K = 3$, too. The signal parameters for the third-order PPS are $f_0 = 0$, $f_1 = 300Hz$, $f_2 = 100\frac{Hz}{s}$, $f_3 = 100\frac{Hz}{s^2}$, $\theta = 40^0$ with a sampling interval $\Delta = 4 \cdot 10^{-3}$, number of samples $N = 255$ and delay $\tau = 85$. The spatiotemporal smoothing factors are $m = 42$ and $L' = 3$. The RMSE results are plotted against SNR in Fig. 4.3.

4.6.2 Discussion

The proposed algorithm demonstrates the possible application of ESPRIT-based JAFE algorithm in estimation of DOA and frequency parameters of wideband PPS. This is primarily due to SHIM, which transforms the originally wideband PPS into a narrowband sinusoidal signal. In other words, SHIM has two-fold advantages, i.e. the wideband steering array is transformed to the classical narrowband steering

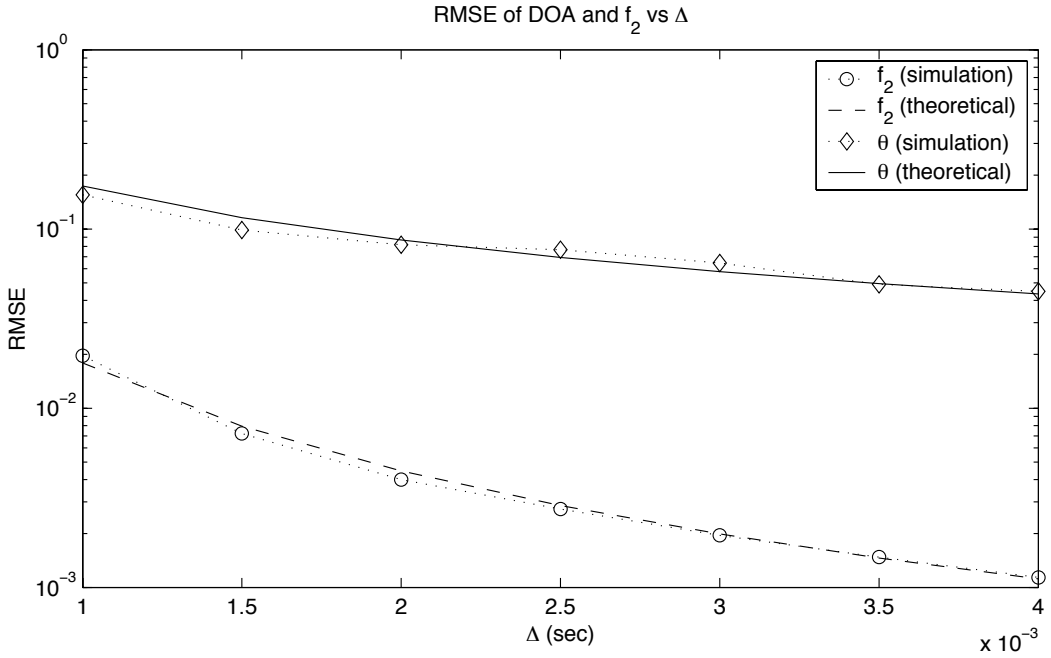

Figure 4.2: RMSE of f_2 (Hz/s) and DOA (^{o}) as function of Δ at SNR=30dB

array and simultaneously PPS is transformed to sinusoidal signal.

Since the algorithm here is only applied to single source, the JAFE algorithm becomes very simple, because the joint diagonalization step is not necessary. However, the joint QR-decompositions in Eqn. (4.42) are still needed to obtain $\mathbf{E}_{x\phi}$, $\mathbf{E}_{y\phi}$, $\mathbf{E}_{x\psi}$ and $\mathbf{E}_{y\psi}$, which are all complex scalars in single source scenario. Therefore, we obtain $\lambda_\phi = \mathbf{E}_{y\phi}/\mathbf{E}_{x\phi}$ and $\lambda_\psi = \mathbf{E}_{y\psi}/\mathbf{E}_{x\psi}$.

The optimum choice of τ is assumed to have the same value as that of DPT. There is no optimum choice of Δ, but the value that is closer to upper bound of its identifiable condition would give lower MSE. Large value of Δ is not recommended because one does not know how large is the actual value of a_K. If it is too large such that user's choice of Δ is larger than the upper bound, then the estimated

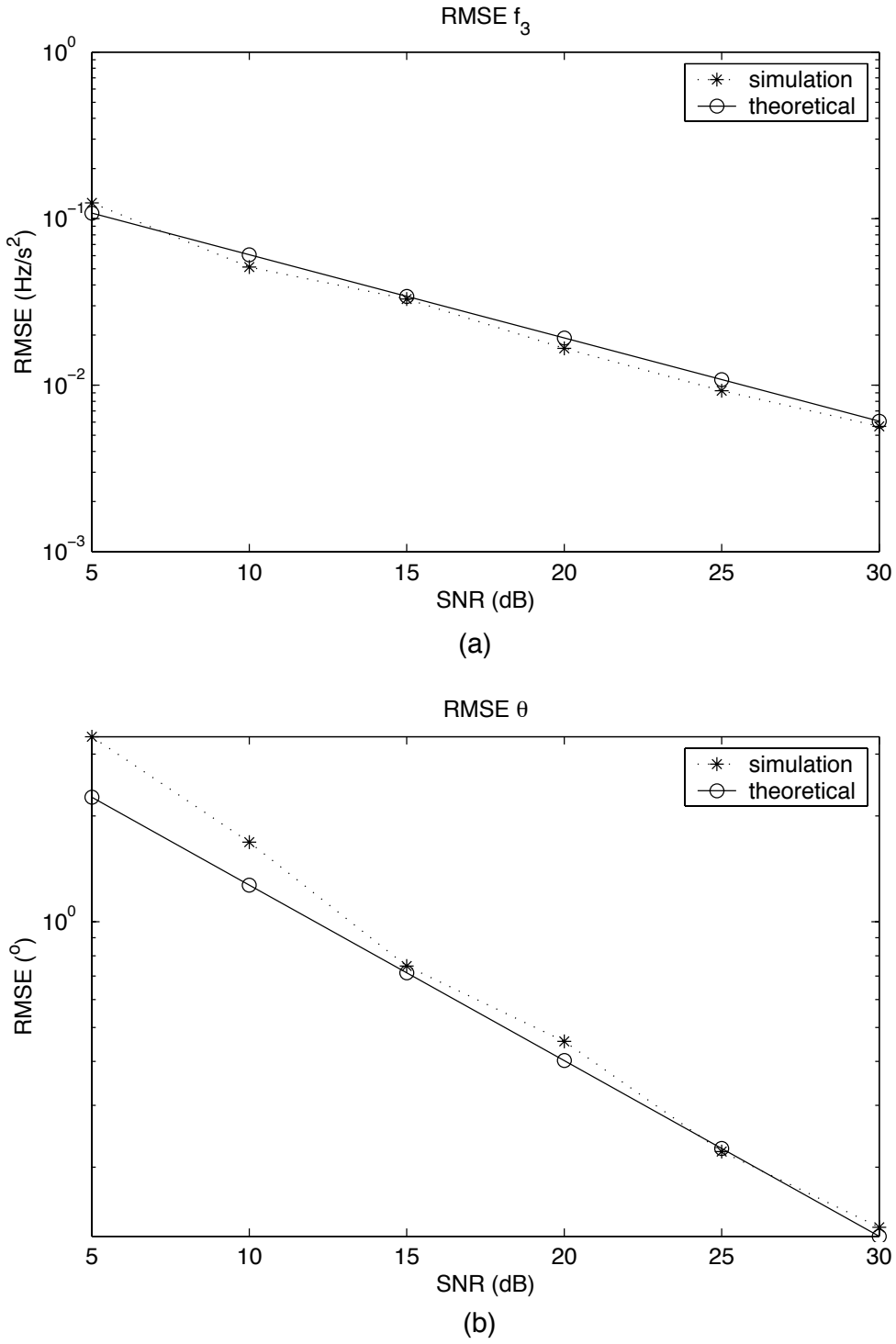

Figure 4.3: RMSE of $f_3(Hz/s^2)$ and DOA (^{o}) as function of SNR are in (a) and (b), respectively

results will be completely erroneous.

The algorithm presented here is primarily for single component PPS. Directly extending SHIM to multi-component PPS and using JAFE algorithm to extract all of their parameters simultaneously is problematic due to the cross-terms that arise because of the non-linearity of SHIM (See Property 3 in Section 2.1.4).

4.7 Summary

In this chapter we have demonstrated a new algorithm to estimate the parameters of PPS that impinged on ULA, whilst in Chapter 3, the proposed algorithm only applicable for the second-order PPS. Here, we introduced the SHIM operator, which essentially transforms the wideband array problem with one PPS into the classical narrowband array problem with one sinusoidal signal. In the following chapter, we will deal with array processing with a more general class of signals, i.e. time-frequency signals, and the mixing channel is assumed to be unknown.

Chapter 5

Underdetermined Blind Source Separation of Time-Frequency Signals

5.1 Introduction

In the previous two chapters, the problems that have been dealt with are array processing of PPS with mixing matrix (channel or array manifold) of known structure. In this chapter, we deal with a more generalized class of signals than PPS, the non-stationary signals that are highly localized in TF plane or have distinct TF signatures. Also, the mixing matrix (the channel) is also completely unknown, but it is assumed to be stationary and non-convolutive. The sources are essentially non-stationary signals. Subsequently, the channel linearly mixes these signals and thereafter a set of linear sensors is used to collect the observations. The objective in this chapter is to recover each source signal from these multiple observations (mixtures) without the knowledge of the mixing matrix and

sources. This is called the blind source separation (BSS) problem. Blind source separation has been used in many applications in engineering [75], for example in radar and sonar applications [76], in communication for multiuser detection [77], in audio/speech processing [78], and in biomedical signal processing to separate EEG signals [79]. Useful theories of BSS can be found in [5,80–84]. A fundamental and necessary assumption of the BSS problem is that these sources are statistically independent, and many solutions are obtained using HOS information [82]. However, one possibly can use only SOS without resolving to HOS, if the other source information is available, such as temporal coherency [27], non-stationarity of signals [85], cyclostationarity of the signals [86] and constant modulus property of the signals [73].

BSS for non-stationary signals were introduced in [85, 87]. They are based on the methods that combine spatial diversity with TF diversity by using STFD. The first advantage of using STFD is the capability of STFD in exploiting non-stationary information of the signals. The second advantage is the capability of STFD in separating Gaussian sources with identical spectral shape but without identical TF localization, where the use of second- and higher-order statistics algorithms are likely to fail (see [83]). The third advantage is the increase in SNR due to its capability to spread noise energy but yet localize signal energy in the TF plane [88]. Since we are dealing with the non-stationary signals, the TF-based BSS method is used in this chapter.

Generally BSS algorithms, such as those mentioned in the references above, work only in determined and overdetermined systems, where the original sources are unmixed by multiplying the inverse or pseudoinverse of the mixing matrix to the received signal vector. In these algorithms, BSS is achieved by first performing blind identification of the mixing matrix, and subsequently the signals are recovered by applying the inverse (pseudoinverse) of the determined (overdetermined) mixing matrix to the received signals. In the determined (overdetermined) system, the mixing matrix is a full column rank square (tall) matrix.

BSS of underdetermined system is a challenging problem even if its wide matrix is available through blind identification algorithms such as those reported in [89–91], because separating sources by the inversion of the mixing matrix is impossible. Hence, obtaining the unmixed source signals would require additional assumptions and processing steps. The problem of underdetermined BSS in general has been studied in [92–97]. Source signals are assumed and limited to a finite set of possible signals [92, 93]. Prior knowledge of source probability density function is needed in [94–96]. However, none of them are TF-based. Similar to [97], sparseness in the TF domain is exploited in the proposed algorithm here, which does not require to know the pdf of source signals or to assume a finite set of possible signals. The methods in [85,87] could be extended for the underdetermined system with condition that the signal signatures in the TF plane could be masked, or partitioned into groups so that each group contains only fewer or equal number of

signals than antenna array sensors. After performing the masking or partitioning, the BSS techniques [85, 87] are then applied to each of the partitioned group. In addition, there are some extra algorithmic steps that are essential to mitigate the cross-terms (CTs) generated between these groups. In this chapter we propose a method for source separation of underdetermined systems with the possibility that signal signatures in the TF plane are non-disjoint, where grouping and partitioning are not applicable.

Authors in [98] proposed to mitigate UBSS problem with disjoint signal signatures in TF plane. The algorithm uses the clustering algorithm and exploits the STFD [85] structure at the single auto points (SAPs), i.e., the location in TF plane where individual source exists alone. Also, the same algorithm was applied to signals with few overlappings in their TF signatures, and it performs well except at the multiple auto points (MAP), i.e., the location where TFDs of two or more sources intersect in TF plane [98]. In [99], the authors proposed a new subspace-based algorithm to perform separation on both the SAPs and MAPs, assuming at MAPs there are fewer number of overlapped sources than the number of sensors. However, applying this subspace-based algorithm to each of the SAPs and the MAPs could be computationally expensive.

In this chapter, we propose a separation technique which relies on pseudoinverse of the virtual array structure [100] of the vectorized STFD matrices of the SAPs. It assumes that the mixing matrix has been obtained through other means, such

as [101]. Herein, we also extend the method in [99] to cross points (CPs). The CPs are the locations of CTs. In addition to that, we also propose a new method for selecting mixtures of CPs and MAPs. With the mixture of MAPs and CPs selected, STFD matrices at these TF points are processed in a way similar to [99], at the lower computational cost resulting from fewer points being processed since only MAPs and CPs, and not the SAPs, need to be processed. Because of lower computation cost and extensibility of the subspace method, we have the luxury to use Wigner-Ville (WV)-based STFD. WV-based STFD has many unsuppressed CTs, which are advantageous for source synthesis of multicomponent signal from any single source, such as audio sources which contain harmonics.

5.2 Signal Model

Assume instantaneous mixing matrix $\mathbf{A} \triangleq [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_L]$ with L narrowband signals impinging on a set of M sensors. Since we are dealing with underdetermined system, we therefore have $L > M$. It is also assumed that any M of the L columns of $\mathbf{A}$ are linearly independent. The received signal is modeled as

$$\mathbf{x}(t) \triangleq \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t), \tag{5.1}$$

where $\mathbf{v}(t)$ is the $M \times 1$ additive white Gaussian noise vector with zero mean. The vector $\mathbf{s}(t) \triangleq [s_1(t), \ldots, s_L(t)]^T$ is the source signal vector of size $L \times 1$ and each of

the $s_i(t)$ is a non-stationary source signal. Without loss of generality, the first row of $\mathbf{A}$ is assumed to be real-valued and each column of $\mathbf{A}$ has been normalized. This is to provide unique solution when estimating the mixing matrix. Before making assumptions on the sources, we first define STFD of the received signals, $\mathbf{x}(t)$, as follows:

$$\mathbf{D_{xx}}(t,f) \triangleq \sum_{l=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} \phi(m,l)\mathbf{x}(t+m+l)\mathbf{x}^H(t+m-l)e^{-j4\pi fl} \tag{5.2}$$

where $\phi(m,l)$ is the TFD time-lag kernel which is applied to all received sensors equally and $(\cdot)^H$ denotes the Hermitian transpose. There are various TFD time-lag kernels to be chosen from, depending on how the cross-terms are to be suppressed. Assuming from now no noise is presence, which is also a practical assumption because noise power always spreads out evenly on the TF plane, the received signal STFD is related to the source signal STFD, $\mathbf{D_{ss}}(t,f)$, in the following way,

$$\mathbf{D_{xx}}(t,f) = \mathbf{A}\mathbf{D_{ss}}(t,f)\mathbf{A}^H \tag{5.3}$$

Basically, elements of the STFDs, e.g. $[\mathbf{D_{ss}}(t,f)]_{i,j} \triangleq D_{s_is_j}(t,f) = \sum_l\sum_m \phi(m,l)s_i(t+m+l)s_j^*(t+m-l)e^{-j4\pi fl}$, is an auto-TFD (if $i=j$) or cross-TFD (if $i \neq j$).

Definition 3. *The two sources, $s_i(t)$ and $s_j(t)$, are disjoint if and only if $\Omega_i \cap \Omega_j = \varnothing$, where Ω_k is the TF support* of the source k's TFD. Conversely, the two sources*

*Assume source $s_i(t)$ has TFD, $D_{s_i,s_i}(t,f)$, then Ω_i denotes its TF support, if and only if

are called non-disjoint.

Definition 4. *Suppose there are only two sources, $s_i(t)$ and $s_j(t)$, then the TF points such that $(t,f) \in \{(\Omega_i \cup \Omega_j) - (\Omega_i \cap \Omega_j)\}$ is called the SAP, and the TF such that $(t,f) \in \{\Omega_i \cap \Omega_i\}$ is called the MAP. (Note that if the sources are disjoint, then $\{\Omega_i \cap \Omega_i\} = \{\emptyset\}$.)*

Source signals in this chapter are allowed to be either disjoint or non-disjoint. It is assumed that the SAPs of each source exist, which is the requirement when estimating $\mathbf{A}$. We further assumed that at most $M-1$ sources intersect at any MAPs, or mathematically $\Omega_{i_1} \cap \Omega_{i_2} \cap \ldots \cap \Omega_{i_M} = \varnothing$ for any sets of M sources. This assumption is essential for estimating the TFDs of sources at the MAPs. Before we proceed to the next section, we define the following,

Definition 5. *The (t,f) point such that $D_{s_i s_j}(t,f) \neq 0$, for $i \neq j$, and is not a MAP, is defined as the CP. Both the $D_{s_i s_j}(t,f)$ which are evaluated at the CP and at the MAP are called the CT.*

It is important to note that, the definitions for CP and CT arise when there are at least two sources. However, CT also arise due to multicomponent signal within one source, and, it will appear as the SAP. From now on, unless it is specifically mentioned, CP and CT refer to the definition above and CT of a multicomponent signal is processed in the same way as processing STFD at the SAPs.

$\forall (t,f) \in \Omega_i$, $D_{s_i,s_i}(t,f) \neq 0$.

5.3 Properties of Distributions at the Time-Frequency Points

We first look at the property of STFDs at the SAPs. It has been studied in [102] that $\mathbf{D_{ss}}(t,f)$ have a diagonal structure only at the SAPs. In fact, only at the i^{th} diagonal element where $D_{s_i s_i}(t,f)$ is non-zero, and the rest of the entries are zero. Hence, Eqn. (5.3) becomes $\mathbf{D_{xx}}(t,f) = \mathbf{A}\text{diag}\{[0\ldots,0,D_{s_i s_i}(t,f),0,\ldots 0]\}\mathbf{A}^H = \mathbf{a}_i\mathbf{a}_i^H D_{s_i s_i}(t,f)$, which is rank one and semi-positive definite since $D_{s_i s_i}(t,f) > 0$. Note that the CT due to the multicomponent signal within one source will also have this property. We next look at the property of $\mathbf{D_{ss}}(t,f)$ at the MAPs. In general $\mathbf{D_{ss}}(t,f)$ at the MAPs are not diagonal because the CTs at the MAPs are non-zeros, and $\text{rank}\{\mathbf{D_{ss}}(t,f)\} = \text{rank}\{\mathbf{D_{xx}}(t,f)\} = k$ if it is at the MAP of k sources, i.e. $(t,f) \in \{\Omega_{i_1} \cap \Omega_{i_2} \cap \ldots \cap \Omega_{i_k}\}$. In addition to that, STFDs at the MAPs are Hermitian symmetric and indefinite matrices. Lastly, we study the property of $\mathbf{D_{ss}}(t,f)$ at the CPs. It will only have off-diagonal entries, because only the CTs are non-zeros. The $\text{rank}\{\mathbf{D_{ss}}(t,f)\} = \text{rank}\{\mathbf{D_{xx}}(t,f)\} = k$ if there exist CTs resulting from k sources. Note that, regardless of the kernel, the CTs near or at the MAPs are difficult to be suppressed without suppressing the signal TFD itself. We will see how these CPs will not affect the performance of the source separation in the Section 5.5. This gives the flexibility to use the original WV distribution without any suppression of the CTs.

5.4 TF Points for Blind Identification

There is no intention to propose a new blind identification method in this chapter, however, we will discuss briefly the method of selecting TF points for the blind identification algorithms, such as [101]. These blind identification algorithms rely on the diagonal form of the sources' STFD matrices. The objective of blind identification is to have sufficient TF points such that their sources' STFDs can be in diagonal form. This implies we need to select STFDs at the SAPs. However, only sufficient and small numbers of the SAPs are needed for blind identification. Hence, we could use a detection scheme that has low error probability when detecting for the SAPs, such as the scheme reported in [102], to avoid STFDs, which are not approximately diagonal matrices, from being processed in the blind identification. This high selectivity of the detection will also lower the computation cost of the blind identification, as a result of fewer STFDs are being processed.

The rationale behind the detection scheme in [102] is to exploit the rank one property of the STFDs at the SAPs. Therefore, we have

$$C(t,f) \triangleq \frac{\max|\text{eig}\{\mathbf{D}_{\mathbf{xx}}(t,f)\}|}{\sum|\text{eig}\{\mathbf{D}_{\mathbf{xx}}(t,f)\}|} \tag{5.4}$$

with $C(t,f) = 1$ for t, f that belong to the SAPs. However, in practice there are almost no TF points which will satisfy $C(t,f) = 1$, hence the authors in [102] suggested to search for the TF locations such that $C(t,f)$ is high, i.e., the local

maxima points of $C(t,f)$. Thus, the criteria is to categorize TF points as the SAPs if they satisfy the following

$$\begin{cases} \|\text{Grad}_{C(t,f)}\|_2 & \leq \epsilon_{Grad} \\ \mathbf{H}_{C(t,f)} & \leq \mathbf{0} \end{cases} \tag{5.5}$$

where $\text{Grad}_{C(t,f)}$ and $\mathbf{H}_{C(t,f)}$ are the gradient function and the Hessian matrix of $C(t,f)$, respectively. The threshold, ϵ_{Grad} adjusts the number of the TF points, that are categorized as the SAPs, in the neighborhood of a local maximum.

5.5 Proposed Source Separation Algorithm

5.5.1 Algorithm Overview

The objective of source separation is to estimate all the individual time-domain source signals. However, if one has the source's TFD, one can invert it uniquely, up to a complex constant, to yield the source signal in time domain [103]. Thus, the estimation of sources' TFDs from STFDs is the main issue in this chapter. Initially, TF points that contain only noise are ignored and their TFDs have to be zeroed out. This is called noise-thresholding. The remaining TF points would be either SAPs, MAPs or CPs. Here, we propose a method to use the STFDs at the SAPs to obtain the individual source's TFDs at the SAPs. Apparently,

the separation method also prompts a new technique to separate the SAPs from MAPs and CPs. Now, only the STFDs at MAPs and CPs have not been treated yet. A subspace method, which is originally meant for TFD separation at MAPs and SAPs only [99], is analyzed at the CPs and then its property are exploited to process STFDs at mixture of MAPs and CPs that remained from the previous step. Finally, one could construct individual TFDs at SAPs, MAPs and CPs (the other TF points are zeros after noise-thresholding) and inverting them to estimate the source signals in time domain. In the following subsections, we will elaborate the procedure of the algorithm.

5.5.2 Proposed Simultaneous TFDs Separation at SAPs

Preceding any processing, the noise-thresholding step is performed by selecting the TF points that satisfy the following,

$$\text{trace}\{\mathbf{D_{xx}}(t,f)\} \geq \epsilon_1 \underset{(t,f)}{\text{mean}} \{\text{trace}\{\mathbf{D_{xx}}(t,f)\}\} \tag{5.6}$$

where the value of ϵ_1 typically is 1 (see [102]). Following the noise-thresholding, blind identification of $\mathbf{A}$ is performed as discussed in Subsection 5.4. Thereafter, the proposed source separation algorithm is performed on the SAPs, in which the method of separating SAPs from the mixture of SAPs, MAPs and CPs will be presented in the next subsection. The algorithm exploits the diagonal structure at

the SAPs, and hence, by vectorizing Eqn. (5.3) gives

$$\mathbf{y}(t,f) \triangleq \text{vec}\{\mathbf{D_{xx}}(t,f)\} = (\mathbf{A}^* \odot \mathbf{A})\mathbf{z}(t,f) \tag{5.7}$$

where $\mathbf{z}(t,f) \triangleq \text{diag}\{\mathbf{D_{ss}}(t,f)\}$ is the vector that contains the diagonal entries of sources STFD, and $\odot$ is the Khatri-Rao product (see [104]). Note that the size of the virtual array, $\mathbf{A}^* \odot \mathbf{A}$, is $M^2 \times L$. Even when $M < L$, the condition $M^2 > L$ is easily achievable to form a full rank virtual array matrix [100], and hence solving for $\mathbf{z}(t,f)$ in Eqn (5.7) becomes a full-rank (overdetermined) least squares problem now. For example, with only three sensors, it is possible to perform separation of TFDs up to eight sources at all SAPs. Mathematically, the estimate of separated TFDs at SAPs is simply

$$\hat{\mathbf{z}}(t,f) = (\hat{\mathbf{A}}^* \odot \hat{\mathbf{A}})^\dagger \mathbf{y}(t,f) \tag{5.8}$$

where $(\hat{\mathbf{A}}^* \odot \hat{\mathbf{A}})^\dagger$ is the pseudoinverse of the virtual array in this full-rank least squares case, i.e. $(\hat{\mathbf{A}}^* \odot \hat{\mathbf{A}})^\dagger = [(\hat{\mathbf{A}}^* \odot \hat{\mathbf{A}})^H (\hat{\mathbf{A}}^* \odot \hat{\mathbf{A}})]^{-1} (\hat{\mathbf{A}}^* \odot \hat{\mathbf{A}})^H$, where $\hat{\mathbf{A}}$ denotes an estimate of $\mathbf{A}$. Note also, one could stack $\mathbf{y}(t,f)$ from different SAPs column-wise into matrix, as follows

$$\mathbf{Y} \triangleq [\mathbf{y}(t_1,f_1), \ldots, \mathbf{y}(t_{K_5},f_{K_5})] \tag{5.9}$$

where K_5 is the number of SAPs selected by the technique to be discussed in next subsection. Hence, with Eqn. (5.9), we can batch processes $\mathbf{Y}$ to obtain the stacked $\hat{\mathbf{z}}(t,f)$ from different SAPs, $\hat{\mathbf{Z}} \triangleq [\hat{\mathbf{z}}(t_1,f_1),\ldots,\hat{\mathbf{z}}(t_{K_5},f_{K_5})]$, just by one matrix multiplication, as follows

$$\hat{\mathbf{Z}} = (\hat{\mathbf{A}}^* \odot \hat{\mathbf{A}})^{\dagger}\mathbf{Y}. \tag{5.10}$$

5.5.3 Proposed SAPs, MAPs and CPs Detection

Now, suppose that pseudoinverse of the virtual array is applied to the vectorized STFD matrix at the MAPs or the CPs as in Eqn. (5.8), then it will lead to the following equation

$$\hat{\mathbf{z}}(t,f) = (\hat{\mathbf{A}}^* \odot \hat{\mathbf{A}})^{\dagger}(\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})\mathbf{w}(t,f) \tag{5.11}$$

where $\otimes$ is the Kronecker product. Without lost of generality, we assumed that the MAPs or the CPs are the points where the first L' sources are overlapping. Hence, $\tilde{\mathbf{A}} \triangleq [\mathbf{a}_1,\ldots,\mathbf{a}_{L'}]$ and the vectorized non-diagonal STFD matrix of the first L'-sources at the MAPs or the CPs is given by $\mathbf{w}(t,f) \triangleq \text{vec}\{\tilde{\mathbf{D}}_{\mathbf{ss}}(t,f)\}$. The Kronecker product arises due to non-diagonal structure of the sources' STFD at MAPs and CPs. Assuming perfect estimation of $\mathbf{A}$, some of the columns of the virtual array $\mathbf{A}^* \odot \mathbf{A} = [\mathbf{a}_1^* \otimes \mathbf{a}_1, \mathbf{a}_2^* \otimes \mathbf{a}_2, \ldots, \mathbf{a}_L^* \otimes \mathbf{a}_L]$ are contained in $\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}} =$

$[\mathbf{a}_1^* \otimes \mathbf{a}_1, \mathbf{a}_1^* \otimes \mathbf{a}_2, \ldots, \mathbf{a}_1^* \otimes \mathbf{a}_{L'}, \ldots, \mathbf{a}_{L'}^* \otimes \mathbf{a}_{L'}]$. This leads to

$$(\mathbf{A}^* \odot \mathbf{A})^\dagger (\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}) = [\mathbf{e}_1, \star, \ldots, \star, \mathbf{e}_2, \star, \ldots, \star, \mathbf{e}_{L'}] \tag{5.12}$$

where $\star$'s are the arbitrary column vectors and $\mathbf{e}_k = [0, \ldots, 0, 1, 0, \ldots 0]^T$ is a unit vector with all the elements equal zeros except the k^{th} element. Every $(k + (k-1)L')^{\text{th}}$ column of $(\mathbf{A}^* \odot \mathbf{A})^\dagger (\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})$ gives $\mathbf{e}_k$. Substituting Eqn. (5.12) into Eqn. (5.11) results in

$$\hat{\mathbf{z}}(t,f) = \begin{bmatrix} D_{s_1 s_1}(t,f) + \text{cross-terms} \\ D_{s_2 s_2}(t,f) + \text{cross-terms} \\ \vdots \\ D_{s_{L'} s_{L'}}(t,f) + \text{cross-terms} \end{bmatrix} \tag{5.13}$$

at the MAPs and similarly for the CPs except that $D_{s_k s_k}(t,f) = 0$ for all k. This prompts a new way of separating the MAPs and the CPs out from the mixture of MAPs, CPs and SAPs. Keep (t,f) as the SAPs and $\max_i\{\hat{z}_i(t,f)\}$ as $\hat{D}_{s_{i_{\max}} s_{i_{\max}}}(t,f)$ if,

$$\frac{\max_i\{\hat{z}_i(t,f)\}}{\sum_i |\hat{z}_i(t,f)|} \geq 1 - \epsilon_2, \tag{5.14}$$

otherwise group (t,f) as a mixture of MAPs and CPs. Here, $\hat{z}_i(t,f)$ is the i^{th} element of $\hat{\mathbf{z}}_i(t,f)$ and $i_{\max}$ is the index that maximizes the numerator of the Eqn (5.14). The value of ϵ_2 is chosen to be a small value less than 1, typically is chosen to be $0.1 \sim 0.5$. We will see in the next subsection the reason that this value is not that critical.

5.5.4 Subspace Separation Method at MAPs and CPs and Its Property

Originally, the subspace method is intended for source separation at the SAPs and the MAPs, but not at the CPs, which was described [99]. However, here we will show that subspace method is also applicable for source separation at the CPs as well, which means that suppression of CTs is not that crucial now. Another advantage when dealing with a source with multicomponent signal is in estimating its original TFD that contains the CTs, where suppression of CTs in this case is in disadvantage. This also means the choice of ϵ_2 is not that crucial if all the SAPs, MAPs and CPs can be processed by using the subspace method. However, the subspace algorithm processes STFDs at each of the TF points and hence it is computationally expensive. Thus, processing more SAPs by the proposed method in subsection 5.5.2 can reduce the computational load due to its batch processing nature in Eqn. (5.10).

Now, we will observe the property of the subspace method at the MPs and the CPs. It is assumed that the number of signal sources involved in the CT at CPs, L', are less than $M-1$, which is the same assumption used for MAPs previously. Thus, at any MPs or CPs, we perform the EVD to obtain the subspace of $\tilde{\mathbf{A}}$,

$$\mathbf{D_{xx}}(t,f) = \tilde{\mathbf{A}}\tilde{\mathbf{D}}_{\mathbf{ss}}(t,f)\tilde{\mathbf{A}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \tag{5.15}$$

where $\mathbf{U}$ corresponds to the L'-largest eigenvalues in magnitude. Magnitude of the eigenvalues is used due to the Hermitian symmetric indefiniteness of MPs and CPs. Next, $\tilde{\mathbf{A}}$ could be identified by

$$\tilde{\mathbf{A}} = \min_{\{i_1,\ldots,i_{L'}\}} \|(\mathbf{I} - \mathbf{U}\mathbf{U}^H)\mathbf{a}_i\| \tag{5.16}$$

which, basically finding a set of L $\mathbf{a}_i$'s, which is obtained from $\hat{\mathbf{A}}$, such that their orthogonal projections to subspace of $\tilde{\mathbf{A}}$ are minimized. Following that, the TFDs at the MAPs or CPs can be extracted from the diagonal elements of the following,

$$\tilde{\mathbf{D}}_{\mathbf{ss}}(t,f) = \tilde{\mathbf{A}}^{\dagger}\mathbf{D}_{\mathbf{xx}}(t,f)(\tilde{\mathbf{A}}^{\dagger})^H \tag{5.17}$$

If it is the STFD of the CPs, then the diagonal entries will be small and near to zero. This is the property that allows the subspace method to be applied to STFDs at CPs. Note that CT due to multicomponent signal will not be zero because it has the spatial structure of STFDs at the SAPs as mentioned in previous section.

5.5.5 Synthesis of Sources

Finally the source separated TFDs are formed as follows,

$$\hat{D}_{s_r s_r}(t,f) = \begin{cases} \hat{z}_r(t,f) & \text{at SAPs by (5.8)} \\ \tilde{D}_{s_{i_r} s_{i_r}}(t,f) & \text{at MAPs/CPs by (5.17)} \\ 0 & \text{elsewhere} \end{cases} \tag{5.18}$$

where all the STFDs used in the algorithm is chosen to be Wigner-Ville-based (WV) or Modified WV-based (MWV) [88], which is needed in order to perform the inversion. Finally, source signals could be synthesized from the separated TFDs, by inverting the WVD as follows, $s_i(t) = \frac{1}{s_i^*(0)} \int_{-\infty}^{\infty} D_{s_i s_i}(\frac{t}{2}, f) e^{j2\pi ft} df$ where its discrete time implementation could be found in [103]. It is also noteworthy to use WVD rather than the modified WVD (MWVD) in the case when sources have multicomponent signal, because the MWVD suppresses the CTs while WVD does not. The proposed algorithm is summarized in Table 5.1.

Note that the algorithm in [99] only involves steps 1-3 and 6-7, and step 6 in [99] is used to obtain TFDs at the SAPs and MAPs (c.f. step 6 of proposed method is applied to MAPs and CPs).

Table 5.1: Summary of the new STFD-based underdetermined BSS

Given sensors output $\mathbf{x}(n)$

1. Compute WV-based or MWV-based STFD in Eqn. (5.2)
2. Noise thresholding using Eqn. (5.6) to obtain signal TF points
3. Select STFDs at SAPs by [102] and estimate $\mathbf{A}$ by [101]
4. Separate MAPs/CPs from SAPs for BSS by applying Eqn. (5.10) and (5.14) to the STFDs at signal TF points obtained from step 2
5. Obtain source separated TFDs at SAPs using $\max_i \hat{z}_i(t,f)$, which has been evaluated in the previous step
6. Obtain TFDs at MAPs/CPs using Eqn. (5.15), (5.16) and (5.17)
7. Form source separated TFDs as in Eqn. (5.18)
8. Synthesize the source separated signals by inverting TFDs [103]

5.6 Simulation Results

In this section, the simulations are performed to show the effectiveness of the proposed algorithm and it is compared to the existing subspace algorithm proposed in [99]. The subspace algorithm is used for comparison, because other existing

algorithms are not applicable to the situation where TF signatures are non-disjoint. In the proposed algorithm, WV-based STFD is used while in [99] the MWV-based STFD is used. In order to make a fair comparison, both algorithms are assumed to have perfect estimation of $\mathbf{A}$, which is randomly generated. There are four sources ($L = 4$) and three sensors ($M = 3$). Three sensors are needed for both algorithms to work, because there are two sources involved at MPs/CPs and more sensors than sources are required for the subspace method to work at MPs/CPs.

In the first example, the additive noise is assumed to be zero-mean white Gaussian and the SNR is assumed to be 5 dB. There are 256 number of snapshots collected each with sampling rate one sample every second and $\epsilon_2 = 0.4$. The parameter ϵ_1 is chosen to be the same for both algorithms. The sources are two single tones with frequencies of $0.2\frac{rad}{s}$ and $2.4\frac{rad}{s}$ and two linear FM signals with instantaneous frequencies of $-0.007t + 2.7\frac{rad}{s}$ and $0.006t\frac{rad}{s}$. With these types of sources, it is impossible to mask or to partition the TF plane into TF planes that contain less than four sources without partitioning any source's TF signature. Thus, one cannot apply the method in [85] onto each partition.

Figure 5.1 shows the original sources' TFDs in the first row, the TFDs at each sensor output in the second row, the estimated sources' TFDs using the proposed method in the third row, and the estimated sources' TFDs using the method similar to [99] in the last row. The results of both algorithms are almost identical if one observes from these plots. In fact, based on the normalized mean

square error (NMSE) with $N_{mc} = 100$ Monte Carlo runs, the performance in estimating of sources 1, 2, 3, 4 are -17.64, -16.95, -17.85, -17.07 dB, respectively for the proposed algorithm, and -16.88, -16.37, -17.21, -16.45 dB, respectively for the existing subspace algorithm. The NMSE of source i is defined as

$$NMSE = \frac{1}{N_{mc}} \sum_{r=1}^{N_{mc}} \frac{\|\hat{\mathbf{s}}_i^{(\Omega_r)}(\mathbf{t}) - \alpha_i \mathbf{s}_i(\mathbf{t})\|^2}{\|\mathbf{s}_i(\mathbf{t})\|^2}$$

where α_i is a complex valued scalar to take care of the scaling invariance in the BSS, since the objective of BSS is to estimate the signals as accurate as possible while allowing scaling ambiguity. The notation $\mathbf{s}_i(\mathbf{t}) \triangleq [\mathbf{s}_i(0), \ldots, \mathbf{s}_i(N-1)]$ defines the N snapshots of the source i signal. The N snapshots estimated source i signal for the r^{th} Monte-Carlo run is denoted as $\hat{\mathbf{s}}_i^{(\Omega_r)}(\mathbf{t})$. From the results of the first example, it can be concluded that their performances are almost the same. However, in this example the proposed algorithm is demonstrated to have the advantage of less computation speed while keeping the same performances. The reduced computation speed is due to the significantly less number of TF points at which STFD matrices need to be eigen-decomposed as in step 6 of Table 5.1, which causes most of the computational load. Denote the number of TF points selected from step 2 as K_2, and the total number of MAPs and CPs selected from step 4 as K_4, then the number of flops in step 6 in order to obtain the TFDs of MAPs and SAPs are dominated by K_2 times the complexity of EVD $\mathcal{O}(K_2 M^3)$ for the algorithm in [99]. On the other hand, the number of flops in step 6 in order to

obtain TFDs of MAPs and CPs is K_4 times the complexity of EVD ($\mathcal{O}(K_4M^3)$) in the proposed algorithm. All the other steps are the same in both algorithms except the extra steps 4 and 5 used in the proposed algorithm. However, these two steps contribute very little to the complexity. From the MATLAB profiling function, step 6 of the proposed algorithm is 2.7 times faster than that of algorithm in [99] because of the difference in computational complexity of EVD in step 6. The additional step 4 and 5 only took up 7% of the total computational time. Although step 4 and 5 cost more due to its complexity, these steps are fast, because we can perform batch processing through Eqn. (5.10).

In the second and third examples, the additive noise is assumed to be zero-mean Gaussian and the SNR is varied up to 30 dB. Furthermore, the values of ϵ_1 and ϵ_2 are tuned such that the computational speeds of both algorithms are the same (using MATLAB profiling function) for comparison of their performance gains. The number of snapshots collected is the same as the first example. In the second example, the four sources are two single tones with frequencies of $0.2\frac{rad}{s}$ and $2.2\frac{rad}{s}$ and two linear FM signals with instantaneous frequencies of $0.0074t\frac{rad}{s}$ and $-0.008t+2.7\frac{rad}{s}$, respectively. In the third example, the four sources are two single tones with respective frequencies of $0.2\frac{rad}{s}$ and $2.2\frac{rad}{s}$, one linear FM signal with instantaneous frequency of $-0.008t + 2.7\frac{rad}{s}$ and one multicomponent source that consists of two linear FM signals crossing each other with instantaneous frequencies of $0.009t\frac{rad}{s}$ and $-0.009t + 1.148\frac{rad}{s}$. With these types of sources, it is impossible

to partition the TF plane into smaller TF planes such that each plane contains less than four sources without partitioning any source's TF signature. Thus, one also cannot apply the method in [85] to each partition in these examples.

Figures 5.2 and 5.4 show the overall normalized mean square error (ONMSE) performance over $N_{mc} = 100$ Monte Carlo runs. The ONMSE is just

$$ONMSE = \frac{1}{N_{mc}} \sum_{r=1}^{N_{mc}} \frac{\|\hat{\mathbf{S}}^{(\Omega_r)} - \text{diag}\{\alpha_1, \ldots, \alpha_K\}\mathbf{S}\|_F^2}{\|\mathbf{S}(\mathbf{t})\|_F^2}$$

where, $\hat{\mathbf{S}}^{(\Omega_r)} = [\hat{\mathbf{s}}_1^{(\Omega_r)}(\mathbf{t})^T, \ldots, \hat{\mathbf{s}}_L^{(\Omega_r)}(\mathbf{t})^T]^T$ and $\mathbf{S} = [\mathbf{s}_1(\mathbf{t})^T, \ldots, \mathbf{s}_L(\mathbf{t})^T]^T$. In Fig. 5.2, the results of the second example show that the proposed method is better than the existing subspace method [99] by about 1 dB at SNR 30 dB. In Fig. 5.3, the result of the existing subspace method suffers from restoring the original sources TFDs, particularly for sources 2 and 4. The sources in this example are non-disjoint, however, each of them is still mono-component LFM signal. The results of the proposed method are expected to be even more dramatic in the third example as shown in Fig. 5.4. The performance gains at SNR 20 dB to 30 dB are almost 3 dB. This is because the proposed algorithm uses WV-based STFD, where the useful CTs—the CTs belonging to multicomponent signal source—are not suppressed at all; on the contrary, MWV-based STFD suppresses all of the CTs indiscriminately. The strength of the proposed algorithm is its capability to retain the useful CTs and to dispose the unwanted CTs. With WV-based STFD, some of the useful TF points at the CTs are processed as SAPs because they have the same STFD

structure as SAPs. The other useful TF points at the CTs, with the proposed algorithm speed maintained at the same speed as the existing subspace algorithm through the control of ϵ_1 and ϵ_2, could be allowed to be included in the subspace processing in Subsection 5.5.4 (or in step 6). Fig. 5.5 shows the severe degradation in restoring source 2 occurs to the existing subspace algorithm. Especially in the region proximity to the MAPs of source 2, TFD is almost nulled completely.

5.7 Discussions

This chapter has demonstrated a better underdetermined source separation technique by the use of time-frequency distributions. The gain in performance is achieved without loss of the computational speed is mainly attributed by two factors. The first factor is caused by the use of WV-based STFD which do not suppress CTs resulting from either the product between different sources or from the multicomponent signal of a given source, where the latter type of CTs are not nuisance signals. To exploit these CTs from multicomponent signal, we leverage on the spatial structure of STFD, which can reveal the distinctions between these two types of CTs, by utilizing the subspace method at CPs which has not been exploited in [99]. This property of using subspace method to distinguish the CPs, whether are they resulting from multicomponent, was not discovered in [99]. With the WV-based STFD and subspace algorithm applied to SAPs, MPs and CPs, the source separation will perform well even when there is a multicomponent source.

However, now the computational speed is severely reduced if one uses the subspace algorithm alone because many CPs are included for processing. The second factor is caused by the new batch processing of (5.8), which off-loads the computational burden of the subspace method, especially for the TF points, where they are detected as SAPs.

5.8 Summary

In this chapter, an underdetermined blind source separation method was proposed for signals that have distinct time-frequency signatures. It exploited the sparseness of time-frequency signatures in time-frequency domain. The algorithm is robust averse to signals which overlapped in their time-frequency signatures. There are several main ideas in this proposed algorithm. Firstly, the TFDs that contains cross-terms are useful if they can be exploited properly. We have demonstrated by applying the subspace method to the CTs, one can distinguish and keep the CTs due to multicomponent signal from the CTs due to the product of two different sources. Secondly, we proposed a batch separation method at SAPs, which allowed us to increase the processing speed.

In this chapter and previous two chapters, the noise has always been assumed to be additive white Gaussian noise. In the following chapter, we will deal with Gaussian noise that could be either white or colored. Also the signal sources are

not assumed to be time-frequency signals, but just zero-mean signals.

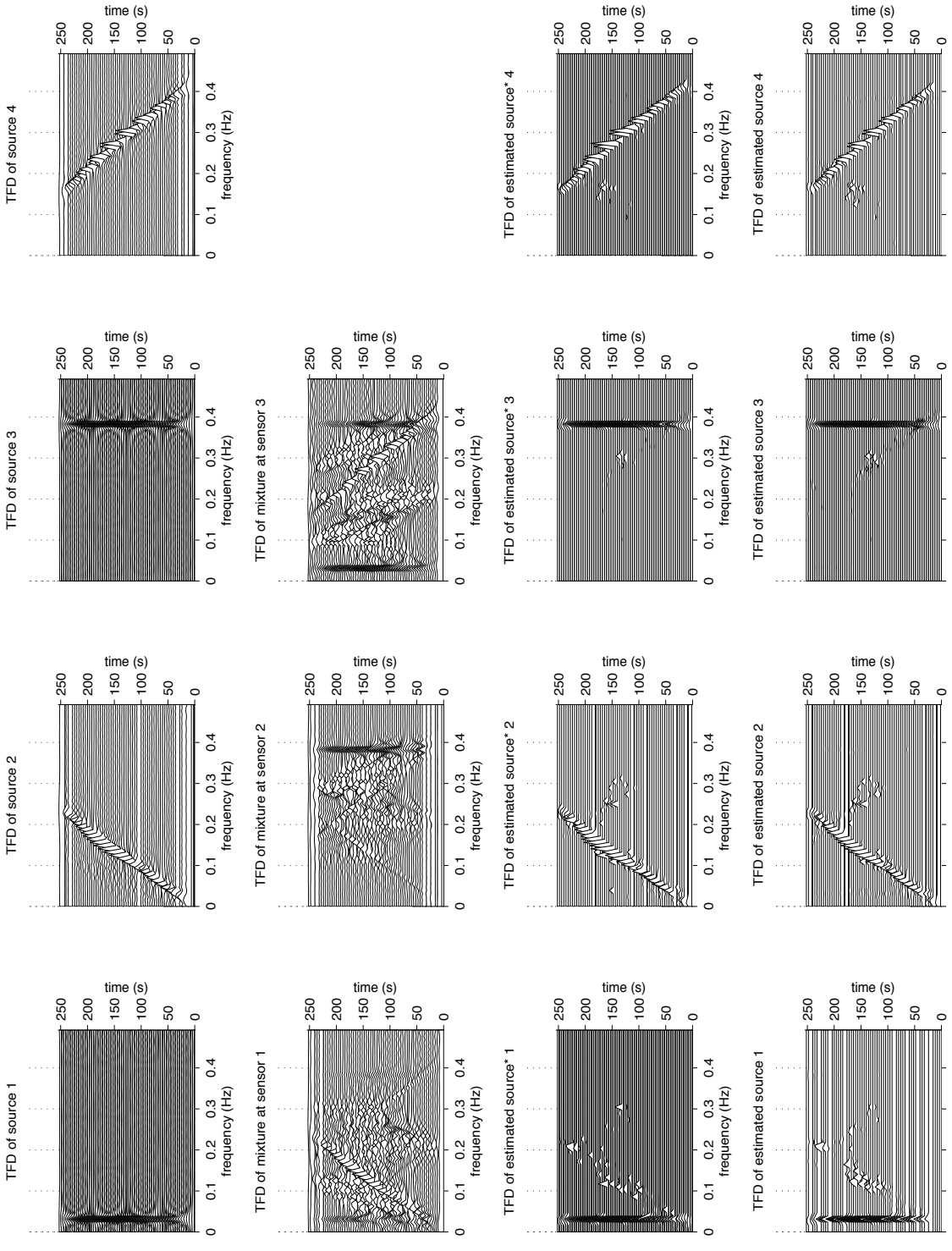

Figure 5.1: TFD for one realization of example 1. The first row is TFDs of the original sources; the second is of the mixtures at each sensor; the third is of the estimated sources by the proposed method and the last is of the estimated sources by existing subspace method

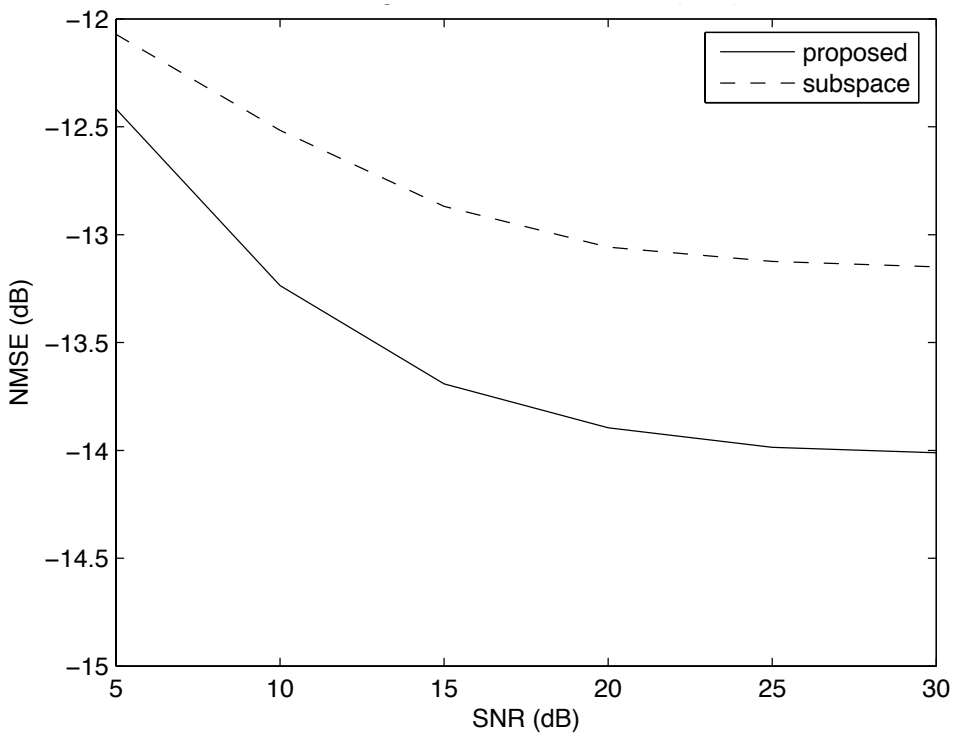

Figure 5.2: NMSE for example 2. All sources are llinear FMs

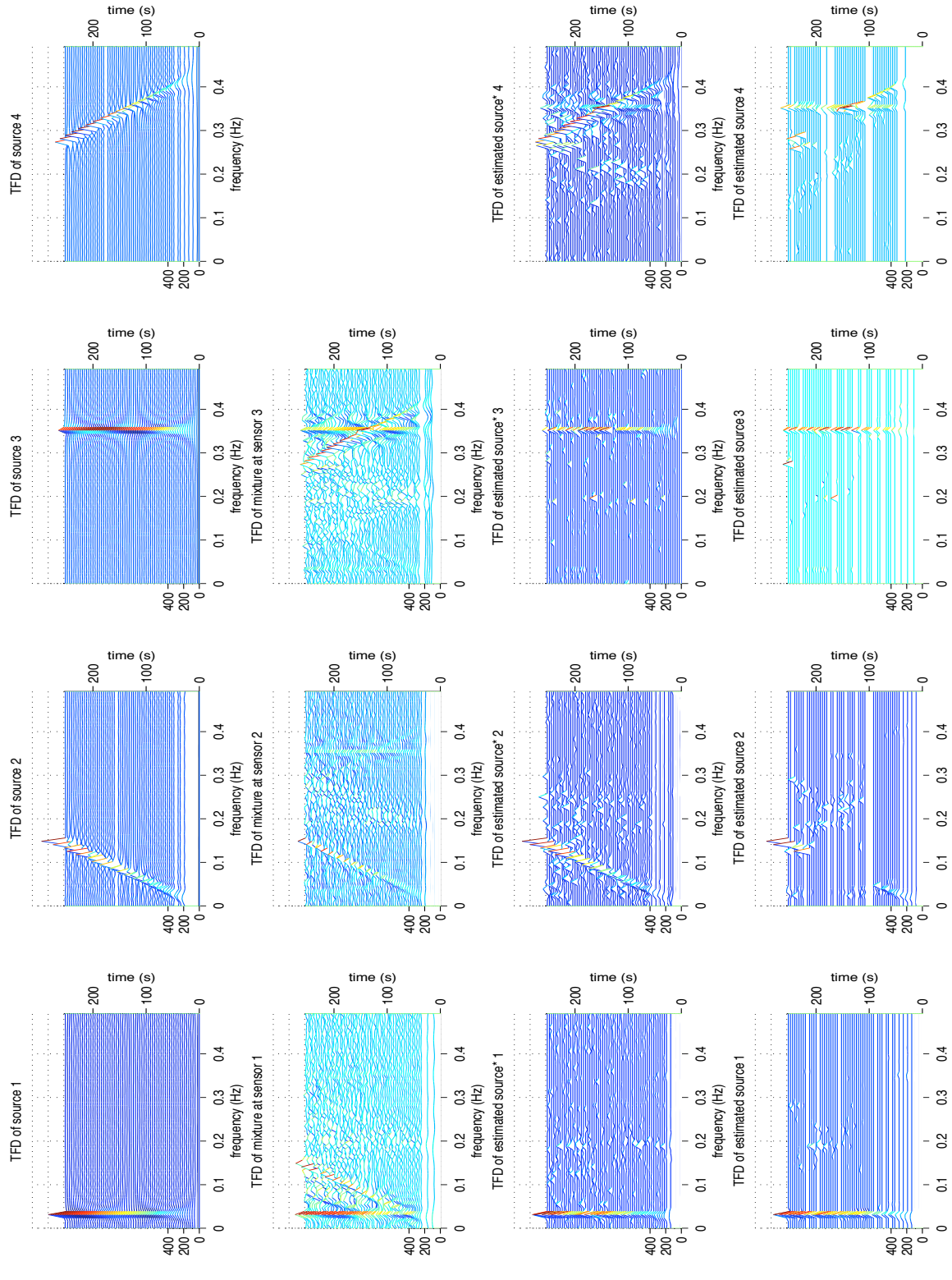

Figure 5.3: TFD for one realization of example 2. The first row is TFDs of the original sources; the second is of the mixtures at each sensor; the third is of the estimated sources by the proposed method and the last is of the estimated sources by existing subspace method

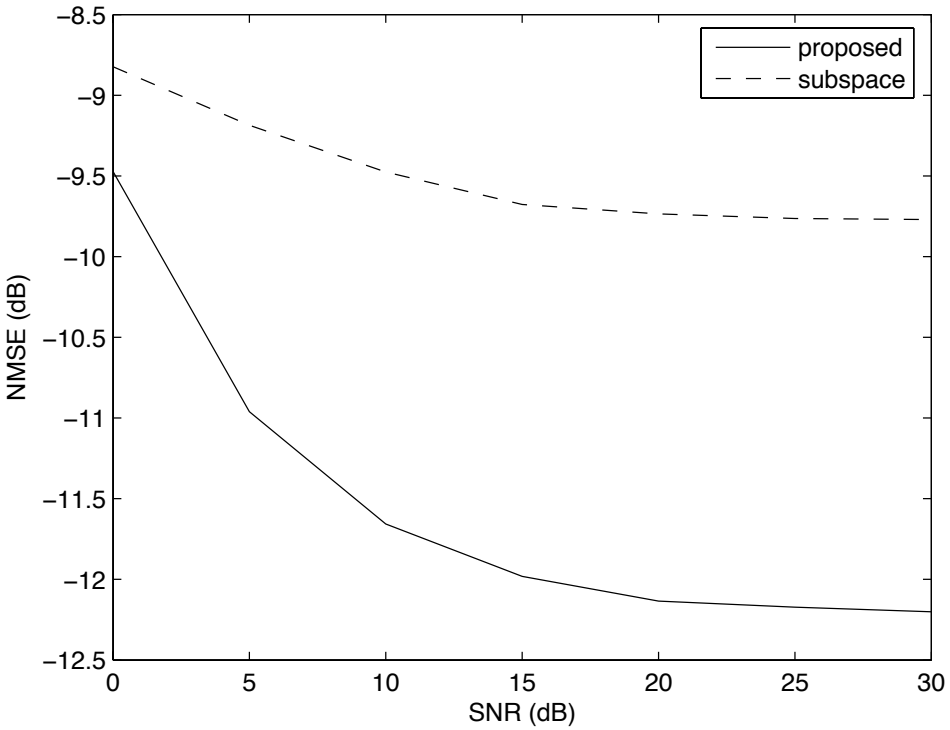

Figure 5.4: NMSE for example 3. Sources are 3 linear FMs and one multicomponent signal

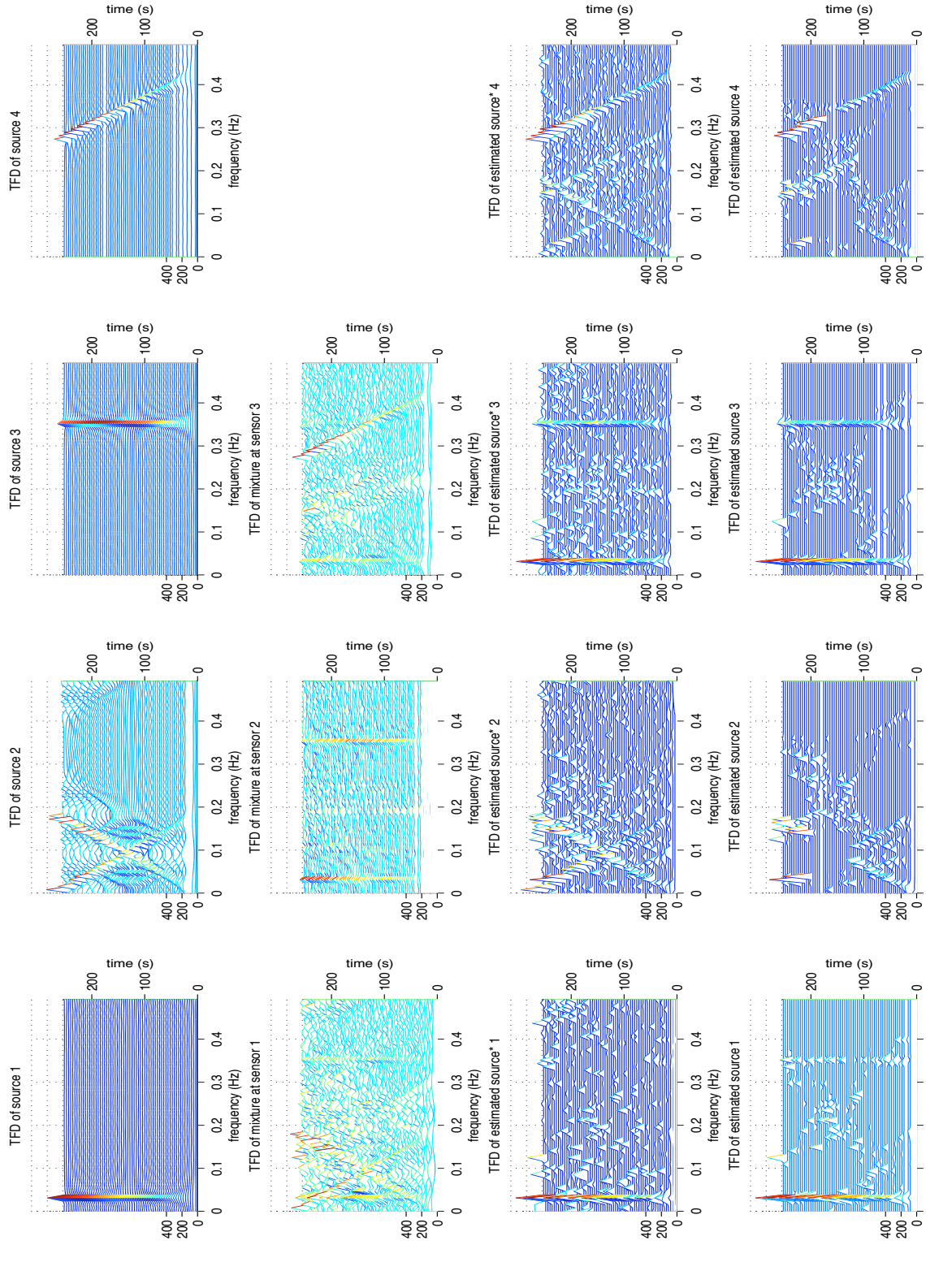

Figure 5.5: TFD for one realization of example 3. The first row are TFDs of the original sources, the second are of the mixtures at each sensor, the third are of the estimated sources by the proposed method and the last are of the estimated sources by existing subspace method

Chapter 6

Higher- and Mixed-Order Statistics based DOA Estimators

6.1 Introduction

In this chapter, we deal with parametric DOA estimation of non-Gaussian signal sources. The main difference between the problem in this chapter and the classical DOA estimation problem is that the noise observed by each sensor is assumed to be unknown Gaussian and spatially correlated or colored. In many applications such as in sonar, radar and seismology, neither the statistics of the signal nor the noise covariance is known [105–108]. Additionally, in many wireless communication applications, the signals usually have non-Gaussian statistics while the noise statistics often remain unknown. Although from central limit theorem, it is reasonable to assume the noise observed by each sensor to be Gaussian distributed, however the noise observed from one sensor to another may be correlated, which means that the noise is non-white.

Fourth-order cumulant statistics have been exploited for non-Gaussian signals in many situations especially in the presence of Gaussian noise with unknown correlation [109]. In this situation, SOS-based DOA estimator fails to perform well. Another key motivation of using FOS is its ability to estimate DOA in a scenario with more sources than sensors. Furthermore, we can observe an increase in the resolution of these estimates by using FOS [100,110]. The existing FO DOA estimators which make use of the full set of FO cumulants, as in [111], are computationally intensive due to the very large set of statistics to be processed. On the other hand, downsizing the FO cumulants to a single contracted quadricovariance matrix or a diagonal slice quadricovariance matrix [112] always leads to poor estimates. In general, contracted quadricovariance can be *steered* to optimally estimate one particular DOA. However, this choice of contracted quadricovariance is not necessarily optimal to estimate other DOAs. In [113], a *steering* technique was proposed by using the inverse of signal covariance matrix. However, the SO covariance of the signal cannot be consistently estimated if the Gaussian noise has a peak power from a certain direction. This can lead to a suboptimal choice of quadricovariance when performing DOA estimation.

In this chapter, we propose a new FO DOA estimator that is based on the multiple contracted quadricovariance matrices i.e. *eigenmatrices*. Following that, we extend the proposed FO DOA estimator to a joint second- and fourth-order DOA estimator, which will compromise the disadvantages of SOS and FOS, and lead to

a more robust algorithm. This is particularly useful in the case when FO DOA estimator alone degrades or performs worse than SO DOA estimator, such as in the situation when the signal statistics are nearly Gaussian and in the situation when the Gaussian noise covariance is known or nearly white [42,114]. This is because in these situations, the estimates of HOS have higher variances than the estimates of SOS. In [115], the concept of exploiting both SOS and FOS is applied to multiuser detection. Similar concept has also been applied to equalization [116], segmentation of textured surfaces [117,118], and blind source separation [119]. In the blind source separation applications [119], the FOS and SOS are weighted optimally to obtain the best estimate of the channel mixing matrix, however, the algorithm requires one column of the non-parametric channel mixing matrix to be known, which might not always available in practice. In [120], root-MUSIC algorithm is used twice, one to the quadricovariance matrix and the other to the covariance matrix, to obtain their roots respectively. These roots are classified and used to determine whether to exploit the DOA estimates from the quadricovariance-based root-MUSIC, or from the covariance-based root-MUSIC, or from both. Hence, the estimates could come from either the SOS-based estimation, the FOS-based estimation, or the average of both estimations. In this chapter, our estimates come from the weighted average of covariance and quadricovariance matrices. It is not simply by choosing one among SOS, FOS and the averaging of both (three discrete choices), but rather a linear combination of SOS and FOS. Finally, simulation results are included to justify our claims.

6.2 Signal Model

Assume an ULA of M sensors and L narrowband signals, with $L < M$. The random circular vector of the sensors output is modeled as

$$\mathbf{x}(n) \triangleq \mathbf{A}\mathbf{s}(n) + \mathbf{v}(n), \qquad n = 1, \ldots, N \tag{6.1}$$

where $\mathbf{A} \triangleq [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_L)]$ is the $M \times L$ array manifold matrix with $\{\theta_i\}$ denotes the DOAs. The steering vector is defined as $\mathbf{a}(\theta) \triangleq [1, e^{j\frac{2\pi d}{\lambda}\sin\theta}, \ldots, e^{j\frac{2\pi d}{\lambda}(L-1)\sin\theta}]^T$. The zero-mean non-Gaussian source signals $\mathbf{s}(n)$ are arranged in a $L \times 1$ vector and $\mathbf{v}(n)$ is the $M \times 1$ complex zero-mean Gaussian noise vector. The variable N is the number of snapshots, d is the inter-element spacing of the array, λ is the signal source wavelength and $(\cdot)^T$ denotes the transpose. The array covariance matrix can therefore be expressed as

$$\mathbf{R} \triangleq E[\mathbf{x}(n)\mathbf{x}^H(n)] = \mathbf{A}\mathbf{R}_\mathbf{s}\mathbf{A}^H + \mathbf{R}_\mathbf{v}, \tag{6.2}$$

where $\mathbf{R}_\mathbf{s} \triangleq E[\mathbf{s}(n)\mathbf{s}^H(n)]$ is the $L \times L$ covariance matrix of the signal sources, $\mathbf{R}_\mathbf{v} \triangleq E[\mathbf{v}(n)\mathbf{v}^H(n)]$ is the $M \times M$ noise covariance matrix, $E[\cdot]$ denotes the expectation operator and $(\cdot)^H$ denotes the Hermitian transpose. The quadricovariance is given

by the set of FO circular cumulants, which are defined as

$$\kappa_{\mathbf{x}}(i,j,k,l) \triangleq Cum(x_i, x_j^*, x_k, x_l^*) \triangleq E[x_i x_j^* x_k x_l^*] - E[x_i x_j^*]E[x_k x_l^*] \\ - E[x_i x_l^*]E[x_k x_j^*] - E[x_i x_k]E[x_l^* x_j^*] \quad \text{for } 1 \leq i,j,k,l \leq M \quad (6.3)$$

where $(\cdot)^*$ denotes the complex conjugate, x_i denotes the i^{th} element of vector $\mathbf{x}$ and the last term vanishes if the random multivariate is circularly complex. The contracted quadricovariance matrix of $\mathbf{x}$ is defined as $\mathbf{Q}_{\mathbf{x}}(\mathbf{M})$ with its $(i,j)^{\text{th}}$ element given by

$$[\mathbf{Q}_{\mathbf{x}}(\mathbf{M})]_{i,j} = \sum_{k,l=1}^{M} \kappa_{\mathbf{x}}(i,j,k,l)\,[\mathbf{M}]_{l,k}\,, \quad (6.4)$$

where $[\mathbf{M}]_{i,j}$ is the $(i,j)^{\text{th}}$ element of the free parameter matrix $\mathbf{M}$. A full quadricovariance is formed by the set of parallel slices of quadricovariance matrices defined by $\mathcal{Q}^P \triangleq \{\mathbf{Q}_{\mathbf{x}}(\mathbf{Z}_{l,m}), 1 \leq l,m \leq M\}$ where $\mathbf{Z}_{l,m}$ is a matrix with zero entries everywhere except the $(l,m)^{\text{th}}$ entry is one.

6.3 Second-Order Estimator

As we have reviewed in Section 2.3, the SO DOA algorithms rely on the covariance matrix $\mathbf{R}$ which can be estimated (consistently) from its samples

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}(n)\mathbf{x}^H(n) \quad (6.5)$$

Its eigen-decomposition is given by

$$\hat{\mathbf{R}} \triangleq \begin{bmatrix} \mathbf{U_s} & \mathbf{U_v} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda_v} \end{bmatrix} \begin{bmatrix} \mathbf{U_s}^H \\ \mathbf{U_v}^H \end{bmatrix} = \mathbf{U_s}\mathbf{\Lambda_s}\mathbf{U_s}^H + \mathbf{U_v}\mathbf{\Lambda_v}\mathbf{U_v}^H \tag{6.6}$$

where matrices $\mathbf{\Lambda_s}, \mathbf{U_s}, \mathbf{\Lambda_v}, \mathbf{U_v}$ contain the eigenvalues and eigenvectors of signal subspace and the eigenvalues and eigenvectors of noise subspace, respectively. Subspace algorithms exploit either the matrix $\mathbf{U_s}$ or $\mathbf{U_v}$ to estimate the DOAs. This is because columns of $\mathbf{U_s}$ and $\mathbf{U_v}$ form orthonormal bases for $span\{\mathbf{A}\}$ and noise subspace, respectively. In this chapter, we use root-MUSIC algorithm because it is a search-free approach. The DOA of ULA structure can be found by computing the L roots which are closest to the unit circle of the following polynomial,

$$f(z) = \mathbf{a}^T(1/z)\mathbf{U_v}\mathbf{U_v}^H\mathbf{a}(z) \tag{6.7}$$

which is the denominator of the search-based MUSIC spectrum given in Eqn. (2.51) with $z = e^{j\frac{2\pi d}{\lambda}\sin\theta}$. This algorithm performs well in the situation where either white or known colored noise is present. However, in the situation when the noise is unknown colored Gaussian noise, the FO cumulant-based DOA algorithms will perform better. In the following section, we propose a new FO cumulant-based DOA algorithm.

6.4 Proposed Fourth-Order DOA Estimator

The FO DOA algorithms hinge on the quadricovariance statistics which are estimated consistently by their sample estimates as follows,

$$\hat{\kappa}_{\mathbf{x}}(i,j,k,l) = \hat{\mu}_{ijkl} - \hat{\sigma}_{ij}\hat{\sigma}_{kl} - \hat{\sigma}_{il}\hat{\sigma}_{kj} - \hat{\zeta}_{ik}\hat{\zeta}^*_{lj} \tag{6.8}$$

where $\hat{\mu}_{ijkl} \triangleq \frac{1}{N}\sum_{n=1}^{N} x_i(n)x_j^*(n)x_k(n)x_l^*(n)$, $\hat{\sigma}_{ij} \triangleq \frac{1}{N}\sum_{n=1}^{N} x_i(n)x_j^*(n)$ and $\hat{\zeta}_{ij} \triangleq \frac{1}{N}\sum_{n=1}^{N} x_i(n)x_j(n)$. For clarity, we have ignored the noise, which is a reasonable assumption because FO cumulant statistics are insensitive to Gaussian noise. The contracted quadricovariance of $\mathbf{x}$ defined in Eqn. (6.4) is related to the quadricovariance of $\mathbf{s}$ (after Gaussian noise rejection) by

$$[\mathbf{Q}_{\mathbf{x}}(\mathbf{M})]_{i,j} = \sum_{k,l,\alpha,\beta,\gamma,\delta} \kappa_{\mathbf{s}}(\alpha,\beta,\gamma,\delta) a_{i,\alpha} a^*_{j,\beta} a_{k,\gamma} a^*_{l,\delta} [\mathbf{M}]_{l,k}$$

with $a_{i,j} = [\mathbf{A}]_{i,j}$. The equation above could be written in matrix form as

$$\mathbf{Q}_{\mathbf{x}}(\mathbf{M}) = \mathbf{A}\mathbf{W}\mathbf{A}^H \tag{6.9}$$

where $[\mathbf{W}]_{i,j} \triangleq \sum_{k,l} \kappa_{\mathbf{s}}(i,j,k,l)\mathbf{a}^H(\theta_l)\mathbf{M}\mathbf{a}(\theta_k)$. This is of the same form as in the SO covariance matrix case, except that $\mathbf{W}$ might not necessarily be Hermitian symmetric. In general, because of this Hermitian property, the covariance matrix possesses symmetric eigen-decomposition, which gives unitary eigenvectors that

are used by many subspace-based DOA estimation methods. Furthermore, due to positive definiteness, the covariance matrix possesses positive real eigenvalues, which can be used to determine the dominant signal subspace in the presence of noise. However, the contracted quadricovariance, $\mathbf{W}$, is Hermitian *if and only if* $\mathbf{M}$ is Hermitian, which is due to the super-symmetry property of the FO *circular cumulant*. Hence, if $\mathbf{M}$ is Hermitian, then $\mathbf{Q}_{\mathbf{x}}(\mathbf{M})$ will have real eigenvalues which may be either positive or negative. However, for a weaker condition of $\mathbf{Q}_{\mathbf{x}}(\mathbf{M})$ that does not possess the Hermitian symmetry, it would then be a *normal* matrix (i.e. $\mathbf{B}$ is normal matrix when $\mathbf{B}\mathbf{B}^H = \mathbf{B}^H\mathbf{B}$). This property allows $\mathbf{Q}_{\mathbf{x}}(\mathbf{M})$ to possess unitary eigenvectors with possibly complex eigenvalues.

Contracted quadricovariance is a method of downsizing the full quadricovariance to $M \times M$ matrix with $\mathbf{M}$ as the free parameter matrix. For example, the contracted quadricovariance $\mathbf{Q}_{\mathbf{x}}(\mathbf{I})$ is used for DOA estimations in [112] by choosing $\mathbf{M} = \mathbf{I}$. Another method of downsizing the quadricovariance is by *diagonal slicing*. The subspace algorithms based on the diagonal slice of quadricovariance perform worse than the subspace algorithms based on $\mathbf{Q}_{\mathbf{x}}(\mathbf{I})$. In spite of that, choosing $\mathbf{M} = \mathbf{I}$ does not exploit the freedom to choose the suitable $\mathbf{M}$'s which can be derived from *eigenmatrices* using the received signals. We will explain this shortly.

Eigenmatrices of FO tensor were introduced in [113]. Since the circular cumulant $\kappa_{\mathbf{x}}(i,j,k,l)$ acts as a linear operator $\mathbf{Q}_{\mathbf{x}}(\cdot) : \mathbb{C}^{M\times M} \mapsto \mathbb{C}^{M\times M}$, which is a

compact[†] and self-adjoint[‡] operator with inner-product $\langle \mathbf{A}, \mathbf{B} \rangle = trace\{\mathbf{A}\mathbf{B}^H\}$, it then takes the following eigen-structure,

$$\kappa_{\mathbf{x}}(i,j,k,l) = \sum_{r=1}^{M^2} \lambda_r [\mathbf{E}^{(r)}]_{i,j} [\mathbf{E}^{(r)}]^*_{l,k} \tag{6.10}$$

where λ_r is the eigenvalue and $\mathbf{E}^{(r)}$ is the eigenmatrix pair. The eigenmatrices are orthonormal, i.e. $trace\{\mathbf{E}^{(r)}\mathbf{E}^{(s)H}\} = \delta_{r,s}$ and invariant, i.e. $\mathbf{Q}(\mathbf{E}^{(r)}) = \lambda_r \mathbf{E}^{(r)}$. In general the eigenmatrices span $\mathbb{C}^{M \times M}$. However, assuming no noise and the steering vectors of the signal are linearly independent, the dimension of the subspace spanned by the eigenmatrices is L^2, which could be much smaller than M^2 [113]. In the special case of statistically independent sources, the dimension of the subspace spanned by eigenmatrices is only L. Hence, the dimension of this subspace is in $[L, L^2]$. The subspace spanned by these eigenmatrices is called FO signal subspace, while the orthogonal subspace from the signal subspace is called FO noise subspace, though the noise is asymptotically zero. This leads to the MUSIC-like algorithm in tensor domain [111, 113]. This type of processing is computationally expensive because it is search-based, which requires large computational power even for single search due to the large matrix multiplication involved in each search.

This chapter utilizes the tensor eigen-decomposition in a different way. The eigenmatrices are also the contracted quadricovariances of the form of Eqn. (6.9).

[†]because the range of the operator is finite dimension space

[‡]self-adjoint if the operator satisfies $\langle \mathbf{Q}(\mathbf{A}), \mathbf{B} \rangle = \langle \mathbf{Q}(\mathbf{B}), \mathbf{A} \rangle, \forall (\mathbf{A}, \mathbf{B})$, this is satisfied by super-symmetry property of the circular cumulant

This suggests us to find an approach to estimate the common subspace, $span\{\mathbf{A}\}$, or its bases from the eigenmatrices. Following that, the DOAs can be extracted in the similar fashion as the SO subspace algorithms. With $\mathcal{Q}^e \triangleq \{\lambda_r \mathbf{E}^{(r)}, 1 \leq r \leq L^2\}$, we only need to find the common subspace from L^2 matrices (or L matrices if sources are known to be statistically independent) instead of M^2 matrices from $\mathcal{Q}^P$. Finding orthonormal bases of $span\{\mathbf{A}\}$ could be made through performing simultaneous eigen-decomposition of

$$\mathbf{Q}(\mathbf{E}^{(r)}) = \lambda_r \mathbf{E}^{(r)} = \mathbf{A}\mathbf{W}^{(r)}\mathbf{A}^H = \mathbf{U}_s \mathbf{\Gamma}^{(r)} \mathbf{U}_s^H, \;\; 1 \leq r \leq L^2 \tag{6.11}$$

where $\mathbf{U}_s$ is a $M \times L$ matrix that contains the eigenvectors of signal subspace and $\mathbf{\Gamma}^{(r)}$ is a $L \times L$ diagonal matrix that contains the eigenvalues of the signal subspace of the r^{th} eigenmatrix. This eigen-decomposition with unitary eigenvector is guaranteed by the normality of the eigenmatrices. In fact, because of the horizontal symmetry of FO circular cumulants, the eigenmatrices could be chosen to be Hermitian [113], but the eigenmatrices are unique up to a phase factor. Suppose $\mathbf{E}^{(r)}$ is Hermitian then there exists an $\mathbf{E}'^{(r)}$ that is not Hermitian such that $\mathbf{E}'^{(r)} = e^{j\phi}\mathbf{E}^{(r)}$. However, if $\mathbf{E}^{(r)}$ is Hermitian, then we have $\mathbf{E}'^{(r)}\mathbf{E}'^{(r)H} = \mathbf{E}'^{(r)H}\mathbf{E}'^{(r)}$. This guarantees that eigenmatrices obtained through Eqn. (6.10) are normal.

In practice, the simultaneous eigen-decomposition is performed by the joint approximate diagonalization of eigenmatrices (JADE) algorithm [26], which has been successfully exploited for blind source separation. The eigenvectors of signal sub-

space, $\mathbf{U}_s$, is determined by a slightly different manner from the SO counterpart. Assuming the full eigenvectors as $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M]$, then

$$\mathbf{U}_s = [\mathbf{u}_{s_1}, \ldots, \mathbf{u}_{s_L}] \tag{6.12}$$

In Eqn. (6.12), $\{s_1, \ldots, s_L\}$ is taken such that $\gamma_{s_1} \geq \gamma_{s_2} \geq \ldots \geq \gamma_{s_L} \geq \ldots \geq \gamma_{s_M}$, where $\gamma_i \triangleq \sum_r |[\mathbf{\Gamma}^{(r)}]_{i,i}|$. The sum of the magnitude of the eigenvalues across different eigenmatrices is used because the eigenvalues are complex numbers. It can be shown that the magnitudes of the eigenvalues of a *normal* eigenmatrix are the same as the eigenvalues of the its equivalent Hermitian eigenmatrix, except for a difference in sign. After obtaining $\mathbf{U}_s$, DOAs could be estimated, in general, regardless of the structure of the array, by many existing subspace algorithms. Particularly, in this chapter root-MUSIC in (6.7) is used (assuming array structure is ULA).

6.5 Joint Second- and Fourth-Order DOA Estimator

Observing the structure in Eqn. (6.6) and Eqn. (6.11), a new estimator based on SOS and FOS could be derived by simultaneously performing eigen-decomposition of the eigenmatrices and autocovariance matrix, i.e. $\mathcal{Q}^{FSO} \triangleq \{\mathbf{R}, \lambda_1\mathbf{E}^{(1)}, \ldots, \lambda_{L^2}\mathbf{E}^{(L^2)}\}$.

In fact, as we will see later in the cost function of the JADE algorithm, it is possible to weigh the matrices, and hence, it is possible to ascertain whether can SOS be used more predominantly than FOS or vice versa, depending on whether the noise is white or colored Gaussian noise.

Assume joint diagonalization algorithm for 2×2 matrices (it can be easily extended to square matrices of any size, see Appendix E for the complete unweighted algorithm), for example, matrices of set $\mathcal{Q}^{uw} = \{\mathbf{G}^{(r)}, 1 \leq r \leq L^2\}$ with the entries,

$$\mathbf{G}^{(r)} = \begin{bmatrix} a_r & b_r \\ c_r & d_r \end{bmatrix} \tag{6.13}$$

The objective of joint diagonalization is to get the unitary matrix $\mathbf{V}$ such that matrices $\mathbf{G}'^{(r)} = \mathbf{V}^H \mathbf{G}^{(r)} \mathbf{V}$ are as "diagonal" as possible. Mathematically, it is the same as maximizing $\mathcal{C}$, where

$$\mathcal{C} \triangleq \sum_r |a'_r|^2 + |d'_r|^2 \overset{1}{=} \frac{1}{2} \sum_r |a'_r - d'_r|^2 \tag{6.14}$$

where a'_r, d'_r are the diagonal elements of $\mathbf{G}'^{(r)}$. Equality 1 above is due to the invariance in the trace of $\mathbf{G}'^{(r)}$. The joint diagonalization uses complex Givens rotation technique,

$$\mathbf{V} = \begin{bmatrix} \cos\theta & -e^{j\phi}\sin\theta \\ e^{-j\phi}\sin\theta & \cos\theta \end{bmatrix}$$

and hence $\mathcal{C}$ relates the new parameters to the earlier ones by,

$$a'_r - d'_r = (a_r - d_r)\cos 2\theta + (b_r + c_r)\sin 2\theta \cos\phi + j(c_r - b_r)\sin 2\theta \sin\phi \quad (6.15)$$

From Eqns.(6.14) and (6.15), weighting each diagonalization of $\mathbf{G}_r$ by w_r could be done by multiplying it with w_r such that the weighted joint diagonalization algorithm is just a joint diagonalization on set $\mathcal{Q}^w = \{w_r \mathbf{G}^{(r)}, 1 \leq r \leq L^2\}$. Table. 6.1 enumerates the steps of the proposed algorithms, i.e., the new fourth-order (NFO) statistics-based algorithm and the joint fourth- and second-order (FSO) statistics-based algorithm.

Table 6.1: Summary of the new fourth-order (NFO) and mixed fourth- and second-order (FSO) algorithms steps

Given sensors output $\mathbf{x}(n)$

1. Estimate $\hat{\mathbf{R}}$ and $\hat{\kappa}(i,j,k,l)$ as in (6.5) and (6.8)
2. Compute L or L^2 largest eigen-value and -matrices of $\hat{\kappa}(i,j,k,l)$ as in [26]
3. Perform JADE [26] on matrices in:
   (a) $\mathcal{Q}^e$ for NFO algorithm
   (b) $\mathcal{Q}^{FSO}$ with weight for FSO algorithm
4. Determine $\mathbf{U_s}$ from (6.12), and then perform the subspace-based DOA estimation on $\mathbf{U_s}$, such as root-MUSIC

6.6 Simulation Results

In this section, two simulations are provided to illustrate the effectiveness of the NFO algorithm and the FSO algorithm. For both simulations, root-MUSIC algorithm is used on the estimated subspace and the ULA of 13 sensors with $\lambda/2$ spacing is employed. It is assumed that two sources are emitting from directions $-25^o, 17^o$ normal to broadside of antenna and the number of samples collected is 1000. Weights for the FSO algorithm are chosen to be 1 for covariance matrix and 10^{-3} for eigenmatrices such that JADE is performed on $\mathcal{Q}^{FSO} = \{\mathbf{R}, 10^{-3}\lambda_1\mathbf{E}^{(1)}, \ldots, 10^{-3}\lambda_{L^2}\mathbf{E}^{(L^2)}\}$.

For the first simulation, signals emitted by the non-Gaussian sources are modeled as

$$\begin{aligned} \mathbf{s}(n) &= \text{diag}\{f_1(n), \ldots, f_L(n)\}\mathbf{r}(n) \\ \mathbf{r}(n) &= [r_1(n), \ldots, r_L(n)]^T \end{aligned}$$

The zero-mean Gaussian processes $f_i(n)$ and $r_i(n)$ have unit-variance and σ_s^2-variance, respectively. This type of non-Gaussian signal is common in sonar and other applications [107]. Note that the source signals are independent and have equal power in the first simulation. The unknown colored Gaussian additive noise is assumed to have covariance matrix $[\mathbf{R}_\mathbf{v}]_{lk} = \sigma^2\rho^{|l-k|}$, with $\rho = 0.9$. The noise is further assumed to be temporally correlated as, $E\{\mathbf{v}_i(n)\mathbf{v}_i(n+m)\} = \sigma\tilde{\rho}^m$ with

$\tilde{\rho} = 0.5$. The SNR is defined as $SNR \triangleq 10\log(\sigma_s^2/\sigma^2)\, dB$ which is varied from $-10\, dB$ to $15\, dB$. The simulation results in Fig. 6.1 show that at lower SNR, the NFO algorithm and FSO algorithm perform better than both the SO root-MUSIC algorithm and the fourth-order contracted quadricovariance (FO QC)-based root-MUSIC algorithm. The FO QC algorithm has been previously described in Section 6.4, which is based on evaluating the root-MUSIC of $\mathbf{Q}_{\mathbf{x}}(\mathbf{I})$ [112]. The performance of the SO algorithm is severely affected in low SNR by the colored Gaussian noise while all FO cumulant statistics based algorithms are not significantly affected by the noise. The FSO algorithm is not affected because the weights allow the inference drawn from covariance to be minimized, yet still useful. The FO QC algorithm performed worse than the NFO and FSO algorithm because it relies on a single slice of the quadricovariance matrix which might not necessary be the best slice chosen. Note that only RMSE of one DOA is shown and the RMSE of the other DOA is similar to what is shown here.

For the second simulation, signals emitted by the sources are modeled similarly as before, except that the sources are correlated with correlation coefficient of 0.4. The noise is also assumed to be the same as in the previous simulation, except that the SNR is now fixed at $0\, dB$, while the correlation coefficient, ρ, is varied from 0 to 1. The results in Fig. 6.2 show the robustness of the proposed FSO algorithm with only slight performance degradation in the region of $\rho < 0.6$ as compared to the SO algorithm. The NFO and FO QC algorithms are performing reasonably

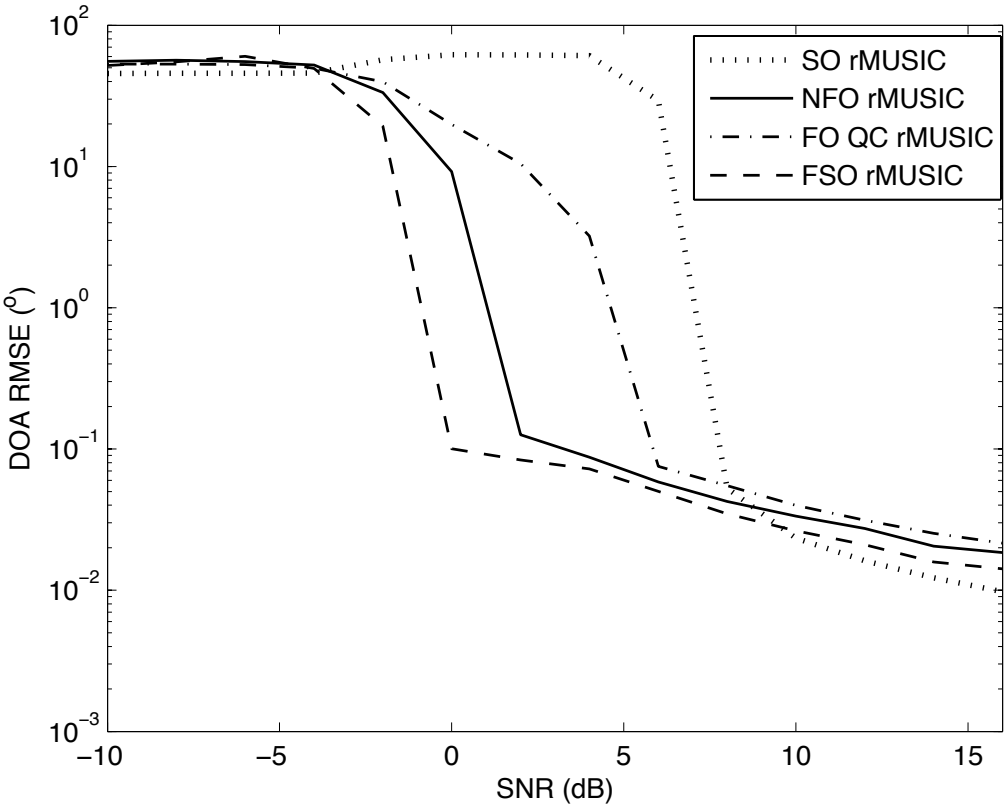

Figure 6.1: DOA estimation RMSE's vs. SNR for two independent sources.

well for higher ρ values, because they are using FOS in the presence of unknown correlated Gaussian noise. Nevertheless, NFO still performs better than FO QC.

6.7 Discussion

In this chapter, we have proposed two algorithms for DOA estimations in the presence of unknown Gaussian noise. The SO root-MUSIC, which is used in the comparison study, is search-free and requires $\mathcal{O}(M^3)$ computational complexity. If $M < L^2$, the computational complexity of the SO root-MUSIC is smaller than

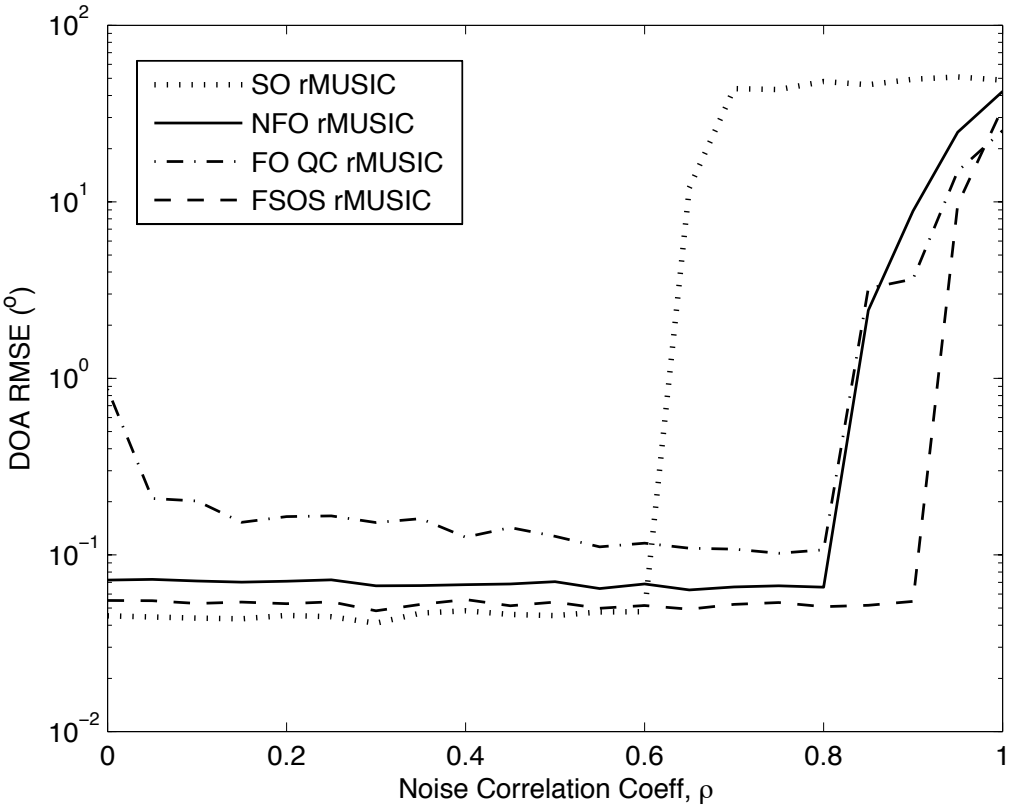

Figure 6.2: DOA estimation RMSE's vs. spatial correlation coefficient of noise

the computational complexity of our proposed methods, i.e. $\mathcal{O}(L^2M^2)$, which are also search-free. Whenever robustness against unknown Gaussian noise is required, FO based algorithms would be preferred at the expense of having additional computational complexity. Note that our proposed FO and FSO algorithms are search-free, hence, they are computationally less demanding than FO search-based algorithms [111,113], which require $\mathcal{O}(L^2M^2)$ computational load for every search. Therefore, for N_{search}, they require $\mathcal{O}(N_{\text{search}}L^2M^2)$ computational load.

We have shown that the application of (weighted) joint approximate diagonalization applied directly to FOS or both FOS and SOS could be utilized for

estimating DOAs. The performance of the proposed algorithms, especially the FSO algorithm, are very promising. In the development of FSO, it is discovered that different weightings could improve the performance of the algorithm. We will discuss the optimum weights and its analysis in our future work. Another possible extension of this weighting is to weigh various autocorrelation matrices $\{\mathbf{R}(\tau),\ \tau = 0, \ldots, L^2\}$, as in SO blind identification (SOBI) algorithm [27], and various eigenmatrices (quadricovariance matrices).

Albeit [121] proposed a weighted joint diagonalization in LS formulation, the weighting in this chapter is different from [121]. Suppose that $\mathbf{H}$ is any matrix, the weighting in [121], $\mathcal{W}$, is based on weighted LS with norm, $\|\mathbf{H}\|_{\mathcal{W}} \triangleq \sqrt{\text{vec}\{\mathbf{H}\}^H \mathcal{W} \text{vec}\{\mathbf{H}\}}$. In this chapter, it is based on the weighted LS with norm, $\|\mathbf{H}\|_{\mathsf{W}} \triangleq \sqrt{tr\{\mathbf{H}^H \mathsf{W} \mathbf{H}\}}$ and W is restricted to be a diagonal matrix with w_r as the diagonal entries.

Generally, the SOS subspace algorithms require the number of sources to be less than the number of sensors. Therefore, because the proposed NFO and FSO algorithms, as well as the existing contracted quadricovariance algorithms, use the quadricovariance matrices which are analogous to the structure of SOS's autocorrelation (see Eqn. (6.9)), the proposed algorithms may not work properly when there are more sources than sensors. The reason is that the SOS autocorrelation matrix and contracted quadricovariance matrices of the source signals are full-rank and generally non-diagonal. Therefore, the noise subspace cannot be estimated and

subsequently, DOA also cannot be estimated. However, if the sources are independent, which make $\mathbf{W}$ becomes diagonal, estimating $\mathbf{A}$ for more sources than sensors using the proposed FO and FSO algorithm is possible by utilizing the diagonalization algorithms given in [101,121]. These algorithms formulate and solve for mixing matrix $\mathbf{A}$ through minimizing the LS error, hence solving wide matrix $\mathbf{A}$ is still possible. Assuming source signals are independent, there are also other blind identification methods which can be used to estimate the wide-matrix $\mathbf{A}$, such as those reported in [90, 122, 123]. Following that, by techniques similar to MUSIC or weighted subspace fitting algorithms, DOAs can be elicited from the estimate of $\mathbf{A}$, i.e., $\hat{\mathbf{A}}$. For example, using MUSIC-like algorithm given by

$$\hat{\theta}_i = \arg\max_{\theta_i} \frac{1}{\tilde{\mathbf{a}}(\theta_i)^H(\mathbf{I} - \mathbf{P}_{\hat{\mathbf{A}}^* \odot \hat{\mathbf{A}}})\tilde{\mathbf{a}}(\theta_i)} \tag{6.16}$$

where $\tilde{\mathbf{a}}(\theta_i) = \mathbf{a}^*(\theta_i) \odot \mathbf{a}(\theta_i)$.

6.8 Summary

In this chapter, two new algorithms for DOA estimation that are robust to any unknown Gaussian noise have been proposed. Both algorithms exploit the insensitivity of the FOS to the Gaussian noise. Compared to existing algorithms which exploit FOS through a contracted quadricovarince matrix, our algorithms perform better, because we exploit more than one quadricovariance matrices. The

eigenmatrices we use, parsimoniously capture all the FO statistics of the observations that have significant power, which is analogous to low-rank approximation of autocorrelation matrix.

Chapter 7

Conclusions and Future Works

7.1 Conclusions

In this thesis, we have presented a variety of novel array processing algorithms to handle three problems. The approaches are mainly based on higher-order statistics and time-frequency distributions.

In Chapter 3, we considered the case where wideband LFM signals from multiple sources impinging on sensor arrays. The objective is to estimate DOAs, frequency rates and frequencies for each of the sources. The proposed solution is devised by using the DPT. The properties that received signal frequencies vary linearly and frequency rates remain constant across the antenna sensors, are exploited to estimate their DOAs, frequency rates and initial frequencies. The proposed algorithm is efficient because it does not require a multidimensional search as compared to the ML estimation method and CBF. Furthermore, it is search-free by incorporating root-MUSIC algorithm, except for the part to estimate DOAs

which uses the variant of CBF, and only one-dimensional search is needed instead of three-dimensional search as in the original CBF. From the presented simulation results, the algorithm essentially can perform near to the Cramer-Rao Bound.

In Chapter 4, the SHIM operator is formulated and derived from HIM. HIM is also using HOS, because it is basically a product between more than or equal to two realizations each of different delays. The non-linear SHIM operator transforms a wideband PPS of any order that impinged on an antenna array into a narrowband sinusoidal signal that impinged on an antenna array, which then can be processed directly by many classical array processing and harmonic retrieval algorithms. With the harmonic retrieval algorithm, one can estimate the highest-order frequency rate, followed by using array processing algorithm to estimate the spatial parameter. Classical array processing and harmonic retrieval algorithms are the same except for the difference in rearranging the data before processing. However, we opted to use a closed-form subspace algorithm that simultaneously estimates the highest-order frequency parameter and DOA because it is a search-free algorithm and performs joint estimation in one step. We also examined the theoretical performance and the identifiability of the algorithm. It was discovered that lower sampling rate would deliver better performance in DOA estimation, provided that the identifiability conditions are satisfied.

In Chapter 5, we turned to the problem in separating/recovering more time-frequency signal sources received than receiving antenna sensors. Note that a PPS

is subset of the time-frequency signals. Assuming that the mixing matrix is non-parametric, we could apply the solution to this type of problem to the problem of recovering signals either when the sensor array is poorly calibrated or when the sensor array's structure/geometry is unknown. The structures of STFD at three different TF locations, i.e. SAPs, MAPs and CPs, are analyzed and exploited. The structures of STFDs at SAPs were exploited to form a computationally efficient algorithm and to formulate a detection algorithm to separate SAPs from MAPs and CPs. Following that, the existing subspace method, which is meant for SAPs and MAPs [99], was analyzed for its property at CPs. With a newly discovered property, the abundance of CPs in WVD-based STFD, and the simplicity of the proposed technique for SAPs, we devised a superior algorithm. While keeping the computational speed the same as the existing subspace algorithm, the proposed algorithm outperforms the existing algorithms especially when there is a multicomponent source. The existing algorithms assumed that all of the CTs are useless interfering source signals, and hence they are suppressed. However, we have selectively taken CTs that belong to multicomponent source into account through the newly discovered property of subspace algorithm at CPs, as presented in Chapter 5.

Finally in Chapter 6, we developed a new FOS-based and a mixed SOS-FOS-based DOA estimation algorithm. The FOS is advantageous because it is insensitive towards any Gaussian noise and applicable to the situation with more sources

than antenna sensors. However, many of the FOS-based DOA estimation algorithms are either too computationally expensive or performing too badly. When one uses all of the FOS, it gets computationally expensive. On the other hand, some FOS-based algorithms perform poorly because they only use a smaller set of FOS, known as the quadricovariance matrix. There are useful FOS that are not contained in the quadricovariance, and are not being utilized. In Chapter 6, we proposed a better way to choose FOS, i.e. through eigenmatrix decomposition, which gives multiple quadricovariance matrices. This method parsimoniously captures all the FO statistics that have significant energy, in the analogous way that low-rank approximation of autocorrelation matrix parsimoniously captures the statistics of the signals. Therefore, the resulting new FOS-based DOA algorithm is not too computationally expensive and yet the set of the FOS chosen is not too small. Following that, because of the similar structure between quadricovariance of FOS and covariance of SOS, we developed a mixed-order DOA estimator in addition to the FOS-based approach. This mixed-order algorithm indeed inherits the advantages of both FOS and SOS. We showed using simulations that this algorithm performs well in spatially white Gaussian noise environment as well as in spatially colored Gaussian noise environment. Although FOS-based algorithms work well in spatially white Gaussian environment, SOS-based algorithms are normally superior to FOS counterpart. The mixed-order algorithm, on the other hand, outperforms FOS-based algorithm in spatially white Gaussian noise environment. In Chapter 6, we also introduced weighting to weigh the quadricovariance and co-

variance matrices. This framework allows one to adjust the statistics information according to the level of significance between the SOS and FOS.

7.2 Future Works

From the comparison of the work in Chapter 3 and Chapter 4, we see that parameter estimation of multiple PPS (with order larger than two) impinging on sensor array from multiple sources has not been completely solved. The hindrance of using SHIM for multiple sources is the cross-terms generated. Further investigation is needed to understand: whether the estimation of highest-order frequency parameters alone from SHIM is always possible even in the presence of cross-terms; whether the estimation of DOAs alone from SHIM is possible. If one of these is possible, we could devise an algorithm that can estimate the highest-order frequency parameters and DOAs, of course additional processing may incur.

The work in Chapter 5 has assumed that the estimates of a wide mixing matrix is available. New algorithm can be developed to estimate the mixing matrix. In addition to that, since we only utilized quadratic TF distribution, we can explore the application of the linear TF distributions or other transformations, such as fractional Fourier transform. Linear transformation is appealing because naturally there is no cross-term interference.

In Chapter 6, the performance analysis and optimal weights were not provided.

The future work is to analyze the performance of the NFO and FSO algorithms, which will involve sixth- and eighth-order statistics. With this analysis, we could find the optimal weights, which will allow us to achieve the optimum performance regardless of whether the noise is spatially white. The optimal performance, regardless of whether the Gaussian is colored, is possible if the optimal weights rely on the data or on the statistics that could be estimated directly from the data [33], such as sample cumulants of order sixth. With these optimum weights, there is a possibility to devise mixed FO-SO blind source separation algorithm, which is expected to be more robust in unknown Gaussian noise.

Bibliography

[1] X. Zhuge, J. M. Centrella, and S. L. W. McMillan, "Gravitational radiation from coalescing binary neutron stars," *Physical Review D*, vol. 50, no. 10, pp. 6247–6261, Nov. 1994.

[2] M. I. Skolnik, *Radar Handbook*. McGraw-Hill, Jan. 1990.

[3] S. G. Schock and L. R. LeBlanc, "Some applications of the chirp sonar," in *OCEANS '90*, Sept. 1990, pp. 69–75.

[4] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, July 1996.

[5] S. Haykin, Ed., *Unsupervised Adaptive Filtering Vol I: Blind Source Seperation*. John Wiley & Sons, Inc., 2000.

[6] S. Haykin and B. V. Veen, *Signals and Systems*. John Wiley & Sons, Inc., 1999.

[7] L. Cohen, *Time-Frequency Analysis*. Prentice-Hall Inc., 1995.

[8] ——, "Generalized phase-space distribution functions," *J. Math. Phys.*, vol. 7, no. 5, pp. 781–786, 1966.

[9] ——, "Time-frequency distributions: A review," in *Proc. IEEE*, vol. 77, no. 7, July 1989, pp. 941–981.

[10] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, 1999.

[11] H. I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. on Signal Processing*, vol. 37, no. 6, pp. 862–871, June 1989.

[12] H. Margenau and R. N. Hill, "Correlation between measurements in quantum theory," in *Quantum Theory and Reality*, M. Bunge, Ed. Springer, 1967.

[13] Y. Zhao, L. E. Atlas, and R. J. Marks, "The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 1084–1091, 1990.

[14] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation: optimal kernel design," *IEEE Trans. on Signal Processing*, vol. 41, pp. 1589–1602, 1993.

[15] L. Stankovic, "A method for time-frequency analysis," *IEEE Trans. on Signal Processing*, vol. 42, no. 1, pp. 225–229, Jan. 1994.

[16] B. Porat and B. Friedlander, "Asymptotic statistical analysis of the high-order ambiguity function for parameter estimation of polynomial-phase signals," *IEEE Trans. on Info. Theory*, vol. 42, no. 3, pp. 995–1001, May 1996.

[17] Y. Wang and G. Zhou, "On the use of higher order ambiguity function for multi-component polynomial phase signals," *Signal Processing*, vol. 65, pp. 283–296, Mar. 1998.

[18] S. Peleg and B. Friedlander, "The discrete polynomial-phase transform," *IEEE Trans. on Signal Processing*, vol. 43, no. 8, pp. 1901–1914, Aug. 1995.

[19] G. Zhou, G. B. Giannakis, and A. Swami, "On polynomial phase signals with time-varying amplitudes," *IEEE Trans. on Signal Processing*, vol. 44, no. 4, pp. 848–861, May 1996.

[20] R. H. Bhattacharya and R. R. Rao, *Normal Approximation and Asymptotic Expansions*. Wiley and Sons, 1976.

[21] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic ideas and selected topics*, 2nd ed. Prentice Hall, 2001.

[22] B. Boashash, *Higher-Order Statistical Signal Processing*. John Wiley & Sons, Inc., 1995.

[23] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," in *Proc. IEEE*, vol. 79, no. 3, Mar. 1991, pp. 278–305.

[24] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall Inc, 1993.

[25] J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge, 2006.

[26] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[27] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[28] A. N. Lemma, A.-J. van der Veen, and E. F. Deprettere, "Analysis of joint angle-frequency estimation using ESPRIT," *IEEE Trans. on Signal Processing*, vol. 51, no. 5, pp. 1264–1283, May 2003.

[29] ——, "Multiresolution ESPRIT algorithm," *IEEE Trans. on Signal Processing*, vol. 47, no. 6, pp. 1722–1726, June 1999.

[30] H. Chiang and C. Nikias, "The ESPRIT algorithm with higher-order statistics," in *Proc. Int. Workshop Higher-Order Stat.*, Vail, 1989, pp. 163–168.

[31] A. Zeira and B. Friedlander, "Joint direction finding, signal and channel response estimation for a polynomial phase signal in a multipath channel," in *Proc. 30th Asilomar Conf. on Sign., Syst. and Comp.*, Pacific Grove, CA, Nov. 1996, pp. 733–737.

[32] A. Belouchrani and M. G. Amin, "Time-frequency MUSIC," *IEEE Signal Process. Lett.*, vol. 6, pp. 109–110, 1999.

[33] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. on Signal Processing*, vol. 40, no. 3, pp. 590–600, Mar. 1992.

[34] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. on Signal Processing*, vol. 39, no. 5, pp. 1110–1121, May 1991.

[35] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford University, California, USA, 1981.

[36] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-83*, Boston, 1983, pp. 336–339.

[37] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

[38] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques – ESPRIT," in *Proc. 19th Asilomar Conf. on Sign., Syst. and Comp.*, Pacific Grove, CA, Nov. 1985, pp. 83–89.

[39] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, July 1990.

[40] J. H. Holland, *Adaptation in Natural and Artificial Systems*, 2nd ed. MIT Press, 1992.

[41] G. B. Dantzig, *Linear Programming and Extensions.* Princeton University Press, 1998.

[42] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectral Estim. Workshop*, NY, 1979, pp. 234–258.

[43] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice-Hall Inc., 1993.

[44] Scharf, *Statistical Signal Processing.* Addison-Wesley, 1990.

[45] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Prentice-Hall Inc, 1996.

[46] H. Jeffreys and B. Jeffreys, *Methods of Mathematical Physics*, 3rd ed. Cambridge University Press, 1988, ch. Weierstrass's Theorem on Approximation by Polynomials, pp. 446–448.

[47] G. Galati, *Advanced Radar Techniques and Systems*, G. Galati, Ed. London: IEE, 1993.

[48] J. Chatillon, M. E. Bouhier, and M. E. Zakharia, "Synthetic aperture sonar for seabed imaging: Relative merits of narrowband and wideband approaces," *IEEE J. Ocean. Eng.*, vol. 17, no. 1, pp. 95–105, Jan. 1992.

[49] S. Barbarossa and A. Scaglione, "Product high-order ambiguity function for multicomponent polynomial-phase signal modeling," *IEEE Trans. on Signal Processing*, vol. 46, no. 3, pp. 691–707, Mar. 1998.

[50] X.-G. Xia, "Discrete chirp-fourier transform and its application to chirp rate estimation," *IEEE Trans. on Signal Processing*, vol. 48, no. 11, pp. 3122–3133, Nov. 2000.

[51] P. M. Djuric and S. M. Kay, "Parameter estimation of chirp signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2118–2126, Dec. 1990.

[52] S. Peleg and B. Porat, "Estimation and classification of polynomial phase signals," *IEEE Trans. on Info. Theory*, vol. 37, no. 2, pp. 422–430, Mar. 1991.

[53] S. Shamsunder, G. Giannakis, and B. Friedlander, "Estimating random-amplitude polynomial phase signals: A cyclostationary approach," *IEEE Trans. on Signal Processing*, vol. 43, no. 2, pp. 492–505, Feb. 1995.

[54] O. Besson, M. Ghogho, and A. Swami, "Parameter estimation for random amplitude chirp signals," *IEEE Trans. on Signal Processing*, vol. 47, no. 12, pp. 3208–3219, Dec. 1999.

[55] B. Völcker and B. Ottersten, "Chirp parameter estimation from a sample covariance matrix," *IEEE Trans. on Signal Processing*, vol. 49, no. 3, pp. 603–612, Mar. 2001.

[56] D. S. Pham and A. M. Zoubir, "analysis of multicomponent polynomial phase signals," *IEEE Trans. on Signal Processing*, vol. 55, no. 1, pp. 56–65, Jan. 2007.

[57] M. G. Amin, "Spatial time-frequency distributions for direction finding and blind source seperation," in *Proc. SPIE*, vol. 3723, 1999, pp. 62–70.

[58] B. Völcker and B. Ottersten, "Linear chirp parameter estimation from multichannel data," in *Proc. 32nd Asilomar Conf. on Sign., Syst. and Comp.*, Pacific Grove, CA, Nov. 1998, pp. 238–242.

[59] A. B. Gershman and M. G. Amin, "Coherent wideband doa estimation of multiple FM signals using spatial time-frequency distributions," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-00*, Istanbul, Turkey, June 2000, pp. 3065–3068.

[60] ——, "Wideband direction of arrival estimation of multiple chirp signals using spatial time-frequency distributions," *IEEE Signal Process. Lett.*, vol. 7, no. 6, pp. 152–155, June 2000.

[61] N. Ma and T. J. Goh, "Ambiguity-function-based techniques to estimate DOA of broadband chirps signals," *IEEE Trans. on Signal Processing*, vol. 54, no. 5, pp. 1826–1839, May 2006.

[62] G. Wang and X.-G. Xia, "Iterative algorithm for direction of arrival estimation with wideband chirp signals," in *IEE Proc. Radar, Sonar, Navig.*, vol. 147, Oct. 2000, pp. 233–238.

[63] A. B. Gershman, M. Pesavento, and M. G. Amin, "Estimating parameters of multiple wideband polynomial-phase sources in sensor arrays," *IEEE Trans. on Signal Processing*, vol. 49, no. 12, pp. 2924–2934, Dec. 2001.

[64] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources,"

IEEE Trans. Acoust., Speech, Signal Process., vol. 33, no. 4, pp. 823–831, Aug. 1985.

[65] B. Friedlander and A. J. Weiss, "Direction finding for wide-band signals using an interpolated array," *IEEE Trans. on Signal Processing*, vol. 41, no. 4, pp. 1618–1634, Apr. 1993.

[66] J. F. Böhme, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Processing*, vol. 10, pp. 329–337, 1986.

[67] S. Peleg and B. Friedlander, "Multicomponent signal analysis using the polynomial-phase transform," *IEEE Trans. on Aerosp. Electron. Syst.*, vol. 32, no. 1, pp. 378–387, Jan. 1996.

[68] M. Wang, A. K. Chan, and C. K. Chui, "Linear frequency-modulated signal detection using radon-ambiguity transform," *IEEE Trans. on Signal Processing*, vol. 46, no. 3, pp. 571–586, Mar. 1998.

[69] S. V. Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis.* Philadelphia, PA: SIAM, 1991.

[70] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1997.

[71] A. B. Gershman, V. I. Turchin, and V. A. Zverev, "Experimental results of localization of moving underwater signal by adaptive beamforming," *IEEE Trans. on Signal Processing*, vol. 43, no. 10, pp. 2249–2257, Oct. 1995.

[72] W. C. Knight, R. G. Pridham, and S. M. Kay, "Digital signal processing for sonar," *Proc. of the IEEE*, vol. 69, no. 11, pp. 1451–1508, 1981.

[73] A.-J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, May 1996.

[74] M. Vanderveen, "Estimation of parametric channel models in wireless communication networks," Ph.D. dissertation, Stanford University, California, USA, Dec. 1997.

[75] A. K. Nandi, Ed., *Blind estimation using higher-order statistics.* Boston: Kluwer Academic Publishers, 1999.

[76] V. Varajarajan and J. L. Krolik, "Multichannel system identification methods for sensor array calibration in uncertain multipath environment," in *Proc. 11th IEEE Signal Processing Workshop on Statistical Signal Processing (SSP'01)*, Singapore, August 2001, pp. 297–300.

[77] K. Abed-Meraim, S. Attallah, T. J. Lim, and M. O. Damen, "A blind interference canceller in DS-CDMA," in *Proc. IEEE 6th International Symposium on Spread Spectrum Techniques and Applications*, vol. 2, Parsippany, NJ, USA, Sept. 2000, pp. 358–362.

[78] S. Rickard, R. Balan, and J. Rosca, "Real-time time-frequency based blind source separation," in *Proc. 3rd International Conference on Independent Component Analysis and Blind Source Separation (ICA'01)*, San Diego, CA, USA, December 2001, pp. 651–656.

[79] A. Rouxel, D. L. Guennec, and O. Macchi, "Unsupervised adaptive separation of impulse signals applied to eeg analysis," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-00*, vol. 1, Istanbul, Turkey, June 2000, pp. 420–423.

[80] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[81] T. W. Lee, *Independent Component Analysis: Theory and Applications.* Boston: Kluwer Academic Publishers, 1998.

[82] J.-F. Cardoso, "Blind signal separation: statistical principles," in *Proc. IEEE*, vol. 86, no. 10, 1998, pp. 2009–2025.

[83] V. Zarzoso and A. K. Nandi, "Blind source separation," in *Blind estimation using higher-order statistics*, A. K. Nandi, Ed. Boston: Kluwer Academic Publishers, January 1999, pp. 167–252.

[84] F. J. Theis, A. Jung, C. G. Puntonet, and E. W. Lang, "Linear geometric ica: Fundamentals and algorithms," *Neural Computation*, vol. 15, no. 2, pp. 419–439, 2003.

[85] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Trans. on Signal Processing*, vol. 46, no. 11, pp. 2888–2897, 1998.

[86] K. Abed-Meraim, Y. Xiang, J. Manton, and Y. Hua, "Blind source separation using second order cyclostationary statistics," *IEEE Trans. on Signal Processing*, vol. 49, no. 4, pp. 694–701, April 2001.

[87] K. Abed-Meraim, A. Belouchrani, and A. R. Leyman, "Blind source separation using time frequency distributions," in *Time-Frequency Signal Analysis and Processing.* Elsevier, 2003, pp. 324–332.

[88] B. Boashash, *Time-Frequency Signal Analysis and Processing: Method and Applications.* Elsevier, 2003.

[89] L. D. Lathauwer and J. Castaing, "Second-order blind identification of underdetermined mixtures," in *6th Int. Conference on Independent Component Analysis and Blind Signal Separation*, J. R. et al., Ed. Charleston, SC, USA: Springer, Mar. 2006, pp. 40–47.

[90] L. Albera, A. Ferreol, P. Comon, and P. Chevalier, "Blind identification of overcomplete mixtures of sources (biome)," *Lin. Algebra Appl.*, vol. 391, pp. 1–30, Nov. 2004.

[91] A. Ferreol, L. Albera, and P. Chevalier, "Fourth order blind identification of underdetermined mixtures of sources (FOBIUM)," *IEEE Trans. on Signal Processing*, vol. 53, no. 5, pp. 1640–1653, May 2005.

[92] A. Belouchrani and J.-F. Cardoso, "A maximum likelihood source separation for discrete sources," in *Proc. European Signal Processing Conference (EUSIPCO '94)*, vol. 2, Edinburg, Scotland, Sept. 1994, pp. 768–771.

[93] A. Taleb and C. Jutten, "On underdetermined source separation," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-99*, vol. 3, Phoenix, AZ, USA, Mar. 1999, pp. 1445–1448.

[94] T. W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, "Blind source separation of more sources than mixtures using over complete representations," *IEEE Trans. on Signal Processing*, vol. 6, no. 4, pp. 87–90, 1999.

[95] P. Comon and O. Grellier, "Non-linear inversion of underdetermined mixtures," in *Proc. 1st International Conference on Independent Component Analysis and Blind Source Separation (ICA'99)*, Aussois, France, Jan. 1999, pp. 461–465.

[96] K. L. Diamantaras, "Blind separation of multiple binary sources using a single linear mixture," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-00*, vol. 5, Istanbul, Turkey, June 2000, pp. 2889–2892.

[97] P. Bofil and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, 2001.

[98] N. Linh-Trung, A. Belouchrani, K. Abed-Meraim, and B. Boashash, "Separating more sources than sensors using time-frequency distributions," *EURASIP J. Appl. Sig. Proc.*, no. 17, pp. 2828–2847, 2005.

[99] N. Linh-Trung, Aissa-El-Bey, K. Abed-Meraim, and A. Belouchrani, "Underdetermined blind source seperation of non-disjoint nonstationary sources in the time-frequency domain," in *Proc. 8th Int. Symp. Signl. Proc. and Appl.*, vol. 1, Aug. 2005, pp. 46–49.

[100] P. Chevalier, L. Albera, A. Ferreol, and P. Comon, "On the virtual array concept for the higher-order array processing," *IEEE Trans. on Signal Processing*, vol. 53, no. 4, pp. 1254–1271, Apr. 2005.

[101] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. on Signal Processing*, vol. 50, no. 7, pp. 1545–1553, July 2002.

[102] C. Fevotte and C. Doncarli, "Two contributions to blind source separation using time-frequency distributions," *IEEE Signal Process. Lett.*, vol. 11, no. 3, pp. 386–389, Mar. 2004.

[103] G. F. Boudreaux-Bartels and T. W. Marks, "Time-varying filtering and signal estimation using Wigner distributions," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 422–430, Mar. 1986.

[104] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, Sept. 1978.

[105] D. Kletter and H. Messer, "Suboptimal detection of non-gaussian signals by third-order spectral analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 6, pp. 901–909, June 1990.

[106] H. Linhart and W. Zucchini, *Model Selection*. New York: Wiley and Sons, 1986.

[107] A. Swami, "Multiplicative noise models: parameter estimation using cumulants," *Signal Proc.*, vol. 36, pp. 355–373, 1994.

[108] R. J. Urick, *Principles of Underwater Sound*, 2nd ed. McGraw-Hill, 1975.

[109] J.-F. Cardoso and E. Moulines, "Asymptotic performance analysis on direction-finding algorithms based on fourth-order cumulants," *IEEE Trans. on Signal Processing*, vol. 43, no. 1, pp. 214–224, Jan. 1995.

[110] M. C. Dogan and J. M. Mendel, "Applications of cumulants to array processing - Part I: Aperture extension and array calibration," *IEEE Trans. on Signal Processing*, vol. 43, no. 5, pp. 1200–1216, May 1995.

[111] B. Porat and B. Friedlander, "Direction finding algorithms based on high-order statistics," *IEEE Trans. on Signal Processing*, vol. 39, no. 9, pp. 2016–2025, Jan. 1991.

[112] R. Pan and C. Nikias, "Harmonic decomposition in cumulant domains," in *Proc. Int. Conf. on Acoust., Speech, Signal Process., ICASSP*, 1988, pp. 2356–2359.

[113] J.-F. Cardoso, "Higher-order narrow band array processing," in *Proc. Int. Workshop Higher-Order Stat.*, Chamrousse, France, 1991, pp. 121–130.

[114] B. D. Rao and K. V. S. Hari, "Performance analysis of root-MUSIC," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.

[115] A. J. Caamano, R. Boloix-Tortosa, J. Ramos, and J. J. Murillo-Fuentes, "Hybrid higher-order statistics learning in multiuser detection," *IEEE Trans. Syst., Man and Cybernetics, Part C*, vol. 34, no. 4, pp. 417–424, Nov. 2004.

[116] G. Jacovitti, G. Panci, and G. Scarano, "Bussgang-zero crossing equalization: an integrated hos-sos approach," *IEEE Trans. on Signal Processing*, vol. 49, no. 11, pp. 2798–2812, Nov. 2001.

[117] T. Stathaki and A. G. Constantinides, "Segmentation and modelling of textured images through combined second and third order statistical models," in *Proceedings of the SPIE, The International Society for Optical Engineering*, vol. 2308, 1994, pp. 1977–1986.

[118] T. Stathaki, "Segmentation of textured surfaces using mixed order statistics and neural network classifiers," *IEE Electronic Letters*, vol. 31, no. 23, pp. 1990–1991, Nov. 1995.

[119] M. Klajman and A. G. Constantinides, "A combined statistics cost function for blind and semi-blind source separation," in *Proc. 3rd International Conference on Independent Component Analysis and Blind Source Separation (ICA'01)*, San Diego, CA, USA, Dec. 2001, pp. 605–610.

[120] A. B. Gershman and H. Messer, "Robust mixed-order root-MUSIC," *Circuits Systems Signal Process.*, vol. 19, no. 5, pp. 451–466, 2000.

[121] A.-J. van der Veen, "Joint diagonalization via subspace fitting techniques," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-01*, Salt Lake City, Utah, 2001, pp. 2773–2776.

[122] J.-F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. blind identification of more sources than sensors," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-91*, 1991, pp. 3109–3112.

[123] L. De Lathauwer, B. D. Moor, and J. Vandewalle, "An algebraic algorithm for ica with more sources than sensors," in *Mathematics in Signal Processing V.* Clarendon Press, Oxford, 2002, pp. 37–46.

[124] S. Peleg, "Estimation and detection with the discrete polynomial transform," Ph.D. dissertation, University of California at Davis, California, USA, 1993.

[125] S. Peleg and B. Porat, "Linear FM signal parameter estimation from discrete-time observations," *IEEE Trans. on Aerosp. Electron. Syst.*, vol. 27, no. 4, pp. 607–616, Jan. 1991.

[126] J. G. Kirkwood, "Quantum statistics of almost classical ensembles," *Phys. Rev.*, vol. 44, pp. 31–37, 1933.

[127] A. B. Gershman and H. Messer, "Mixed-order root-MUSIC algorithms with improved robustness," in *Proc. 9th IEEE SP Workshop on Statistical Signal and Array Processing*, Sept. 1998, pp. 33–36.

[128] S. Lie, A. R. Leyman, and Y. H. Chew, "Underdetermined source separation for non-stationary signal," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-07*, Honolulu, Hawaii, April 2007, to be published.

[129] J. O. Toole, M. Mesbah, and B. Boashash, "A discrete time and frequency Wigner distribution: Properties and implementation," in *Proc. Int. Symp. Digital Signl. Proc. and Comm. Syst.*, Nov. 2005, pp. 34–37.

[130] P. Chevalier and A. Ferreol, "On the virtual array concept for the fourth-order direction finding problem," *IEEE Trans. on Signal Processing*, vol. 47, no. 9, pp. 2592–2595, Sept. 1999.

[131] A. R. Leyman, Z. M. Kamran, and K. Abed-Meraim, "Higher-order time frequency-based blind source separation technique," *IEEE Signal Process. Lett.*, vol. 7, no. 7, pp. 193–196, July 2000.

[132] R. G. Baraniuk, M. Coates, and P. Steeghs, "Hybrid linear/quadratic time-frequency attributes," *IEEE Trans. on Signal Processing*, vol. 49, no. 4, pp. 760–766, Apr. 2001.

[133] R. Bernardini and J. Kovacevic, "Arbitrary tilings of the time-frequency plane using local bases," *IEEE Trans. on Signal Processing*, vol. 47, no. 8, pp. 2293–2304, Aug. 1999.

[134] C. Herley, J. Kovacevic, K. Ramchandran, and M. Vetterli, "Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tilings algorithms," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3341–3359, Dec. 1993.

[135] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 806–811, Aug. 1985.

[136] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, "Parameter estimation in the presence of bounded data uncertainties," *SIAM J. Matrix Anal. Appl.*, vol. 19, no. 1, pp. 235–252, Jan. 1998.

[137] ——, "Parameter estimation in the presence of bounded modeling errors," *IEEE Signal Process. Lett.*, vol. 4, no. 7, pp. 195–197, July 1997.

[138] G. H. Golub, A. Hoffman, and G. W. Stewart, "A generalization of the Eckart-Young-Mirsky matrix approximation theorem," *Lin. Algebra and Its Applic.*, pp. 317–327, 1987.

[139] S. Shahbazpanahi, S. Valaee, and M. H. Bastani, "Distributed source localization using ESPRIT algorithm," *IEEE Trans. on Signal Processing*, vol. 49, no. 10, pp. 2169–2178, Oct. 2001.

[140] J.-F. Cardoso, "Source separation using higher order moments," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-89*, Glasgow, 1989, pp. 2109–2112.

[141] S. Shamsunder and G. Giannakis, "Detection and parameter estimation of multiple sources via HOS," in *Proc. Int. Workshop Higher-Order Stat.*, Chamrousse, France, 1991, pp. 265–268.

[142] A. Hassanien, S. Shahbazpanahi, and A. B. Gershman, "A generalized Capon estimator for localization of multiple spread sources," *IEEE Trans. on Signal Processing*, vol. 52, no. 1, pp. 280–283, Jan. 2004.

[143] S. Lie, A. R. Leyman, and Y. H. Chew, "Wideband chirp parameter estimation in sensor arrays through dpt," *IEE Electronic Letters*, vol. 39, no. 23, pp. 1633–1634, Nov. 2003.

[144] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. on Info. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[145] M. Costa, "Writing on dirty paper," *IEEE Trans. on Info. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[146] S. Gelfand and M. Pinsker, "Coding for channels with random parameters," *Probl. Control Inform. Theory*, vol. 9, pp. 19–31, Jan. 1980.

[147] S. Shahbazpanahi, S. Valaee, and A. B. Gershman, "Parametric localization of multiple incoherently distributed sources using covariance fitting," in *Proc. 3rd IEEE Workshop Sensor Array Multichannel Signal Process.*, Rosslyn, VA, Aug. 2002, pp. 332–336.

[148] M. Wang, A. K. Chan, and C. K. Chui, "Wigner-Ville distribution decomposition via wavelet paket transform," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.*, Paris, France, June 1996, pp. 413–416.

[149] I. Cohen, S. Raz, and D. Malah, "Eliminating interference terms in Wigner distribution using extended libraries of bases," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-97*, Munich, Germany, Apr. 1997, pp. 2133–2136.

[150] P. Comon and E. Moreau, "Improved contrast dedicated to blind separation in communications," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-97*, Munich, Germany, Apr. 1997, pp. 3453–3456.

[151] S. Lie, A. R. Leyman, and Y. H. Chew, "Parameter estimation of wideband chirp signals in sensor arrays through DPT," in *Proc. 37th Asilomar Conf. on Sign., Syst. and Comp.*, Pacific Grove, CA, Nov. 2003, pp. 1878–1881.

[152] ——, "Wideband polynomial-phase parameter estimation in sensor array," in *Proc. of IEEE International Symposium on Signal Processing and Information Technology*, Darmstadt, Germany, Dec. 2003, pp. 314–317.

[153] B. Völcker, "Performance analysis of parametric spectral estimators," Ph.D. dissertation, Royal Institute of Technology, Stockholm, Sweden, 2002.

[154] C. K. Chui, *An Introduction to Wavelets*. Academic Press, 1992.

Appendix A

Proof that Cumulants of a Gaussian distribution is zero

We prove that the cumulant of a Gaussian distribution is zero. Suppose $\mathbf{x}$ is a random vector and is Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{R}$, then its pdf is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{R}^{-1}(\mathbf{x}-\mathbf{m})\right\} \tag{A.1}$$

If $\mathbf{y}$ is a random vector of length n, then its moment generating function is defined as

$$M_{\mathbf{y}}(\boldsymbol{\lambda}) \triangleq E\{\exp(\boldsymbol{\lambda}^T\mathbf{y})\} \tag{A.2}$$

and its cumulant generating function is defined as

$$C_{\mathbf{y}}(\boldsymbol{\lambda}) \triangleq \ln M_{\mathbf{y}}(\boldsymbol{\lambda}) \tag{A.3}$$

where $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_n]^T$ is a real-valued vector of size n. The k^{th} order cumulant of $\mathbf{y}$ is defined as

$$\text{Cum}(\mathbf{y}) \triangleq \left.\frac{\partial^k C_{\mathbf{y}}(\boldsymbol{\lambda})}{\partial\lambda_{i_1}\cdots\partial\lambda_{i_k}}\right|_{\boldsymbol{\lambda}=\mathbf{0}} \tag{A.4}$$

with $i_1, \cdots, i_k \in \{1, \cdots, n\}$ and repetitive indices are allowed, e.g. $i_1 = \cdots = i_k$. Hence, to prove that the cumulant of a Gaussian distribution of order greater than two is zero, one only need to show Eqn. (A.4) equals zero for $n > 2$. The moment

generating function of a Gaussian vector $\mathbf{x}$ is derived as follows (by substitution of Eqn. (A.1) into Eqn. (A.2))

$$\begin{aligned}
M_{\mathbf{x}}(\boldsymbol{\lambda}) &= \int \exp(\boldsymbol{\lambda}^T\mathbf{x})\frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}}\exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{R}^{-1}(\mathbf{x}-\mathbf{m})\}d\mathbf{x} \\
&= \int \frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}}\exp\{-\frac{1}{2}[(\mathbf{x}-\mathbf{m})^T\mathbf{R}^{-1}(\mathbf{x}-\mathbf{m}) - 2\boldsymbol{\lambda}^T\mathbf{x}]\}d\mathbf{x} \\
&= \int \frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}}\exp\{-\frac{1}{2}[\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x} - 2\mathbf{m}^T\mathbf{R}^{-1}\mathbf{x} + \mathbf{m}^T\mathbf{R}^{-1}\mathbf{m} - 2\boldsymbol{\lambda}^T\mathbf{x}]\}d\mathbf{x} \\
&= \int \frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}}\exp\{-\frac{1}{2}[\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x} - 2(\mathbf{m}^T\mathbf{R}^{-1} + \boldsymbol{\lambda}^T)\mathbf{x} + \mathbf{m}^T\mathbf{R}^{-1}\mathbf{m}]\}d\mathbf{x} \\
&= \int \frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}}\exp\{-\frac{1}{2}[\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x} - 2(\mathbf{m}^T + \boldsymbol{\lambda}^T\mathbf{R})\mathbf{R}^{-1}\mathbf{x} + \mathbf{m}^T\mathbf{R}^{-1}\mathbf{m}]\}d\mathbf{x} \\
&= \int \frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}}\exp\{-\frac{1}{2}[(\mathbf{x}-(\mathbf{m}+\mathbf{R}\boldsymbol{\lambda}))^T\mathbf{R}^{-1}(\mathbf{x}-(\mathbf{m}+\mathbf{R}\boldsymbol{\lambda})) \\
&\quad +\mathbf{m}^T\mathbf{R}^{-1}\mathbf{m} - (\mathbf{m}+\mathbf{R}\boldsymbol{\lambda})^T\mathbf{R}^{-1}(\mathbf{m}+\mathbf{R}\boldsymbol{\lambda})]\}d\mathbf{x} \\
&= \int \frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}}\exp\{-\frac{1}{2}(\mathbf{x}-(\mathbf{m}+\mathbf{R}\boldsymbol{\lambda}))^T\mathbf{R}^{-1}(\mathbf{x}-(\mathbf{m}+\mathbf{R}\boldsymbol{\lambda}))\}d\mathbf{x} \\
&\quad \times \exp(\boldsymbol{\lambda}^T\mathbf{m} + 0.5\boldsymbol{\lambda}^T\mathbf{R}\boldsymbol{\lambda}) \\
&= \exp(\boldsymbol{\lambda}^T\mathbf{m} + 0.5\boldsymbol{\lambda}^T\mathbf{R}\boldsymbol{\lambda}) \qquad \text{(A.5)}
\end{aligned}$$

Thus, the cumulant generating function is given by

$$C_{\mathbf{x}}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^T\mathbf{m} + 0.5\boldsymbol{\lambda}^T\mathbf{R}\boldsymbol{\lambda} \qquad \text{(A.6)}$$

Since Eqn. (A.6) is a quadratic function of $\boldsymbol{\lambda}$, then its third-order and higher-order derivatives vanish. Q.E.D.

Appendix B

Derivation of Cramer-Rao Bound for Array PPS Estimation

In this appendix, we derive Cramer-Rao bound (CRB) when estimating the parameters of wideband PPS signals impinged on ULA. We first introduce the notations used in the derivation. We rearrange Eqn. (3.3) such that,

$$s_i(n) = A_i e^{j\alpha_i} g(\boldsymbol{\varpi}_i, n) \tag{B.1}$$

where

$$g(\boldsymbol{\varpi}_i, n) \triangleq \exp\left(j \sum_{k=1}^{K} a_{i,k}(n\Delta)^k\right) \tag{B.2}$$

and $\boldsymbol{\varpi}_i \triangleq [a_{i,1}, \cdots, a_{i,K}]^T$. Define $\mathbf{G}(\boldsymbol{\varpi}, n) \triangleq \text{diag}\{g(\boldsymbol{\varpi}_1, n), \cdots, g(\boldsymbol{\varpi}_L, n)\}$ and $\boldsymbol{\beta} \triangleq [A_1 e^{j\alpha_1}, \cdots, A_L e^{j\alpha_L}]^T$ such that, the PPS impinged on ULA is given by

$$\mathbf{x}(n) = \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n)\mathbf{G}(\boldsymbol{\varpi}, n)\boldsymbol{\beta} + \mathbf{v}(n) \tag{B.3}$$

where $\boldsymbol{\theta} \triangleq [\theta_1, \cdots, \theta_L]^T$, $\boldsymbol{\varpi} \triangleq [\boldsymbol{\varpi}_1^T, \cdots, \boldsymbol{\varpi}_L^T]^T$, and $\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, t)$ is the same as* $\mathbf{A}(n)$ in Eqn. (3.4). Since we have assumed that $\mathbf{v}(n) \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I})$ (complex circularly and white Gaussian distributed), then

$$\mathbf{x}(n) \sim \mathcal{N}(\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n)\mathbf{G}(\boldsymbol{\varpi}, n)\boldsymbol{\beta}, \sigma_n^2\mathbf{I}) \tag{B.4}$$

*Here we explicitly denote its dependency on $\boldsymbol{\theta}$ and $\boldsymbol{\varpi}$ for derivation purpose

We group all the unknown parameters, including σ_n^2 which is not of our interest in this estimation, as follows

$$\boldsymbol{\psi} \triangleq [\boldsymbol{\beta}_{\parallel}^T, \boldsymbol{\beta}_{\angle}^T, \boldsymbol{\theta}^T, \boldsymbol{\varpi}^T, \sigma_n^2]^T \tag{B.5}$$

where $\boldsymbol{\beta}_{\parallel} \triangleq |\boldsymbol{\beta}| = [A_1, \cdots, A_L]^T$ and $\boldsymbol{\beta}_{\angle} \triangleq \angle\{\boldsymbol{\beta}\} = [\alpha_1, \alpha_2, \cdots, \alpha_L]^T$

In order to get CRB, we need Fisher Information Matrix (FIM) because its inverse is the CRB. The elements of the FIM, $\mathbf{F}_{l,k}$, of a complex circularly Gaussian process $\mathbf{x}(n) \sim \mathcal{N}(\boldsymbol{\mu}(n), \mathbf{R})$ are given by [24, 44],

$$\mathbf{F}_{l,k} = N\text{tr}\left(\mathbf{R}^{-1}\frac{\mathbf{R}}{\partial\boldsymbol{\psi}_l}\mathbf{R}^{-1}\frac{\mathbf{R}}{\partial\boldsymbol{\psi}_k}\right) + 2\Re\left(\sum_{n=0}^{N-1}\frac{\partial\boldsymbol{\mu}(n)^H}{\partial\boldsymbol{\psi}_l}\mathbf{R}^{-1}\frac{\partial\boldsymbol{\mu}(n)^H}{\partial\boldsymbol{\psi}_k}\right) \tag{B.6}$$

Applying above Eqn. (B.6) to our model in Eqn. (B.4), gives

$$\begin{aligned}\mathbf{F}_{l,k} = &\frac{MN}{\sigma_n^4}\frac{\partial\sigma_n^2}{\partial\boldsymbol{\psi}_l}\frac{\partial\sigma_n^2}{\partial\boldsymbol{\psi}_k} \\ &+ \frac{2}{\sigma_n^2}\Re\left(\sum_{n=0}^{N-1}\frac{\partial\{\boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial\boldsymbol{\psi}_l}\frac{\partial\{\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathbf{G}(\boldsymbol{\varpi},n)\boldsymbol{\beta}\}}{\partial\boldsymbol{\psi}_k}\right)\end{aligned} \tag{B.7}$$

Below, we will derive the elements that is required to build each FIM sub-block.

First, we derive the expression related to the derivative with respect to $\boldsymbol{\beta}_{\parallel}$. Direct computations yield,

$$\frac{\partial\{\boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial\boldsymbol{\beta}_{\parallel}} = \text{diag}(e^{-j\boldsymbol{\beta}_{\angle}})\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H \tag{B.8}$$

where $e^{-j\boldsymbol{\beta}_{\angle}} \triangleq [e^{-j\alpha_1}, e^{-j\alpha_2}, \cdots, e^{-j\alpha_L}]^T$ and, hence, $\text{diag}(e^{-j\boldsymbol{\beta}_{\angle}})$ is a diagonal matrix containing elements of vector $e^{-j\boldsymbol{\beta}_{\angle}}$ on its diagonal.

Next expression related to the derivative with respect to $\boldsymbol{\beta}_{\angle}$, also by straightforward computations, is given by

$$\frac{\partial\{\boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial\boldsymbol{\beta}_{\angle}} = -j\mathcal{B}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H \tag{B.9}$$

where $\mathcal{B} \triangleq \text{diag}(\boldsymbol{\beta})$. Subsequent expression related to the derivative with respect

to θ_l is given by

$$\begin{aligned}\frac{\partial\{\boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial\theta_l} &= \boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\left\{\frac{\partial\mathbf{a}(\theta_l,\boldsymbol{\varpi}_l,n)^H}{\partial\theta_l}\right\} \\ &= \boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{e}_l\mathbf{e}_l^T\left\{\frac{\partial}{\partial\boldsymbol{\theta}^T}\odot\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\right\}^H \\ &= \boldsymbol{\beta}^H\mathbf{e}_l\mathbf{e}_l^T\mathbf{G}(\boldsymbol{\varpi},n)^H\left\{\frac{\partial}{\partial\boldsymbol{\theta}^T}\odot\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\right\}^H\end{aligned} \tag{B.10}$$

where $\odot$ is Khatri-Rao column-wise Kronecker product and $\mathbf{e}_l$ is a vector containing one in the l^{th} position and zeros elsewhere. Note that,

$$\frac{\partial}{\partial\boldsymbol{\theta}^T}\odot\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)=\left[\frac{\partial\mathbf{a}(\theta_1,\boldsymbol{\varpi}_1,n)}{\partial\theta_1},\cdots,\frac{\partial\mathbf{a}(\theta_L,\boldsymbol{\varpi}_L,n)}{\partial\theta_L}\right] \tag{B.11}$$

where

$$\frac{\partial\mathbf{a}(\theta_l,\boldsymbol{\varpi}_l,n)}{\partial\theta_l}=p(\theta_l,\boldsymbol{\varpi}_l,n)\mathbf{U}\mathbf{a}(\theta_l,\boldsymbol{\varpi}_l,n) \tag{B.12}$$

In Eqn. (B.12) above, we have used the following definitions

$$\mathbf{U} \triangleq \frac{d}{c}\,\text{diag}\{0,1,\cdots,M-1\} \tag{B.13}$$

$$p(\theta_l,\boldsymbol{\varpi}_l,n) \triangleq j\cos\theta_l\left(\sum_{k=1}^{K}ka_{l,k}(\Delta n)^{k-1}\right) \tag{B.14}$$

By stacking Eqn. (B.10) over different l's and using Eqn. (B.13) and (B.14), yields

$$\begin{aligned}\frac{\partial\{\boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial\boldsymbol{\theta}} &= \begin{bmatrix}\boldsymbol{\beta}^H\mathbf{e}_1\mathbf{e}_1^T\mathbf{G}(\boldsymbol{\varpi},n)^H\left\{\frac{\partial}{\partial\boldsymbol{\theta}^T}\odot\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\right\}^H\\ \vdots\\ \boldsymbol{\beta}^H\mathbf{e}_L\mathbf{e}_L^T\mathbf{G}(\boldsymbol{\varpi},n)^H\left\{\frac{\partial}{\partial\boldsymbol{\theta}^T}\odot\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\right\}^H\end{bmatrix} \\ &= \mathcal{B}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\left(\frac{\partial}{\partial\boldsymbol{\theta}^T}\odot\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\right)^H \\ &= \mathcal{B}^H\mathbf{P}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\mathbf{U}^H\end{aligned} \tag{B.15}$$

where

$$\mathbf{P}(\boldsymbol{\theta},\boldsymbol{\varpi},n)=\text{diag}\{p(\theta_1,\boldsymbol{\varpi}_1,n),\cdots,p(\theta_L,\boldsymbol{\varpi}_L,n)\} \tag{B.16}$$

Before proceeding with the last expression related to the derivatives with respect

to $a_{l,k}$, let us define the following notations:

$$\begin{aligned}
\boldsymbol{\eta}_k &\triangleq [a_{1,k}, a_{2,k}, \cdots, a_{L,k}]^T && \text{(B.17)}\\
h_k(\boldsymbol{\varpi}_l, n) &\triangleq \frac{\partial g(\boldsymbol{\varpi}_l, n)}{\partial a_{l,k}} \\
&= j(\Delta n)^k g(\boldsymbol{\varpi}_l, n) && \text{(B.18)}\\
\mathbf{g}(\boldsymbol{\varpi}, n) &\triangleq [g(\boldsymbol{\varpi}_1, n), g(\boldsymbol{\varpi}_2, n), \ldots, g(\boldsymbol{\varpi}_L, n)]^T && \text{(B.19)}\\
\mathbf{H}_k(\boldsymbol{\varpi}, n) &\triangleq \frac{\partial \mathbf{g}^T(\boldsymbol{\varpi}, n)}{\partial \boldsymbol{\eta}_k} \\
&= \mathrm{diag}\{h_k(\boldsymbol{\varpi}_1, n), h_k(\boldsymbol{\varpi}_2, n), \cdots, h_k(\boldsymbol{\varpi}_L, n)\} \\
&= j(\Delta n)^k \mathbf{G}(\boldsymbol{\varpi}, n) && \text{(B.20)}\\
\mathbf{b}_k(\theta_l, \boldsymbol{\varpi}_l, n) &\triangleq \frac{\partial \mathbf{a}(\theta_l, \boldsymbol{\varpi}_l, n)}{\partial a_{l,k}} \\
&= jk(\Delta n)^{k-1} \sin\theta_l \mathbf{U}\mathbf{a}(\theta_l, \boldsymbol{\varpi}_l, n) && \text{(B.21)}\\
\mathbf{B}_k(\boldsymbol{\theta}, \boldsymbol{\varpi}, n) &\triangleq [\mathbf{b}_k(\theta_1, \boldsymbol{\varpi}_1, n), \cdots, \mathbf{b}_k(\theta_L, \boldsymbol{\varpi}_L, n)] \\
&= jk(\Delta n)^{k-1} \mathbf{U}\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\varpi}, n)\mathbf{C} && \text{(B.22)}\\
\mathbf{C} &\triangleq \mathrm{diag}\{\sin\theta_1, \sin\theta_2 \cdots, \sin\theta_L\} && \text{(B.23)}\\
\mathbf{T}_k(n) &\triangleq jk(\Delta n)^{k-1}\left(\frac{\Delta n}{k}\mathbf{1}_M\mathbf{1}_L^T + \mathbf{u}\mathbf{c}^T\right) && \text{(B.24)}\\
\mathbf{u} &\triangleq \mathrm{diag}\{\mathbf{U}\} = \frac{d}{c}[0, 1, \cdots, M-1]^T && \text{(B.25)}\\
\mathbf{c} &\triangleq \mathrm{diag}\{\mathbf{C}\} = [\sin\theta_1, \sin\theta_2 \cdots, \sin\theta_L]^T && \text{(B.26)}
\end{aligned}$$

where $\mathbf{1}_M$ is a $(M\times 1)$-vector containing ones in all rows. Using above Eqns.(B.17)–(B.26), we derive the following

$$\begin{aligned}
\frac{\partial\{\boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial a_{l,k}} &= \frac{\partial\{\mathbf{g}(\boldsymbol{\varpi},n)^H\mathcal{B}^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial a_{l,k}} \\
&= \frac{\partial\{\mathbf{g}(\boldsymbol{\varpi},n)^H\}}{\partial a_{l,k}}\mathcal{B}^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H \\
&\quad +\mathbf{g}(\boldsymbol{\varpi},n)^H\mathcal{B}^H\frac{\partial\{\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial a_{l,k}} \\
&= \mathbf{e}_l^T\mathbf{H}_k(\boldsymbol{\varpi},n)^H\mathcal{B}^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H \\
&\quad +\mathbf{g}(\boldsymbol{\varpi},n)^H\mathcal{B}^H\mathbf{e}_l\mathbf{e}_l^T\mathbf{B}_k(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H && \text{(B.27)}
\end{aligned}$$

By stacking Eqn. (B.27) over different l's into vector form, we obtain

$$\begin{aligned}\frac{\partial\{\boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial\boldsymbol{\eta}_k} &= \mathbf{H}_k(\boldsymbol{\varpi},n)^H\mathcal{B}^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H \\ &\quad +\mathbf{G}(\boldsymbol{\varpi},n)^H\mathcal{B}^H\mathbf{B}_k(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H \end{aligned} \tag{B.28}$$

By using the relationship given in Eqn. (B.20) and (B.22), Eqn. (B.28) simplifies to

$$\begin{aligned}\frac{\partial\{\boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial\boldsymbol{\eta}_k} &= -j(\Delta n)^k\mathbf{G}_k(\boldsymbol{\varpi},n)^H\mathcal{B}^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H \\ &\quad -jk(\Delta n)^{k-1}\mathbf{G}(\boldsymbol{\varpi},n)^H\mathcal{B}^H\mathbf{C}^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\mathbf{U}^H \end{aligned} \tag{B.29}$$

By exploiting diagonal structure of $\mathbf{C}$, $\mathbf{U}$ and $\mathcal{B}$, as well as the definitions given in Eqn. (B.25) and (B.26), we can rewrite

$$\mathbf{G}(\boldsymbol{\varpi},n)^H\mathcal{B}^H\mathbf{C}^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\mathbf{U}^H = (\mathcal{B}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H)\circ(\mathbf{c}\mathbf{u}^T) \tag{B.30}$$

where $\circ$ is Schur-Hadamard (element-wise) matrix product. Using Eqn. (B.30) above, we obtain Eqn. (B.29) as follows,

$$\frac{\partial\{\boldsymbol{\beta}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\}}{\partial\boldsymbol{\eta}_k} = (\mathcal{B}^H\mathbf{G}(\boldsymbol{\varpi},n)^H\mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H)\circ\mathbf{T}_k(n)^H \tag{B.31}$$

Now, we are ready to derive sub-blocks of FIM by using Eqn. (B.7), (B.8), (B.9), (B.15), and (B.31), as follows:

$$F_{\sigma_n^2\sigma_n^2} = \frac{NM}{\sigma_n^4} \tag{B.32}$$

$$\mathbf{F}_{\boldsymbol{\theta}\sigma_n^2} = \mathbf{F}_{\boldsymbol{\beta}_{\parallel}\sigma_n^2} = \mathbf{F}_{\boldsymbol{\beta}_{\angle}\sigma_n^2} = \mathbf{F}_{\boldsymbol{\eta}_k\sigma_n^2} = \mathbf{0} \tag{B.33}$$

$$\mathbf{F}_{\boldsymbol{\beta}_{\parallel}\boldsymbol{\beta}_{\parallel}} = \frac{2}{\sigma_n^2}\Re\left\{\sum_{n=0}^{N-1}\mathrm{diag}(e^{-j\boldsymbol{\beta}_{\angle}})\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathrm{diag}(e^{j\boldsymbol{\beta}_{\angle}})\right\} \tag{B.34}$$

$$\mathbf{F}_{\boldsymbol{\beta}_{\parallel}\boldsymbol{\beta}_{\angle}} = \frac{2}{\sigma_n^2}\Re\left\{j\sum_{n=0}^{N-1}\mathrm{diag}(e^{-j\boldsymbol{\beta}_{\angle}})\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathcal{B}\right\} \tag{B.35}$$

$$\mathbf{F}_{\boldsymbol{\beta}_{\parallel}\boldsymbol{\theta}} = \frac{2}{\sigma_n^2}\Re\left\{j\sum_{n=0}^{N-1}\mathrm{diag}(e^{-j\boldsymbol{\beta}_{\angle}})\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\mathbf{U}\tilde{\tilde{\mathbf{A}}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathcal{B}\right\} \tag{B.36}$$

$$\mathbf{F}_{\boldsymbol{\beta}_{\|}\boldsymbol{\eta}_k} = \frac{2}{\sigma_n^2}\Re\left\{ j\sum_{n=0}^{N-1} \mathrm{diag}(e^{-j\boldsymbol{\beta}_{\angle}})\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H \cdot\{\mathbf{T}_k(n)\circ(\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathcal{B})\}\right\} \quad \text{(B.37)}$$

$$\mathbf{F}_{\boldsymbol{\beta}_{\angle}\boldsymbol{\beta}_{\angle}} = \frac{2}{\sigma_n^2}\Re\left\{\sum_{n=0}^{N-1}\mathcal{B}^H\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathcal{B}\right\} \quad \text{(B.38)}$$

$$\mathbf{F}_{\boldsymbol{\beta}_{\angle}\boldsymbol{\theta}} = \frac{2}{\sigma_n^2}\Re\left\{\sum_{n=0}^{N-1}\mathcal{B}^H\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\mathbf{U}\tilde{\tilde{\mathbf{A}}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathcal{B}\right\} \quad \text{(B.39)}$$

$$\mathbf{F}_{\boldsymbol{\beta}_{\angle}\boldsymbol{\eta}_k} = \frac{2}{\sigma_n^2}\Re\left\{\sum_{n=0}^{N-1}\mathcal{B}^H\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\{\mathbf{T}_k(n)\circ(\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathcal{B})\}\right\} \quad \text{(B.40)}$$

$$\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{2}{\sigma_n^2}\Re\left\{\sum_{n=0}^{N-1}\mathcal{B}^H\tilde{\tilde{\mathbf{A}}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\mathbf{U}^H\mathbf{U}\tilde{\tilde{\mathbf{A}}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathcal{B}\right\} \quad \text{(B.41)}$$

$$\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\eta}_k} = \frac{2}{\sigma_n^2}\Re\left\{\sum_{n=0}^{N-1}\mathcal{B}^H\tilde{\tilde{\mathbf{A}}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H\mathbf{U}^H\{\mathbf{T}_k(n)\circ(\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathcal{B})\}\right\} \quad \text{(B.42)}$$

$$\mathbf{F}_{\boldsymbol{\eta}_m\boldsymbol{\eta}_k} = \frac{2}{\sigma_n^2}\Re\left\{\sum_{n=0}^{N-1}\{\mathcal{B}^H(\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)^H)\circ\mathbf{T}_m(n)^H\} \cdot\{\mathbf{T}_k(n)\circ(\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathcal{B})\}\right\} \quad \text{(B.43)}$$

where

$$\tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n) \triangleq \mathbf{A}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathbf{G}(\boldsymbol{\varpi},n) \quad \text{(B.44)}$$

$$\tilde{\tilde{\mathbf{A}}}(\boldsymbol{\theta},\boldsymbol{\varpi},n) \triangleq \tilde{\mathbf{A}}(\boldsymbol{\theta},\boldsymbol{\varpi},n)\mathbf{P}(\boldsymbol{\theta},\boldsymbol{\varpi},n) \quad \text{(B.45)}$$

With Eqn. (B.7) and Eqns.(B.32)–(B.43), the FIM possesses the following block structure

$$\mathbf{F} = \begin{bmatrix} \boldsymbol{\mathcal{F}} & \mathbf{0} \\ \mathbf{0}^T & F_{\sigma_n^2\sigma_n^2} \end{bmatrix} \quad \text{(B.46)}$$

where

$$\boldsymbol{\mathcal{F}} = \begin{bmatrix} \mathbf{F}_{\boldsymbol{\beta}_{\|}\boldsymbol{\beta}_{\|}} & \mathbf{F}_{\boldsymbol{\beta}_{\|}\boldsymbol{\beta}_{\angle}} & \mathbf{F}_{\boldsymbol{\beta}_{\|}\boldsymbol{\theta}} & \mathbf{F}_{\boldsymbol{\beta}_{\|}\boldsymbol{\eta}_1} & \cdots & \mathbf{F}_{\boldsymbol{\beta}_{\|}\boldsymbol{\eta}_K} \\ \mathbf{F}^T_{\boldsymbol{\beta}_{\|}\boldsymbol{\beta}_{\angle}} & \mathbf{F}_{\boldsymbol{\beta}_{\angle}\boldsymbol{\beta}_{\angle}} & \mathbf{F}_{\boldsymbol{\beta}_{\angle}\boldsymbol{\theta}} & \mathbf{F}_{\boldsymbol{\beta}_{\angle}\boldsymbol{\eta}_1} & \cdots & \mathbf{F}_{\boldsymbol{\beta}_{\angle}\boldsymbol{\eta}_K} \\ \mathbf{F}^T_{\boldsymbol{\beta}_{\|}\boldsymbol{\theta}} & \mathbf{F}^T_{\boldsymbol{\beta}_{\angle}\boldsymbol{\theta}} & \mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\eta}_1} & \cdots & \mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\eta}_K} \\ \mathbf{F}^T_{\boldsymbol{\beta}_{\|}\boldsymbol{\eta}_1} & \mathbf{F}^T_{\boldsymbol{\beta}_{\angle}\boldsymbol{\eta}_1} & \mathbf{F}^T_{\boldsymbol{\theta}\boldsymbol{\eta}_1} & \mathbf{F}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_1} & \cdots & \mathbf{F}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_K} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}^T_{\boldsymbol{\beta}_{\|}\boldsymbol{\eta}_K} & \mathbf{F}^T_{\boldsymbol{\beta}_{\angle}\boldsymbol{\eta}_K} & \mathbf{F}^T_{\boldsymbol{\theta}\boldsymbol{\eta}_K} & \mathbf{F}^T_{\boldsymbol{\eta}_1\boldsymbol{\eta}_K} & \cdots & \mathbf{F}_{\boldsymbol{\eta}_K\boldsymbol{\eta}_K} \end{bmatrix} \tag{B.47}$$

Finally, the CRB could be obtained as follows

$$\text{CRB} \triangleq \mathbf{F}^{-1} = \begin{bmatrix} \boldsymbol{\mathcal{F}}^{-1} & \mathbf{0} \\ \mathbf{0}^T & \frac{1}{F_{\sigma_n^2\sigma_n^2}} \end{bmatrix} \tag{B.48}$$

where the second equality came from the partitioned matrix inversion formula [44].

Appendix C

Statistical Analysis for the Estimates of PPS frequency parameters in Array Setting

We analyze the asymptotic mean and variance of the estimated highest-order frequency parameter, which is the estimated frequency rate in case of LFM. Following that, we will analyze the asymptotic mean and variance of the estimated frequency in LFM. The analysis in this appendix is assumed focused on the single source case. The multicomponent or multiple sources case is too complicated to be analyzed. However, we can still make inferences from the result of single component case.

C.1 Statistical Analysis of Estimated Highest-order Frequency Parameters

Let us define

$$\delta a_{i,K}^{(m)} \triangleq \hat{a}_{i,K,(m)} - a_{i,K,(m)} \tag{C.1}$$

which is the error perturbation on the estimated highest-order (K^{th} order) frequency parameter of source i^{th} at sensor m (advert to Eqn. (3.3) for PPS signal model). We will drop notation i from here on, since we consider only single component. Under first-order perturbation analysis, we recall the explicit expression of (C.2) from [18], which is for single sensor case and being adopted to multi-sensor

case here*, as follows

$$\begin{aligned}\delta a_{K,(m)} \;\approx\; & \frac{12}{K!\tau^{K-1}L'(L'^2-1)\Delta^K} \\ & \cdot\Im\left\{\sum_{n=(K-1)\tau}^{N-1}((K-1)\tau+0.5L'-0.5-n)\eta^*_{(m)}(n)\right\}\end{aligned} \tag{C.2}$$

where $L \triangleq N-(K-1)\tau$,

$$\eta_{(m)}(n) \triangleq \prod_{q=0}^{K-1}\left[\sum_{i=0}^{\binom{K-1}{q}}\binom{\binom{K-1}{q}}{i}\left(\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{s^{\{q\}}(n-q\tau)[\mathbf{a}^{\{q\}}(\theta,n-q\tau)]_m}\right)^i\right]-1 \tag{C.3}$$

and $s(n)$ is the PPS given in Eqn. (3.3), $\mathbf{a}(\theta,n)$ is the wideband steering array given in Eqn. (3.4). We have also used the following notations: the even/odd-conjugation notation is given as

$$s^{\{q\}}(n) = \begin{cases} s(n) & , \quad q \quad \text{even} \\ s^*(n) & , \quad q \quad \text{odd} \end{cases} \tag{C.4}$$

and the binomial coefficient is given as $\binom{p}{q} = \frac{p!}{q!(p-q)!}$, as well as the m^{th} element of vector $\mathbf{v}(n)$ is $[\mathbf{v}(n)]_m$. Under high SNR, where SNR is defined as SNR $\triangleq A^2/\sigma_n^2$ with A is the PPS amplitude as defined in Eqn. (3.3), Eqn. (C.3) could be approximated as follows,

$$\eta_{(m)}(n) \approx \sum_{q=0}^{K-1}\binom{K-1}{q}\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{s^{\{q\}}(n-q\tau)[\mathbf{a}^{\{q\}}(\theta,n-q\tau)]_m} \tag{C.5}$$

Hence, under high SNR assumption, $E\{\eta_{(m)}(n)\} = 0$. Consequently, from Eqn. (C.2), we have

$$E\{\delta a_{K,(m)}\} = 0 \qquad \text{for all } m \tag{C.6}$$

which means the estimate of the K^{th} order frequency parameter is asymptotically unbiased. Similarly, by taking expectation of the squared Eqn. (C.2) and with Eqn. (C.5) substituted in Eqn. (C.2), we could obtain the covariance of the perturbation error as

$$E\{\delta a_{K,(m)}\delta a_{K,(l)}\} = \delta_{m,l}E\{(\delta a_{K,(m)})^2\} \tag{C.7}$$

*The notation $_{(m)}$ is appended here and it denotes the observation at the m^{th} sensor.

where

$$E\{(\delta a_{K,(m)})^2\} \approx \frac{6}{(K!)^2\tau^{2K-2}L'(L'^2-1)\Delta^{2K}\mathrm{SNR}}\binom{2K-2}{K-1}\mathbf{1}(L-1) - \frac{12}{(K!)^2\tau^{2K-2}L'^2(L'^2-1)^2\Delta^{2K}\mathrm{SNR}}\sum_{k=1}^{K-1}(-1)^{k+1}(L'-k\tau) \times(L'^2-2L'k\tau-2k^2\tau^2-1)\binom{2K-2}{K-1-k}\mathbf{1}(L'-k\tau-1), \quad \text{(C.8)}$$

$\delta_{m,l}$ is the Kronecker delta function (it equals to one if $m = l$, and zero otherwise) and $\mathbf{1}(n)$ is the discrete step function (it equals to one if $n \geq 0$, and zero ,otherwise). The derivation of Eqn. (C.8) is lengthy, hence, is not going to be reproduced here, but the details can be found in [18]. Note that, because the noise is spatially uncorrelated the $E\{\eta_{(m)}(n)\eta_{(l)}(n)\} = 0$ and $E\{\eta_{(m)}(n)\eta^*_{(l)}(n)\} = 0$. Consequently, the covariance of the perturbation error is also spatially uncorrelated as denoted by $\delta_{m,l}$ in Eqn. (C.7).

The asymptotic mean and variance (MSE) of the estimate derived above are under the assumption SNR is high and for any values of τ. However, it has been studied in [124] that the optimum value for τ, which produces smallest MSE of the estimate in Eqn. (C.8), is $\tau \approx N/K$. With these optimum values of τ, the authors in [18] have derived the MSE of the estimates without assuming high SNR.

Without assuming high SNR, the mean of Eqn. (C.3) is given by

$$\begin{aligned} E\{\eta_{(m)}(n)\} &= E\left\{\prod_{q=0}^{K-1}\left[\sum_{i=0}^{\binom{K-1}{q}}\binom{\binom{K-1}{q}}{i}\left(\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{s^{\{q\}}(n-q\tau)[\mathbf{a}^{\{q\}}(\theta,n-q\tau)]_m}\right)^i\right]\right\} - 1 \\ &= \prod_{q=0}^{K-1}\left[\sum_{i=0}^{\binom{K-1}{q}}\binom{\binom{K-1}{q}}{i}E\left\{\left(\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{s^{\{q\}}(n-q\tau)[\mathbf{a}^{\{q\}}(\theta,n-q\tau)]_m}\right)^i\right\}\right] - 1 \\ &= 1-1=0 \end{aligned} \quad \text{(C.9)}$$

where the equality in the second line above is due to $[\mathbf{v}^{\{q\}}(n-q\tau)]_m$ are statistically independent for different q's, and the equality in the third line is because for any i, $E\{\mathbf{v}^{\{q\}}(n-q\tau)]_m\} = 0$ except for $i = 0$ takes a value of 1. Hence, Eqn. (C.2) is also of zero mean and consequently the estimates of $a_{K,(m)}$ are unbiased.

Similarly, by using the fact that $[\mathbf{v}(n)]_m$ and $[\mathbf{v}(l)]_m$ are independent for $n \neq l$

and the facts that $E\{([\mathbf{v}(n)]_m)^i\} = 0$ and $E\{([\mathbf{v}(n)]_m)^i([\mathbf{v}^*(n)]_m)^i\} = i!\sigma_n^{2i}$ for all $i > 0$, using the results from [18] gives

$$\begin{aligned} E\{(\delta a_{K,(m)})^2\} &\approx -\frac{36K^{2K+4}}{(K!)^2N^{2K}(N^2-K^2)^2\Delta^{2K}} \\ &\quad \times \sum_{n=(K-1)\tau}^{N-1}\sum_{l=(K-1)\tau}^{N-1}((K-1)\tau+0.5L'-0.5-n) \\ &\quad \times ((K-1)\tau+0.5L'-0.5-l) \qquad \text{(C.10)} \\ &\quad \times \left(E\{\eta_{(m)}(n)\eta_{(m)}(l)\} - 2E\{\eta_{(m)}(n)\eta^*_{(m)}(l)\} + E\{\eta^*_{(m)}(n)\eta^*_{(m)}(l)\}\right) \\ &\approx -\frac{36K^{2K+4}}{(K!)^2N^{2K}(N^2-K^2)^2\Delta^{2K}} \\ &\quad \times \sum_{n=(K-1)\tau}^{N-1}((K-1)\tau+0.5L'-0.5-n)^2 \qquad \text{(C.11)} \\ &\quad \times \left(E\{[\eta_{(m)}(n)]^2\} - 2E\{\eta_{(m)}(n)\eta^*_{(m)}(n)\} + E\{[\eta^*_{(m)}(n)]^2\}\right) \end{aligned}$$

We have that

$$\begin{aligned} &E\{[\eta_{(m)}(n)]^2\} = \\ &E\left\{\left(\prod_{q=0}^{K-1}\left[\sum_{i=0}^{\binom{K-1}{q}}\binom{\binom{K-1}{q}}{i}\left(\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{s^{\{q\}}(n-q\tau)[\mathbf{a}^{\{q\}}(\theta,n-q\tau)]_m}\right)^i\right]-1\right)^2\right\} \\ &= 0 \qquad \text{(C.12)} \end{aligned}$$

and

$$
\begin{aligned}
&E\{\eta_{(m)}(n)\eta^{*}_{(m)}(n)\} = \\
&E\Bigg\{\Bigg(\prod_{q=0}^{K-1}\Bigg[\sum_{i=0}^{\binom{K-1}{q}}\binom{\binom{K-1}{q}}{i}\Bigg(\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{s^{\{q\}}(n-q\tau)[\mathbf{a}^{\{q\}}(\theta,n-q\tau)]_m}\Bigg)^{i}\Bigg]-1\Bigg) \\
&\times\Bigg(\prod_{q=0}^{K-1}\Bigg[\sum_{j=0}^{\binom{K-1}{q}}\binom{\binom{K-1}{q}}{j}\Bigg(\frac{[\mathbf{v}^{\{q+1\}}(n-q\tau)]_m}{s^{\{q+1\}}(n-q\tau)[\mathbf{a}^{\{q+1\}}(\theta,n-q\tau)]_m}\Bigg)^{j}\Bigg]-1\Bigg)\Bigg\} \\
&= E\Bigg\{\prod_{q=0}^{K-1}\sum_{i=0}^{\binom{K-1}{q}}\sum_{j=0}^{\binom{K-1}{q}}\binom{\binom{K-1}{q}}{j}\binom{\binom{K-1}{q}}{i}\Bigg(\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{s^{\{q\}}(n-q\tau)[\mathbf{a}^{\{q\}}(\theta,n-q\tau)]_m}\Bigg)^{i} \\
&\Bigg(\frac{[\mathbf{v}^{\{q+1\}}(n-q\tau)]_m}{s^{\{q+1\}}(n-q\tau)[\mathbf{a}^{\{q+1\}}(\theta,n-q\tau)]_m}\Bigg)^{j}\Bigg\}-1-1+1 \\
&= \mathcal{K}_{a_K}(K,\mathrm{SNR}) \qquad \text{(C.13)}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{K}_{a_K}(K,\mathrm{SNR}) &\triangleq \prod_{q=0}^{K-1}\sum_{i=0}^{\binom{K-1}{q}}\binom{\binom{K-1}{q}}{i}^{2}E\Bigg\{\Bigg(\frac{[\mathbf{v}^{\{q\}}(n-q\tau)]_m}{s^{\{q\}}(n-q\tau)[\mathbf{a}^{\{q\}}(\theta,n-q\tau)]_m}\Bigg)^{i} \\
&\quad\times\Bigg(\frac{[\mathbf{v}^{\{q+1\}}(n-q\tau)]_m}{s^{\{q+1\}}(n-q\tau)[\mathbf{a}^{\{q+1\}}(\theta,n-q\tau)]_m}\Bigg)^{i}\Bigg\}-1 \\
&= \Bigg[\prod_{q=0}^{K-1}\sum_{i=0}^{\binom{K-1}{q}}\binom{\binom{K-1}{q}}{i}^{2} i!\Bigg(\frac{1}{\mathrm{SNR}}\Bigg)^{i}\Bigg]-1 \qquad \text{(C.14)}
\end{aligned}
$$

Hence, by assuming $\tau = \frac{N}{K}$ and using summation of series equality given in Eqn. (3.23), and Eqn. (C.14), we can obtain MSE estimate of the highest order frequency parameter as follows,

$$
E\{(\delta a_{K,(m)})^2\} \approx \frac{6K^{2K+1}}{(K!)^2N^{2K-1}(N^2-K^2)\Delta^{2K}}\mathcal{K}_{a_K}(K,\mathrm{SNR}) \qquad \text{(C.15)}
$$

Considering only the second order PPS as in Chapter 3, we will derive the bias and MSE estimate of the frequency rate of the signal using the optimal value of $\tau = N/2$. Combining Eqns.(3.17) and (C.9) with $K = 2$, we obtain the bias of the

estimate as follows,

$$\begin{aligned} E\{\delta b\} &= \hat{b} - b \\ &= \frac{1}{M}\sum_{m=0}^{M-1} E(\nu_m + \delta\nu_m) - b \\ &= \frac{1}{M}\sum_{m=0}^{M-1} E\{\delta\nu_m\} \\ &= E\{\delta\nu_m\} = 0 \end{aligned} \tag{C.16}$$

where $\nu_m \triangleq a_{2,(m)}$ and $\delta\nu_m \triangleq \delta a_{2,(m)}$ are the frequency rate measured at sensor m and its perturbation error, respectively. Hence, the estimate of b is unbiased. Finally, by using Eqn. (C.15) with $K = 2$, the MSE of estimate of the frequency rate of the signal is straightforwardly given by

$$\begin{aligned} E\{(\delta b)^2\} &= E\left\{\left(\frac{1}{M}\sum_{m=0}^{M-1}\delta\nu_m\right)^2\right\} \\ &= \frac{1}{M^2}\sum_{m=0}^{M-1} E\{(\delta\nu_m)^2\} \\ &\approx \frac{48}{MN^3(N^2-4)\Delta^4}\mathcal{K}_{a_2}(2,\mathrm{SNR}) \\ &\approx \left(1+\frac{1}{2\mathrm{SNR}}\right)\frac{96}{MN^3(N^2-4)\Delta^4} \\ &\approx \left(1+\frac{1}{2\mathrm{SNR}}\right)\frac{96}{MN^5\Delta^4} \end{aligned} \tag{C.17}$$

C.2 Statistical Analysis of Estimated Initial Frequency Parameters

Statistical performance analysis of the estimated initial frequency for second-order discrete polynomial transform (DPT) algorithm in single sensor case has been derived in [125]. Let us recall from [125] the error perturbation analysis on the initial frequency estimate at the m^{th} sensor, $\delta\omega_m \triangleq \hat{\omega}_m - \omega_m$, which is adopted for

multi-sensors case, as follows

$$\delta\omega_m = \frac{12}{\Delta A^2 N^3}(\Im\{\vartheta_m\} - 8\Im\{\chi_m\}) \tag{C.18}$$

where A is the amplitude of PPS as defined in Eqn. (3.3) with the subscript i, which denotes i^{th} source, dropped. Similarly, the subscripts i in the observed initial frequency at sensor m^{th} and its estimate, ω_m and $\hat{\omega}_m$, are also dropped (see Definition 1 in Chapter 3 for the original definition). The other parameters, ϑ_m and χ_m, are defined as follows,

$$\vartheta_m \triangleq \sum_{n=1}^{N}(n - \frac{N}{2})s^*(n)[\mathbf{v}(n)]_m \tag{C.19}$$

$$\begin{aligned}\chi_m \triangleq\ & \kappa \sum_{n=1}^{N-\tau}\left[n - \frac{1}{2}(N-\tau)\right] \times \\ & (s^*(n+\tau)[\mathbf{v}(n)]_m + s(n)[\mathbf{v}^*(n+\tau)]_m + [\mathbf{v}^*(n+\tau)]_m[\mathbf{v}(n)]_m)\, e^{j2a_{2(m)}\tau\Delta^2 n}\end{aligned} \tag{C.20}$$

where $\kappa \triangleq e^{ja_1\tau\Delta + a_2\tau^2\Delta^2}$. The means and joint moments the last two parameters are given as follows,

$$E\{\chi_m\} = E\{\vartheta_m\} = 0 \tag{C.21}$$

$$E\{\chi_m\chi_l^*\} \approx \frac{1}{96}(2A^2 + \sigma_n^2)\sigma_n^2 N^3 \delta_{m,l} \tag{C.22}$$

$$E\{\vartheta_m\vartheta_l^*\} \approx \frac{1}{12}A^2\sigma_n^2 N^3 \delta_{m,l} \tag{C.23}$$

$$E\{\chi_m^2\} = E\{\vartheta_m^2\} = 0 \tag{C.24}$$

$$E\{\chi_m\vartheta_l^*\} = E\{\chi_m\vartheta_l\} \approx \frac{1}{96}A^2\sigma_n^2 N^3 \delta_{m,l} \tag{C.25}$$

where $\delta_{m,l}$ is being used again due to spatially uncorrelated noise assumption in our problem. Thus, by Eqn. (C.21), the initial frequency estimate at sensor m, $\hat{\omega}_m$, is unbiased and the mean of the vectorized errors in initial frequencies estimation, $\boldsymbol{\delta\omega} = [\delta\omega_0, \cdots, \delta\omega_{M-1}]^T$, is given by

$$E\{\boldsymbol{\delta\omega}\} = E\{\hat{\boldsymbol{\omega}}\} - \boldsymbol{\omega} = \mathbf{0} \tag{C.26}$$

where $\boldsymbol{\omega} = [\omega_0, \cdots, \omega_{M-1}]^T$.

The joint covariance of the perturbation error between two sensors could be ob-

tained by substituting Eqns.(C.22)–(C.25) into Eqn. (C.18) as follows,

$$E\{\delta\omega_m\delta\omega_n\} = E\{\delta\omega_m^2\}\delta_{m,n} \tag{C.27}$$

where the MSE of the initial frequency estimate at sensor m^{th} is given by

$$\begin{aligned}
E\{\delta\omega_m^2\} &= E\left\{\left[\frac{6}{\Delta A^2 N^3}2(\Im\{\vartheta_m\} - 8\Im\{\chi_m\})\right]^2\right\} \\
&\approx \frac{36}{\Delta^2 A^4 N^6} E\left\{[-j(\vartheta_m - \vartheta_m^*) + j8(\chi_m - \chi_m^*)]^2\right\} \\
&\approx \frac{36}{\Delta^2 A^4 N^6} E\left\{-(\vartheta_m - \vartheta_m^*)^2 - 8(\chi_m - \chi_m^*)^2 + 2(\vartheta_m - \vartheta_m^*)8(\chi_m - \chi_m^*)\right\} \\
&\approx \frac{36}{\Delta^2 A^4 N^6}\left[E\left\{2\vartheta_m\vartheta_m^*\right\} + E\left\{128\chi_m\chi_m^*\right\}\right] \\
&\approx \left(\frac{17}{16} + \frac{1}{2\text{SNR}}\right)\frac{96}{\text{SNR}\Delta^2 N^3}
\end{aligned} \tag{C.28}$$

Hence, the covariance matrix of the perturbation error is given as,

$$E\{\boldsymbol{\delta\omega}\boldsymbol{\delta\omega}^T\} \approx \left(\frac{17}{16} + \frac{1}{2\text{SNR}}\right)\frac{96}{\text{SNR}\Delta^2 N^3}\mathbf{I} = \mathcal{K}_\omega(N, \text{SNR})\mathbf{I} \tag{C.29}$$

where $\mathcal{K}_\omega(N, \text{SNR}) \triangleq \left(\frac{17}{16} + \frac{1}{2\text{SNR}}\right)\frac{96}{\text{SNR}\Delta^2 N^3}$.

Finally, given the perturbation error of the initial frequency at each sensor, we can now derive the perturbation error of the initial frequency of signal a_i. Looking at Eqn. (3.29), the perturbation error of the initial frequency, a_i, is given by the first row of $\hat{\mathcal{A}}^\dagger\boldsymbol{\delta\omega_i}$. To simplify notation, we drop the subscript $_i$. Using the result in Eqn. (C.28) and Eqn. (3.24), we get

$$\hat{\mathcal{A}}^\dagger\boldsymbol{\delta\omega} = \frac{1}{\varrho}\begin{bmatrix} \frac{2}{3}\hat{b}^2(M-1)M(2M-1) & -\hat{b}M(M-1) \\ -\hat{b}M(M-1) & M \end{bmatrix}\begin{bmatrix} \mathbf{1}^T \\ 2\hat{b}\boldsymbol{\zeta}^T \end{bmatrix}\boldsymbol{\delta\omega} \tag{C.30}$$

Thus, the perturbation error of the initial frequency of the signal is given by the

first row of Eqn. (C.30) above as follows

$$\begin{aligned}
\delta a &= \frac{1}{\varrho}\left[\begin{array}{cc} \frac{2}{3}\hat{b}^2(M-1)M(2M-1) & -\hat{b}M(M-1) \end{array}\right]\left[\begin{array}{c} \mathbf{1}^T\boldsymbol{\delta\omega} \\ 2\hat{b}\boldsymbol{\zeta}^T\boldsymbol{\delta\omega} \end{array}\right] \\
&= \frac{3}{\hat{b}^2M^2(M^2-1)}\left[\frac{2}{3}\hat{b}^2(M-1)M(2M-1)\mathbf{1}^T\boldsymbol{\delta\omega} - 2\hat{b}^2M(M-1)\boldsymbol{\zeta}^T\boldsymbol{\delta\omega}\right] \\
&= \frac{3}{M(M+1)}\left[\frac{2}{3}(2M-1)\mathbf{1}^T\boldsymbol{\delta\omega} - 2\boldsymbol{\zeta}^T\boldsymbol{\delta\omega}\right] \\
&= \frac{6}{M(M+1)}\left[\frac{(2M-1)}{3}\sum_{m=0}^{M-1}\delta\omega_m - \sum_{m=0}^{M-1}m\delta\omega_m\right] \qquad \text{(C.31)}
\end{aligned}$$

Subsequently, by the Eqn. (C.26), the bias of initial frequency estimate of the source signal is given by

$$E\{\delta a\} = \frac{6}{M(M+1)}\left[\frac{(2M-1)}{3}\sum_{m=0}^{M-1}E\{\delta\omega_m\} - \sum_{m=0}^{M-1}mE\{\delta\omega_m\}\right] = 0 \qquad \text{(C.32)}$$

Using Eqn. (C.29), the MSE of $\hat{a}$ is, therefore, given by

$$\begin{aligned}
E\{\delta a^2\} &= \frac{36}{M^2(M+1)^2}E\left\{\left[\sum_{m=0}^{M-1}\left(\frac{(2M-1)}{3}-m\right)\delta\omega_m\right]^2\right\} \\
&= \frac{36}{M^2(M+1)^2}\left[\sum_{m=0}^{M-1}\left(\frac{(2M-1)}{3}-m\right)^2 E\left\{\delta\omega_m^2\right\}\right] \\
&\approx \frac{36\mathcal{K}_\omega(N,\mathrm{SNR})}{M^2(M+1)^2}\sum_{m=0}^{M-1}\frac{(2M-1)^2}{9} - 2m\frac{(2M-1)}{3} + m^2 \qquad \text{(C.33)}
\end{aligned}$$

By using summation of series equalities in Eqn. (3.23), we get

$$\begin{aligned}
E\{\delta a^2\} &\approx \mathcal{K}_\omega(N,\mathrm{SNR})\frac{2(2M-1)}{M(M+1)} \\
&\approx \left(\frac{17}{16}+\frac{1}{2\mathrm{SNR}}\right)\frac{192(2M-1)}{\mathrm{SNR}M(M+1)\Delta^2N^3} \qquad \text{(C.34)}
\end{aligned}$$

In summary, the proposed estimation method in Chapter 3, which uses the second-order DPT, gives unbiased estimates for frequency rate and initial frequency. Their MSE are given by Eqns.(C.17) and (C.34), respectively.

Appendix D

Statistical Analysis for DOA Estimate of DPT-based Algorithm

We analyze the asymptotic performance of the proposed algorithm presented in Chapter 3 in estimating DOA. Firstly, since DOA estimation method in Eqn. (3.32) is given by the maxima of a random function, we will review the perturbation analysis for maxima of random functions in general. Secondly, first-order perturbation analysis of the non-parametric estimate of source k's data is detailed. Finally, based on perturbation analysis on the maxima of random functions, we derive the perturbation analysis of the proposed DOA estimation in Chapter 3.

D.1 First Order Perturbation Analysis of Maxima of Random Functions

Let $g(\psi)$ be a complex function of a real variable ψ, and

$$f(\psi) \triangleq |g(\psi)|^2 = g(\psi)g^*(\psi). \tag{D.1}$$

Supposed that $f(\psi)$ achieves global maximum at $\psi = \psi_k$. Let us denote the random perturbation of $g(\psi)$ as $\delta g(\psi)$, which is relatively small, i.e.,

$$\lim_{N\to\infty} \left| \frac{\delta g(\psi)}{g(\psi_k)} \right| = 0 \quad \text{with probability 1, for all } \psi \tag{D.2}$$

Note that the dependency of the functions $f(\psi)$, $g(\psi)$ and $\delta g(\psi)$ on the number of samples N will not be explicitly expressed hereafter to simplify the notations.

Since ψ_k is a maximum point, we have

$$\left. \frac{\partial f(\psi)}{\partial \psi} \right|_{\psi=\psi_k} = 2\Re \left\{ g(\psi_k) \frac{\partial g(\psi_k)}{\partial \psi} \right\} = 0 \tag{D.3}$$

If the random perturbation function $\delta g(\psi)$ is added to $g(\psi)$, then the global maxima's point will shift accordingly, say to $\psi_k + \delta\psi$. Hence,

$$\left[\frac{\partial f(\psi)}{\partial \psi} + \frac{\partial \delta f(\psi)}{\partial \psi} \right]_{\psi=\psi_k+\delta\psi} = 0 \tag{D.4}$$

The first-order perturbation $\delta f(\psi)$ is given by

$$\begin{aligned} \delta f(\psi) &\approx g(\psi)\delta g^*(\psi) + g^*(\psi)\delta g(\psi) \\ &= 2\Re\{g(\psi)\delta g^*(\psi)\} \end{aligned} \tag{D.5}$$

By applying first-order Taylor expansion of Eqn. (D.4) about ψ_k we get

$$\frac{\partial f(\psi_k)}{\partial \psi} + \frac{\partial^2 f(\psi_k)}{\partial \psi^2}\delta\psi + \frac{\partial \delta f(\psi_k)}{\partial \psi} \approx 0 \tag{D.6}$$

By rearranging Eqn. (D.6) and by substitution of Eqn. (D.3), we obtain the perturbation error of the parameter estimate,

$$\delta\psi \approx -\left[\frac{\partial^2 f(\psi_k)}{\partial \psi^2} \right]^{-1} \frac{\partial \delta f(\psi_k)}{\partial \psi} \tag{D.7}$$

where

$$\frac{\partial^2 f(\psi_k)}{\partial \psi^2} = 2\Re \left\{ g(\psi_k) \frac{\partial^2 g^*(\psi_k)}{\partial \psi^2} + \frac{\partial g(\psi_k)}{\partial \psi} \frac{\partial g^*(\psi_k)}{\partial \psi} \right\} \tag{D.8}$$

$$\frac{\partial \delta f(\psi_k)}{\partial \psi} = 2\Re \left\{ g(\psi_k) \frac{\partial \delta g^*(\psi_k)}{\partial \psi} + \frac{\partial g(\psi_k)}{\partial \psi} \delta g^*(\psi_k) \right\} \tag{D.9}$$

Therefore, the bias and the mean square error of the estimate are given by

$$E\{\delta\psi\} \approx -\left[\frac{\partial^2 f(\psi_k)}{\partial\psi^2}\right]^{-1} E\left\{\frac{\partial\delta f(\psi_k)}{\partial\psi}\right\} \tag{D.10}$$

and

$$E\{(\delta\psi)^2\} \approx -\left[\frac{\partial^2 f(\psi_k)}{\partial\psi^2}\right]^{-2} E\left\{\left[\frac{\partial\delta f(\psi_k)}{\partial\psi}\right]^2\right\}, \tag{D.11}$$

respectively.

D.2 First Order Perturbation Analysis of Non-parametric Estimate of k^{th} Source's Data

In this section, we will first analyze the first-order perturbation analysis of $\hat{\mathbf{x}}_k(n)$, the non-parametric estimate of the source k's data, (see Eqn. (3.32)), which will be used in the next section. Meanwhile, assuming there are only two sources, i.e. source k and l, then the non-parametric estimate of the source l's data obtained by using Eqns.(3.33), (3.34) and (3.35) is summarized as follows,

$$[\hat{\mathbf{x}}_l(n)]_m = [\mathbf{x}_l(n)]_m + [\mathbf{x}_k(n)]_m + [\mathbf{v}(n)]_m - \frac{1}{N}\sum_{n'}\{[\mathbf{x}(n')]_m + [\mathbf{v}(n')]_m\} e^{-j\tilde{\varphi}_k(n')} e^{j\tilde{\varphi}_k(n)} \tag{D.12}$$

where

$$\tilde{\varphi}_l(n) = \hat{b}_l(\Delta n)^2 + \hat{\omega}_{m,l}\Delta n \tag{D.13}$$

and $\hat{b}_l$ and $\hat{\omega}_{m,l}$ are the estimated frequency rate of the source signal l and the estimated initial frequency of source signal l for m^{th} sensor, respectively (refer to Definition 1 in Chapter 3). Note that, if we use the non-parametric estimate of source l's data to get the non-parametric estimate of source k, we will obtain the

following,

$$\begin{aligned}
[\hat{\mathbf{x}}_k(n)]_m &= [\mathbf{x}_k(n)]_m + [\mathbf{x}_l(n)]_m + [\mathbf{v}(n)]_m - \frac{1}{N}\sum_{n'}[\hat{\mathbf{x}}_l(n')]_m e^{-j\tilde{\varphi}_l(n')}e^{j\tilde{\varphi}_l(n)} \\
&= [\mathbf{x}_k(n)]_m + [\mathbf{x}_l(n)]_m + [\mathbf{v}(n)]_m \\
&\quad -\frac{1}{N}\sum_{n'}\{[\mathbf{x}_l(n')]_m + [\mathbf{x}_k(n')]_m + [\mathbf{v}(n')]_m\}e^{-j\tilde{\varphi}_l(n')}e^{j\tilde{\varphi}_l(n)} \\
&\quad +\frac{1}{N^2}\sum_{n''}\sum_{n'}\{[\mathbf{x}(n')]_m + [\mathbf{v}(n')]_m\}e^{-j\tilde{\varphi}_k(n')}e^{j\tilde{\varphi}_k(n'')}e^{-j\tilde{\varphi}_l(n'')}e^{j\tilde{\varphi}_l(n)}
\end{aligned}$$

$$[\hat{\mathbf{x}}_k(n)]_m \approx [\mathbf{x}_k(n)]_m + [\mathbf{x}_l(n)]_m + [\mathbf{v}(n)]_m - \frac{1}{N}\sum_{n'}[\mathbf{x}_l(n')]_m e^{-j\tilde{\varphi}_l(n')}e^{j\tilde{\varphi}_l(n)} \quad \text{(D.14)}$$

The last approximation gives the estimated source l's data, $\hat{\mathbf{x}}_l(n')$, is assumed to be approximately equal to its source data. This assumption is reasonable for large value of N. For large N, the following terms in the second equality of Eqn. (D.14), i.e. $\frac{1}{N^2}\sum_{n'',n'}[\mathbf{x}(n')]_m e^{-j\tilde{\varphi}_k(n')}e^{j\tilde{\varphi}_k(n'')}e^{-j\tilde{\varphi}_l(n'')}e^{j\tilde{\varphi}_l(n)}$ and $\frac{1}{N}\sum_{n'}[\mathbf{x}_k(n')]_m e^{-j\tilde{\varphi}_l(n')}e^{j\tilde{\varphi}_l(n)}$, diminish asymptotically. This is the result of applying the Absolute Convergence Test using the facts that the absolute values of the terms inside these two sums, i.e., $\left|[\mathbf{x}_k(n')]_m e^{-j\tilde{\varphi}_l(n')}\right|$ and $\left|[\mathbf{x}(n')]_m e^{-j\tilde{\varphi}_k(n')}e^{j\tilde{\varphi}_k(n'')}e^{-j\tilde{\varphi}_l(n'')}e^{j\tilde{\varphi}_l(n)}\right|$ are finite and the fact that $\frac{1}{N} \to 0$ and $\frac{1}{N^2} \to 0$ for large N.

By the Weak Law of Large Numbers, the following random variable terms in the second equality of Eqn. (D.14) are given by,

$$\frac{1}{N}\sum_{n'}[\mathbf{v}(n')]_m e^{-j\tilde{\varphi}_l(n')}e^{j\tilde{\varphi}_l(n)} \xrightarrow{p} 0 \quad \text{(D.15)}$$

$$\frac{1}{N^2}\sum_{n'',n'}[\mathbf{v}(n')]_m e^{-j\tilde{\varphi}_k(n')}e^{j\tilde{\varphi}_k(n'')}e^{-j\tilde{\varphi}_l(n'')}e^{j\tilde{\varphi}_l(n)} \xrightarrow{p} 0 \quad \text{(D.16)}$$

i.e., converge in law (probability) to zeros as $N \to \infty$. By rearranging the Eqn. (D.14), we obtain

$$[\delta\mathbf{x}_k(n)]_m \triangleq [\hat{\mathbf{x}}_k(n)]_m - [\mathbf{x}_k(n)]_m \approx [\mathbf{x}_l(n)]_m - \frac{1}{N}\sum_{n'}[\mathbf{x}_l(n')]_m e^{-j\tilde{\varphi}_l(n')}e^{j\tilde{\varphi}_l(n)} + [\mathbf{v}(n)]_m \quad \text{(D.17)}$$

If there are more than two sources, the equation above could be simply modified

to

$$[\delta \mathbf{x}_k(n)]_m \approx \sum_{\substack{l=1 \\ l \neq k}}^{L} \left\{ [\mathbf{x}_l(n)]_m - \frac{1}{N} \sum_{n'} [\mathbf{x}_l(n')]_m e^{-j\tilde{\varphi}_l(n')} e^{j\tilde{\varphi}_l(n)} \right\} + [\mathbf{v}(n)]_m \quad \text{(D.18)}$$

where L is the number of sources.

Note that $\sum_{n'} [\mathbf{x}_l(n')]_m e^{-j\tilde{\varphi}_l(n')}$ is intuitively estimate of $A_l e^{j\phi_{m,l}}$, mathematically

$$\hat{A}_l e^{j\hat{\phi}_{m,l}} = \frac{1}{N} \sum_{n'} [\mathbf{x}_l(n')]_m e^{-j\tilde{\varphi}_l(n')} \quad \text{(D.19)}$$

because, $e^{-j\tilde{\varphi}_l(n')}$ removes the second order phase and first-order phase using the estimates of frequency rate and initial frequency and $\frac{1}{N}\sum_{n'}$ is just sample averaging that is applied to complex variables that approximately contain only the zeroth order phase. The bias of the estimates, $\hat{A}_l$ and $\hat{\phi}_{m,l}$, are asymptotically zeros (for derivation see [125]), mathematically

$$E\{\delta A_l\} \triangleq E\{\hat{A}_l - A_l\} = 0 \qquad \text{and} \qquad E\{\delta \phi_{m,l}\} \triangleq E\{\hat{\phi}_{m,l} - \phi_{m,l}\} = 0 \quad \text{(D.20)}$$

Therefore, the first two terms on the right-hand side of Eqn. (D.18) give the estimation error between source l's data, $[\mathbf{x}_l(n)]_m$, and its estimate, $\hat{A}_l e^{j\hat{\phi}_{m,l}} e^{j\tilde{\varphi}_l(n)}$, which is constructed by the estimated parameters, i.e., $\hat{A}_l, \hat{\phi}_{m,l}, \hat{\omega}_{m,l}$ and $\hat{b}_l$. Hence, we can approximate these two terms with first-order perturbation analysis of estimation error of source l's data, as follows,

$$\frac{1}{N}\sum_{n'} [\mathbf{x}_l(n')]_m e^{-j\{\tilde{\varphi}_l(n') + \tilde{\varphi}_l(n)\}} - [\mathbf{x}_l(n)]_m \approx \left(\frac{\delta A_l}{A_l} + j(\delta \hat{\phi}_{m,l} + \delta b_l (\Delta n)^2 + \delta \omega_{l,m} \Delta n) \right) [\mathbf{x}_l(n)]_m \quad \text{(D.21)}$$

where

$$E\{\frac{1}{N}\sum_{n'} [\mathbf{x}_l(n')]_m e^{-j\{\tilde{\varphi}_l(n') + \tilde{\varphi}_l(n)\}} - [\mathbf{x}_l(n)]_m\} \approx 0 \qquad N \to \infty \quad \text{(D.22)}$$

as a result of Eqns.(D.20), (C.32) and (C.16). Subsequently, we have unbiased estimate of source k's data

$$E\{[\delta \mathbf{x}_k(n)]_m\} \approx 0 \quad \text{(D.23)}$$

as a result of Eqn. (D.22) and $E\{[\mathbf{v}(n)]_m\} = 0$.

D.3 First Order Perturbation Analysis of DOA Estimate

In the first section of this appendix, we have seen the first-order perturbation analysis of maxima of random functions. Herein this section, we will apply it to DOA estimate. Observing Eqn. (3.32), we identified that

$$g(\psi) = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}[\mathbf{x}_k(n)]_m^* e^{j\{\psi\varphi_k'(n)m+\varphi_k(n)\}} \tag{D.24}$$

where $\psi \triangleq \frac{d}{c}\sin\theta$, $\varphi_k'(n) \triangleq a_k + 2b_k\Delta n$ and $\varphi_k(n) \triangleq a_k\Delta n + b_k(\Delta n)^2$. Hence, by evaluating $g(\psi)$ at $\psi = \psi_k$, we obtain the following

$$g(\psi_k) = MA_k e^{-j\alpha_k} \tag{D.25}$$

because $[\mathbf{x}_k(n)]_m^* e^{j\{\psi_k\varphi_k'(n)m+\varphi_k(n)\}} = A_k e^{-j\alpha_k}$. Subsequently, the first-order error perturbation on $g(\psi)$ is given as

$$\begin{aligned}\delta g(\psi) \;=\; & \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}[\delta\mathbf{x}_k(n)]_m^* e^{j\{\psi\varphi_k'(n)m+\varphi_k(n)\}} \\ & +[\mathbf{x}_k(n)]_m^* j\left[(2\psi\Delta mn + (\Delta n)^2)\delta b_k + (\psi m + \Delta n)\delta a_k\right] e^{j\{\psi\varphi_k'(n)m+\varphi_k(n)\}}\end{aligned} \tag{D.26}$$

and, for $\psi = \psi_k$, simplified to

$$\delta g(\psi_k) = \frac{A_k e^{-j\alpha_k}}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\frac{[\delta\mathbf{x}_k(n)]_m^*}{[\mathbf{x}_k(n)]_m^*} + j\left[(2\psi\Delta mn + (\Delta n)^2)\delta b_k + (\psi m + \Delta n)\delta a_k\right] \tag{D.27}$$

The first-order partial derivative of $g(\psi)$ with respect to ψ is given by

$$\frac{\partial g(\psi)}{\partial\psi} = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}[\mathbf{x}_k(n)]_m^* j\varphi_k'(n)m e^{j\{\psi\varphi_k'(n)m+\varphi_k(n)\}} \tag{D.28}$$

By evaluating Eqn. (D.28) at $\psi = \psi_k$ and using Eqn. (3.23), we obtain

$$\frac{\partial g(\psi_k)}{\partial \psi} = jA_k e^{-j\alpha_k} \frac{M(M-1)(a_k + b_k\Delta(N-1)}{2} \tag{D.29}$$

Next, the second order partial derivative of $g(\psi)$ with respect to ψ is given by

$$\frac{\partial^2 g(\psi)}{\partial \psi^2} = -\frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}[\mathbf{x}_k(n)]_m^*(\varphi_k'(n))^2 m^2 e^{j\{\psi\varphi_k'(n)m+\varphi_k(n)\}} \tag{D.30}$$

Again by evaluating it at $\psi = \psi_k$ and using the series summation formulas in Eqn. (3.23), we obtain

$$\frac{\partial^2 g(\psi_k)}{\partial \psi^2} = -\frac{A_k e^{-j\alpha_k}(M-1)M(2M-1)\{3a_k^2+6a_k b_k\Delta(N-1)+2b_k^2\Delta^2(N-1)(2N-1)\}}{18} \tag{D.31}$$

Therefore, Eqn. (D.8) for DOA estimation is derived, by using Eqns.(D.25), (D.29), (D.31) and (3.23), as follows

$$\begin{aligned}\frac{\partial^2 f(\psi_k)}{\partial \psi^2} &= \frac{A_k^2 M^2(M-1)}{6}\Big[(7M-5)a_k^2 + 2(7M-5)a_k b_k\Delta(N-1) \\ &\quad +\frac{1}{3}(25NM-17N-17M+13)(N-1)b_k^2\Delta^2\Big]\end{aligned} \tag{D.32}$$

which could be approximated by

$$\frac{\partial^2 f(\psi_k)}{\partial \psi^2} \approx \frac{A_k^2 M^2(M-1)}{6}\left[(7M-5)a_k^2+2(7M-5)a_k b_k\Delta N+\frac{1}{3}(25M-17)N^2 b_k^2\Delta^2\right] \tag{D.33}$$

for $N \gg M$, i.e. the number of samples is much greater than the number of sensors (which is typically the case in practice). The partial derivative of the perturbation error $\delta g(\psi)$ with respect to ψ is given by

$$\begin{aligned}\frac{\partial \delta g(\psi)}{\partial \psi} &= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\Bigg\{[\delta\mathbf{x}_k(n)]_m^* e^{j\{\psi\varphi_k'(n)m+\varphi_k(n)\}} j\varphi_k'(n)m \\ &+jm[\mathbf{x}_k(n)]_m^*\big[2\Delta n\delta b_k+\delta a_k+j\varphi_k'(n)\big[(2\psi\Delta mn+(\Delta n)^2)\delta b_k+(\psi m+\Delta n)\delta a_k\big]\big] \\ &\times e^{j\{\psi\varphi_k'(n)m+\varphi_k(n)\}}\Bigg\}\end{aligned} \tag{D.34}$$

and for $\psi = \psi_k$ it is simplified to

$$\frac{\partial \delta g(\psi_k)}{\partial \psi} = \frac{A_k e^{-j\alpha_k}}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left\{ \frac{[\delta \mathbf{x}_k(n)]_m^*}{[\mathbf{x}_k(n)]_m^*} j\varphi_k'(n)m \right. \tag{D.35}$$
$$\left. + jm(2\Delta n \delta b_k + \delta a_k) - m\varphi_k'(n)\left[(2\psi\Delta mn + (\Delta n)^2)\delta b_k + (\psi m + \Delta n)\delta a_k\right] \right\}$$

Therefore, Eqn. (D.9) for DOA estimation is derived, by using Eqns.(D.25), (D.29), (D.27), (D.35) and (3.23), and by assuming $N \gg M$,

$$\begin{aligned} \frac{\partial \delta f(\psi_k)}{\partial \psi} \quad &\approx \quad \frac{A_k^2 M}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \frac{[\delta \mathbf{x}_k(n)]_m^*}{[\mathbf{x}_k(n)]_m^*} \left\{ \frac{M-1}{2}(a_k + b_k \Delta N) - (a_k + 2b_k \Delta n)m \right\} \\ &\quad - \mathcal{K}_a \delta a_k - \mathcal{K}_b \delta b_k \end{aligned} \tag{D.36}$$

where

$$\mathcal{K}_a \triangleq \frac{A_k^2 M^2 (M-1)}{12} \left[a_k \psi_k (M+1) + b_k \psi_k \Delta N (M+1) - b_k \Delta^2 N^2 \right] \tag{D.37}$$

$$\mathcal{K}_b \triangleq \frac{A_k^2 M^2 (M-1) N \Delta}{36} \left[3a_k \psi_k (M+1) + b_k \psi_k \Delta N (7M+1) - 3b_k \Delta^2 N^2 \right] \tag{D.38}$$

By Eqn. (C.32) and (C.16), the mean of Eqn. (D.36) is simplified to

$$E\left\{ \frac{\partial \delta f(\psi_k)}{\partial \psi} \right\} \approx \frac{A_k^2 M}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \frac{E\{[\delta \mathbf{x}_k(n)]_m^*\}}{[\mathbf{x}_k(n)]_m^*} \left\{ \frac{M-1}{2}(a_k + b_k \Delta N) - (a_k + 2b_k \Delta n)m \right\}. \tag{D.39}$$

However, due to Eqn. (D.23),

$$E\left\{ \frac{\partial \delta f(\psi_k)}{\partial \psi} \right\} = 0 \tag{D.40}$$

Subsequently, the bias of the estimate of ψ is given by Eqn. (D.10) for large N, as follows

$$E\{\delta\psi\} \approx -\left[\frac{\partial^2 f(\psi_k)}{\partial \psi^2} \right]^{-1} E\left\{ \frac{\partial \delta f(\psi_k)}{\partial \psi} \right\} = 0 \tag{D.41}$$

and hence $\hat{\psi}_k$ is asymptotically unbiased. By Taylor expansion, we have $\delta\psi \approx \frac{d}{c} \cos\theta_k \delta\theta$, which implies that $\hat{\theta}_k$ is also asymptotically unbiased.

The analysis of the MSE of the estimate of ψ is not performed here because of the

complexity in calculating the cross correlations between many error perturbation parameters.

Appendix E

Joint Approximate Diagonalization Algorithm

In this appendix we review the joint diagonalization algorithm [26] for 2×2 matrices only. However, it can be easily extended to square matrices of any size in an analogous way as Jacobi technique (see [70]). Suppose that we want to diagonalize a set matrices $\mathcal{Q}^{uw} = \{\mathbf{G}^{(r)}, 1 \leq r \leq L'\}$ with the entries,

$$\mathbf{G}^{(r)} = \begin{bmatrix} a_r & b_r \\ c_r & d_r \end{bmatrix} \tag{E.1}$$

The objective of the joint diagonalization is to get unitary matrix $\mathbf{V}$ such that $\mathbf{G}'^{(r)} = \mathbf{V}^H \mathbf{G}^{(r)} \mathbf{V}$ is as diagonal as possible. Mathematically, it is the same as maximizing $\mathcal{C}$, where

$$\mathcal{C} \triangleq \sum_r |a'_r|^2 + |d'_r|^2 \stackrel{\ddagger}{=} \frac{1}{2} \sum_r |a'_r - d'_r|^2 \tag{E.2}$$

where a'_r, d'_r are the diagonal elements of $\mathbf{G}'^{(r)}$. Equality ‡ above is due to the invariance of the trace of $\mathbf{G}'^{(r)}$. The joint diagonalization uses complex Givens rotation technique,

$$\mathbf{V} = \begin{bmatrix} \cos\theta & -e^{j\phi}\sin\theta \\ e^{-j\phi}\sin\theta & \cos\theta \end{bmatrix}$$

and hence $\mathcal{C}$ relates the new parameters to the old ones as follows,

$$a'_r - d'_r = (a_r - d_r)\cos 2\theta + (b_r + c_r)\sin 2\theta \cos\phi + j(c_r - b_r)\sin 2\theta \sin\phi \quad \text{(E.3)}$$

for $r = 1, \ldots, L'$.

Let us define the following notations,

$$\begin{aligned}
\mathbf{u} &\triangleq [a'_1 - d'_1, \ldots, a'_{L'} - d'_{L'}]^T && \text{(E.4)}\\
\mathbf{v} &\triangleq [\cos 2\theta, \sin 2\theta \cos\phi, \sin 2\theta \sin\phi]^T && \text{(E.5)}\\
\mathbf{g}_r &\triangleq [a_r - d_r, b_r + c_r, j(c_r - b_r)]^T && \text{(E.6)}\\
\mathbf{G}^T &\triangleq [\mathbf{g}_1, \ldots, \mathbf{g}_{L'}] && \text{(E.7)}
\end{aligned}$$

With these definitions, we can rewrite Eqn. (E.3) as $\mathbf{u} = \mathbf{G}\mathbf{v}$. Therefore, the cost function, $\mathcal{C}$, in Eqn. (E.2) can be rewritten as follows

$$\mathcal{C} = \mathbf{u}^H\mathbf{u} = \mathbf{v}^T\mathbf{G}^H\mathbf{G}\mathbf{v} = \mathbf{v}^T\Re\{\mathbf{G}^H\mathbf{G}\}\mathbf{v} \quad \text{(E.8)}$$

where the last equality of the Eqn. (E.8) is due to the fact that $\mathbf{G}^H\mathbf{G}$ is Hermitian by construction, which means its imaginary part is antisymmetric, and hence contributes nothing to the above quadratic form.

Finally, since $\|\mathbf{v}\| = 1$, finding $\mathbf{v}$ that maximize Eqn. (E.8) is equivalent to solving for eigenvector that corresponds to largest eigenvalue of $\Re\{\mathbf{G}^H\mathbf{G}\}$. With this optimum $\mathbf{v} \triangleq [v_1, v_2, v_3]^T$, we can get the entries of $\mathbf{V}$, without the need to find θ and ϕ, as follows

$$\begin{aligned}
\cos\theta &= \sqrt{\frac{1+v_1}{2}} && \text{(E.9)}\\
e^{-j\phi}\sin\theta &= \frac{v_2 - jv_3}{2\cos\theta} = \frac{v_2 - jv_3}{\sqrt{2(1+v_1)}} && \text{(E.10)}
\end{aligned}$$

To get Eqn. (E.9), one could use $\cos 2\theta = 1 + 2\cos^2\theta$, and to get Eqn. (E.10) one could use $e^{-j\phi} = \cos\phi - j\sin\phi$ and $\sin 2\theta = 2\sin\theta\cos\theta$.

The algorithm could be summarized as follows:

1. Construct $\Re\{\mathbf{G}^H\mathbf{G}\}$ by using Eqns.(E.6) and (E.7)
2. Find eigenvector, $\mathbf{v} = [v_1, v_2, v_3]^T$, that corresponds to the largest eigenvalue

of $\Re\{\mathbf{G}^H\mathbf{G}\}$

3. Form $\mathbf{V}$ by its entries by using Eqns.(E.9) and (E.10).

S. Lie, A. R. Leyman and Y. H. Chew, "Underdetermined source separation for non-stationary signal," *The 32nd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Hawaii, USA, April 2007.

S. Lie, A. R. Leyman and Y. H. Chew, "Fourth-order and weighted mixed order direction of arrival estimators," *IEEE Signal Processing Letters*, vol. 13, no.11, Nov 2006.

S. Lie, A. R. Leyman and Y. H. Chew, "Wideband polynomial-phase parameter estimation in sensor array," in *Proc. of the 3rd IEEE International Symposium on Sign. Proc. and Info. Tech.*, Darmstadt, Germany, Dec 2003.

S. Lie, A. R. Leyman and Y. H. Chew, "Parameter estimation of wideband chirp signals in sensor arrays through DPT," in *Proc. 37th Asilomar Conf. on Sign., Syst. and Comp.*, Pacific Grove, CA, Nov. 2003.

S. Lie, A. R. Leyman and Y. H. Chew, "Wideband chirp parameter estimation in sensor arrays through DPT," *IEE Electronic Letters*, vol. 39, no. 23, pp. 1633-1634, Nov. 2003.