Parameter Estimation in Low-Rank Models from Small Sample Size and Undersampled Data: DOA and Spectrum Estimation

by

Mahdi Shaghaghi

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in
Communications

Department of Electrical and Computer Engineering
University of Alberta

Abstract

In estimation theory, a set of parameters are estimated from a finite number of measurements (samples). In general, the quality of estimation degrades as the number of samples is reduced. In this thesis, the problem of parameter estimation in low-rank models from a small number of samples is studied. Specifically, we consider two related problems that fit in this system model: direction-of-arrival (DOA) and spectrum estimation. We focus on subspace based DOA estimation methods which present a good compromise between performance and complexity. However, these methods are exposed to performance breakdown for a small number of samples. The reason is identified to be the intersubspace leakage where some portion of the true signal subspace resides in the estimated noise subspace. A two-step algorithm is proposed to reduce the amount of the subspace leakage. Theoretical derivations and simulation results are given to show the improvement achieved by the introduced method. Furthermore, the dynamics of the DOA estimation method in the breakdown region has been investigated, which led to identification of a problem named root-swap where a root associated with noise is mistakenly taken for a root associated with the signal. Then, an improved method is introduced to remedy this issue.

Spectrum estimation from undersampled data (samples obtained at a rate lower than the Nyquist rate) is studied next. Specifically, the performance of the averaged correlogram for undersampled data is theoretically analyzed for the finite length sample size as well as asymptotically. This method partitions the spectrum into a number of segments and estimates the average power within each segment from samples obtained at a rate lower than the Nyquist rate. However, the frequency resolution of the estimator is restricted

to the number of spectral segments, and the estimation made for each segment has also limited accuracy. Therefore, it is of significant importance to analyze the performance of this method especially in the case that only a finite number of samples is available. We derive the bias and variance of the averaged correlogram for undersampled data for finite-length signals, and we show the associated tradeoffs among the resolution, the accuracy, and the complexity of the method.

Finally, spectrum estimation from compressive measurements is studied. The number of such measurements is much less than the number of Nyquist samples, and they are obtained by correlating the signal with a number of sensing waveforms. We specifically consider signals composed of linear combinations of sinusoids. Albeit these type of signals have a sparse model, their representation in the Fourier basis exhibits frequency leakage. This problem results in the poor performance of the conventional compressive sensing recovery algorithms that rely on the Fourier basis. We introduce an improved model-based reconstruction algorithm which has a performance close to the Cramér-Rao bound, which we also derive.

Preface

Chapter 3 has been submitted to *IEEE Transaction on Signal Processing*. Some preliminary results have been published as M. Shaghaghi and S. A. Vorobyov, "Iterative root-MUSIC algorithm for DOA estimation," invited paper, in *Proc. 5th Inter. Workshop Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP'2013*, The Friendly Island, Saint Martin, Dec. 2013, pp. 53–56.

Chapter 4 has been submitted to *Applied and Computational Harmonic Analysis*. A summary of results has been published as M. Shaghaghi and S. A. Vorobyov, "Correlogram for undersampled data: bias and variance analysis," in *Proc. 37th IEEE Inter. Conf. Acoustics, Speech, and Signal Processing, ICASSP'2012*, Kyoto, Japan, Mar. 2012, pp. 3513-3516.

Chapter 5 has been published as M. Shaghaghi and S. A. Vorobyov, "Improved model-based spectral compressive sensing via nested Least Squares," in *Proc. 36th IEEE Inter. Conf. Acoustics, Speech, and Signal Processing, ICASSP'2011*, Prague, Czech Republic, May 2011, pp. 3904-3907.

I have been the first author of the aforementioned papers. I have performed the literature review, algorithm design, theoretical derivations, computer simulations, and manuscript writing of the papers with the guidance and supervision of Dr. Sergiy A. Vorobyov.

Table of Contents

List of Tables

List of Figures

List of Symbols

Symbol	Description	First use
$*$	Convolution operation	57
$\odot$	Hadamard product	10
$<,>$	Inner product	6
$(\cdot)^*$	Complex conjugate operator	9
$(\cdot)^H$	Hermitian transposition operator	8
$(\cdot)^T$	Transposition operator	7
β	Step size of least squares algorithm	74
γ	Scaling factor	17
Γ	Set of candidate direction-of-arrivals	19
$\delta\boldsymbol{P}$	First order term in $\widehat{\boldsymbol{P}}$ series expansion	22
$\delta^n\boldsymbol{P}$	n-th order term in $\widehat{\boldsymbol{P}}$ series expansion	22
$\delta(k)$	Kronecker delta function	48
$\Delta\omega_k$	$\hat{\omega}_k - \omega_k$	25
Δr_i	$\hat{r}_i - r_i$	19
$\Delta\boldsymbol{R}$	$\widehat{\boldsymbol{R}} - \boldsymbol{R}$	22
$\Delta\boldsymbol{R}^{(2)}$	$\widehat{\boldsymbol{R}}^{(2)} - \boldsymbol{R}$	24

$\boldsymbol{S}$	Source covariance matrix	8
T	Niquist period	45
$\mathrm{Tr}\{\cdot\}$	Trace operator	10
$\boldsymbol{T}$	By-product matrix in $\widehat{\boldsymbol{R}}$	16
$\boldsymbol{u}$	Vector containing elements of $\boldsymbol{R}_z$	46
$\breve{\boldsymbol{u}}$	Real-valued $\boldsymbol{u}$	46
$\widehat{\breve{\boldsymbol{u}}}$	Estimated $\breve{\boldsymbol{u}}$	48
$\boldsymbol{U}$	$E\{\widehat{\breve{\boldsymbol{u}}}\widehat{\breve{\boldsymbol{u}}}^T\}$	53
$\boldsymbol{v}$	Vector containing average power in each spectral segment	46
$\widehat{\boldsymbol{v}}$	Estimated $\boldsymbol{v}$	48
$\boldsymbol{V}$	$\boldsymbol{R}-\sigma_{\mathrm{n}}^2\boldsymbol{I}_M$	22
$\boldsymbol{V}^\dagger$	Pseudo-inverse of $\boldsymbol{V}$	22
$\boldsymbol{w}$	Vector of compressive sensing measurement noise	74
W	Bandwidth	45
$W_D(n)$	Window of length N	48
$x(m)$	Nyquist rate samples of $x(t)$	48
$x(t)$	Continuous-time process	45
$\boldsymbol{x}_a(t)$	Received vector	7
$y_i(n)$	Output of i-th sampling channel	45
$\boldsymbol{y}$	Vector of compressive sensing measurements	73
$\widehat{z}_a(n)$	Estimated $z_a(n)$ from finite samples	46

List of Abbreviations

Abbreviation	Description	First use
CMSE	Conditional MSE	30
CRB	Cramér-Rao bound	4
CS	Compressive sensing	4
DFT	Discrete Fourier transform	4
DOA	Direction-of-arrival	1
DTFT	Discrete-time Fourier transform	4
ESPRIT	Estimation of signal parameters via rotational invariance techniques	12
FD	Fractional delay	47
FIR	Finite impulse response	47
LS	Least squares	74
MC	Multi-coset	45
MIMO	Multiple-input multiple-output	88
ML	Maximum likelihood	2
MSE	Mean squared error	10
MUSIC	Multiple signal classification	2
NCRB	Normalized CRB	81

NMSE	Normalized MSE	81
PSD	Power spectral density	43
R-MUSIC	Root-MUSIC	35
RSUR-MUSIC	Root-swap unitary root-MUSIC	35
SIHT	Spectral iterative hard thresholding	75
SML	Stochastic maximum likelihood	17
SNR	Signal-to-noise ratio	12
ULA	Uniform linear array	7
UR-MUSIC	Unitary root-MUSIC	35
WSS	Wide-sense stationary	45

Chapter 1

Introduction

Parameter estimation from a finite number of measurements is a fundamental problem in signal processing. The quality of estimation highly depends on the number of available samples. In general, the availability of more samples translates into lower estimation error. On the other hand, the difference between the estimated and true values can grow large for a small number of samples. In some cases, there may be a minimum number of samples required for a successful estimation. In this thesis, we consider parameter estimation in low-rank system models from a small number of samples.

In a low-rank system, the vector of measured data is modeled as a signal component corrupted with additive noise. The signal term is a vector equal to the multiplication of a tall matrix by a vector containing the amplitudes of the sources. A tall matrix refers to a matrix with the number of rows being much larger than the number of columns. This matrix can be fully known, known up a number of unknown parameters, or completely unknown. In this thesis, we study the second case where the matrix contains unknown parameters, but the structure of the matrix is known. Specifically, we consider direction-of-arrival (DOA) and spectrum estimation problems which fit in this system model. For DOA estimation, the unknown parameters are the angles between far-field sources and an array of sensors, and for spectrum estimation, the unknown parameters are the frequencies of signals. Therefore, the system models of the source location problem and the problem of temporal spectral estimation are closely related. Furthermore, the DOA estimation problem

deals with determining the position of sources of energy in air, water or the earth, and can be viewed as estimating the distribution of energy over the space. Therefore, it can be called a spatial estimation problem. Because of these ties, most of the methods used for one problem are applicable to the other one as well.

In the case of DOA estimation, we consider the problem of locating K radiating or reflecting sources by using an array of M sensors. The emitted energy from the sources can be electromagnetic, acoustic, and so on, and the receiving sensors may be any transducers that convert the received energy to electrical signals. Examples of sensors include electromagnetic antennas, hydrophones, and seismometers. This type of problem has applications in radar and sonar systems, communications, astrophysics, biomedical research, seismology, underwater surveillance, and so on [1].

The goal is to perform an estimation from the spatial and temporal samples collected at the sensors (space-time processing). Spatial filtering (conventional beamformer) was the first approach to perform space-time processing from data sampled at an array of sensors, which dates back to the second world-war [2]. Later, classical time delay estimation methods [3] were introduced to enhance the ability of resolving closely spaced signal sources. Furthermore, an early application of the maximum likelihood (ML) principle for multiple-signal sources appeared in [4]. The ML method is an efficient estimator. However, it suffers from high computational complexity. Then, the emergence of subspace-based estimation techniques [5, 6] started a new era in the sensor array signal processing literature, as they provided a compromise between performance and complexity.

1.1 Proposed research problems

In this dissertation, we concentrate on small sample size and undersampled data scenarios. A brief description of the tackled problems is as follows.

Classical methods of DOA estimation based on the sample covariance matrix, such as the multiple signal classification (MUSIC) algorithm [7], are based

on estimating the signal and noise subspaces from the sample covariance matrix. For a small number of samples, such methods are exposed to performance breakdown, as the sample covariance matrix can largely deviate from the true covariance matrix. In Chapter 3, the problem of DOA estimation performance breakdown is investigated.

We consider the structure of the sample covariance matrix and the dynamics of the DOA estimation algorithm. It is shown in [8] that the performance breakdown problem is associated with the intersubspace leakage "whereby a small portion of the true signal eigenvector resides in the sample noise subspace (and vice versa)". In Chapter 3, the *subspace leakage* notion is formally defined, and its theoretical derivation is given. We propose a two-step method which improves the performance by modifying the sample covariance matrix such that the amount of the subspace leakage is reduced. Furthermore, we introduce a phenomenon named as *root-swap* which occurs in the DOA estimation algorithm in the low sample size region and degrades the performance of the DOA estimation. A new method is then proposed to alleviate this problem.

In the case of spectrum estimation, not only the number of available samples is finite, but also the rate at which these samples are collected is limited. The latter can lead to aliasing. A sufficient condition for alias-free sampling is to sample the signal of interest at the Nyquist rate which is twice the maximum frequency of the signal. In some cases, such as wideband spectrum sensing, the Nyquist rate sampling can be too costly with the current technology. Therefore, it is desirable to make spectrum estimation from measurements obtained at a rate lower than the Nyquist rate.

Chapter 4 gives the finite-length analysis of a spectrum estimation method for the case when the samples are obtained at a rate lower than the Nyquist rate. The method is referred to as the averaged correlogram for undersampled data. It is based on partitioning the spectrum into a number of segments and estimating the average power within each spectral segment. In this method, samples are collected using multiple channels, each operating at a rate L times lower than the Nyquist rate. The frequency resolution of the estimator is re-

stricted to the number of spectral segments, and the estimation made for each segment has also limited accuracy. Therefore, it is of significant importance to analyze the performance of this method especially in the case when only a finite number of samples is available. We derive the bias and variance of the spectrum estimator, and show that there is a tradeoff between the accuracy of the estimation, the frequency resolution, and the complexity of the estimator. A closed-form approximation of the estimation variance is derived, which clearly shows how the variance is related to different parameters. The asymptotic behavior of the estimator is also investigated.

In Chapter 5, we consider signals with sparse representations. These signals can be recovered from a number of measurements much less than the number of samples given by the Nyquist sampling rate using the compressive sensing (CS) method [9–11]. Such measurements are obtained by correlating the signal with a number of sensing waveforms. The algorithms used for recovering the signal from these measurements exploit the sparsity of the signal in a proper basis. In Chapter 5, sparse signals composed of linear combinations of sinusoids are studied. Albeit these type of signals generate sparse coefficients by the discrete-time Fourier transform (DTFT), their representation in the Fourier basis obtained by the discrete Fourier transform (DFT) exhibits frequency leakage. This problem results in the poor performance of the conventional CS recovery algorithms that rely on the Fourier basis. Although these signals do not have a sparse representation in the Fourier basis, they possess a sparse model in terms of the DTFT. We introduce an improved model-based algorithm that takes the signal structure into account to estimate the unknown parameters (the frequencies and amplitudes of the linearly combined sinusoidal signals). Furthermore, we derive the Cramér-Rao bound (CRB) for spectral compressive sensing, and show that the proposed algorithm approaches the CRB.

Chapter 2

Preliminaries

2.1 Array processing

In this section, we briefly review the assumptions that lead to the derivation of the system model for DOA estimation. Two approximations are considered to simplify the model of the antenna array.

The first approximation is that the array is in the far-field of the signal source. In the far-field, the direction of propagation is approximately the same for each element of the array. Modeling an emitter as far-field depends on the size of the array and on the desired accuracy of the approximation. Consider the geometry shown in Fig. 2.1 [12]. An emitter at the center of the circle is at distance r away from a linear array of length 2ℓ. In this example, the direction of propagation is normal to the array at the center, but deviates from normal incidence by θ at the end of the array. If we desire to keep the deviation from normal incidence to be less than a tenth of a degree, then $r/\ell > \cot(0.1\pi/180) \approx 573$. Therefore, the distance from the center of the array to the emitter must be at least $573/2$ times the length of the array for the far-field approximation to hold [12].

The second approximation is about the bandwidth of the received signal. When the far-field approximation holds, the direction of propagation is the same at each array element. Consider an array with two elements. If $\boldsymbol{u}$ is a unit-length vector in the direction of propagation and $\boldsymbol{d}$ is the vector from the first array element to the second as shown in Fig. 2.2, then the signal at the second array element will be delayed from that received at the first

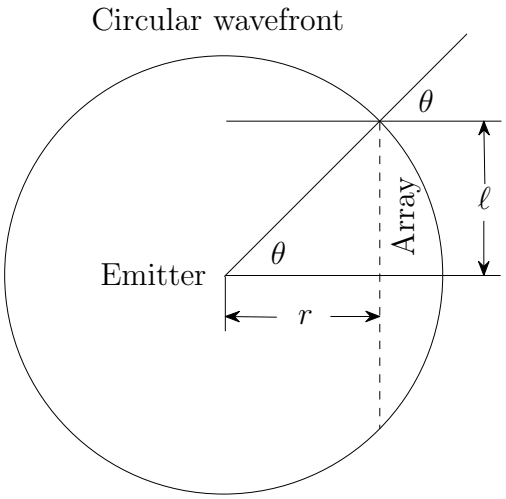

Figure 2.1: Far-field approximation.

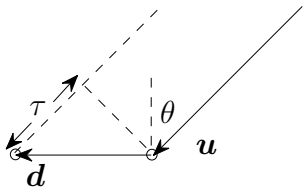

Figure 2.2: Narrowband approximation.

element by $\tau =< \boldsymbol{d}, \boldsymbol{u} > /c = d\sin(\theta)/c$, where $<,>$ represents the innner product of two vectors, d is the distance between the two array elements, and c is the speed of the traveling wave. If the signal received at the first element is $x_1(t) = m(t)e^{j\omega t}$, then the signal received at the second element is $x_2(t) = x_1(t-\tau) = m(t-\tau)e^{j\omega(t-\tau)} = m(t-\tau)e^{j\omega t}e^{-j\omega\tau}$. The narrowband approximation is used here. The rate at which a signal changes is related to the bandwidth of the signal. If the bandwidth of $m(t)$ is much smaller than $1/\tau$, we can use the approximation $m(t-\tau) \approx m(t)$, so that $x_2(t) = m(t-\tau)e^{j\omega t}e^{-j\omega\tau} \approx m(t)e^{j\omega t}e^{-j\omega\tau} = e^{-j\omega\tau}x_1(t)$. With this approximation the delay operator is replaced with a simple phase shift. The center frequency ω can be written as $\omega = 2\pi c/\lambda$, where λ is the wavelength of the plane wave impinging on the array. Therefore, the received signal at the second element

of the array can be written as $x_2(t) = e^{-j2\pi(d/\lambda)\sin(\theta)}x_1(t)$.

We next consider the model of a uniform linear array (ULA). The array consists of M number of antennas separated from each other by a distance of $d \leq \lambda/2$. The array receives K number of narrowband plane waves from directions $\theta_1,\ \theta_2, \cdots,\ \theta_K$. Without loss of generality, assume $-\pi/2 \leq \theta_1 \leq \theta_2 \leq \cdots \leq \theta_K \leq \pi/2$. The received waves are down-converted, low-pass filtered, and sampled in time. Each sample is a vector containing the measured data at the antenna elements. If we consider the first element of the array as the reference, the received signal from a given source is delayed by τ at the second element, by 2τ at the third element, by 3τ at the forth element and so on. Then, by defining the steering vector of the array $\boldsymbol{a}(\theta) \in \mathbb{C}^{M\times 1}$ as

$$\boldsymbol{a}(\theta) \triangleq \left[1,\ e^{-j2\pi(d/\lambda)\sin(\theta)}, \cdots,\ e^{-j2\pi(M-1)(d/\lambda)\sin(\theta)}\right]^T \tag{2.1}$$

(here $(\cdot)^T$ stands for the transposition operator) and also using the superposition principle for multiple sources, at time instant $t \in \mathbb{N}$, the received vector $\boldsymbol{x}_a(t) \in \mathbb{C}^{M\times 1}$ is given by

$$\boldsymbol{x}_a(t) = \sum_{i=1}^{K} \boldsymbol{a}(\theta_i)s_i(t) + \boldsymbol{n}(t) \tag{2.2}$$

where $s_i(t) \in \mathbb{C}$ is the amplitude of the i-th source and $\boldsymbol{n}(t) \in \mathbb{C}^{M\times 1}$ is the noise vector at time t. By arranging the amplitudes of the sources in the vector $\boldsymbol{s}(t) = [s_1(t),\ s_2(t), \cdots,\ s_K(t)]^T \in \mathbb{C}^{K\times 1}$ and forming the Vandermonde matrix $\boldsymbol{A} = [\boldsymbol{a}(\theta_1),\ \boldsymbol{a}(\theta_2), \cdots,\ \boldsymbol{a}(\theta_K)] \in \mathbb{C}^{M\times K}$, the model (2.2) can be rewritten in matrix-vector form as

$$\boldsymbol{x}_a(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t). \tag{2.3}$$

We consider the noise vector $\boldsymbol{n}(t)$ to be independent from the sources and other noise vectors and to have the circularly-symmetric complex jointly-Gaussian distribution $\mathcal{N}_C(0, \sigma_{\mathrm{n}}^2\boldsymbol{I}_M)$ where $\boldsymbol{I}_M$ is the identity matrix of size M.

2.2 DOA estimation problem

Subspace based DOA estimation methods are based on the structure of the covariance matrix of the received signal. Considering the system model (2.3),

the data covariance matrix $\boldsymbol{R} \in \mathbb{C}^{M\times M}$ is given by

$$\boldsymbol{R} \triangleq E\left\{\boldsymbol{x}_a(t)\boldsymbol{x}_a^H(t)\right\} = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H + \sigma_{\mathrm{n}}^2\boldsymbol{I}_M \tag{2.4}$$

where $\boldsymbol{S} = E\left\{\boldsymbol{s}(t)\boldsymbol{s}^H(t)\right\} \in \mathbb{C}^{K\times K}$ is the source covariance matrix, and $(\cdot)^H$ and $E\{\cdot\}$ stand for the Hermitian transposition and the expectation operators, respectively.

Let N number of snapshots (samples) be available. The conventional method for estimating the covariance matrix from the samples $\boldsymbol{x}_a(t)$ $(1 \leq t \leq N)$ is given by

$$\widehat{\boldsymbol{R}} \triangleq \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{x}_a(t)\boldsymbol{x}_a^H(t) \tag{2.5}$$

where $\widehat{\boldsymbol{R}} \in \mathbb{C}^{M\times M}$ is the sample data covariance matrix. Let $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \cdots \leq \hat{\lambda}_M$ be the eigenvalues of $\widehat{\boldsymbol{R}}$ in nondecreasing order, and let $\hat{\boldsymbol{g}}_1,\ \hat{\boldsymbol{g}}_2, \cdots,\ \hat{\boldsymbol{g}}_{M-K}$ be the noise eigenvectors associated with $\hat{\lambda}_1,\ \hat{\lambda}_2, \cdots,\ \hat{\lambda}_{M-K}$ and $\hat{\boldsymbol{e}}_1,\ \hat{\boldsymbol{e}}_2, \cdots, \hat{\boldsymbol{e}}_K$ be the signal eigenvectors corresponding to $\hat{\lambda}_{M-K+1},\ \hat{\lambda}_{M-K+2}, \cdots,\ \hat{\lambda}_M$. Let also $\widehat{\boldsymbol{G}} \in \mathbb{C}^{M\times(M-K)}$ and $\widehat{\boldsymbol{E}} \in \mathbb{C}^{M\times K}$ be defined as

$$\widehat{\boldsymbol{G}} \triangleq \left[\hat{\boldsymbol{g}}_1,\ \hat{\boldsymbol{g}}_2, \cdots,\ \hat{\boldsymbol{g}}_{M-K}\right] \tag{2.6}$$

$$\widehat{\boldsymbol{E}} \triangleq \left[\hat{\boldsymbol{e}}_1,\ \hat{\boldsymbol{e}}_2, \cdots,\ \hat{\boldsymbol{e}}_K\right]. \tag{2.7}$$

The range spaces of $\widehat{\boldsymbol{G}}$ and $\widehat{\boldsymbol{E}}$ represent the estimations of the noise and signal subspaces, respectively. Recalling (2.1) and defining $z \triangleq e^{j2\pi(d/\lambda)\sin(\theta)}$, the steering vector can be rewritten as

$$\boldsymbol{a}(z) = \left[1,\ z^{-1}, \cdots,\ z^{-(M-1)}\right]^T. \tag{2.8}$$

2.3 Root-MUSIC method

In the root-MUSIC method, the roots of the equation $\boldsymbol{a}^T(z^{-1})\widehat{\boldsymbol{G}}\widehat{\boldsymbol{G}}^H\boldsymbol{a}(z) = 0$ which are inside the unit circle are considered. These roots are sorted based on their distance to the unit circle, and the first K number of the roots which are closer to the unit circle are picked. The estimates of the DOAs denoted by $\hat{\theta}_1,\ \hat{\theta}_2, \cdots,\ \hat{\theta}_K$ are then obtained by multiplying the angles of the selected roots by $\lambda/(2\pi d)$ and taking the inverse sinusoid function of the results.

2.4 Unitary root-MUSIC algorithm

The unitary root-MUSIC algorithm [13] has a lower computational complexity compared to the root-MUSIC method, as it uses the eigendecomposition of a real-valued covariance matrix. Furthermore, the unitary root-MUSIC algorithm has better performance for the case that the sources are correlated. The real-valued covariance matrix $\widehat{\boldsymbol{C}} \in \mathbb{R}^{M \times M}$ is given by

$$\begin{aligned} \widehat{\boldsymbol{C}} &= \frac{1}{2}\boldsymbol{Q}_M^H \left(\widehat{\boldsymbol{R}} + \boldsymbol{J}_M \widehat{\boldsymbol{R}}^* \boldsymbol{J}_M \right) \boldsymbol{Q}_M \\ &= Re\left\{ \boldsymbol{Q}_M^H \widehat{\boldsymbol{R}} \boldsymbol{Q}_M \right\} \end{aligned} \tag{2.9}$$

where $(\cdot)^*$ and $Re\{\cdot\}$ stand for the complex conjugate and real part operators, respectively. Furthermore, the matrix $\boldsymbol{J}_M$ is the exchange matrix (all zeros except for the antidiagonal elements which are equal to one) of size M. In the case that M is even ($M = 2i$), $\boldsymbol{Q}_M$ is given by

$$\boldsymbol{Q}_M = \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{I}_i & j\boldsymbol{I}_i \\ \boldsymbol{J}_i & -j\boldsymbol{J}_i \end{bmatrix} \tag{2.10}$$

and in the case that M is odd ($M = 2i + 1$), $\boldsymbol{Q}_M$ is given by

$$\boldsymbol{Q}_M = \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{I}_i & \boldsymbol{0}_i & j\boldsymbol{I}_i \\ \boldsymbol{0}_i^T & \sqrt{2} & \boldsymbol{0}_i^T \\ \boldsymbol{J}_i & \boldsymbol{0}_i & -j\boldsymbol{J}_i \end{bmatrix} \tag{2.11}$$

where $\boldsymbol{0}_i$ is an $i \times 1$ vector of all zeros.

The rest of the steps of the unitary root-MUSIC algorithm are the same as the root-MUSIC method except for that $\boldsymbol{Q}_M \widehat{\boldsymbol{G}}_c$ is used instead of $\widehat{\boldsymbol{G}}$ to form the unitary root-MUSIC polynomial, where $\widehat{\boldsymbol{G}}_c \in \mathbb{R}^{M \times (M-K)}$ is composed of the noise eigenvectors of $\widehat{\boldsymbol{C}}$.

2.5 Cramér-Rao bound

The CRB, first introduced in [14, 15], is a useful lower bound on the variance of any unbiased estimator, and it provides a benchmark for comparisons of performance. The CRB for array processing is obtained in [16,17] in an indirect manner by an asymptotic analysis. A direct derivation of the CRB for array

processing is given in [18]. Let $\boldsymbol{\theta} \triangleq [\theta_1,\ \theta_2, \cdots,\ \theta_K]^T \in \mathbb{R}^{K\times 1}$ be the vector of parameters. Then, the covariance matrix of any unbiased estimator $\boldsymbol{C}_{\widehat{\boldsymbol{\theta}}} \in \mathbb{R}^{K\times K}$ follows

$$\boldsymbol{C}_{\widehat{\boldsymbol{\theta}}} \geq \mathrm{CRB}\left(\boldsymbol{\theta}\right) \tag{2.12}$$

where $\boldsymbol{C}_{\widehat{\boldsymbol{\theta}}} = E\left\{\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^T\right\}$ and $\boldsymbol{C}_{\widehat{\boldsymbol{\theta}}} \geq \mathrm{CRB}\left(\boldsymbol{\theta}\right)$ means that $\boldsymbol{C}_{\widehat{\boldsymbol{\theta}}} - \mathrm{CRB}\left(\boldsymbol{\theta}\right)$ is a positive-semidefinite matrix. The matrix $\mathrm{CRB}\left(\boldsymbol{\theta}\right)$ is given by [18]

$$\mathrm{CRB}\left(\boldsymbol{\theta}\right) = \frac{\sigma_{\mathrm{n}}^2}{2N}\left\{Re\left[\left(\boldsymbol{C}^H\boldsymbol{P}_A^{\perp}\boldsymbol{C}\right)\odot\left(\boldsymbol{S}\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{S}\right)^T\right]\right\}^{-1} \tag{2.13}$$

where $\boldsymbol{C} \triangleq [d\boldsymbol{a}(\theta_1)/d\theta_1\ \cdots\ d\boldsymbol{a}(\theta_K)/d\theta_K]$ and $\odot$ stands for the Hadamard product. The matrix $\boldsymbol{P}_A^{\perp} \in \mathbb{C}^{M\times M}$ is the true noise projection matrix given by $\boldsymbol{P}_A^{\perp} = \boldsymbol{I}_M - \boldsymbol{A}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H$.

In Chapter 3, the performance of various methods is investigated by comparing the mean squared error (MSE) of the estimators. The MSE is given by $\mathrm{Tr}\left\{\boldsymbol{C}_{\widehat{\boldsymbol{\theta}}}\right\}$ where $\mathrm{Tr}\left\{\cdot\right\}$ stands for the trace operator. Since $\boldsymbol{C}_{\widehat{\boldsymbol{\theta}}} - \mathrm{CRB}\left(\boldsymbol{\theta}\right)$ is a positive-semidefinite matrix, all of its eigenvalues are nonnegative. Therefore, $\mathrm{Tr}\left\{\boldsymbol{C}_{\widehat{\boldsymbol{\theta}}} - \mathrm{CRB}\left(\boldsymbol{\theta}\right)\right\} \geq 0$ since the trace of a matrix is equal to the summation of its eigenvalues. This gives the result that the MSE of any unbiased DOA estimator is lower bounded by $\mathrm{Tr}\left\{\mathrm{CRB}\left(\boldsymbol{\theta}\right)\right\}$.

2.6 Pseudo-noise resampling

Pseudo-noise resampling has been proposed in [19] to improve the DOA estimation performance at the breakdown region. The pseudo-noise resampling method is motivated by resampling schemes such as bootstrap [20,21] and uses synthetically generated pseudo-noise in order to perturb the original noise [22]. The conventional beamformer has also been combined with the pseudo-noise resampling technique in [23]. Furthermore, an improved unitary root-MUSIC method which uses the pseudo-noise resampling and beamforming techniques is introduced in [24]. In this section, we summarize the steps of the method in [24]. This algorithm will be later used in Chapter 3.

The method in [24] starts with estimating the DOAs using the unitary root-MUSIC algorithm. Next, a hypothesis is tested for the DOA estimates. If the

test is passed, the algorithm halts with the current DOA estimates as the final estimates. Otherwise, the pseudo-noise resampling technique is implemented to obtain improved estimates of the DOAs. The hypothesis test $\mathcal{H}$ checks if the estimated DOAs belong to a set of predetermined angular sectors. These sectors are obtained by the conventional beamformer [23, 25, 26]. There are K number of sectors. Each sector is centered around a peak of the beamformer output which corresponds to a source. The left and right boundaries of each sector is chosen as the points with 3 dB drop compared to the corresponding peak of the beamformer output.

The resampling process is next performed P number of times, and the DOAs are estimated at each resampling. The resampling is done by adding pseudo-noise vectors to the received vectors $\boldsymbol{x}_a(t)$ $(1 \leq t \leq N)$. Each pseudo-noise vector is independent from other pseudo-noise vectors and is drawn from the circularly-symmetric complex jointly-Gaussian distribution $\mathcal{N}_C(0, \hat{\sigma}_{\mathrm{n}}^2 \boldsymbol{I}_M)$. Here, $\hat{\sigma}_{\mathrm{n}}^2$ is equal to the average value of the noise eigenvalues of the real-valued covariance matrix $\widehat{\boldsymbol{C}}$. Then, for each resampling run, the unitary root-MUSIC algorithm is used to obtain a set of DOA estimates. The corresponding magnitudes of the roots associated with these DOA estimates are also recorded.

After the resampling step is done, there are P sets of DOA estimates (and their corresponding root magnitudes). First, the elements of each set are sorted, so that the k-th element of each set is associated with the k-th source. Then, each set is tested by the hypothesis $\mathcal{H}$, and two disjoint sets are formed: one containing the sets that passed the test, and the other one containing the sets that were rejected by the test. There are two possible scenarios. In case that the first set containing accepted DOA estimates is not empty, the final DOAs are obtained as the average value of the DOA estimates in this set (the k-th DOA estimate is equal to the average value of the k-th elements of the accepted DOA sets). In the other case when there are no accepted sets, the final DOA estimates are obtained in the following way. For the k-th DOA estimate, the corresponding magnitudes for the DOA estimates of the k-th source are compared, and the DOA estimate with the largest corresponding magnitude is picked as the final DOA estimate for the k-th source.

Chapter 3

Subspace Leakage Analysis and Improved DOA Estimation Methods

Classical methods DOA estimation such as the MUSIC [7], root-MUSIC [27], and estimation of signal parameters via rotational invariance techniques (ESPRIT) [28] are based on estimating the signal and noise subspaces from the sample data covariance matrix. It is well-known that these methods suffer from performance breakdown for a small number of samples or a low signal-to-noise ratio (SNR) value [29,30] where the expected estimation error departs from the CRB. The SNR region at which this phenomenon happens is known as the threshold region.

The fidelity of the sample data covariance matrix to the true data covariance matrix plays a critical role in a successful estimation. At the low SNR and/or small sample size region, the sample data covariance matrix can largely deviate from the true data covariance matrix. There are various methods introduced in the literature which target at improving the estimation of the data covariance matrix [19, 23, 24, 31–33].

Diagonal loading [31] and shrinkage [32] methods improve the estimate of the covariance matrix by scaling and shifting the eigenvalues of the sample data covariance matrix. However, the eigenvectors are kept unchanged. As a result, the estimated signal and noise projection matrices from the improved covariance matrices are exactly the same as those obtained from the sample

data covariance matrix. Therefore, these methods are not beneficial for the subspace-based DOA estimation algorithms.

In [33], techniques from random matrix theory have been deployed to improve the performance of the MUSIC algorithm. The introduced method considers the asymptotic situation when both the sample size and the number of array elements tend to infinity at the same rate. It is then inferred that the improved method gives a more accurate description of the situation where these two quantities are finite and comparable in magnitude. However, the performance of the introduced method is not satisfactory at the small sample size scenario [34].

A more promising approach to remedy the performance breakdown at the threshold region was introduced in [19] and has been further improved in [23] and [24]. These methods are based on a technique called pseudo-noise resampling which uses synthetically generated pseudo-noise to perturb the original noise. The pseudo-noise is added to the observed data, and a new estimate of the covariance matrix is obtained, which leads to new DOA estimates. This process is repeated for a number of times, and the final DOAs are determined based on the bank of the DOA estimates.

In this chapter, we tackle the problem of the performance breakdown at the threshold region by considering the structure of the sample data covariance matrix and the dynamics of the root-MUSIC algorithm. It is shown in [8] that the performance breakdown problem is associated with the intersubspace leakage "whereby a small portion of the true signal eigenvector resides in the sample noise subspace (and vice versa)". In this chapter, we formally define the *subspace leakage* notion, and we present its theoretical derivation. We propose a two-step method which improves the performance of the root-MUSIC algorithm by modifying the sample data covariance matrix such that the amount of the subspace leakage is reduced. Furthermore, we introduce a phenomenon named as *root-swap* which occurs in the root-MUSIC algorithm at the threshold region and degrades the performance of the DOA estimation. A new method is then proposed to alleviate this problem.

It will be shown that there are undesirable by-products in the sample data

covariance matrix that tend to zero as the number of samples goes to infinity. However, for a limited number of samples, these terms can have significant values leading to a large amount of subspace leakage. One possible approach to remedy the effect of the undesirable components is to consider the *eigenvalue perturbation* caused by these terms. The incorporation of this knowledge into the estimation method can result in better estimates of the signal and noise subspaces. In this chapter, we propose a two-step algorithm in order to reduce the effect of the undesirable terms. The introduced method is based on estimating the DOAs at the first step and modifying the covariance matrix using the estimated DOAs at the second step. We will theoretically derive the subspace leakage at both steps. Then, it will be shown using numerical examples that the subspace leakage is reduced at the second step leading to better performance.

In the root-MUSIC method, the estimation error of the roots follows a chi-square distribution with a variance which is proportional to the variance of noise over the number of samples [35]. Therefore, at the threshold region, the variance of the estimation error can have a significant value which in turn can result in a swap between a root corresponding to a signal source with a root associated with the noise. We dub this phenomenon as a root-swap. Then, a new method is proposed to remedy this problem. The introduced method considers different combinations of the roots as the candidates for the signal sources. These candidates are then evaluated using the stochastic maximum likelihood function, and the combination that minimizes the objective function is picked up for the DOA estimates [36].

The rest of the chapter is organized as follows. First, The two-step and root-swap algorithms are proposed in Section 3.1. Definition and theoretical derivations of the subspace leakage are presented in Section 3.2. Finally, numerical examples and simulation results are given in Section 3.3.

3.1 Proposed improved methods

3.1.1 Two-step root-MUSIC algorithm

Let us start by expanding (2.5) using (2.3) as follows

$$\begin{aligned}\widehat{\boldsymbol{R}} &= \frac{1}{N}\sum_{t=1}^{N}\left(\boldsymbol{A}\boldsymbol{s}(t)+\boldsymbol{n}(t)\right)\left(\boldsymbol{A}\boldsymbol{s}(t)+\boldsymbol{n}(t)\right)^{H} \\ &= \boldsymbol{A}\left\{\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{s}(t)\boldsymbol{s}^{H}(t)\right\}\boldsymbol{A}^{H}+\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{n}(t)\boldsymbol{n}^{H}(t) \\ &\quad +\boldsymbol{A}\left\{\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{s}(t)\boldsymbol{n}^{H}(t)\right\}+\left\{\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{n}(t)\boldsymbol{s}^{H}(t)\right\}\boldsymbol{A}^{H}. \end{aligned} \tag{3.1}$$

Comparing (3.1) with (2.4), it can be observed that the expansion of $\widehat{\boldsymbol{R}}$ consists of four terms while the model for $\boldsymbol{R}$ comprises two summands. The first two terms of $\widehat{\boldsymbol{R}}$ given by (3.1) can be considered as estimations for the two summands of $\boldsymbol{R}$, which represent the signal and noise components, respectively. The last two terms of $\widehat{\boldsymbol{R}}$ in (3.1) are undesirable by-products which can be viewed as estimations for the correlation between the signal and noise vectors. In the system model under study, we consider the noise vectors to be zero-mean and also independent of the signal vectors. Therefore, the signal and noise components are uncorrelated to each other. As a result, for large enough number of samples N, the last two terms in (3.1) tend to zero. However, the number of available samples can be limited in practical applications, and in some cases, it is favorable to estimate the DOAs based on a finite number of samples, so that real-time decisions can be made as fast as possible. In this case, the last two terms in (3.1) may have significant values, which causes the estimations of the signal and noise subspaces to deviate from the true signal and noise subspaces.

The main idea of the two-step root-MUSIC algorithm is that at the first step, DOAs are estimated based on the sample data covariance matrix $\widehat{\boldsymbol{R}}$, and at the second step, DOAs are estimated again based on a modified covariance matrix. The new covariance matrix is obtained by deducting a scaled version of the estimated undesirable terms from the sample data covariance

matrix. The steps of the proposed method are listed in Table 3.1. The algorithm starts by computing the sample data covariance matrix $\widehat{\boldsymbol{R}}$. Next, the DOAs are estimated using the root-MUSIC algorithm. The superscript $(\cdot)^{(1)}$ refers to the estimation made at the first step. Then, at the second step, the Vandermonde matrix is formed using the available estimations of the DOAs. Next, we estimate the amplitudes of the sources such that the squared norm of the difference between the observation and the estimation is minimized. This step can be formulated in terms of the following optimization problem

$$\hat{\boldsymbol{s}}(t) = \arg \min_{\boldsymbol{s}} \|\boldsymbol{x}_a(t) - \widehat{\boldsymbol{A}}\boldsymbol{s}\|_2^2. \tag{3.2}$$

The minimization of (3.2) can be performed using the least squares (LS) technique and the corresponding solution is given as

$$\hat{\boldsymbol{s}}(t) = \left(\widehat{\boldsymbol{A}}^H \widehat{\boldsymbol{A}}\right)^{-1} \widehat{\boldsymbol{A}}^H \boldsymbol{x}_a(t). \tag{3.3}$$

The noise component is then estimated as the difference between the estimated signal and the observation made by the array, i.e.,

$$\hat{\boldsymbol{n}}(t) = \boldsymbol{x}_a(t) - \widehat{\boldsymbol{A}}\hat{\boldsymbol{s}}(t). \tag{3.4}$$

After estimating the signal and noise vectors, the third term in (3.1) can be found as

$$\begin{aligned}
\boldsymbol{T} &\triangleq \widehat{\boldsymbol{A}} \left\{ \frac{1}{N} \sum_{t=1}^{N} \hat{\boldsymbol{s}}(t) \hat{\boldsymbol{n}}^H(t) \right\} \\
&= \widehat{\boldsymbol{A}} \left\{ \frac{1}{N} \sum_{t=1}^{N} \left(\widehat{\boldsymbol{A}}^H \widehat{\boldsymbol{A}}\right)^{-1} \widehat{\boldsymbol{A}}^H \boldsymbol{x}_a(t) \left(\boldsymbol{x}_a^H(t) - \boldsymbol{x}_a^H(t) \widehat{\boldsymbol{A}} \left(\widehat{\boldsymbol{A}}^H \widehat{\boldsymbol{A}}\right)^{-1} \widehat{\boldsymbol{A}}^H \right) \right\} \\
&= \widehat{\boldsymbol{P}}_A \left\{ \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{x}_a(t) \boldsymbol{x}_a^H(t) \left(\boldsymbol{I}_M - \widehat{\boldsymbol{P}}_A \right) \right\} \\
&= \widehat{\boldsymbol{P}}_A \widehat{\boldsymbol{R}} \widehat{\boldsymbol{P}}_A^{\perp}
\end{aligned} \tag{3.5}$$

where

$$\widehat{\boldsymbol{P}}_A \triangleq \widehat{\boldsymbol{A}} \left(\widehat{\boldsymbol{A}}^H \widehat{\boldsymbol{A}}\right)^{-1} \widehat{\boldsymbol{A}}^H \tag{3.6}$$

is an estimation for the projection matrix of the signal subspace, and

$$\widehat{\boldsymbol{P}}_A^{\perp} \triangleq \boldsymbol{I}_M - \widehat{\boldsymbol{P}}_A \tag{3.7}$$

is an estimation for the projection matrix of the noise subspace. The forth term in (3.1) is equal to the Hermitian of the third term, i.e., $\boldsymbol{T}^H$. Next, the modified covariance matrix is obtained by deducting a scaled version of the estimated terms from the sample data covariance matrix as follows

$$\widehat{\boldsymbol{R}}^{(2)} = \widehat{\boldsymbol{R}} - \gamma \left(\boldsymbol{T} + \boldsymbol{T}^H \right). \tag{3.8}$$

The scaling factor γ is considered to be a real number between zero and one, and it is introduced due to the fact that the estimations of the undesirable terms are not perfect, and therefore, these estimations are scaled down before being subtracted from the sample data covariance matrix.

Next, given the modified data covariance matrix $\widehat{\boldsymbol{R}}^{(2)}$, the DOAs are estimated again using the root-MUSIC algorithm. The value of γ can be predetermined before running the algorithm, or it can be obtained to optimize an objective function. Here, we consider the minimization of the *stochastic maximum likelihood* (SML) function given by [16]

$$F_{SML}(\gamma) = \ln \det \left(\widehat{\boldsymbol{P}}_A^{(2)} \widehat{\boldsymbol{R}} \widehat{\boldsymbol{P}}_A^{(2)} + \frac{\mathrm{Tr}\left\{ \widehat{\boldsymbol{P}}_A^{\perp(2)} \widehat{\boldsymbol{R}} \right\}}{M-K} \widehat{\boldsymbol{P}}_A^{\perp(2)} \right) \tag{3.9}$$

where $\widehat{\boldsymbol{P}}_A^{(2)}$ is an estimation of the projection matrix of the signal subspace obtained from $\hat{\theta}_1^{(2)},\ \hat{\theta}_2^{(2)}, \cdots,\ \hat{\theta}_K^{(2)}$, and $\widehat{\boldsymbol{P}}_A^{\perp(2)} = \boldsymbol{I}_M - \widehat{\boldsymbol{P}}_A^{(2)}$. The minimization of (3.9) is performed by considering different values for γ taken on a grid (e.g. $\gamma = 0,\ 0.1,\ 0.2, \cdots,\ 1$) and making DOA estimations for each value of γ. Then, the set of DOA estimations corresponding to the value of γ that minimizes (3.9) is chosen as the output of the algorithm.

Table 3.1: Two-step root-MUSIC algorithm

Inputs:

$M,\ d,\ \lambda,\ N,\ K$, and
received vectors $\boldsymbol{x}_a(1),\ \boldsymbol{x}_a(2), \cdots,\ \boldsymbol{x}_a(N)$

Outputs:

Estimations $\hat{\theta}_1^{(2)},\ \hat{\theta}_2^{(2)}, \cdots,\ \hat{\theta}_K^{(2)}$

Step 1:

$\widehat{\boldsymbol{R}} = \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{x}_a(t) \boldsymbol{x}_a^H(t)$

$\left\{ \hat{\theta}_1^{(1)},\ \hat{\theta}_2^{(1)}, \cdots,\ \hat{\theta}_K^{(1)} \right\} \leftarrow \text{root-MUSIC} \left(\widehat{\boldsymbol{R}}, K, d, \lambda \right)$

Step 2:

$\widehat{\boldsymbol{A}} = \left[\boldsymbol{a}\left(\hat{\theta}_1^{(1)} \right),\ \boldsymbol{a}\left(\hat{\theta}_2^{(1)} \right), \cdots,\ \boldsymbol{a}\left(\hat{\theta}_K^{(1)} \right) \right]$

$\widehat{\boldsymbol{P}}_A = \widehat{\boldsymbol{A}} \left(\widehat{\boldsymbol{A}}^H \widehat{\boldsymbol{A}} \right)^{-1} \widehat{\boldsymbol{A}}^H$

$\widehat{\boldsymbol{P}}_A^{\perp} = \boldsymbol{I}_M - \widehat{\boldsymbol{P}}_A$

$\boldsymbol{T} = \widehat{\boldsymbol{P}}_A \widehat{\boldsymbol{R}} \widehat{\boldsymbol{P}}_A^{\perp}$

Determine γ **as the minimizer of** (3.9)

$\widehat{\boldsymbol{R}}^{(2)} = \widehat{\boldsymbol{R}} - \gamma \left(\boldsymbol{T} + \boldsymbol{T}^H \right)$

$\left\{ \hat{\theta}_1^{(2)},\ \hat{\theta}_2^{(2)}, \cdots,\ \hat{\theta}_K^{(2)} \right\} \leftarrow \text{root-MUSIC} \left(\widehat{\boldsymbol{R}}^{(2)}, K, d, \lambda \right)$

3.1.2 Root-swap root-MUSIC algorithm

In the root-MUSIC method, the roots inside the unit circle are considered, and the first K number of them which are closer to the unit circle are picked. Then, the DOAs are estimated based on the angles of these roots. Due to the finiteness of the available samples, the estimated roots from the sample data covariance matrix $\widehat{\boldsymbol{R}}$ deviate from the true roots corresponding to the actual data covariance matrix $\boldsymbol{R}$. Specifically, let $\Delta r_i \triangleq \hat{r}_i - r_i$ be the difference between the magnitude of the i-th estimated root $\hat{r}_i$ and the magnitude of the corresponding true root r_i. It is shown in [35] that $(\Delta r_i)^2$ follows a $\left(\sigma_y^2/2\right)\chi^2\left(2(M-K)-1\right)$ distribution, where $\chi^2(\ell)$ denotes a chi-square distribution with ℓ degrees of freedom, and σ_y^2 is computed from the eigenvalues and eigenvectors of $\boldsymbol{R}$ and the true DOAs. The variance of $(\Delta r_i)^2$ depends on σ_y^2 which is proportional to σ_{n}^2/N [35]. Therefore, the variance of $(\Delta r_i)^2$ can be significant for a small number of samples and a large value of σ_{n}^2 (low SNR region). Consequently, there can be a considerable probability that an estimated root associated with noise takes a larger magnitude than an estimated root associated with a signal source. We will refer to this phenomenon as a *root-swap*. To deal with this problem, we propose an algorithm that considers different combinations of the roots to be the candidates for the DOA estimates. The method is dubbed the root-swap root-MUSIC algorithm.

The root-MUSIC polynomial has $M-1$ number of roots inside the unit circle. Our goal is to find the roots which have a higher likelihood of being associated with the K sources. Consider choosing K number of roots out of all the roots. There are $N_c \triangleq (M-1)!/\left(K!(M-K-1)!\right)$ different possible combinations. Let

$$\Gamma \triangleq \{\Theta_1,\ \Theta_2,\ \cdots,\ \Theta_{N_c}\} \tag{3.10}$$

where Θ_i $(1 \leq i \leq N_c)$ is a set containing the DOA estimates obtained from the i-th combination. Then, the root-swap root-MUSIC method estimates the DOAs as

$$\left\{\hat{\theta}_1,\ \hat{\theta}_2, \cdots,\ \hat{\theta}_K\right\} = \arg\ \min_{\Theta \in \Gamma} F_{SML}(\Theta) \tag{3.11}$$

where $F_{SML}(\Theta)$ is the stochastic maximum likelihood function given in (3.9)

with $\widehat{\boldsymbol{P}}_A^{(2)}$ replaced with the signal projection matrix obtained from Θ.

The complexity of the introduced root-swap root-MUSIC method can be reduced by pre-eliminating some of the roots. Specifically, let $p \leq K$ closest roots to the unit circle be picked, and let q number of roots closest to the origin (furthest from the unit circle) be ignored. Then, our task is to choose $K-p$ number of roots out of $M-p-q-1$ roots. Now, there are $N_r \triangleq (M-p-q-1)!/\left((K-p)!(M-K-q-1)!\right)$ different possible combinations. The rest of the algorithm is the same as above except for that here each combination contains $K-p$ number of roots. Therefore, in order to evaluate the SML function, the fixed p pre-selected roots are added to each combination.

3.2 Subspace leakage

3.2.1 Definition

Consider the eigendecomposition of the data covariance matrix $\boldsymbol{R}$. Form $\boldsymbol{G} \in \mathbb{C}^{M\times(M-K)}$ and $\boldsymbol{E} \in \mathbb{C}^{M\times K}$ by placing the noise and signal eigenvectors as the columns of $\boldsymbol{G}$ and $\boldsymbol{E}$, respectively. The range spaces of $\boldsymbol{G}$ and $\boldsymbol{E}$ represent the true noise and signal subspaces. Note that the matrix of the eigenvectors $\boldsymbol{Q}_R = [\boldsymbol{G}\ \boldsymbol{E}] \in \mathbb{C}^{M\times M}$ is a unitary matrix $\left(\boldsymbol{Q}_R\boldsymbol{Q}_R^H = \boldsymbol{I}_M\right)$, therefore

$$\boldsymbol{G}\boldsymbol{G}^H + \boldsymbol{E}\boldsymbol{E}^H = \boldsymbol{I}_M \tag{3.12}$$

or

$$\boldsymbol{P}^{\perp} + \boldsymbol{P} = \boldsymbol{I}_M \tag{3.13}$$

where, $\boldsymbol{P}^{\perp} \triangleq \boldsymbol{G}\boldsymbol{G}^H$ and $\boldsymbol{P} \triangleq \boldsymbol{E}\boldsymbol{E}^H$ are the true projection matrices into the noise and signal subspaces.

Ideally, the estimation of each signal eigenvector $\hat{\boldsymbol{e}}_k$ $(1 \leq k \leq K)$ would perfectly fall in the true signal subspace. In practice, however, the energy of the projection of $\hat{\boldsymbol{e}}_k$ into the noise subspace $\left(\|\boldsymbol{P}^{\perp}\hat{\boldsymbol{e}}_k\|_2^2\right)$ is almost surely nonzero, which can be viewed as the leakage of $\hat{\boldsymbol{e}}_k$ into the true noise subspace. Here, we define the *subspace leakage* as the average value of the energy of the estimated

signal eigenvectors leaked into the true noise subspace, i.e.,

$$\rho \triangleq \frac{1}{K}\sum_{k=1}^{K} \|\boldsymbol{P}^{\perp}\hat{\boldsymbol{e}}_k\|_2^2. \tag{3.14}$$

Note that $\boldsymbol{P}^{\perp}$ is a Hermitian matrix and $\boldsymbol{P}^{\perp}\boldsymbol{P}^{\perp} = \boldsymbol{P}^{\perp}$. Therefore, ρ can be written as

$$\rho = \frac{1}{K}\sum_{k=1}^{K} \hat{\boldsymbol{e}}_k^H \boldsymbol{P}^{\perp}\hat{\boldsymbol{e}}_k. \tag{3.15}$$

Using (3.13), the expression (3.15) can be simplified to

$$\begin{aligned}
\rho &= \frac{1}{K}\sum_{k=1}^{K} \hat{\boldsymbol{e}}_k^H \left(\boldsymbol{I}_M - \boldsymbol{P}\right)\hat{\boldsymbol{e}}_k \\
&= \frac{1}{K}\left(K - \sum_{k=1}^{K} \hat{\boldsymbol{e}}_k^H \boldsymbol{P}\hat{\boldsymbol{e}}_k\right) \\
&= 1 - \frac{1}{K}\sum_{k=1}^{K} \mathrm{Tr}\left\{\hat{\boldsymbol{e}}_k\hat{\boldsymbol{e}}_k^H\boldsymbol{P}\right\} \\
&= 1 - \frac{1}{K}\mathrm{Tr}\left\{\left(\sum_{k=1}^{K}\hat{\boldsymbol{e}}_k\hat{\boldsymbol{e}}_k^H\right)\boldsymbol{P}\right\} \\
&= 1 - \frac{1}{K}\mathrm{Tr}\left\{\widehat{\boldsymbol{E}}\widehat{\boldsymbol{E}}^H\boldsymbol{P}\right\} \\
&= 1 - \frac{1}{K}\mathrm{Tr}\left\{\widehat{\boldsymbol{P}}\boldsymbol{P}\right\}
\end{aligned} \tag{3.16}$$

where $\widehat{\boldsymbol{P}} \triangleq \widehat{\boldsymbol{E}}\widehat{\boldsymbol{E}}^H$ is the estimated signal projection matrix.

3.2.2 Two-step root-MUSIC algorithm

The estimated signal and noise projection matrices obtained from the eigendecomposition of the sample data covariance matrix $\widehat{\boldsymbol{R}}$ and the modified covariance matrix $\widehat{\boldsymbol{R}}^{(2)}$ are deviated from the true signal and noise projection matrices. Let ρ_1 and ρ_2 be the subspace leakage due to the estimations obtained from $\widehat{\boldsymbol{R}}$ and $\widehat{\boldsymbol{R}}^{(2)}$, respectively. In this section, we derive the expected value of ρ_1 and ρ_2.

Subspace leakage at the first step of the proposed algorithm

Let us start with the computation of ρ_1. Let $\Delta\boldsymbol{P} \triangleq \widehat{\boldsymbol{P}} - \boldsymbol{P}$ be the estimation error of the signal projection matrix. Then, using the properties that $\boldsymbol{P}^2 = \boldsymbol{P}$

and $\mathrm{Tr}\{\boldsymbol{P}\} = K$, the expression (3.16) can be simplified as

$$\begin{aligned}\rho_1 &= 1 - \frac{1}{K}\mathrm{Tr}\left\{\left(\boldsymbol{P} + \Delta\boldsymbol{P}\right)\boldsymbol{P}\right\} \\ &= 1 - \frac{1}{K}\left(K + \mathrm{Tr}\left\{\Delta\boldsymbol{P}\boldsymbol{P}\right\}\right) \\ &= -\frac{1}{K}\mathrm{Tr}\left\{\Delta\boldsymbol{P}\boldsymbol{P}\right\}. \end{aligned} \tag{3.17}$$

Now, let $\Delta\boldsymbol{R} \triangleq \widehat{\boldsymbol{R}} - \boldsymbol{R}$ be the estimation error of the covariance matrix. Define also

$$\begin{aligned}\boldsymbol{V} &\triangleq \boldsymbol{R} - \sigma_{\mathrm{n}}^2\boldsymbol{I}_M \\ &= \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H \\ &= \sum_{k=1}^{K}\left(\lambda_{M-K+k} - \sigma_{\mathrm{n}}^2\right)\boldsymbol{e}_k\boldsymbol{e}_k^H \end{aligned} \tag{3.18}$$

and let $\boldsymbol{V}^\dagger \in \mathbb{C}^{M\times M}$ denote the pseudo-inverse of $\boldsymbol{V}$ given by

$$\boldsymbol{V}^\dagger = \sum_{k=1}^{K}\frac{1}{\lambda_{M-K+k} - \sigma_{\mathrm{n}}^2}\boldsymbol{e}_k\boldsymbol{e}_k^H \tag{3.19}$$

where $\lambda_{M-K+1} \le \lambda_{M-K+2} \le \cdots \le \lambda_M$ are the K largest eigenvalues of $\boldsymbol{R}$, and $\boldsymbol{e}_1,\ \boldsymbol{e}_2, \cdots,\ \boldsymbol{e}_K$ are their corresponding eigenvectors. It is shown in [35] that the series expansion of $\widehat{\boldsymbol{P}}$ based on $\Delta\boldsymbol{R}$ is given by

$$\widehat{\boldsymbol{P}} = \boldsymbol{P} + \delta\boldsymbol{P} + \cdots + \delta^n\boldsymbol{P} + \cdots \tag{3.20}$$

where

$$\delta\boldsymbol{P} = \boldsymbol{P}^\perp\Delta\boldsymbol{R}\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger\Delta\boldsymbol{R}\boldsymbol{P}^\perp \tag{3.21}$$

and the rest of the terms are related by the following recurrence

$$\begin{aligned}\delta^n\boldsymbol{P} = &-\boldsymbol{P}^\perp\left(\delta^{n-1}\boldsymbol{P}\right)\Delta\boldsymbol{R}\boldsymbol{V}^\dagger + \boldsymbol{P}^\perp\Delta\boldsymbol{R}\left(\delta^{n-1}\boldsymbol{P}\right)\boldsymbol{V}^\dagger \\ &-\boldsymbol{V}^\dagger\Delta\boldsymbol{R}\left(\delta^{n-1}\boldsymbol{P}\right)\boldsymbol{P}^\perp + \boldsymbol{V}^\dagger\left(\delta^{n-1}\boldsymbol{P}\right)\Delta\boldsymbol{R}\boldsymbol{P}^\perp \\ &-\sum_{i=1}^{n-1}\boldsymbol{P}\left(\delta^i\boldsymbol{P}\right)\left(\delta^{n-i}\boldsymbol{P}\right)\boldsymbol{P} + \sum_{i=1}^{n-1}\boldsymbol{P}^\perp\left(\delta^i\boldsymbol{P}\right)\left(\delta^{n-i}\boldsymbol{P}\right)\boldsymbol{P}^\perp. \end{aligned} \tag{3.22}$$

The following lemma regarding the columns of $\boldsymbol{V}^\dagger$ is in order.

Lemma 3.1. *The columns of $\boldsymbol{V}^\dagger$ belong to the signal subspace, i.e., $\boldsymbol{P}\boldsymbol{V}^\dagger = \boldsymbol{V}^\dagger$.*

Proof. The following train of equalities is valid.

$$\begin{aligned}\boldsymbol{P}\boldsymbol{V}^{\dagger} &= \boldsymbol{E}\boldsymbol{E}^{H}\sum_{k=1}^{K}\frac{1}{\lambda_{M-K+k}-\sigma_{\mathrm{n}}^{2}}\boldsymbol{e}_{k}\boldsymbol{e}_{k}^{H}\\ &= \sum_{i=1}^{K}\boldsymbol{e}_{i}\boldsymbol{e}_{i}^{H}\sum_{k=1}^{K}\frac{1}{\lambda_{M-K+k}-\sigma_{\mathrm{n}}^{2}}\boldsymbol{e}_{k}\boldsymbol{e}_{k}^{H}\\ &= \sum_{k=1}^{K}\frac{1}{\lambda_{M-K+k}-\sigma_{\mathrm{n}}^{2}}\boldsymbol{e}_{k}\boldsymbol{e}_{k}^{H} = \boldsymbol{V}^{\dagger} \end{aligned} \tag{3.23}$$

In the last step, we used the fact that $\boldsymbol{e}_i^H\boldsymbol{e}_k$ is equal to 1 for $i = k$ and it equals zero otherwise. □

In a similar way to Lemma 3.1, it can also be shown that

$$\boldsymbol{V}\boldsymbol{V}^{\dagger} = \boldsymbol{V}^{\dagger}\boldsymbol{V} = \boldsymbol{P}. \tag{3.24}$$

Using (3.17) together with the series expansion of $\widehat{\boldsymbol{P}}$ in (3.20), (3.21), and (3.22) up to the $\delta^2\boldsymbol{P}$ term, the facts that $\boldsymbol{P}\boldsymbol{P}^{\perp} = \boldsymbol{P}^{\perp}\boldsymbol{P} = \boldsymbol{0}$, $\boldsymbol{P}^{\perp}\boldsymbol{P}^{\perp} = \boldsymbol{P}^{\perp}$, $\boldsymbol{P} = \boldsymbol{P}\boldsymbol{P}$, and Lemma 3.1, we can compute ρ_1 as

$$\begin{aligned}\rho_1 = & -\frac{1}{K}\mathrm{Tr}\left\{-\boldsymbol{P}\left(\delta\boldsymbol{P}\right)\left(\delta\boldsymbol{P}\right)\right\}\\ &= \frac{1}{K}\mathrm{Tr}\left\{\boldsymbol{P}\left(\boldsymbol{P}^{\perp}\Delta\boldsymbol{R}\boldsymbol{V}^{\dagger}+\boldsymbol{V}^{\dagger}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\right)\left(\boldsymbol{P}^{\perp}\Delta\boldsymbol{R}\boldsymbol{V}^{\dagger}+\boldsymbol{V}^{\dagger}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\right)\right\}\\ &= \frac{1}{K}\mathrm{Tr}\left\{\boldsymbol{P}\boldsymbol{V}^{\dagger}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\boldsymbol{P}^{\perp}\Delta\boldsymbol{R}\boldsymbol{V}^{\dagger}\right\}\\ &= \frac{1}{K}\mathrm{Tr}\left\{\boldsymbol{V}^{\dagger}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\Delta\boldsymbol{R}\boldsymbol{V}^{\dagger}\right\}. \end{aligned} \tag{3.25}$$

Computation of the expected value of the subspace leakage requires considering the statistical properties of $\Delta\boldsymbol{R}$. We use the following two properties in our derivations [35].

Lemma 3.2. *For all matrices* $\boldsymbol{A}_1,\ \boldsymbol{A}_2 \in \mathbb{C}^{M\times M}$*, we have*

$$E\left\{\Delta\boldsymbol{R}\boldsymbol{A}_1\Delta\boldsymbol{R}\right\} = \frac{1}{N}Tr\{\boldsymbol{R}\boldsymbol{A}_1\}\,\boldsymbol{R} \tag{3.26}$$

and

$$E\left\{Tr\{\Delta\boldsymbol{R}\boldsymbol{A}_1\}\,Tr\{\Delta\boldsymbol{R}\boldsymbol{A}_2\}\right\} = \frac{1}{N}Tr\{\boldsymbol{R}\boldsymbol{A}_1\boldsymbol{R}\boldsymbol{A}_2\}. \tag{3.27}$$

Using (3.25) and (3.26), the expected value of ρ_1 can be computed as

$$\begin{aligned} E\left\{\rho_1\right\} &= \frac{1}{K}\mathrm{Tr}\left\{\boldsymbol{V}^\dagger E\left\{\Delta\boldsymbol{R}\boldsymbol{P}^\perp\Delta\boldsymbol{R}\right\}\boldsymbol{V}^\dagger\right\} \\ &= \frac{1}{K}\mathrm{Tr}\left\{\boldsymbol{V}^\dagger\frac{1}{N}\mathrm{Tr}\left\{\boldsymbol{R}\boldsymbol{P}^\perp\right\}\boldsymbol{R}\boldsymbol{V}^\dagger\right\} \\ &= \frac{1}{NK}\mathrm{Tr}\left\{\boldsymbol{P}^\perp\boldsymbol{R}\right\}\mathrm{Tr}\left\{\boldsymbol{V}^\dagger\boldsymbol{V}^\dagger\boldsymbol{R}\right\}. \end{aligned} \tag{3.28}$$

Since the range space of the matrix $\boldsymbol{A}$ is the same as the signal subspace, we have $\boldsymbol{P}^\perp\boldsymbol{A} = \boldsymbol{0}$. As a result, $\mathrm{Tr}\left\{\boldsymbol{P}^\perp\boldsymbol{R}\right\}$ can be simplified as

$$\begin{aligned} \mathrm{Tr}\left\{\boldsymbol{P}^\perp\boldsymbol{R}\right\} &= \mathrm{Tr}\left\{\boldsymbol{P}^\perp\left(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H + \sigma_{\mathrm{n}}^2\boldsymbol{I}_M\right)\right\} \\ &= \mathrm{Tr}\left\{\sigma_{\mathrm{n}}^2\boldsymbol{P}^\perp\right\} \\ &= \sigma_{\mathrm{n}}^2\mathrm{Tr}\left\{\boldsymbol{I}_M - \boldsymbol{P}\right\} \\ &= \sigma_{\mathrm{n}}^2\left(M-K\right). \end{aligned} \tag{3.29}$$

Furthermore, using (3.19) and the fact that the eigenvectors of $\boldsymbol{R}$ are orthonormal, $\boldsymbol{V}^\dagger\boldsymbol{V}^\dagger\boldsymbol{R}$ can be written as

$$\boldsymbol{V}^\dagger\boldsymbol{V}^\dagger\boldsymbol{R} = \sum_{k=1}^{K}\frac{\lambda_{M-K+k}}{\left(\lambda_{M-K+k}-\sigma_{\mathrm{n}}^2\right)^2}\boldsymbol{e}_k\boldsymbol{e}_k^H \tag{3.30}$$

which results in

$$\mathrm{Tr}\left\{\boldsymbol{V}^\dagger\boldsymbol{V}^\dagger\boldsymbol{R}\right\} = \sum_{k=1}^{K}\frac{\lambda_{M-K+k}}{\left(\lambda_{M-K+k}-\sigma_{\mathrm{n}}^2\right)^2}. \tag{3.31}$$

Finally, $E\left\{\rho_1\right\}$ is obtained by substituting (3.29) and (3.31) in (3.28) as

$$E\left\{\rho_1\right\} = \frac{\sigma_{\mathrm{n}}^2\left(M-K\right)}{NK}\sum_{k=1}^{K}\frac{\lambda_{M-K+k}}{\left(\lambda_{M-K+k}-\sigma_{\mathrm{n}}^2\right)^2}. \tag{3.32}$$

Subspace leakage at the second step of the proposed algorithm

The subspace leakage at the second step can be obtained through the same steps taken for the computation of ρ_1. Referring to (3.25), ρ_2 is given by

$$\rho_2 = \frac{1}{K}\mathrm{Tr}\left\{\boldsymbol{V}^\dagger\Delta\boldsymbol{R}^{(2)}\boldsymbol{P}^\perp\Delta\boldsymbol{R}^{(2)}\boldsymbol{V}^\dagger\right\} \tag{3.33}$$

where $\Delta\boldsymbol{R}^{(2)} \triangleq \widehat{\boldsymbol{R}}^{(2)} - \boldsymbol{R}$ is the estimation error of the covariance matrix at the second step. Using (3.8), $\Delta\boldsymbol{R}^{(2)}$ is given by

$$\Delta\boldsymbol{R}^{(2)} = \Delta\boldsymbol{R} - \gamma\left(\boldsymbol{T} + \boldsymbol{T}^H\right). \tag{3.34}$$

Recalling (3.5), we have $T = \widehat{\boldsymbol{P}}_A \widehat{\boldsymbol{R}} \widehat{\boldsymbol{P}}_A^{\perp}$. Next, we consider the first order Taylor series expansion of $\widehat{\boldsymbol{P}}_A$ around the true DOAs given by

$$\widehat{\boldsymbol{P}}_A \approx \boldsymbol{P}_A + d\boldsymbol{P} \tag{3.35}$$

where $\boldsymbol{P}_A \triangleq \boldsymbol{A}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H$ is equal to the true signal projection matrix, i.e., $\boldsymbol{P}_A = \boldsymbol{P}$, and $d\boldsymbol{P}$ is given by

$$d\boldsymbol{P} = \sum_{k=1}^{K} \frac{\partial \boldsymbol{P}_A}{\partial \omega_k} \Delta\omega_k \tag{3.36}$$

where $\omega_k \triangleq 2\pi(d/\lambda)\sin(\theta_k)$, and $\Delta\omega_k \triangleq \hat{\omega}_k - \omega_k$ is the error of estimation of ω_k with $\hat{\omega}_k \triangleq 2\pi(d/\lambda)\sin(\hat{\theta}_k)$. The following lemma is used for the computation of $\partial \boldsymbol{P}_A/\partial\omega_k$.

Lemma 3.3. *For any square and invertible matrix $\boldsymbol{B}$, the partial derivative of $\boldsymbol{B}^{-1}$ with respect to the variable ω is given by*

$$\frac{\partial \boldsymbol{B}^{-1}}{\partial \omega} = -\boldsymbol{B}^{-1}\frac{\partial \boldsymbol{B}}{\partial \omega}\boldsymbol{B}^{-1}. \tag{3.37}$$

Proof. We have $\boldsymbol{B}\boldsymbol{B}^{-1} = \boldsymbol{I}$. Therefore,

$$\frac{\partial \boldsymbol{B}}{\partial \omega}\boldsymbol{B}^{-1} + \boldsymbol{B}\frac{\partial \boldsymbol{B}^{-1}}{\partial \omega} = \boldsymbol{0}. \tag{3.38}$$

Then, by left multiplying all terms by $\boldsymbol{B}^{-1}$ and moving the first term to the right-hand-side, we complete the proof. □

Using Lemma 3.3, $\partial \boldsymbol{P}_A/\partial\omega_k$ can be computed as

$$\begin{aligned}
\frac{\partial \boldsymbol{P}_A}{\partial \omega_k} &= \frac{\partial \boldsymbol{A}}{\partial \omega_k}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H + \boldsymbol{A}\frac{\partial\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}}{\partial \omega_k}\boldsymbol{A}^H + \boldsymbol{A}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial \omega_k}\right)^H \\
&= \frac{\partial \boldsymbol{A}}{\partial \omega_k}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H + \boldsymbol{A}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial \omega_k}\right)^H \\
&\quad -\boldsymbol{A}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\left(\left(\frac{\partial \boldsymbol{A}}{\partial \omega_k}\right)^H\boldsymbol{A} + \boldsymbol{A}^H\frac{\partial \boldsymbol{A}}{\partial \omega_k}\right)\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H \\
&= \boldsymbol{P}^{\perp}\frac{\partial \boldsymbol{A}}{\partial \omega_k}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H + \boldsymbol{A}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial \omega_k}\right)^H\boldsymbol{P}^{\perp}.
\end{aligned} \tag{3.39}$$

The estimation error of ω_k, i.e., $\Delta\omega_k$ in (3.36) can be written based on $\Delta\boldsymbol{R}$ as [35]

$$\Delta\omega_k = \frac{\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\Delta\boldsymbol{R}\boldsymbol{V}^{\dagger}\boldsymbol{a}_k - \boldsymbol{a}_k^{H}\boldsymbol{V}^{\dagger}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\boldsymbol{a}_k^{(1)}}{2j\left(\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_k^{(1)}\right)} \tag{3.40}$$

where $\boldsymbol{a}_k$ is a shorthand notation for $\boldsymbol{a}(\theta_k)$, and $\boldsymbol{a}_k^{(1)} \in \mathbb{C}^{M\times 1}$ is defined as

$$\boldsymbol{a}_k^{(1)} \triangleq -\left[0,\ e^{-j\omega_k},\ 2e^{-j2\omega_k},\ \cdots,\ (M-1)e^{-j(M-1)\omega_k}\right]^T. \tag{3.41}$$

The first order Taylor series expansion of $\widehat{\boldsymbol{P}}_A^{\perp}$ is obtained using (3.7) and (3.35) as

$$\widehat{\boldsymbol{P}}_A^{\perp} \approx \boldsymbol{P}_A^{\perp} - d\boldsymbol{P} \tag{3.42}$$

where

$$\boldsymbol{P}_A^{\perp} \triangleq \boldsymbol{I}_M - \boldsymbol{P}_A. \tag{3.43}$$

The matrix $\boldsymbol{T}$ can be then computed using (3.5), (3.35), and (3.42) with keeping only the first order terms and noting that $\boldsymbol{P}_A = \boldsymbol{P}$, $\boldsymbol{P}_A^{\perp} = \boldsymbol{P}^{\perp}$, and $\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^{\perp} = \boldsymbol{0}$ as

$$\begin{aligned}\boldsymbol{T} &= \left(\boldsymbol{P}_A + d\boldsymbol{P}\right)\left(\boldsymbol{R} + \Delta R\right)\left(\boldsymbol{P}_A^{\perp} - d\boldsymbol{P}\right)\\ &\approx -\boldsymbol{P}\boldsymbol{R}d\boldsymbol{P} + \boldsymbol{P}\Delta R\boldsymbol{P}^{\perp} + d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^{\perp}.\end{aligned} \tag{3.44}$$

We can now compute ρ_2 using (3.33), (3.34), and (3.44) as

$$\begin{aligned}\rho_2 &= \frac{1}{K}\mathrm{Tr}\left\{\boldsymbol{V}^{\dagger}\left(\Delta\boldsymbol{R} - \gamma\left(\boldsymbol{T}+\boldsymbol{T}^H\right)\right)\boldsymbol{P}^{\perp}\left(\Delta\boldsymbol{R} - \gamma\left(\boldsymbol{T}+\boldsymbol{T}^H\right)\right)\boldsymbol{V}^{\dagger}\right\}\\
&= \frac{1}{K}\mathrm{Tr}\Big\{\boldsymbol{V}^{\dagger}\big(\Delta\boldsymbol{R} - \gamma\big(-\boldsymbol{P}\boldsymbol{R}d\boldsymbol{P} + \boldsymbol{P}\Delta R\boldsymbol{P}^{\perp} + d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^{\perp}\\
&\qquad\qquad -d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P} + \boldsymbol{P}^{\perp}\Delta R\boldsymbol{P} + \boldsymbol{P}^{\perp}\boldsymbol{R}d\boldsymbol{P}\big)\big)\\
&\qquad \times\boldsymbol{P}^{\perp}\big(\Delta\boldsymbol{R} - \gamma\big(-\boldsymbol{P}\boldsymbol{R}d\boldsymbol{P} + \boldsymbol{P}\Delta R\boldsymbol{P}^{\perp} + d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^{\perp}\\
&\qquad\qquad -d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P} + \boldsymbol{P}^{\perp}\Delta R\boldsymbol{P} + \boldsymbol{P}^{\perp}\boldsymbol{R}d\boldsymbol{P}\big)\big)\boldsymbol{V}^{\dagger}\Big\}\\
&= \frac{1}{K}\mathrm{Tr}\Big\{\boldsymbol{V}^{\dagger}\big(\Delta\boldsymbol{R} - \gamma\big(-\boldsymbol{P}\boldsymbol{R}d\boldsymbol{P} + \boldsymbol{P}\Delta R\boldsymbol{P}^{\perp} + d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^{\perp}\big)\big)\\
&\qquad \times\boldsymbol{P}^{\perp}\big(\Delta\boldsymbol{R} - \gamma\big(-d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P} + \boldsymbol{P}^{\perp}\Delta R\boldsymbol{P} + \boldsymbol{P}^{\perp}\boldsymbol{R}d\boldsymbol{P}\big)\big)\boldsymbol{V}^{\dagger}\Big\}\end{aligned} \tag{3.45}$$

where in the last step, we used (3.36), (3.39), and the facts that $\boldsymbol{P}\boldsymbol{P}^{\perp} = \boldsymbol{P}^{\perp}\boldsymbol{P} = \boldsymbol{V}^{\dagger}\boldsymbol{P}^{\perp} = \boldsymbol{P}^{\perp}\boldsymbol{V}^{\dagger} = \boldsymbol{0}$ to eliminate the terms that equal zero. Expanding the terms in (3.45) and using the fact that $\boldsymbol{P}\boldsymbol{V}^{\dagger} = \boldsymbol{V}^{\dagger}\boldsymbol{P} = \boldsymbol{V}^{\dagger}$ results in

the following expression for ρ_2

$$\begin{aligned}\rho_2 = \frac{1}{K}\mathrm{Tr}\Big\{ & \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger - \gamma\big(-\boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{R}\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger \\ & + \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp \boldsymbol{R} d\boldsymbol{P}\boldsymbol{V}^\dagger - \boldsymbol{V}^\dagger \boldsymbol{R} d\boldsymbol{P}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger \\ & + \boldsymbol{V}^\dagger d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger\big) + \gamma^2\big(\boldsymbol{V}^\dagger \boldsymbol{R} d\boldsymbol{P}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{R}\boldsymbol{V}^\dagger \\ & - \boldsymbol{V}^\dagger \boldsymbol{R} d\boldsymbol{P}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger - \boldsymbol{V}^\dagger \boldsymbol{R} d\boldsymbol{P}\boldsymbol{P}^\perp \boldsymbol{R} d\boldsymbol{P}\boldsymbol{V}^\dagger - \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{R}\boldsymbol{V}^\dagger \\ & + \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp \boldsymbol{R} d\boldsymbol{P}\boldsymbol{V}^\dagger - \boldsymbol{V}^\dagger d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{R}\boldsymbol{V}^\dagger \\ & + \boldsymbol{V}^\dagger d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^\perp \boldsymbol{R} d\boldsymbol{P}\boldsymbol{V}^\dagger\big)\Big\}. \end{aligned} \tag{3.46}$$

By reordering the terms in (3.46), ρ_2 can be further rewritten as

$$\begin{aligned}\rho_2 = \frac{1}{K}\mathrm{Tr}\Big\{ & \left(1 - 2\gamma + \gamma^2\right) \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger + \left(\gamma^2 - \gamma\right) \\ & \times\big(- \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{R}\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp \boldsymbol{R} d\boldsymbol{P}\boldsymbol{V}^\dagger \\ & - \boldsymbol{V}^\dagger \boldsymbol{R} d\boldsymbol{P}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger\big) \\ & + \gamma^2\big(\boldsymbol{V}^\dagger \boldsymbol{R} d\boldsymbol{P}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{R}\boldsymbol{V}^\dagger - \boldsymbol{V}^\dagger \boldsymbol{R} d\boldsymbol{P}\boldsymbol{P}^\perp \boldsymbol{R} d\boldsymbol{P}\boldsymbol{V}^\dagger \\ & - \boldsymbol{V}^\dagger d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{R}\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^\perp \boldsymbol{R} d\boldsymbol{P}\boldsymbol{V}^\dagger\big)\Big\}. \end{aligned} \tag{3.47}$$

The terms multiplied by $(\gamma^2 - \gamma)$ in (3.47) can be simplified using (3.18), (3.24), and the fact that $\boldsymbol{P}^\perp \boldsymbol{V} = \boldsymbol{0}$ as

$$\begin{aligned} & -\boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp d\boldsymbol{P}\left(\boldsymbol{V} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}_M\right)\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp \left(\boldsymbol{V} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}_M\right) d\boldsymbol{P}\boldsymbol{V}^\dagger \\ & -\boldsymbol{V}^\dagger \left(\boldsymbol{V} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}_M\right) d\boldsymbol{P}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger d\boldsymbol{P}\left(\boldsymbol{V} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}_M\right)\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger \\ & = -\boldsymbol{V}^\dagger \Delta\boldsymbol{R}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{P} - \boldsymbol{P} d\boldsymbol{P}\boldsymbol{P}^\perp \Delta\boldsymbol{R}\boldsymbol{V}^\dagger. \end{aligned} \tag{3.48}$$

In a similar way, the terms multiplied by γ^2 in (3.47) are simplified as

$$\begin{aligned} & \boldsymbol{V}^\dagger \boldsymbol{R} d\boldsymbol{P}\boldsymbol{P}^\perp d\boldsymbol{P}\left(\boldsymbol{V} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}_M\right)\boldsymbol{V}^\dagger - \boldsymbol{V}^\dagger \boldsymbol{R} d\boldsymbol{P}\boldsymbol{P}^\perp \left(\boldsymbol{V} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}_M\right) d\boldsymbol{P}\boldsymbol{V}^\dagger \\ & -\boldsymbol{V}^\dagger d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^\perp d\boldsymbol{P}\left(\boldsymbol{V} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}_M\right)\boldsymbol{V}^\dagger + \boldsymbol{V}^\dagger d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^\perp \left(\boldsymbol{V} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}_M\right) d\boldsymbol{P}\boldsymbol{V}^\dagger \\ & = \boldsymbol{V}^\dagger \boldsymbol{R} d\boldsymbol{P}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{P} - \boldsymbol{V}^\dagger d\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{P} \\ & = \boldsymbol{V}^\dagger \left(\boldsymbol{V} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}_M\right) d\boldsymbol{P}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{P} - \boldsymbol{V}^\dagger d\boldsymbol{P}\left(\boldsymbol{V} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}_M\right)\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{P} \\ & = \boldsymbol{P} d\boldsymbol{P}\boldsymbol{P}^\perp d\boldsymbol{P}\boldsymbol{P} \\ & = \left(\boldsymbol{I}_M - \boldsymbol{P}^\perp\right) d\boldsymbol{P}\boldsymbol{P}^\perp d\boldsymbol{P}\left(\boldsymbol{I}_M - \boldsymbol{P}^\perp\right) \\ & = d\boldsymbol{P}\boldsymbol{P}^\perp d\boldsymbol{P} \end{aligned} \tag{3.49}$$

where in the last step, we used the fact that $\boldsymbol{P}^{\perp} d\boldsymbol{P}\boldsymbol{P}^{\perp} = \boldsymbol{0}$ (see (3.36) and (3.39)).

Finally, using (3.25), (3.47), (3.48), (3.49), and Lemma 3.1, ρ_2 is computed as

$$\begin{aligned}\rho_2 = \left(1 - 2\gamma + \gamma^2\right)\rho_1 + \frac{2\left(\gamma - \gamma^2\right)}{K} Re\left\{\mathrm{Tr}\left\{\boldsymbol{V}^{\dagger}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp} d\boldsymbol{P}\right\}\right\} \\ + \frac{\gamma^2}{K}\mathrm{Tr}\left\{d\boldsymbol{P}\boldsymbol{P}^{\perp} d\boldsymbol{P}\right\}. \end{aligned} \tag{3.50}$$

Computation of the expected value of ρ_2 involves finding the expected value of the two trace functions in (3.50). Using (3.36), (3.39), and (3.40), the expected value of the first trace function in (3.50) is given by

$$\begin{aligned} & E\left\{\mathrm{Tr}\left\{\boldsymbol{V}^{\dagger}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp} d\boldsymbol{P}\right\}\right\} \\ & \quad = E\left\{\mathrm{Tr}\left\{\Delta\boldsymbol{R}\sum_{k=1}^{K}\boldsymbol{P}^{\perp}\frac{\partial\boldsymbol{A}}{\partial\omega_k}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H\Delta\omega_k\boldsymbol{V}^{\dagger}\right\}\right\} \\ & \quad = E\Bigg\{\mathrm{Tr}\Bigg\{\sum_{k=1}^{K}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\frac{\partial\boldsymbol{A}}{\partial\omega_k}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H\boldsymbol{V}^{\dagger}\frac{1}{2j\left(\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_k^{(1)}\right)} \\ & \qquad\qquad \times\left(\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\Delta\boldsymbol{R}\boldsymbol{V}^{\dagger}\boldsymbol{a}_k - \boldsymbol{a}_k^{H}\boldsymbol{V}^{\dagger}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\boldsymbol{a}_k^{(1)}\right)\Bigg\}\Bigg\} \\ & \quad = E\Bigg\{\sum_{k=1}^{K}\frac{1}{2j\left(\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_k^{(1)}\right)}\mathrm{Tr}\Bigg\{\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\frac{\partial\boldsymbol{A}}{\partial\omega_k}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H\boldsymbol{V}^{\dagger}\Bigg\} \\ & \qquad\qquad \times\left(\mathrm{Tr}\left\{\Delta\boldsymbol{R}\boldsymbol{V}^{\dagger}\boldsymbol{a}_k\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\right\} - \mathrm{Tr}\left\{\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\boldsymbol{a}_k^{(1)}\boldsymbol{a}_k^{H}\boldsymbol{V}^{\dagger}\right\}\right)\Bigg\} \end{aligned} \tag{3.51}$$

which is computed using (3.27) as

$$\begin{aligned} & E\left\{\mathrm{Tr}\left\{\boldsymbol{V}^{\dagger}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp} d\boldsymbol{P}\right\}\right\} \\ & \quad = \frac{1}{N}\sum_{k=1}^{K}\frac{1}{2j\left(\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_k^{(1)}\right)} \\ & \qquad\qquad \times\Bigg(\mathrm{Tr}\Bigg\{\boldsymbol{R}\boldsymbol{P}^{\perp}\frac{\partial\boldsymbol{A}}{\partial\omega_k}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H\boldsymbol{V}^{\dagger}\boldsymbol{R}\boldsymbol{V}^{\dagger}\boldsymbol{a}_k\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\Bigg\} \\ & \qquad\qquad\quad -\mathrm{Tr}\Bigg\{\boldsymbol{R}\boldsymbol{P}^{\perp}\frac{\partial\boldsymbol{A}}{\partial\omega_k}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H\boldsymbol{V}^{\dagger}\boldsymbol{R}\boldsymbol{P}^{\perp}\boldsymbol{a}_k^{(1)}\boldsymbol{a}_k^{H}\boldsymbol{V}^{\dagger}\Bigg\}\Bigg). \end{aligned} \tag{3.52}$$

The second trace function in (3.52) equals zero as $\boldsymbol{V}^{\dagger}\boldsymbol{R}\boldsymbol{P}^{\perp}=\boldsymbol{0}$. Then, (3.52) can be rewritten as

$$E\left\{\mathrm{Tr}\left\{\boldsymbol{V}^{\dagger}\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}d\boldsymbol{P}\right\}\right\}=\frac{\sigma_{\mathrm{n}}^{2}}{N}\sum_{k=1}^{K}\frac{\boldsymbol{a}_{k}^{(1)H}\boldsymbol{P}^{\perp}\frac{\partial\boldsymbol{A}}{\partial\omega_{k}}\left(\boldsymbol{A}^{H}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{H}\boldsymbol{V}^{\dagger}\boldsymbol{R}\boldsymbol{V}^{\dagger}\boldsymbol{a}_{k}}{2j\left(\boldsymbol{a}_{k}^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_{k}^{(1)}\right)} \tag{3.53}$$

where we used the equality $\boldsymbol{P}^{\perp}\boldsymbol{R}=\sigma_{\mathrm{n}}^{2}\boldsymbol{P}^{\perp}$. In a similar way, the expected value of the second trace function in (3.50) is given by

$$\begin{aligned}
&E\left\{\mathrm{Tr}\left\{d\boldsymbol{P}\boldsymbol{P}^{\perp}d\boldsymbol{P}\right\}\right\}\\
&\quad=E\Bigg\{\mathrm{Tr}\Bigg\{\sum_{k=1}^{K}\sum_{i=1}^{K}\boldsymbol{A}\left(\boldsymbol{A}^{H}\boldsymbol{A}\right)^{-1}\left(\frac{\partial\boldsymbol{A}}{\partial\omega_{k}}\right)^{H}\boldsymbol{P}^{\perp}\frac{\partial\boldsymbol{A}}{\partial\omega_{i}}\left(\boldsymbol{A}^{H}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{H}\Delta\omega_{k}\Delta\omega_{i}\Bigg\}\Bigg\}\\
&\quad=E\Bigg\{\mathrm{Tr}\Bigg\{\sum_{k=1}^{K}\sum_{i=1}^{K}\boldsymbol{A}\left(\boldsymbol{A}^{H}\boldsymbol{A}\right)^{-1}\left(\frac{\partial\boldsymbol{A}}{\partial\omega_{k}}\right)^{H}\boldsymbol{P}^{\perp}\frac{\partial\boldsymbol{A}}{\partial\omega_{i}}\left(\boldsymbol{A}^{H}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{H}\\
&\qquad\qquad\times\frac{1}{2j\left(\boldsymbol{a}_{k}^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_{k}^{(1)}\right)}\times\frac{1}{2j\left(\boldsymbol{a}_{i}^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_{i}^{(1)}\right)}\\
&\qquad\qquad\times\left(\mathrm{Tr}\left\{\Delta\boldsymbol{R}\boldsymbol{V}^{\dagger}\boldsymbol{a}_{k}\boldsymbol{a}_{k}^{(1)H}\boldsymbol{P}^{\perp}\right\}-\mathrm{Tr}\left\{\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\boldsymbol{a}_{k}^{(1)}\boldsymbol{a}_{k}^{H}\boldsymbol{V}^{\dagger}\right\}\right)\\
&\qquad\qquad\times\left(\mathrm{Tr}\left\{\Delta\boldsymbol{R}\boldsymbol{V}^{\dagger}\boldsymbol{a}_{i}\boldsymbol{a}_{i}^{(1)H}\boldsymbol{P}^{\perp}\right\}-\mathrm{Tr}\left\{\Delta\boldsymbol{R}\boldsymbol{P}^{\perp}\boldsymbol{a}_{i}^{(1)}\boldsymbol{a}_{i}^{H}\boldsymbol{V}^{\dagger}\right\}\right)\Bigg\}\Bigg\}
\end{aligned} \tag{3.54}$$

which is computed using (3.27) and the fact that $\boldsymbol{P}^{\perp}\boldsymbol{R}\boldsymbol{V}^{\dagger}=\boldsymbol{0}$ as

$$\begin{aligned}
E\left\{\mathrm{Tr}\left\{d\boldsymbol{P}\boldsymbol{P}^{\perp}d\boldsymbol{P}\right\}\right\}=\frac{\sigma_{\mathrm{n}}^{2}}{2N}\sum_{k=1}^{K}\sum_{i=1}^{K}&\frac{\mathrm{Tr}\left\{\left(\frac{\partial\boldsymbol{A}}{\partial\omega_{k}}\right)^{H}\boldsymbol{P}^{\perp}\frac{\partial\boldsymbol{A}}{\partial\omega_{i}}\left(\boldsymbol{A}^{H}\boldsymbol{A}\right)^{-1}\right\}}{\left(\boldsymbol{a}_{k}^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_{k}^{(1)}\right)\left(\boldsymbol{a}_{i}^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_{i}^{(1)}\right)}\\
&\times Re\left\{\boldsymbol{a}_{i}^{H}\boldsymbol{V}^{\dagger}\boldsymbol{R}\boldsymbol{V}^{\dagger}\boldsymbol{a}_{k}\boldsymbol{a}_{k}^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_{i}^{(1)}\right\}.
\end{aligned} \tag{3.55}$$

Finally, the expected value of ρ_2 for a fixed value of γ is obtained using (3.50),

(3.53), and (3.55) as

$$E\{\rho_2\} = \left(1-2\gamma+\gamma^2\right)E\{\rho_1\} + \frac{2\left(\gamma-\gamma^2\right)\sigma_{\mathrm{n}}^2}{NK} Re\left\{\sum_{k=1}^{K}\frac{\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\frac{\partial \boldsymbol{A}}{\partial \omega_k}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H\boldsymbol{V}^{\dagger}\boldsymbol{R}\boldsymbol{V}^{\dagger}\boldsymbol{a}_k}{2j\left(\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_k^{(1)}\right)}\right\} + \frac{\gamma^2\sigma_{\mathrm{n}}^2}{2NK}\sum_{k=1}^{K}\sum_{i=1}^{K}\frac{\mathrm{Tr}\left\{\left(\frac{\partial \boldsymbol{A}}{\partial \omega_k}\right)^H\boldsymbol{P}^{\perp}\frac{\partial \boldsymbol{A}}{\partial \omega_i}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\right\}}{\left(\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_k^{(1)}\right)\left(\boldsymbol{a}_i^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_i^{(1)}\right)} \times Re\left\{\boldsymbol{a}_i^H\boldsymbol{V}^{\dagger}\boldsymbol{R}\boldsymbol{V}^{\dagger}\boldsymbol{a}_k\boldsymbol{a}_k^{(1)H}\boldsymbol{P}^{\perp}\boldsymbol{a}_i^{(1)}\right\}. \quad (3.56)$$

3.3 Numerical Examples and Simulation Results

In this section, the performance of the proposed two-step root-MUSIC and the root-swap root-MUSIC algorithms is investigated and compared with the performance of the unitary root-MUSIC method and the improved unitary root-MUSIC algorithm based on pseudo-noise resampling [24]. We also consider the combination of the proposed methods with the other methods.

We consider $K = 2$ sources impinging on an array of $M = 10$ antenna elements from directions $\theta_1 = 35^\circ \times (\pi/180)$ and $\theta_2 = 37^\circ \times (\pi/180)$. The interelement spacing is set to $d = \lambda/2$ and the number of snapshots is $N = 10$. Each source vector $\boldsymbol{s}(t)$ is considered to be independent from the source vectors at other time instances and to have the circularly-symmetric complex jointly-Gaussian distribution $\mathcal{N}_C(0, \boldsymbol{S})$. The source covariance matrix $\boldsymbol{S}$ is given by

$$\boldsymbol{S} = \sigma_{\mathrm{s}}^2\begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \quad (3.57)$$

where $0 \leq r \leq 1$ is the correlation coefficient. The SNR is defined as $\mathrm{SNR} \triangleq 10\log_{10}\left(\sigma_{\mathrm{s}}^2/\sigma_{\mathrm{n}}^2\right)$.

The performance of the proposed algorithms is investigated by considering the subspace leakage, the MSE, the detection probability, and the conditional MSE (CMSE). The detection probability is defined as the probability of successful detection which is in turn defined as the event that both DOA's are estimated within one degree of their corresponding true values, i.e., the difference

between the true value of each DOA and its estimated value is less than $1^\circ \times (\pi/180)$. The CMSE is defined as $E\left\{\sum_{k=1}^{K} \|\hat{\theta}_k - \theta_k\|_2^2 \;\middle|\; \text{successful detection}\right\}$ which is the expected value of the estimation error conditioned on successful detection. The reason for defining the CMSE in the above manner is to further investigate the accuracy of the algorithms after making successful detection. We estimate the subspace leakage, the MSE, the detection probability, and the CMSE using the Monte Carlo method with 10^5 number of trials. Two cases are considered in the simulations: 1) the two sources are uncorrelated, i.e., $r = 0$, and 2) the two sources are correlated with a correlation coefficient of $r = 0.9$.

Let us start by investigating the subspace leakage in the two-step root-MUSIC algorithm for the case of the uncorrelated sources. The expected value of the subspace leakage is estimated using (3.16) and the Monte Carlo simulations. The approximate value for the subspace leakage is also obtained from the theoretical derivations in (3.32) and (3.56). The value of γ is fixed at 0.5. The results are shown in Fig. 3.1. The solid lines represent the subspace leakage at the first step, and the dashed lines depict the subspace leakage at the second step of the proposed two-step root-MUSIC algorithm. It can be seen that the curves obtained from the simulations are very close to those obtained from our theoretical derivations at high SNR values. At the low SNR region, the curve associated with the theoretical approximation at the second step deviates from the curve obtained by simulations. The reason is that in the derivations, only the first order Taylor series expansion is used. More accurate results can be obtained by using higher order Taylor series. However, the computations can become intractable. In Fig. 3.1, it can be observed from both theoretical and simulation results that the subspace leakage from the modified covariance matrix at the second step is significantly smaller than the subspace leakage from the sample data covariance matrix at the first step. This is achieved by removing the undesirable terms from the sample data covariance matrix leading to an estimate of the signal projection matrix that is closer to the true signal projection matrix, which is equivalent to a lower subspace leakage at the second step.

We next consider the performance of the proposed two-step algorithm when applied to the root-MUSIC, unitary root-MUSIC, improved unitary root-MUSIC with pseudo-noise resampling, root-swap unitary root-MUSIC, and root-swap unitary root-MUSIC with pseudo-noise resampling methods. The value of the scaling factor γ is obtained by minimizing the SML function as described in the steps of the two-step root-MUSIC method. In the root-swap algorithm, the parameters p and q are set to $p = 1$ and $q = 0$, which means the closest root to the unit circle is picked up and paired with the other roots one at a time in order to find the pair of DOA estimates that minimizes the SML function. In this case, the number of different possible combinations of the roots is $N_r = 8$. The number of samples used for the pseudo-noise resampling method is set to $P = 50$. According to our simulations, using more number of samples would not yield in any considerable improvement in the performance.

The MSE versus SNR performance of the methods tested for the case of the uncorrelated sources is presented in Fig. 3.2. The CRB [18] is also shown in the figure. For the root-MUSIC method, the modification of the covariance matrix in the second step of the introduced two-step method shifts the MSE curve by almost half a dB to the left. For the unitary root-MUSIC method the improvement is more significant and is about one dB. For the rest of the methods, there is no considerable change in the MSE performance. However, as it will be shown in the next figures, the modification of the covariance matrix has benefits in terms of the CMSE performance and the detection probability for these methods. It can also be seen from Fig. 3.2 that the proposed root-swap unitary root-MUSIC algorithm performs about 2 dB better than the unitary root-MUSIC method, while imposing only a small amount of computational complexity for evaluating the SML function for $N_r = 8$ different combinations of the roots. The best performance is achieved by the root-swap unitary root-MUSIC algorithm combined with the pseudo-noise resampling method.

Fig. 3.3 shows the detection probability versus SNR for the uncorrelated sources. For the root-MUSIC method, the second step of the two-step algorithm improves the performance by 1 to 2 dB. The rest of the algorithms have almost the same performance with the root-swap based methods slightly out-

performing the other algorithms at low SNR values. It is observed that the second step of the two-step algorithm results in about 1 dB improvement in the performance.

Finally, Fig. 3.4 illustrates the performance of the algorithms for the uncorrelated sources in terms of the CMSE. The root-MUSIC method is significantly improved by the two-step method with an improvement ranging from 5 dB at low SNR values to 1 dB at high SNR values. The rest of the algorithms show similar performance, and the application of the two-step method leads to up to 2 dB improvement in the CMSE performance.

The results for the case of the correlated sources with $r = 0.9$ are depicted in Figs. 3.5, 3.6, 3.7, and 3.8. Similar observations are made from these figures as those discussed for the case of uncorrelated sources.

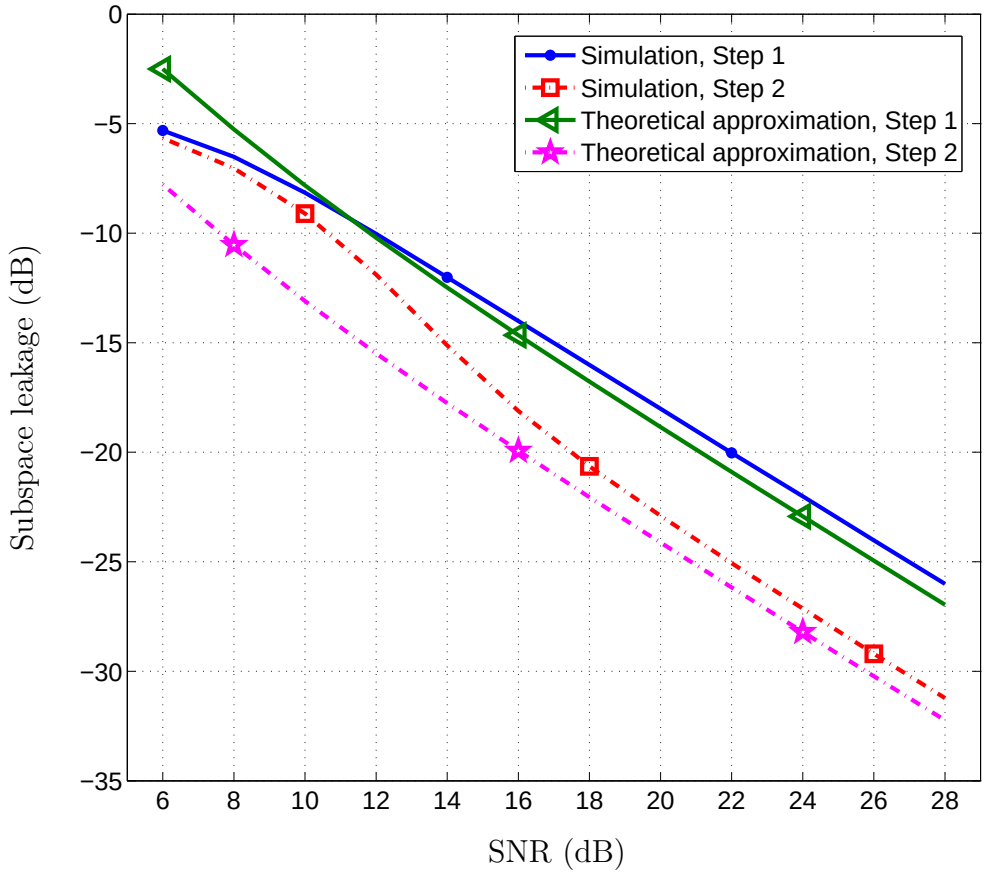

Figure 3.1: Subspace leakage versus SNR for uncorrelated sources. The solid and dashed lines represent the subspace leakage at the first and second steps of the proposed two-step root-MUSIC algorithm, respectively.

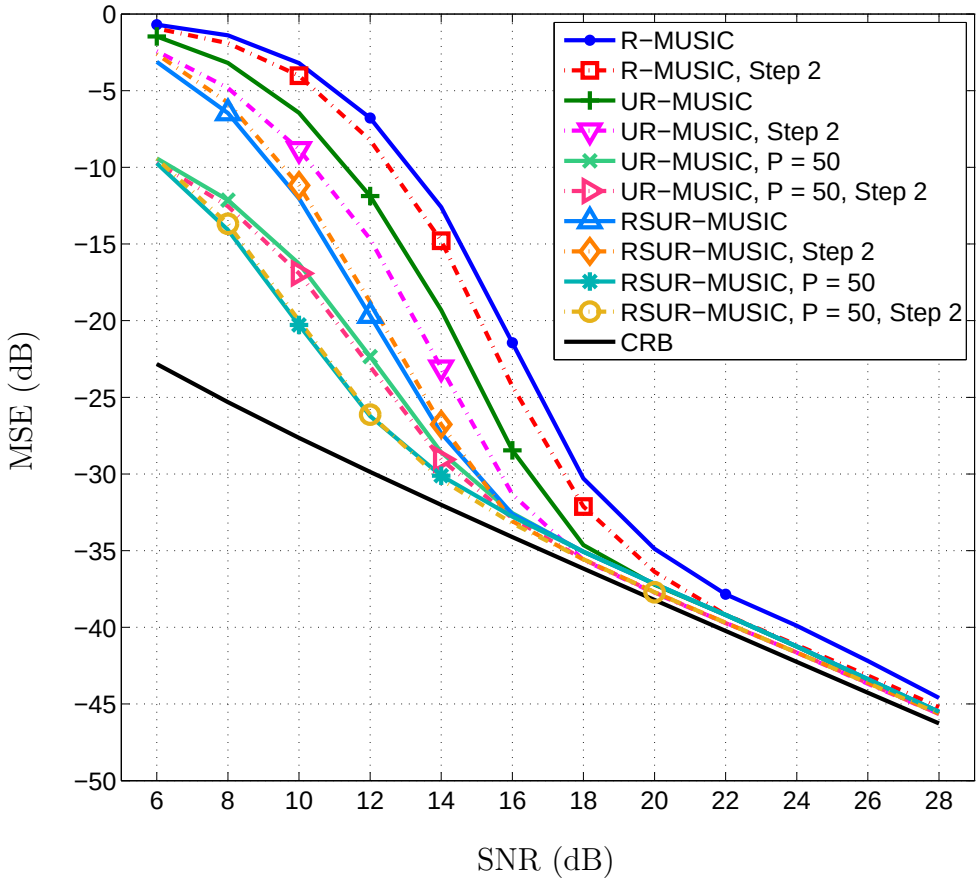

Figure 3.2: MSE versus SNR for uncorrelated sources. The solid and dashed lines are based on the first and second steps of the proposed two-step method, respectively. The methods used in the two-step algorithm are root-MUSIC (R-MUSIC), unitary root-MUSIC (UR-MUSIC), and root-swap unitary root-MUSIC (RSUR-MUSIC) methods. P is the number of samples used for the pseudo-noise resampling algorithm.

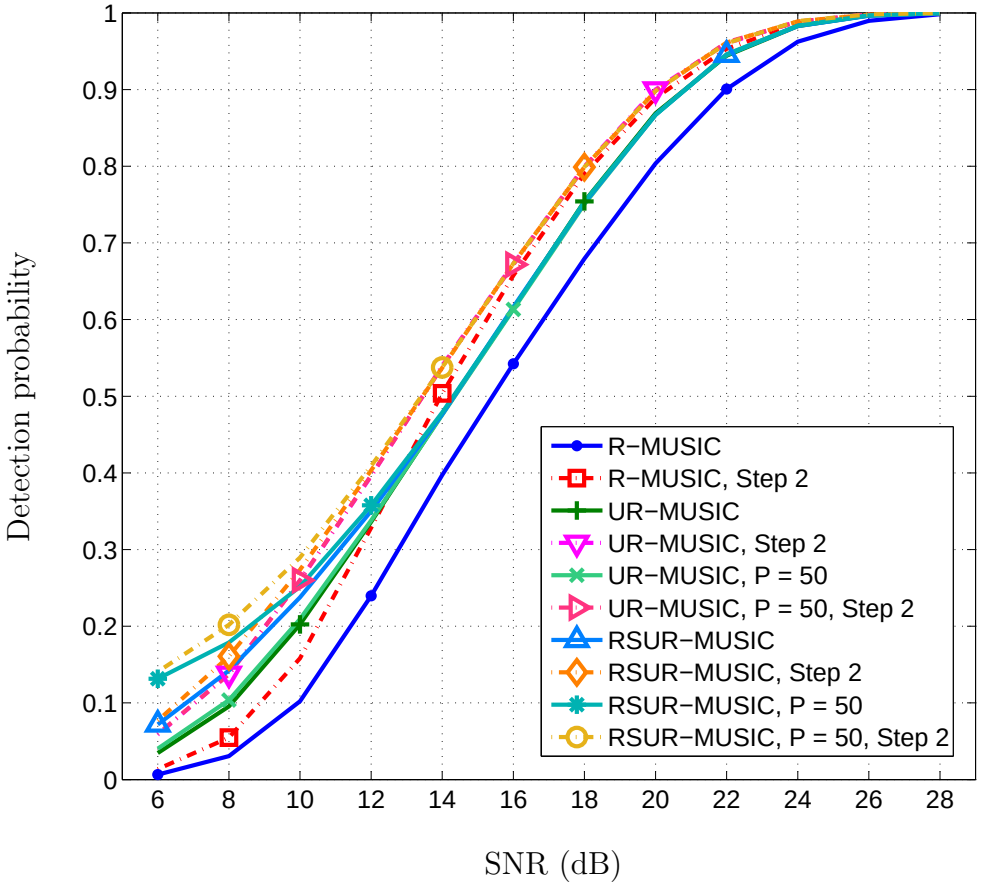

Figure 3.3: Detection probability versus SNR for uncorrelated sources. The solid and dashed lines are based on the first and second steps of the proposed two-step method, respectively. The methods used in the two-step algorithm are R-MUSIC, UR-MUSIC, and RSUR-MUSIC methods.

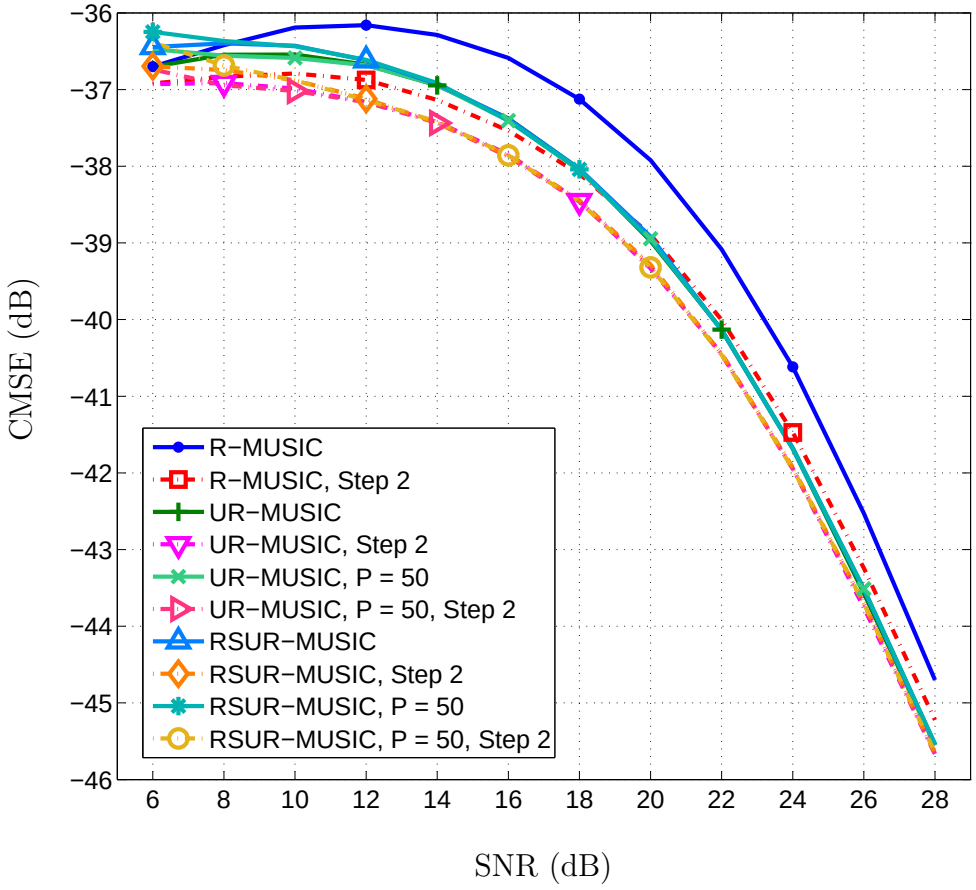

Figure 3.4: CMSE versus SNR for uncorrelated sources. The solid and dashed lines are based on the first and second steps of the proposed two-step method, respectively. The methods used in the two-step algorithm are R-MUSIC, UR-MUSIC, and RSUR-MUSIC methods.

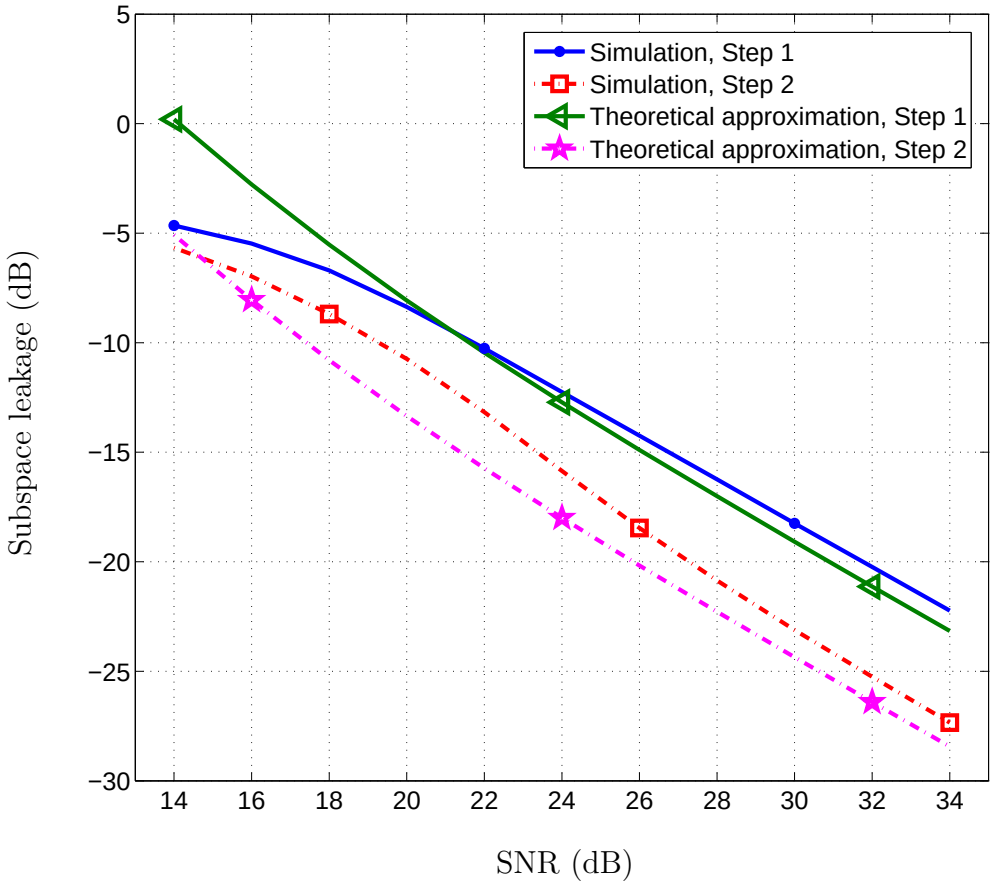

Figure 3.5: Subspace leakage versus SNR for correlated sources with $r = 0.9$. The solid and dashed lines represent the subspace leakage at the first and second steps of the proposed two-step R-MUSIC algorithm, respectively.

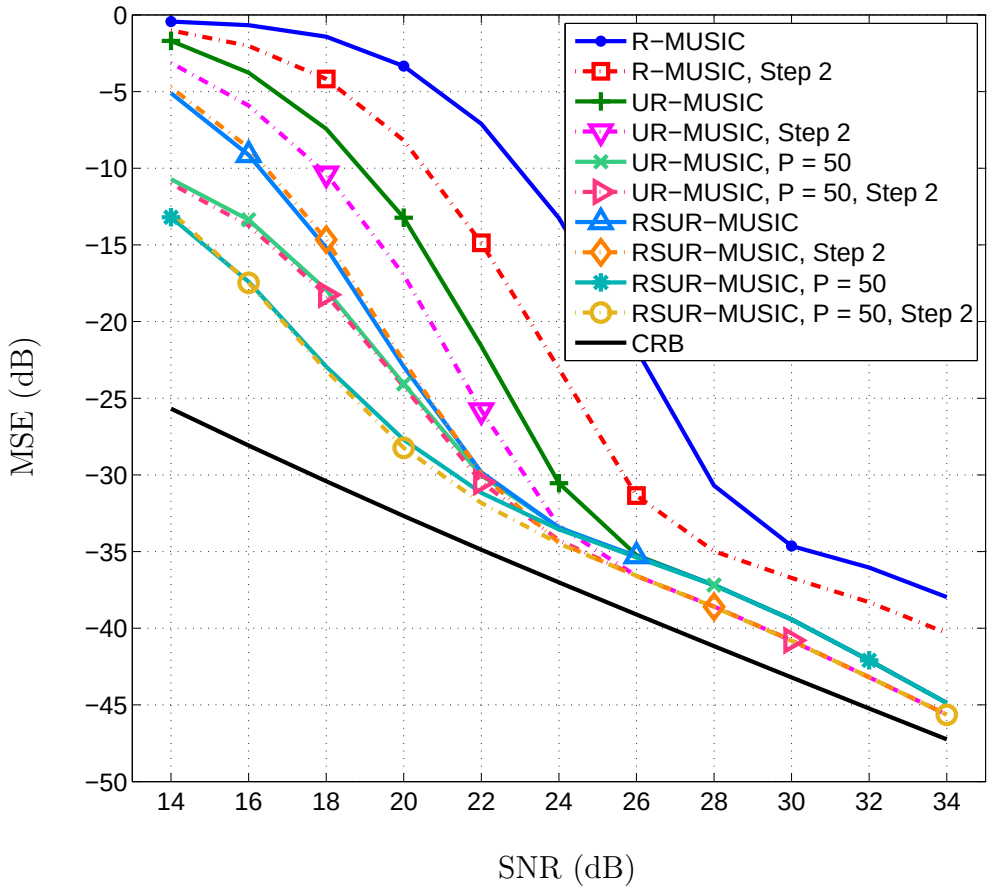

Figure 3.6: MSE versus SNR for correlated sources with $r = 0.9$. The solid and dashed lines are based on the first and second steps of the proposed two-step method, respectively. The methods used in the two-step algorithm are R-MUSIC, UR-MUSIC, and RSUR-MUSIC methods.

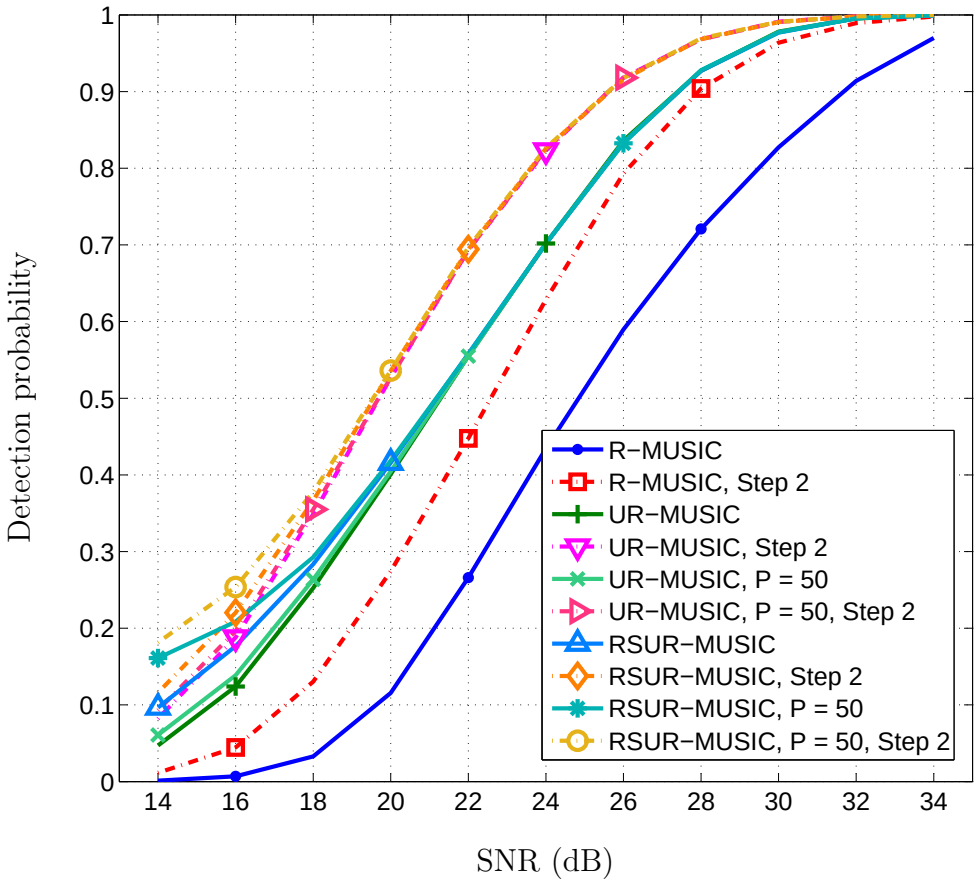

Figure 3.7: Detection probability versus SNR for correlated sources with $r = 0.9$. The solid and dashed lines are based on the first and second steps of the proposed two-step method, respectively. The methods used in the two-step algorithm are R-MUSIC, UR-MUSIC, and RSUR-MUSIC methods.

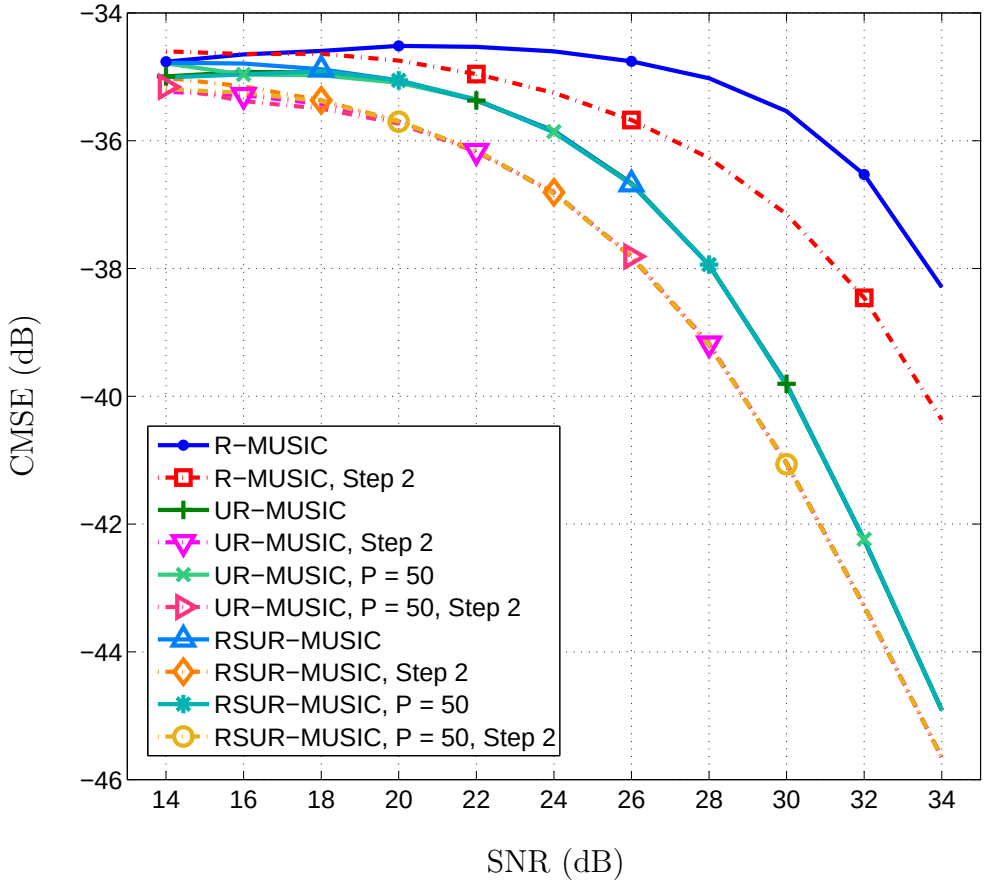

Figure 3.8: CMSE versus SNR for correlated sources with $r = 0.9$. The solid and dashed lines are based on the first and second steps of the proposed two-step method, respectively. The methods used in the two-step algorithm are R-MUSIC, UR-MUSIC, and RSUR-MUSIC methods.

Chapter 4

Averaged Correlogram for Undersampled Data and Its Finite-Length and Asymptotic Analysis

Related to DOA estimation in spacial domain, the spectrum estimation problem in frequency domain from a finite set of noisy measurements (samples) is a classical problem with wide applications in communications, astronomy, seismology, radar, sonar signal processing, etc. [37], [38]. For practical applications, not only the number of available samples is finite, but also the rate at which these samples are collected can be limited. The latter can lead to aliasing. A sufficient condition for alias-free sampling is to sample the signal of interest at the Nyquist rate which is twice the maximum frequency of the signal. In practice however, the sampling rate can be restricted. As an example, in the case of wideband spectrum sensing, there may not exist a fast enough sampling hardware based on the current technology. If such a high rate sampler exists, it can be quite costly. Therefore, it is desirable to make spectrum estimation from measurements obtained at a rate lower than the Nyquist rate.

Alias-free sampling was first introduced in [39], and further studied in [40] and [41]. In [39], it is demonstrated that if the average sampling rate is below the Nyquist rate, alias-free sampling cannot be obtained by uniform sampling, and therefore nonuniform sampling is considered. In [40], it is shown that

sampling at an arbitrarily small fraction of the Nyquist rate can be alias-free, but error-free estimation of the spectrum may not be generally possible. In [41], a new definition of alias-free sampling is given, and based on it, consistent estimators are derived. However, the samples are considered to be collected at random times which do not necessarily lie on a time grid or lattice. This makes the hardware implementation of such methods very difficult.

In [42] and [43], authors have studied signal reconstruction from sub-Nyquist samples which form a subset of Nyquist samples (samples obtained at the Nyquist rate). The methods in these works aim at reconstructing the signal, whereas one might be only interested in recovering the spectral content of the signal. In [44], authors have shown that for signals with sparse Fourier representations, i.e., signals which have only a few nonzero coefficients in the Fourier basis, the Fourier coefficients can be estimated using a subset of Nyquist samples. In [45], power spectral density (PSD) estimation based on CS techniques [9] with applications to wideband cognitive radios has been introduced.

For all of the above mentioned methods, the sparsity of the signal is a requirement for successful recovery of the spectrum. In [46, 47], PSD estimation from a subset of Nyquist samples has been considered. The introduced method is able to estimate the PSD from undersampled data without essentially requiring the signal to be sparse. We refer to this method as the *averaged correlogram for undersampled data*. In this method, samples are collected using multiple channels, each operating at a rate L times lower than the Nyquist rate. This method of sampling is known as the *multi-coset sampling* [48]. The averaged correlogram for undersampled data partitions the spectrum into L segments (subbands), and it estimates the average power within each spectral segment. The frequency resolution of the estimator is given by the width of each spectral segment. In this chapter, we equivalently use the number of spectral segments L as the frequency resolution of the estimator (with larger L meaning higher resolution or narrower segments).

Similar methods to the averaged correlogram for undersampled data which are based on multi-coset sampling have also been reported in [49] and [50].

The method in [49] relies on the sparsity of the signal, and in the case that the bandwidth of the active subbands is equal to the total bandwidth, the sampling is performed at the Nyquist rate. In [50], prior to the sampling stage, the input signal is multiplied by a waveform which has a bandwidth larger than the original signal. Furthermore, the introduced methods estimate the PSD at each individual frequency, whereas the averaged correlogram for undersampled data estimates the average power within each subband. As a result, the averaged correlogram for undersampled data is less complex in terms of implementation and computational burden.

The averaged correlogram for undersampled data as mentioned above estimates the PSD from sub-Nyquist samples without necessarily imposing sparsity conditions on the signal. However, the frequency resolution of the estimator is restricted to the number of spectral segments, and the estimation made for each segment also has limited accuracy. Therefore, it is of significant importance to analyze the performance of this method especially in the case that only a finite number of samples is available. Our major contribution is that we derive the bias and variance of the averaged correlogram for undersampled data for finite-length signals, and we show how they are related to various parameters by formulating the associated tradeoffs among the resolution, the accuracy, and the complexity of the method.

We first study the averaged correlogram for undersampled data by computing the bias of the estimator in Section 4.3. Next, the covariance matrix of the estimator is derived in Section 4.4, and using these derivations, we show that for finite-length signals, there exists a tradeoff between the estimation accuracy, the frequency resolution, and the complexity of the estimator. For the case of a white Gaussian process, we derive a closed-form expression for the estimation variance, which clearly shows how the variance is related to different parameters. Moreover, we prove that the estimation bias and variance tend to zero asymptotically. Therefore, in the case of a white Gaussian process, the averaged correlogram for undersampled data is a consistent estimator. Furthermore, we show that this method makes uncorrelated estimations for different spectral segments as the signal length goes to infinity [51].

4.1 Averaged Correlogram for Undersampled Data

Consider a wide-sense stationary (WSS) stochastic process $x(t)$ bandlimited to $W/2$ Hz with PSD $P_x(f)$. Let $x(t)$ be sampled using a multi-coset (MC) sampler which collects the samples by a multi-channel system. The i-th channel ($1 \leq i \leq q$) samples $x(t)$ at the time instants $t = (nL + c_i)T$ for $n = 0,\ 1,\ 2,\ \ldots$, where T is the Nyquist period ($T = 1/W$), L is a suitable positive integer, and $q < L$ is the number of sampling channels. The time offsets c_i ($1 \leq i \leq q$) are distinct nonnegative integer numbers less than L, and the set $\{c_i\}$ is referred to as the *sampling pattern*. Let the output of the i-th channel be denoted by $y_i(n) = x\left((nL + c_i)T\right)$. The i-th channel can be implemented by a system that shifts $x(t)$ by c_iT seconds and then samples uniformly at a rate of $1/(LT)$ Hz. The samples obtained in this manner form a subset of Nyquist samples. The average sampling rate is $q/(LT)$ Hz, and it is less than the Nyquist rate since $q < L$.

Given the MC samples, the first step of the averaged correlogram for undersampled data method is to undo the time shift that each channel imposes on the signal. Let $z_i(n)$ be defined as $y_i(n)$ delayed by a fractional delay equal to c_i/L. Let also a and b denote two channel indices. It is shown in [47] that the cross-correlation function $r_{z_az_b}(k) = E\{z_a(n+k)z_b^*(n)\}$ at $k = 0$ is given by

$$r_{z_az_b}(0) = \sum_{l=1}^{L} e^{-j\frac{2\pi}{L}(c_a - c_b)m_l} P_x(m_l) \tag{4.1}$$

where $E\{\cdot\}$ stands for the expectation operator, L is an odd number, $m_l = -\frac{1}{2}(L+1) + l$, and $P_x(m_l)$ is defined as

$$P_x(m_l) \triangleq \int_{-\frac{W}{2L}}^{\frac{W}{2L}} P_x\left(f - \frac{W}{L}m_l\right) df. \tag{4.2}$$

Consider partitioning the bandwidth of $x(t)$ into L equal segments. Then, for a given m_l, $\frac{L}{W}P_x(m_l)$ is equal to the average power of the process $x(t)$ within the spectral segment $\left[\frac{W}{2} - \frac{W}{L}l,\ \frac{W}{2} - \frac{W}{L}(l-1)\right)$.

Let us arrange the elements of the cross-correlation function $r_{z_az_b}(0)$ ($1 \leq a, b \leq q$) in a matrix $\boldsymbol{R}_z \in \mathbb{C}^{q\times q}$ such that $[\boldsymbol{R}_z]_{a,b} = r_{z_az_b}(0)$. Note that $\boldsymbol{R}_z$ is

a Hermitian matrix with equal diagonal elements. Then, it is sufficient to let the indices a and b just refer to the elements of the upper triangle and the first diagonal element of $\boldsymbol{R}_z$. Therefore, there are $Q = q(q-1)/2 + 1$ equations of type (4.1). In matrix-vector form, (4.1) can be rewritten as

$$\boldsymbol{u} = \boldsymbol{\Psi}\boldsymbol{v} \tag{4.3}$$

where $\boldsymbol{v} = [v_1,\ v_2,\ \ldots,\ v_L]^T \in \mathbb{R}^{L\times 1}$ consists of the elements $v_l = P_x(m_l)$, $\boldsymbol{u} = [u_1,\ u_2,\ \ldots,\ u_Q]^T \in \mathbb{C}^{Q\times 1}$ is composed of $u_1 = [\boldsymbol{R}_z]_{1,1}$ and $u_2, \ldots,\ u_Q$ corresponding to the elements of the upper triangle of $\boldsymbol{R}_z$, and $\boldsymbol{\Psi} \in \mathbb{C}^{Q\times L}$ consists of the elements given by

$$[\boldsymbol{\Psi}]_{k,l} = e^{-j\nu_k m_l} \tag{4.4}$$

where $\nu_k = \frac{2\pi}{L}(c_a - c_b)$, $(1 \le l \le L$ and $1 \le k \le Q)$. Note that a and b are obtained from k based on the arrangement of the elements of $\boldsymbol{R}_z$ in $\boldsymbol{u}$.

Since the elements of $\boldsymbol{v}$ are real-valued, the number of equations in (4.3) can be doubled by solving $\breve{\boldsymbol{u}} = \breve{\boldsymbol{\Psi}}\boldsymbol{v}$, where $\breve{\boldsymbol{u}} \triangleq [Re(\boldsymbol{u}),\ Im(\boldsymbol{u})]^T \in \mathbb{R}^{2Q\times 1}$ and $\breve{\boldsymbol{\Psi}} \triangleq [Re(\boldsymbol{\Psi}),\ Im(\boldsymbol{\Psi})]^T \in \mathbb{R}^{2Q\times L}$. Doubling the number of equations is beneficial in turning an underdetermined system of equations into an overdetermined system.

Suppose $\breve{\boldsymbol{\Psi}}$ is full rank and $2Q \ge L$. Then, $\breve{\boldsymbol{u}} = \breve{\boldsymbol{\Psi}}\boldsymbol{v}$ is an overdetermined system and $\boldsymbol{v}$ can be obtained using the pseudoinverse of $\breve{\boldsymbol{\Psi}}$ as

$$\boldsymbol{v} = (\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{u}}. \tag{4.5}$$

The cross-correlation function $r_{z_a z_b}(k)$ can be estimated from a finite number of samples as

$$\widehat{r}_{z_a z_b}(k) = \frac{1}{N}\sum_{n=0}^{N-|k|-1} \widehat{z}_a(n+k)\widehat{z}_b(n) \tag{4.6}$$

where N is the number of samples obtained from each channel, and $\widehat{z}_a(n+k)$ and $\widehat{z}_b(n)$ are obtained by delaying $y_a(n+k)$ and $y_b(n)$ for c_a/L and c_b/L fractions, respectively. Next, the elements of the matrix $\boldsymbol{R}_z$ are estimated as

$$[\widehat{\boldsymbol{R}}_z]_{a,b} = \widehat{r}_{z_a z_b}(0) = \frac{1}{N}\sum_{n=0}^{N-1} \widehat{z}_a(n)\widehat{z}_b^*(n). \tag{4.7}$$

The fractional delays c_a/L and c_b/L can be implemented by fractional delay (FD) filters. In [47], authors consider using ideal FD filters which have infinite impulse responses. Then, for the purpose of implementation, these filters are truncated using a rectangular window whose width is twice the signal length N. Consequently, the length of the filters can be quite large as N increases. Here, we consider using causal finite impulse response (FIR) filters which have two practical advantages [52]: first, the length of the filters are fixed, and second, they enjoy causality. As for the analysis, we will use a general formulation for the FIR FD filters, and for numerical examples, we will use the Lagrange interpolator [53].

FIR FD filters perform the best when the total delay is approximately equal to half of the order of the filter [54]. The fractional delays c_a/L and c_b/L are positive numbers less than one, and the performance of the FIR FD filters is very poor with such delays. To remedy this problem, a suitable integer delay can be added to the fractional part. Note that $\widehat{r}_{z_a z_b}(k)$ is the inverse DTFT of $(1/N)\widehat{Z}_a\left(e^{j2\pi fL/W}\right)\widehat{Z}_b^*\left(e^{j2\pi fL/W}\right)$, where $\widehat{Z}_a\left(e^{j2\pi fL/W}\right)$ and $\widehat{Z}_b\left(e^{j2\pi fL/W}\right)$ are the DTFT of $\widehat{z}_a(n)$ and $\widehat{z}_b(n)$, respectively [55]. Then, considering that

$$\begin{aligned}&\widehat{Z}_a\left(e^{j2\pi f\frac{L}{W}}\right)\widehat{Z}_b^*\left(e^{j2\pi f\frac{L}{W}}\right)= \\ &\left[\widehat{Z}_a\left(e^{j2\pi f\frac{L}{W}}\right)e^{-jD2\pi f\frac{L}{W}}\right]\left[\widehat{Z}_b\left(e^{j2\pi f\frac{L}{W}}\right)e^{-jD2\pi f\frac{L}{W}}\right]^*\end{aligned} \tag{4.8}$$

we can rewrite (4.7) as

$$[\widehat{\boldsymbol{R}}_z]_{a,b}=\frac{1}{N}\sum_{n=0}^{N-1}\widehat{z}_a(n-D)\widehat{z}_b^*(n-D) \tag{4.9}$$

where D is a suitable integer number close to half of the order of the FD filter.

Let $h_a(n)$ be the impulse response of a causal filter that delays a signal for $c_a/L+D$. Furthermore, let us assume that the length of $h_a(n)$ is large enough, so that its deviation from an ideal FD filter can be ignored. Therefore, $z_a(n-D)$ can be written as

$$z_a(n-D)=\sum_{r=0}^{N_h-1}h_a(r)y_a(n-r) \tag{4.10}$$

where N_h is the length of the filter's impulse response. For a limited number of samples, we have

$$\begin{aligned}\widehat{z}_a(n-D) &= \sum_{r=0}^{N_h-1} h_a(r)y_a(n-r)W_D(n-r) \\ &= \sum_{r=n-N_h+1}^{n} h_a(n-r)y_a(r)W_D(r) \end{aligned} \tag{4.11}$$

where $W_D(n)$ is a window of length N which equals 1 for $0 \leq n \leq N-1$ and is equal to zero elsewhere. Using the elements of $\widehat{\boldsymbol{R}}_z$, the vector $\widehat{\breve{\boldsymbol{u}}}$ is formed as an estimation for $\breve{\boldsymbol{u}}$. Next, $\widehat{\boldsymbol{v}}$ (the estimation for $\boldsymbol{v}$) is formed by replacing $\breve{\boldsymbol{u}}$ with $\widehat{\breve{\boldsymbol{u}}}$ in (4.5) as

$$\widehat{\boldsymbol{v}} = (\breve{\boldsymbol{\Psi}}^T \breve{\boldsymbol{\Psi}})^{-1} \breve{\boldsymbol{\Psi}}^T \widehat{\breve{\boldsymbol{u}}}. \tag{4.12}$$

Finally, let us define $\widehat{\boldsymbol{p}} \in \mathbb{R}^{L\times 1}$ as

$$\widehat{\boldsymbol{p}} \triangleq \frac{L}{W}\widehat{\boldsymbol{v}}. \tag{4.13}$$

The elements of $\widehat{\boldsymbol{p}}$ give an estimation for the average power within each spectral segment.

4.2 Preliminary Formulations for Bias and Variance Analysis

Consider a Gaussian WSS signal $x(t)$ bandlimited to $W/2$ Hz, and let $x(m)$ be the samples of the signal obtained at the Nyquist rate ($m \in \mathbb{Z}$). Let also $r_x(k) = E\{x(m+k)x^*(m)\}$ and $P_x(e^{j2\pi f/W}) = \text{DTFT}\{r_x(k)\}$ be the autocorrelation function and the PSD of $x(m)$, respectively. Furthermore, consider a zero-mean Gaussian random process $e(t)$ bandlimited to $W/2$ Hz with a flat PSD $P_e(f) = \sigma^2/W$. The autocorrelation function of $e(t)$ is $r_e(\tau) = \sigma^2 \text{sinc}(W\tau)$. Let $e(m)$ be the samples of $e(t)$ obtained at the Nyquist rate. Then, the autocorrelation function of $e(m)$ is given by

$$r_e(k) = \sigma^2 \text{sinc}(Wk/W) = \sigma^2\delta(k) \tag{4.14}$$

where $\delta(k)$ is the Kronecker delta function. Therefore, the PSD of $e(m)$ is given by $P_e(e^{j2\pi f/W}) = \sigma^2$. Consider a filter $h_x(m)$ such that $\sigma^2|H_x(e^{j2\pi f/W})|^2$

is equal to $P_x(e^{j2\pi f/W})$, where $H_x(e^{j2\pi f/W})$ is the DTFT of $h_x(m)$. Therefore, we have

$$P_x(e^{j2\pi f/W}) = |H_x(e^{j2\pi f/W})|^2 P_e(e^{j2\pi f/W}). \tag{4.15}$$

As a result, $x(m)$ can be considered as $e(m)$ filtered by $h_x(m)$ since the output of the filter has the same PSD as $P_x(e^{j2\pi f/W})$. Then, the output of the i-th sampling channel can be written as

$$y_i(n) = x(nL + c_i) = \sum_{m\in\mathbb{Z}} h_x(m)e(nL + c_i - m). \tag{4.16}$$

Let a and b denote two channel indices. Then, the cross-correlation function $r_{y_a y_b}(k) = E\{y_a(n+k)y_b^*(n)\}$ is given by

$$\begin{aligned} r_{y_a y_b}(k) &= \sum_{m\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} h_x(m)h_x^*(l) \\ &\quad \times E\left\{e((n+k)L + c_a - m)e^*(nL + c_b - l)\right\} \\ &= \sum_{m\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} h_x(m)h_x^*(l)r_e(kL + l - m + c_a - c_b) \\ &= \sigma^2 \sum_{m\in\mathbb{Z}} h_x(m)h_x^*(-kL + m + c_b - c_a). \end{aligned} \tag{4.17}$$

Furthermore, using (4.10), $[\boldsymbol{R_z}]_{a,b}$ can be written as

$$\begin{aligned} [\boldsymbol{R_z}]_{a,b} &= E\{z_a(n-D)z_b^*(n-D)\} \\ &= \sum_{r=0}^{N_h-1}\sum_{p=0}^{N_h-1} h_a(r)h_b(p)r_{y_a y_b}(p-r). \end{aligned} \tag{4.18}$$

4.3 Bias Analysis of Averaged Correlogram for Undersampled Data

The bias of the averaged correlogram for undersampled data estimator is given by

$$E\{\widehat{\boldsymbol{p}}\} - \boldsymbol{p} = \frac{L}{W}\left(E\{\widehat{\boldsymbol{v}}\} - \boldsymbol{v}\right) \tag{4.19}$$

where $\boldsymbol{p} = (L/W)\boldsymbol{v}$. The expected value of $\widehat{\boldsymbol{v}}$ is obtained using (4.12) as

$$E\{\widehat{\boldsymbol{v}}\} = (\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}\breve{\boldsymbol{\Psi}}^T E\{\widehat{\breve{\boldsymbol{u}}}\}. \tag{4.20}$$

Computing $E\{\widehat{\breve{\boldsymbol{u}}}\}$ requires finding the expected value of the real and imaginary parts of $\widehat{\boldsymbol{R}}_{\boldsymbol{z}}$. The expectation operation can be performed before taking the real or imaginary parts of $\widehat{\boldsymbol{R}}_{\boldsymbol{z}}$, as these operators are linear. Moreover, (4.9) is used to form $\widehat{\boldsymbol{R}}_{\boldsymbol{z}}$. Taking expectation from both sides of (4.9) along with using (4.11) results in

$$\begin{aligned} E\{[\widehat{\boldsymbol{R}}_{\boldsymbol{z}}]_{a,b}\} &= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{r=0}^{N_h-1}\sum_{p=0}^{N_h-1} h_a(r)h_b(p) \\ &\qquad \times W_D(n-r)W_D(n-p)E\{y_a(n-r)y_b^*(n-p)\} \\ &= \sum_{r=0}^{N_h-1}\sum_{p=0}^{N_h-1} h_a(r)h_b(p)r_{y_a y_b}(p-r) \\ &\qquad \times\frac{1}{N}\sum_{n=0}^{N-1} W_D(n-r)W_D(n-p) \end{aligned} \tag{4.21}$$

With the assumption that the number of samples N is larger than the length of the fractional delay filters N_h, the last summation of (4.21) can be simplified to

$$\sum_{n=0}^{N-1} W_D(n-r)W_D(n-p) = N - \max(r,p). \tag{4.22}$$

Therefore, (4.21) can be rewritten as

$$E\{[\widehat{\boldsymbol{R}}_{\boldsymbol{z}}]_{a,b}\} = [\boldsymbol{R}_{\boldsymbol{z}}]_{a,b} - \frac{1}{N}\sum_{r=0}^{N_h-1}\sum_{p=0}^{N_h-1} h_a(r)h_b(p)r_{y_a y_b}(p-r)\max(r,p) \tag{4.23}$$

where $[\boldsymbol{R}_{\boldsymbol{z}}]_{a,b}$ is given by (4.18).

Theorem 4.1. *The averaged correlogram for undersampled data estimator $\widehat{\boldsymbol{p}}$ is asymptotically unbiased.*

Proof. It can be seen from (4.23) that as N tends to infinity, $E\{[\widehat{\boldsymbol{R}}_{\boldsymbol{z}}]_{a,b}\}$ tends to $[\boldsymbol{R}_{\boldsymbol{z}}]_{a,b}$. Therefore, $\widehat{\boldsymbol{R}}_z$ is an asymptotically unbiased estimator of $\boldsymbol{R}_z$. Since $\widehat{\breve{\boldsymbol{u}}}$ consists of the elements of $\widehat{\boldsymbol{R}}_z$ and the operation of taking the real and imaginary parts are linear, it follows that $\widehat{\breve{\boldsymbol{u}}}$ is also an asymptotically unbiased estimator of $\breve{\boldsymbol{u}}$. Furthermore, letting the number of samples tend to infinity in (4.20) and using (4.5), we find that

$$\begin{aligned} \lim_{N\to\infty} E\{\widehat{\boldsymbol{v}}\} &= (\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}\breve{\boldsymbol{\Psi}}^T \lim_{N\to\infty} E\{\widehat{\breve{\boldsymbol{u}}}\} \\ &= (\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{u}} = \boldsymbol{v}. \end{aligned} \tag{4.24}$$

In other words, $\widehat{\boldsymbol{v}}$ is also an asymptotically unbiased estimator of $\boldsymbol{v}$. Finally, it can be concluded from (4.19) that the averaged correlogram for undersampled data estimator $\widehat{\boldsymbol{p}}$ is asymptotically unbiased. □

Theorem 4.2. *For the case that the input signal $x(t)$ is equal to the white Gaussian random process $e(t)$, the bias of the averaged correlogram for undersampled data estimator is given by*

$$E\{\widehat{\boldsymbol{p}}\} - \boldsymbol{p} = (H_1 - 1)\frac{\sigma^2}{W}\mathbf{1}_L. \tag{4.25}$$

where $\mathbf{1}_L$ is the column vector of length L with all its elements equal to 1, and H_1 is given by

$$H_1 = \frac{1}{N}\sum_{r=0}^{N_h-1}(N-r)h_1^2(r). \tag{4.26}$$

Proof. In the case that $x(t)$ is equal to $e(t)$, we have $h_x(m) = \delta(m)$. Then, using (4.17), the cross-correlation function $r_{y_a y_b}(k)$ is given by

$$r_{y_a y_b}(k) = \sigma^2\delta(k)\delta(a-b). \tag{4.27}$$

Applying (4.27) to (4.23), we find that

$$E\{[\widehat{\boldsymbol{R}}_{\boldsymbol{z}}]_{a,b}\} = [\boldsymbol{R}_{\boldsymbol{z}}]_{a,b} - \frac{1}{N}\sum_{r=0}^{N_h-1}h_a^2(r)r\sigma^2\delta(a-b). \tag{4.28}$$

Next, $[\boldsymbol{R}_{\boldsymbol{z}}]_{a,b}$ is obtained using (4.18) and (4.27) as

$$[\boldsymbol{R}_{\boldsymbol{z}}]_{a,b} = \sum_{r=0}^{N_h-1}h_a^2(r)\sigma^2\delta(a-b). \tag{4.29}$$

Replacing (4.29) into (4.28) results in

$$E\{[\widehat{\boldsymbol{R}}_z]_{a,b}\} = 0 \tag{4.30}$$

for $a \neq b$, and

$$E\{[\widehat{\boldsymbol{R}}_z]_{a,b}\} = \sigma^2\frac{1}{N}\sum_{r=0}^{N_h-1}(N-r)h_a^2(r) = H_a\sigma^2 \tag{4.31}$$

for $a = b$, where

$$H_a \triangleq \frac{1}{N}\sum_{r=0}^{N_h-1}(N-r)h_a^2(r). \tag{4.32}$$

Recalling that the first diagonal element of $\widehat{\boldsymbol{R}}_z$ is used in $\widehat{\breve{\boldsymbol{u}}}$ and taking the real and imaginary parts of (4.30) and (4.31), $E\{\widehat{\breve{\boldsymbol{u}}}\}$ can be obtained as

$$E\{\widehat{\breve{\boldsymbol{u}}}\} = H_1 \sigma^2 \boldsymbol{e}_1 \tag{4.33}$$

where $\boldsymbol{e}_1$ is a column vector of length $q(q-1)+2$ with all its elements equal to zero except for the first element which is 1. The expected value of $\widehat{\boldsymbol{v}}$ can be found using (4.20) and (4.33) as

$$E\{\widehat{\boldsymbol{v}}\} = H_1 \sigma^2 (\breve{\boldsymbol{\Psi}}^T \breve{\boldsymbol{\Psi}})^{-1} \breve{\boldsymbol{\Psi}}^T \boldsymbol{e}_1. \tag{4.34}$$

Next, Consider the fact that $x(t)$ has equal power in all spectral segments (the elements of $\boldsymbol{v}$ are all the same). Since $\widehat{\boldsymbol{v}}$ is asymptotically unbiased, it follows that the elements of $\lim_{N\to\infty} E\{\widehat{\boldsymbol{v}}\}$ are also equal.

Replacing the true values in (4.1) with the estimated values for $a = b = 1$, taking expectation from both sides, and letting the number of samples tend to infinity, we obtain that

$$\begin{aligned} \lim_{N\to\infty} E\{[\widehat{\boldsymbol{R}}_z]_{1,1}\} &= \sum_{l=1}^{L} \lim_{N\to\infty} E\{\widehat{v}_l\} \\ &= \mathbf{1}_L^T \lim_{N\to\infty} E\{\widehat{\boldsymbol{v}}\} \end{aligned} \tag{4.35}$$

where $\widehat{v}_l$ $(1 \le l \le L)$ are the elements of $\widehat{\boldsymbol{v}}$. Considering normalized FD filters $(\sum_{r=0}^{N_h-1} h_a^2(r) = 1)$ and referring to (4.32), we also find that

$$\lim_{N\to\infty} H_a = 1. \tag{4.36}$$

Therefore, using (4.31), we can find that

$$\lim_{N\to\infty} E\{[\widehat{\boldsymbol{R}}_z]_{1,1}\} = \sigma^2. \tag{4.37}$$

Combining (4.35) with (4.37) results in

$$\lim_{N\to\infty} E\{\widehat{\boldsymbol{v}}\} = \frac{\sigma^2}{L} \mathbf{1}_L. \tag{4.38}$$

Letting the number of samples tend to infinity in (4.34) and using (4.38), we obtain

$$\lim_{N\to\infty} E\{\widehat{\boldsymbol{v}}\} = \sigma^2 (\breve{\boldsymbol{\Psi}}^T \breve{\boldsymbol{\Psi}})^{-1} \breve{\boldsymbol{\Psi}}^T \boldsymbol{e}_1 = \frac{\sigma^2}{L} \mathbf{1}_L. \tag{4.39}$$

It follows from (4.39) that all the elements of the first column of $(\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}\breve{\boldsymbol{\Psi}}^T$ are equal to $1/L$. Therefore, (4.34) can be simplified as

$$E\{\widehat{\boldsymbol{v}}\} = H_1\frac{\sigma^2}{L}\mathbf{1}_L. \tag{4.40}$$

Finally, using (4.13), we have

$$E\{\widehat{\boldsymbol{p}}\} = H_1\frac{\sigma^2}{W}\mathbf{1}_L. \tag{4.41}$$

□

4.4 Variance Analysis of Averaged Correlogram for Undersampled Data

The covariance matrix of the averaged correlogram for undersampled data is given by

$$\begin{aligned}\mathcal{C}_{\widehat{\boldsymbol{p}}} &= E\left\{\left(\widehat{\boldsymbol{p}} - E\{\widehat{\boldsymbol{p}}\}\right)\left(\widehat{\boldsymbol{p}} - E\{\widehat{\boldsymbol{p}}\}\right)^T\right\} \\ &= E\{\widehat{\boldsymbol{p}}\widehat{\boldsymbol{p}}^T\} - E\{\widehat{\boldsymbol{p}}\}E\{\widehat{\boldsymbol{p}}\}^T. \end{aligned}\tag{4.42}$$

The diagonal elements of $\mathcal{C}_{\widehat{\boldsymbol{p}}}$ are the estimation variance of each spectral segment. The off-diagonal elements of $\mathcal{C}_{\widehat{\boldsymbol{p}}}$ represent the correlation between pairs of the estimations made for different spectral segments.

It follows from (4.12) and (4.13) that

$$E\{\widehat{\boldsymbol{p}}\widehat{\boldsymbol{p}}^T\} = \left(\frac{L}{W}\right)^2(\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}\breve{\boldsymbol{\Psi}}^T\boldsymbol{U}\breve{\boldsymbol{\Psi}}(\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1} \tag{4.43}$$

where $\boldsymbol{U} \triangleq E\{\widehat{\breve{\boldsymbol{u}}}\widehat{\breve{\boldsymbol{u}}}^T\} \in \mathbb{R}^{2Q\times 2Q}$. Computation of the elements of $\boldsymbol{U}$ involves taking expectation of the multiplication of the real or imaginary parts of the elements of $\widehat{\boldsymbol{R}}_z$. We will use the following lemma [56] for interchanging the expectation and the operation of taking real or imaginary parts.

Lemma 4.1. *Let x and y be two arbitrary complex numbers. The following equations hold*

$$Re(x)Re(y) = \frac{1}{2}\left(Re(xy) + Re(xy^*)\right) \tag{4.44}$$

$$Im(x)Im(y) = -\frac{1}{2}\left(Re(xy) - Re(xy^*)\right) \tag{4.45}$$

$$Re(x)Im(y) = \frac{1}{2}\left(Im(xy) - Im(xy^*)\right). \tag{4.46}$$

The elements of $\boldsymbol{U}$ can be obtained using Lemma 4.1, $E\{[\widehat{\boldsymbol{R}}_z]_{a,b}[\widehat{\boldsymbol{R}}_z]_{c,d}\}$, and $E\{[\widehat{\boldsymbol{R}}_z]_{a,b}[\widehat{\boldsymbol{R}}_z]^*_{c,d}\}$, where $[\widehat{\boldsymbol{R}}_z]_{a,b}$ and $[\widehat{\boldsymbol{R}}_z]_{c,d}$ are the elements of $\widehat{\boldsymbol{R}}_z$ used for forming $\widehat{\breve{\boldsymbol{u}}}$. Let the outputs of the sampling channels be given by (4.16). Using (4.9) and (4.11), we obtain

$$\begin{aligned}
E\{[\widehat{\boldsymbol{R}}_z]_{a,b}[\widehat{\boldsymbol{R}}_z]_{c,d}\} &= \frac{1}{N^2} \\
&\times \sum_{n=0}^{N-1} \sum_{\substack{r= \\ (n-N_h+1)}}^{n} \sum_{\substack{p= \\ (n-N_h+1)}}^{n} \sum_{u=0}^{N-1} \sum_{\substack{s= \\ (u-N_h+1)}}^{u} \sum_{\substack{m= \\ (u-N_h+1)}}^{u} h_a(n-r)h_b(n-p) \\
&\quad \times h_c(u-s)h_d(u-m)W_D(r)W_D(p)W_D(s)W_D(m) \\
&\quad \times E\{y_a(r)y_b^*(p)y_c(s)y_d^*(m)\} \\
&= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{\substack{r=\max \\ (0,n-N_h+1)}}^{n} \sum_{\substack{p=\max \\ (0,n-N_h+1)}}^{n} \sum_{u=0}^{N-1} \sum_{\substack{s=\max \\ (0,u-N_h+1)}}^{u} \sum_{\substack{m=\max \\ (0,u-N_h+1)}}^{u} h_a(n-r) \\
&\quad \times h_b(n-p)h_c(u-s)h_d(u-m) \\
&\quad \times \left(r_{y_a y_b}(r-p)r_{y_c y_d}(s-m) + r_{y_a y_d}(r-m)r_{y_c y_b}(s-p)\right). \quad (4.47)
\end{aligned}$$

The last line in (4.47) is obtained using the forth-order moment of Gaussian random processes.

In a similar way, $E\{[\widehat{\boldsymbol{R}}_z]_{a,b}[\widehat{\boldsymbol{R}}_z]^*_{c,d}\}$ can be obtained as

$$\begin{aligned}
E\{[\widehat{\boldsymbol{R}}_z]_{a,b}[\widehat{\boldsymbol{R}}_z]^*_{c,d}\} &= \frac{1}{N^2} \\
&\times \sum_{n=0}^{N-1} \sum_{\substack{r=\max \\ (0,n-N_h+1)}}^{n} \sum_{\substack{p=\max \\ (0,n-N_h+1)}}^{n} \sum_{u=0}^{N-1} \sum_{\substack{s=\max \\ (0,u-N_h+1)}}^{u} \sum_{\substack{m=\max \\ (0,u-N_h+1)}}^{u} h_a(n-r) \\
&\quad \times h_b(n-p)h_c(u-s)h_d(u-m) \\
&\quad \times \left(r_{y_a y_b}(r-p)r_{y_d y_c}(m-s) + r_{y_a y_c}(r-s)r_{y_d y_b}(m-p)\right). \quad (4.48)
\end{aligned}$$

Theorem 4.3. *For the case that the input signal $x(t)$ is equal to the white Gaussian random process $e(t)$, $\boldsymbol{U}$ is a diagonal matrix with*

$$\begin{aligned}
[\boldsymbol{U}]_{1,1} &= \frac{\sigma^4}{N^2}\Big(N^2H_1^2 + (N-2N_h+2)G_1 + \Sigma_1\Big) \\
[\boldsymbol{U}]_{Q+1,Q+1} &= 0 \\
[\boldsymbol{U}]_{k,k} &= \frac{\sigma^4}{2N^2}\left((N-2N_h+2)G_k + \Sigma_k\right) \quad (4.49)
\end{aligned}$$

where G_1, Σ_1, G_k, and Σ_k ($2 \le k \le 2Q$ and $k \ne Q+1$) are independent of the signal length and depend on the FD filters.

Proof. In the case that $x(t)$ is equal to $e(t)$, we have $h_x(m) = \delta(m)$. Then, the cross-correlation functions in (4.47) are simplified as

$$\begin{aligned} E_1 &\triangleq r_{y_a y_b}(r-p) r_{y_c y_d}(s-m) + r_{y_a y_d}(r-m) r_{y_c y_b}(s-p) \\ &= \sigma^4 \big(\delta(r-p)\delta(a-b)\delta(s-m)\delta(c-d) \\ &\qquad + \delta(r-m)\delta(a-d)\delta(s-p)\delta(c-b) \big). \end{aligned} \tag{4.50}$$

Similarly, the cross-correlation functions in (4.48) are simplified as

$$\begin{aligned} E_2 &\triangleq r_{y_a y_b}(r-p) r_{y_d y_c}(m-s) + r_{y_a y_c}(r-s) r_{y_d y_b}(m-p) \\ &= \sigma^4 \big(\delta(r-p)\delta(a-b)\delta(m-s)\delta(d-c) \\ &\qquad + \delta(r-s)\delta(a-c)\delta(m-p)\delta(d-b) \big). \end{aligned} \tag{4.51}$$

Recalling that only the first diagonal element of $\widehat{\boldsymbol{R}}_z$ is present in $\widehat{\breve{\boldsymbol{u}}}$, E_1 can be found to be equal to

$$E_1 = \sigma^4 \big(\delta(r-p)\delta(s-m) + \delta(r-m)\delta(s-p) \big) \tag{4.52}$$

for $a = b = c = d = 1$, and it equals to zero otherwise. Similarly, E_2 can be found to be equal to

$$E_2 = \sigma^4 \delta(r-s)\delta(m-p) \tag{4.53}$$

for $a = c$ and $b = d$, and it equals zero otherwise (excluding the case when $a = b = c = d = 1$ since $[\widehat{\boldsymbol{R}}_z]_{1,1}$ is real-valued, and therefore, we do not need to compute (4.48)). Noting that $E\{[\widehat{\boldsymbol{R}}_z]_{a,b}[\widehat{\boldsymbol{R}}_z]_{c,d}\}$ and $E\{[\widehat{\boldsymbol{R}}_z]_{a,b}[\widehat{\boldsymbol{R}}_z]^*_{c,d}\}$ are real-valued and using (4.46), (4.52), and (4.53), we can find that all the off-diagonal elements of $\boldsymbol{U}$ are equal to zero.

Let us start computing the diagonal elements of $\boldsymbol{U}$ by setting $a = b = c = d = 1$. It follows from (4.47) and (4.52) that

$$\begin{aligned} E\{[\widehat{\boldsymbol{R}}_z]_{1,1}[\widehat{\boldsymbol{R}}_z]_{1,1}\} = \frac{\sigma^4}{N^2} \Big(\sum_{n=0}^{N-1} \sum_{\substack{r=\max \\ (0,n-N_h+1)}}^{n} \sum_{u=0}^{N-1} \sum_{\substack{s=\max \\ (0,u-N_h+1)}}^{u} h_1^2(n-r) \\ \times h_1^2(u-s) + \sum_{n=0}^{N-1} S_1(n) \Big) \end{aligned} \tag{4.54}$$

where $S_1(n)$ is defined as

$$S_1(n) \triangleq \sum_{\substack{r=\max \\ (0,n-N_h+1)}}^{n} \sum_{\substack{p=\max \\ (0,n-N_h+1)}}^{n} \sum_{u=0}^{N-1} \sum_{\substack{s=\max \\ (0,u-N_h+1)}}^{u} \sum_{\substack{m=\max \\ (0,u-N_h+1)}}^{u} \delta(r-m)\delta(s-p) \times h_1(n-r)h_1(n-p)h_1(u-s)h_1(u-m). \quad (4.55)$$

For $N_h - 1 \leq n \leq N - N_h$, $S_1(n)$ is given by

$$S_1(n) = \sum_{\substack{u= \\ n-N_h+1}}^{n+N_h-1} \left[\sum_{\substack{r= \\ n-N_h+1}}^{n} \sum_{\substack{m=\max \\ (0,u-N_h+1)}}^{u} \delta(r-m)h_1(n-r)h_1(u-m) \right] \times \left[\sum_{\substack{p= \\ n-N_h+1}}^{n} \sum_{\substack{s=\max \\ (0,u-N_h+1)}}^{u} \delta(s-p)h_1(n-p)h_1(u-s) \right]. \quad (4.56)$$

Note that the summations in the brackets are equivalent to each other, which leads to the following simplification

$$\begin{aligned} S_1(n) &= \sum_{\substack{u= \\ n-N_h+1}}^{n+N_h-1} \left[\sum_{\substack{r= \\ n-N_h+1}}^{n} \sum_{\substack{m=\max \\ (0,u-N_h+1)}}^{u} \delta(r-m)h_1(n-r)h_1(u-m) \right]^2 \\ &= \sum_{\substack{u= \\ n-N_h+1}}^{n+N_h-1} \left[\sum_{r=\max(n,u)-N_h+1}^{\min(n,u)} h_1(n-r)h_1(u-r) \right]^2. \end{aligned} \quad (4.57)$$

Next, a change of variable ($g = u - n + N_h - 1$) is used, which results in

$$S_1(n) = \sum_{g=0}^{2N_h-2} \left[\sum_{r=\max(0,g-N_h+1)+n-N_h+1}^{\min(0,g-N_h+1)+n} h_1(n-r)h_1(n-r+g-N_h+1) \right]^2. \quad (4.58)$$

With another change of variable ($p = n - r + g - N_h + 1$), we obtain the following

$$S_1(n) = \sum_{g=0}^{2N_h-2} \left[\sum_{p=\max(0,g-N_h+1)}^{\min(g,N_h-1)} h_1(p-g+N_h-1)h_1(p) \right]^2 \quad (4.59)$$

which is equal to

$$G_1 \triangleq S_1(n) = \sum_{g=0}^{2N_h-2} \left[h_1(i) * h_1(N_h-1-i)|_g \right]^2 \quad (4.60)$$

where $*$ denotes the convolution operation. Note that G_1 is not a function of n. In a similar way, $S_1(n)$ for $0 \leq n < N_h - 1$ is given by

$$S_1(n) = \sum_{g=0}^{n+N_h-1} \left[(h_1(i)W_n(i)) * h_1(N_h - 1 - i)|_g\right]^2 \tag{4.61}$$

where $W_n(i)$ is equal to 1 for $0 \leq i \leq n$ and zero elsewhere. For $N - N_h < n \leq N - 1$, $S_1(n)$ is given by

$$S_1(n) = \sum_{g=0}^{N-n+N_h-2} \left[h_1(i) * h_1(N_h - 1 - i)|_g\right]^2. \tag{4.62}$$

Next, (4.54) can be rewritten as

$$\begin{aligned} E\{[\widehat{\boldsymbol{R}}_z]_{1,1}[\widehat{\boldsymbol{R}}_z]_{1,1}\} &= \frac{\sigma^4}{N^2} \times \\ \Big(\sum_{n=0}^{N-1} \sum_{\substack{r=\max \\ (0,n-N_h+1)}}^{n} h_1^2(n-r) &\sum_{u=0}^{N-1} \sum_{\substack{s=\max \\ (0,u-N_h+1)}}^{u} h_1^2(u-s) \\ +(N-2N_h+2)G_1 + \sum_{n=0}^{N_h-2} S_1(n) &+ \sum_{n=N-N_h+1}^{N-1} S_1(n)\Big). \end{aligned} \tag{4.63}$$

Using (4.32), we have

$$\sum_{n=0}^{N-1} \sum_{\substack{r=\max \\ (0,n-N_h+1)}}^{n} h_1^2(n-r) = \sum_{r=0}^{N_h-1} (N-r)h_1^2(r) = NH_1. \tag{4.64}$$

Therefore, (4.63) can be simplified as

$$E\{[\widehat{\boldsymbol{R}}_z]_{1,1}[\widehat{\boldsymbol{R}}_z]_{1,1}\} = \frac{\sigma^4}{N^2}\Big(N^2H_1^2 + (N-2N_h+2)G_1 + \Sigma_1\Big) \tag{4.65}$$

where $\Sigma_1 \triangleq \sum_{n=0}^{N_h-2} S_1(n) + \sum_{n=N-N_h+1}^{N-1} S_1(n)$. Note that $[\widehat{\boldsymbol{R}}_z]_{1,1}$ is real-valued. Therefore, $[\boldsymbol{U}]_{1,1}$ is equal to $E\{[\widehat{\boldsymbol{R}}_z]_{1,1}[\widehat{\boldsymbol{R}}_z]_{1,1}\}$ as given in (4.65) and $[\boldsymbol{U}]_{Q+1,Q+1}$ equals zero since the imaginary part of $[\widehat{\boldsymbol{R}}_z]_{1,1}$ is zero.

For the rest of the diagonal elements of $\boldsymbol{U}$, $E\{[\widehat{\boldsymbol{R}}_z]_{a,b}[\widehat{\boldsymbol{R}}_z]_{a,b}\}$ equals zero, as E_1 is zero. Therefore, $[\boldsymbol{U}]_{k,k}$ ($2 \leq k \leq 2Q$ and $k \neq Q+1$) can be obtained using (4.44) and (4.45) as

$$[\boldsymbol{U}]_{k,k} = \frac{1}{2} Re\left(E\{[\widehat{\boldsymbol{R}}_z]_{a,b}[\widehat{\boldsymbol{R}}_z]^*_{a,b}\}\right). \tag{4.66}$$

From (4.48) and (4.53) we have

$$[\boldsymbol{U}]_{k,k} = \frac{\sigma^4}{2N^2} \sum_n S_k(n) \tag{4.67}$$

where $S_k(n)$ is defined as

$$S_k(n) \triangleq \sum_{\substack{r=\max \\ (0,n-N_h+1)}}^{n} \sum_{\substack{p=\max \\ (0,n-N_h+1)}}^{n} \sum_{u=0}^{N-1} \sum_{\substack{s=\max \\ (0,u-N_h+1)}}^{u} \sum_{\substack{m=\max \\ (0,u-N_h+1)}}^{u} \delta(r-s)\delta(m-p) \times h_a(n-r)h_b(n-p)h_a(u-s)h_b(u-m). \tag{4.68}$$

It can be shown that for $N_h - 1 \leq n \leq N - N_h$, $S_k(n)$ is given by

$$G_k \triangleq S_k(n) = \sum_{g=0}^{2N_h-2} \left(h_a(i) * h_a(N_h-1-i)\right)|_g \times \left(h_b(i) * h_b(N_h-1-i)\right)|_g. \tag{4.69}$$

For $0 \leq n < N_h - 1$, $S_k(n)$ is given by

$$S_k(n) = \sum_{g=0}^{n+N_h-1} \left((h_a(i)W_n(i)) * h_a(N_h-1-i)\right)|_g \times \left((h_b(i)W_n(i)) * h_b(N_h-1-i)\right)|_g. \tag{4.70}$$

For $N - N_h < n \leq N - 1$, $S_k(n)$ is given by

$$S_k(n) = \sum_{g=0}^{N-n+N_h-2} \left(h_a(i) * h_a(N_h-1-i)\right)|_g \times \left(h_b(i) * h_b(N_h-1-i)\right)|_g. \tag{4.71}$$

Thus, (4.67) can be rewritten as

$$[\boldsymbol{U}]_{k,k} = \frac{\sigma^4}{2N^2} \left((N-2N_h+2)G_k + \Sigma_k\right) \tag{4.72}$$

where $\Sigma_k \triangleq \sum_{n=0}^{N_h-2} S_k(n) + \sum_{n=N-N_h+1}^{N-1} S_k(n)$. □

The equations for computing the covariance matrix $\mathcal{C}_{\widehat{\boldsymbol{p}}}$ as given by (4.42) to (4.49) are in the matrix form. Next, we simplify these formulas to show the dependence of the estimation variance on different parameters more clearly.

Theorem 4.4. *For the white Gaussian process, the diagonal elements of* $\mathcal{C}_{\widehat{\boldsymbol{p}}}$ *can be approximated by*

$$[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{l,l} \approx \frac{\sigma^4}{2W^2N_x^2}\left(\frac{L^3}{Q}+L\right)((N_x-2N_hL+2L)G_1+L\Sigma_1) \tag{4.73}$$

where N_x *is the number of Nyquist samples.*

Proof. Referring to (4.43), computation of the l-th diagonal element of the covariance matrix requires the knowledge of the elements of the l-th row of $\boldsymbol{X} \triangleq (\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}\breve{\boldsymbol{\Psi}}^T$. The diagonal elements of $\boldsymbol{U}$ for $2 \leq k \leq 2Q$ and $k \neq Q+1$ as given by (4.49) differ from each other in G_k and Σ_k. However, the values of G_k and Σ_k for different values of k almost remain the same as they are related to the energy of the FD filters which are normalized to one. Let us approximate G_k and Σ_k by G_1 and Σ_1. Then, $[\boldsymbol{U}]_{k,k}$ can be approximated by

$$\eta \triangleq \frac{\sigma^4}{2N^2}\left((N-2N_h+2)G_1+\Sigma_1\right). \tag{4.74}$$

The approximation in (4.74) relaxes the problem of computing the l-th diagonal element of the covariance matrix to just finding the Euclidean norm of the l-th row of $\boldsymbol{X}$. The squared norm of the l-th row of $\boldsymbol{X}$ can be obtained as

$$\begin{aligned}\phi_l &\triangleq \left[\boldsymbol{X}\boldsymbol{X}^T\right]_{l,l}\\ &= \left[(\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}\right]_{l,l}\\ &= \left[\left(Re(\boldsymbol{\Psi}^H\boldsymbol{\Psi})\right)^{-1}\right]_{l,l}.\end{aligned} \tag{4.75}$$

Referring to (4.4), the diagonal elements of $Re(\boldsymbol{\Psi}^H\boldsymbol{\Psi})$ are all equal to Q, and the off-diagonal elements are given as

$$\left[Re(\boldsymbol{\Psi}^H\boldsymbol{\Psi})\right]_{i,j} = 1+\sum_{k=2}^{Q}\cos((i-j)\nu_k) \tag{4.76}$$

where $1 \leq i,j \leq L$ and $i \neq j$. Noting that the frequencies ν_k are randomly obtained based on the sampling pattern, the value of the off-diagonal elements of $Re(\boldsymbol{\Psi}^H\boldsymbol{\Psi})$ are negligible compared to the value of the diagonal elements. Therefore, $Re(\boldsymbol{\Psi}^H\boldsymbol{\Psi})$ can be approximated by a diagonal matrix with elements equal to Q, which results in

$$\phi_l \approx \frac{1}{Q}. \tag{4.77}$$

It is shown in the proof of Theorem 4.2 that all the elements of the first column of $\boldsymbol{X}$ are equal to $1/L$. Furthermore, all the elements of the $(Q+1)$-th column of $\boldsymbol{X}$ are equal to zero, as all the elements of the $(Q+1)$-th row of $\breve{\boldsymbol{\Psi}}$ are zero. Then, using (4.41), (4.42), (4.43), and (4.74), $[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{l,l}$ can be approximated as

$$[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{l,l} \approx \left(\frac{L}{W}\right)^2 \left[\eta\phi_l + \frac{1}{L^2}\left([\boldsymbol{U}]_{1,1} - \eta\right)\right] - \left(H_1\frac{\sigma^2}{W}\right)^2. \tag{4.78}$$

Next, using (4.49), (4.74), and (4.77), we can simplify (4.78) to

$$[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{l,l} \approx \frac{\sigma^4}{2W^2N_x^2}\left(\frac{L^3}{Q} + L\right)\left((N_x - 2N_hL + 2L)G_1 + L\Sigma_1\right) \tag{4.79}$$

where $N_x \triangleq NL$ is the number of Nyquist samples. □

Remark 4.1. Considering a large enough N_x, it can be seen from (4.73) that the estimation variance is a cubic function of the number of spectral segments L as $(L^3/Q + L)$. Moreover, the variance is inversely proportional to Q, which means that the variance decreases quadratically with the number of sampling channels q. Furthermore, at a fixed average sampling rate $(q/L)W$ and a given signal length N_x, the variance increases almost linearly with the number of spectral segments. Finally, it can be seen that the estimation variance decreases as the signal length increases at an approximate rate of $1/N_x$.

We next consider the asymptotic behavior of the averaged correlogram for undersampled data for the case of a white Gaussian process. The following theorem studies the covariance matrix of the estimator as the length of the signal tends to infinity.

Theorem 4.5. *In the case of a white Gaussian process, the averaged correlogram for undersampled data is a consistent estimator of the average power in each spectral segment. Furthermore, the estimations made for different spectral segments are asymptotically uncorrelated.*

Proof. Letting the number of samples tend to infinity in (4.42) yields

$$\lim_{N\to\infty} \mathcal{C}_{\widehat{\boldsymbol{p}}} = \lim_{N\to\infty} E\{\widehat{\boldsymbol{p}}\widehat{\boldsymbol{p}}^T\} - \lim_{N\to\infty} E\{\widehat{\boldsymbol{p}}\}E\{\widehat{\boldsymbol{p}}\}^T. \tag{4.80}$$

Since the averaged correlogram for undersampled data estimator is asymptotically unbiased, we have

$$\lim_{N\to\infty} E\{\widehat{\boldsymbol{p}}\} = \boldsymbol{p} = \frac{\sigma^2}{W}\mathbf{1}_L. \tag{4.81}$$

From (4.43), we obtain

$$\lim_{N\to\infty} E\{\widehat{\boldsymbol{p}}\widehat{\boldsymbol{p}}^T\} = \left(\frac{L}{W}\right)^2 (\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}\breve{\boldsymbol{\Psi}}^T \left(\lim_{N\to\infty} \boldsymbol{U}\right) \breve{\boldsymbol{\Psi}}(\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}. \tag{4.82}$$

Recall that all the off-diagonal elements of $\boldsymbol{U}$ are zeros, and the first diagonal element of $\boldsymbol{U}$ is given by (4.49). Letting the number of samples tend to infinity in (4.49), we obtain

$$\lim_{N\to\infty} E\{[\boldsymbol{U}]_{1,1}\} = \sigma^4. \tag{4.83}$$

The $(Q+1)$-th element of $\boldsymbol{U}$ is zero, and if the number of samples tend to infinity in (4.72), $\lim_{N\to\infty}[\boldsymbol{U}]_{k,k} = 0$. Therefore, all the elements of $\lim_{N\to\infty}\boldsymbol{U}$ are equal to zero except for its first diagonal element which is equal to σ^4.

In order to further simplify (4.82), only the elements of the first column of $(\breve{\boldsymbol{\Psi}}^T\breve{\boldsymbol{\Psi}})^{-1}\breve{\boldsymbol{\Psi}}^T$ are required. We have shown in the proof of Theorem 4.2 that these elements are all equal to $1/L$. Therefore, (4.82) can be simplified to

$$\lim_{N\to\infty} E\{\widehat{\boldsymbol{p}}\widehat{\boldsymbol{p}}^T\} = \left(\frac{L}{W}\right)^2 \left(\frac{\sigma^4}{L^2}\right) \mathbf{1}_{\boldsymbol{LL}} = \left(\frac{\sigma^4}{W^2}\right) \mathbf{1}_{LL} \tag{4.84}$$

where $\mathbf{1}_{LL}$ is an $L \times L$ matrix with all its elements equal to 1. It follows from (4.80), (4.81), and (4.84) that

$$\lim_{N\to\infty} \mathcal{C}_{\widehat{\boldsymbol{p}}} = 0. \tag{4.85}$$

In other words, the variance of the averaged correlogram for undersampled data tends to zero as the number of samples goes to infinity, which proves the consistency of the estimator. Moreover, all the elements of $\mathcal{C}_{\widehat{\boldsymbol{p}}}$ tend to zero, which implies that the estimations made for different spectral segments are asymptotically uncorrelated. □

Remark 4.2. Note that the conventional correlogram method is inconsistent for the case of a white Gaussian process. This is not in contradiction with

the result given by Theorem 4.5. Note that first of all, the averaged correlogram for undersampled data does not simplify to the conventional correlogram method if the average sampling rate is made equal to the Nyquist rate (due to the separation of the real and imaginary equations and also the presence of the fractional delay filters). Furthermore, the averaged correlogram for undersampled data estimates the average power in each spectral segment, and the conventional correlogram method also becomes a consistent estimator if the PSD estimation is averaged over a frequency band using a spectral window [57].

4.5 Numerical Examples

In this section, we investigate the behavior of the averaged correlogram for undersampled data for finite-length signals based on the analytical results obtained in Sections 4.3 and 4.4. Monte Carlo simulations have also been performed to validate the theoretical results.

The estimation bias and variance of the averaged correlogram method depends on the number of sampling channels q, the number of spectral segments L, and the number of samples per channel N. Here, the Nyquist sampling rate is considered to be $W = 1000$ Hz. The time offsets c_i $(1 \leq i \leq q)$ are distinct positive integer numbers less than L which are generated with equal probability for each (L, q)-pair. After generating the time offsets c_i, the matrix $\breve{\mathbf{\Psi}}$ is formed and its rank is checked. In the case that $\breve{\mathbf{\Psi}}$ is rank deficient, a new set of time offsets is generated until a full rank matrix $\breve{\mathbf{\Psi}}$ is obtained or a maximum number of tries is performed. In the latter case, the given (L, q)-pair is considered as unfeasible. Once a full rank matrix $\breve{\mathbf{\Psi}}$ is obtained, it is kept unchanged for different signal lengths.

We present six examples to illustrate the bias and variance of the averaged correlogram for undersampled data. For the first four examples, we consider a white Gaussian process with its PSD equal to $\sigma^2/W = 1$. For the last two examples, a filtered Gaussian process is used.

The estimation bias is investigated first. We consider the case when the average sampling rate $(q/L)W$ is kept unchanged. Therefore, for a given num-

ber of Nyquist samples, the overall number of samples available for estimation is the same for different (L, q)-pairs. Fig. 4.1 depicts the bias of the estimator versus the number of Nyquist samples N_x. The curve marked with squares is obtained by Monte Carlo simulations for comparison with the theoretical results. The rest of the curves are obtained from (4.25). Referring to (4.25) and (4.26), it can be seen that the bias is proportional to the inverse of the signal length N_x (consider multiplying (4.26) by L/L, and note that $N_x = NL$). Moreover, at a given signal length, the bias increases linearly with the number of spectral segments. It can also be seen that the estimation bias tends to zero as the length of the signal tends to infinity.

Fig. 4.2 depicts the variance of the estimator $[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{1,1}$ versus the number of sampling channels q for different values of spectral segments L. The signal length is fixed at $N_x = 10^5$. The curves drawn with solid lines represent the exact variance obtained from (4.42) to (4.49), and the curves plotted with dashed lines are the approximate values obtained from (4.73). Increasing q at a fixed L is equivalent to increasing the average sampling rate $(q/L)W$. According to the approximate variance as given in (4.73), the variance decreases quadratically with the number of sampling channels q. Therefore, the performance of the estimator improves by increasing q, but this comes at the price of adding to the complexity of the system by using more sampling channels.

Fig. 4.3 shows the variance of the estimator $[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{1,1}$ versus the number of spectral segments L for different numbers of sampling channels q. The signal length is fixed at $N_x = 10^5$. Again, the curves drawn with solid lines are obtained from (4.42) to (4.49), and the curves plotted with dashed lines are obtained from (4.73). According to the approximate variance as given in (4.73), the variance increases cubicly with the number of spectral segments L. Therefore, at a fixed signal length and fixed number of sampling channels, the performance of the estimator is degraded by increasing the number of spectral segments L, i.e., by increasing the frequency resolution.

The variance of the estimator $[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{1,1}$ versus the signal length N_x is illustrated in Fig. 4.4. Here, the average sampling rate $(q/L)W$ is kept unchanged. Therefore, for a given number of Nyquist samples, the overall number of sam-

ples available for estimation is the same for different (L, q)-pairs. The curve marked with squares is obtained by Monte Carlo simulations for comparison with the theoretical results. Again, the curves drawn with solid lines are obtained from (4.42) to (4.49), and the curves plotted with dashed lines are obtained from (4.73). Referring to the approximate variance as given in (4.73), the variance is almost proportional to the inverse of the signal length N_x. From the curves corresponding to the $(51, 12)$, $(101, 25)$, and $(201, 50)$-pairs in Fig. 4.4, it can be seen that the performance of the estimator degrades when increasing the number of spectral segments, i.e., when increasing the frequency resolution. The average sampling rate is kept almost the same in this scenario. It can also be seen that the estimation variance tends to zero as the length of the signal tends to infinity.

For the next two examples, we consider a more general case with a filtered Gaussian process. The signal is obtained by passing a white Gaussian signal through a bandlimited filter with cutoff frequencies set at $W/10$ and $W/5$ Hz. Through our experiments, we found that the estimation variance at each spectral segment depends not only on the power of signal at that frequency band, but also it is dependant on the power of the signal at other spectral segments. As noticed from the analytical derivations for the white Gaussian process (see (4.43), (4.49), and (4.73)), the estimation variance is proportional to the square of the signal power (σ^4/W^2). Therefore, we set the gain of the filter so that the square of the power averaged over all spectral segments for both the white Gaussian process at the input of the filter and the filtered signal is the same.

In Fig. 4.5, the variance of the estimator $[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{1,1}$ versus the number of spectral segments L is depicted. The number of sampling channels is set to $q = 45$, and the signal length is fixed at $N_x = 10^5$. The curve for the white Gaussian signal is based on (4.42)–(4.49), and the curve for the filtered Gaussian signal is obtained by Monte Carlo simulations. The latter curve is the average estimation variance of the spectral segments that pass through the filter. It can be seen in Fig. 4.5 that the variance of the estimator for the white and the filtered signals are close to each other.

Finally, the variance of the estimator $[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{1,1}$ versus the signal length N_x

for the white and the filtered signals is investigated. The number of spectral segments is set to $L = 101$, and the number of sampling channels is set to $q = 25$. Again, the curve for the white Gaussian signal is based on (4.42)–(4.49), and the curve for the filtered Gaussian signal is obtained by Monte Carlo simulations. Similar to the previous example, it can be seen in Fig. 4.6 that the estimation variance for the white and the filtered signals are close to each other. It can also be seen that the estimation variance tends to zero as the length of the signal tends to infinity.

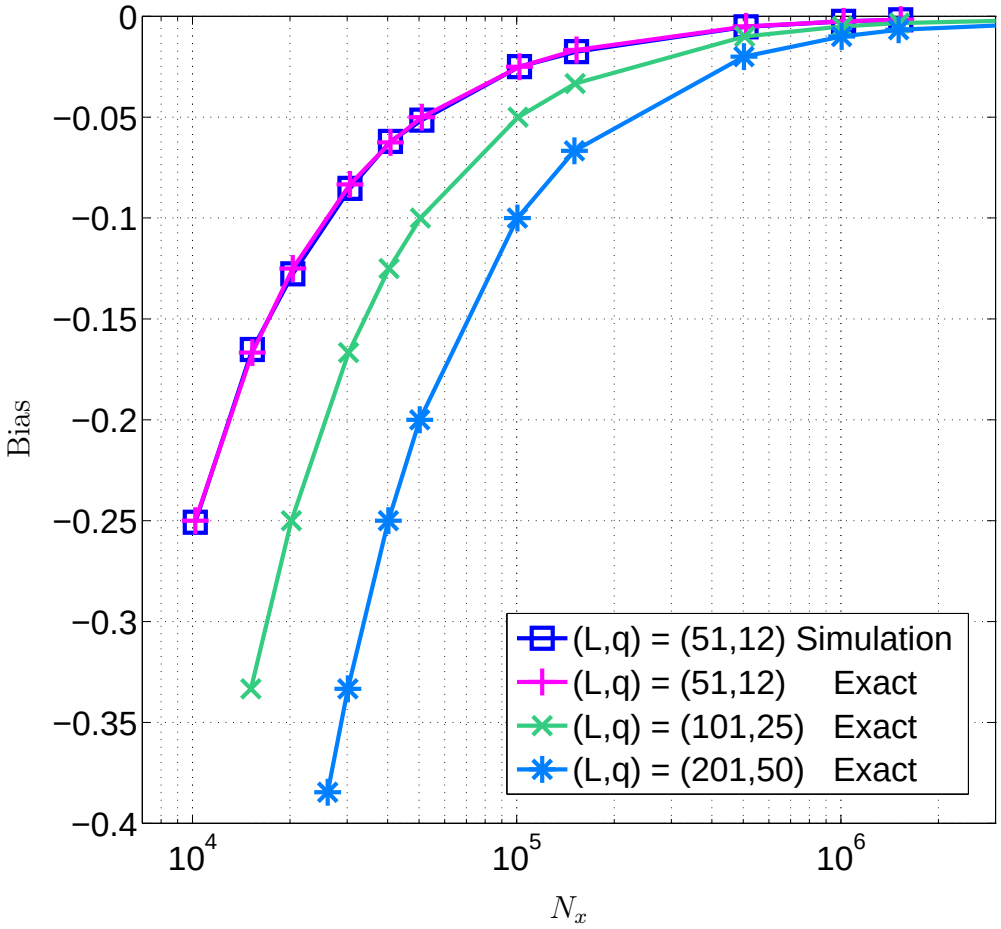

Figure 4.1: Bias versus Nyquist signal length N_x. The average sampling rate $(q/L)W$ for the $(L, q) = (51, 12)$, $(101, 25)$, and $(201, 50)$ pairs are 235 Hz, 247 Hz, and 248 Hz, respectively. The curve marked with squares is obtained by Monte Carlo simulations. The rest of the curves are based on (4.25).

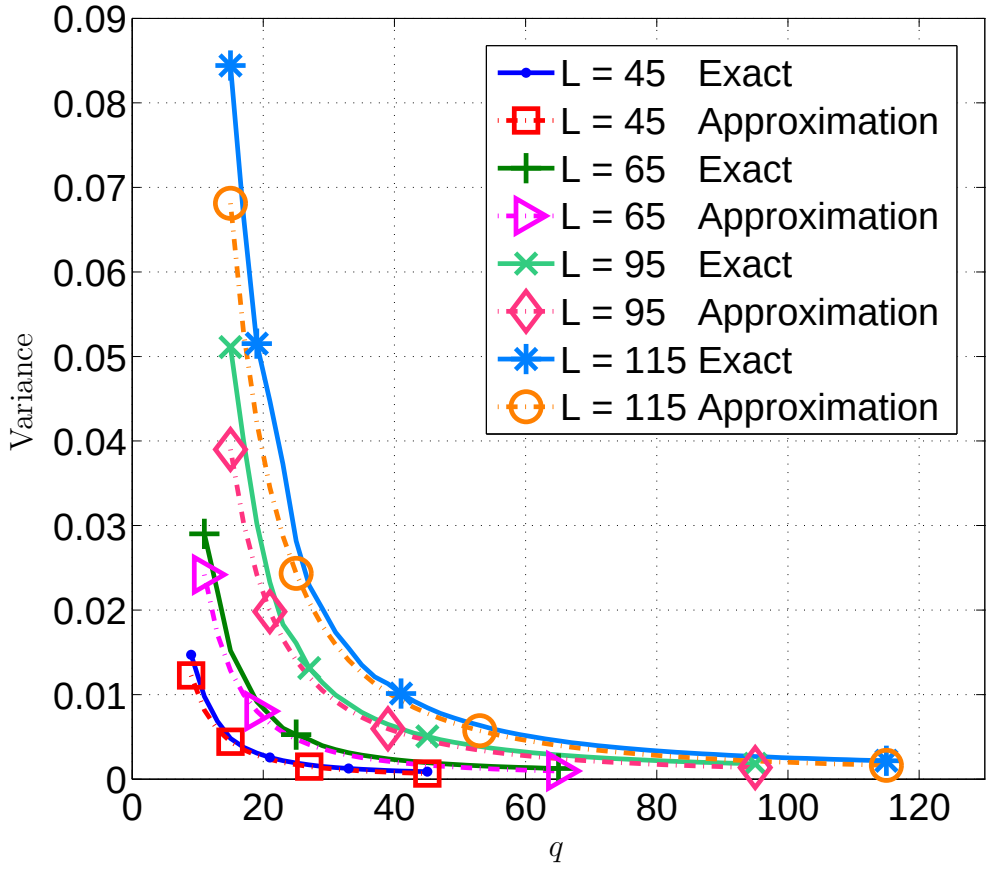

Figure 4.2: Variance $[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{1,1}$ versus number of sampling channels q at a fixed number of spectral segments L. The number of Nyquist samples is set to $N_x = 10^5$. Solid lines are based on (4.42)–(4.49) and dashed lines are based on (4.73).

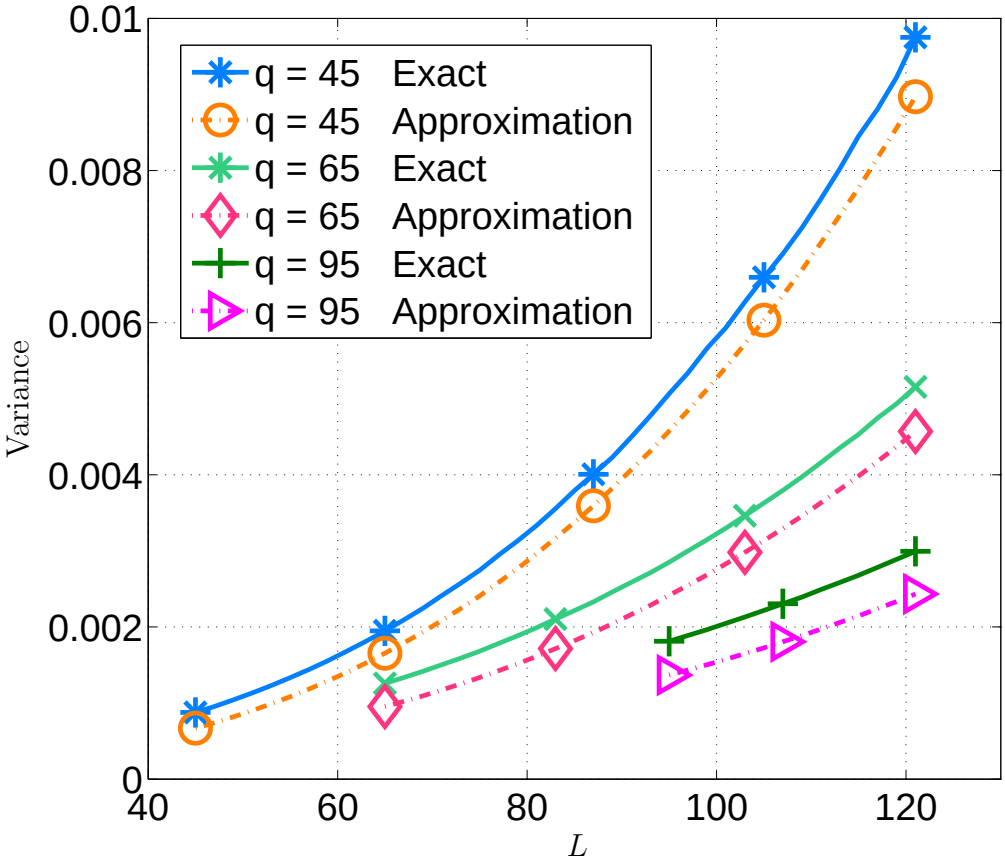

Figure 4.3: Variance $[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{1,1}$ versus number of spectral segments L at a fixed number of sampling channels q. The number of Nyquist samples is set to $N_x = 10^5$. Solid lines are based on (4.42)–(4.49) and dashed lines are based on (4.73).

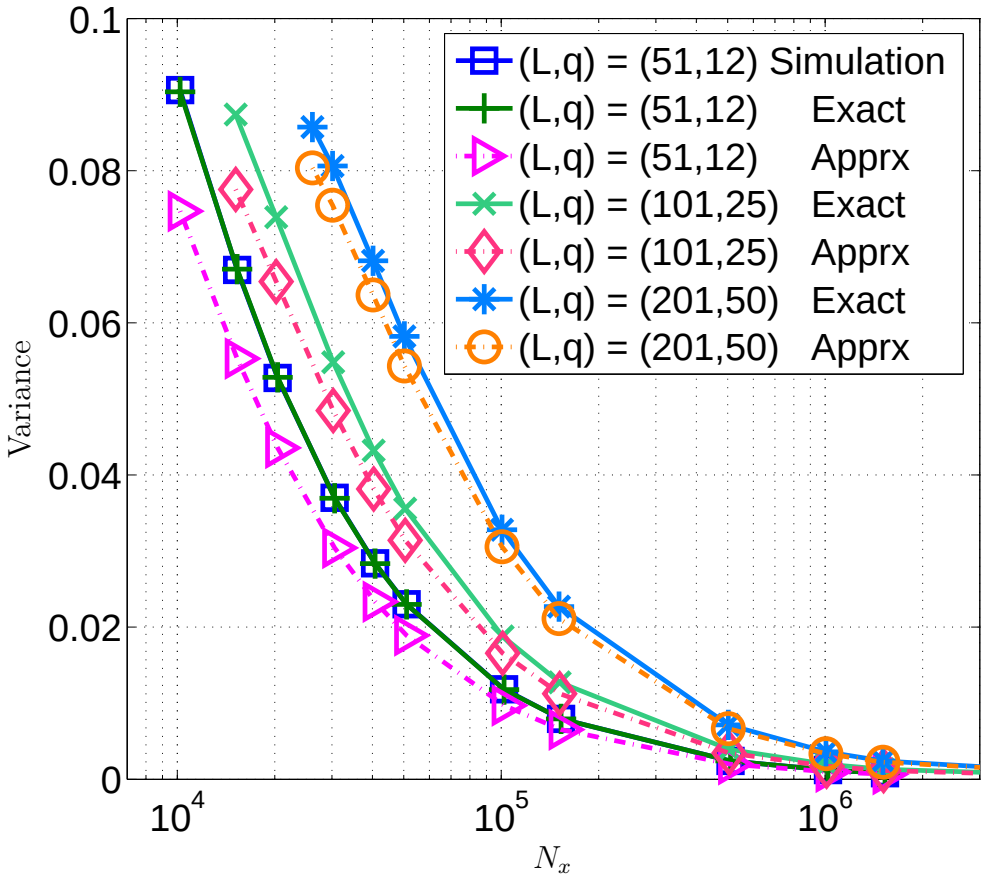

Figure 4.4: Variance $[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{1,1}$ versus Nyquist signal length N_x. The average sampling rate $(q/L)W$ for the $(L, q) = (51, 12)$, $(101, 25)$, and $(201, 50)$ pairs are 235 Hz, 247 Hz, and 248 Hz, respectively. The curve marked with squares is obtained by Monte Carlo simulations. Solid lines are based on (4.42)–(4.49), and dashed lines are based on (4.73).

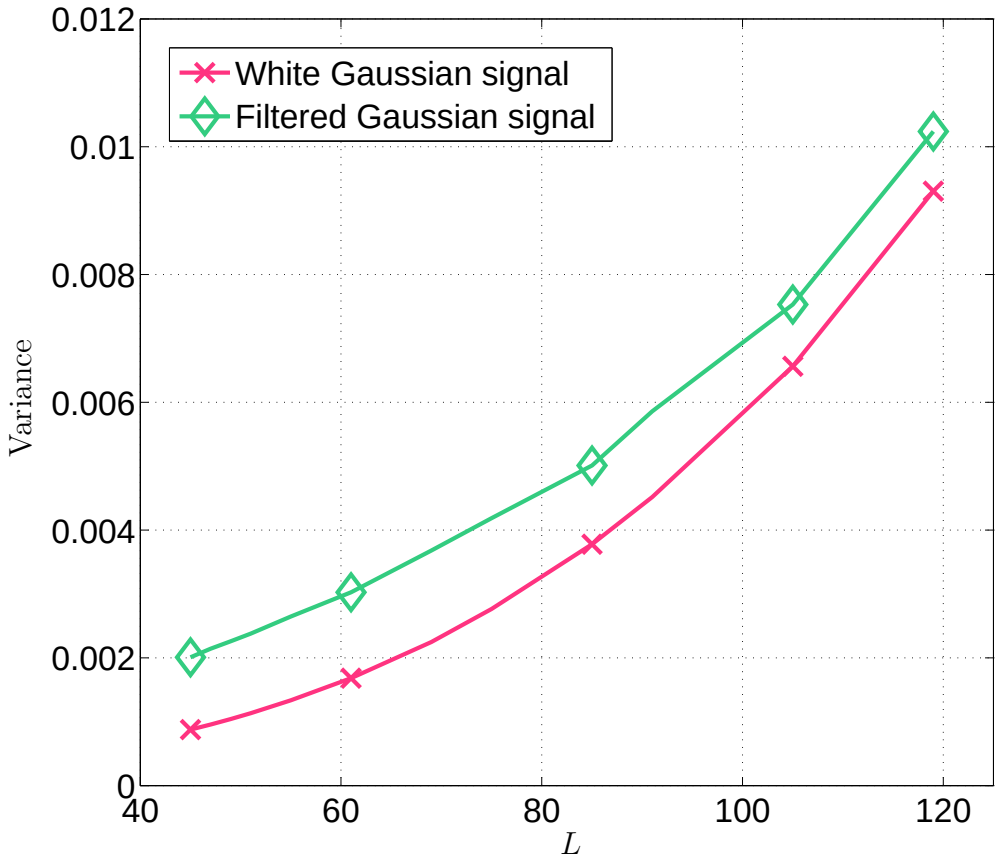

Figure 4.5: Variance $[\mathcal{C}_{\widehat{\mathbf{p}}}]_{1,1}$ versus number of spectral segments L at a fixed number of sampling channels $q = 45$. The number of Nyquist samples is set to $N_x = 10^5$. The curve for the white Gaussian signal is based on (4.42)–(4.49), and the curve for the filtered Gaussian signal is obtained by Monte Carlo simulations.

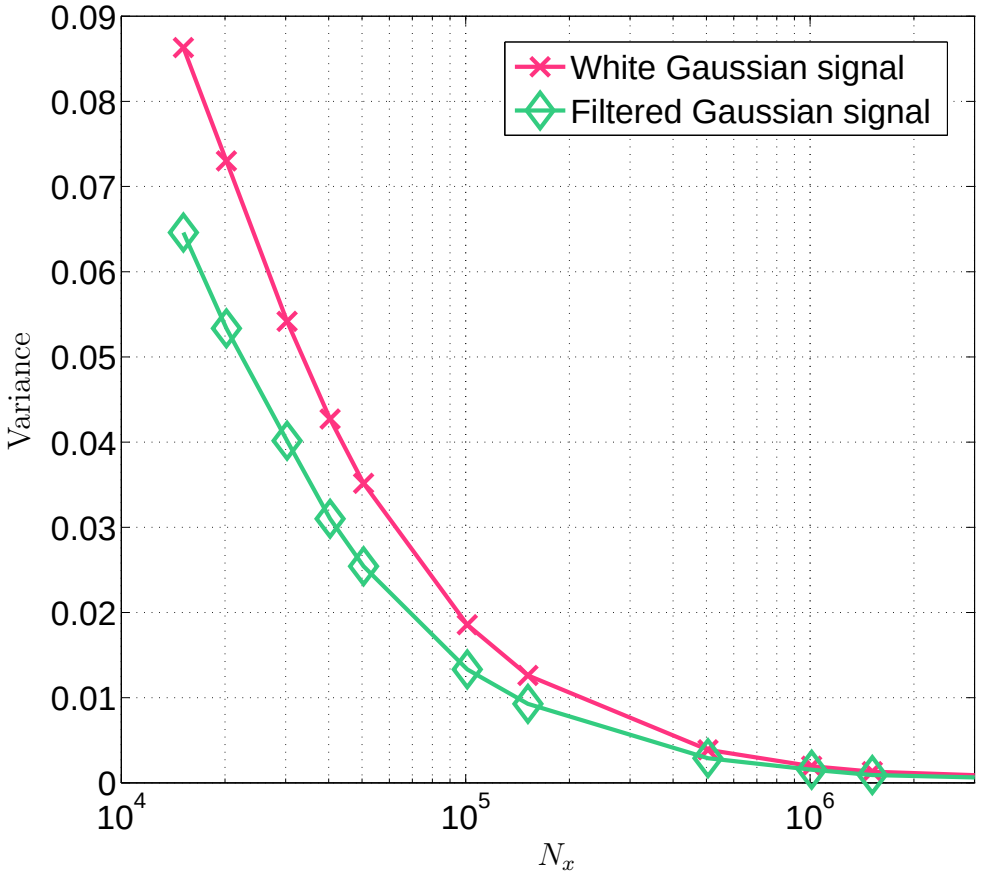

Figure 4.6: Variance $[\mathcal{C}_{\widehat{\boldsymbol{p}}}]_{1,1}$ versus Nyquist signal length N_x for $(L, q) = (101, 25)$ pair. The curve for the white Gaussian signal is based on (4.42)–(4.49), and the curve for the filtered Gaussian signal is obtained by Monte Carlo simulations.

Chapter 5

Improved Model-Based Spectral Compressive Sensing via Nested Least Squares

Signals with sparse representations can be recovered from a number of measurements much less than the number of samples given by the Nyquist sampling rate using the CS methods [9], [10]. Such measurements are obtained by correlating the signal with a number of sensing waveforms. The algorithms used for recovering the signal from these measurements exploit the sparsity or compressibility of the signal in a proper basis. There exist two main approaches for signal reconstruction: convex relaxation and greedy pursuits [9, 10, 58–60].

A special type of sparse signals that appears frequently in signal processing and digital communications is the class of signals composed of linear combinations of sinusoids [61], [62]. Albeit these type of signals generate sparse coefficients by the DTFT, their representation in the Fourier basis obtained by the DFT exhibits frequency leakage. This problem results in the poor performance of the conventional CS recovery algorithms that rely on the Fourier basis (see [63]). Although these signals do not have a sparse representation in the Fourier basis, they possess a sparse model in terms of the DTFT. In [64], the advantages of taking a signal model into account for signal reconstruction have been demonstrated and the name *model-based CS* has been coined. Although the model-based CS method still requires the sparsity or compressibility of the signal, it can be modified to handle the class of signals of our

interest. In [63], two model-based reconstruction methods have been considered for the signals of interest, and it has been shown that they outperform conventional CS recovery methods. According to these methods, the signal is reconstructed in an iterative manner, where at each iteration, a signal estimate is formed and pruned according to the signal structure.

In this chapter, we introduce a new spectral CS recovery method. The important difference of our method from that of [63] is the approach used for estimating the amplitudes of the signal elements. In [63], the unknown amplitudes are estimated using the DTFT, while we estimate the amplitudes by minimizing the squared norm of the compressed estimation error. Furthermore, we analyze the proposed method, derive the CRB for spectral CS, and show that the proposed algorithm approaches the CRB.

5.1 Model-Based Nested Least Squares Method

Let the signal $\boldsymbol{x} = [x(0),\ x(1),\ \ldots,\ x(N_x-1)]^T \in \mathbb{C}^{N_x \times 1}$ be a linear combination of K sinusoids ($K \ll N_x$) where $x(n)$ ($0 \leq n < N_x$) are the samples of the signal obtained at the Nyquist rate. Here the sample $x(n)$ is given by

$$x(n) = \sum_{k=1}^{K} d_k e^{-j\omega_k n} \tag{5.1}$$

where d_k and ω_k ($1 \leq k \leq K$) are unknown amplitudes and frequencies of the K sinusoids, respectively. By arranging the amplitude parameters in the vector $\boldsymbol{d} = [d_1,\ d_2,\ \ldots,\ d_K]^T \in \mathbb{C}^{K \times 1}$ and forming the matrix $\boldsymbol{A} = [\boldsymbol{a}(\omega_1),\ \boldsymbol{a}(\omega_2),\ \ldots,\ \boldsymbol{a}(\omega_K)] \in \mathbb{C}^{N_x \times K}$ with the frequency parameters, the model (5.1) can be rewritten in the matrix-vector form as

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{d} \tag{5.2}$$

where $\boldsymbol{a}(\omega) = [1,\ e^{-j\omega},\ e^{-j2\omega},\ \ldots,\ e^{-j(N_x-1)\omega}]^T \in \mathbb{C}^{N_x \times 1}$ is the Vandermonde vector.

Let the vector of the measurements $\boldsymbol{y} \in \mathbb{C}^{N_y \times 1}$ be given by

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{w} \tag{5.3}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{N_y \times N_x}$ is the measurement matrix, and $\boldsymbol{w} \in \mathbb{C}^{N_y \times 1}$ is the measurement noise with circularly-symmetric complex jointly-Gaussian distribution $\mathcal{N}_C(0, \sigma_w^2 I)$. The elements of the measurement matrix $\boldsymbol{\Phi}$ are drawn independently from, for example, the Gaussian distribution $\mathcal{N}(0, 1/N_y)$.

The goal is to estimate the unknown amplitudes and frequencies of the signal (5.2) from the noisy compressive measurements (5.3).

Two criteria are taken into consideration for developing the estimation algorithm: minimization of the estimation error and matching the estimated signal to the sparsity model. The squared norm of the compressed estimation error is $\|\boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{x}}\|_2^2$ where $\hat{\boldsymbol{x}}$ is the estimated signal. Thus, the problem of finding the estimate $\hat{\boldsymbol{x}}$ can be formulated as

$$\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2. \tag{5.4}$$

The estimation error is a convex function, and the minimization of (5.4) can be obtained using the least squares (LS) technique with the iterative solution

$$\hat{\boldsymbol{x}}_i = \hat{\boldsymbol{x}}_{i-1} + \beta \boldsymbol{\Phi}^T(\boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{x}}_{i-1}) \tag{5.5}$$

where $\hat{\boldsymbol{x}}_i$ is the estimated signal at the ith iteration and β represents the step size of the LS algorithm or equivalently the scaling factor for the residual signal of the previous iteration, that is, $\boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{x}}_{i-1}$.

The LS problem of (5.4) is underdetermined and has many solutions. In order to match the estimated signal to the model in (5.1), a pruning step is inserted in the iterative solution of (5.5). Specifically, let $\boldsymbol{x}^e = \hat{\boldsymbol{x}}_{i-1} + \beta \boldsymbol{\Phi}^T(\boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{x}}_{i-1})$, then the frequencies $\omega_1, \omega_2, \ldots, \omega_K$ can be estimated from $\boldsymbol{x}^e$ using, for example, the root-MUSIC technique [27]. This method needs the knowledge of the autocorrelation matrix of the data $\boldsymbol{R}_{\boldsymbol{x}}$ for estimating the frequencies.

Consider windowing $\boldsymbol{x}^e$ by overlapping frames of length W_x. Then, the elements of $\boldsymbol{R}_{\boldsymbol{x}}$ can be estimated as

$$[\widehat{\boldsymbol{R}}_{\boldsymbol{x}}]_{a,b} = \frac{1}{N_x - W_x + 1} \sum_{n=W_x}^{N_x} x^{e*}(n-a)x^e(n-b) \tag{5.6}$$

where $\widehat{\boldsymbol{R}}_{\boldsymbol{x}}$ is an estimation for $\boldsymbol{R}_{\boldsymbol{x}}$, $x^e(n)$ $(0 \leq n < N_x)$ are the elements of $\boldsymbol{x}^e$, and $1 \leq a, b \leq W_x$. Let $\widehat{\boldsymbol{\Omega}} = [\hat{\omega}_1, \hat{\omega}_2, \ \ldots, \ \hat{\omega}_K]^T \in [-\pi, \pi]^{K \times 1}$ be the vector of the estimated frequencies. Then, the estimate of the Vandermonde matrix $\boldsymbol{A}$ (denoted by $\widehat{\boldsymbol{A}}$) can be straightforwardly computed based on $\widehat{\boldsymbol{\Omega}}$.

Recalling the objective of minimizing the squared norm of the compressed estimation error, the vector of the amplitudes $\boldsymbol{d}$ can be estimated by minimizing $\|\boldsymbol{y} - \boldsymbol{\Phi}\widehat{\boldsymbol{A}}\boldsymbol{d}\|_2^2$ [65]. The solution for this problem is given by

$$\hat{\boldsymbol{d}}_i = (\widehat{\boldsymbol{B}}^H \widehat{\boldsymbol{B}})^{-1} \widehat{\boldsymbol{B}}^H \boldsymbol{y} \tag{5.7}$$

where $\widehat{\boldsymbol{B}} = \boldsymbol{\Phi}\widehat{\boldsymbol{A}}$. Note that in [63], the amplitudes are estimated as

$$\hat{\boldsymbol{d}}_i = \widehat{\boldsymbol{A}}^H \boldsymbol{x}^e \tag{5.8}$$

where $\hat{\boldsymbol{d}}_i$ is the vector of the estimated amplitudes at the i-th iteration. The algorithm based on (5.8) is referred to as *spectral iterative hard thresholding* (SIHT) via root-MUSIC.

Finally, $\hat{\boldsymbol{x}}_i$ can be obtained using the estimated frequencies and amplitudes as $\hat{\boldsymbol{x}}_i = \widehat{\boldsymbol{A}}\hat{\boldsymbol{d}}_i$. The steps of the algorithm are summarized in Table 5.1.

The algorithm consists of the outer and the inner least squares steps along with the root-MUSIC method. In each iteration, the algorithm converges to the true signal in three steps. First, the outer least squares makes an estimation of the subspace in which the original signal lies. This is done by minimizing the squared norm of the compressed estimation error. Note that due to the fact that the problem is underdetermined, the signal $\boldsymbol{x}$ cannot be estimated, but only an improved estimate of the subspace to which the signal $\boldsymbol{x}$ belongs can be found. Then, the signal estimate $\hat{\boldsymbol{x}}$ is enhanced in the second and the third steps of the algorithm. In the second step, the estimation is forced to match the signal model by applying the root-MUSIC method. The frequencies are estimated at this stage. Note that each frequency represents one of the dimensions of the signal subspace. In the first few iterations of the algorithm, some of the frequencies might be estimated incorrectly, as the output of the outer least squares step might not be close enough to the true signal subspace. In the third step of the algorithm, the amplitudes are estimated by applying

the inner least squares. The last two steps are building the signal subspace according to the signal model, and then, estimating the projection coefficients for each dimension of the subspace. Finally, the estimated signal is fed back to the outer least squares step for the next iteration. The algorithm continues until some stopping criterion is satisfied. For example, the criterion can be satisfied when a predetermined fixed number of iterations is performed or the normalized compressed estimation error ($\|\boldsymbol{y}-\boldsymbol{\Phi}\hat{\boldsymbol{x}}\|_2^2/\|\boldsymbol{y}\|_2^2$) is less than a given threshold value.

Table 5.1: Model-based nested least squares algorithm

Initialize: $\hat{\boldsymbol{x}}_0 = 0,\ i = 1$
repeat
$\boldsymbol{x}^e \leftarrow \hat{\boldsymbol{x}}_{i-1} + \beta \boldsymbol{\Phi}^T (\boldsymbol{y} - \boldsymbol{\Phi} \hat{\boldsymbol{x}}_{i-1})$
$\widehat{\boldsymbol{\Omega}} \leftarrow \text{root-MUSIC}(\boldsymbol{x}^e, K)$
$\widehat{\boldsymbol{A}} \leftarrow [\boldsymbol{a}(\hat{\omega}_1)\ \boldsymbol{a}(\hat{\omega}_2)\ \ldots\ \boldsymbol{a}(\hat{\omega}_K)]$
$\widehat{\boldsymbol{B}} \leftarrow \boldsymbol{\Phi} \widehat{\boldsymbol{A}}$
$\hat{\boldsymbol{d}}_i \leftarrow (\widehat{\boldsymbol{B}}^H \widehat{\boldsymbol{B}})^{-1} \widehat{\boldsymbol{B}}^H \boldsymbol{y}$
$\hat{\boldsymbol{x}}_i \leftarrow \widehat{\boldsymbol{A}} \hat{\boldsymbol{d}}_i$
$i \leftarrow i + 1$
until stopping criterion is satisfied

5.2 Cramér-Rao Bound for Spectral Compressive Sensing

The CRB for the problem of estimating the parameters of multiple superimposed exponential signals in noise has been derived in [56]. In this section, the CRB for spectral compressive sensing is derived by considering the system model (5.2) and (5.3).

First, let the vector of parameters be defined by $\boldsymbol{\vartheta} = [\bar{\boldsymbol{d}}^T \ \tilde{\boldsymbol{d}}^T \ \boldsymbol{\Omega}^T]^T$ where $\bar{\boldsymbol{d}}$ and $\tilde{\boldsymbol{d}}$ represent the real and imaginary parts of $\boldsymbol{d}$, respectively. Furthermore, let

$$\boldsymbol{D} = \text{diag}(\boldsymbol{d}) = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_K \end{bmatrix} \tag{5.9}$$

and $\boldsymbol{C} = [\boldsymbol{c}(\omega_1) \ \ldots \ \boldsymbol{c}(\omega_K)]$ where $\boldsymbol{c}(\omega) = d\boldsymbol{a}(\omega)/d\omega$.

The likelihood function of the measurement vector $\boldsymbol{y}$ is

$$L(\boldsymbol{y}) = \frac{1}{(\pi\sigma_w^2)^{N_y}} \exp\big\{ -\frac{1}{\sigma_w^2}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{d})^H(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{d})\big\} \tag{5.10}$$

where $\boldsymbol{B} = \boldsymbol{\Phi}\boldsymbol{A}$.

The inverse of the Fisher information matrix is given by

$$I^{-1}(\boldsymbol{\vartheta}) = (E\{\boldsymbol{\psi}\boldsymbol{\psi}^T\})^{-1} \tag{5.11}$$

where $\boldsymbol{\psi} = \partial \ln L/\partial\boldsymbol{\vartheta}$.

The log-likelihood function is

$$\ln L = -N_y \ln \pi - N_y \ln \sigma_w^2 - \frac{1}{\sigma_w^2}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{d})^H(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{d}) \tag{5.12}$$

and its derivatives with respect to $\bar{\boldsymbol{d}}$ and $\tilde{\boldsymbol{d}}$ are

$$\frac{\partial \ln L}{\partial \bar{\boldsymbol{d}}} = \frac{1}{\sigma_w^2}\left(\boldsymbol{B}^H\boldsymbol{w} + \left(\boldsymbol{w}^H\boldsymbol{B}\right)^T\right) = \frac{2}{\sigma_w^2} Re\left\{\boldsymbol{B}^H\boldsymbol{w}\right\} \tag{5.13}$$

and

$$\frac{\partial \ln L}{\partial \tilde{\boldsymbol{d}}} = \frac{1}{\sigma_w^2}\left(-j\boldsymbol{B}^H\boldsymbol{w} + j\left(\boldsymbol{w}^H\boldsymbol{B}\right)^T\right) = \frac{2}{\sigma_w^2} Im\left\{\boldsymbol{B}^H\boldsymbol{w}\right\} \tag{5.14}$$

respectively. Recall that $\boldsymbol{w} = \boldsymbol{y} - \boldsymbol{B}\boldsymbol{d}$ is the measurement noise introduced in (5.3).

The derivative of the log-likelihood function with respect to ω_k for $1 \leq k \leq K$ is

$$\begin{aligned}\frac{\partial \ln L}{\partial \omega_k} &= \frac{2}{\sigma_w^2} Re\left\{\boldsymbol{d}^H \frac{d\boldsymbol{B}^H}{d\omega_k}\boldsymbol{w}\right\} \\ &= \frac{2}{\sigma_w^2} Re\left\{\boldsymbol{d}^H \frac{d\boldsymbol{A}^H}{d\omega_k}\boldsymbol{\Phi}^T\boldsymbol{w}\right\} \\ &= \frac{2}{\sigma_w^2} Re\left\{d_k^* \boldsymbol{c}^H(\omega_k)\boldsymbol{\Phi}^T\boldsymbol{w}\right\}. \end{aligned} \tag{5.15}$$

The derivatives of the log-likelihood function with respect to the frequencies can be written in the matrix form as

$$\frac{\partial \ln L}{\partial \boldsymbol{\Omega}} = \frac{2}{\sigma_w^2} Re\left\{\boldsymbol{D}^H \boldsymbol{C}^H \boldsymbol{\Phi}^T \boldsymbol{w}\right\}. \tag{5.16}$$

To proceed, we use the extension of Lemma 1 to the vector case. Then, the submatrices of $I(\boldsymbol{\vartheta})$ can be computed as

$$\begin{aligned} E\left[\frac{\partial \ln L}{\partial \bar{\boldsymbol{d}}}\right]\left[\frac{\partial \ln L}{\partial \bar{\boldsymbol{d}}}\right]^T &= \frac{4}{\sigma_w^4}\frac{1}{2} Re\left\{E\left\{\boldsymbol{B}^H \boldsymbol{w}\boldsymbol{w}^H \boldsymbol{B}\right\}\right\} \\ &= \frac{2}{\sigma_w^2} Re\left\{\boldsymbol{B}^H \boldsymbol{B}\right\} \end{aligned} \tag{5.17}$$

$$E\left[\frac{\partial \ln L}{\partial \bar{\boldsymbol{d}}}\right]\left[\frac{\partial \ln L}{\partial \tilde{\boldsymbol{d}}}\right]^T = -\frac{2}{\sigma_w^2} Im\left\{\boldsymbol{B}^H \boldsymbol{B}\right\} \tag{5.18}$$

$$E\left[\frac{\partial \ln L}{\partial \bar{\boldsymbol{d}}}\right]\left[\frac{\partial \ln L}{\partial \boldsymbol{\Omega}}\right]^T = \frac{2}{\sigma_w^2} Re\left\{\boldsymbol{B}^H \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{D}\right\} \tag{5.19}$$

$$E\left[\frac{\partial \ln L}{\partial \tilde{\boldsymbol{d}}}\right]\left[\frac{\partial \ln L}{\partial \tilde{\boldsymbol{d}}}\right]^T = \frac{2}{\sigma_w^2} Re\left\{\boldsymbol{B}^H \boldsymbol{B}\right\} \tag{5.20}$$

$$E\left[\frac{\partial \ln L}{\partial \tilde{\boldsymbol{d}}}\right]\left[\frac{\partial \ln L}{\partial \boldsymbol{\Omega}}\right]^T = \frac{2}{\sigma_w^2} Im\left\{\boldsymbol{B}^H \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{D}\right\} \tag{5.21}$$

$$E\left[\frac{\partial \ln L}{\partial \boldsymbol{\Omega}}\right]\left[\frac{\partial \ln L}{\partial \boldsymbol{\Omega}}\right]^T = \frac{2}{\sigma_w^2} Re\left\{\boldsymbol{D}^H \boldsymbol{C}^H \boldsymbol{\Phi}^T \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{D}\right\}. \tag{5.22}$$

Note that $E\left\{\boldsymbol{w}\boldsymbol{w}^T\right\} = \boldsymbol{0}$. Then, $I(\boldsymbol{\vartheta})$ is given by

$$I(\boldsymbol{\vartheta}) = \begin{bmatrix} \bar{\boldsymbol{F}} & -\tilde{\boldsymbol{F}} & \bar{\boldsymbol{\Delta}} \\ \tilde{\boldsymbol{F}} & \bar{\boldsymbol{F}} & \tilde{\boldsymbol{\Delta}} \\ \bar{\boldsymbol{\Delta}}^T & \tilde{\boldsymbol{\Delta}}^T & \boldsymbol{\Lambda} \end{bmatrix} \tag{5.23}$$

where $\bar{(\cdot)}$ and $\tilde{(\cdot)}$ stand for the real and imaginary parts of a matrix, and

$$\boldsymbol{F} = \frac{2}{\sigma_w^2}\boldsymbol{B}^H\boldsymbol{B} \tag{5.24}$$

$$\boldsymbol{\Delta} = \frac{2}{\sigma_w^2}\boldsymbol{B}^H\boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{D} \tag{5.25}$$

$$\boldsymbol{\Lambda} = \frac{2}{\sigma_w^2}Re\left\{\boldsymbol{D}^H\boldsymbol{C}^H\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{D}\right\}. \tag{5.26}$$

The signal $\boldsymbol{x}$ can be considered as a function of $\boldsymbol{\vartheta}$, and therefore, the covariance matrix of any unbiased estimator of $\boldsymbol{x}$, that is, $\mathcal{C}_{\hat{\boldsymbol{x}}}$, satisfies the inequality [66]

$$\mathcal{C}_{\hat{\boldsymbol{x}}} - \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\vartheta}} I^{-1}(\boldsymbol{\vartheta}) \frac{\partial \boldsymbol{x}^H}{\partial \boldsymbol{\vartheta}} \geq \boldsymbol{0}. \tag{5.27}$$

Moreover, the signal $\boldsymbol{x}$ can be written as

$$\boldsymbol{x} = \boldsymbol{A}\bar{\boldsymbol{d}} + j\boldsymbol{A}\tilde{\boldsymbol{d}} = d_1\boldsymbol{a}(\omega_1) + \ldots + d_K\boldsymbol{a}(\omega_K). \tag{5.28}$$

Then the derivative of $\boldsymbol{x}$ with respect to the whole vector of unknown parameters $\boldsymbol{\vartheta}$ can be found as

$$\begin{aligned} \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\vartheta}} &= [\boldsymbol{A} \;\; j\boldsymbol{A} \;\; d_1\boldsymbol{c}(\omega_1) \;\; \ldots \;\; d_K\boldsymbol{c}(\omega_K)] \\ &= [\boldsymbol{A} \;\; j\boldsymbol{A} \;\; \boldsymbol{C}\boldsymbol{D}]. \end{aligned} \tag{5.29}$$

Finally, by summing over the diagonal elements of (5.27), we obtain

$$E\left\{\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2\right\} \geq \mathrm{Tr}\left\{\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\vartheta}} I^{-1}(\boldsymbol{\vartheta}) \frac{\partial \boldsymbol{x}^H}{\partial \boldsymbol{\vartheta}}\right\} \triangleq CRB. \tag{5.30}$$

5.3 Simulation Results

Consider a signal consisting of $K = 20$ complex-valued sinusoids with a length of $N_x = 1024$ samples. The window size in the computation of $\widehat{\boldsymbol{R}}_{\boldsymbol{x}}$ is set to $W_x = 200$. The frequencies $(\omega_1, \omega_2, \ldots, \omega_K)$ are drawn randomly from the $[0, 2\pi)$ interval with the constraint that the pairs of the frequencies are spaced by at least $10\pi/N_x$ radians/sample. Furthermore, the amplitudes $(s_1, s_2, \ldots, s_K)$ are uniformly drawn at random from the $[1, 2]$ interval. In all of our simulations, the step size of the outer LS algorithm is set to 1 ($\beta = 1$). The number of Monte Carlo trials is also set to 300.

The normalized MSE (NMSE) is defined as

$$NMSE = 10 \log \left(\frac{E\{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2\}}{E\{\|\boldsymbol{x}\|_2^2\}} \right). \tag{5.31}$$

Recalling (5.30), the normalized CRB (NCRB) is defined as

$$NCRB = 10 \log \left(\frac{CRB}{E\{\|\boldsymbol{x}\|_2^2\}} \right). \tag{5.32}$$

The first experiment explores the performance of the model-based nested LS and the SIHT via root-MUSIC algorithms [63] over 10 iterations. The number of measurements is set to 300 ($N_y = 300$) and the noise standard deviation to 2 ($\sigma_w = 2$). The simulation results are illustrated in Fig. 5.1. It can be seen that the model-based nested LS algorithm converges after 5 iterations, while the SIHT via root-MUSIC method requires more iterations to converge. At the 5th iteration, the proposed algorithm performs 3 dB better than the SIHT method, and it is 1 dB away from the NCRB. After 10 iterations, the algorithm still performs 1 dB better than the SIHT method.

Next, the performance of the algorithm is investigated for a range of noise variances. The result of the second experiment is depicted in Fig. 5.2. The number of measurements is set to 300 ($N_y = 300$) and the number of iterations of the algorithm to 10. Similar to the previous example, the proposed method outperforms the SIHT via root-MUSIC algorithm. Moreover, it can be seen that the performance of the proposed algorithm approaches the bound at high signal to noise ratio values.

The third experiment investigates the performance of the algorithms for different numbers of measurements. The noise standard deviation is set to 2 ($\sigma_w = 2$), and the number of iterations is set to 10. The results are shown in Fig. 5.3. It can be seen that with 200 measurements, the proposed algorithm is able to recover the signal, while the performance of the SIHT method is significantly far from the NCRB. For larger number of measurements, the model-based nested LS algorithm performs about 1 dB better than the SIHT via root-MUSIC method, and it is about 1 dB away from the NCRB.

Finally, we investigate the convergence of the proposed algorithm by counting the number of missed signal frequencies over the iterations of the algorithm.

A frequency of the true signal is considered as missing in the estimated signal when the root-MUSIC algorithm does not output any frequency within a distance of less than $5\pi/N_x$ radians/sample (which is the resolution limit under the simulation set-up) to the true frequency.

The average number of missed signal frequencies over 4 iterations is presented in Table 5.2. The number of the measurements is set to 300 ($N_y = 300$). It can be seen that after 3 iterations, the root-MUSIC algorithm is able to find all the signal frequencies (for $\sigma_w = 2,\ 3,$ and 4). This indicates that the outer LS step of the algorithm is converging to the true signal subspace, as the root-MUSIC algorithm is able to distinguish more accurately between the signal and noise subspaces.

Table 5.2: Average number of missed signal frequencies

Iteration number	1	2	3	4
$\sigma_w = 2$	2.89	0.05	0	0
$\sigma_w = 3$	3.11	0.12	0	0
$\sigma_w = 4$	3.24	0.23	0.01	0

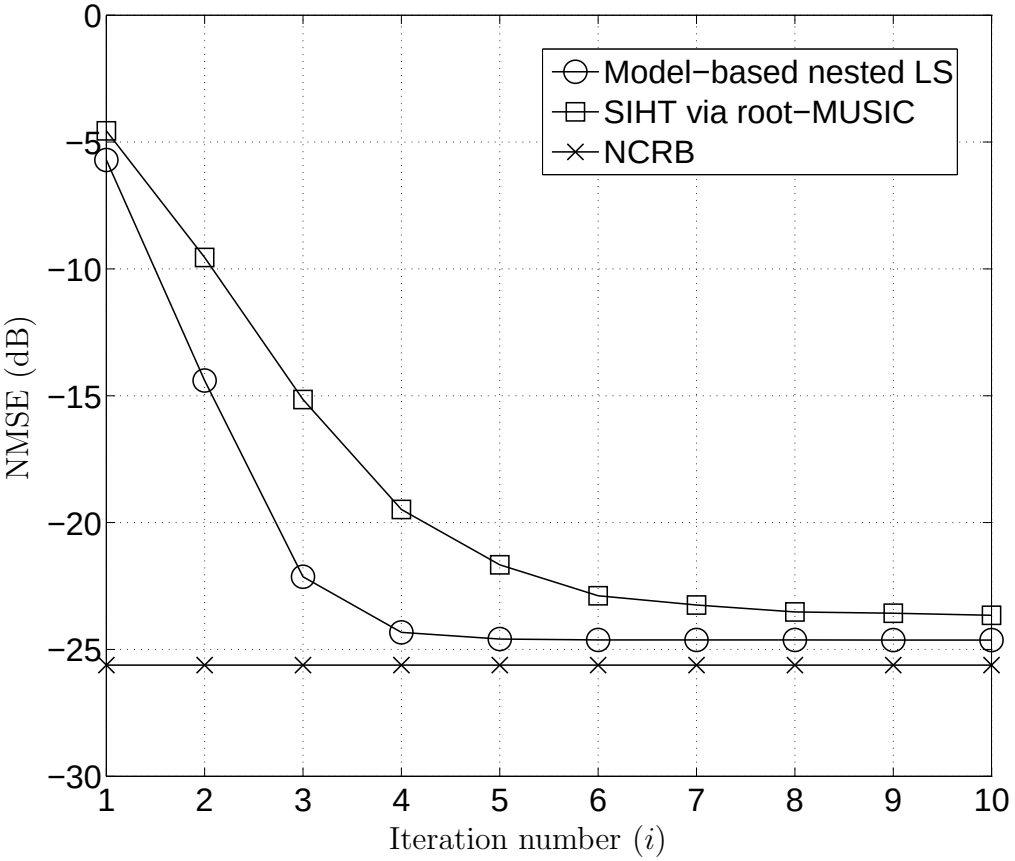

Figure 5.1: Normalized mean squared error versus iteration number for $\sigma_w = 2$ and $N_y = 300$.

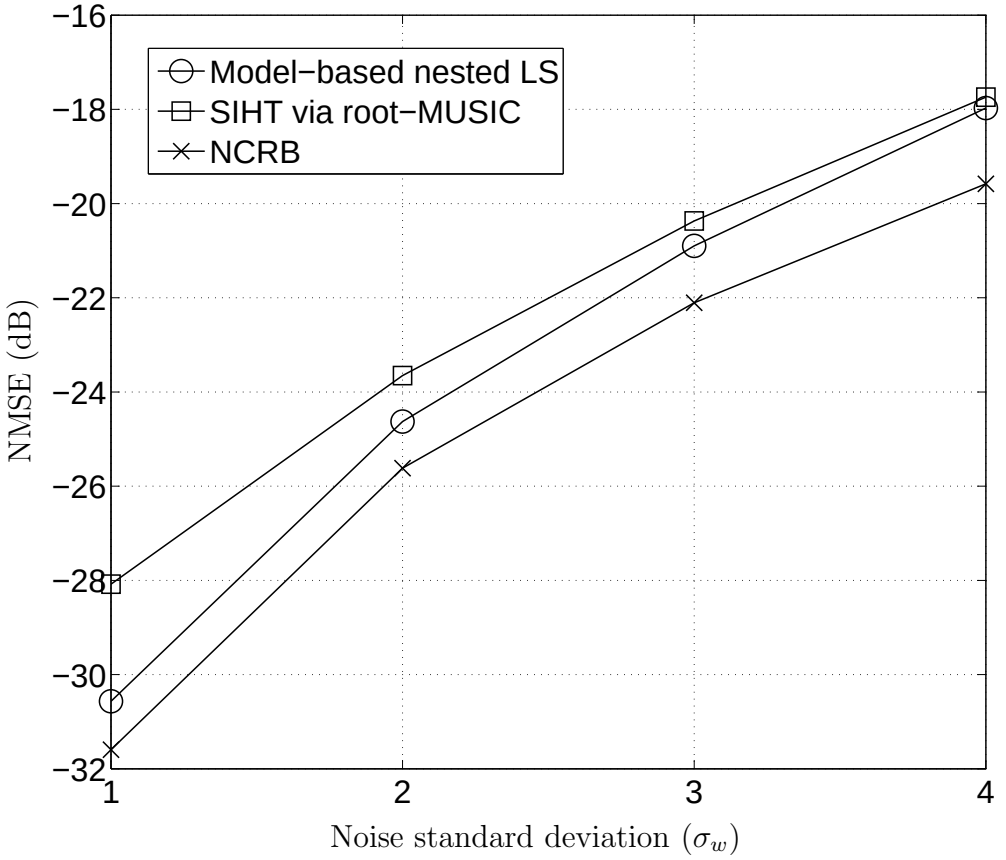

Figure 5.2: Normalized mean squared error versus noise standard deviation after 10 iterations for $N_y = 300$.

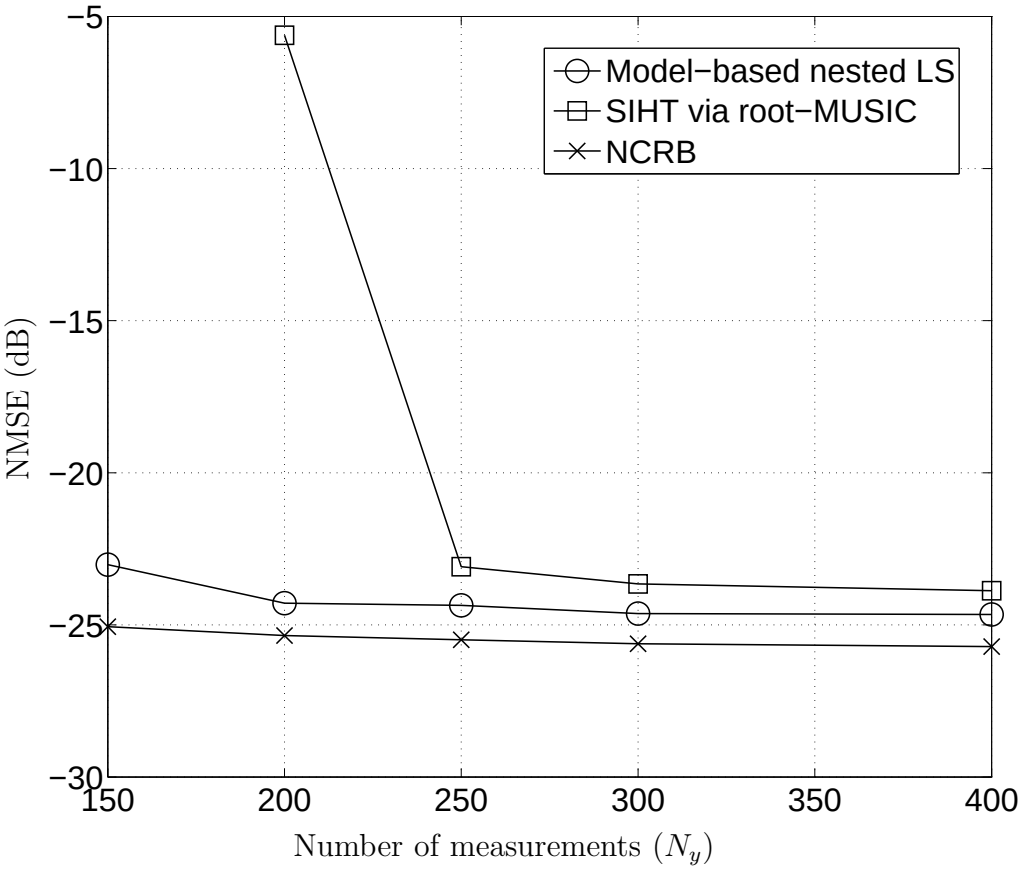

Figure 5.3: Normalized mean squared error versus number of measurements after 10 iterations for $\sigma_w = 2$.

Chapter 6

Conclusion

A summary of the contributions of the thesis is given in this chapter. We also comment on the future work that can be done regarding different research problems considered.

6.1 Summary of contributions

In this dissertation, we tackled the problem of parameter estimation in the small sample size region and undersampled data scenarios. The contributions of the thesis are as follows.

6.1.1 DOA estimation in the low sample size region

The performance breakdown of the subspace based DOA estimation methods in the threshold region where the SNR and/or sample size is low has been studied in Chapter 3. The subspace leakage as the main cause of the performance breakdown was formally defined and theoretically derived. We proposed a two-step algorithm in order to reduce the amount of subspace leakage. The introduced method is based on estimating the DOAs at the first step and modifying the covariance matrix using the estimated DOAs at the second step. We theoretically derived the subspace leakage at both steps, and it was shown that the subspace leakage was reduced at the second step leading to better performance.

The behavior of the root-MUSIC algorithm in the threshold region has been also studied, and a phenomenon called root-swap was observed to contribute

to the performance breakdown. Then, an improved method was introduced to remedy this problem by considering different combinations of the roots and picking up the one that minimizes the stochastic maximum likelihood function.

The performance improvement achieved by the proposed method was presented using numerical examples and simulation results. We also combined the proposed algorithm with the previously introduced methods in the literature, which resulted in further improvement in the performance.

6.1.2 Analysis of averaged correlogram for undersampled data

We considered the averaged correlogram for undersampled data which estimates the spectrum from a subset of Nyquist samples. This method has been analyzed in Chapter 4 by computing the bias and variance of the estimator.

The behavior of the estimator for finite-length signals has been investigated. It has been shown that at a given signal length, the estimation accuracy increases as the average sampling rate is increased (either by decreasing the frequency resolution or by increasing the complexity of the system). It has also been shown that at a fixed average sampling rate, the performance of the estimator degrades for the estimation with higher frequency resolution. To sum up, it has been illustrated that there is a tradeoff between the accuracy of the estimator (the estimation variance), the frequency resolution (the number of spectral segments), and the complexity of the estimator (the number of sampling channels).

It has also been shown that in the case of a white Gaussian process, this method is consistent, as the bias and variance of the estimator tend to zero asymptotically. Furthermore, it has been shown that the estimation made for different spectral segments becomes uncorrelated as the signal length goes to infinity.

6.1.3 Model-based spectral compressive sensing

A new signal recovery algorithm for model-based spectral compressive sensing has been introduced in Chapter 5. We considered a general signal model

consisting of complex-valued sinusoids with unknown frequencies and amplitudes. Although the signal model is inherently sparse, its representation in the Fourier basis does not offer much sparsity. For this reason, the conventional CS recovery algorithms do not perform well for such signals.

The proposed algorithm estimates the signal iteratively by performing three steps at each iteration. First, the outer LS makes an estimation of the subspace in which the original signal lies. This is done by minimizing the squared norm of the compressed estimation error. Next, the unknown frequencies are estimated using the root-MUSIC algorithm. Then, the amplitudes of the signal elements are estimated by the inner LS, and the result is fed back to the outer LS for the next iteration.

The CRB for the given signal model has been also derived. Finally, the simulation results have been presented, and it has been shown that the proposed algorithm is able to converge after 5 iterations for the given settings and it approaches the CRB at high signal to noise ratio values.

6.2 Probable future research

In Chapter 3, an improved subspace based DOA estimation method was introduced. Our main focus was on the MUSIC algorithm. One possible route for future research is to study other subspace based DOA estimation methods such as the ESPRIT [28] and Min-Norm [67] algorithms at the threshold region. Then, the application of the proposed methods to these algorithms can be investigated. Furthermore, we considered uniform linear arrays. There are more sophisticated array configurations such as phased-array radars and systems that employ multiple transmit antennas. In the phased-array radars, scaled versions of a single waveform are transmitted from an array of antenna elements, which allows quick scanning of space using digital beamforming [68]. In multiple-input multiple-output (MIMO) radars, more degrees of freedom are provided by transmitting different waveforms via the antenna elements [69]. In the case of orthogonal waveforms, a virtual radar with more antenna elements can be formed, which improves the angular resolution of the radar.

A special form of the MIMO radar is the phased-MIMO radar which gives a tradeoff between phased-array and MIMO radars [70]. Extension of the proposed methods to these systems can also be an interesting research topic to be studied.

Spectrum estimation from compressive measurements was studied in Chapter 5. The elements of the measurement matrix were assumed to be drawn from a Gaussian distribution, which is a common assumption in the CS theory. Another possibility for future research is to consider sparse measurement matrices consisting of only zeros and ones. Such a measurement matrix can represent measurements from a sparse network with sensors randomly placed in a field. Then, the introduced method will need to be modified to handle the new system model.

Bibliography

[1] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice Hall, 2005.

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

[3] G. Carter, "Time delay estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, no. 3, pp. 461–462, Jun. 1981.

[4] F. C. Schweppe, "Sensor-array data processing for multiple-signal sources," *IEEE Trans. Inform. Theory*, vol. IT-14, no. 2, pp. 294–305, Mar. 1968.

[5] G. Bienvenu and L. Kopp, "Adaptivity to background noise spatial coherence for high resolution passive methods," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 1980)*, Denver, CO, Apr. 1980, pp. 307–310.

[6] R. O. Schmidt, *A signal subspace approach to multiple emitter location and spectral estimation*. Ph.D. thesis, Stanford Univ., Stanford, CA, 1981.

[7] ——, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[8] B. A. Johnson, Y. I. Abramovich, and X. Mestre, "MUSIC, G-MUSIC, and maximum-likelihood performance breakdown," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3944–3958, Aug. 2008.

[9] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[10] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[11] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[12] M. Sullivan, *Practical Array Processing*. New York: McGraw-Hill, 2009.

[13] M. Pesavento, A. B. Gershman, and M. Haardt, "Unitary root-MUSIC with a real-valued eigendecomposition: A theoretical and experimental performance study," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1306–1314, May 2000.

[14] H. Cramér, *Mathematical Methods of Statistics*. NJ, Princeton University Press: Princeton, 1946.

[15] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 37, 1945.

[16] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

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

[18] P. Stoica, E. G. Larsson, and A. B. Gershman, "The stochastic CRB for array processing: A textbook derivation," *IEEE Signal Process. Lett.*, vol. 8, no. 5, pp. 148–150, May 2001.

[19] A. B. Gershman and J. F. Böhme, "Improved DOA estimation via pseudorandom resampling of spatial spectrum," *IEEE Signal Process. Lett.*, vol. 4, no. 2, pp. 54–57, Feb. 1997.

[20] A. M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *IEEE Signal Processing Mag.*, vol. 15, no. 1, pp. 56–76, Jan. 1998.

[21] A. M. Zoubir, *Bootstrap Techniques for Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[22] A. B. Gershman and J. F. Böhme, "A pseudo-noise approach to direction finding," *Signal Process.*, vol. 71, no. 1, pp. 1–13, Nov. 1998.

[23] V. Vasylyshyn, "Removing the outliers in root-MUSIC via pseudo-noise resampling and conventional beamformer," *Signal Process.*, vol. 93, no. 12, pp. 3423–3429, Dec. 2013.

[24] C. Qian, L. Huang, and H. C. So, "Improved unitary root-MUSIC for DOA estimation based on pseudo-noise resampling," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 140–144, Feb. 2014.

[25] V. Vasylyshyn, "Improved beamspace ESPRIT-based DOA estimation via pseudo-noise resampling," in *IEEE 9th Eur. Radar Conf.*, Amsterdam, The Netherlands, 2012, pp. 238–241.

[26] A. B. Gershman, "Pseudo-randomly generated estimator banks: A new tool for improving the threshold performance of direction finding," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1351–1364, May 1998.

[27] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 1983)*, Boston, MA, Apr. 1983, pp. 336–339.

[28] R. Roy and T. Kailath, "ESPRIT–estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[29] J. Thomas, L. Scharf, and D. Tufts, "The probability of a subspace swap in the SVD," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 730–736, Mar. 1995.

[30] M. Hawkes, A. Nehorai, and P. Stoica, "Performance breakdown of subspace-based methods: Prediction and cure," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 2001)*, Salt Lake City, UT, May 2001, pp. 4005–4008.

[31] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, Jul. 1988.

[32] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5028, Oct. 2010.

[33] X. Mestre and M. A. Lagunas, "Modified subspace algorithms for DOA estimation with large arrays," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 598–614, Feb. 2008.

[34] M. Shaghaghi and S. A. Vorobyov, "Iterative root-MUSIC algorithm for DOA estimation," in *Proc. 5th Inter. Workshop Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2013)*, The Friendly Island, Saint Martin, Dec. 2013, pp. 53–56.

[35] H. Krim, P. Forster, and J. G. Proakis, "Operator approach to performance analysis of root-MUSIC and root-min-norm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1687–1696, May 1992.

[36] M. Shaghaghi and S. A. Vorobyov, "Subspace leakage analysis and improved DOA estimation with small sample size," *Submitted to IEEE Trans. Signal Process.*, Jul. 2014.

[37] H. L. V. Trees, *Optimum Array Processing: Detection, Estimation, and Modulation Theory, Part IV*. New York: Wiley, 2002.

[38] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*. Boston, MA: McGraw-Hill, 2000.

[39] H. S. Shapiro and R. A. Silverman, "Alias-free sampling of random noise," *J. Soc. Indust. Appl. Math.*, vol. 8, no. 2, pp. 225–248, Jun. 1960.

[40] F. J. Beutler, "Alias free randomly timed sampling of stochastic processes," *IEEE Trans. Inform. Theory*, vol. IT-16, no. 2, pp. 147–152, Mar. 1970.

[41] E. Masry, "Alias free sampling: An alternative conceptualization and its applications," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 3, pp. 317–324, May 1978.

[42] P. P. Vaidyanathan and V. C. Liu, "Efficient reconstruction of band-limited sequences from nonuniformly decimated versions by use of polyphase filter banks," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1927–1936, Nov. 1990.

[43] C. Herley and P. W. Wong, "Minimum rate sampling and reconstruction of signals with arbitrary frequency support," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1555–1564, Jul. 1999.

[44] A. C. Gilbert, M. J. Strauss, and J. A. Tropp, "A tutorial on fast Fourier sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 57–66, Mar. 2008.

[45] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 2007)*, Honolulu, HI, Apr. 2007, pp. 1357–1360.

[46] M. Lexa, M. Davies, J. Thompson, and J. Nikolic, "Compressive power spectral density estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 2011)*, Prague, Czech Republic, May 2011, pp. 3884–3887.

[47] M. A. Lexa, M. E. Davies, and J. S. Thompson, "Compressive and non-compressive power spectral density estimation from periodic nonuniform samples," *[Online] arXiv:1110.2722v2 [cs.IT]*, May 2012.

[48] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 1996)*, Atlanta, GA, May 1996, pp. 1688–1691.

[49] C. P. Yen, Y. Tsai, and X. Wang, "Wideband spectrum sensing based on sub-Nyquist sampling," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3028–3040, Jun. 2013.

[50] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, Sept. 2012.

[51] M. Shaghaghi and S. A. Vorobyov, "Finite-length and asymptotic analysis of averaged correlogram for undersampled data," *Submitted to Appl. Comput. Harmon. Anal.*, Nov. 2013.

[52] ——, "Correlogram for undersampled data: Bias and variance analysis," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 2012)*, Kyoto, Japan, Mar. 2012, pp. 3513–3516.

[53] T. I. Laakso, V. Valimaki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay: Tools for fractional delay filter design," *IEEE Signal Processing Mag.*, vol. 13, no. 1, pp. 30–60, Jan. 1996.

[54] V. Valimaki and T. I. Laakso, "Principles of fractional delay filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 2000)*, Istanbul, Turkey, Jun. 2000, pp. 3870–3873.

[55] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Upper Saddle River, NJ: Prentice Hall, 1993.

[56] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.

[57] W. Wei, *Time series: Univariate and Multivariate Methods, 2nd ed.* New York: Pearson Education, 2006.

[58] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[59] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[60] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2009.

[61] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss, "Near-optimal sparse Fourier representations via sampling," in *Proc. 34th ACM Symp. Theory of Computing*, Montréal, QC, Canada, May 2002, pp. 152–161.

[62] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[63] M. F. Duarte and R. G. Baraniuk, "Recovery of frequency-sparse signals from compressive measurements," in *Proc. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sept. 2010, pp. 599–606.

[64] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[65] M. Shaghaghi and S. A. Vorobyov, "Improved model-based spectral compressive sensing via nested least squares," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 2011)*, Prague, Czech Republic, May 2011, pp. 3904–3907.

[66] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1993, pp. 45–46.

[67] R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, no. 1, pp. 134–139, Jan. 1983.

[68] A. J. Fenn, D. H. Temme, W. P. Delaney, and W. E. Courtney, "The development of phased-array radar technology," *Lincoln Lab. Journal*, vol. 12, no. 2, pp. 321–340, 2000.

[69] E. Fishler, A. M. Haimovich, R. S. Blum, L. Cimini, D. Chizhik, and R. Valenzuela, "MIMO radar: An idea whose time has come," in *Proc. 2004 IEEE Int. Conf. Radar*, Apr. 2004, pp. 71–78.

[70] A. Hassanien and S. A. Vorobyov, "Phased-MIMO radar: a tradeoff between phased-array and MIMO radars," *IEEE Trans. Signal Process.*, vol. 58, pp. 3137–3151, Jun. 2010.