

A PRE-FILTERING MAXIMUM LIKELIHOOD APPROACH TO MULTIPLE SOURCE DIRECTION ESTIMATION

Nopporn Chotikakamthorn

A thesis submitted for the
Degree of Doctor of Philosophy of the University of London
and the
Diploma of Membership of Imperial College

January 1996

Signal Processing Section
Department of Electrical and Electronic Engineering
Imperial College of Science, Technology and Medicine
University of London



Abstract

The problem of narrow-band source direction estimation is considered. A new temporal pre-filtering approach to DOA estimation is proposed. Four types of pre-filters, with varying degrees of prior knowledge of the signal waveforms, are discussed. Maximum likelihood DOA estimators, based on the statistical properties of the filtered data, are formulated. This Pre-Filtering Maximum Likelihood (PFML) framework provides an alternative and unified treatment of an established maximum likelihood method with known signal waveforms and instrumental variable-based DOA estimators. Such methods are limited by the need to have knowledge of the signal waveforms, which may not be possible in practice. A novel Beamforming-based PFML (BPFML) approach is therefore suggested. The BPFML DOA estimators avoid the need to know a priori the signal waveforms by extracting them directly from the observed data. Analytical and simulation studies show that the BPFML methods offer performance comparable with that of the PFML techniques with known signal waveforms.

The Iterative Quadratic Maximum Likelihood (IQML) algorithm is next considered to obtain a maximum likelihood solution with reduced complexity. By employing a new consistent nontriviality constraint, the algorithm performance is improved. In addition, a formulation of the IQML algorithm which incorporates prior information of parameter values is provided. This constrained algorithm is shown to improve the estimation performance as well as the convergence properties of the algorithm. The extension of the IQML algorithm to solve a 2-D parameter estimation problem is additionally considered. The problem of pairing two sets of 1-D parameters from two different dimensions is examined and a sufficient condition for consistent pairing is provided.

Finally, the DOA estimation problem using the IQML algorithm with an interpolated uniform linear array is addressed. A new interpolated array design procedure is proposed. The performance degradation of the Weighted Subspace Fitting (WSF) DOA estimator in the presence of out-of-sector signals is studied. A solution based on one of the BPFML methods is suggested for the case of uncorrelated out-of-sector signals.

Statement of Originality

The main contributions of the thesis are listed below.

1. A novel Pre-Filtering Maximum Likelihood (PFML) approach to narrow-band source direction estimation has been proposed. Four PFML DOA estimators have been formulated, based on the filtered data using different pre-filtering processors. In addition, the proposed framework provides an alternative and unified treatment of some of the existing DOA estimators, as well as leading to the possible improvement of these existing methods.
2. The analytical formula for the asymptotic bias of the PFML estimators based on the type-4 pre-filtering reference signal, has been derived.
3. Beamforming-based PFML (BPFML) methods have been proposed as the realisations of the PFML estimators based on the type-4 pre-filtering reference signal.
4. An improved version of the IQML algorithm has been proposed by using a new consistent nontriviality constraint.
5. Formulation of the IQML algorithm for the case where some of the parameter values are known a priori, has been provided.
6. A sufficient condition for consistent pairing of two 1-D parameters in the 2-D parameter estimation problem has been provided.

7. Based on the proposed 2-D parameter pairing condition, the 1-D based 2-D parameter estimation method using the IQML algorithm has been proposed for the case of uniform-sampled observed data.
8. A new interpolated array design procedure has been proposed. Compared with the existing least squares procedure due to Friedlander, the proposed procedure has two novel features. First, the weighting function is introduced into a design equation to reduce the maximum interpolation error. Secondly, a mild constraint is imposed onto the out-of-sector directivity gain of the interpolated array.
9. Analysis of the estimation bias due to the array interpolation error has been provided.

Additional contributions include the following.

1. The asymptotic analysis of the DEMML estimator, provided by other researchers, has been extended to a more general source scenario.
2. The proposed combined classical and LCMV beamforming procedure, which is part of the BPFML DOA estimators, has been shown to provide adaptive beamforming which is robust to pointing error as well as to the presence of correlated signals.
3. A compact CRLB of the parameter estimates, under the deterministic signal assumption with prior knowledge of some parameter values, has been given.
4. A compact CRLB of the 2-D parameter estimates for arbitrary multiplication of the parameter values in each of the two dimensions, has been derived.

5. The performance of the proposed CIQML algorithm and the 1-D based 2-D parameter estimation method has been studied and compared with some of the existing methods, using computer simulation.
6. Performance study of the PFML and BPFML DOA estimators has been carried out. The analytical result has been validated by the computer simulation result.
7. It has been suggested that, by replacing the estimated data covariance matrix $\hat{\mathbf{R}}$ with the estimated noise covariance matrix $\hat{\hat{\mathbf{R}}}_n$, performance improvement of the instrumental variable-based DOA estimators can be obtained.
8. The performance degradation of DOA estimation using the interpolated array, in the presence of out-of-sector signals, has been addressed. The use of the BPFML methods has been proposed as a solution to the problem, for the case of uncorrelated out-of-sector signals.

Acknowledgements

First, I would like to thank Dr. Jonathon A. Chambers, my supervisor, for his helpful discussion and encouragement. He, along with financial support from a Royal Thai Government scholarship, made possible attendance at two important conferences. In addition, I would like to thank Dr. Athanassios Manikas for useful discussion on beamforming. Special thanks go to Rob Arnott and ERA Technology for providing real data sets and an opportunity to collect field trial data at the company site. Finally, financial support throughout my PhD study by a Royal Thai Government scholarship is very gratefully acknowledged.

To my mother and our family

Contents

Abstract	2
Statement of Originality.....	3
Acknowledgements	6
List of Figures	12
List of Tables.....	19
List of Symbols.....	20
1 Introduction	24
2 Array Signal Processing for Narrow-band Source Direction Estimation...	29
2.1 Concepts.....	29
2.2 Narrow-band signal propagation model.....	30
2.3 Source direction estimation methods	33
2.3.1 A non-parametric approach.....	33
2.3.2 A parametric approach	35
2.4 Discussions.....	45
3 Pre-Filtering Maximum Likelihood Methods for Source Direction Estimation	47
3.1 A pre-filtering approach to narrow-band source DOA estimation.....	48
3.2 Pre-filtering maximum likelihood DOA estimation	53
3.2.1 PFML DOA estimation with the type-1 pre-filtering reference signal	54

3.2.2	PFML DOA estimation with the type-2 pre-filtering reference signal	63
3.2.3	PFML DOA estimation with the type-3 pre-filtering reference signal	67
3.2.4	PFML DOA estimation with the type-4 pre-filtering reference signal	75
3.3	Some practical aspects of the PFML methods	83
3.4	Conclusions.....	85
4	Beamforming-based Pre-Filtering Maximum Likelihood DOA Estimators	87
4.1	Adaptive beamforming techniques for signal waveform estimation	87
4.2	Beamforming-based PFML DOA estimation in the well-separated multiple cluster source scenario.....	91
4.2.1	The multiple cluster source model	91
4.2.2	Beamforming-based PFML DOA estimation.....	94
4.2.3	An improved beamforming-based PFML estimator for correlated cluster sources	98
4.3	Bias correction	100
4.4	Robust adaptive beamformer	102
4.5	Conclusions.....	105
5	The IQML Algorithm for Maximum Likelihood Parameter Estimation ...	106
5.1	Problem formulation	108
5.2	The generalised IQML algorithm and the consistent nontriviality constraint.....	110
5.2.1	The generalised IQML algorithm.....	110

5.2.2	The IQML algorithm with the consistent nontriviality constraint.	113
5.3	The constrained IQML algorithm	118
5.4	Two-dimensional parameter estimation based on the IQML algorithm ...	119
5.5	Numerical results	129
5.6	Conclusions.....	151
6	Performance Study of the PFML DOA Estimators	153
6.1	Simulation and analytical results	154
6.2	Summary of the results and further discussion	193
6.3	Conclusions.....	198
7	Source Direction Estimation with Interpolated Arrays	199
7.1	Introduction.....	199
7.2	The weighted least squares approach to array interpolation	200
7.2.1	The weighted least squares interpolated array design	201
7.2.2	Design example and the analysis of estimation bias	205
7.3	Source direction estimation by the IQML algorithm with the inter- polated uniform linear array	212
7.4	Conclusions.....	218
8	Conclusions and Future Work	220
8.1	Conclusions.....	220
8.2	Future work.....	225
Appendices		228
A	Partitioned complex Wishart matrices	228
B	Proof of Theorem 3.1	230
C	Proof of Theorem 3.2	231
D	Proof of Theorem 3.3	235

E	Proof of Theorem 3.4	237
F	Proof of Theorem 3.5	239
G	Proof of Theorem 5.1	241
H	Proof of Theorem 5.2	245
I	Proof of the 2-D CRLB formula Eq. (5.61)	247
J	Published papers	249
References		250

List of Figures

2.1	Array signal processing concept.....	30
2.2	A uniform linear array of N sensors, with d as an inter-sensor spacing and the array length l . An impinging plane-wave signal $s_i(t)$ comes from the direction θ_i , with the distance across the array $l_{\text{eff}}(\theta_i)$	32
3.1	The pre-filtering approach to narrow-band DOA estimation.....	48
4.1	The three cluster source scenario with the cluster angular widths $\Delta_{c;k}$, $k = 1, \dots, 3$, and the cluster nominal directions of arrival $\theta_{c;k}$, $k = 1, \dots, 3$	93
4.2	The cluster nominal angular location, and the left and right cluster-edge angular locations.	95
4.3	Plots of the CORRs corresponding to the LCMV method, the proposed beamforming method, and the signal received by the 1st sensor, at eight SNRs.	104
5.1	The 2-D 5×5 data matrix	127
5.2	The estimated RMSE_{av} s of the DOA estimates in EXP 5.1 using three IQML algorithms with the DML cost function, and the RMNORM method, compared with the average CRLB on the std at various SNRs with a 1.25 dB step size	133
5.3	The estimated RMSE_{av} s of the DOA estimates in EXP 5.1 using three IQML algorithms with the DML and WSF cost functions, and the RMNORM method, compared with the average CRLB on the std at various SNRs with a 1.25 dB step size	133

5.4	The average number of iterations of three IQML algorithms with the DML and WSF cost functions, at various SNRs with a step size of 1.25 dB	134
5.5	The percentages of convergence of three IQML algorithms with the DML and WSF cost functions, at various SNRs with a step size of 1.25 dB	134
5.6	From EXP 5.2, the estimated RMSE_{av} s of the correlated source DOA estimates using three IQML algorithms with the DML cost function and the RMNORM method, compared with the average CRLB at various SNRs with a 1.25 dB step size	136
5.7	From EXP 5.2, the RMSE_{av} s of the correlated source DOA estimates using three IQML algorithms with the WSF cost function and the RMNORM method, compared with the average CRLB at various SNRs with a 1.25 dB step size	136
5.8	From EXP 5.3, the estimated RMSE_{av} s of the two sinusoidal frequency estimates using three IQML algorithms and MFBLP method, compared with the average CRLB on the std at various SNRs with a 1.25 dB step size	138
5.9	From EXP 5.3, the average number of iterations of three IQML algorithms, at various SNRs with a step size of 1.25 dB	138
5.10	From EXP 5.4, the estimated RMSE_{av} s of two sinusoidal frequency estimates using three IQML algorithms and MFBLP method, at nine different values of the second phase factor α_2 . (a) SNR = 15 dB; (b) SNR = 20 dB	140
5.11	From EXP 5.5, the estimated RMSE_{av} s of two sinusoidal frequency estimates using three IQML algorithms and the MFBLP method, compared with the average CRLB on the std for 20 values of the data length N . (a) SNR = 15 dB; (b) SNR = 20 dB	142

5.12	From EXP 5.6, the estimated RMSE_{av} s of three frequency estimates using two unconstrained IQML algorithms and the MFBLP method, compared with the average CRLB on the std at $N = 25$ and various SNRs with a step size of 1.25 dB	144
5.13	From EXP 5.6, the estimated RMSE_{av} s of two frequency estimates using four unconstrained and constrained IQML algorithms, and the constrained RMNORM method, at $N = 25$ and various SNRs with a step size of 1.25 dB	144
5.14	The average numbers of iterations required by two unconstrained and two constrained IQML algorithms	145
5.15	Estimates of three 2-D parameters by the MEMP method, based on the subdata matrix of size 5×5 at SNR of 20 dB	149
5.16	Estimates of three 2-D parameters by the proposed method with the pairing matrix $\hat{\mathbf{R}}_{\text{av},1}^{\text{fb}}$ at SNR of 20 dB	149
6.1	The estimated RMSE_{av} of two coherent-source DOA estimates obtained by each of five estimators, shown with the RMS value of the CRLBs on the standard deviations of the estimates at eight SNRs.....	159
6.2	The estimated RMSEs of two coherent-source DOA estimates obtained by the PFML-2 and BPFML-2 methods, shown with the CRLBs on the standard deviations of the estimates at eight SNRs	160
6.3	The biases of two coherent-source DOA estimates obtained by the PFML-2, BPFML-2, and BPFML-2+bc methods, at eight SNRs.....	160
6.4	The ensemble average of the array output spatial spectrum, and the asymptotic beam pattern of the beamformer used by the BPFML-2 method in EXP 6.2, at the SNR of 10 dB	164

6.5	The estimated RMSEs of two 1st-cluster DOA estimates obtained by five ML-based methods and the RMNORM method, under the two-cluster four-source scenario of EXP 6.2, shown with the analytical RMSEs of the DEML and PFML-2 estimates at eight SNRs. (a) DOA = 20 degrees; (b) DOA = 15 degrees.....	164
6.6	The RMSEs of two 1st-cluster DOA estimates obtained by the PFML-2, BPFML-2, and BPFML-2a methods, under the two-cluster four-source scenario of EXP 6.2, shown with the analytical RMSEs of the PFML-2 estimates at eight SNRs. (a) DOA = 20 degrees; (b) DOA = 15 degrees .	165
6.7	The estimates of the 1st-cluster angular location, and the corresponding left and right cluster-edge locations, under the two-cluster four-source scenario of EXP 6.2 with the SNR of 30 dB.....	166
6.8	The estimated RMSEs of two 1st-cluster DOA estimates obtained by six estimators under the two-cluster four-source scenario of EXP 6.3, shown with the asymptotic analytical RMSEs due to the DEML and PFML-2 methods at eight SNRs. (a) DOA = 20 degrees; (b) DOA = 15 degrees.....	170
6.9	The ensemble average of the array output spatial spectrum and the asymptotic beam pattern of the beamformer used by the BPFML-2 and BPFML-2+bc methods, under the two-cluster four-source scenario of EXP 6.3 and the SNR of 10 dB.....	171
6.10	The estimated RMSEs of two 1st-cluster DOA estimates obtained by the WSF method and three BPFML methods under the two-cluster four-source scenario of EXP 6.3, shown with the asymptotic analytical RMSEs due to the BPFML-2 method at eight SNRs. (a) DOA = 20 degrees; (b) DOA = 15 degrees.....	171
6.11	The biases of the first and second DOA (20 and 15 degrees) estimates, obtained by the PFML-2 and two BPFML methods, under the two-cluster four-source scenario of EXP 6.3 at eight SNRs	172

6.12	The average numbers of iterations required by five PFML estimators, performed in EXP 6.3 at eight SNRs	173
6.13	The estimated and asymptotic analytical RMSEs of two 1st-cluster DOA estimates obtained by four PFML methods under the signal scenario of EXP 6.4, at ten values of L with the SNR of 5 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.....	174
6.14	The estimated and asymptotic analytical RMSEs of two 1st-cluster DOA estimates obtained by four PFML methods under the signal scenario of EXP 6.4, at ten values of L with the SNR of 10 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.....	175
6.15	The estimated and asymptotic analytical RMSEs of two 1st-cluster DOA estimates obtained by the DEML, PFML-2, BPFML-2, and BPFML-2+bc methods, under the signal scenario of EXP 6.5, with six values of L and the SNR of -5 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.....	178
6.16	The estimated and asymptotic analytical RMSEs of two 1st-cluster DOA estimates obtained by the DEML, PFML-2, BPFML-2, and BPFML-2+bc methods, under the signal scenario of EXP 6.5, with six values of L and the SNR of 0 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.....	179
6.17	The biases of the first and second DOA (20 and 15 degrees) estimates obtained by the PFML-2, BPFML-2, and BPFML-2+bc methods, under the signal scenario of EXP 6.5, with six different values of L . (a) SNR = -5 dB; (b) SNR = 0 dB	180
6.18	The estimated RMSEs of two 1st-cluster DOA estimates obtained by eight ML-based methods, under the signal scenario of EXP 6.6, at nine different values of CORR and the SNR of 5 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees	183

6.19	The estimated RMSEs of two 1st-cluster DOA estimates obtained by eight ML-based methods, under the signal scenario of EXP 6.6, at nine different values of CORR and the SNR of 10 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.....	184
6.20	The ensemble average of the array output spatial spectrum, and two asymptotic beam patterns corresponding to the beamformers used by the BPFML-4 and BPFML-4+bc methods in EXP 6.6 at the SNR of 10 dB..	185
6.21	The estimated RMSEs of ten DOA estimates using the PFML-2 and BPFML-2 methods, under the ten-source scenario of EXP 6.7, compared with the analytical RMSEs of the PFML-2 estimates at eight SNRs. (a) DOAs = -65 and -60 degrees; (b) DOAs = -35 and -30 degrees; (c) DOAs = -5 and 0 degrees; (d) DOAs = 20 and 25 degrees; (e) DOAs = 55 and 60 degrees.....	187
6.22	The estimated RMSEs of two coherent-source DOA estimates obtained by four ML-based methods, under the unknown correlated noise scenario of EXP 6.8, shown with the analytical RMSEs of the PFML-2 estimates at nine different values of ρ (RHO) and the SNR of 5 dB. (a) DOA = 5 degrees; (b) DOA = 10 degrees.....	191
6.23	The noise spatial spectrum at three different values of ρ (RHO).....	192
6.24	The ensemble average of the array output spatial spectrum at three different values of ρ (RHO)	192
6.25	The asymptotic beam pattern of the beamformer used by the BPFML-2 method, at three different values of ρ (RHO).....	193
7.1	Array interpolation error ($e_{i,k}$) corresponding to each of two interpolated array designs, Designs #1 and #2, as a function of the angular location within the ISOV.....	207

7.2	Array interpolation error ($e_{i,k}$) corresponding to each of two interpolated array designs, Designs #2 and #3, as a function of the angular location within the ISOV.....	207
7.3	The estimated and analytical biases of the DOA estimate as a function of the single source direction using two interpolated arrays obtained by Designs #1 and #2	209
7.4	The analytical bias of the DOA estimate as a function of the single source direction using two interpolated arrays obtained by Designs #2 and #3	209
7.5	The biases of the two DOA estimates in EXP 7.2 for various angular locations using two interpolated arrays obtained by Designs #1 and #2. (a) 1st source DOA; (b) 2nd source DOA.....	211
7.6	The analytical biases of two coherent source DOA estimates as a function of the first source direction using two interpolated arrays obtained by Designs #2 and #3	212
7.7	The estimated RMSEs of two coherent source DOA estimates in EXP 7.3 using the optimal WSF cost function with the CIQML algorithm. Two sets of the DOA estimates were obtained based on the real and interpolated arrays	214
7.8	The estimated biases of two coherent source DOA estimates in EXP 7.3 using the optimal WSF cost function with the CIQML algorithm. Two sets of the DOA estimates were obtained based on the real and interpolated arrays	214

List of Tables

3.1	Summary of the applied filters and PFML DOA estimators for the four pre-filtering reference signals.....	83
4.1	The signal directions of arrival (in degrees, compared with broad side) ..	103
5.1	The $RMSE_{av}$ s of the frequency estimates by four estimation methods, together with the $CRLB_{av;1}$ and $CRLB_{av;2}$. (a) The y-dimension estimates; (b) The z-dimension estimates	148
5.2	The $RMSE_{av}$ s of the frequency estimates using four estimation methods, together with the $CRLB_{av;1}$ (on the standard deviation). (a) The y-dimension estimates; (b) The z-dimension estimates.....	151
6.1	The directions of arrival (in degrees) of ten signals, with respect to the array broadside.....	186
7.1	List of 11 source DOA configurations corresponding to two coherent signals in EXP 7.3	213
7.2	The estimated RMSEs (in degrees) of two coherent source DOA estimates in EXP 7.4, obtained by three ML methods at four DOAs of the 3rd uncorrelated signal with the SNR of 20 dB. (a) 1st DOA = -5 degrees; (b) 2nd DOA = 0 degrees.....	217

List of Symbols

Abbreviations

AR	Autoregressive
ARMA	Autoregressive Moving Average
BIQML	Bresler's IQML
BPFML	Beamforming-based Pre-Filtering Maximum Likelihood
CIQML	Consistent IQML
CRLB	Cramér-Rao Lower Bound
DEML	Decoupled Maximum Likelihood
DML	Deterministic Maximum Likelihood
DOA	Direction Of Arrival
FBLP	Forward/Backward Linear Prediction
FFT	Fast Fourier Transform
FOV	Field Of View
IQML	Iterative Quadratic Maximum Likelihood
ISOV	Interpolated Sector Of View
IV	Instrumental Variable
IV-WSF	Instrumental Variable-based Weighted Subspace Fitting
LCMV	Linearly Constrained Minimum Variance
LMMSE	Linear Minimum Mean-Square Error
MA	Moving Average
MEMP	Matrix Enhancement and Matrix Pencil

MFBLP	Modified Forward/Backward Linear Prediction
ML	Maximum Likelihood
MNORM	Minimum Norm
MSE	Mean-Square Error
MSINR	Maximum Signal-to-Interference plus Noise Ratio
MUSIC	Multiple Signal Classification
NIQML	Nagesha's IQML
PDF	Probability Density Function
PFML	Pre-Filtering Maximum Likelihood
PSD	Power Spectral Density
RMS	Root-Mean-Square
RMSE	Root-Mean-Square Error
RMNORM	Root-Minimum Norm
SINR	Signal-to-Interference plus Noise Ratio
SIV	Spatial Instrumental Variable
SML	Stochastic Maximum Likelihood
SNR	Signal-to-Noise Ratio
std	standard deviation
SVD	Singular Value Decomposition
TBWP	Time Bandwidth Product
TIV	Temporal Instrumental Variable
WSF	Weighted Subspace Fitting
1-D	1-Dimensional
2-D	2-Dimensional

Notation

x	a scalar number x
$\mathbf{x}$	a vector $\mathbf{x}$
$\mathbf{X}$	a matrix $\mathbf{X}$
$\text{diag}(\mathbf{x})$	a diagonal matrix with the i th elements of $\mathbf{x}$ as its (i, i) th diagonal components
$[\mathbf{X}]_{i,k}$	the (i, k) th element of a matrix $\mathbf{X}$
$[\mathbf{x}]_i$	the i th element of a column vector $\mathbf{x}$
$\hat{\theta}$	an estimate of the unknown parameter θ
$\bar{\theta}$	a true parameter value of θ
$\mathbb{E}\{\cdot\}$	expectation operation
$(\cdot)^T$	transposition
$(\cdot)^H$	Hermitian transposition
$(\cdot)^*$	complex conjugation
$\mathbf{I}_m$	the $m \times m$ identity matrix
$\tilde{\mathbf{I}}_m$	the $m \times m$ reverse matrix, whose $(i, m - i + 1)$ th elements, $i = 1, \dots, K$, are 1's with 0's elsewhere
$\mathbf{1}_m$	the $m \times 1$ vector with 1's as its elements
$\mathbf{0}_{m \times n}$	the $m \times n$ zero matrix
$\text{vec}(\cdot)$	a vector formed by stacking the column vectors (from left to right) of the matrix inside the bracket
$\text{Re}(\cdot)$	a real part of a complex quantity (scalar, vector, or matrix)
$\text{Im}(\cdot)$	an imaginary part of a complex quantity
$\text{Re}^{-1}(\cdot)$	an inverse of a real part of a complex square matrix
$ \cdot $	determinant
$\odot$	element-wise matrix product

$\otimes$	Kronecker product
$\nabla_{\mathbf{x}}(\cdot)$	the gradient vector, of which its i th element is $\frac{\partial(\cdot)}{\partial x_i}$, where x_i is the i th element of a vector $\mathbf{x}$ (a scalar quantity inside the bracket)
$\nabla_{\mathbf{x}}^2(\cdot)$	the Hessian matrix, of which its (i, k) th element is $\frac{\partial^2(\cdot)}{\partial x_i \partial x_k^*}$, where x_i and x_k are, respectively, the i th and k th elements of a vector $\mathbf{x}$ (a scalar quantity inside the bracket)
$\text{tr}\{\cdot\}$	trace operator
$\ \cdot\ $	square norm (2-norm)
$\ \cdot\ _{\infty}$	infinite norm (∞ -norm)
$O(\cdot)$	O -order (the same order of magnitude as the quantity inside the bracket)
$O_p(\cdot)$	a probabilistic version of O -order
$o(\cdot)$	o -order (small in the order of magnitude, compared with the quantity inside the bracket)
$o_p(\cdot)$	a probabilistic version of o -order
$p(\cdot)$	a probability density function
$\triangleq$	define as

Chapter 1

Introduction

Array signal processing appertains to the processing of a data set, measured by an array of spatially distinct sensors, to extract information related to the signal sources and surrounding environment from which the data set is collected. An early array signal processing technique, called a beamformer, adopts a simple coherent combining scheme to a measured data set. The signals from the outputs of array sensors are weighted, and linearly combined, to extract (implicitly or explicitly) information such as the source locations or the waveform of the desired signal. By suitable choice of the array weights, the beamformer functions as a directive antenna. Unlike a single directive antenna, however, the array beam pattern can be modified by applying different weights to the outputs of the array. This offers flexibility not found in a single directive antenna such as the ability to steer the beam in different directions without the need for physical movement of the antenna. It also allows for adaptation of the array beam pattern to maintain a performance level in a dynamic environment. When the array weights are not fixed, but varied according to some criterion, a beamformer is said to be adaptive.

Probably the first adaptive beamforming system was Van Atta's retrospective antenna system for satellite communication [1]. Following this pioneering work, adaptive beamforming has been extensively studied by many researchers. Major contributions in this area are due to Howell, Applebaum, and Widrow [1-4]. Instead of a simple coherent combining scheme as used in a classical beamformer, the work of these researchers contains more sophisticated beamforming criteria.

In recent years, many sophisticated array processing techniques have been proposed and have received more attention, due partly to the availability of low cost high speed digital signal processors. In the area of source localisation, there has been major interest in so-called high resolution methods [5-8], which are based on the parameter estimation approach. Probably the most well-known parametric-based array processing technique is the MUSIC method, due to Schmidt [8]. It was shown in [9] that the MUSIC method has certain statistical optimality.

Although the one-dimensional-search subspace-based methods[†] such as MUSIC achieve higher resolution than the classical beamforming spectral estimation method, their performance can be poor at low snapshots times Signal-to-Noise Ratio (SNR). Moreover, these methods generally fail in the presence of coherent sources. Pre-processing methods such as the spatial smoothing (for uniform linear arrays) and wide-band focusing techniques [10-12], generally result in sub-optimal performance.

A Maximum Likelihood (ML) method, of which the underlying principle is generally credited to Fisher [13-14], is one of the most widely used techniques for parameter estimation in various application areas. Its use in array signal processing has received attention due to its performance generally being superior to the subspace-based methods with respect to Root-Mean-Square Error (RMSE). Unlike the MUSIC method, the ML method is applicable in the presence of coherent sources. There have been two different versions of the ML method for parameter estimation in array signal processing. One is called the Deterministic Maximum Likelihood (DML) method, whereas the other is the pseudo Stochastic Maximum Likelihood (SML) method. The difference between the two ML methods lies in the way in which the signal waveforms of the emitting sources are modelled.

Recently, a new source Direction-Of-Arrival (DOA) estimator, based on the ML principle, was studied in [15-17]. In these papers, a ML DOA estimator was formulated,

[†] Later, the one-dimensional-search subspace-based methods will be simply referred to as the 'subspace-based' methods.

under the assumption that the waveforms of the signals whose directions are of interest, are known a priori. This ML method with known signal waveforms has many interesting properties. For instance, the maximum number of sources whose directions can be estimated by the method is not limited by the number of sensors when the vectors of the signal waveforms are independent, e.g., sinusoidal signals with distinct frequencies. This is not the case for most other DOA estimators. In addition, the method can be applied in an unknown spatially correlated noise field. Furthermore, the estimator is asymptotically statistically efficient, i.e., its estimation error covariance matrix attains the corresponding Cramér-Rao Lower Bound (CRLB) when the number of data snapshots is large. This is in contrast with the DML method, which is, for a finite number of sensors, asymptotically statistically inefficient [9].

Despite the previously described advantages of the ML method with known signal waveforms over other DOA estimators, the method has a major drawback of requiring knowledge of the signal waveforms. Although such knowledge may be available, for instance, from the training sequence in digital radio communication systems, the duration of the sequence is generally limited to avoid bandwidth inefficiency. Of course, in many other applications such as passive sonar systems, this knowledge may not be available.

In this thesis, a new approach for narrow-band source DOA estimation is proposed. It consists of a pre-processing step performing a temporal filtering operation on the observed array data, followed by a step performing DOA estimation. In Chapter 3, four different types of filters are proposed as temporal processors in the pre-filtering step. Based on the data filtered by each of the four filters, a ML DOA estimator is next formulated. Asymptotic analyses of the formulated Pre-Filtering Maximum Likelihood (PFML) methods are provided. With suitable choices of the filter in the first pre-filtering step and the DOA estimator in the second step, this PFML approach includes the ML method with known signal waveforms and the Instrumental Variable-based (IV-based)

DOA estimators [18-20], as special cases. Using this pre-filtering framework, we provide alternative interpretation of the IV-based DOA estimators and the ML method with known signal waveforms.

In Chapter 4, based on the PFML framework, the Beamforming-based Pre-Filtering Maximum Likelihood (BPFML) DOA estimators are proposed. These estimators share some desirable properties of the ML method with known signal waveforms. Unlike the ML method with known signal waveforms, however, the BPFML methods avoid the need for prior knowledge of the signal waveforms, by extracting the waveforms from the observed data using a newly developed adaptive beamforming technique. This beamforming technique, used as part of the BPFML methods, can be applied both in the coherent source scenario and in the presence of steering error. In either of the two cases, the standard Linearly Constrained Minimum Variance (LCMV) beamformer suffers performance degradation due to the signal cancellation problem [21].

In Chapter 5, the computational aspect of a ML source DOA estimator is considered. Generally, a ML method requires intensive computing power due to the need to perform non-linear multi-dimensional optimisation. For the case of a uniform linear array, Bresler and Macovski [22] proposed a computational less intensive algorithm known as the Iterative Quadratic Maximum Likelihood (IQML) algorithm, to solve the ML exponential signal parameter estimation problem. In spite of computational saving, the IQML algorithm can perform badly for a certain signal parameter configuration, as remarked in [23-24]. In [24], a modified version of the algorithm was proposed by Nagesha and Kay to alleviate the problem. However, both Bresler's and Nagesha's versions of the algorithm do not guarantee consistent parameter estimates. This can result in inefficient parameter estimation when the number of available data snapshots is large. In Chapter 5, an alternative version of the algorithm is proposed. The inconsistency problem which arises in Bresler's and Nagesha's algorithms

is avoided by employing a new nontriviality constraint. As a result, the proposed algorithm outperforms both Bresler's and Nagesha's algorithms. In addition, when some of the parameter values to be estimated are known a priori, the IQML algorithm which incorporates this prior information is formulated. The application of the IQML algorithm to solve a two-dimensional (2-D) parameter estimation problem is also considered. A 2-D parameter estimation method based on the one-dimensional (1-D) IQML algorithm is proposed. The 2-D parameter pairing problem is addressed. A sufficient condition for consistent parameter pairing is provided. The performance of the proposed 2-D parameter estimation method and the improved IQML algorithm, in comparison with some existing methods, is studied by means of computer simulation.

By applying the improved IQML algorithm, developed in Chapter 5, the performance of the PFML and BPFML methods is studied in Chapter 6 through computer simulation. The simulation result is also used to validate the analytical result, which is based on the asymptotic analytical study in Chapters 3 and 4. In Chapter 7, the application of the IQML algorithm to a real irregular uniform linear array is considered. A new interpolated array design procedure, which is a modification of the procedure in [25], is proposed. The out-of-sector-signal problem, which arises when there is an interfering signal arriving from the direction outside the angular region over which the array is interpolated, is addressed. A solution to the problem is provided for the case of uncorrelated out-of-sector interfering signals.

Finally, concluding remarks are given in Chapter 8. Future research topics and remaining problems are also mentioned.

Chapter 2

Array Signal Processing for Narrow-band Source Direction Estimation

2.1 Concepts

The role of an array signal processor is to extract information related to the signal sources and surrounding environment. The array processing concept is visualised in Figure 2.1, where K signal sources emit the (electromagnetic or acoustic) signals which are received by an array of N spatially distinct sensors. The required information contained in the received data snapshots $\mathbf{x}(t) = [x_1(t) \ x_2(t) \ \cdots \ x_N(t)]^T$, $t = 0, 1, \dots, L - 1$, is extracted by a suitable array processing algorithm. The data snapshots $\mathbf{x}(t)$, $t = 0, 1, \dots, L - 1$, are obtained by sampling simultaneously the N sensor outputs. The sampling interval is not necessary uniform. Thus, here ' t ' simply denotes the order of the data snapshots. In this thesis, $\mathbf{x}(t)$ is assumed to be complex.

By using an array of sensors to measure the information-containing signals, more information is available for processing than would be present in a single-channel measurement. This advantage of array processing has been utilised to solve various scientific and engineering problems. Some of its applications are listed [26-27]:

1. Radar/Sonar systems
2. Radio communications
3. Biomedical and hearing aid equipment
4. Seismology

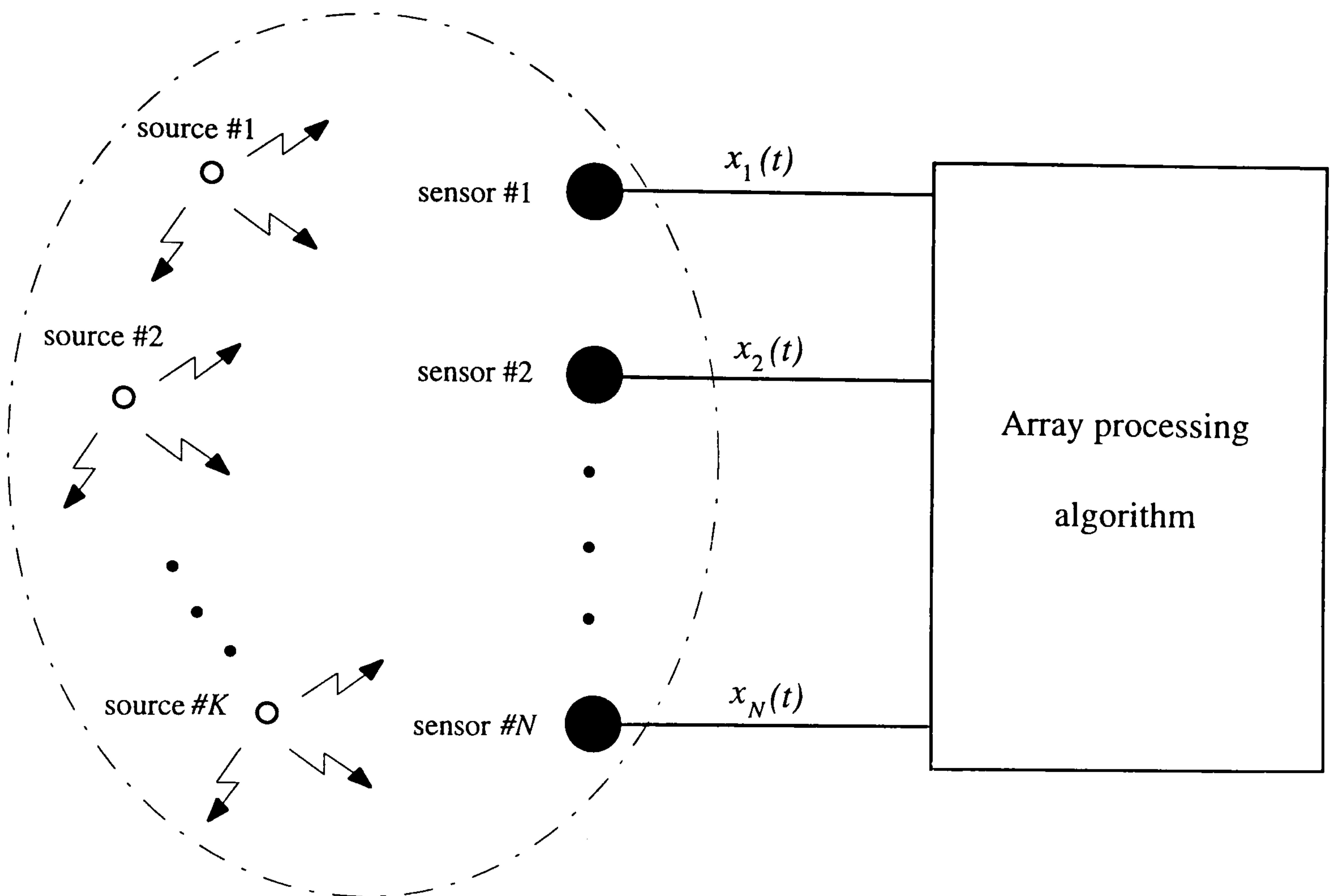


Figure 2.1: Array signal processing concept.

Although from Figure 2.1 no specific information is given relating to the signal sources, the emitted waves, the surrounding environment, and the sensing device characteristics, some prior knowledge of these components is essential in processing and interpreting the observed data. In the next section, we will restrict our interest to a particular source, environment, and array configuration.

2.2 Narrow-band signal propagation model

In this thesis, the following assumptions on the signal sources, the propagation media, and the array of sensors, are implied.

- A. 2.1 The propagation media is linear, homogeneous and, within the signal bandwidth of interest, non-dispersive. The wave propagation speed c is assumed to be known.

A. 2.2 All signals whose directions are of interest are emitted from point sources, located in the far-field. The signals are also assumed to be narrow-band with common frequency bandwidth. Generally, for a signal to be considered narrow-band, the signal bandwidth should be small relative to its centre frequency. In array processing, it is required that the time a signal takes to travel across the array, is short relative to the reciprocal of its bandwidth. This is equivalent to the requirement that the observation Time Bandwidth Product (TBWP) defined by [28]

$$\text{TBWP}(\theta) \triangleq \frac{l_{\text{eff}}(\theta)\Delta_B}{c} \quad (2.1)$$

is much less than 1 for any signal direction θ within the array Field Of View (FOV). Here $l_{\text{eff}}(\theta)$ is the distance across the array of a signal under consideration, which arrives from the direction θ (see Figure 2.2 for the case of a uniform linear array), and Δ_B is the signal bandwidth.

- A. 2.3 In addition to the narrow-band signals from far-field point sources, the additive noise component at each sensor output is assumed to be uncorrelated with the point-source signals.
- A. 2.4 Within the FOV, the array is assumed to be unambiguous in the sense that any $K < N$ array response vectors (to be discussed) which correspond to K distinct source DOAs, span a unique K -dimensional subspace.
- A. 2.5 The array manifold, defined as the set of the array response vectors corresponding to all possible DOAs within the FOV, is assumed known.
- A. 2.6 The array response vector $\mathbf{a}(\theta)$ is assumed to be 1st- and 2nd-order differentiable, with respect to the DOA θ .

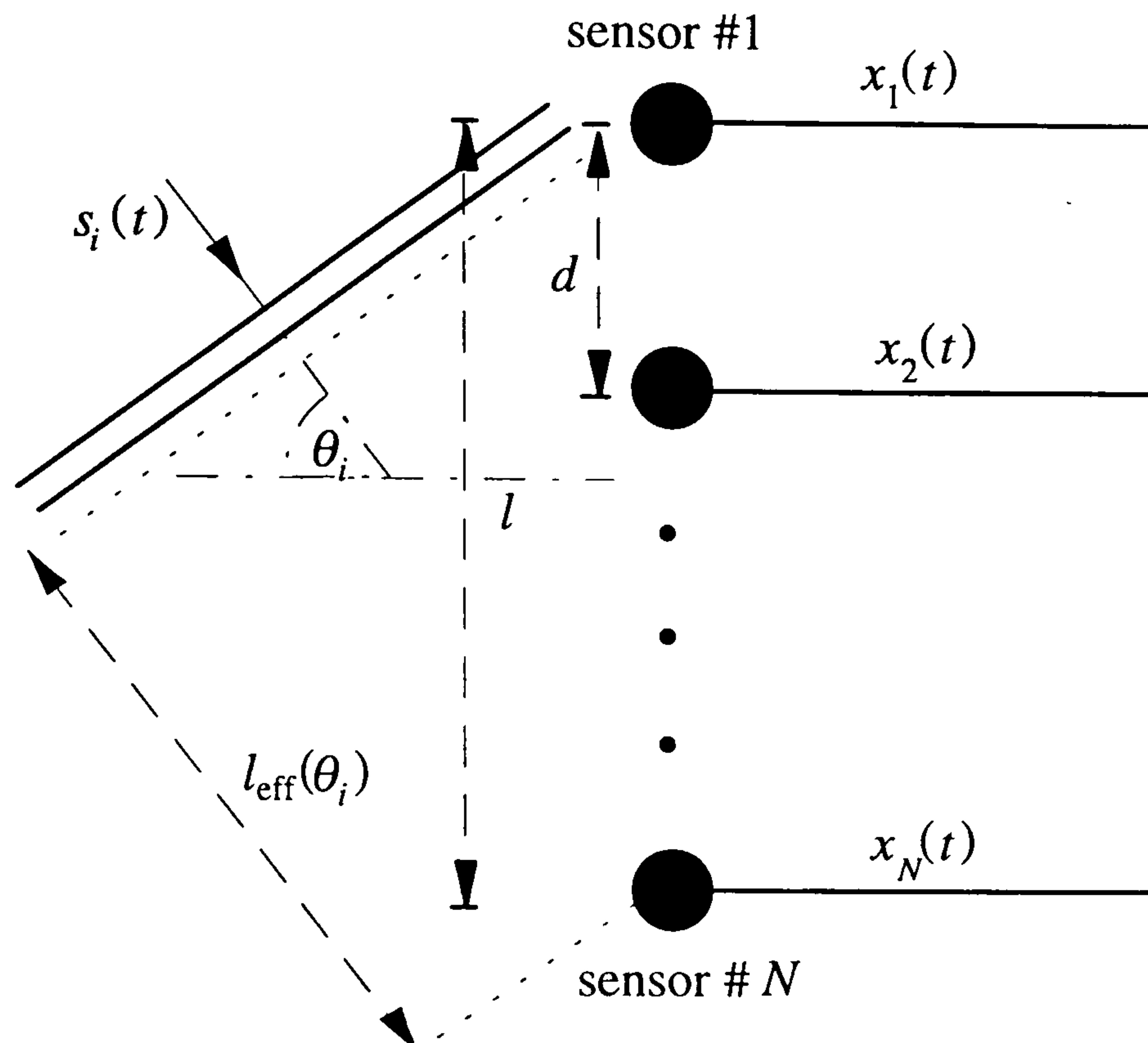


Figure 2.2: A uniform linear array of N sensors, with d as an inter-sensor spacing and the array length l . An impinging plane-wave signal $s_i(t)$ comes from the direction θ_i , with the distance across the array $l_{\text{eff}}(\theta_i)$.

Under the conditions as stated, with the number of sources K , the array data vector $\mathbf{x}(t)$ is written as

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t) \quad (2.2)$$

where $\mathbf{s}(t) = [s_1(t) \ s_2(t) \ \cdots \ s_K(t)]^T$ is composed of K signal waveforms, and $\mathbf{n}(t)$ is the additive noise vector. The response matrix $\mathbf{A}(\boldsymbol{\theta})$ is given by

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}_1(\theta_1) \ \mathbf{a}_2(\theta_2) \ \cdots \ \mathbf{a}_K(\theta_K)] \quad (2.3)$$

where the parameter vector

$$\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \cdots \ \theta_K]^T \quad (2.4)$$

and each of the response vectors $\mathbf{a}_k(\theta_k)$, $k = 1, \dots, K$, is a function of the k th source DOA θ_k . For the uniform linear array which consists of identical omni-directional sensors, with the inter-sensor spacing d , $\mathbf{a}_k(\theta_k)$ has the analytical form given by

$$\mathbf{a}_k(\theta_k) = \alpha_k \begin{bmatrix} 1 & e^{-j\frac{2d}{\lambda}\pi\sin(\theta_k)} & e^{-j\frac{2d}{\lambda}2\pi\sin(\theta_k)} & e^{-j\frac{2d}{\lambda}(N-1)\pi\sin(\theta_k)} \end{bmatrix}^T \quad (2.5)$$

where λ is the signal wavelength at the centre frequency, and α_k is the phase factor whose value depends on a chosen phase reference point. If the centre of the uniform linear array is chosen as the reference point,

$$\alpha_k = e^{j\frac{2d}{\lambda}\pi\sin(\theta_k)(N+1)/2} \quad (2.6)$$

2.3 Source direction estimation methods

One of the main uses of array processing is to estimate a source location. When source directions are of interest, the problem is referred to as the DOA estimation problem. In this thesis, we are mainly concentrated with the use of array signal processing for this purpose, although its uses for signal waveform estimation, and 1-D and 2-D frequency estimation are also considered, respectively, in Chapters 4 and 5. In the following, we review some of the main existing methods for DOA estimation.

Thus far source direction estimation methods can be categorised into two different approaches as follows.

2.3.1 A non-parametric approach

A 'non-parametric' DOA estimator is the estimator which does not rely upon the knowledge of the signal propagation model Eq. (2.2), and the structure of $\mathbf{A}(\theta)$ as described in Eq. (2.3). Two important non-parametric methods are described below.

1) *Classical beamforming method* [29]

This method obtains the K DOA estimates as $\hat{\theta}_k$, $k = 1, \dots, K$, which correspond to the K highest peaks in the beamforming spatial spectrum

$$P_B(\theta) = \mathbf{w}^H(\theta)\hat{\mathbf{R}}\mathbf{w}(\theta) \quad (2.7)$$

where the sample data covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{x}(t) \mathbf{x}^H(t) \quad (2.8)$$

and $\mathbf{w}(\theta)$ is the beamforming weight vector, which should be designed to pass a signal arriving from a look direction θ , but to minimise signals from other directions. In designing $\mathbf{w}(\theta)$, there is a trade-off between resolution and bias properties of the estimator. Note that $\mathbf{w}(\theta)$ in this classical beamforming method is data *independent*, which is in contrast with the Capon DOA estimator, to be described next.

2) Capon DOA estimator [7]

Generally, this DOA estimator is not regarded as a non-parametric method, nor can it be fully considered as a parametric-based method, to be discussed. Here, we include this method within the same class as the classical beamforming method because, as in the classical beamforming method, source directions are indirectly estimated using the beamforming spatial spectrum Eq. (2.7). Unlike the classical beamforming method, however, the weight vector $\mathbf{w}(\theta)$ in the Capon estimator is data *dependent*. It is obtained as

$$\mathbf{w}(\theta) = \hat{\mathbf{R}}^{-1} \mathbf{a}(\theta) (\mathbf{a}^H(\theta) \hat{\mathbf{R}}^{-1} \mathbf{a}(\theta))^{-1} \quad (2.9)$$

which is in fact the LCMV beamforming weight vector, with a unity gain constraint in the look direction θ . By substituting Eq. (2.9) into Eq. (2.7), we arrive at the Capon spatial spectrum

$$P_C(\theta) = \frac{1}{\mathbf{a}^H(\theta) \hat{\mathbf{R}}^{-1} \mathbf{a}(\theta)} \quad (2.10)$$

It should be mentioned that this DOA estimator has also been referred to as the ML method. This misuse of terminology is caused by the relation between the waveform estimate of the signal from a look direction by the LCMV beamformer, and that

obtained by a method based on the ML principle. Under the assumption that the signal from a look direction is uncorrelated with other signals including noise, and that the exact ensemble averaged data covariance matrix is known, the output of the LCMV beamformer is the ML estimate of the signal from the look direction for a Gaussian process. As this relation is connected with a signal waveform, not a signal direction, the DOA estimates by the Capon estimator are generally not the same as those obtained by parametric-based ML DOA estimators, to be described later.

2.3.2 A parametric approach

The source DOA estimation methods which lie in this class are based on the knowledge of the data model Eq. (2.2), and the structure of $A(\theta)$ as described in Eq. (2.3). Due to the ability of the parametric-based methods to resolve two closely angularly spaced sources which are unresolvable by the classical beamforming method, they have also been referred to as high-resolution or super-resolution methods. Important parametric-based DOA estimation methods are described:

1) *Subspace-based methods*

The main principle of this class of DOA estimation methods lies in the decomposition of the vector space, spanned by the eigenvectors of the data covariance matrix, into two orthogonal subspaces. The first one is associated with the signals and the second one is a noise subspace. The first subspace-based parameter estimator is the Pisarenko harmonic retrieval method [5]. Its improved version called the MUSIC (Multiple Signal Classification) algorithm [8] is probably the most well-known high-resolution DOA estimation method. In the following we consider these two methods and some important variants.

a) Pisarenko and MUSIC methods

A subspace approach to parameter estimation originated in the work of V. F. Pisarenko on harmonic retrieval in 1973 [5]. Due to a mathematical equivalence, the method, originally proposed for exponential signal parameter estimation, can also be applied to estimate source directions. Following [5], other methods which are based on the same subspace decomposition principle, including the well-known MUSIC method, were proposed in 1970s and 1980s. In the following, we consider first the basic principle underlying Pisarenko's and other subspace-based methods. Next, the MUSIC method will be described.

From the data model Eq. (2.2), with the number of sensors $N = K + 1$, and the spatial white noise assumption, the data covariance matrix $\mathbf{R} = \mathbb{E}\{\mathbf{x}(t)\mathbf{x}^H(t)\}$ can be written as

$$\mathbf{R} = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \sigma^2\mathbf{I}_{K+1} \quad (2.11)$$

where the signal covariance matrix $\mathbf{R}_s = \mathbb{E}\{\mathbf{s}(t)\mathbf{s}^H(t)\}$ and σ^2 is the additive noise power at each sensor. Note that, the explicit dependence of $\mathbf{A}$ on θ is not shown in Eq. (2.11) for notational convenience. If $\mathbf{R}_s$ is of full rank K (which is the case for non-coherent sources) and that $\mathbf{A}$ has full rank K , the first matrix term in Eq. (2.11) has its column vectors spanning a K -dimensional subspace. By eigendecomposition of $\mathbf{R}$, it can be shown that the K eigenvectors which correspond to the K largest eigenvalues, span the same subspace as the column space of $\mathbf{A}$. The source DOAs can be obtained by finding $\hat{\theta}_k$, $k = 1, \dots, K$, which correspond to the K highest peaks in the following pseudospectrum

$$P_p(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{u}_{K+1}\mathbf{u}_{K+1}^H\mathbf{a}(\theta)} \quad (2.12)$$

where $\mathbf{u}_{K+1}$ is the eigenvector of $\mathbf{R}$ which corresponds to the smallest eigenvalue. This vector is commonly referred to as the noise-subspace vector, although the correct terminology for $\mathbf{u}_{K+1}$ should be the 'noise-only subspace vector'.

The Pisarenko method has one major advantage over the classical beamforming method. Unlike the classical beamformer, whose resolution capability is limited to the array beamwidth[†], the Pisarenko method is capable of estimating the DOAs of two (or more) sources whose angular spacing is smaller than one array beamwidth. Despite this high-resolution capability, the Pisarenko method yields poor result at low L .SNR. In [8], the MUSIC method, based on the same subspace decomposition principle, was proposed. It offers performance improvement over the Pisarenko method by increasing the data dimension N beyond $K + 1$. In doing so, the eigendecomposition of the data covariance matrix $\mathbf{R}$ results in more than one noise-subspace vector. In [8], Schmidt suggested to use all of the noise-subspace vectors, which are the eigenvectors of $\mathbf{R}$ corresponding to the $N - K$ smallest eigenvalues. The DOA estimates are obtained as $\hat{\theta}_k$, $k = 1, \dots, K$, which correspond to the K highest peaks in the following pseudospectrum

$$P_m(\theta) = \frac{1}{\mathbf{a}^H(\theta) \mathbf{U}_n \mathbf{U}_n^H \mathbf{a}(\theta)} \quad (2.13)$$

where $\mathbf{U}_n$ is the $N \times (N - K)$ matrix which is composed of the $N - K$ noise-subspace vectors. This peak-picking MUSIC method, which is sometimes referred to as the Spectral-MUSIC method, has an important property of being applicable to an array with arbitrary geometry. For a uniform linear array, however, the so-called Root-MUSIC method offers improved performance [30-31]. Instead of performing a 1-D search over the pseudospectrum Eq. (2.13), the Root-MUSIC method obtains the DOA estimates by

[†] For the uniform linear array with half signal wavelength inter-sensor spacing, two sources are said to be angularly spaced by one array beamwidth if $|\sin(\theta_1) - \sin(\theta_2)| = 2/N$, where θ_1 and θ_2 are their DOAs and N is the number of sensors.

choosing K (double) roots which lie closest to the unit circle from the $2(N - 1)$ roots of the $2(N - 1)$ -order polynomial $\mathbf{a}^H(z)\mathbf{U}_n\mathbf{U}_n^H\mathbf{a}(z)$, where $\mathbf{a}(z) = [1 \ z \ \cdots \ z^{N-1}]^T$. By mapping the resulting K distinct roots $z_k^\dagger$, $k = 1, \dots, K$, onto the unit circle, the DOA estimates can be obtained from the (mapped) roots $\tilde{z}_k (= z_k / |z_k|)$ by using the relation

$$\tilde{z}_k = e^{-j\frac{2d}{\lambda}\pi\sin(\theta_k)} \quad (2.14)$$

The discussion so far is based on the assumption that the (theoretical) data covariance matrix of the array $\mathbf{R}$ is known. In practice, where $\mathbf{R}$ is generally not known, its consistent estimate such as the sample data covariance matrix $\hat{\mathbf{R}}$, given by Eq. (2.8), can be used instead. In addition, when the additive noise is not spatially white, both the Pisarenko and MUSIC methods are still applicable, provided that the noise spatial covariance structure $\mathbf{R}_{\text{ns}}$, which is a scaled version of the noise covariance matrix, is known. In this case, the eigendecomposition is performed instead on the pre-whitened covariance matrix $\mathbf{R}_{\text{ns}}^{-1/2}\mathbf{R}\mathbf{R}_{\text{ns}}^{-1/2}$ (or $\mathbf{R}_{\text{ns}}^{-1/2}\hat{\mathbf{R}}\mathbf{R}_{\text{ns}}^{-1/2}$, when $\hat{\mathbf{R}}$ is used in place of $\mathbf{R}$).

b) Minimum Norm (MNORM) method

For a uniform linear array, the performance of the MUSIC method can be improved by replacing $\mathbf{U}_n$ in the MUSIC method by the minimum-norm vector

$$\mathbf{u}_{\text{mn}} = (\tilde{\mathbf{u}}_1^H \tilde{\mathbf{u}}_1)^{-1} \mathbf{U}_n \tilde{\mathbf{u}}_1 \quad (2.15)$$

where $\tilde{\mathbf{u}}_1$ is the first column vector of $\mathbf{U}_n^H$.

By applying $\mathbf{u}_{\text{mn}}$ instead of $\mathbf{U}_n$ to the Root-MUSIC method, the spurious roots tend to be uniformly distributed around the unit circle in the region away from the K roots which correspond to the signal directions. This helps differentiating spurious roots, due to noise, from those due to the signals of interest. A peak-picking version of the

[†] When the exact data covariance matrix $\mathbf{R}$ is used, the K roots lie *on* the unit circle. This is generally not the case if $\mathbf{R}$ is replaced by a consistent estimate.

minimum norm method can also be similarly obtained by substituting U_n in Eq. (2.13) by u_{mn} .

c) Modified Forward/Backward Linear Prediction (MFBLP) method

The linear prediction method for DOA estimation is known to give poor result at low SNR [32]. This is due to incorrect modelling of the received sensor outputs $x_i(t)$, $i = 1, \dots, N$, as being generated from the AR process. A correct model for $x_i(t)$ should be the special ARMA model where the AR and MA parameters are the same (see, for example, Chapter 3 of [7]). In [32], a modification to the Forward/Backward Linear Prediction (FBLP) method [7] was proposed to alleviate this problem. The performance of the FBLP method is enhanced by increasing the prediction order beyond K , and improving the effective SNR by eigendecomposition of the prediction matrix. The MFBLP method can be explained as follows.

First, consider the FBLP method. The M prediction filter coefficients are obtained by solving the following problem

$$(\mathbf{R}_{LP} + \tilde{\mathbf{R}}_{LP})\mathbf{g} = -(\mathbf{r}_{LP} + \tilde{\mathbf{r}}_{LP}) \quad (2.16)$$

where

$$\mathbf{R}_{LP} = \frac{1}{N-M} \sum_{i=2}^{N-M+1} \mathbf{E}\{ \mathbf{x}_i(t) \mathbf{x}_i^H(t) \} \quad (2.17)$$

$$\tilde{\mathbf{R}}_{LP} = \frac{1}{N-M} \sum_{i=1}^{N-M} \mathbf{E}\{ \tilde{\mathbf{I}}_M \mathbf{x}_i^*(t) \mathbf{x}_i^T(t) \tilde{\mathbf{I}}_M \} \quad (2.18)$$

$$\mathbf{r}_{LP} = \frac{1}{N-M} \sum_{i=1}^{N-M} \mathbf{E}\{ \mathbf{x}_{i+1}(t) \mathbf{x}_i^*(t) \} \quad (2.19)$$

$$\tilde{\mathbf{r}}_{LP} = \frac{1}{N-M} \sum_{i=1}^{N-M} \mathbf{E}\{ \tilde{\mathbf{I}}_M \mathbf{x}_i^*(t) \mathbf{x}_{i+M}(t) \} \quad (2.20)$$

$$\mathbf{g} = [g_1 \ g_2 \ \cdots \ g_M]^T \quad (2.21)$$

$$\mathbf{x}_i(t) = [x_i(t) \ x_{i+1}(t) \ \cdots \ x_{i+M-1}(t)]^T, \ i \geq 1, \ M \geq K \quad (2.22)$$

When the knowledge of $\mathbf{R}_{\text{LP}}$, $\tilde{\mathbf{R}}_{\text{LP}}$, $\mathbf{r}_{\text{LP}}$, and $\tilde{\mathbf{r}}_{\text{LP}}$ is not available, the estimate $\hat{\mathbf{g}}$ of $\mathbf{g}$ is obtained as a least squares solution of Eq. (2.16) with $\mathbf{R}_{\text{LP}}$, $\tilde{\mathbf{R}}_{\text{LP}}$, $\mathbf{r}_{\text{LP}}$, and $\tilde{\mathbf{r}}_{\text{LP}}$ being replaced by their consistent estimates. The source directions can be estimated from the roots of the prediction polynomial, formed from the estimate $\begin{bmatrix} 1 & \hat{\mathbf{g}}^T \end{bmatrix}^T$ of the true prediction vector $\begin{bmatrix} 1 & \mathbf{g}^T \end{bmatrix}^T$, by replacing $\mathbf{U}_n$ in the Root-MUSIC method with $\begin{bmatrix} 1 & \hat{\mathbf{g}}^T \end{bmatrix}^T$. Note that, in this case, the number of sensors N in the Root-MUSIC method is equal to $M + 1$. If the prediction order M is chosen as K , there are exactly K roots which correspond to the directions of the K signals. In practice, M is generally chosen to be higher than K to improve the estimation performance. Further improvement can be gained by replacing the estimate $\hat{\mathbf{R}}_{\text{FB}} = \hat{\mathbf{R}}_{\text{LP}} + \hat{\tilde{\mathbf{R}}}_{\text{LP}}$ by $\hat{\mathbf{R}}_{\text{MFB}}$ which is formed from the K eigenvectors of $\hat{\mathbf{R}}_{\text{FB}}$ corresponding to the K largest eigenvalues [32]. By doing so, the bias due to the additive noise and incorrect modelling can be reduced, resulting in performance improvement.

2) *Maximum likelihood methods*

The ML approach for parameter estimation has been applied in diverse subjects such as economics, forecasting, and engineering. Its use in array signal processing has gained research interest due to its superior performance in the RMSE sense to the subspace-based methods. Thus far, there have been three different DOA estimators which are based on the ML principle. The difference between these methods lies in the assumed prior knowledge of the data model Eq. (2.2). These ML methods, namely the deterministic maximum likelihood (DML) method, stochastic maximum likelihood (SML) method, and the ML method with known signal waveforms, are described as follows.

a) *Deterministic Maximum Likelihood (DML) method [22]*

In this ML formulation, the following assumptions are applied:

1. The additive noise $\mathbf{n}(t)$ is circular complex zero-mean multivariate Gaussian distributed with known noise covariance structure $\mathbf{R}_{\text{ns}} = \sigma^{-2} \mathbf{R}_n$, where σ^2 is an unknown scaling factor. In addition, $\mathbf{n}(t)$ is temporally uncorrelated, i.e., $\mathbb{E}\{\mathbf{n}(t_1)\mathbf{n}^H(t_2)\} = \mathbf{0}$, $t_1 \neq t_2$.
2. The signal waveforms $s_k(t)$, $k = 1, \dots, K$, are modelled as unknown deterministic parameters.

Under the above assumptions, the likelihood function of the unknown parameters, given L array data snapshots and the number of signal sources K , is [†]

$$\mathcal{L}(\mathbf{x}(0), \dots, \mathbf{x}(L-1); \theta, \sigma^2, s(0), \dots, s(L-1)) = \prod_{t=0}^{L-1} \frac{1}{|\pi\sigma^2 \mathbf{R}_{\text{ns}}|} \exp\{-\sigma^{-2}\{(\mathbf{x}(t) - \mathbf{A}s(t))^H \mathbf{R}_{\text{ns}}^{-1}(\mathbf{x}(t) - \mathbf{A}s(t))\}\} \quad (2.23)$$

The ML DOA estimates $\hat{\theta}_k$, $k = 1, \dots, K$, can be obtained by minimising, with respect to θ , the following concentrated cost function [22]

$$f_{\text{DML}}(\theta) = \text{tr}\{\mathbf{P}_{\mathbf{R}_{\text{ns}}^{-1/2}\mathbf{A}}^\perp \mathbf{R}_{\text{ns}}^{-1/2} \hat{\mathbf{R}} \mathbf{R}_{\text{ns}}^{-1/2}\} \quad (2.24)$$

where the null projection matrix $\mathbf{P}_{\mathbf{R}_{\text{ns}}^{-1/2}\mathbf{A}}^\perp = \mathbf{I}_N - \mathbf{R}_{\text{ns}}^{-1/2} \mathbf{A}(\mathbf{A}^H \mathbf{R}_{\text{ns}}^{-1} \mathbf{A})^{-1} \mathbf{A}^H \mathbf{R}_{\text{ns}}^{-1/2}$, and $\hat{\mathbf{R}}$ is the sample data covariance matrix as given in Eq. (2.8). Note that the explicit expression of $\mathbf{A}$ as a function of θ is dropped for notational convenience.

b) Pseudo Stochastic Maximum Likelihood (SML) method and WSF framework

The second formulation of the ML DOA estimator assumes that the signal vector $\mathbf{s}(t)$ is circular complex zero-mean multivariate Gaussian distributed and is uncorrelated with the additive noise $\mathbf{n}(t)$. Other assumptions are the same as those of the DML method. The estimates of the source directions are obtained by solving the following optimisation problem [33-34]:

[†] With some abuse of notation, $\mathbf{x}(t)$ appeared in Eq. (2.23) is a realisation of a random vector.

$$\hat{\theta} = \arg \min_{\theta \in \Theta^{K \times 1}} f_{\text{SML}}(\theta) \quad (2.25)$$

where $\Theta^{K \times 1}$ is the set of feasible parameter values, and the concentrated cost function

$$f_{\text{SML}}(\theta) = \ln |A \tilde{R}_s(\theta) A^H + \tilde{\sigma}^2(\theta) R_{\text{ns}}| \quad (2.26)$$

$$\tilde{R}_s(\theta) = (A^H R_{\text{ns}}^{-1} A)^{-1} A^H R_{\text{ns}}^{-1} (\hat{R} - \tilde{\sigma}^2(\theta) R_{\text{ns}}) R_{\text{ns}}^{-1} A (A^H R_{\text{ns}}^{-1} A)^{-1} \quad (2.27)$$

$$\tilde{\sigma}^2(\theta) = \frac{1}{N-K} \text{tr} \{ P_{R_{\text{ns}}^{-1/2} A}^\perp R_{\text{ns}}^{-1/2} \hat{R} R_{\text{ns}}^{-1/2} \} \quad (2.28)$$

To obtain $\hat{\theta}$ from Eq. (2.25), high computational complexity is generally required. In [33-34], an approximate approach called the Weighted Subspace Fitting (WSF) framework was proposed. The WSF method obtains $\hat{\theta}$ from

$$\hat{\theta} = \arg \min_{\theta \in \Theta^{K \times 1}} f_{\text{wsf}}(\theta) \quad (2.29)$$

where

$$f_{\text{wsf}}(\theta) = \text{tr}(P_{R_{\text{ns}}^{-1/2} A}^\perp \hat{U}_s W_{\text{wsf}} \hat{U}_s^H) \quad (2.30)$$

The matrix $\hat{U}_s$ is composed of the d_s eigenvectors of $R_{\text{ns}}^{-1/2} \hat{R} R_{\text{ns}}^{-1/2}$ which correspond to the d_s largest eigenvalues, where d_s is the rank of the signal covariance matrix. The $d_s \times d_s$ weighting matrix W_{wsf} is assumed to be positive definite. It was shown in [33] that the estimate $\hat{\theta}$, obtained from Eq. (2.29), is a large sample SML estimate, if the weighting matrix W_{wsf} is chosen as

$$W_{\text{wsf}} = \tilde{\Lambda}^2 \Lambda_s^{-1} \quad (2.31)$$

where Λ_s is a diagonal matrix whose diagonal elements are the d_s largest eigenvalues of $R_{\text{ns}}^{-1/2} \hat{R} R_{\text{ns}}^{-1/2}$ and $\tilde{\Lambda} = \Lambda_s - \sigma^2 I_{d_s}$. This asymptotic result still holds when W_{wsf} is replaced by the consistent estimate $\hat{W}_{\text{wsf}} = \hat{\tilde{\Lambda}}^2 \hat{\Lambda}_s^{-1}$, where $\hat{\tilde{\Lambda}}$ and $\hat{\Lambda}_s$ are the consistent estimates of $\tilde{\Lambda}$ and Λ_s . Note that a similar result was also derived in [35]. By comparing Eqs. (2.25-2.26) with Eqs. (2.29-2.30), it is seen that solving the WSF problem is a much simpler, although computationally complex, task than finding a solution to the SML problem.

c) Maximum likelihood method with known signal waveforms [15-17]

In some engineering problems such as digital radio communications, the knowledge of the signal waveforms is available. For instance, this can be obtained in a form of the training sequence. Interestingly, the use of the training sequence has long been suggested for other purposes such as channel equalisation and optimal beamforming. The utility of this prior knowledge for DOA estimation has received comparatively less attention.

In [15], Li and Compton studied various formulations of the ML DOA estimators, based on the knowledge of the signal waveforms. In [16-17], a simplified ML method with known signal waveforms was proposed. Many interesting properties of this ML method with known signal waveforms were pointed out in [15-17]. For instance, the number of non-coherent sources can be greater than $N - 1$. In addition, the performance of the method is largely unaffected by the angular spacing between any two non-coherent sources. Furthermore, the method is applicable in an unknown spatially correlated noise field.

In [16-17], it is assumed that $s(t)$ is composed of non-coherent signals and $\tilde{s}(t)$ is known, where

$$s(t) = \Gamma \tilde{s}(t) \quad (2.32)$$

and Γ is the diagonal matrix whose main diagonal elements γ_k , $k = 1, \dots, K$, are the unknown signal amplitudes. Given $\tilde{s}(t)$, the data snapshots $\mathbf{x}(t)$, and the noise covariance structure $\mathbf{R}_{\text{ns}}$, the estimates of the signal directions and amplitudes are obtained by minimising

$$f_{\text{ks}}(\boldsymbol{\theta}, \boldsymbol{\gamma}) \Big|_{\mathbf{R}_{\text{ns}}} = \sum_{t=0}^{L-1} (\mathbf{x}(t) - \mathbf{A}\Gamma\tilde{\mathbf{s}}(t))^H \mathbf{R}_{\text{ns}}^{-1} (\mathbf{x}(t) - \mathbf{A}\Gamma\tilde{\mathbf{s}}(t)) \quad (2.33)$$

where $\boldsymbol{\gamma} = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_K]^T$. When the noise covariance structure is not known, the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ can be obtained by minimising the following cost function

$$f_{\text{KS}}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \left| \frac{1}{L} \sum_{t=0}^{L-1} (\mathbf{x}(t) - \mathbf{A}\boldsymbol{\Gamma}\tilde{\mathbf{s}}(t))(\mathbf{x}(t) - \mathbf{A}\boldsymbol{\Gamma}\tilde{\mathbf{s}}(t))^H \right| \quad (2.34)$$

Minimisation of Eq. (2.34) is a difficult task. A simpler cost function which is asymptotically equivalent to Eq. (2.34) is given by [16-17]

$$f_{\text{KS}}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \text{tr} \{ \hat{\tilde{\mathbf{R}}}_{\text{s}} (\hat{\tilde{\mathbf{R}}}_{\text{xs}} \hat{\tilde{\mathbf{R}}}_{\text{s}}^{-1} - \mathbf{A}\boldsymbol{\Gamma})^H \hat{\tilde{\mathbf{R}}}_{\text{n}}^{-1} (\hat{\tilde{\mathbf{R}}}_{\text{xs}} \hat{\tilde{\mathbf{R}}}_{\text{s}}^{-1} - \mathbf{A}\boldsymbol{\Gamma}) \} \quad (2.35)$$

where

$$\hat{\tilde{\mathbf{R}}}_{\text{s}} = \frac{1}{L} \sum_{t=0}^{L-1} \tilde{\mathbf{s}}(t) \tilde{\mathbf{s}}^H(t) \quad (2.36)$$

$$\hat{\tilde{\mathbf{R}}}_{\text{xs}} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{x}(t) \tilde{\mathbf{s}}^H(t) \quad (2.37)$$

and

$$\hat{\tilde{\mathbf{R}}}_{\text{n}} = \hat{\mathbf{R}} - \hat{\tilde{\mathbf{R}}}_{\text{xs}} \hat{\tilde{\mathbf{R}}}_{\text{s}}^{-1} \hat{\tilde{\mathbf{R}}}_{\text{xs}}^H \quad (2.38)$$

In [16-17], it was shown that the cost function given by Eq. (2.35) is asymptotically equivalent to

$$f_{\text{KS}}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \text{tr} \{ \tilde{\mathbf{R}}_{\text{s}} (\hat{\tilde{\mathbf{R}}}_{\text{xs}} \hat{\tilde{\mathbf{R}}}_{\text{s}}^{-1} - \mathbf{A}\boldsymbol{\Gamma})^H \hat{\tilde{\mathbf{R}}}_{\text{n}}^{-1} (\hat{\tilde{\mathbf{R}}}_{\text{xs}} \hat{\tilde{\mathbf{R}}}_{\text{s}}^{-1} - \mathbf{A}\boldsymbol{\Gamma}) \} \quad (2.39)$$

where $\tilde{\mathbf{R}}_{\text{s}} = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{t=0}^{L-1} \tilde{\mathbf{s}}(t) \tilde{\mathbf{s}}^H(t)$. For the case of uncorrelated signals where $\tilde{\mathbf{R}}_{\text{s}}$ is diagonal, each of the K source directions can be separately estimated by minimising

$$f_{\text{DEML}}(\theta_k, \gamma_k) = (\hat{\tilde{\mathbf{r}}}_{\text{xs};k} - \gamma_k \mathbf{a}_k)^H \hat{\tilde{\mathbf{R}}}_{\text{n}}^{-1} (\hat{\tilde{\mathbf{r}}}_{\text{xs};k} - \gamma_k \mathbf{a}_k) \quad (2.40)$$

where $\hat{\tilde{\mathbf{r}}}_{\text{xs};k}$ is the k th column vector of $\hat{\tilde{\mathbf{R}}}_{\text{xs}} \hat{\tilde{\mathbf{R}}}_{\text{s}}^{-1}$. The estimator based on the cost function Eq. (2.40) was proposed in [16-17] and named the Decoupled Maximum Likelihood (DEML) method. The DEML cost function for the case of known noise covariance structure can be similarly derived and is given by

$$f_{\text{DEML}}(\theta_k, \gamma_k) = (\hat{\tilde{\mathbf{r}}}_{\text{xs};k} - \gamma_k \mathbf{a}_k)^H \mathbf{R}_{\text{ns}}^{-1} (\hat{\tilde{\mathbf{r}}}_{\text{xs};k} - \gamma_k \mathbf{a}_k) \quad (2.41)$$

2.4 Discussions

In comparing the classical beamforming method and the parametric-based methods, the latter methods offer high-resolution capability not found in the classical method. The classical method is, however, generally more robust to modelling error. In addition, the performance of the classical methods does not suffer with a performance threshold effect with reducing SNR. This is in contrast with the parametric-based methods which have a threshold effect, the term used to describe an abrupt estimation performance degradation as the SNR value falls below a certain level.

From analytical results in [31, 36], it is concluded that the spectral and root-finding versions of either the MUSIC or MNORM method have the same asymptotic performance. The asymptotic performance of the class of weighted MUSIC methods, which includes the MNORM method as a special case, was shown to be bounded by that of the (un-weighted) MUSIC method [37]. For uncorrelated signals, it was shown in [9] that the MUSIC method is asymptotically equivalent to the DML method. For partially correlated signals, however, the performance of the MUSIC method is generally inferior to that of the DML method [37]. The performance degradation also arises in the Capon estimator for correlated sources. Both the subspace-based and Capon DOA estimation methods fail when the signals are fully correlated. A spatial smoothing pre-processor (for uniform linear arrays) is required to improve the performance of these estimators in the highly correlated source scenario [10-11]. No such pre-processor is needed, however, for the described ML methods to operate under this scenario. Analytical performance study of the Capon estimator can be found in [38].

In comparing the error covariance matrices of the DML and SML estimates with the corresponding Cramér-Rao lower bounds, the SML error covariance matrix attains the corresponding CRLB for a large number of data snapshots. However, for a finite number of sensors, the DML error covariance matrix never attains the corresponding CRLB. Note that because the WSF method with the optimal weight Eq. (2.31) is a large

sample realisation of the SML method, its asymptotic properties are the same as those of the SML method.

The comparatively new ML method with known signal waveforms, considered in [15-17], has some distinct properties not shared by the above mentioned ML methods. First, the method is able to estimate more non-coherent source directions than the number of sensors, by using only second-order statistics. Secondly, its performance in estimating two (or more) uncorrelated sources with sufficient length of the known signal waveforms, is unaffected by the angular separation between two (or more) sources. In addition, it can be used to estimate source directions in an unknown spatially correlated noise field. It was shown in [16] that, for uncorrelated sources, the DEMML method, which is a simplified ML method with known signal waveforms, is asymptotically statistically efficient, i.e., its error covariance matrix approaches the corresponding CRLB for a large number of data snapshots.

Chapter 3

Pre-Filtering Maximum Likelihood Methods for Source Direction Estimation

In Chapter 2, important DOA estimators, which include ML-based and subspace-based methods, have been considered. However, apart from a choice of a DOA estimator, some pre-processing schemes can sometimes improve the accuracy of DOA estimation, or make the task possible. For instance, the use of spatial averaging allows subspace-based methods to operate in the coherent source scenario [10-11, 39]. Another form of pre-processing is the beamspace pre-processor [40-42], which is based on spatial filtering of the received array data. It has been reported that using the beamspace data for DOA estimation can result in improved performance [40-42]. Although the application of these spatial pre-processors has been widely studied, the use of temporal pre-processors in narrow-band DOA estimation receives less attention. Note that, in the context of wide-band source localisation, spatio-temporal processing was considered in [43] and references therein.

In this chapter, a new approach to narrow-band source DOA estimation is proposed. It consists of a temporal pre-processing step, followed by a step performing DOA estimation. Four different types of filters are then proposed as a pre-processor in the first step. The filter coefficients for each filter type are obtained from a known pre-filtering reference signal, which is required to be highly correlated with the signal $s(t)$. Four different pre-filtering reference signals which may be obtainable in practice are next considered. Based on the statistical properties of the data pre-processed by each of

the four filters, a ML DOA estimator is then formulated. Analysis of the asymptotic properties of the DOA estimates obtained by each estimator is provided. Relation between this class of the Pre-Filtering Maximum Likelihood (PFML) DOA estimators and the ML method with known signal waveforms [15-17], as well as the optimal IV-based DOA estimators in [18-20], will be addressed.

3.1 A pre-filtering approach to narrow-band source DOA estimation

Detail of the temporal pre-filtering approach to narrow-band source direction estimation is considered in this section. Under this framework, knowledge of the signal and noise temporal properties are utilised to improve the accuracy of the DOA estimation. A basic concept of the pre-filtering approach can be explained by first considering the block diagram described in Figure 3.1.

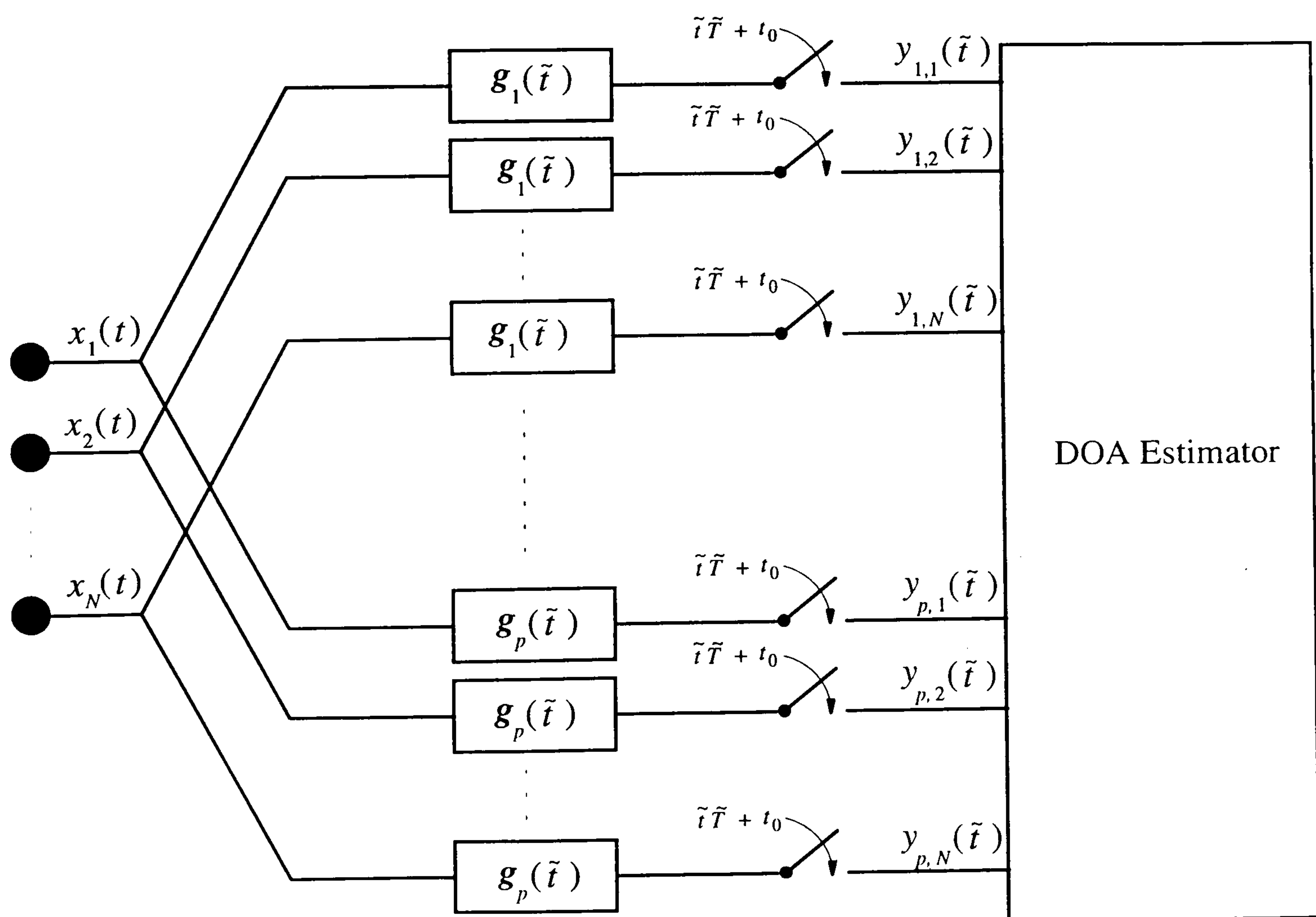


Figure 3.1: The pre-filtering approach to narrow-band DOA estimation.

From the figure, t is the discrete time index at the array output and $\tilde{t}$ is the second, resampled, discrete time index. The data vector $\mathbf{x}(t)$ is passed through p sets of filters, whose coefficients may be varied as a function of $\tilde{t}$. Filtered by p filter sets, the pre-processed signals are subsequently re-sampled simultaneously with the sampling interval $\tilde{T}$, and the timing offset t_0 . The sampled data vectors $\mathbf{y}_k(\tilde{t}) = [y_{k;1}(\tilde{t}) \ y_{k;2}(\tilde{t}) \ \cdots \ y_{k;N}(\tilde{t})]^T, k = 1, \dots, p$, are then applied to a DOA estimator. The filters $\mathbf{g}_k(\tilde{t}), k = 1, \dots, p$, are given by

$$\begin{aligned} \mathbf{G}(\tilde{t}) &= \mathbf{M}_r(\tilde{t})\mathbf{W}_{\text{pf}}(\tilde{t}) \\ &\triangleq [\mathbf{g}_1(\tilde{t}) \ \mathbf{g}_2(\tilde{t}) \ \cdots \ \mathbf{g}_p(\tilde{t})] \end{aligned} \quad (3.1)$$

where

$$\mathbf{g}_k(\tilde{t}) = [g_{k,0}(\tilde{t}) \ g_{k,1}(\tilde{t}) \ \cdots \ g_{k,q}(\tilde{t})]^T, k = 1, \dots, p \quad (3.2)$$

$$\mathbf{M}_r(\tilde{t}) = \tilde{\mathbf{I}}_{q+1} \cdot [\mathbf{r}(\tilde{t}\tilde{T}) \ \mathbf{r}(\tilde{t}\tilde{T}+1) \ \cdots \ \mathbf{r}(\tilde{t}\tilde{T}+q)]^H \quad (3.3)$$

$$\mathbf{r}(t) = [r_1(t) \ r_2(t) \ \cdots \ r_p(t)]^T \quad (3.4)$$

Given $\mathbf{G}(\tilde{t})$, the filtered data is obtained from

$$\begin{aligned} \mathbf{Y}(\tilde{t}) &= \frac{1}{q+1} \tilde{\mathbf{X}}(\tilde{t})\mathbf{G}(\tilde{t}) \\ &\triangleq [\mathbf{y}_1(\tilde{t}) \ \mathbf{y}_2(\tilde{t}) \ \cdots \ \mathbf{y}_p(\tilde{t})] \end{aligned} \quad (3.5)$$

where

$$\tilde{\mathbf{X}}(\tilde{t}) = [\mathbf{x}(\tilde{t}\tilde{T}) \ \mathbf{x}(\tilde{t}\tilde{T}+1) \ \cdots \ \mathbf{x}(\tilde{t}\tilde{T}+q)] \cdot \tilde{\mathbf{I}}_{q+1} \quad (3.6)$$

The vector $\mathbf{r}(t)$ is composed of p known signal sequences each of length $q+1$, and will later be referred to as the pre-filtering reference signal. The choice of $\mathbf{W}_{\text{pf}}(\tilde{t})$, which is assumed to be positive definite and is a function of $\mathbf{r}(t)$, defines the type of



filtering operation as performed in Eq. (3.5). It is assumed at the moment that $\mathbf{r}(t) = \mathbf{s}_{\text{nc}}(t)$, where $\mathbf{s}_{\text{nc}}(t)$ is the result of the following decomposition of $\mathbf{s}(t)$ [†].

$$\mathbf{s}(t) = \Gamma_s \mathbf{s}_{\text{nc}}(t) \quad (3.7)$$

The $K \times d_s$ matrix Γ_s is composed of d_s independent column vectors. The $d_s \times 1$ vector $\mathbf{s}_{\text{nc}}(t) = [s_{\text{nc};1}(t) \ s_{\text{nc};2}(t) \ \cdots \ s_{\text{nc};d_s}(t)]^T$ is assumed to be composed of non-coherent signals. Thus its covariance matrix $\mathbf{R}_{\text{nc}} = \text{E}\{\mathbf{s}_{\text{nc}}(t)\mathbf{s}_{\text{nc}}^H(t)\}$ is of full rank d_s . Note that, as defined in Chapter 2, d_s is the rank of the signal covariance matrix $\mathbf{R}_s$. With this pre-filtering reference signal, four different filters, based on different choices of $\mathbf{W}_{\text{pf}}(\tilde{t})$, are discussed below.

a) Unitary filter

For this filter type, $\mathbf{W}_{\text{pf}}(\tilde{t})$ is given by

$$\mathbf{W}_{\text{pf}}(\tilde{t}) = \left(\frac{1}{q+1} \mathbf{M}_r^H(\tilde{t}) \mathbf{M}_r(\tilde{t}) \right)^{-1/2} \quad (3.8)$$

This choice of $\mathbf{W}_{\text{pf}}(\tilde{t})$ gives $\mathbf{g}_k(\tilde{t})$, $k = 1, \dots, p$, which preserve the power of the signal component at each of the sensor outputs, with minimum white noise gain. To see this, notice that $\mathbf{G}(\tilde{t})/\sqrt{q+1}$ with $\mathbf{W}_{\text{pf}}(\tilde{t})$ as given by Eq. (3.8) is composed of the orthonormal vectors which form the basis of the column space of $\mathbf{M}_r(\tilde{t})$. Thus, with $\mathbf{r}(t) = \mathbf{s}_{\text{nc}}(t)$, at each of the N sensor outputs, the power of each of the signal components

$$[a_{i,1} \ a_{i,2} \ \cdots \ a_{i,K}] \cdot \Gamma_s \cdot \mathbf{M}_r^H(\tilde{t}), \ i = 1, \dots, N$$

remains unchanged after filtering. Note that, $a_{i,k}$ denotes the (i, k) th element of the response matrix $\mathbf{A}$. In addition, the filtering operation in Eq. (3.5) with $\mathbf{W}_{\text{pf}}(\tilde{t})$ as given by Eq. (3.8) performs the unitary transformation (filtering) operation to the scaled data matrix $\tilde{\mathbf{X}}(\tilde{t})/\sqrt{q+1}$, rather than $\tilde{\mathbf{X}}(\tilde{t})$.

[†] The decomposition in Eq. (3.7) is not unique. However, this issue is irrelevant to our discussion.

b) Nulling filter

For this filter type,

$$\mathbf{W}_{\text{pf}}(\tilde{t}) = \left(\frac{1}{q+1} \mathbf{M}_{\text{r}}^H(\tilde{t}) \mathbf{M}_{\text{r}}(\tilde{t}) \right)^{-1} \quad (3.9)$$

The filter $\mathbf{g}_k(\tilde{t})$ is chosen to minimise the white noise gain, subject to

$$\mathbf{M}_{\text{r}}^H(\tilde{t}) \mathbf{g}_k(\tilde{t}) = \begin{bmatrix} 0 & \cdots & \underset{\text{kth}}{q+1} & \cdots & 0 \end{bmatrix}^T \quad (3.10)$$

With the constraint Eq. (3.10), the filter $\mathbf{g}_k(\tilde{t})$ preserves the k th signal power while removing other signal components. By solving for the solution of this linearly constrained optimisation problem, we obtain

$$\mathbf{G}(\tilde{t}) = \mathbf{M}_{\text{r}}(\tilde{t}) \left(\frac{1}{q+1} \mathbf{M}_{\text{r}}^H(\tilde{t}) \mathbf{M}_{\text{r}}(\tilde{t}) \right)^{-1} \quad (3.11)$$

which is Eq. (3.1) with $\mathbf{W}_{\text{pf}}$ as given by Eq. (3.9).

c) Matched filter

For this filter, $\mathbf{W}_{\text{pf}}(\tilde{t})$ is the diagonal matrix whose diagonal elements are given by

$$\left[\mathbf{W}_{\text{pf}}(\tilde{t}) \right]_{k,k} = \left(\frac{1}{q+1} \left[\mathbf{M}_{\text{r}}(\tilde{t}) \right]_k^H \left[\mathbf{M}_{\text{r}}(\tilde{t}) \right]_k \right)^{-1/2}, \quad k = 1, \dots, p \quad (3.12)$$

It is seen that with $\mathbf{W}_{\text{pf}}(\tilde{t})$ as given by Eq. (3.12), $\mathbf{g}_k(\tilde{t})$ is the matched filter which maximises at the filter output the power ratio of the k th signal sequence $\tilde{\mathbf{I}}_{q+1} \left[\mathbf{M}_{\text{r}}(\tilde{t}) \right]_k^*$, and the temporally white noise uncorrelated with $s_{\text{nc};k}(t)$. For the k th filter $\mathbf{g}_k(\tilde{t})$, as well as the additive noise $\mathbf{n}(t)$, the signal components in $\mathbf{s}(t)$ which are uncorrelated with $s_{\text{nc};k}(t)$ are treated as the uncorrelated temporally white noise.

d) Unitary-Matched filter

For this filter type,

$$\mathbf{W}_{\text{pf}}(\tilde{t}) = \begin{bmatrix} \mathbf{W}_{\text{pf};1}(\tilde{t}) & & & \mathbf{0} \\ & \mathbf{W}_{\text{pf};2}(\tilde{t}) & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{W}_{\text{pf};K_c}(\tilde{t}) \end{bmatrix} \quad (3.13)$$

where

$$\mathbf{W}_{\text{pf};k}(\tilde{t}) = \left(\frac{1}{q+1} \mathbf{M}_k^H(\tilde{t}) \mathbf{M}_k(\tilde{t}) \right)^{-1/2} \quad (3.14)$$

$$\mathbf{M}_k(\tilde{t}) = \tilde{\mathbf{I}}_{q+1} \cdot \left[\mathbf{r}_k(\tilde{t}\tilde{T}) \quad \mathbf{r}_k(\tilde{t}\tilde{T}+1) \quad \cdots \quad \mathbf{r}_k(\tilde{t}\tilde{T}+q) \right]^H, k = 1, \dots, K_c \quad (3.15)$$

and $\mathbf{r}_k(t)$, $k = 1, \dots, K_c$, are the $K_k \times 1$ subvectors of $\mathbf{r}(t)$ as defined by

$$\left[\mathbf{r}_1^T(t) \quad \mathbf{r}_2^T(t) \quad \cdots \quad \mathbf{r}_{K_c}^T(t) \right]^T = \mathbf{r}(t) \quad (3.16)$$

If each of $\mathbf{r}_k(t)$, $k = 1, \dots, K_c$, is assumed to be composed of non-coherent sources, and $\mathbf{E}\{\mathbf{r}_k(t)\mathbf{r}_l^H(t)\} = \mathbf{E}\{\mathbf{r}_k(t)\mathbf{r}_l^T(t)\} = \mathbf{0}$ for $k \neq l$, the K signals are filtered into K_c groups of partially correlated signals. The filter may be seen as a combination of the Unitary filter and the Matched filter.

In the above discussion of the four filter types, it is assumed that $\mathbf{r}(t) = \mathbf{s}_{\text{nc}}(t)$. The optimality for each filter type is thus claimed under this choice of the pre-filtering reference signal. Unfortunately, $\mathbf{s}_{\text{nc}}(t)$ is not always available in practice. Nevertheless, some other choices of $\mathbf{r}(t)$ can still provide useful filtering operation, provided that $\mathbf{r}(t)$ is highly correlated with $\mathbf{s}_{\text{nc}}(t)$ and a suitable choice of $\mathbf{W}_{\text{pf}}$, which in turn defines the filter type, is applied. In this thesis, we consider a practically obtainable pre-filtering reference signal of the form

$$\mathbf{r}(t) = \tilde{\mathbf{A}}\mathbf{s}_{\text{nc}}(t) + \mathbf{Q}_n\mathbf{n}(t) + \tilde{\mathbf{n}}(t) \quad (3.17)$$

where $\tilde{\mathbf{A}}$ and $\mathbf{Q}_n$ are, respectively, the $p \times d_s$ and $p \times N$ matrices. The $N \times 1$ vector $\mathbf{n}(t)$ is the additive noise in Eq. (2.2). The term $\tilde{\mathbf{n}}(t)$ in Eq. (3.17) is the second noise source which is assumed to be uncorrelated with both $s_{nc}(t)$ and $\mathbf{n}(t)$. In the following section, we consider four pre-filtering reference signals of the form Eq. (3.17). The formulations of the ML DOA estimators, based on the filtered data using each of the four filters, are provided as well as the analyses of the asymptotic properties of the estimators.

3.2 Pre-filtering maximum likelihood DOA estimation

In this and subsequent discussions, we limit our interest to the case of a single snapshot of the filtered data matrix $\mathbf{Y}(\tilde{t})$ which is obtained by filtering L snapshots of the observed data vector $\mathbf{x}(t)$. Thus, the filter order q is equal to $L - 1$. The data matrix $\mathbf{Y}(\tilde{t})$ will be simply written as $\mathbf{Y}$. This notational simplification also applies to other functions of $\tilde{t}$. It is assumed that the number of signals K and the rank d_s of the signal covariance matrix are known. In addition to the assumptions A.2.1-2.6 in Chapter 2, the ergodic stochastic signals and noises $\mathbf{x}(t)$, $s(t)$, $\mathbf{n}(t)$, $\tilde{\mathbf{n}}(t)$, and the pre-filtering reference signal $\mathbf{r}(t)$ are assumed to be circular complex zero-mean jointly Gaussian distributed. Sometimes, however, interesting interpretations based on the deterministic assumption of $s(t)$ are mentioned. Both $\mathbf{n}(t)$, and $\tilde{\mathbf{n}}(t)$ are assumed to be temporally white. If not stated otherwise, the spatial noise covariance structures[†] of $\mathbf{n}(t)$ and $\tilde{\mathbf{n}}(t)$, denoted by $\mathbf{R}_{ns}$ and $\tilde{\mathbf{R}}_{ns}$, are unknown positive definite matrices. Under the described assumptions, the PFML DOA estimators based on the four filtering pre-processors, for four pre-filtering reference signals of the form of Eq. (3.17), are described as follows.

[†] Here, the noise covariance structure is the matrix which is a scaled version of the noise covariance matrix. The scaling factor is an arbitrary unknown positive scalar number.

3.2.1 PFML DOA estimation with the type-1 pre-filtering reference signal

The type-1 pre-filtering reference signal $\mathbf{r}_1(t)$ is given by

$$\begin{aligned}\mathbf{r}_1(t) &= \mathbf{s}_{\text{nc}}(t) \\ &\triangleq \begin{bmatrix} r_{1;1}(t) & r_{1;2}(t) & \cdots & r_{1;d_s}(t) \end{bmatrix}^T\end{aligned}\quad (3.18)$$

The signal $\mathbf{r}_1(t)$ is $\mathbf{r}(t)$ in Eq. (3.17) with $p = d_s$, $\tilde{\mathbf{A}} = \mathbf{I}_{d_s}$, $\mathbf{Q}_n = \mathbf{0}$, and $\tilde{\mathbf{n}}(t) = \mathbf{0}$. To formulate the ML DOA estimator from the filtered data matrix $\mathbf{Y}$, using this type-1 pre-filtering reference signal, we first provide in the following theorem a general ML DOA estimator based on the filtered data matrix $\mathbf{Y}$ using the pre-filtering reference signal $\mathbf{r}(t)$ of the form given by Eq. (3.17) with $\mathbf{Q}_n = \mathbf{0}$.

Theorem 3.1: Under the signal and noise condition as stated, the estimate of θ is obtained by maximising the following likelihood function

$$\begin{aligned}\mathcal{L}(\mathbf{Y}, \mathbf{W}_{\text{pf}}, \hat{\mathbf{R}}_r, \hat{\mathbf{R}}_n; \theta, \tilde{\mathbf{R}}_n, \mathbf{T}) &= \\ &= \frac{\left|L\hat{\mathbf{R}}_r\right|^N \left|L\hat{\mathbf{R}}_n\right|^{L-N-d_s}}{\pi^{d_s N} \Gamma_N^f(L-d_s) \left|\tilde{\mathbf{R}}_n\right|^L} \exp\{-\text{tr}\{ \{ (\mathbf{Y} - \mathbf{A}\mathbf{T})\mathbf{W}_{\text{pf}}^{-1} \hat{\mathbf{R}}_r^{-1} \mathbf{W}_{\text{pf}}^{-1} (\mathbf{Y} - \mathbf{A}\mathbf{T})^H + \hat{\mathbf{R}}_n \} \cdot L\tilde{\mathbf{R}}_n^{-1} \} \}\end{aligned}\quad (3.19)$$

with respect to θ , $\tilde{\mathbf{R}}_n^\dagger$, and $\mathbf{T}$, under the conditions that $\theta \in \Theta^{K \times 1}$, where $\Theta^{K \times 1}$ is the set of feasible parameter values, $\tilde{\mathbf{R}}_n$ is positive definite, and the $K \times d_s$ matrix $\mathbf{T}$ has the form given by

$$\mathbf{T} = \mathbf{R}_{\text{sr}} \mathbf{R}_r^{-1} \hat{\mathbf{R}}_r \mathbf{W}_{\text{pf}} \quad (3.20)$$

where

$$\begin{aligned}\hat{\mathbf{R}}_r &= \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{r}(t) \mathbf{r}^H(t) \\ &= \frac{1}{L} \mathbf{M}_r^H \mathbf{M}_r\end{aligned}\quad (3.21)$$

[†] In this problem, $\tilde{\mathbf{R}}_n$ is an unknown parameter, not an ensemble averaged covariance matrix of a stochastic matrix.

and $\mathbf{R}_{\text{sr}} = \text{E}\{\mathbf{s}(t)\mathbf{r}^H(t)\}$, $\mathbf{R}_{\text{r}} = \text{E}\{\hat{\mathbf{R}}_{\text{r}}\}$. The matrix $\hat{\tilde{\mathbf{R}}}_{\text{n}}$ in Eq. (3.19) is given by

$$\hat{\tilde{\mathbf{R}}}_{\text{n}} = \hat{\mathbf{R}} - \mathbf{Y}\mathbf{W}_{\text{pf}}^{-1}\hat{\mathbf{R}}_{\text{r}}^{-1}\mathbf{W}_{\text{pf}}^{-1}\mathbf{Y}^H \quad (3.22)$$

By taking logarithms of both sides of Eq. (3.19) and neglecting unrelated terms, the ML estimates of $\boldsymbol{\theta}$ and $\mathbf{T}$ are alternatively obtained by minimising the following cost function

$$f(\boldsymbol{\theta}, \mathbf{T}) = \left| (\mathbf{Y} - \mathbf{A}\mathbf{T})\mathbf{W}_{\text{pf}}^{-1}\hat{\mathbf{R}}_{\text{r}}^{-1}\mathbf{W}_{\text{pf}}^{-1}(\mathbf{Y} - \mathbf{A}\mathbf{T})^H + \hat{\tilde{\mathbf{R}}}_{\text{n}} \right| \quad (3.23)$$

Proof: See Appendix B

From Eq. (3.23), it is seen that any positive definite matrix $\mathbf{W}_{\text{pf}}$ yields the same ML estimate of $\boldsymbol{\theta}$. However, the cost function Eq. (3.23) is perhaps at its simplest form when $\mathbf{W}_{\text{pf}}$ is chosen as

$$\mathbf{W}_{\text{pf}} = \hat{\mathbf{R}}_{\text{r}}^{-1/2} \quad (3.24)$$

This choice of $\mathbf{W}_{\text{pf}}$ corresponds to the use of the Unitary filter in the pre-filtering step. With $\mathbf{W}_{\text{pf}}$ as given by Eq. (3.24), the cost function Eq. (3.23) can be written as

$$f(\boldsymbol{\theta}, \mathbf{T}) = \left| (\mathbf{Y} - \mathbf{A}\mathbf{T})(\mathbf{Y} - \mathbf{A}\mathbf{T})^H + \hat{\tilde{\mathbf{R}}}_{\text{n}} \right| \quad (3.25)$$

where

$$\mathbf{T} = \mathbf{R}_{\text{sr}}\mathbf{R}_{\text{r}}^{-1}\hat{\mathbf{R}}_{\text{r}}^{1/2} \quad (3.26)$$

and

$$\hat{\tilde{\mathbf{R}}}_{\text{n}} = \hat{\mathbf{R}} - \mathbf{Y}\mathbf{Y}^H \quad (3.27)$$

Now, let $\mathbf{r}(t) = \mathbf{r}_1(t)$, we obtain

$$\mathbf{T} = \boldsymbol{\Gamma}_{\text{s}}\hat{\mathbf{R}}_{\text{r}_1}^{1/2} \quad (3.28)$$

where

$$\hat{\mathbf{R}}_{r_1} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{r}_1(t) \mathbf{r}_1^H(t) \quad (3.29)$$

which is equal to $\hat{\mathbf{R}}_{nc} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{s}_{nc}(t) \mathbf{s}_{nc}^H(t)$.

With this choice of the filter and the pre-filtering reference signal, the described ML estimator, which is based on the minimisation of Eq. (3.25), coincides with the exact ML estimator with known signal waveforms in [16-17], although the derivation in [16] is different from the one given here. Minimisation of Eq. (3.25) with respect to θ and $\mathbf{T}$ is, however, a computational intensive task. In [16-17], a somewhat simpler large sample cost function which is an approximation to Eq. (3.25) was derived and is given by

$$f_0(\theta, \mathbf{T}) = \text{tr}\{(\mathbf{Y} - \mathbf{A}\mathbf{T})^H \hat{\mathbf{R}}_n^{-1} (\mathbf{Y} - \mathbf{A}\mathbf{T})\} \quad (3.30)$$

The above cost function is noted to be slightly different in form from the original cost function in [16-17], which is shown in Eq. (2.35). For the case of known noise covariance structure $\mathbf{R}_{ns}$, the *exact* (not asymptotic) ML estimates of θ and $\mathbf{T}$ can be obtained by minimising Eq. (3.30), with $\hat{\mathbf{R}}_n$ being replaced by $\mathbf{R}_{ns}$. In addition, because $\hat{\mathbf{R}}_n$ as given in Eq. (3.27) is a consistent estimate of $\mathbf{R}_n$, it was shown in [16] that the estimates of θ and $\mathbf{T}$, obtained by minimising Eq. (3.30), are asymptotically equivalent to the ones obtained under the assumption that $\mathbf{R}_n$ or $\mathbf{R}_{ns}$ is known a priori.

When no knowledge of $\mathbf{T}$ is available, i.e., $\mathbf{T}$ is a completely unknown complex matrix, the ML estimate of θ can be obtained by noticing that Eq. (3.30) has the same form as the DML cost function. Thus, the asymptotic ML estimate of θ can be obtained by minimising the cost function similar to the concentrated DML cost function Eq. (2.24). However, if at least partial knowledge of Γ_s is available, the estimation accuracy can be improved. Of special interest here is the case where Γ_s has the form given by

$$\Gamma_s = \begin{bmatrix} \gamma_1 & & & 0 \\ & \gamma_2 & & \\ & & \ddots & \\ 0 & & & \gamma_{d_s} \end{bmatrix} \quad (3.31)$$

where $\gamma_k, k = 1, \dots, d_s$, are the $K_k \times 1$ complex amplitude vectors. The case where $K_k = 1, k = 1, \dots, d_s$, corresponds to a non-coherent source scenario, with the number of signals $K = d_s$. When any $K_k > 1$, Γ_s with the above structure can be found in a mixed scenario consisting of both uncorrelated (or partially correlated) and coherent signals. From Eq. (3.31), it is seen that under either of the two source scenarios, the number of the unknown complex parameters in Γ_s reduces from Kd_s to $\sum_{k=1}^{d_s} K_k (= K)$. Note that, sometimes further reduction of the number of the unknown parameters corresponding to Γ_s is possible. For instance, if Γ_s with the form Eq. (3.31) is also known to be a real matrix, there are $\sum_{k=1}^{d_s} K_k$ *real* rather than complex parameters which are unknown. However, here $\gamma_k, k = 1, \dots, d_s$, are assumed to be unknown complex vectors. The case where additional information of γ_k is known a priori was considered in [15-17].

Despite the reduction of the number of the unknown parameters in Γ_s which has the structure given by Eq. (3.31), minimisation of Eq. (3.30) with Γ_s of the form Eq. (3.31) can still be a computationally complicated problem due to the presence of the term $\hat{\mathbf{R}}_{r_1}^{1/2} (= \hat{\mathbf{R}}_{nc}^{1/2})$ in Eq. (3.28). Nevertheless, it is possible to obtain a computationally more efficient estimator by applying the Nulling filter instead of the Unitary filter in the pre-filtering step. With this filter type,

$$\mathbf{W}_{pf} = \hat{\mathbf{R}}_{nc}^{-1} \quad (3.32)$$

Due to this filtering operation, the column vectors of $\mathbf{Y}$ can be expressed as

$$\mathbf{y}_k = \mathbf{A}_k(\boldsymbol{\theta}_k)\gamma_k + \mathbf{e}_k, k = 1, \dots, d_s \quad (3.33)$$

where the $K \times K_k$ sub-response matrices $\mathbf{A}_k(\boldsymbol{\theta}_k), k = 1, \dots, d_s$, are defined by

$$\begin{bmatrix} A_1(\theta_1) & A_2(\theta_2) & \cdots & A_{d_s}(\theta_{d_s}) \end{bmatrix} = A(\theta) \quad (3.34)$$

From Eqs. (2.3, 3.34), it is noted that $\theta = [\theta_1^T \ \theta_2^T \ \cdots \ \theta_{d_s}^T]^T$. Later, the explicit expression of A_k as a function of θ_k will be dropped for notational convenience. In Eq. (3.33), the filtered noise vector e_k is obtained from

$$e_k = \frac{1}{L} \tilde{N} g_k \quad (3.35)$$

$$\tilde{N} = [n(0) \ n(1) \ \cdots \ n(L-1)] \cdot \tilde{I}_L \quad (3.36)$$

Note that g_k is the k th filter vector as given by Eq. (3.1). Because the filtering operation Eq. (3.35) is performed in the temporal domain, the filtered noise e_k has the same spatial covariance structure as that of $n(t)$. By noticing that g_k is a linear combination of d_s sequences of the signals which are uncorrelated with $n(t)$, $e_k = O_p(1/\sqrt{L})$ for large L . Lets assume at the moment that the noise covariance structure R_{ns} is known. With this, the ML estimates of θ_k and γ_k , given y_k , are obtained by maximising the log likelihood function (with the unrelated terms being neglected)

$$\ell_k = -\text{tr}\{ (y_k - A_k \gamma_k) R_{ns}^{-1} (y_k - A_k \gamma_k)^H \} \quad (3.37)$$

with respect to θ_k and γ_k . Alternatively, the ML estimate of θ_k , given y_k , can be obtained from

$$\hat{\theta}_k = \arg \min_{\theta_k \in \Theta^{K_k \times 1}} f_{1;k}(\theta_k), \quad k = 1, \dots, d_s \quad (3.38)$$

where

$$f_{1;k}(\theta_k) = \text{tr}\{ P_{R_{ns}^{-1/2} A_k}^\perp R_{ns}^{-1/2} y_k y_k^H R_{ns}^{-1/2} \} \quad (3.39)$$

$$P_{R_{ns}^{-1/2} A_k}^\perp = I_N - R_{ns}^{-1/2} A_k (A_k^H R_{ns}^{-1} A_k)^{-1} A_k^H R_{ns}^{-1/2} \quad (3.40)$$

and $\Theta^{K_k \times 1}$, $k = 1, \dots, d_s$, are the sets of feasible parameter values. As can be seen from the proof of Theorem 3.2 (to be given) in Appendix C, for up to a scaling factor, R_{ns} in

Eqs. (3.39-3.40) can be replaced by its consistent estimate without affecting the asymptotic properties of the described estimator. For instance, $\mathbf{R}_{\text{ns}}$ can be replaced by

$$\hat{\tilde{\mathbf{R}}}_{\text{n}} = \hat{\mathbf{R}} - \mathbf{Y}\hat{\mathbf{R}}_{\text{nc}}\mathbf{Y}^H \quad (3.41)$$

For large L , it is noted that $\hat{\tilde{\mathbf{R}}}_{\text{n}} = \mathbf{R}_{\text{n}} + O_p(1/\sqrt{L})$. In addition, the estimate $\hat{\theta}_k$ obtained by Eq. (3.38) is not affected by the scaling of $\mathbf{R}_{\text{ns}}$ (or $\hat{\tilde{\mathbf{R}}}_{\text{n}}$).

The estimator Eq. (3.38) was first derived in [16-17] (see also Eq. (2.40)) for the special non-coherent source case, i.e., $K_k = 1$, for $k = 1, \dots, K$ ($= d_s$ in this case). The estimator Eq. (3.38), under the more general case where Γ_s has the form Eq. (3.31), will thus be referred to as the DEML (Decoupled Maximum Likelihood) estimator, the acronym first used in [16]. It should be noted, however, that although $\hat{\theta}_k$, obtained from Eq. (3.38), is the asymptotic ML estimate when $\mathbf{y}_k$ is given, the estimates $\hat{\theta}_k$, $k = 1, \dots, d_s$, are not necessarily the asymptotic ML estimates for the given $\mathbf{Y}$, obtained by minimising Eq. (3.30). Because here all of the column vectors of $\mathbf{Y}$ are given at the same time, the estimator Eq. (3.38), which uses each of the column vectors of $\mathbf{Y}$, one at a time, is thus sub-optimal. Nevertheless, it was proved in [16] for the non-coherent source case ($K_k = 1$) that, if $s_{\text{nc}}(t)$ is composed of uncorrelated signals, the estimates $\hat{\theta}_k$, $k = 1, \dots, d_s$, which are obtained from Eq. (3.38) are also the asymptotic ML estimates for the given $\mathbf{Y}$, and thus asymptotically statistically efficient under this condition. Later, it will be shown that this is also true for the more general case $K_k > 1$. It is also worth noting that, when compared with the estimator based on Eq. (3.30), the DEML estimator was found in [16] to yield only slightly inferior performance when $s_{\text{nc}}(t)$ is composed of partially correlated signals.

If required, in addition to the estimates of θ_k , $k = 1, \dots, d_s$, the DEML estimates of γ_k , $k = 1, \dots, d_s$, can be obtained from

$$\hat{\gamma}_k = (\hat{\mathbf{A}}_{\text{de};k}^H \hat{\tilde{\mathbf{R}}}_{\text{n}}^{-1} \hat{\mathbf{A}}_{\text{de};k})^{-1} \hat{\mathbf{A}}_{\text{de};k}^H \hat{\tilde{\mathbf{R}}}_{\text{n}}^{-1} \mathbf{y}_k \quad (3.42)$$

where $\hat{\mathbf{A}}_{\text{de};k}$ is the sub-response matrix corresponding to the DEML estimate $\hat{\boldsymbol{\theta}}_k$, which is obtained from Eq. (3.38) with $\mathbf{R}_{\text{ns}}$ being replaced by $\hat{\mathbf{R}}_n$.

Asymptotic analysis

In [16-17], the asymptotic properties of the DEML estimator were analysed. Results from similar analysis for the more general case where Γ_s has the form Eq. (3.31), are summarised in the following theorem.

Theorem 3.2: Under the described signal and noise model assumptions, the DEML estimate $\hat{\boldsymbol{\theta}}_k$, obtained from Eq. (3.38) with $\mathbf{r}(t) = \mathbf{r}_1(t)$ and Γ_s of the form Eq. (3.31), is asymptotically Gaussian distributed with mean $\bar{\boldsymbol{\theta}}_k$ and the covariance matrix

$$\mathbb{E}\{(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k)(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k)^T\} = \frac{[\mathbf{R}_{\text{nc}}^{-1}]_{k,k}}{2L} \text{Re}^{-1}\{(\bar{\mathbf{D}}_k^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{D}}_k) \odot (\bar{\boldsymbol{\gamma}}_k \bar{\boldsymbol{\gamma}}_k^H)^T\},$$

$$k = 1, \dots, d_s \quad (3.43)$$

where $\mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp$ is the null projection of the column space of $\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k$, $\bar{\mathbf{R}}_n = \mathbf{R} - \bar{\mathbf{A}} \mathbf{R}_s \bar{\mathbf{A}}^H = \mathbf{R}_n$,

$$\bar{\mathbf{D}}_k = \begin{bmatrix} \left. \frac{\partial \mathbf{a}_{k;1}}{\partial \theta_{k;1}} \right|_{\bar{\theta}_{k;1}} & \left. \frac{\partial \mathbf{a}_{k;2}}{\partial \theta_{k;2}} \right|_{\bar{\theta}_{k;2}} & \cdots & \left. \frac{\partial \mathbf{a}_{k;K_k}}{\partial \theta_{k;K_k}} \right|_{\bar{\theta}_{k;K_k}} \end{bmatrix} \quad (3.44)$$

and $[\theta_{k;1} \ \theta_{k;2} \ \cdots \ \theta_{k;K_k}]^T = \boldsymbol{\theta}_k$, $[\mathbf{a}_{k;1} \ \mathbf{a}_{k;2} \ \cdots \ \mathbf{a}_{k;K_k}] = \mathbf{A}_k$.

Proof: See Appendix C.

As previously noted, the DEML method is also the asymptotic ML estimator if $\mathbf{s}_{\text{nc}}(t)$ is composed of uncorrelated signals. This implies also that the DEML estimator is asymptotically statistically efficient. This property of the DEML estimator was proved in [16] for the special case where Γ_s is a diagonal matrix. With slight modification of the CRLB of the DOA estimates with known $\mathbf{s}_{\text{nc}}(t)$, derived in [16], the bound for the more general case where Γ_s has the form Eq. (3.31) is given by

$$\text{CRLB}(\hat{\theta}) = \frac{1}{2L} \text{Re}^{-1} \{ \{ (\bar{\mathbf{D}}^H \mathbf{R}_n^{-1} \bar{\mathbf{D}}) \odot (\bar{\Gamma}_s \hat{\mathbf{R}}_{\text{nc}} \bar{\Gamma}_s^H)^T \} - \Phi^H \Psi^{-1} \Phi \} \quad (3.45)$$

where $\bar{\mathbf{D}}$ is given by Eq. (E.9) in Appendix E;

$$\Phi = (\bar{\mathbf{A}}^H \mathbf{R}_n^{-1} \bar{\mathbf{D}}) \odot (\bar{\Gamma}_s \hat{\mathbf{R}}_{\text{nc}} \Gamma_o^H)^T \quad (3.46)$$

$$\Psi = (\bar{\mathbf{A}}^H \mathbf{R}_n^{-1} \bar{\mathbf{A}}) \odot (\Gamma_o \hat{\mathbf{R}}_{\text{nc}} \Gamma_o^H)^T \quad (3.47)$$

$$\Gamma_o = \begin{bmatrix} \mathbf{I}_{K_1} & & \mathbf{0} \\ & \mathbf{I}_{K_2} & \\ \mathbf{0} & & \mathbf{I}_{K_{d_s}} \end{bmatrix} \quad (3.48)$$

and $\bar{\Gamma}_s$ is Γ_s which corresponds to $\bar{\gamma}_k$, $k = 1, \dots, d_s$. If $s_{\text{nc}}(t)$ is composed of uncorrelated signals, for large L ,

$$\begin{aligned} \text{CRLB}(\hat{\theta}_k) &\simeq \frac{1}{2L} \text{Re}^{-1} \{ \{ (\bar{\mathbf{D}}_k^H \mathbf{R}_n^{-1} \bar{\mathbf{D}}_k) \odot ([\mathbf{R}_{\text{nc}}]_{k,k} (\bar{\gamma}_k \bar{\gamma}_k^H)^T) \} - \\ &\quad \{ (\bar{\mathbf{D}}_k^H \mathbf{R}_n^{-1} \bar{\mathbf{A}}_k) (\bar{\mathbf{A}}_k^H \mathbf{R}_n^{-1} \bar{\mathbf{A}}_k)^{-1} (\bar{\mathbf{A}}_k^H \mathbf{R}_n^{-1} \bar{\mathbf{D}}_k) \odot ([\mathbf{R}_{\text{nc}}]_{k,k} (\bar{\gamma}_k \bar{\gamma}_k^H)^T) \} \} \\ &= \frac{[\mathbf{R}_{\text{nc}}]_{k,k}^{-1}}{2L} \text{Re}^{-1} \{ (\bar{\mathbf{D}}_k^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{D}}_k) \odot (\bar{\gamma}_k \bar{\gamma}_k^H)^T \} \end{aligned} \quad (3.49)$$

which coincides with Eq. (3.43) ($[\mathbf{R}_{\text{nc}}]_{k,k}^{-1} = [\mathbf{R}_{\text{nc}}^{-1}]_{k,k}$ in this uncorrelated-source case). It is thus concluded that the DEML estimator with Γ_s of the form given by Eq. (3.31) is asymptotically statistically efficient if $s_{\text{nc}}(t)$ is composed of uncorrelated signals.

Discussions

Some interesting remarks are given below.

1. The derivation of Eq. (3.23) is based on the assumption that the signal $s(t)$ is Gaussian distributed. An alternative derivation of Eq. (3.23) in [16], however, does not make any statistical model assumption of $s(t)$. These two different derivations result in the same cost function because, in both approaches, $s_{\text{nc}}(t)$ is assumed known a priori. Thus, any assigned probability density function which has non-zero value for any

realisation of $s_{\text{nc}}(t)$ (and of $s(t)$ through the relation Eq. (3.7)), will result in the same cost function Eq. (3.23). An example of such a function is the Gaussian density function.

2. From Eq. (3.43), with Γ_s as given by Eq. (3.31), it should be noted that the performance of the DEML method is largely unaffected by the angular separation between any two non-coherent sources. The reason for this can be understood by noticing that the filtered data vector y_k , obtained by using the Nulling filter ($W_{\text{pf}} = \hat{R}_{\text{nc}}^{-1}$) with the type-1 pre-filtering reference signal $r_1(t) = s_{\text{nc}}(t)$, can be written by Eq. (3.33). From Eq. (3.33), as a result of null filtering, y_k is entirely free of terms due to other non-coherent signal components of $s_{\text{nc}}(t)$. Thus, the estimation accuracy of the DEML estimate $\hat{\theta}_k$ is essentially unaffected by the directions of other non-coherent signals[†]. It should be noted that this property of the DEML estimator remains even if $\hat{R}_{\text{nc}}$ is not a diagonal matrix. However, when $\hat{R}_{\text{nc}}$ is non-diagonal, two filtered noise terms e_k and e_l , $k \neq l$, are correlated. As a result, $\hat{\theta}_k$, $k = 1, \dots, d_s$, which are obtained from Eq. (3.38), are no longer the asymptotic ML estimates for the given filtered data matrix Y . Nevertheless, as noted before, analytical and simulation results in [16-17] suggest that the performance of the DEML estimator is only gradually degraded when the correlation between the signal components of $s_{\text{nc}}(t)$ is increased.

3. The Null pre-filtering operation is performed entirely in the temporal domain. Thus, as shown in [16-17], the DEML method can estimate more source directions than those achievable by other existing methods such as the DML and SML methods. Note also that, unlike this temporal pre-filtering, the beamspace processor, which is based on spatial filtering, does not have this property.

[†] The DEML estimate $\hat{\theta}_k$ is slightly dependent on other non-coherent source directions through the use of the estimate Eq. (3.41) of the noise covariance matrix.

4. Unlike most existing DOA estimators, the DEML method does not require the knowledge of the spatial noise covariance structure. Thus, it can be used in an unknown spatially correlated noise field.

3.2.2 PFML DOA estimation with the type-2 pre-filtering reference signal

The type-2 pre-filtering reference signal $\mathbf{r}_2(t)$ is given by

$$\begin{aligned}\mathbf{r}_2(t) &= \mathbf{s}_{\text{nc}}(t) + \tilde{\mathbf{n}}(t) \\ &\triangleq \begin{bmatrix} r_{2;1}(t) & r_{2;2}(t) & \cdots & r_{2;d_s}(t) \end{bmatrix}^T\end{aligned}\quad (3.50)$$

This is the pre-filtering reference signal Eq. (3.17) with $p = d_s$, $\tilde{\mathbf{A}} = \mathbf{I}_{d_s}$, and $\mathbf{Q}_n = \mathbf{0}$. Due to the presence of the noise term $\tilde{\mathbf{n}}(t)$ in $\mathbf{r}_2(t)$, there is insufficient information to form the Nulling filter to decouple d_s non-coherent signals, as was the case when the type-1 pre-filtering reference signal is given. Although the asymptotic ML cost function Eq. (3.30) (with the Unitary filter) is still applicable in this case, high computation may be required to obtain the DOA estimates, subject to the constraint on the structure of $\mathbf{T}$, which is given by

$$\mathbf{T} = \mathbf{\Gamma}_s \mathbf{R}_{\text{nc}} \mathbf{R}_{\mathbf{r}_2}^{-1} \hat{\mathbf{R}}_{\mathbf{r}_2}^{1/2} \quad (3.51)$$

where

$$\hat{\mathbf{R}}_{\mathbf{r}_2} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{r}_2(t) \mathbf{r}_2^H(t) \quad (3.52)$$

The expression of $\mathbf{T}$ in Eq. (3.51) is obtained from Eq. (3.20) with $\mathbf{W}_{\text{pf}} = \hat{\mathbf{R}}_{\mathbf{r}_2}^{-1/2}$ and $\mathbf{r}(t) = \mathbf{r}_2(t)$. If $\mathbf{s}_{\text{nc}}(t)$ is composed of uncorrelated signals, however, a simpler sub-optimal estimator can be obtained by filtering the observed data by the Matched filter. With this filter type, $\mathbf{W}_{\text{pf}}$ is a diagonal matrix of which the diagonal elements are given by (see Eq. (3.12))

$$[\mathbf{W}_{\text{pf}}]_{k,k} = \left\{ \frac{1}{L} \sum_{t=0}^{L-1} |r_{2;k}(t)|^2 \right\}^{-1/2}, k = 1, \dots, d_s \quad (3.53)$$

With this filter type, each of the filtered data vectors $\mathbf{y}_k$, $k = 1, \dots, d_s$, has the form given by Eq. (3.33), with $\mathbf{e}_k = O_p(1/\sqrt{L})$ for large L due to the assumption that $s_{nc}(t)$ is composed of uncorrelated signals. Note that, in this case, the covariance structure of each $\mathbf{e}_k$ is not $\mathbf{R}_{ns}$. In addition, two noise vectors $\mathbf{e}_k$ and $\mathbf{e}_l$ generally have different noise covariance structures. Given $\bar{\mathbf{A}}_k$ and $\bar{\boldsymbol{\gamma}}_k$, the spatial covariance structure of each of $\mathbf{e}_k$, $k = 1, \dots, d_s$, is a scaled version of

$$\bar{\tilde{\mathbf{R}}}_{n;k} = \mathbf{R} - (\text{corr}_{r_2;k})^2 [\mathbf{R}_{nc}]_{k,k} \bar{\mathbf{A}}_k \bar{\boldsymbol{\gamma}}_k \bar{\boldsymbol{\gamma}}_k^H \bar{\mathbf{A}}_k^H \quad (3.54)$$

where

$$\begin{aligned} \text{corr}_{r_2;k} &= \left| E\{s_{nc;k}(t)r_{2;k}^*(t)\} \right| \cdot [\mathbf{R}_{nc}]_{k,k}^{-1/2} [\mathbf{R}_{r_2}]_{k,k}^{-1/2} \\ &= [\mathbf{R}_{nc}]_{k,k}^{1/2} [\mathbf{R}_{r_2}]_{k,k}^{-1/2} \end{aligned} \quad (3.55)$$

In a similar manner to the derivation of the DEML estimator, given $\mathbf{y}_k$, the ML estimate of $\boldsymbol{\theta}_k$ can be obtained from

$$\hat{\boldsymbol{\theta}}_k = \arg \min_{\boldsymbol{\theta}_k \in \Theta^{K_k \times 1}} f_{2;k}(\boldsymbol{\theta}_k), \quad k = 1, \dots, d_s \quad (3.56)$$

where

$$f_{2;k}(\boldsymbol{\theta}_k) = \text{tr}\{ \mathbf{P}_{\hat{\tilde{\mathbf{R}}}_{n;k}^{-1/2} \mathbf{A}_k}^\perp \hat{\tilde{\mathbf{R}}}_{n;k}^{-1/2} \mathbf{y}_k \mathbf{y}_k^H \hat{\tilde{\mathbf{R}}}_{n;k}^{-1/2} \} \quad (3.57)$$

and $\mathbf{A}_k$ is as defined in Eq. (3.34). The matrix $\hat{\tilde{\mathbf{R}}}_{n;k}$ is a consistent estimate of $\tilde{\mathbf{R}}_{n;k}$ or its scaled version. Here, $\hat{\tilde{\mathbf{R}}}_{n;k}$ is chosen as

$$\hat{\tilde{\mathbf{R}}}_{n;k} = \hat{\mathbf{R}} - \mathbf{y}_k \mathbf{y}_k^H \quad (3.58)$$

As for the case of the DEML estimator, the above estimator is sub-optimal because it uses each of the data vectors $\mathbf{y}_k$, $k = 1, \dots, d_s$, only one at a time. For ease of reference, we will refer to the estimator Eq. (3.56), based on the data filtered by the

Matched filter, as the PFML-2 method[†].

Asymptotic analysis

Due to the similarity between Eqs. (3.56-3.58) and Eqs. (3.38-3.39, 3.41), the asymptotic properties of the PFML-2 estimator can be obtained in a similar manner to those of the DEML estimator, and are summarised in the following theorem.

Theorem 3.3: Under the described signal and noise model assumptions, with $\mathbf{r}(t) = \mathbf{r}_2(t)$, Γ_s as given by Eq. (3.31), and $\mathbf{s}_{nc}(t)$ being composed of uncorrelated signals, the PFML-2 DOA estimate $\hat{\theta}_k$ is asymptotically Gaussian distributed with mean $\bar{\theta}_k$ and the covariance matrix

$$\mathbb{E}\{(\hat{\theta}_k - \bar{\theta}_k)(\hat{\theta}_k - \bar{\theta}_k)^T\} = \frac{[\mathbf{R}_{nc}]_{k,k}^{-1}}{2L(\text{corr}_{r_2;k})^2} \text{Re}^{-1}\{(\bar{\mathbf{D}}_k^H \bar{\mathbf{R}}_{n;k}^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_{n;k}^{-1/2} \bar{\mathbf{A}}_k}^\perp \bar{\mathbf{R}}_{n;k}^{-1/2} \bar{\mathbf{D}}_k) \odot (\bar{\boldsymbol{\gamma}}_k \bar{\boldsymbol{\gamma}}_k^H)^T\},$$

$$k = 1, \dots, d_s \quad (3.59)$$

where $\mathbf{R}_{r_2} = \mathbb{E}\{\hat{\mathbf{R}}_{r_2}\}$, $\mathbf{P}_{\bar{\mathbf{R}}_{n;k}^{-1/2} \bar{\mathbf{A}}_k}^\perp$ is the null projection of the column space of $\bar{\mathbf{R}}_{n;k}^{-1/2} \bar{\mathbf{A}}_k$, and $\bar{\mathbf{D}}_k$, $\bar{\mathbf{R}}_{n;k}$, and $\text{corr}_{r_2;k}$ are given, respectively, in Eqs. (3.44, 3.54, 3.55).

Proof: See Appendix D.

From Eq. (3.59), it is seen that the estimation Mean-Square Error (MSE) is inversely proportional to $(\text{corr}_{r_2;k})^2$ and $[\mathbf{R}_{nc}]_{k,k}$. The value of $\text{corr}_{r_2;k}$ measures the degree of correlation between the k th component of the pre-filtering reference signal $\mathbf{r}_2(t)$ and the k th signal component of $\mathbf{s}_{nc}(t)$, which, in this case, is uncorrelated with other signal components in $\mathbf{s}_{nc}(t)$. The value of $\text{corr}_{r_2;k}$ lies between 0 and 1, with the higher value indicating higher correlation between the two signals. It is seen from Eq. (3.59) that $\text{corr}_{r_2;k} = 1$, which corresponds to the case where $\tilde{\mathbf{n}}(t) = \mathbf{0}$, achieves the

[†] The DEML method should have been named the PFML-1 method. However, the acronym 'DEML' is chosen instead to acknowledge the original work. In fact, the PFML-2 estimator may have been called instead the DEML-2 method, due to its similarity to the DEML method.

lowest estimation error. Note, however, that the estimation accuracy of the PFML-2 estimator can not be better than that of the DEML method, even for this best signal correlation case. This can be seen by first noticing that

$$\bar{\tilde{R}}_{n;k} \geq R_n \quad (3.60)$$

with equality for the case where $d_s = 1$ and $\tilde{n}(t) = 0$. To prove the above performance order, it will be shown that

$$\bar{D}_k^H R_n^{-1/2} P_{R_n^{-1/2} \bar{A}_k}^\perp R_n^{-1/2} \bar{D}_k \geq \bar{D}_k^H \bar{\tilde{R}}_{n;k}^{-1/2} P_{\bar{\tilde{R}}_{n;k}^{-1/2} \bar{A}_k}^\perp \bar{\tilde{R}}_{n;k}^{-1/2} \bar{D}_k \quad (3.61)$$

Let $\bar{B}_k$ be the matrix whose column vectors form the null space of $\bar{A}_k^H$. Thus,

$$R_n^{-1/2} P_{R_n^{-1/2} \bar{A}_k}^\perp R_n^{-1/2} = \bar{B}_k (\bar{B}_k^H R_n \bar{B}_k)^{-1} \bar{B}_k^H \quad (3.62)$$

and

$$\bar{\tilde{R}}_{n;k}^{-1/2} P_{\bar{\tilde{R}}_{n;k}^{-1/2} \bar{A}_k}^\perp \bar{\tilde{R}}_{n;k}^{-1/2} = \bar{B}_k (\bar{B}_k^H \bar{\tilde{R}}_{n;k} \bar{B}_k)^{-1} \bar{B}_k^H \quad (3.63)$$

Because $\bar{\tilde{R}}_{n;k} \geq R_n$, and by the matrix inversion lemma (see, for example, Appendix A.2.4 of [44]), we obtain

$$\bar{B}_k (\bar{B}_k^H \bar{\tilde{R}}_{n;k} \bar{B}_k)^{-1} \bar{B}_k^H = \bar{B}_k (\bar{B}_k^H R_n \bar{B}_k)^{-1} \bar{B}_k^H - \bar{B}_k C \bar{B}_k^H \quad (3.64)$$

where C is a suitable positive semi-definite matrix. By multiplying both sides of Eq. (3.64) by $\bar{D}_k^H$ on the left and $\bar{D}_k$ on the right of each side, the inequality (3.61) is proved. Next, notice that the term $(\bar{\gamma}_k \bar{\gamma}_k^H)^T$ in Eq. (3.43) and in Eq. (3.59) is positive semi-definite. Thus, for large L with $s_{nc}(t)$ being composed of uncorrelated signals, it is seen that the MSE of each of the PFML-2 DOA estimates (each of the diagonal elements of the error covariance matrix Eq. (3.59)) using either $r_1(t)$ or $r_2(t)$, can not be smaller than the MSE of the estimate using the DEML method with the pre-filtering reference signal $r_1(t)$.

In addition to the described performance comparison, it should be noted that, unlike the DEML estimator, the MSE of $\hat{\theta}_k$ using the PFML-2 estimator is dependent of the directions of the signals which are uncorrelated with $r_{2;k}(t)$ and $s_{nc;k}(t)$. This can be seen by observing that

$$\bar{\tilde{R}}_{n;k} = \mathbf{R}_n + \sum_{l=1; l \neq k}^{d_s} \{ [\mathbf{R}_{nc}]_{l,l} (\bar{\mathbf{A}}_l \bar{\gamma}_l \bar{\gamma}_l^H \bar{\mathbf{A}}_l^H) \} + \{ (1 - (\text{corr}_{r_2;k})^2) [\mathbf{R}_{nc}]_{k,k} \} \cdot \{ \bar{\mathbf{A}}_k \bar{\gamma}_k \bar{\gamma}_k^H \bar{\mathbf{A}}_k^H \} \quad (3.65)$$

As a result,

$$\begin{aligned} \bar{\mathbf{D}}_k^H \bar{\tilde{R}}_{n;k}^{-1/2} \mathbf{P}_{\bar{\tilde{R}}_{n;k}^{-1/2} \bar{\mathbf{A}}_k}^\perp \bar{\tilde{R}}_{n;k}^{-1/2} \bar{\mathbf{D}}_k &= \bar{\mathbf{D}}_k^H \bar{\mathbf{B}}_k (\bar{\mathbf{B}}_k^H \bar{\tilde{R}}_{n;k} \bar{\mathbf{B}}_k)^{-1} \bar{\mathbf{B}}_k^H \bar{\mathbf{D}}_k \\ &= \bar{\mathbf{D}}_k^H \bar{\mathbf{B}}_k \{ \bar{\mathbf{B}}_k^H \mathbf{R}_n \bar{\mathbf{B}}_k + \bar{\mathbf{B}}_k^H \{ \sum_{l=1; l \neq k}^{d_s} [\mathbf{R}_{nc}]_{l,l} (\bar{\mathbf{A}}_l \bar{\gamma}_l \bar{\gamma}_l^H \bar{\mathbf{A}}_l^H) \} \bar{\mathbf{B}}_k \}^{-1} \bar{\mathbf{B}}_k^H \bar{\mathbf{D}}_k \end{aligned} \quad (3.66)$$

where $\bar{\mathbf{B}}_k$ is as previously defined. Note that, the last term in Eq. (3.65) disappears in Eq. (3.66) because $\bar{\mathbf{A}}_k^H \bar{\mathbf{B}}_k = \mathbf{0}$, by definition. From Eq. (3.59) and Eq. (3.66), the dependence of the estimation error to all source directions is seen.

3.2.3 PFML DOA estimation with the type-3 pre-filtering reference signal

The type-3 pre-filtering reference signal $\mathbf{r}_3(t)$ has the form given by

$$\begin{aligned} \mathbf{r}_3(t) &= \tilde{\mathbf{A}} s_{nc}(t) + \tilde{\mathbf{n}}(t) \\ &\triangleq \begin{bmatrix} r_{3;1}(t) & r_{3;2}(t) & \cdots & r_{3;d_s}(t) \end{bmatrix}^T \end{aligned} \quad (3.67)$$

where $\tilde{\mathbf{A}}$ is an unknown complex square matrix and is of full rank. When compared with Eq. (3.17), $\mathbf{r}_3(t)$ is $\mathbf{r}(t)$ with $p = d_s$, and $\mathbf{Q}_n = \mathbf{0}$. Because $\tilde{\mathbf{A}}$ is not known, neither the Nulling filter nor the Matched filter can be formed. Based on the data filtered by the Unitary filter using $\mathbf{r}_3(t)$, the DOA estimates can be obtained by minimising Eq. (3.25)

or the asymptotically equivalent cost function Eq. (3.30). With this type-3 pre-filtering reference signal, the conditional mean

$$\begin{aligned}
 E(Y | \hat{\mathbf{R}}_{r_3}, W_{\text{pf}} = \hat{\mathbf{R}}_{r_3}^{-1/2}, \tilde{\mathbf{A}}, \mathbf{A}, \Gamma_s) &= \mathbf{A} \Gamma_s \mathbf{R}_{\text{nc}} \tilde{\mathbf{A}}^H \mathbf{R}_{r_3}^{-1} \hat{\mathbf{R}}_{r_3} W_{\text{pf}} \\
 &= \mathbf{A} \Gamma_s \mathbf{R}_{\text{nc}} \tilde{\mathbf{A}}^H \mathbf{R}_{r_3}^{-1} \hat{\mathbf{R}}_{r_3}^{1/2} \\
 &\triangleq \mathbf{AT}
 \end{aligned} \tag{3.68}$$

where

$$\hat{\mathbf{R}}_{r_3} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{r}_3(t) \mathbf{r}_3^H(t) \tag{3.69}$$

and $\mathbf{R}_{r_3} = E\{\hat{\mathbf{R}}_{r_3}\}$. Except for the case where $\tilde{\mathbf{A}}$ is a diagonal matrix, minimisation of either Eq. (3.25) or Eq. (3.30) with respect to $\boldsymbol{\theta}$, and $\mathbf{T}$ of the form described in Eq. (3.68), can be a computationally intensive task. However, a simpler minimisation problem can be obtained by treating $\mathbf{T}$ as a completely unknown complex matrix. By doing so, given $\mathbf{Y}$, the ML estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta^{K \times 1}} f_3(\boldsymbol{\theta}) \tag{3.70}$$

where

$$f_3(\boldsymbol{\theta}) = \text{tr} \{ \mathbf{P}_{\hat{\mathbf{R}}_{r_3}^{-1/2} \mathbf{A}}^\perp \hat{\mathbf{R}}_{r_3}^{-1/2} \mathbf{Y} \mathbf{Y}^H \hat{\mathbf{R}}_{r_3}^{-1/2} \} \tag{3.71}$$

and $\hat{\mathbf{R}}_{r_3}$ is as given by Eq. (3.27). The estimator Eq. (3.70), based on the data filtered by the Unitary filter, will be referred to as the PFML-3 method.

Asymptotic analysis

The following theorem provides the asymptotic properties of the PFML-3 estimator.

Theorem 3.4: Under the signal and noise model as stated, the estimate $\hat{\theta}$, obtained by Eq. (3.70) with $\mathbf{r}(t) = \mathbf{r}_3(t)$, is asymptotically multivariate Gaussian distributed with mean $\bar{\theta}$ and the covariance matrix

$$\mathbb{E}\{(\hat{\theta} - \bar{\theta})(\hat{\theta} - \bar{\theta})^T\} = \frac{1}{2L} \text{Re}^{-1}\{ \bar{\mathbf{D}}^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^\perp \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{D}} \odot (\bar{\mathbf{T}} \bar{\mathbf{T}}^H)^T \} \quad (3.72)$$

where $\bar{\mathbf{T}} = \bar{\Gamma}_s \mathbf{R}_{\text{nc}} \tilde{\mathbf{A}}^H \mathbf{R}_{r_3}^{-1/2}$; $\bar{\mathbf{R}}_n = \mathbf{R} - (\bar{\mathbf{A}} \bar{\mathbf{T}})(\bar{\mathbf{A}} \bar{\mathbf{T}})^H$, and $\mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^\perp$ is the null projection of the column space of $\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}$.

Proof: See Appendix E.

Note that, unlike in the DEML method, $\bar{\mathbf{R}}_n$ in the above theorem is generally not equal to $\mathbf{R}_n$. However, $\bar{\mathbf{R}}_n$ can be replaced in Eq. (3.72) by $\mathbf{R}_n$. This can be shown by using the argument of the proof of Theorem 3 in [19]. Alternatively, this equivalence can be proved by using a slightly more general result, given in the following lemma.

Lemma 3.1: Let a positive definite matrix $\mathbf{C} = \mathbf{B} + \sum_i \mathbf{F}_i$, where $\mathbf{B}$ is positive definite, and $\mathbf{F}_i$ is a matrix whose column (or row) vectors lie within the space spanned by the column (or row) vectors of an arbitrary finite-value matrix $\mathbf{A}$ (or $\mathbf{A}^H$). Then,

$$\mathbf{C}^{-1/2} \mathbf{P}_{\mathbf{C}^{-1/2} \mathbf{A}}^\perp \mathbf{C}^{-1/2} = \mathbf{B}^{-1/2} \mathbf{P}_{\mathbf{B}^{-1/2} \mathbf{A}}^\perp \mathbf{B}^{-1/2} \quad (3.73)$$

Proof: The result can be proved by the same argument used in Proof of Theorem 3 in [19].

To gain more insight into the effect of the noise term $\tilde{\mathbf{n}}(t)$, we next express $\mathbf{R}_{r_3}$ as

$$\mathbf{R}_{r_3} = \tilde{\mathbf{A}} \mathbf{R}_{\text{nc}} \tilde{\mathbf{A}}^H + \tilde{\mathbf{R}}_n \quad (3.74)$$

where $\tilde{\mathbf{R}}_n = \mathbb{E}\{\tilde{\mathbf{n}}(t)\tilde{\mathbf{n}}^H(t)\}$. By using the matrix inversion lemma, we obtain

$$\begin{aligned}
\tilde{\mathbf{A}}^H \mathbf{R}_{r_3}^{-1} \tilde{\mathbf{A}} &= \tilde{\mathbf{A}}^H \tilde{\mathbf{R}}_n^{-1} \tilde{\mathbf{A}} - \tilde{\mathbf{A}}^H \tilde{\mathbf{R}}_n^{-1} \tilde{\mathbf{A}} (\mathbf{R}_{nc}^{-1} + \tilde{\mathbf{A}}^H \tilde{\mathbf{R}}_n^{-1} \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^H \tilde{\mathbf{R}}_n^{-1} \tilde{\mathbf{A}} \\
&= (\mathbf{R}_{nc} + \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{R}}_n (\tilde{\mathbf{A}}^H)^{-1})^{-1}
\end{aligned} \tag{3.75}$$

Next, by using Eq. (3.75), $\overline{\mathbf{T}}\overline{\mathbf{T}}^H$ can be expressed as

$$\begin{aligned}
\overline{\mathbf{T}}\overline{\mathbf{T}}^H &= \overline{\Gamma}_s \mathbf{R}_{nc} \tilde{\mathbf{A}}^H \mathbf{R}_{r_3}^{-1} \tilde{\mathbf{A}} \mathbf{R}_{nc} \overline{\Gamma}_s^H \\
&= \overline{\Gamma}_s \mathbf{R}_{nc} (\mathbf{R}_{nc} + \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{R}}_n (\tilde{\mathbf{A}}^H)^{-1})^{-1} \mathbf{R}_{nc} \overline{\Gamma}_s^H \\
&= \overline{\Gamma}_s (\mathbf{R}_{nc} + \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{R}}_n (\tilde{\mathbf{A}}^H)^{-1}) (\mathbf{R}_{nc} + \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{R}}_n (\tilde{\mathbf{A}}^H)^{-1})^{-1} \mathbf{R}_{nc} \overline{\Gamma}_s^H - \\
&\quad \overline{\Gamma}_s \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{R}}_n (\tilde{\mathbf{A}}^H)^{-1} (\mathbf{R}_{nc} + \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{R}}_n (\tilde{\mathbf{A}}^H)^{-1})^{-1} \mathbf{R}_{nc} \overline{\Gamma}_s^H \\
&= \overline{\Gamma}_s \{ \mathbf{R}_{nc} - (\tilde{\mathbf{A}}^H \tilde{\mathbf{R}}_n^{-1} \tilde{\mathbf{A}} + \mathbf{R}_{nc}^{-1})^{-1} \} \overline{\Gamma}_s^H
\end{aligned} \tag{3.76}$$

Because $\mathbf{R}_{nc}$ and $\tilde{\mathbf{A}}^H \tilde{\mathbf{R}}_n^{-1} \tilde{\mathbf{A}}$ are positive definite,

$$\mathbf{R}_{nc} > \mathbf{R}_{nc} - (\tilde{\mathbf{A}}^H \tilde{\mathbf{R}}_n^{-1} \tilde{\mathbf{A}} + \mathbf{R}_{nc}^{-1})^{-1} \tag{3.77}$$

If $\tilde{\mathbf{R}}_n$ is allowed to approach a zero matrix, the inequality (3.77) will approach equality. Because $\overline{\mathbf{T}} = \overline{\Gamma}_s \mathbf{R}_{nc} \tilde{\mathbf{A}}^H \mathbf{R}_{r_3}^{-1/2}$, for the case where $\tilde{\mathbf{R}}_n = \mathbf{0}$, it is seen that $\overline{\mathbf{T}}\overline{\mathbf{T}}^H = \overline{\Gamma}_s \mathbf{R}_{nc} \overline{\Gamma}_s^H$. Thus, from Eqs. (3.76-3.77),

$$\overline{\mathbf{T}}\overline{\mathbf{T}}^H \Big|_{\tilde{\mathbf{R}}_n=\mathbf{0}} > \overline{\mathbf{T}}\overline{\mathbf{T}}^H \Big|_{\tilde{\mathbf{R}}_n>\mathbf{0}} \tag{3.78}$$

which, in combination with Eq. (3.72), implies that

$$\mathbb{E}\{ (\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^T \} \Big|_{\tilde{\mathbf{R}}_n=\mathbf{0}} < \mathbb{E}\{ (\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^T \} \Big|_{\tilde{\mathbf{R}}_n>\mathbf{0}} \tag{3.79}$$

Discussions

First, the relation between the described PFML-3 estimator and the IV-based DOA estimators [18-20] should be noted. There have been two versions of the IV-based DOA estimators, namely the optimal temporal and spatial IV DOA estimation methods

[20]. In both the optimal SIV (Spatial Instrumental Variable) and optimal TIV (Temporal Instrumental Variable) DOA estimators, the instrumental variable vector of the similar form to that of $\mathbf{r}_3(t)$ is assumed available. For the optimal SIV estimator, the spatial instrumental variable vector is the data observed by the secondary array whose additive noise is uncorrelated with that of the primary array, where the actual DOA estimation takes place. In the optimal TIV estimator, the temporal instrumental variable vector is the time-delay version of $\mathbf{x}(t)$. Despite the difference in approach, it was shown in [20] that the optimal TIV and SIV estimators are mathematically equivalent. It will be shown that the two optimal IV-based methods can be cast in the form of the PFML-3 method. To show this, let the instrumental variable vector $\mathbf{s}_{iv}(t)$ be obtained under one of the following conditions.

a) The $M \times 1$ vector $\mathbf{s}_{iv}(t)$ is the output of the secondary array of M sensors. The signal component at this second array is fully correlated with that of the primary array. However, the noise components at the two arrays are uncorrelated. In addition, the secondary array response matrix is assumed to be composed of K independent column vectors.

b) The $M \times 1$ vector $\mathbf{s}_{iv}(t)$ is composed of M_τ delayed vectors of the observed data vector $\mathbf{x}(t)$. With $M = N \cdot M_\tau$, it is given by

$$\mathbf{s}_{iv}(t) = [\mathbf{x}^T(t-1) \quad \mathbf{x}^T(t-2) \quad \cdots \quad \mathbf{x}^T(t-M_\tau)]^T \quad (3.80)$$

under the assumption that the signal $s(t)$ is temporally coherent, while $\mathbf{n}(t)$ is temporally uncorrelated.

To prove the equivalence of the PFML-3 method and the optimal IV-based method with $\mathbf{s}_{iv}(t)$ being obtained under one of the above two conditions, consider first the optimal IV-WSF (Instrumental Variable-based Weighted Subspace Fitting) cost function [19]

$$f_{\text{iv-wsf}}(\theta) = \text{tr}\{ \mathbf{P}_{\hat{\mathbf{R}}^{-1/2}\mathbf{A}}^\perp \hat{\mathbf{R}}^{-1/2} \mathbf{Y}_{\text{iv}} \mathbf{Y}_{\text{iv}}^H \hat{\mathbf{R}}^{-1/2} \} \quad (3.81)$$

where $\mathbf{Y}_{\text{iv}} = \hat{\mathbf{U}}_{\text{iv}} \hat{\mathbf{\Lambda}}_{\text{iv}}$ with $\hat{\mathbf{U}}_{\text{iv}}$ and $\hat{\mathbf{\Lambda}}_{\text{iv}}$ being obtained from the following SVD decomposition

$$\begin{aligned} \hat{\mathbf{R}}_{\text{iv}} &= \hat{\tilde{\mathbf{R}}}_{\text{iv}} \hat{\mathbf{R}}_{\text{siv}}^{-1/2} \\ &= \hat{\mathbf{U}} \hat{\mathbf{\Lambda}} \hat{\mathbf{V}}^H \end{aligned} \quad (3.82)$$

where

$$\hat{\tilde{\mathbf{R}}}_{\text{iv}} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{x}(t) \mathbf{s}_{\text{iv}}^H(t) \quad (3.83)$$

$$\hat{\mathbf{R}}_{\text{siv}} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{s}_{\text{iv}}(t) \mathbf{s}_{\text{iv}}^H(t) \quad (3.84)$$

The matrix $\hat{\mathbf{U}}_{\text{iv}}$ is composed of the d_s left singular vectors which correspond to the diagonal matrix $\hat{\mathbf{\Lambda}}_{\text{iv}}$ of the d_s largest singular values. The difference between the optimal IV-WSF cost function and the PFML-3 cost function is that the former cost function uses $\mathbf{Y}_{\text{iv}}$ and $\hat{\mathbf{R}}$ instead of $\mathbf{Y}$ and $\hat{\tilde{\mathbf{R}}}_{\text{n}}$ in the latter cost function. From Lemma 3.1, however, it is seen that either $\hat{\tilde{\mathbf{R}}}_{\text{n}}$ or $\hat{\mathbf{R}}$ results in the same asymptotic properties of the DOA estimates. In addition, it will be shown that $\mathbf{Y}_{\text{iv}}$ can, in fact, be formulated under the PFML framework with the pre-filtering reference signal of the form Eq. (3.67), by applying certain spatio-temporal filtering to the instrumental variable vector $\mathbf{s}_{\text{iv}}(t)$. To see this, notice that $\mathbf{s}_{\text{iv}}(t)$ can be expressed as

$$\mathbf{s}_{\text{iv}}(t) = \tilde{\tilde{\mathbf{A}}} \mathbf{s}_{\text{nc}}(t) + \mathbf{n}_{\text{iv}}(t) \quad (3.85)$$

where $\tilde{\tilde{\mathbf{A}}}$ is a full rank- d_s matrix, $\text{E}\{\mathbf{n}_{\text{iv}}(t) \mathbf{n}^H(t)\} = \text{E}\{\mathbf{n}_{\text{iv}}(t) \mathbf{n}^T(t)\} = \mathbf{0}$, and $\text{E}\{\mathbf{s}_{\text{nc}}(t) \mathbf{n}_{\text{iv}}^H(t)\} = \text{E}\{\mathbf{s}_{\text{nc}}(t) \mathbf{n}_{\text{iv}}^T(t)\} = \mathbf{0}$.

Next, let

$$\mathbf{r}_3(t) = \mathbf{H}^H \mathbf{s}_{iv}(t) \quad (3.86)$$

where

$$\mathbf{H} = \hat{\mathbf{R}}_{s_{iv}}^{-1} \hat{\tilde{\mathbf{R}}}_{iv}^H \hat{\mathbf{U}}_{iv} \quad (3.87)$$

With $\mathbf{H}$ as given by Eq. (3.87), we obtain

$$\mathbf{r}_3(t) = \hat{\mathbf{U}}_{iv}^H \hat{\tilde{\mathbf{R}}}_{iv} \hat{\mathbf{R}}_{s_{iv}}^{-1} \mathbf{s}_{iv}(t) \quad (3.88)$$

which can be written in the form Eq. (3.67) with $\tilde{\mathbf{A}} = \hat{\mathbf{U}}_{iv}^H \hat{\tilde{\mathbf{R}}}_{iv} \hat{\mathbf{R}}_{s_{iv}}^{-1} \tilde{\mathbf{A}}$ and $\tilde{\mathbf{n}}(t) = \hat{\mathbf{U}}_{iv}^H \hat{\tilde{\mathbf{R}}}_{iv} \hat{\mathbf{R}}_{s_{iv}}^{-1} \mathbf{n}_{iv}(t)$. Based on $\mathbf{r}_3(t)$ in Eq. (3.88), we obtain the filtered data matrix

$$\mathbf{Y} = \hat{\mathbf{U}}_{iv} \hat{\mathbf{A}}_{iv} \quad (3.89)$$

which is recognised as the optimal IV-WSF data matrix $\mathbf{Y}_{iv}$. Thus, it is concluded that the optimal IV-WSF method, with $\mathbf{s}_{iv}(t)$ being obtained under one of the above conditions, coincides with the PFML-3 method with $\mathbf{r}_3(t)$ being obtained by certain spatio-temporal filtering of $\mathbf{s}_{iv}(t)$, as described in Eqs. (3.86-87). Note that, $\tilde{\mathbf{A}}$ in $\mathbf{r}_3(t)$ which is obtained by Eq. (3.88) is data dependent. In addition, it is suggested that, when additional information of the signal sources or noise is available, there may exist a better choice of the filter matrix $\mathbf{H}$ (than the one in Eq. (3.87)) which, by filtering $\mathbf{s}_{iv}(t)$ by $\mathbf{H}$ in Eq. (3.86), gives $\mathbf{r}_3(t)$ with higher signal to noise ratio. For instance, if approximate information of source directions is available, a beamspace pre-processor based on this prior knowledge may be applied in place of (or in combination with) $\hat{\mathbf{U}}_{iv}$ in Eq. (3.87). By doing so, improved noise rejection capability of $\mathbf{H}$ may be gained. Note also that, due to the generality of the form Eq. (3.67), the optimal IV-WSF method can still be cast as the PFML-3 method for the case where the signal component in $\mathbf{s}_{iv}(t)$ is not fully correlated with $\mathbf{s}(t)$, by a suitable choice of $\tilde{\mathbf{A}}$ and the noise term $\tilde{\mathbf{n}}(t)$ in Eq. (3.67).

In addition to the above discussion, some interesting remarks are next given.

1. It was shown in [19] that, as performed in Eq. (3.81), the use of the data covariance matrix $\mathbf{R}$ (or its consistent estimate) in place of the noise covariance matrix $\mathbf{R}_n$ (or its consistent estimate) does not affect the asymptotic properties of the DOA estimates. With the previously shown relation between the IV-based methods and the PFML-3 method, it is seen that, without affecting the asymptotic properties, the matrix $\hat{\mathbf{R}}$ in the optimal IV-WSF cost function Eq. (3.81) may also be replaced by $\hat{\tilde{\mathbf{R}}}_n (= \hat{\mathbf{R}} - \mathbf{Y}_{iv}\mathbf{Y}_{iv}^H$ in this case). On the other hand, $\hat{\mathbf{R}}$ may as well be used in place of $\hat{\tilde{\mathbf{R}}}_n$ in the PFML-3 cost function Eq. (3.71). The use of $\hat{\mathbf{R}}$ instead of $\hat{\tilde{\mathbf{R}}}_n$ has the advantage of being slightly simpler to compute. However, as noted in Chapter 6, the use of $\hat{\tilde{\mathbf{R}}}_n$ instead of $\hat{\mathbf{R}}$ in the best-case PFML-3 method ($\tilde{\mathbf{n}}(t) = \mathbf{0}$) gives slightly improvement in the estimation accuracy at low SNR. It also results in faster convergence rate when the IQML algorithm is used to compute the DOA estimates. Because the IV-based methods are related to the PFML-3 method as shown before, these improvements for the IV-based methods when using $\hat{\tilde{\mathbf{R}}}_n$ instead of $\hat{\mathbf{R}}$ are also expected.

2. Because the PFML-3 method does not assume any prior knowledge of $\mathbf{T}$, its performance is expected to be inferior to those of the DEML and PFML-2 methods when $\mathbf{\Gamma}_s$ has the structure shown in Eq. (3.31), and $s_{nc}(t)$ is composed of uncorrelated signals. This remark will be confirmed by results from the performance study, discussed in Chapter 6.

3. As for the case of the DEML and PFML-2 methods, the PFML-3 method is applicable when the noise covariance structure is not known. However, unlike the DEML and PFML-2 methods, the number of source directions which can be estimated by the PFML-3 method is less than the number of array sensors.

3.2.4 PFML DOA estimation with the type-4 pre-filtering reference signal

The type-4 pre-filtering reference signal is given by

$$\begin{aligned} \mathbf{r}_4(t) &= \tilde{\mathbf{A}}\mathbf{s}_{\text{nc}}(t) + \mathbf{Q}_n\mathbf{n}(t) \\ &\triangleq \begin{bmatrix} r_{4;1}(t) & r_{4;2}(t) & \cdots & r_{4;d_s}(t) \end{bmatrix}^T \end{aligned} \quad (3.90)$$

where the $d_s \times d_s$ matrix $\tilde{\mathbf{A}}$ is assumed to be of full rank. The complete knowledge of $\tilde{\mathbf{A}}$ may not be available. However, under a certain situation to be described in Chapter 4, $\tilde{\mathbf{A}}$ may be well approximated as a diagonal or block-diagonal matrix. Let's assume at the moment that $\mathbf{Q}_n\mathbf{n}(t) = \mathbf{0}$. With this, the formulations of the DOA estimators, based on three different configurations of $\tilde{\mathbf{A}}$, $\mathbf{\Gamma}_s$, and $\mathbf{s}_{\text{nc}}(t)$, are considered below.

CASE-1: $\tilde{\mathbf{A}}$ is an unknown full-rank matrix; $\mathbf{s}_{\text{nc}}(t)$ is composed of non-coherent signals, and $\mathbf{\Gamma}_s$ is of full rank.

In this case, $\mathbf{r}_4(t)$ is a special case of $\mathbf{r}_2(t)$ with $\tilde{\mathbf{n}}(t) = \mathbf{0}$. Thus, with the Unitary filter being employed in the pre-filtering step, the PFML-3 estimator should be applied.

CASE-2: $\tilde{\mathbf{A}} = \check{\mathbf{A}}$, where $\check{\mathbf{A}}$ is an unknown diagonal matrix; $\mathbf{s}_{\text{nc}}(t)$ is composed of uncorrelated signals, and $\mathbf{\Gamma}_s$ has the structure given by Eq. (3.31).

In this case, either the DEMML estimator (with the Nulling filter) or the PFML-2 estimator (with the Matched filter) may be applied. However, in practice when $\tilde{\mathbf{A}}$ is only approximately diagonal or $\mathbf{Q}_n\mathbf{n}(t) \neq \mathbf{0}$, the PFML-2 estimator is preferred because, from empirical study, the use of the DEMML estimator in this case generally results in poorer DOA estimates than those obtained by the PFML-2 estimator.

CASE-3: $\tilde{\mathbf{A}} = \check{\mathbf{A}}$, where $\check{\mathbf{A}}$ is a block-diagonal matrix with known structure, and $\mathbf{s}_{\text{nc}}(t)$ is composed of uncorrelated signals. The matrix $\mathbf{\Gamma}_s$ is composed of block matrices.

Let $\check{\mathbf{A}}$ be written as

$$\tilde{\mathbf{A}} = \begin{bmatrix} \tilde{\mathbf{A}}_1 & & & \mathbf{0} \\ & \tilde{\mathbf{A}}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \tilde{\mathbf{A}}_{K_c} \end{bmatrix} \quad (3.91)$$

where $\tilde{\mathbf{A}}_k$ are the $\tilde{K}_k \times \tilde{K}_k$ matrices of known dimensions, and $\sum_{k=1}^{K_c} \tilde{K}_k = d_s$. Under this condition, the Unitary-Matched filter may be employed. With this filter type, $\mathbf{W}_{\text{pf}}(\tilde{t})$ is given by Eqs. (3.13-3.14), with

$$\mathbf{M}_k = \tilde{\mathbf{I}}_L \cdot [\mathbf{r}_{4;k}(0) \quad \mathbf{r}_{4;k}(1) \quad \cdots \quad \mathbf{r}_{4;k}(L-1)]^H, k = 1, \dots, K_c \quad (3.92)$$

where $\mathbf{r}_{4;k}$, $k = 1, \dots, K_c$, are obtained by dividing $\mathbf{r}_4(t)$ into K_c subvectors of length $\tilde{K}_k$ each, i.e.,

$$[\mathbf{r}_{4;1}^T \quad \mathbf{r}_{4;2}^T \quad \cdots \quad \mathbf{r}_{4;K_c}^T]^T = \mathbf{r}_4(t) \quad (3.93)$$

Based on this filter type, in a similar manner to the formulations of the PFML-2 and PFML-3 estimators, a ML DOA estimator can be formulated from each of the $N \times \tilde{K}_k$ filtered data matrices $\mathbf{Y}_k$, $k = 1, \dots, K_c$, which are defined by

$$[\mathbf{Y}_1 \quad \mathbf{Y}_2 \quad \cdots \quad \mathbf{Y}_{K_c}] = \mathbf{Y} \quad (3.94)$$

Because $\mathbf{Q}_n \mathbf{n}(t) = \mathbf{0}$, the conditional means of $\mathbf{Y}_k$, $k = 1, \dots, K_c$, are

$$\mathbb{E}\{\mathbf{Y}_k \mid \hat{\mathbf{R}}_{\mathbf{r}_4;k}, \mathbf{W}_{\text{pf};k} = \hat{\mathbf{R}}_{\mathbf{r}_4;k}^{-1/2}, \mathbf{A}_k(\boldsymbol{\theta}_k), \boldsymbol{\Gamma}_s\} = \mathbf{A}_k(\boldsymbol{\theta}_k) \mathbf{T}_k \quad (3.95)$$

$$\boldsymbol{\Gamma}_s = \begin{bmatrix} \boldsymbol{\Gamma}_{s;1} & & & \mathbf{0} \\ & \boldsymbol{\Gamma}_{s;2} & & \\ & & \ddots & \\ \mathbf{0} & & & \boldsymbol{\Gamma}_{s;K_c} \end{bmatrix} \quad (3.96)$$

$$\mathbf{T}_k = \boldsymbol{\Gamma}_{s;k} \mathbf{R}_{\text{nc};k} \tilde{\mathbf{A}}_k^H \mathbf{R}_{\mathbf{r}_4;k}^{-1} \hat{\mathbf{R}}_{\mathbf{r}_4;k}^{1/2} \quad (3.97)$$

$$\hat{\mathbf{R}}_{\mathbf{r}_4;k} = \frac{1}{L} \sum_{k=0}^{L-1} \mathbf{r}_{4;k}(t) \mathbf{r}_{4;k}^H(t) \quad (3.98)$$

where $\Gamma_{s;k}, k = 1, \dots, K_c$, are the $K_k \times \tilde{K}_k$ matrices, $\sum_{k=1}^{K_c} K_k = K$, and $\mathbf{R}_{r_{4;k}} = \mathbf{E}\{\mathbf{r}_{4;k}(t)\mathbf{r}_{4;k}^H(t)\}$, $\mathbf{R}_{nc;k} = \mathbf{E}\{\mathbf{s}_{nc;k}(t)\mathbf{s}_{nc;k}^H(t)\}$. The matrices $\tilde{\mathbf{A}}_k$ are as given by Eq. (3.91).

The $\tilde{K}_k \times 1$ vectors $\mathbf{s}_{nc;k}(t)$ and the $N \times K_k$ sub-response matrices $\mathbf{A}_k(\theta_k), k = 1, \dots, K_c$, are as given by

$$\begin{bmatrix} \mathbf{A}_1(\theta_1) & \mathbf{A}_2(\theta_2) & \cdots & \mathbf{A}_{K_c}(\theta_{K_c}) \end{bmatrix} = \mathbf{A}(\theta) \quad (3.99)$$

$$\begin{bmatrix} \mathbf{s}_{nc;1}^T(t) & \mathbf{s}_{nc;2}^T(t) & \cdots & \mathbf{s}_{nc;K_c}^T(t) \end{bmatrix}^T = \mathbf{s}_{nc}(t) \quad (3.100)$$

From Eqs. (2.3, 3.99), it is noted that $\theta = [\theta_1^T \ \theta_2^T \ \cdots \ \theta_{K_c}^T]^T$, where θ_k are the $K_k \times 1$ vectors. As before, the explicit expression of $\mathbf{A}_k$ as a function of θ_k will later be dropped. Based on $\mathbf{Y}_k$, the ML estimates of $\theta_k, k = 1, \dots, K_c$, are given by

$$\hat{\theta}_k = \arg \min_{\theta_k \in \Theta^{K_k \times 1}} f_{4;k}(\theta_k), \quad k = 1, \dots, K_c \quad (3.101)$$

where

$$f_{4;k}(\theta_k) = \text{tr}\{\mathbf{P}_{\hat{\mathbf{R}}_{n;k}^{-1/2}\mathbf{A}_k}^\perp \hat{\mathbf{R}}_{n;k}^{-1/2} \mathbf{Y}_k \mathbf{Y}_k^H \hat{\mathbf{R}}_{n;k}^{-1/2}\} \quad (3.102)$$

$$\hat{\mathbf{R}}_{n;k} = \hat{\mathbf{R}} - \mathbf{Y}_k \mathbf{Y}_k^H \quad (3.103)$$

The above estimator, which is based on the data filtered by the Unitary-Matched filter, will be referred to as the PFML-4 estimator. By assuming that $\mathbf{Q}_n \mathbf{n}(t) = \mathbf{0}$, the asymptotic analysis of the above estimator is similar to those of the PFML-2 and PFML-3 estimators. It can be seen that the above estimator is asymptotically unbiased. Under the assumptions that $\mathbf{Q}_n \mathbf{n}(t) = \mathbf{0}$, $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}$, with $\tilde{\mathbf{A}}$ and Γ_s as given by Eqs. (3.91, 3.96), and $\mathbf{s}_{nc}(t)$ is composed of uncorrelated signals, the asymptotic error covariance matrices of the PFML-4 DOA estimates $\hat{\theta}_k, k = 1, \dots, K_c$, are given by

$$\mathbf{E}\{(\hat{\theta}_k - \bar{\theta}_k)(\hat{\theta}_k - \bar{\theta}_k)^T\} = \frac{1}{2L} \text{Re}^{-1}\{(\bar{\mathbf{D}}_k^H \bar{\mathbf{R}}_{n;k}^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_{n;k}^{-1/2} \bar{\mathbf{A}}_k}^\perp \bar{\mathbf{R}}_{n;k}^{-1/2} \bar{\mathbf{D}}_k) \odot (\bar{\mathbf{T}}_k \bar{\mathbf{T}}_k^H)^T\}, \quad k = 1, \dots, K_c \quad (3.104)$$

where $\bar{\mathbf{A}}_k = \mathbf{A}_k(\bar{\boldsymbol{\theta}}_k)$ (see Eq. (3.99)),

$$\bar{\mathbf{T}}_k = \bar{\boldsymbol{\Gamma}}_{s;k} \mathbf{R}_{nc;k} \tilde{\mathbf{A}}_k^H \mathbf{R}_{r_4;k}^{-1/2} \quad (3.105)$$

$$\bar{\mathbf{D}}_k = \begin{bmatrix} \left. \frac{\partial \mathbf{a}_{k;1}}{\partial \theta_{k;1}} \right|_{\bar{\theta}_{k;1}} & \left. \frac{\partial \mathbf{a}_{k;2}}{\partial \theta_{k;2}} \right|_{\bar{\theta}_{k;2}} & \cdots & \left. \frac{\partial \mathbf{a}_{k;K_k}}{\partial \theta_{k;K_k}} \right|_{\bar{\theta}_{k;K_k}} \end{bmatrix} \quad (3.106)$$

and $[\theta_{k;1} \ \theta_{k;2} \ \cdots \ \theta_{k;K_k}]^T = \boldsymbol{\theta}_k$, $[\mathbf{a}_{k;1} \ \mathbf{a}_{k;2} \ \cdots \ \mathbf{a}_{k;K_k}] = \mathbf{A}_k$, $\bar{\mathbf{R}}_{n;k} = \mathbf{R} - \bar{\mathbf{Y}}_k \bar{\mathbf{Y}}_k^H$ with $\bar{\mathbf{Y}}_k = \bar{\mathbf{A}}_k \bar{\mathbf{T}}_k$.

When $\mathbf{Q}_n \mathbf{n}(t) \neq \mathbf{0}$, the bias term $\mathbf{E}_b$ exists in the conditional mean of $\mathbf{Y}$. For the case of the Unitary filter,

$$\mathbb{E}\{\mathbf{Y} | \hat{\mathbf{R}}_{r_4}, \mathbf{W}_{pf} = \hat{\mathbf{R}}_{r_4}^{-1/2}, \tilde{\mathbf{A}}, \mathbf{A}, \boldsymbol{\Gamma}_s\} = \mathbf{A} \mathbf{T} + \mathbf{E}_b \quad (3.107)$$

where

$$\mathbf{T} = \boldsymbol{\Gamma}_s \mathbf{R}_{nc} \tilde{\mathbf{A}}^H \mathbf{R}_{r_4}^{-1} \hat{\mathbf{R}}_{r_4}^{1/2} \quad (3.108)$$

$$\mathbf{E}_b = \mathbf{R}_n \mathbf{Q}_n^H \mathbf{R}_{r_4}^{-1} \hat{\mathbf{R}}_{r_4}^{1/2} \quad (3.109)$$

$$\hat{\mathbf{R}}_{r_4} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{r}_4(t) \mathbf{r}_4^H(t) \quad (3.110)$$

When $\tilde{\mathbf{A}}$ is only approximately diagonal or block-diagonal, the bias term $\mathbf{E}_b$ in the data filtered by either the Matched filter or the Unitary-Matched filter, is due to both $\mathbf{Q}_n \mathbf{n}(t)$ and $\check{\tilde{\mathbf{A}}}$, where $\check{\tilde{\mathbf{A}}} = \tilde{\mathbf{A}} - \tilde{\mathbf{A}}$ with $\tilde{\mathbf{A}}$ being diagonal or block-diagonal. If $s_{nc}(t)$ is composed of uncorrelated signals and $\tilde{\mathbf{A}}$ is diagonal, by using the Matched filter, the conditional means of the filtered data vectors $\mathbf{y}_k$, $k = 1, \dots, d_s$, are given by

$$\mathbb{E}\{\mathbf{y}_k | [\hat{\mathbf{R}}_{r_4}]_{k,k}, [\mathbf{W}_{pf}]_{k,k} = [\hat{\mathbf{R}}_{r_4}]_{k,k}^{-1/2}, \tilde{\mathbf{A}}, \check{\tilde{\mathbf{A}}}, \mathbf{A}, \boldsymbol{\Gamma}_s\} = \mathbf{A}_k \mathbf{t}_k + [\mathbf{E}_b]_k, \quad k = 1, \dots, d_s \quad (3.111)$$

where

$$t_k = \gamma_k [R_{nc}]_{k,k} [\tilde{A}^H]_{k,k} [R_{r_4}]_{k,k}^{-1} [\hat{R}_{r_4}]_{k,k}^{1/2} \quad (3.112)$$

$$[E_b]_k = A \Gamma_s R_{nc} [\tilde{A}^H]_k [R_{r_4}]_{k,k}^{-1} [\hat{R}_{r_4}]_{k,k}^{1/2} + R_n [Q_n^H]_k [R_{r_4}]_{k,k}^{-1} [\hat{R}_{r_4}]_{k,k}^{1/2} \quad (3.113)$$

and Γ_s , γ_k , are as described by Eq. (3.31).

For the case where $s_{nc}(t)$ is composed of uncorrelated signals and $\tilde{A}$ is block-diagonal with the structure shown in Eq. (3.91), the conditional means of the filtered data matrices Y_k , $k = 1, \dots, K_c$, obtained by the Unitary-Matched filtering, are given by

$$E\{Y_k | \hat{R}_{r_4;k}, W_{pf;k} = \hat{R}_{r_4;k}^{-1/2}, \tilde{A}, \tilde{A}^*, A, \Gamma_s\} = A_k T_k + E_{b;k}, \quad k = 1, \dots, K_c \quad (3.114)$$

$$T_k = \Gamma_{s;k} R_{nc;k} \tilde{A}_k^H R_{r_4;k}^{-1} \hat{R}_{r_4;k}^{1/2} \quad (3.115)$$

$$E_{b;k} = A \Gamma_s R_{nc} \tilde{A}_k^* R_{r_4;k}^{-1} \hat{R}_{r_4;k}^{1/2} + R_n Q_{n;k}^* R_{r_4;k}^{-1} \hat{R}_{r_4;k}^{1/2} \quad (3.116)$$

where Γ_s and $\Gamma_{s;k}$ are as described by Eq. (3.96). The matrices $Q_{n;k}$ and $\tilde{A}_k$ are given by

$$[Q_{n;1} \quad Q_{n;2} \quad \dots \quad Q_{n;K_c}]^T = Q_n \quad (3.117)$$

$$[\tilde{A}_1 \quad \tilde{A}_2 \quad \dots \quad \tilde{A}_{K_c}]^T = \tilde{A} \quad (3.118)$$

where $Q_{n;k}$, $k = 1, \dots, K_c$, are the $N \times \tilde{K}_k$ matrices, and $\tilde{A}_k$, $k = 1, \dots, K_c$, are the $d_s \times \tilde{K}_k$ matrices.

Because Eqs. (3.107, 3.111, 3.114) have the same mathematical form, the following discussion based on the Unitary filter case can also be applied to the other two filter cases. From Eq. (3.107), without the knowledge of the bias term E_b , the DOA estimates using the filtered data matrix Y are expected to be biased, even for large L . When at least an approximate E_c to E_b is available, the estimation bias can be reduced by substituting Y by the bias corrected data matrix

$$Y_{bc} = Y - E_c \quad (3.119)$$

The bias term in Y_{bc} is equal to $(E_b - E_c)$, which can be smaller than E_b (in a matrix norm sense) if E_c is close to E_b . In the next chapter, we will show how E_c can be obtained in practice.

An exact knowledge of E_b is generally not available. An immediate question is that how the estimation performance is affected by this bias term. Only the case of no correction term E_c needs to be considered. In the following, we provide the asymptotic bias analysis for the estimate of θ . Results from the following discussion, which deals with the data matrix Y with no bias correction term, can be equally applied to Y_{bc} by substituting E_b by $(E_b - E_c)$.

Asymptotic analysis

The asymptotic condition in the following theorem on the asymptotic estimation bias is referred to the case where the bias term E_b is sufficiently small, and the additive noise term in Y , denoted by E_n , is negligible when compared with E_b , all in the Frobenius norm sense, say. Note that, the condition as described should be interpreted in a probabilistic sense. Let $E_n = \tilde{E}_n / \sqrt{L}$, and $E_b = \tilde{E}_b / \sqrt{\tilde{L}}$, where $\tilde{E}_n$ and $\tilde{E}_b$ are independent of L and $\tilde{L}$. The described asymptotic condition is met by letting $\tilde{L} \rightarrow \infty$ with L approaching infinity faster than $\tilde{L}^\dagger$. Under this condition, the asymptotic bias of the DOA estimate $\hat{\theta}$ by the PFML-3 estimator with $r(t) = r_4(t)$ is given below.

Theorem 3.5: Under the described condition, the asymptotic bias of the PFML-3 estimate $\hat{\theta}$ with $r(t) = r_4(t)$ is given by

$$\tilde{\theta} = \text{Re}^{-1} \{ (\bar{D}^H \bar{R}_n^{-1/2} P_{\bar{R}_n^{-1/2} \bar{A}}^\perp \bar{R}_n^{-1/2} \bar{D}) \odot (\bar{T} \bar{T}^H)^T \}. \text{Re} \{ \{ (\bar{D}^H \bar{R}_n^{-1/2} P_{\bar{R}_n^{-1/2} \bar{A}}^\perp \bar{R}_n^{-1/2}) \odot (\bar{E}_b \bar{T}^H)^T \} I_N \} \quad (3.120)$$

where $\tilde{\theta} = \hat{\theta} - \bar{\theta}$, $\bar{E}_b = R_n Q_n^H R_{r_4}^{-1/2}$, $\bar{T} = \bar{\Gamma}_s R_{nc} \tilde{A}^H R_{r_4}^{-1/2}$, and $\bar{R}_n = R - \bar{Y} \bar{Y}^H$ with

[†] The idea of this asymptotic condition setting is adapted from [45].

$$\bar{Y} = \bar{A}\bar{T} + \bar{E}_b \quad (3.121)$$

Proof: See Appendix F.

For sufficiently small E_b , by letting $L \rightarrow \infty$ at the same rate as $\tilde{L}$, the asymptotic error covariance matrix of the estimate $\hat{\theta}$ using the PFML-3 estimator with $r(t) = r_4(t)$ can be obtained as a combination of $\tilde{\theta}\tilde{\theta}^T$, and the right hand side of Eq. (3.72) which is based on $\bar{T} = \bar{\Gamma}_s R_{nc} \tilde{A}^H R_{r_4}^{-1/2}$, and $\bar{R}_n = R - \bar{Y}\bar{Y}^H$ with $\bar{Y}$ as given by Eq. (3.121).

For the case of the PFML-2 estimator with $r(t) = r_4(t)$, applied under the assumptions that $s_{nc}(t)$ is composed of uncorrelated signals and $\tilde{A}$ is diagonal, the asymptotic bias of the estimate $\hat{\theta}_k$ can be obtained from

$$\begin{aligned} \tilde{\theta}_k = \text{Re}^{-1} \{ (\bar{D}_k^H \bar{R}_{n;k}^{-1/2} P_{\bar{R}_{n;k}^{-1/2} \bar{A}_k}^\perp \bar{R}_{n;k}^{-1/2} \bar{D}_k) \odot (\bar{t}_k \bar{t}_k^H)^T \} \cdot \text{Re} \{ \{ (\bar{D}_k^H \bar{R}_{n;k}^{-1/2} P_{\bar{R}_{n;k}^{-1/2} \bar{A}_k}^\perp \bar{R}_{n;k}^{-1/2}) \odot \\ ([\bar{E}_b]_k \bar{t}_k^H)^T \} I_N \} \quad (3.122) \end{aligned}$$

where $\tilde{\theta}_k = \hat{\theta}_k - \bar{\theta}_k$,

$$[\bar{E}_b]_k = \bar{A} \bar{\Gamma}_s R_{nc} [\tilde{A}^H]_k [R_{r_4}]_{k,k}^{-1/2} + R_n [Q_n^H]_k [R_{r_4}]_{k,k}^{-1/2} \quad (3.123)$$

$$\bar{t}_k = \bar{\gamma}_k [R_{nc}]_{k,k} [\tilde{A}^H]_{k,k} [R_{r_4}]_{k,k}^{-1/2} \quad (3.124)$$

and $\bar{R}_{n;k} = R - \bar{y}_k \bar{y}_k^H$ with $\bar{y}_k = \bar{A}_k \bar{t}_k + [\bar{E}_b]_k$.

The above asymptotic formula can be derived in a similar manner to the derivation of Eq. (3.120). In addition, as in the case of the PFML-3 method, for sufficiently small $[E_b]_k$, the asymptotic error covariance matrix of the estimate $\hat{\theta}_k$ using the PFML-2 estimator can be obtained as a combination of $\tilde{\theta}_k \tilde{\theta}_k^T$, and the right hand side of Eq. (3.59) which is based on $\bar{R}_{n;k} = R - \bar{y}_k \bar{y}_k^H$ with $\{[R_{nc}]_{k,k} (\text{corr}_{r_2;k})^2\} \gamma_k \gamma_k^H$ being replaced by $\bar{t}_k \bar{t}_k^H$.

For the case of the PFML-4 estimator with $\mathbf{r}(t) = \mathbf{r}_4(t)$, applied under the assumptions that $\mathbf{s}_{nc}(t)$ is composed of uncorrelated signals and $\bar{\mathbf{A}}$, Γ_s are given, respectively, by Eq. (3.91) and Eq. (3.96), the asymptotic bias of the estimate $\hat{\boldsymbol{\theta}}_k$ can be obtained from

$$\tilde{\boldsymbol{\theta}}_k = \text{Re}^{-1} \{ (\bar{\mathbf{D}}_k^H \bar{\mathbf{R}}_{n;k}^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_{n;k}^{-1/2} \bar{\mathbf{A}}_k}^\perp \bar{\mathbf{R}}_{n;k}^{-1/2} \bar{\mathbf{D}}_k) \odot (\bar{\mathbf{T}}_k \bar{\mathbf{T}}_k^H)^T \} \cdot \text{Re} \{ \{ (\bar{\mathbf{D}}_k^H \bar{\mathbf{R}}_{n;k}^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_{n;k}^{-1/2} \bar{\mathbf{A}}_k}^\perp \bar{\mathbf{R}}_{n;k}^{-1/2}) \odot (\bar{\mathbf{E}}_{b;k} \bar{\mathbf{T}}_k^H)^T \} \mathbf{I}_N \} \quad (3.125)$$

where $\tilde{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k$; $\bar{\mathbf{T}}_k$ is as given by Eq. (3.105),

$$\bar{\mathbf{E}}_{b;k} = \bar{\mathbf{A}} \bar{\Gamma}_s \mathbf{R}_{nc} \bar{\mathbf{A}}_k^* \mathbf{R}_{r_4;k}^{-1/2} + \mathbf{R}_n \mathbf{Q}_{n;k}^* \mathbf{R}_{r_4;k}^{-1/2} \quad (3.126)$$

and $\bar{\mathbf{R}}_{n;k} = \mathbf{R} - \bar{\mathbf{Y}}_k \bar{\mathbf{Y}}_k^H$ with $\bar{\mathbf{Y}}_k = \bar{\mathbf{A}}_k \bar{\mathbf{T}}_k + \bar{\mathbf{E}}_{b;k}$.

As in the case of the PFML-2 and PFML-3 methods, the asymptotic error covariance matrix of the estimate $\hat{\boldsymbol{\theta}}_k$ using the PFML-4 estimator can be obtained as a combination of $\tilde{\boldsymbol{\theta}}_k \tilde{\boldsymbol{\theta}}_k^T$, and the right hand side of Eq. (3.104) which is based on $\bar{\mathbf{R}}_{n;k} = \mathbf{R} - \bar{\mathbf{Y}}_k \bar{\mathbf{Y}}_k^H$ with $\bar{\mathbf{Y}}_k = \bar{\mathbf{A}}_k \bar{\mathbf{T}}_k + \bar{\mathbf{E}}_{b;k}$.

Finally, the summary of the applied filter types and the resulting DOA estimators for the four pre-filtering reference signals is given in Table 3.1.

Reference signal	Correlatedness of $s_{nc}(t)$	Filter type	Estimator
type-1	non-coherent	Nulling filter	DEML
type-2	uncorrelated	Matched filter	PFML-2
type-3	non-coherent	Unitary filter	PFML-3
type-4	uncorrelated/ non-coherent/ mixed correlated and uncorrelated	Matched filter/ Unitary filter/ Unitary-Matched filter	PFML-2/ PFML-3/ PFML-4

Table 3.1: Summary of the applied filters and PFML DOA estimators for the four pre-filtering reference signals.

3.3 Some practical aspects of the PFML methods

So far we have not yet detailed how to choose and obtain the pre-filtering reference signal $\mathbf{r}(t)$, required for the formation of the filter in the pre-filtering step. In this section, we provide some practical examples of how these reference signals can be obtained.

In some application areas such as digital radio communications, the noise-free type-1 pre-filtering reference signal $\mathbf{r}_1(t)$ may be obtained in a form of the training sequence. However, in practice, the known sequence may not be well synchronised with the incoming signal sequence. As a result, the known signal may not be fully correlated with the signal component in the received data. Similar problem can also arise in the presence of the carrier frequency error. The known type-1 pre-filtering reference signal which is not fully correlated with the received signal component is more accurately modelled as the type-2 pre-filtering reference signal $\mathbf{r}_2(t)$. When this is the case, the PFML-2 method should be applied instead of the DEML method.

Although both the DEML and PFML-2 methods based on, respectively, $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$, offer asymptotically unbiased DOA estimates as well as superior performance to

the PFML-3 and standard DML and SML (or optimal WSF) methods, the duration of the known sequence of the pre-filtering reference signal may, however, be limited. This is the case in digital radio communications where a short training sequence is preferred to avoid bandwidth inefficiency. Furthermore, in many other problems (systems), neither $r_1(t)$ nor $r_2(t)$ is available. Examples of which are passive sonar systems, and medical source localisation and imaging problems [46].

In [19], the IV-based DOA estimator was proposed. The instrumental variable $s_{iv}(t)$ is obtained as a delayed version of the observed data vector $\mathbf{x}(t)$. This temporal instrumental variable approach is based on the assumption that the additive noise is temporally uncorrelated while the signal $s(t)$ is at least to some extent temporally correlated. The optimal IV-WSF method in [19] was shown in Section 3.2.3 to coincide with the PFML-3 method, when certain spatio-temporal filtering is applied to $s_{iv}(t)$. Both the optimal IV-WSF and PFML-3 methods generally yield inferior result to those of the DEML and PFML-2 methods, when all mentioned methods are applicable. In addition, the performance of the optimal IV-WSF method degrades as the temporal correlatedness of $s(t)$ decreases. It was suggested in [19] that oversampling of the observed data can be used to increase the signal correlation at the cost of increasing the noise level in $s_{iv}(t)$. Unfortunately, as demonstrated in [19], the data sampling rate can not be increased indefinitely without, at some point, deteriorating the DOA estimation performance. In addition to this drawback, in some application areas such as radio communications, this solution may not be applicable due to the presence of signals from adjacent frequency bands.

In addition to the above TIV approach, the instrumental variable $s_{iv}(t)$ can be alternatively obtained as the output of the secondary array [18, 20, 47]. This SIV approach is based on the assumption that the additive noise in the primary array, where the actual DOA estimation takes place, and the additive noise in the secondary array, from which the instrumental variable signal $s_{iv}(t)$ is obtained, are uncorrelated while the

signal components in both arrays are highly or fully correlated. The advantage of this SIV approach over the TIV approach is that the former does not rely on the temporal correlatedness of $s(t)$. An obvious drawback of the SIV approach is that it requires the secondary array to supply $s_{iv}(t)$. The secondary array, however, is not used to perform DOA estimation. Although improved efficiency of the SIV approach can be obtained by averaging the DOA estimates using the two arrays, one at a time [47], the scheme is sub-optimal because the two arrays are not used simultaneously as a single array. In addition, the maximum number of source directions which can be estimated by the SIV method is smaller than that achievable when both arrays are used as a single array.

The type-4 pre-filtering reference signal $r_4(t)$ has also been considered in Section 3.2.4. However, it has not yet been mentioned how $r_4(t)$ can be obtained in practice. The detail of acquiring $r_4(t)$ in practice is the subject of Chapter 4. Briefly, $r_4(t)$ may be obtained from the observed data vector $x(t)$ by a suitable spatial filtering (beamforming) scheme. By pointing the beamformer toward each of the signal directions, $s(t)$ can be extracted from $x(t)$. Of course, the exact knowledge of source directions is not available in our case. However, under a certain source scenario, it will be shown in Chapter 4 how $s(t)$ can be effectively estimated from $x(t)$. At the moment, some interesting features of this new beamforming-based approach should be noted. For instance, it does not require a priori the knowledge of the signal waveforms. Unlike the SIV DOA estimation approach, the approach considered in Chapter 4 does not need the secondary array. In addition, $s(t)$ is not required to be temporally correlated, as demanded by the TIV DOA estimator.

3.4 Conclusions

In this chapter, we propose the (temporal) pre-filtering approach to narrow-band source direction estimation. The observed data is first temporally filtered by one of the four proposed filtering operators. The filtered data is subsequently applied to a DOA

estimator. Four different pre-filtering reference signals, from which the filter coefficients are obtained, have been considered. The ML DOA estimators have been formulated based on the filtered data using each of the four filter types. The relation of these PFML estimators with the ML method with known signal waveforms, and the IV-based DOA estimators, has been addressed. The asymptotic properties of the PFML DOA estimators have been analysed. However, detailed performance comparison of the described methods will be deferred until Chapter 6. Before that, in the next chapter we develop the beamforming-based PFML methods, which are practical realisations of the PFML methods based on the type-4 pre-filtering reference signal $\mathbf{r}_4(t)$. In Chapter 5, the IQML algorithm will be studied and modified to allow for efficient computation of the ML DOA estimates. Thus, as well as the analytical results based on the asymptotic analytical study described in this chapter, simulation results using the IQML algorithm will also be provided in Chapter 6 to validate the analytical results.

Before leaving this chapter, we summarise some of the important properties of the PFML DOA estimation methods as follows:

- The DEML, PFML-2, and PFML-4 methods can estimate more non-coherent source directions than most existing DOA estimation methods.
- All PFML methods are applicable in an unknown spatially correlated noise field.
- Performance of these PFML methods are generally superior to those of the DML and SML methods, particularly under the mixed coherent-uncorrelated source scenario.

Chapter 4

Beamforming-based Pre-Filtering Maximum Likelihood DOA Estimators

4.1 Adaptive beamforming techniques for signal waveform estimation

Classical and adaptive beamforming techniques have long been developed and widely applied [1-4, 27]. A beamformer is used to enhance the quality of the received desired signal by means of spatial filtering [27]. In this section, we develop an adaptive beamforming procedure, tailored to extract from the received array data the type-4 pre-filtering reference signal $\mathbf{r}_4(t)$, to be used under the PFML framework for DOA estimation. As in other beamforming methods, the goal of our procedure is, subject to a certain criterion, to obtain at the beamforming output an estimate of the desired signal $\hat{s}_d(t) = \mathbf{w}^H \mathbf{x}(t)$, where $\mathbf{w}$ is the beamforming weight vector. In our problem, the desired signal is $s_{nc;k}(t)$, $k = 1, \dots, d_s$, where $s_{nc;k}(t)$ is the k th component of $\mathbf{s}_{nc}(t)$. The signal $\mathbf{s}_{nc}(t)$ is in turn related to $\mathbf{s}(t)$ by Eq. (3.7). Thus, for $d_s > 1$, more than one beamformer is required.

One of the main issues in designing a beamformer is the choice of a beamforming criterion. In the following, we provide a brief discussion of three commonly known criteria and their suitability for our problem.

a) Maximum Signal-to-Interference plus Noise Ratio (MSINR) criterion

Under this criterion, the weight vector $\mathbf{w}$ is chosen so as to maximise the output SINR. It is given by [2]

$$\mathbf{w} = \mu \mathbf{R}_{\text{in}}^{-1} \mathbf{a}_d \quad (4.1)$$

where $\mathbf{a}_d$ is the array response vector which corresponds to the desired signal, $\mathbf{R}_{\text{in}}$ is the interference plus noise covariance matrix, and μ is a scaling factor. To keep a unity gain in the look direction, the scaling factor μ should be selected as

$$\mu = (\mathbf{a}_d^H \mathbf{R}_{\text{in}}^{-1} \mathbf{a}_d)^{-1} \quad (4.2)$$

It is noted that the described optimality of the method is claimed under the condition that the exact interference plus noise covariance matrix $\mathbf{R}_{\text{in}}$, or its scaled version, is used in Eq. (4.1). When $\mathbf{R}_{\text{in}}$ is not known but, in the absence of the desired signal, the array data snapshots $\mathbf{x}_{\text{in}}(t)$ consisting of only interference and noise are available, the following consistent estimate may be applied instead.

$$\hat{\mathbf{R}}_{\text{in}} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{x}_{\text{in}}(t) \mathbf{x}_{\text{in}}^H(t) \quad (4.3)$$

Whether or not this method can be applied to our problem depends on the availability of either $\mathbf{R}_{\text{in}}$ or $\mathbf{x}_{\text{in}}(t)$, as well as the knowledge of $\mathbf{a}_d$. Although in the DOA estimation problem the exact knowledge of $\mathbf{a}_d$ is not available (otherwise there is no need for DOA estimation), approximate information for the desired signal direction (and for $\mathbf{a}_d$ through the assumed known array manifold) may be sufficient. This is the case when there is no strong interference coming from the direction close to that of $s_d(t)$. In many cases, however, neither $\mathbf{R}_{\text{in}}$ nor $\mathbf{x}_{\text{in}}(t)$ is known.

b) Linear Minimum Mean-Square Error (LMMSE) criterion

Under this beamforming criterion, the objective is to estimate $\hat{s}_d(t)$ which minimises the MSE given by

$$\text{MSE} \triangleq E\{|\hat{s}_d(t) - s_r(t)|^2\} \quad (4.4)$$

where $s_r(t)$ is known as the reference signal. Under the LMMSE criterion, the array weight vector is given by [2, 13]

$$\mathbf{w} = \mathbf{R}^{-1} \mathbf{r}_{xs_r} \quad (4.5)$$

where $\mathbf{r}_{xs_r} = E\{\mathbf{x}(t)s_r^*(t)\}$. When $\mathbf{R}$ and $\mathbf{r}_{xs_r}$ are not known, their consistent estimates are applied instead. In addition, to obtain, up to a scaling factor, an exact LMMSE estimate of $s_d(t)$, the reference signal $s_r(t)$ is a scaled version of $s_d(t)$. In practice, however, it is generally sufficient for $s_r(t)$ to be highly correlated with the desired signal.

From the above discussion, the major drawback of this criterion is the requirement to know $s_d(t)$, or a signal which is highly correlated with $s_d(t)$. It is arguable that if such a signal is known, it can be directly used in either the DEMML or PFML-2 method. However, without prior knowledge of $s_d(t)$, the required reference signal can sometimes be obtained from the beamforming output itself. An example of this is in digital radio communication systems employing a decision feedback beamformer or equaliser [48-49]. The decision feedback system generally consists of two parts. The first part is a beamformer (or equaliser) whose purpose is to improve the quality of the received signal, e.g., to increase the SINR. The beamforming output signal is then passed through a decision device which, for sufficiently high SINR, can give at its output a scaled replica of $s_d(t)$. The signal at the output of the decision device is subsequently used as the reference signal $s_r(t)$ for updating the beamforming weight vector, e.g., by using Eq. (4.5). Note that a good initial beamforming weight vector is generally required. This can be obtained, for instance, by using a known training signal.

c) *Linearly Constrained Minimum Variance (LCMV) criterion*

The LCMV criterion chooses the weight vector to minimise the average beamforming output power, subject to a unity (or constant) gain in the look direction,

which is generally the direction of the desired signal. Under this criterion, $\mathbf{w}$ is obtained as a solution to the following linearly constrained optimisation problem [27]

$$\mathbf{w} = \arg \min_{\mathbf{w}} \mathbf{w}^H \mathbf{R} \mathbf{w} \quad (4.6)$$

subject to

$$\mathbf{w}^H \mathbf{a}_d = 1 \quad (4.7)$$

The solution is given by

$$\mathbf{w} = \mathbf{R}^{-1} \mathbf{a}_d (\mathbf{a}_d^H \mathbf{R}^{-1} \mathbf{a}_d)^{-1} \quad (4.8)$$

In practice, the ensemble averaged array covariance matrix $\mathbf{R}$ is generally not available and has to be replaced by a consistent estimate. We note on passing that, although a single constraint is imposed on $\mathbf{w}$ in Eq. (4.7), it may sometimes be useful to apply instead a set of linear constraints [27].

From Eq. (4.8), the knowledge of the response vector $\mathbf{a}_d$ is required. Although for the MSINR beamformer $\mathbf{a}_d$ need not be precisely known, in the LCMV beamformer, slight difference between the constraint vector in Eq. (4.7) and the true array response vector of the desired signal can lead to severe performance degradation. In addition, the presence of the interfering signals which are correlated or coherent with the desired signal results in similar performance degradation. In either of these two cases, in spite of the constraint Eq. (4.7), the desired signal is partially cancelled to minimise the beamforming output power, a phenomenon known as the signal cancellation effect [21, 27]. Despite these drawbacks, the LCMV beamformer has a practical advantage over the earlier methods because, apart from $\mathbf{a}_d$, its only other requirement is the array covariance matrix $\mathbf{R}$, which can be estimated from the observed data. In the next section, a modified LCMV beamformer which solves the signal cancellation problem is provided. This, with careful application of the beamformer, makes signal waveform estimation possible without the need for the reference signal $s_r(t)$ or the desired signal-free data vector $\mathbf{x}_{in}(t)$.

Although these adaptive beamformers generally have superior interference rejection capability to the classical non-adaptive beamformer, they can effectively reduce interference and noise only up to a certain extent. For instance, under finite input SINR and a limited number of sensors, it is clear that complete recovery of the desired signal from noisy data is not possible by any of the described beamformers. In addition, the closer the DOAs of the interferences are to that of the desired signal, the more difficult it is for a beamformer to reject these closely-spaced interferences. Thus, in the following section, we consider a scenario where the required type-4 pre-filtering reference signal $\mathbf{r}_4(t)$ can be effectively obtained from the observed data by a suitable spatial filtering scheme. Next, to obtain $\mathbf{r}_4(t)$ from the observed data, a beamforming procedure based on the classical and LCMV beamforming methods is proposed. One of the important features of the procedure is that it is not susceptible to the signal cancellation effect due to pointing error and correlated signals. The pre-filtering reference signal $\mathbf{r}_4(t)$, obtained by the mentioned beamforming procedure, is then applied under the PFML framework for DOA estimation. Note that, the signal and noise model as stated in Chapter 3 is also applied in subsequent sections.

4.2 Beamforming-based PFML DOA estimation in the well-separated multiple cluster source scenario

4.2.1 The multiple cluster source model

Due to the limitation of a beamformer in suppressing strong interference whose angular location is close to that of the desired signal, in this section we restrict the use of a beamformer in obtaining $\mathbf{r}_4(t)$ to the following source scenario.

First, it is assumed that $\mathbf{s}(t)$ is composed of groups of one or more signals, with two signals from different groups being uncorrelated. If the signals in each group are fully correlated, $\mathbf{s}(t)$ can be written by Eq. (3.7) with $\mathbf{\Gamma}_s$ as given by Eq. (3.31). When

the signals in each group are only partially correlated such that $d_s = K$, the matrix Γ_s is either diagonal or composed of block matrices.

Next, the angular spacing of two uncorrelated signal sources from different groups is required not to be less than one array beamwidth. However, the angular spacing of any two correlated signal sources within the same group can be less than one array beamwidth. This is the scenario where, for each group of correlated signal sources, the signal source directions are clustered around a nominal direction of arrival[†], as shown in Figure 4.1. The number of correlated signal sources in each cluster must not exceed the maximum number of sources identifiable by the array (see [50-51] for discussion of the conditions for the source DOAs to be identifiable). For arbitrary signal phases, with a uniform linear array, the maximum number of identifiable source directions is $N/2$ for even N , and $(N - 1)/2$ for odd N [50]. Note, however, that the total number of signal sources due to all clusters are allowed to exceed this identifiability limit.

Before we continue the discussion, it is necessary to give practical justification for the described cluster source scenario. One of the application areas where the described signal model is well suited, is in mobile radio communications. Currently, mobile communication systems reuse the radio frequency bands by dividing a service area into small cells [52]. Two cells using the same frequency band are well separated to avoid co-channel interference. However, with an increasing demand for supporting more users and bandwidth requirement, the application of array signal processing techniques has been considered for improving bandwidth efficiency, by reusing the same frequency band within the same cell [53-56]. By dividing the area surrounding a base-station antenna array into several angular sectors, the signals from different mobiles, occupying the same frequency band, can be separated by using an antenna array employing a

[†] This assumption can be relaxed by treating a signal whose direction is well separated from other correlated signal directions, as belonging to another source cluster. The scenario in Figure 4.1 is, however, assumed for ease of presentation.

suitable beamforming scheme. With proper bandwidth management and handover operation, two mobiles using the same frequency band can be kept well-separated angularly. In addition, for a mobile sufficiently far from the base station, the DOAs (at the base station) of the multipath mobile signals, resulting from reflection near the mobile, are generally confined to a narrow angular width [52]. Although this cluster of multipath signals may be more effectively modelled as being generated from a distributed source [55], it may also be well approximated by a small number of point sources, as suggested in [53]. Thus, this mobile radio propagation scenario may be represented by the described cluster source model.

Another practical example of the cluster source model is in radar low-angle estimation and tracking [26]. In this problem, the estimation of the target object direction is made difficult by the presence of the mirror object, created by the reflection of the radar signal from the sea surface. For the case of a smooth reflecting surface, the scenario can also be described by the model.

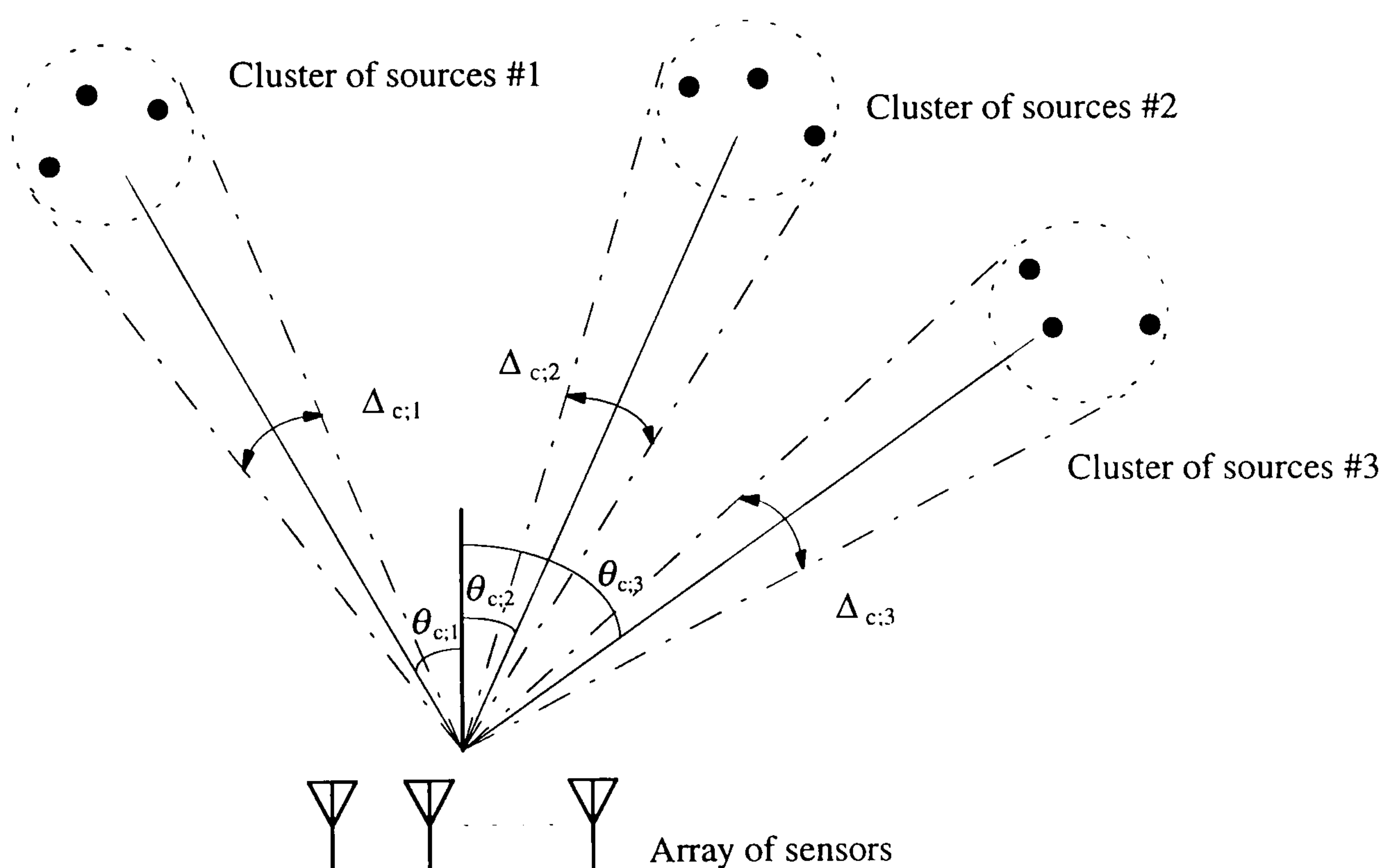


Figure 4.1: The three cluster source scenario with the cluster angular widths $\Delta_{c;k}$, $k = 1, \dots, 3$, and the cluster nominal directions of arrival $\theta_{c;k}$, $k = 1, \dots, 3$.

4.2.2 Beamforming-based PFML DOA estimation

Having provided some justification for practical usefulness of the described cluster source model, we next propose a combined classical and LCMV beamforming procedure, for operating under this source scenario.

A combined classical and LCMV beamforming procedure

1. Over the array FOV, calculate the following classical beamforming spatial spectrum

$$P_{ss}(\theta) = \mathbf{a}^H(\theta)\hat{\mathbf{R}}\mathbf{a}(\theta) \quad (4.9)$$

From $P_{ss}(\theta)$, find the locations $\hat{\theta}_{c;k}$ of, say K_c , spectral peaks corresponding to the nominal directions of the cluster signals. In addition, for each of the K_c spectral peaks, the estimates of the left and right cluster widths ($\hat{\Delta}_{l;k}$ and $\hat{\Delta}_{r;k}$ for the k th peak) may also be obtained from $P_{ss}(\theta)$ at this step. Figure 4.2 illustrates these parameters for the single source cluster case ($K_c = 1$). From the figure, $\hat{\Delta}_{l;k} = \hat{\theta}_{c;k} - \hat{\theta}_{l;k}$, and $\hat{\Delta}_{r;k} = \hat{\theta}_{r;k} - \hat{\theta}_{c;k}$, where $\hat{\theta}_{l;k}$ and $\hat{\theta}_{r;k}$ are the estimates of the left and right cluster-edge locations.

2. Form the matrix

$$\hat{\mathbf{R}}_{p;0} = \left\{ \int_{\theta_{\min}}^{\theta_{\max}} P_{ss}^{\beta}(\theta) W(\theta) \mathbf{a}(\theta) \mathbf{a}^H(\theta) d\theta \right\} / \left\{ \int_{\theta_{\min}}^{\theta_{\max}} P_{ss}^{\beta}(\theta) d\theta \right\} \quad (4.10)$$

where $\theta_{\min}$ and $\theta_{\max}$ form the array FOV. The matrix $W(\theta)$ is an optional weighting function. The positive number β is used to emphasise (or de-emphasise) the spectral peaks. The choice of $\beta > 1$ sometimes improves DOA estimation performance by improving strong directional interference suppression of the beamformer, which is to be formed in the next step. Note that, in practice,

where the array response vector is known only at discrete values of θ , the integral operator in Eq. (4.10) should be replaced by the sum operator.

3. At each spectral peak location $\hat{\theta}_{c;k}$, calculate the array weight vector

$$\mathbf{w}_k = \hat{\mathbf{R}}_{p;k}^{-1} \mathbf{a}(\hat{\theta}_{c;k}) (\mathbf{a}^H(\hat{\theta}_{c;k}) \hat{\mathbf{R}}_{p;k}^{-1} \mathbf{a}(\hat{\theta}_{c;k}))^{-1} \quad (4.11)$$

where

$$\hat{\mathbf{R}}_{p;k} = \hat{\mathbf{R}}_{p;0} - \left\{ \int_{\hat{\theta}_{c;k} - \hat{\Delta}_{l,k}}^{\hat{\theta}_{c;k} + \hat{\Delta}_{r,k}} P_{ss}^{\beta}(\theta) W(\theta) \mathbf{a}(\theta) \mathbf{a}^H(\theta) d\theta \right\} / \left\{ \int_{\theta_{\min}}^{\theta_{\max}} P_{ss}^{\beta}(\theta) d\theta \right\} \quad (4.12)$$

Note that $\hat{\Delta}_{l,k}$ and $\hat{\Delta}_{r,k}$ are the estimates of the left and right cluster widths as mentioned in Step 1. If the k th cluster angular width, denoted by $\Delta_{c;k}$, is known at least approximately, instead of the values obtained by Step 1, $\hat{\Delta}_{l,k}$ and $\hat{\Delta}_{r,k}$ may be alternatively chosen as $\tilde{\Delta}_{c;k}/2$, where $\tilde{\Delta}_{c;k}$ should be close to $\Delta_{c;k}$.

4. Obtain the estimates of K_c nominal cluster signals from

$$s_{c;k}(t) = \mathbf{w}_k^H \mathbf{x}(t), k = 1, \dots, K_c \quad (4.13)$$

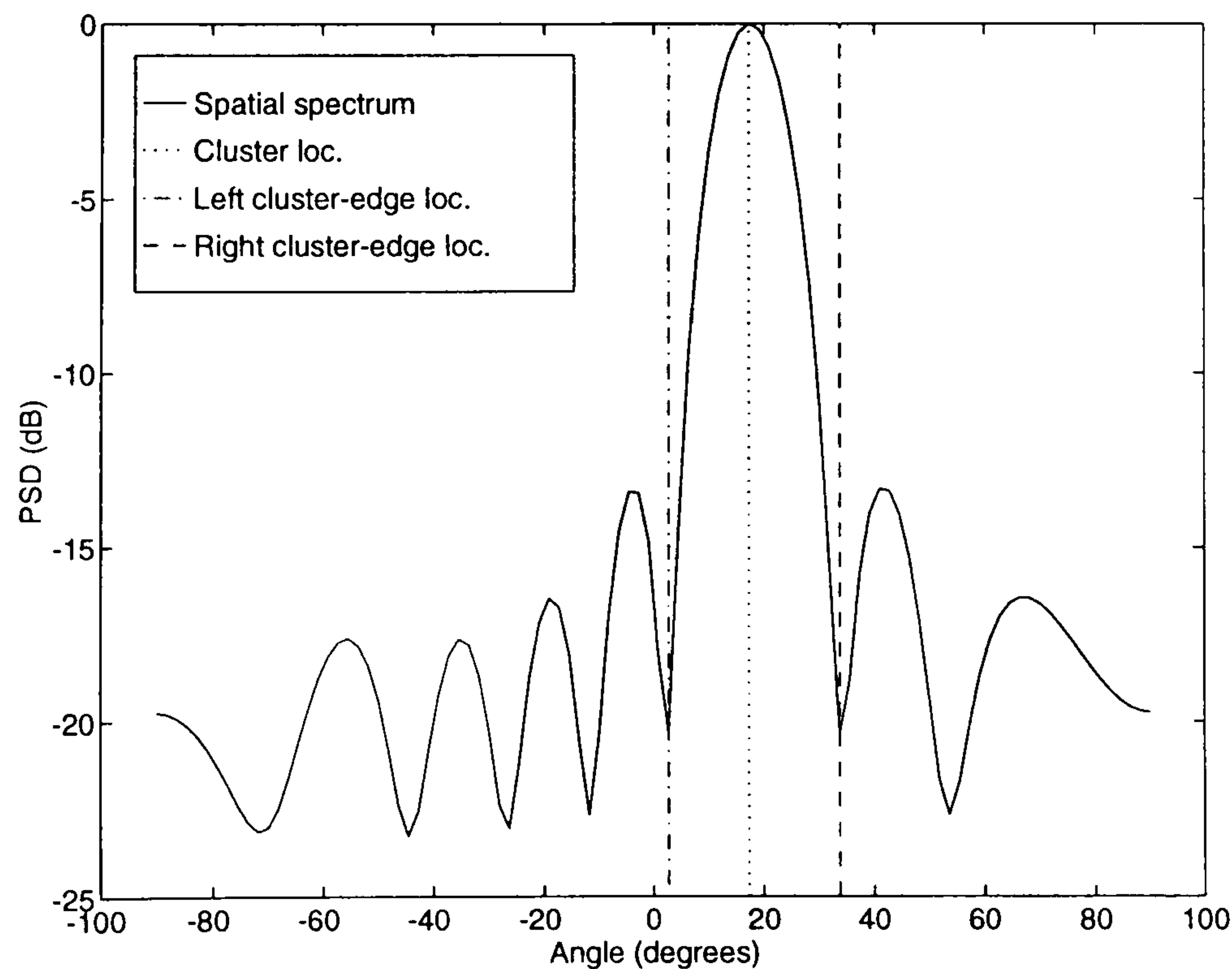


Figure 4.2: The cluster nominal angular location, and the left and right cluster-edge angular locations.

If the signals in each cluster are coherent, i.e., $s(t) = \Gamma_s s_{nc}(t)$ with Γ_s as given by Eq. (3.31), the beamforming signal $s_c(t) = [s_{c;1}(t) \ s_{c;2}(t) \ \cdots \ s_{c;K_c}(t)]^T$ provides the type-4 pre-filtering reference signal ($r_4(t)$) of the form Eq. (3.90). To see this, from Eq. (4.13) and Eq. (2.2), $s_c(t)$ can be written as

$$s_c(t) = W_B^H A s(t) + W_B^H n(t) \quad (4.14)$$

where W_B is the matrix composed of K_c weight vectors from Step 3 of the beamforming procedure. Because $s(t) = \Gamma_s s_{nc}(t)$,

$$s_c(t) = W_B^H A \Gamma_s s_{nc}(t) + W_B^H n(t) \quad (4.15)$$

Provided that $W_B^H A \Gamma_s$ is of full rank d_s , $s_c(t)$ in Eq. (4.15) is recognised as the type-4 pre-filtering reference signal $r_4(t)$ with $\tilde{A} = W_B^H A \Gamma_s$ and $Q_n = W_B^H$.

When, for each cluster, the signals are only partially correlated, useful pre-filtering can still be obtained using $s_c(t)$ as the pre-filtering reference signal, provided that all of the signals $s_k(t)$, $k = 1, \dots, K$, are at least partially correlated with $s_c(t)$. This requirement is necessary to ensure that there is sufficient information left in the filtered data for estimating all source directions. Note that, in this case, $s_c(t)$ does not have the form Eq. (3.90). However, if for each of $s_{c;k}(t)$, $k = 1, \dots, K_c$, the component due to the signals from other clusters can be negligible, $s_c(t)$ can be approximated by

$$s_c(t) \simeq \tilde{s}_{nc}(t) + Q_n n(t) \quad (4.16)$$

where $\tilde{s}_{nc}(t)$ is composed of uncorrelated signals. Based on the described assumption, each of the components of $\tilde{s}_{nc}(t)$ is a combination of the signals within the same cluster. In addition, $s(t)$ can be written as

$$s(t) = \tilde{\Gamma}_s \tilde{s}_{nc}(t) + s'(t) \quad (4.17)$$

where the $K \times K_c$ complex matrix $\tilde{\Gamma}_s$ has the same form as that of Γ_s which is given by Eq. (3.31). In this case, K_c is smaller than d_s , which is the rank of the signal covariance matrix. The term $s'(t)$ is uncorrelated with $\tilde{s}_{nc}(t)$. With some loss of useful information,

$s'(t)$ can be treated as noise under the PFML framework. It is thus seen that $s_c(t)$ in Eq. (4.16) has the form of $r_4(t)$ in Eq. (3.90) if we let $s(t) = \tilde{I}_s \tilde{s}_{nc}(t)$ and treat the term $s'(t)$ as part of the unknown additive noise.

Let $r_4(t) = s_c(t)$, and assume at the moment that the term $Q_n n(t)$ in $r_4(t)$ is negligible. Thus, based on $r_4(t)$, either the Unitary filter or Matched filter may be formed. The DOA estimates can be obtained from either the PFML-2 estimator (with the Matched filter) or the PFML-3 estimator (with the Unitary filter). If $\tilde{A}$ in $r_4(t)$ (see Eqs. (3.90, 4.15)) is approximately diagonal, the PFML-2 estimator should be preferred due to its superior performance to the PFML-3 estimator. The matrix $\tilde{A}$ is approximately diagonal when each of the K_c beamformers, formed according to the described procedure, effectively suppress the signals from other $K_c - 1$ clusters. When this is not the case, however, $\tilde{A}$ is poorly represented by a diagonal matrix and the PFML-3 method should be used instead. Note that, because generally $\tilde{A}$ is not truly diagonal, the DEML estimator should not be applied. It was observed from empirical study that, attempting to do so generally results in poor DOA estimates. In addition, when the term $Q_n n(t)$ in $r_4(t)$ is not negligible, both the PFML-2 and PFML-3 estimators yield biased DOA estimates. This estimation bias can be reduced by the use of the bias correction term, to be described in the next section.

As mentioned in Section 3.2.4, the asymptotic properties of the above Beamforming-based PFML-2 and PFML-3 (BPFML-2 and BPFML-3) estimators can be obtained, respectively, from Theorem 3.3, and Theorem 3.4. For large L , and by neglecting the term of the order $O_p(1/\sqrt{L})$, $s_{c;k}(t) \simeq w_{as;k}^H x(t)$, $k = 1, \dots, K_c$, where

$$w_{as;k} = R_{p;k}^{-1} a(\bar{\theta}_{c;k}) (a^H(\bar{\theta}_{c;k}) R_{p;k}^{-1} a(\bar{\theta}_{c;k}))^{-1} \quad (4.18)$$

and $R_{p;k} = E\{\hat{R}_{p;k}\}$. The cluster locations $\bar{\theta}_{c;k}$, $k = 1, \dots, K_c$, can be obtained by Step 1 of the described beamforming procedure, with $\hat{R}$ in Eq. (4.9) being replaced by R . This approximation of $s_{c;k}(t)$ is based on the fact that, for large L , $\hat{\theta}_{c;k} =$

$\bar{\theta}_{c;k} + O_p(1/\sqrt{L})$ and $\hat{\mathbf{R}}_{p;k} = \mathbf{R}_{p;k} + O_p(1/\sqrt{L})$. Thus, for large L with $\mathbf{r}_4(t) = \mathbf{s}_c(t)$, $\mathbf{R}_{r_4} \simeq \mathbf{W}_{as}^H \mathbf{R} \mathbf{W}_{as}$, where $\mathbf{W}_{as} = [\mathbf{w}_{as;1} \quad \mathbf{w}_{as;2} \quad \cdots \quad \mathbf{w}_{as;K_c}]$. This approximation of $\mathbf{s}_{c;k}(t)$ does not affect the asymptotic properties of the BPFML-2 and BPFML-3 estimators, provided that the bias term $\mathbf{E}_b$, to be discussed next, is sufficiently small.

When the bias term $\mathbf{E}_b$, which is due to $\mathbf{Q}_n \mathbf{n}(t)$ and, for the case of the BPFML-2 estimator, the non-diagonal part of $\tilde{\mathbf{A}}$, is not negligible but is sufficiently small, the asymptotic analysis which takes into account the bias term $\mathbf{E}_b$ differs from the above analysis, which neglects the term $\mathbf{E}_b$, only in the asymptotic biases of the DOA estimates. The asymptotic estimation biases can be separately obtained from Eq. (3.120) for the BPFML-3 method and from Eq. (3.122) for the BPFML-2 method.

4.2.3 An improved beamforming-based PFML estimator for correlated cluster sources

The two beamforming-based PFML estimators as described in Section 4.2.2 have one drawback. Their performance starts to deteriorate as the correlation between the signals in the same cluster decreases. This can be seen by noticing that, as the cluster signal correlation decreases, the term $s'(t)$ in Eq. (4.17) has its amplitude increased while the amplitude of the term $\tilde{\mathbf{I}}_s \tilde{\mathbf{s}}_{nc}(t)$ decreases. Because it is the latter term which is considered as the signal term in either of the two BPFML estimators, the effective SNR is reduced, resulting in poor DOA estimation. In this section, we provide a solution to this problem. In the following, based on approximate information of the signal directions, an improved version of the beamforming procedure in Section 4.2.2 is given.

An improved beamforming procedure

1. Follow Steps 1-2 of the beamforming procedure as described in Section 4.2.2.
2. Calculate $\hat{\mathbf{R}}_{p;k}$ using Eq. (4.12). Let the number of signals at the k th cluster be K_k . The K_k beamforming weight vectors corresponding to the k th cluster are

obtained from

$$\mathbf{w}_{k,l} = \hat{\mathbf{R}}_{p;k}^{-1} \mathbf{C}_k (\mathbf{C}_k^H \hat{\mathbf{R}}_{p;k}^{-1} \mathbf{C}_k)^{-1} \mathbf{f}_l, \quad k = 1, \dots, K_c; l = 1, \dots, K_k \quad (4.19)$$

where

$$\mathbf{C}_k = \begin{bmatrix} \mathbf{a}(\tilde{\boldsymbol{\theta}}_{k,1}) & \mathbf{a}(\tilde{\boldsymbol{\theta}}_{k,2}) & \cdots & \mathbf{a}(\tilde{\boldsymbol{\theta}}_{k,K_k}) \end{bmatrix} \quad (4.20)$$

$$\mathbf{f}_l = [0 \quad \cdots \quad \underset{l\text{th}}{1} \quad \cdots \quad 0]^T \quad (4.21)$$

and $\tilde{\boldsymbol{\theta}}_{k,l}$, $l = 1, \dots, K_k$ are the approximations of the directions of the K_k signals which belong to the k th cluster.

3. Obtain the estimates of the signal waveforms from

$$\hat{s}_{k,l}(t) = \mathbf{w}_{k,l}^H \mathbf{x}(t), \quad k = 1, \dots, K_c; l = 1, \dots, K_k \quad (4.22)$$

where $\hat{s}_{k,l}(t)$ is the estimate of the l th signal in the k th cluster.

From the above beamforming procedure, let

$$\hat{\mathbf{s}}(t) = [\hat{s}_{1,1}(t) \quad \hat{s}_{1,2}(t) \quad \cdots \quad \hat{s}_{1,K_1}(t) \quad \hat{s}_{2,1}(t) \quad \cdots \quad \cdots \quad \hat{s}_{K_c,K_{K_c}}(t)]^T \quad (4.23)$$

It is seen that the signal vector $\hat{\mathbf{s}}(t)$ has the form of the type-4 pre-filtering reference signal $\mathbf{r}_4(t)$ as described by Eq. (3.90), provided that each of the signal components in $\mathbf{s}(t)$ is at least partially correlated with one of the components in $\hat{\mathbf{s}}(t)$. Note that, in this partially correlated signal case, $\mathbf{s}(t) = \boldsymbol{\Gamma}_s \mathbf{s}_{nc}(t)$ with $d_s = K$, and $\boldsymbol{\Gamma}_s$ being diagonal or composed of block matrices. If $\hat{s}_{k,l}(t)$, $l = 1, \dots, K_k$, contain negligible amount of the signals from other clusters, by letting $\mathbf{r}_4(t) = \hat{\mathbf{s}}(t)$, the matrix $\tilde{\mathbf{A}}$ in $\mathbf{r}_4(t)$ ($= \tilde{\mathbf{A}} \mathbf{s}_{nc}(t) + \mathbf{Q}_n \mathbf{n}(t)$) is approximately block-diagonal. When this is the case, and that the term $\mathbf{Q}_n \mathbf{n}(t)$ is sufficiently small, the DOA estimates can be obtained by the PFML-4 estimator (with the Unitary-Matched filter), detailed in Section 3.2.4 of Chapter 3. The

asymptotic properties of this Beamforming-based PFML-4 (BPFML-4) estimator can be obtained from Eqs. (3.104, 3.125) (see the related discussion in Section 3.2.4).

Remark

It should be noted that, in all BPFML methods as described in this section and in Section 4.2.2, the maximum number of signal sources are limited by the interference suppressing ability of each of the beamformers, formed by either of the two beamforming procedures. Thus, for efficient estimation of the signal waveforms, which are subsequently used in the pre-filtering step of a PFML estimator, the number of source clusters K_c should not be larger than the number of sensors. In addition, the cluster signals with a smaller cluster angular width (i.e., less angular spread) are more effectively suppressed by the beamformer than the one with a larger cluster width. Moreover, for the improved beamforming procedure, the use of a set of K_k linear constraints in each beamformer reduces the capability of the beamformer in suppressing the signals from other clusters, especially when K_k is comparatively large compared with the number of sensors. Thus, while the above improved procedure can outperform the procedure in Section 4.2.2 when the signals within the same cluster are partially correlated, the beamformer based on the procedure in Section 4.2.2 is expected to be better than that of the improved procedure in suppressing the signals from other clusters. At each beamforming output, the signals from other clusters need to be sufficiently reduced for the BPFML DOA estimates based on the PFML-2 or PFML-4 cost function to be approximately asymptotically unbiased. This requirement is, however, irrelevant for the BPFML estimator based on the PFML-3 cost function.

4.3 Bias correction

As mentioned in Section 3.2.4 of Chapter 3, the biases of the DOA estimates using the type-4 pre-filtering reference signal $\mathbf{r}_4(t)$ can be reduced if the bias term $\mathbf{E}_b$ is

known at least approximately. We consider first the case of the BPFML-3 estimator with the Unitary filter. If $\mathbf{r}_4(t)$ is obtained by one of the two beamforming procedures, ideally the bias correction term $\mathbf{E}_c$ in Eq. (3.119) should be chosen exactly as $\mathbf{E}_b$, which is given by

$$\mathbf{E}_b = \mathbf{R}_n \mathbf{W}_B \mathbf{R}_{r_4}^{-1} \hat{\mathbf{R}}_{r_4}^{1/2} \quad (4.24)$$

where $\mathbf{W}_B$ is the matrix whose column vectors are the beamforming weight vectors, calculated, for instance, by Eq. (4.11). Unfortunately, in practice complete knowledge of $\mathbf{R}_n$ is generally unavailable. However, it can sometimes be estimated from the observed data. This is the case, for example, if the noise covariance structure $\mathbf{R}_{ns} = \sigma^{-2} \mathbf{R}_n$ is known and $N > d_s$. Without imposing any constraint on the response matrix $\mathbf{A}$ and the signal covariance matrix $\mathbf{R}_s$, apart from that the column vectors of $\mathbf{A}\mathbf{R}_s^{1/2}$ span a d_s -dimensional subspace, the ML estimate of $\mathbf{R}_n$ is given by (see, for example, Chapter 8 of [44])

$$\hat{\mathbf{R}}_n = \hat{\sigma}^2 \mathbf{R}_{ns} \quad (4.25)$$

where

$$\hat{\sigma}^2 = \frac{1}{N - d_s} \sum_{i=d_s+1}^N \lambda_i \quad (4.26)$$

and λ_i , $i = d_s + 1, \dots, N$, are the $N - d_s$ smallest eigenvalues of $\mathbf{R}_{ns}^{-1/2} \hat{\mathbf{R}} \mathbf{R}_{ns}^{-1/2}$. Thus, $\mathbf{E}_c$ may be chosen as

$$\mathbf{E}_c = \hat{\sigma}^2 \mathbf{R}_{ns} \mathbf{W}_B \hat{\mathbf{R}}_{r_4}^{-1/2} \quad (4.27)$$

with $\hat{\sigma}^2$ as given by Eq. (4.26). For the BPFML-2 method, the bias correction term for each of $\mathbf{y}_k$, $k = 1, \dots, K_c$, can be similarly chosen as

$$[\mathbf{E}_c]_k = \hat{\sigma}^2 \mathbf{R}_{ns} \mathbf{w}_k [\hat{\mathbf{R}}_{r_4}]_{k,k}^{-1/2} \quad (4.28)$$

where $\mathbf{w}_k$ is the weight vector obtained from Eq. (4.11). For the case of the BPFML-4 method with the weight vectors obtained from the improved beamforming procedure, the bias correction term for each of $\mathbf{Y}_k$, $k = 1, \dots, K_c$, may be obtained in a similar manner to that of the BPFML-3 method, and is given by

$$\mathbf{E}_{c,k} = \hat{\sigma}^2 \mathbf{R}_{ns} \mathbf{W}_{iB,k} \hat{\mathbf{R}}_{r_4,k}^{-1/2} \quad (4.29)$$

where $\hat{\mathbf{R}}_{r_4,k}$ is as given by Eq. (3.98), and $\mathbf{W}_{iB,k}$ is the matrix whose column vectors are the K_k beamforming weight vectors corresponding to the k th cluster. These beamforming weight vectors are obtained by Step 2 of the improved beamforming procedure (Eq. (4.19)).

4.4 Robust adaptive beamformer

Although the beamforming procedure of Section 4.2.2 is developed for the purpose of DOA estimation, it suggests a new beamforming formation which is robust to pointing error and the presence of correlated sources. Based on the waveform estimate $\hat{s}_d(t)$ of the signal of interest, obtained by the described beamforming procedure, we propose the robust beamformer whose weight vector is given by

$$\mathbf{w} = \hat{\mathbf{R}}_{in}^{-1} \hat{\mathbf{r}}_{xs_d} (\hat{\mathbf{r}}_{xs_d}^H \hat{\mathbf{R}}_{in}^{-1} \hat{\mathbf{r}}_{xs_d})^{-1} \quad (4.30)$$

where

$$\hat{\mathbf{r}}_{xs_d} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{x}(t) \hat{s}_d^*(t) \quad (4.31)$$

and the estimated interference plus noise covariance matrix

$$\hat{\mathbf{R}}_{in} = \hat{\mathbf{R}} - \zeta \cdot (\hat{\mathbf{r}}_{xs_d} \hat{\mathbf{r}}_{xs_d}^H) \quad (4.32)$$

$$\zeta = 1 / \left(\frac{1}{L} \sum_{t=0}^{L-1} |\hat{s}_d(t)|^2 \right) \quad (4.33)$$

In the presence of coherent sources, the performance of the proposed method, compared to that of the LCMV beamformer is demonstrated by the following simulation experiment.

EXP-4.1: In this experiment, a uniform linear array of eight sensors was used. The inter-sensor spacing was half of the signal wavelength. The total of $K = 10$ complex signals impinged on the array. They were grouped into 5 clusters. Any two signals in the same cluster were coherent while uncorrelated with signals from other clusters. All signals had unity power. Their directions of arrival are listed in Table 4.1.

Coherent source no.	Cluster No.				
	1	2	3	4	5
1	-65	-35	-5	20	55
2	-60	-30	0	25	60

Table 4.1: The signal directions of arrival (in degrees, compared with broad side).

All pairs of coherent signals had the phase difference of $\pi/4$. All signal phases were measured with respect to the array centre which is the point in the middle between the fourth and fifth sensors. The second coherent signal from the 3rd cluster was considered the desired signal. In addition to these directional point sources, the additive circular complex white Gaussian noise was added at each sensor to obtain eight different SNRs, defined here as the reciprocal of the noise power. Five hundred simulations were carried out with the number of snapshots in each simulation $L = 100$. The proposed beamforming procedure was performed with $\hat{\Delta}_{l;k}$ and $\hat{\Delta}_{r;k}$ being obtained by the first step of the beamforming procedure described in Section 4.2.2. The power factor β in Eqs. (4.10, 4.12) was chosen to be $\beta = 2$. In addition, the weighting function $W(\theta)$ in Eqs. (4.10, 4.12) was chosen to be uniform, i.e., its value is independent of θ . Apart from the proposed robust beamformer, the standard LCMV beamformer was performed

with the unity gain constraint in the look direction of 0 degrees. A beamforming performance is measured by the correlation between the beamforming output signal and the desired signal. The described performance measure is given in mathematical form by

$$\text{CORR} \triangleq \left| \sum_{t=0}^{L-1} \hat{s}_d(t) s_d^*(t) \right| \cdot \left[\left\{ \sum_{t=0}^{L-1} |s_d(t)|^2 \right\} \left\{ \sum_{t=0}^{L-1} |\hat{s}_d(t)|^2 \right\} \right]^{-1/2} \quad (4.34)$$

where $\hat{s}_d(t)$ is the beamforming output signal. The results from eight different SNRs using the two beamforming methods, as well as the signal received at the reference sensor (no. 1), are plotted in Figure 4.3. From the figure, it is seen that the LCMV beamformer suffers performance degradation due to the signal cancellation effect. The problem becomes more pronounced at high SNR. In contrast, the proposed robust beamformer performed better as SNR was increased. No signal cancellation problem was found for the proposed beamformer.

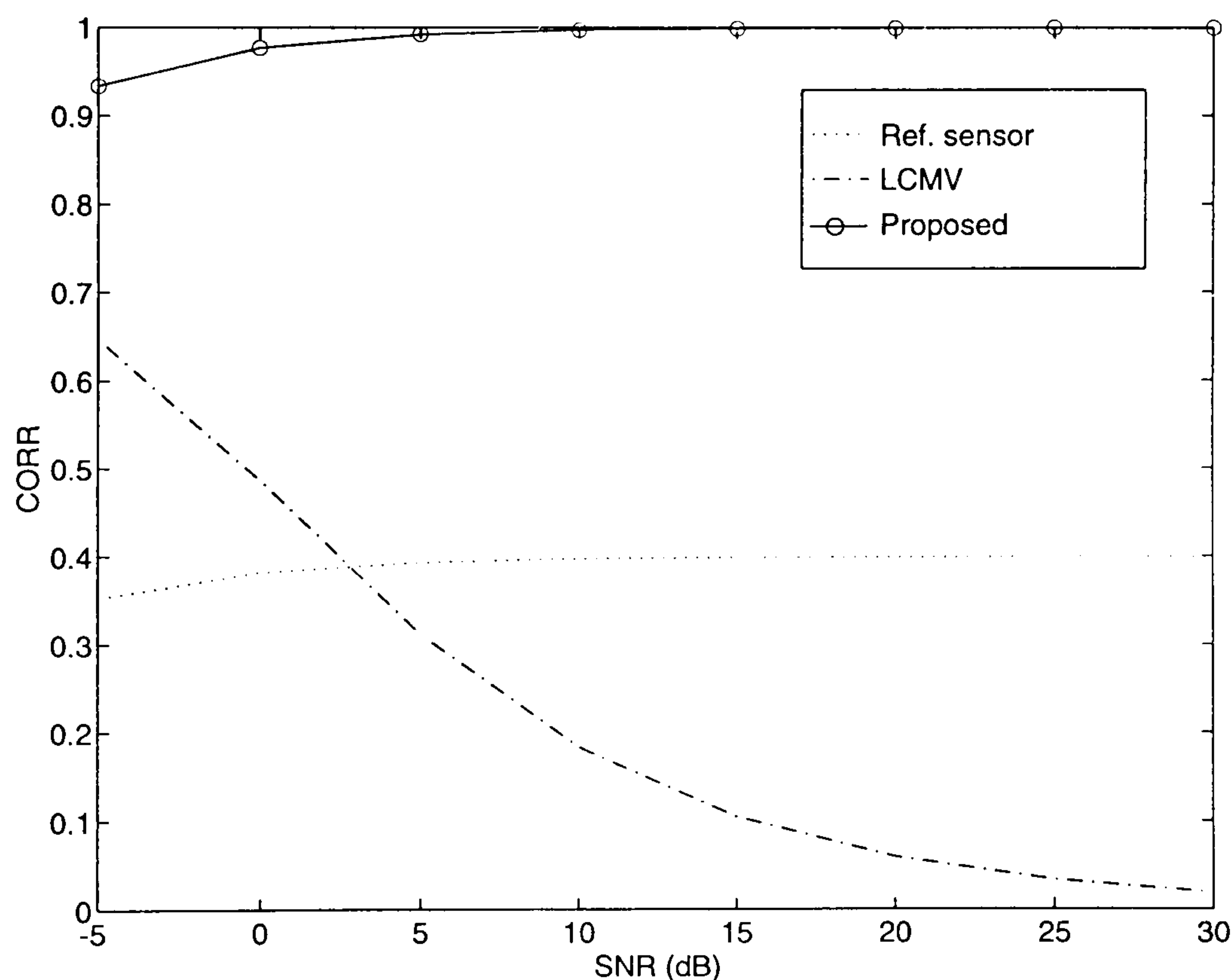


Figure 4.3: Plots of the CORRs corresponding to the LCMV method, the proposed beamforming method, and the signal received by the 1st sensor, at eight SNRs.

4.5 Conclusions

In this chapter, we have developed the combined classical and LCMV beamforming procedure for estimating the desired signal waveforms. The source scenario where the described procedure is effective has been described. The main purpose of this beamforming procedure is to obtain the pre-filtering reference signal $\mathbf{r}_4(t)$ for use under the PFML DOA estimation framework. Based on this beamforming output signal, either the PFML-2 or PFML-3 estimator may be applied for DOA estimation. In addition, the improved beamforming procedure has been given to improve the estimation performance when the cluster signals are only partially correlated. From the pre-filtering reference signal obtained by this improved procedure, the PFML-4 estimator with the Unitary-Matched filter can be applied. When the knowledge of the noise covariance matrix or its structure is known, it has been shown how the bias correction term $\mathbf{E}_c$ in Eq. (3.119) can be obtained in practice. Finally, by means of computer simulation, it has been observed that the adaptive beamformer based on the proposed beamforming procedure of Section 4.2.2 is robust to the presence of coherent sources.

Chapter 5

The IQML Algorithm for Maximum Likelihood Parameter Estimation

In the previous three chapters, we have formulated various ML-based DOA estimators. However, the computational aspect of these methods has not yet been considered. The ML DOA estimation problem can be solved, for instance, by using the algorithms considered in [57-59]. Despite superior performance over the subspace-based methods, the ML methods are in general computationally complicated. For the case of the data set with uniform sampling grid, e.g., the one obtained from a uniform linear array, the computationally efficient Iterative Quadratic Maximum Likelihood (IQML) algorithm was proposed in [22] to solve the DML parameter estimation problem. The algorithm is known to perform well for sufficiently high SNR. Later, the method was shown to be equivalent to the well-known Steiglitz-McBride algorithm [60].

Recently Nagesha and Kay [24] (see also Stoica and Sharman [23]) showed that the IQML algorithm in [22] suffers performance degradation under a certain parameter configuration, to be described later. They (Nagesha and Kay) provided a version of the IQML algorithm which solves the problem. In spite of this success, for the multiple snapshot case, both algorithms have high SNR threshold[†] and are highly inefficient, when measured in terms of the performance departure from the CRLB. Later, it will be pointed out that this is due to an improper choice of the nontriviality constraint, required by both algorithms.

[†] A SNR level below which the estimation performance degrades significantly, when compared with the upper performance bound set by the corresponding CRLB.

In this chapter, we consider the IQML algorithm for ML parameter estimation by minimising a certain cost function. Despite our main interest in narrow-band source DOA estimation, the development in this chapter can also be applied to other related parameter estimation problems such as multiple frequency estimation [22]. In the next section, we first formulate the form of the IQML algorithm which is the generalisation of the original one described in [22]. Pre-processing such as the array interpolator [25, 61] is allowed in this generalised algorithm. We next propose an improved version of the IQML algorithm. For the multiple data snapshot case, the proposed algorithm offers performance improvement over both the original algorithm in [22] and the modified one in [24]. For the single data snapshot case, the proposed algorithm has superior performance to the other two existing IQML algorithms, under difficult situations such as small data length. In addition, when some parameter values are known a priori, the IQML algorithm which makes use of this knowledge is formulated in Section 5.3. This utilisation of the prior information is later shown to improve the estimation performance and convergence properties of the algorithm.

In Section 5.4, we turn our attention from the one-dimensional to the two-dimensional parameter estimation problem. The problem of 2-D parameter estimation can be found, for instance, in source azimuth-elevation localisation and frequency-wavenumber analysis. Due to the increase in the dimensionality of the problem, the straightforward extension of many existing 1-D methods either is infeasible or results in a computationally complicated problem. Much work has concentrated on reducing the complexity of 2-D parameter estimation [41, 62-66]. In [41], the IQML algorithm was applied, in conjunction with classical beamforming, to a rectangular array for the 2-D source localisation problem. The 2-D ML-based method employing the IQML algorithm was suggested in [62]. Although the latter method is based on an exact 2-D ML principle, its use is limited due to the requirement that there is at least one of the two parameter dimensions where its 1-D parameters have distinct values. The method in

[63] also has similar restriction. While this is not the case for the method due to Hua [64], the method in [64], as well as the one in [66], can be highly inefficient because, to avoid high computational complexity, it must be applied to a subdata set with reduced dimensionality. Although the method in [65] is less computationally complicated than Hua's method, its performance is inferior to that of [64].

Based on the IQML algorithm, we propose a 1-D IQML-based 2-D parameter estimation method. The method consists of two IQML algorithms for separate estimation of the parameters in each of the two dimensions. The parameter pairing problem is addressed. The main contribution in this part is to provide a sufficient condition on the construction of the data matrix for consistent parameter pairing. Not only is the proposed method not as restrictive as Clark's method, its performance is found to be comparable or superior to the matrix pencil method of Hua [64].

5.1 Problem formulation

We first consider the 2-D extension of the data model Eq. (2.2), given by

$$\mathbf{x}(t) = \mathbf{A}(\mathbf{y}, \mathbf{z})\mathbf{s}(t) + \mathbf{n}(t), \quad t = 0, \dots, L-1 \quad (5.1)$$

where

$$\mathbf{x}(t) = \text{vec}(\mathbf{X}^T(t)) \quad (5.2)$$

$$\mathbf{X}(t) = \begin{bmatrix} x_{1,1}(t) & x_{1,2}(t) & \cdots & x_{1,N}(t) \\ x_{2,1}(t) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{M,1}(t) & \cdots & \cdots & x_{M,N}(t) \end{bmatrix} \quad (5.3)$$

and

$$\mathbf{y} = [y_1 \quad y_2 \quad \cdots \quad y_K]^T \quad (5.4)$$

$$\mathbf{z} = [z_1 \quad z_2 \quad \cdots \quad z_K]^T \quad (5.5)$$

Let the response matrix $\mathbf{A}(\mathbf{y}, \mathbf{z})$ be expressed as

$$\mathbf{A}(\mathbf{y}, \mathbf{z}) = [\mathbf{a}_1(y_1, z_1) \quad \mathbf{a}_2(y_2, z_2) \quad \cdots \quad \mathbf{a}_K(y_K, z_K)] \quad (5.6)$$

where the response vectors $\mathbf{a}_k(y_k, z_k)$, $k = 1, \dots, K$, are given by

$$\mathbf{a}_k(y_k, z_k) = \mathbf{g}_k(y_k) \otimes \mathbf{h}_k(z_k) \quad (5.7)$$

$$\mathbf{g}_k(y_k) = \varphi_{y_k} [1 \quad y_k^1 \quad y_k^2 \quad \cdots \quad y_k^{M-1}]^T \quad (5.8)$$

$$\mathbf{h}_k(z_k) = \varphi_{z_k} [1 \quad z_k^1 \quad z_k^2 \quad \cdots \quad z_k^{N-1}]^T \quad (5.9)$$

The phase factors φ_{y_k} and φ_{z_k} are the modulus-1 complex numbers whose values depend on the chosen phase reference point. For example, if the signal phases are referenced to the first data sampling point $x_{1,1}(t)$, $\varphi_{y_k} = \varphi_{z_k} = 1$, $k = 1, \dots, K$. In addition, the noise $\mathbf{n}(t)$ is assumed to be zero-mean circular complex temporally white Gaussian distributed. The noise covariance matrix $\mathbf{R}_n = \sigma^2 \mathbf{R}_{ns}$ is assumed to be of full rank. Depending on the estimation criterion, however, the signal vector $\mathbf{s}(t) = [s_1(t) \quad s_2(t) \quad \cdots \quad s_K(t)]^T$ can be treated as being composed of either the deterministic unknown parameters or the stochastic signals. For the latter case, it is assumed that $\mathbf{s}(t)$ is a zero-mean circular complex temporally white stationary random vector with the multivariate Gaussian distribution. It is further assumed to be uncorrelated with $\mathbf{n}(t)$. Both $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are also assumed to be ergodic.

Based on the above data model, we consider in this chapter the problems of 1-D and 2-D frequency estimation, as well as the (1-D) source direction estimation problem. For the 2-D frequency estimation problem, the K 2-D frequency pairs $(\theta_{y,k}, \theta_{z,k})$ are related to (y_k, z_k) by

$$y_k = \exp(j2\pi\theta_{y,k}), \quad k = 1, \dots, K \quad (5.10)$$

$$z_k = \exp(j2\pi\theta_{z,k}), \quad k = 1, \dots, K \quad (5.11)$$

The described data model can also be used for 1-D parameter estimation problems, by letting $\mathbf{g}_k = 1$, $k = 1, \dots, K$, and $M = 1$. For the 1-D frequency estimation problem, the K frequencies $\theta_{z,k}$ are related to z_k by Eq. (5.11). With the uniform linear

array of half signal wavelength inter-sensor spacing, the K source directions $\theta_{z,k}$ are related to z_k by

$$z_k = \exp(-j\pi \sin(\theta_{z,k})), k = 1, \dots, K \quad (5.12)$$

Let θ be the vector of the parameters (signal frequencies or directions) to be estimated. For the 2-D frequency estimation problem,

$$\begin{aligned} \theta &= [\theta_{y,1} \ \theta_{y,2} \ \cdots \ \theta_{y,K} \ \theta_{z,1} \ \cdots \ \theta_{z,K}]^T \\ &\triangleq [\theta_y^T \ \theta_z^T]^T \end{aligned} \quad (5.13)$$

where θ_y and θ_z are the $K \times 1$ vectors, while

$$\theta = [\theta_{z,1} \ \theta_{z,2} \ \cdots \ \theta_{z,K}]^T \quad (5.14)$$

for the 1-D frequency and DOA estimation problems. Only the estimation of θ is of interest here. The problem of estimating $s(t)$ is addressed, for instance, in [62, 67-69]. In subsequent sections, the estimation of θ is based on the assumption that there exists a *one-to-one* relation between (y_k, z_k) and $(\theta_{y,k}, \theta_{z,k})$, i.e., $(y_k, z_k) \Leftrightarrow (\theta_{y,k}, \theta_{z,k})$, $k = 1, \dots, K$. For the DOA estimation problem, this assumption is satisfied for $-\pi/2 \leq \theta_{z,k} \leq \pi/2$. In the following section, the 1-D problem of estimating $\theta_{z,k}$ is considered. A new nontriviality constraint is proposed to improve the algorithm performance.

5.2 The generalised IQML algorithm and the consistent non-triviality constraint

5.2.1 The generalised IQML algorithm

The IQML algorithm was proposed in [22] to solve the DML parameter estimation problem. It can also be used in a similar manner to solve the WSF problem [33], ([34], chapter 4). Let $N_2 = (K + 1)(N_1 - K)$, where $K < N_1 \leq N$. The $N_2 \times N$ matrix $\mathbf{H}$ is given by

$$\mathbf{H} = [\mathbf{H}_1^T \ \mathbf{H}_2^T \ \cdots \ \mathbf{H}_{N_1-K}^T]^T \quad (5.15)$$

where

$$\begin{aligned}
\mathbf{H}_k &= \tilde{\mathbf{H}} \cdot \begin{bmatrix} \mathbf{I}_{K+1} & \mathbf{0}_{(K+1) \times (N_1-K-1)} \end{bmatrix} \cdot \check{\mathbf{H}}, k = 1 \\
&= \tilde{\mathbf{H}} \cdot \begin{bmatrix} \mathbf{0}_{(K+1) \times (k-1)} & \mathbf{I}_{K+1} & \mathbf{0}_{(K+1) \times (N_1-K-k)} \end{bmatrix} \cdot \check{\mathbf{H}}, k = 2, \dots, N_1 - K - 1 \\
&= \tilde{\mathbf{H}} \cdot \begin{bmatrix} \mathbf{0}_{(K+1) \times (N_1-K-1)} & \mathbf{I}_{K+1} \end{bmatrix} \cdot \check{\mathbf{H}}, k = N_1 - K
\end{aligned} \tag{5.16}$$

are the full-rank $(K+1) \times N$ matrices. In Eq. (5.16), the $N_1 \times N$ matrix $\check{\mathbf{H}}$ has full rank N_1 , while the $(K+1) \times (K+1)$ matrix $\tilde{\mathbf{H}}$ is of full rank $(K+1)$. In addition, we let $\mathbf{b} = [b_0 \ b_1 \ \dots \ b_K]^T \in \Omega^{(K+1) \times 1}$, where $\Omega^{(K+1) \times 1}$ is the set of feasible parameter values of $\mathbf{b}$. Based on the $N \times 1$ data snapshots $\mathbf{x}(t)$, $t = 0, 1, \dots, L-1$, as described in Eq. (5.1) with $\mathbf{g}_k = 1$, $k = 1, \dots, K$, and $M = 1$, the IQML algorithm solves the following minimisation problem

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b} \in \Omega^{(K+1) \times 1}} \mathbf{b}^H \left\{ \sum_{t=0}^{L-1} \mathbf{X}_t^H(t) (\mathbf{B}^H \mathbf{H} \mathbf{R}_{\text{ns}} \mathbf{H}^H \mathbf{B})^{-1} \mathbf{X}_t(t) \right\} \mathbf{b} \tag{5.17}$$

where $\mathbf{R}_{\text{ns}}$ is the noise covariance structure, the $N_2 \times (N_1 - K)$ matrix

$$\mathbf{B} = \mathbf{I}_{N_1-K} \otimes \tilde{\mathbf{I}}_{K+1} \mathbf{b}^* \tag{5.18}$$

and

$$\mathbf{X}_t(t) = \begin{bmatrix} \tilde{\mathbf{I}}_{K+1} \mathbf{x}_{t,1}(t) & \tilde{\mathbf{I}}_{K+1} \mathbf{x}_{t,2}(t) & \dots & \tilde{\mathbf{I}}_{K+1} \mathbf{x}_{t,N_1-K}(t) \end{bmatrix}^T \tag{5.19}$$

with $\mathbf{x}_{t,k}(t)$ as the $(K+1) \times 1$ transformed data vectors

$$\mathbf{x}_{t,k}(t) = \mathbf{H}_k \mathbf{x}(t), k = 1, \dots, N_1 - K \tag{5.20}$$

The minimisation problem Eq. (5.17) can be found, for example, in the 1-D DML frequency estimation problem [22], where $N_1 = N$, $\check{\mathbf{H}} = \mathbf{I}_N$, and $\tilde{\mathbf{H}} = \mathbf{I}_{K+1}$. To obtain an exact DML estimate of the parameter θ as given by Eq. (5.14), the vector $\mathbf{b}$ must be constrained so that the K roots z_k , $k = 1, \dots, K$, of the polynomial $g(\mathbf{b}) = b_0 z^K + b_1 z^{K-1} + \dots + b_K$, are distinct and lie on the unit circle. In doing so, the column vectors of $\mathbf{H}^H \mathbf{B}$ form the null space of $\mathbf{A}^H(\mathbf{z})$. The parameter estimates $\hat{\theta}_{z,k}$, $k = 1, \dots$,

K , can be obtained from the K roots $\hat{z}_k$, $k = 1, \dots, K$, of $g(\hat{\mathbf{b}})$ by using either Eq. (5.11) or Eq. (5.12). Based on a uniform linear array, the WSF and PFML problems for DOA estimation can also be similarly cast in the form Eq. (5.17). In comparing Eq. (5.17) with the original form in [22], the form Eq. (5.17) turns out to be equivalent to the original form if $N_1 = N$, $\tilde{\mathbf{H}} = \mathbf{I}_N$, and $\tilde{\mathbf{H}} = \mathbf{I}_{K+1}$. However, Eq. (5.17) is more general because it allows for the inclusion of additional constraints as well as some forms of linear pre-processing. For instance, solving Eq. (5.17), with $\mathbf{b}$ being constrained such that the K roots of $g(\mathbf{b})$ are distinct and lie on the unit circle, is a difficult problem. To simplify the minimisation problem Eq. (5.17), it was proposed in [22] that the described constraint is replaced by the conjugate-symmetric constraint, imposed on $\mathbf{b}$ such that

$$\mathbf{b} = \tilde{\mathbf{I}}_{K+1} \mathbf{b}^* \quad (5.21)$$

With this constraint, $\mathbf{b}$ can be written as

$$\mathbf{b} = \tilde{\mathbf{J}} \tilde{\mathbf{b}} \quad (5.22)$$

where $\tilde{\mathbf{b}}$ is a *real* $(K+1) \times 1$ vector, and

$$\tilde{\mathbf{J}}_{K+1} = \begin{bmatrix} \mathbf{I}_{(K+1)/2} & j\tilde{\mathbf{I}}_{(K+1)/2} \\ \tilde{\mathbf{I}}_{(K+1)/2} & -j\mathbf{I}_{(K+1)/2} \end{bmatrix} \quad (5.23)$$

for the case of odd K , and

$$\tilde{\mathbf{J}}_{K+1} = \begin{bmatrix} \mathbf{I}_{K/2} & \mathbf{0}_{(K/2) \times 1} & j\tilde{\mathbf{I}}_{K/2} \\ \mathbf{0}_{1 \times (K/2)} & \sqrt{2} & \mathbf{0}_{1 \times (K/2)} \\ \tilde{\mathbf{I}}_{K/2} & \mathbf{0}_{(K/2) \times 1} & -j\mathbf{I}_{K/2} \end{bmatrix} \quad (5.24)$$

if K is even. This constraint can be included into the minimisation problem Eq. (5.17) by replacing $\mathbf{b}$ with the real vector $\tilde{\mathbf{b}}$ and letting $\tilde{\mathbf{H}} = \tilde{\mathbf{I}}_{K+1} \tilde{\mathbf{J}}_{K+1}^T \tilde{\mathbf{I}}_{K+1}$, $N_1 = N$, $\tilde{\mathbf{H}} = \mathbf{I}_N$. In addition to this conjugate-symmetric constraint, the linear interpolator as considered, for instance, in [70] can be included into the algorithm, with $\tilde{\mathbf{H}}$ in Eq. (5.16) as the

interpolation matrix. This type of pre-processing allows for the application of the IQML algorithm to a real irregular uniform linear array. Discussion on this topic is given in Chapter 7. Furthermore, as will be discussed in Section 5.3, when some of the parameter components of θ (or z) are known a priori, this prior information can be incorporated into the algorithm by a suitable choice of $\tilde{H}$.

5.2.2 The IQML algorithm with the consistent nontriviality constraint

The IQML algorithm is known to converge quickly with good performance for sufficiently high SNR. However, the original method can yield a poor result for a certain parameter configuration as pointed out in [23-24]. To be specific, when applied with the conjugate-symmetric constraint, the original algorithm breaks down if the parameters $z_k, k = 1, \dots, K$, are chosen such that

$$\operatorname{Re}\{e^{jK\pi/2} \prod_{k=1}^K z_k^{-1/2}\} = 0 \quad (5.25)$$

In [24], a solution to this problem was provided. The difference between the original IQML algorithm and the one in [24] is in the choice of the nontriviality constraint[†], imposed on the vector $\mathbf{b}$ to avoid the nontrivial solution $\hat{\mathbf{b}} = \mathbf{0}$. Although the modified method in [24] offers improved performance, as does the one in [22], it is still inefficient when a few data snapshots are available. This inefficiency can be explained by observing that the first step of Bresler's IQML algorithm is the linear prediction method with the prediction order of $K + 1$. It is well-known that the linear prediction method, particularly with a low prediction order, yields a poor result when applied to noisy data [7, 32]. Subsequent iterations of the algorithm also suffer the same problem. Despite the fact that the first step of Nagesha's IQML algorithm is the Pisarenko harmonic retrieval method with spatial smoothing, and thus gives asymptotically unbiased estimates, the parameter estimates obtained from subsequent iterations of the

[†] The use of the term 'nontriviality constraint' here follows that of [62].

algorithm can be biased. This is because the additive noise covariance matrix in subsequent iterations is modified by the matrix term F_i (see Eq. (5.29)). In the following, by applying a proper nontriviality constraint, these problems found in the two algorithms can be alleviated, resulting in marked performance improvement.

The proposed algorithm and the earlier two methods, based on the generalised form of the IQML algorithm Eq. (5.17) with $\Omega^{(K+1) \times 1}$ as the set of $(K+1) \times 1$ complex vectors, are summarised as follows:

1. Choose the initial condition F_1 . For instance, by letting the $(K+1) \times 1$ vector $\hat{\mathbf{b}}_0 = [1 \ 0 \ \dots \ 0]^T$, F_1 may be selected as

$$F_1 = (\hat{\mathbf{B}}_0^H \mathbf{H} \mathbf{R}_{\text{ns}} \mathbf{H}^H \hat{\mathbf{B}}_0)^{-1} \quad (5.26)$$

where $\hat{\mathbf{B}}_0$ is the matrix $\mathbf{B}$ in Eq. (5.18) with $\mathbf{b}$ in the equation being replaced by $\hat{\mathbf{b}}_0$.

2. At the i th iteration, obtain $\hat{\mathbf{b}}_i = [\hat{b}_{i,0} \ \hat{b}_{i,1} \ \dots \ \hat{b}_{i,K}]^T$ from one of the following methods:

- a) Bresler and Macovski's method [22]

$$\hat{\mathbf{b}}_i = \arg \min_{\mathbf{b} \in \Omega^{(K+1) \times 1}} \mathbf{b}^H \left\{ \sum_{t=0}^{L-1} \mathbf{X}_t^H(t) F_i \mathbf{X}_t(t) \right\} \mathbf{b} \quad (5.27)$$

subject to

$$\mathbf{c}^T \mathbf{b} = 1 \quad (5.28)$$

where

$$F_i = (\hat{\mathbf{B}}_{i-1}^H \mathbf{H} \mathbf{R}_{\text{ns}} \mathbf{H}^H \hat{\mathbf{B}}_{i-1})^{-1}, \quad i > 1 \quad (5.29)$$

$$\mathbf{c} = [1 \ 0 \ \dots \ 0]^T \quad (5.30)$$

and $\hat{\mathbf{B}}_{i-1}$ is the matrix $\mathbf{B}$ in Eq. (5.18) with $\mathbf{b}$ in the equation being replaced by $\hat{\mathbf{b}}_{i-1}$.

b) Nagesha and Kay's method [24]

$\hat{\mathbf{b}}_i$ is obtained by Eq. (5.27), subject to the following nontriviality constraint

$$\|\mathbf{b}\| = 1 \quad (5.31)$$

c) The proposed method

$\hat{\mathbf{b}}_i$ is obtained by Eq. (5.27), subject to

$$\text{tr}\{\mathbf{H}^H \mathbf{B} \mathbf{F}_i \mathbf{B}^H \mathbf{H} \mathbf{R}_{\text{ns}}\} = 1 \quad (5.32)$$

3. Repeat the first step by using the estimate of $\mathbf{b}$ from the previous iteration, i.e., $\hat{\mathbf{b}}_{i-1}$, to form $\mathbf{F}_i$. The algorithm is repeated in this manner until either the number of iterations reaches a maximum limit or when $\|\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i-1}\| < \varepsilon$, where ε is a small threshold number.
4. From the final $\hat{\mathbf{b}}$, map the K roots $\hat{z}_k, k = 1, \dots, K$, of the polynomial $g(\hat{\mathbf{b}}) = \hat{b}_0 z^K + \hat{b}_1 z^{K-1} + \dots + \hat{b}_K$, onto the unit circle. For the DOA estimation problem, the estimates $\hat{\theta}_{z,k}, k = 1, \dots, K$, can be obtained from the mapped roots by using Eq. (5.12). For the 1-D frequency estimation problem, the frequency estimates can be similarly obtained by using Eq. (5.11).

In considering the first step of the described procedure, it can be seen that, at any i th iteration, the estimate $\hat{\mathbf{b}}_i$ obtained by the proposed method (1.c) is asymptotically unbiased (at least for a positive definite $\mathbf{F}_i$). In addition, while the proposed constraint Eq. (5.32) appears rather complicated, its computation can be simplified. This can be done by observing that Eq. (5.32) is equivalent to

$$\mathbf{b}^H \tilde{\mathbf{T}}_i \mathbf{b} = 1 \quad (5.33)$$

where

$$\tilde{\mathbf{T}}_i = \sum_{l=1}^N \mathbf{X}_n^H(l) \mathbf{F}_i \mathbf{X}_n(l) \quad (5.34)$$

and

$$\mathbf{X}_n(l) = \begin{bmatrix} \tilde{\mathbf{I}}_{K+1} \mathbf{x}_{n;1}(l) & \tilde{\mathbf{I}}_{K+1} \mathbf{x}_{n;2}(l) & \cdots & \tilde{\mathbf{I}}_{K+1} \mathbf{x}_{n;N_1-K}(l) \end{bmatrix}^T \quad (5.35)$$

with $\mathbf{x}_{n;k}(l)$ as the l th column vector of

$$\tilde{\mathbf{R}}_{\text{ns};k} = \mathbf{H}_k \mathbf{R}_{\text{ns}}^{1/2} \quad (5.36)$$

Note that, for fixed $\mathbf{H}$ and $\mathbf{R}_{\text{ns}}$, $\mathbf{X}_n(l)$ can be computed off-line.

When the conjugate-symmetric constraint is applied by letting $\tilde{\mathbf{H}} = \tilde{\mathbf{I}}_{K+1} \tilde{\mathbf{J}}_{K+1}^T \tilde{\mathbf{I}}_{K+1}$, the vector $\mathbf{b} = \tilde{\mathbf{J}} \tilde{\mathbf{b}}$. The estimate of the real vector $\tilde{\mathbf{b}} = [\tilde{b}_0 \ \tilde{b}_1 \ \cdots \ \tilde{b}_K]^T$ is obtained from

$$\hat{\tilde{\mathbf{b}}} = \arg \min_{\tilde{\mathbf{b}} \in \Omega^{(K+1) \times 1}} \tilde{\mathbf{b}}^T \left\{ \text{Re} \left(\sum_{t=0}^{L-1} \mathbf{X}_t^H(t) \mathbf{F}_i \mathbf{X}_t(t) \right) \right\} \tilde{\mathbf{b}} \quad (5.37)$$

subject to one of the three nontriviality constraints Eqs. (5.28, 5.31, 5.33) with $\mathbf{b}$ and $\hat{\mathbf{b}}_i$ in these and related equations being replaced by $\tilde{\mathbf{b}}$ and $\hat{\tilde{\mathbf{b}}}_i$. In addition, $\Omega^{(K+1) \times 1}$ is, in this case, the set of $(K+1) \times 1$ real vectors.

For $N_1 \gg K$, the main computational part of the IQML algorithm is in computing $(\mathbf{X}_t^H(t) \mathbf{F}_i \mathbf{X}_t(t))$ at each iteration. If $N_1 = N$, $\mathbf{R}_{\text{ns}} = \tilde{\mathbf{H}} = \mathbf{I}_N$, the matrix term $(\hat{\mathbf{B}}_{i-1}^H \mathbf{H} \mathbf{R}_{\text{ns}} \mathbf{H}^H \hat{\mathbf{B}}_{i-1})$ is banded with the one-side bandwidth K (excluding the main diagonal). In this case, based on Hua's implementation of the algorithm [71], for $J = (N_1 - K) \gg K$, the approximate number of floating-point complex multiplications required by both Bresler's and Nagesha's IQML algorithms is

$$\text{Flop}_1 = (3L + 1)JK^2/2 + (7L + 3)JK/2 + 2LJ \quad (5.38)$$

In addition to Eq. (5.38), both algorithms require the $O((K+1)^3)$ computation to obtain $\hat{\tilde{\mathbf{b}}}_i$ at each iteration from Eq. (5.27), subject to the nontriviality constraint Eq. (5.28) for Bresler's algorithm, and the constraint Eq. (5.31) for Nagesha's algorithm.

For the proposed algorithm, additional computation is required to obtain $\tilde{T}_i$ by Eq. (5.34). If $N_1 = N$, $R_{\text{ns}} = \tilde{H} = I_N$, and $\tilde{H} = I_{K+1}$ (or $\tilde{I}_{K+1}\tilde{J}_{K+1}^T\tilde{I}_{K+1}$), formation of $\tilde{T}_i$ by Eq. (5.34) involves only addition operations and re-indexing of the elements in F_i . Because $(\hat{B}_{i-1}^H H R_{\text{ns}} H^H \hat{B}_{i-1})$ is banded and Toeplitz, its inverse, which is F_i , can be obtained by a slight modification of the Trench algorithm [72], to take into account the band structure of $(\hat{B}_{i-1}^H H R_{\text{ns}} H^H \hat{B}_{i-1})$. Based on this algorithm, the approximate number of required multiplications to obtain F_i is given by

$$\text{Flop}_2 = 5J^2/4 + KJ \quad (5.39)$$

The total number of multiplications required by the proposed algorithm is the combination of Flop_1 and Flop_2 , plus the $O((K+1)^3)$ computation required to obtain $\hat{b}_i$ from Eq. (5.27), subject to the constraint Eq. (5.33).

When $(\hat{B}_{i-1}^H H R_{\text{ns}} H^H \hat{B}_{i-1})$ has no special structure, based on Hua's implementation, the required number of multiplications in Bresler's and Nagesha's algorithms is approximately $J^3/6 + (3K^2 + 7K + 4)JL/2$, while the proposed algorithm requires approximately $J^3/6 + (3K^2 + 7K + 4)(L + N_1)J/2$ multiplication operations. All three algorithms require in addition the $O((K+1)^3)$ computation to obtain the solution to Eq. (5.27).

From this computational count, it is seen that the proposed algorithm is more computationally complicated than the two existing algorithms. The higher complexity of the proposed algorithm is, however, partly compensated for by faster convergence rate. In the following section, the inclusion of additional constraints on b based on prior knowledge of some parameter values will be discussed.

5.3 The constrained IQML algorithm

In this section, we consider the case where some of the parameter values are known a priori. Let $z_l^c, l = 1, \dots, K_1$, which correspond to $\theta_{z,l}^c \in \{\theta_{z,k} \mid k = 1, \dots, K\}$, $l = 1, \dots, K_1$, are known a priori. In addition, the polynomial $f(\mathbf{b}^u) = b_0^u z^{K-K_1} + b_1^u z^{K-K_1-1} + \dots + b_{K-K_1}^u$, has $K - K_1$ roots $z_l^u, l = 1, \dots, K_1$, which correspond to the remaining $K - K_1$ parameters $\theta_{z,l}^u \in \{\theta_{z,k} \mid k = 1, \dots, K\}, l = 1, \dots, K - K_1$. By letting $\mathbf{b}^u = [b_0^u \ b_1^u \ \dots \ b_{K-K_1}^u]^T$, the vector $\mathbf{b}$ can be written as

$$\mathbf{b} = \tilde{\mathbf{I}}_{K+1} \mathbf{B}_c^* \tilde{\mathbf{I}}_{K-K_1+1} \mathbf{b}^u \quad (5.40)$$

$$\mathbf{B}_c = \begin{bmatrix} b_{K_1}^c & b_{K_1-1}^c & \dots & b_0^c & & & 0 \\ & b_{K_1}^c & b_{K_1-1}^c & \dots & b_0^c & & \\ & & \ddots & & & \ddots & \\ 0 & & & b_{K_1}^c & b_{K_1-1}^c & \dots & b_0^c \end{bmatrix}^H \quad (5.41)$$

where $z_l^c, l = 1, \dots, K_1$, are the K_1 roots of the polynomial $g(\mathbf{b}^c) = b_0^c z^{K_1} + b_1^c z^{K_1-1} + \dots + b_{K_1}^c$, with $\mathbf{b}^c = [b_0^c \ b_1^c \ \dots \ b_{K_1}^c]^T$. If $\tilde{\mathbf{H}}$ is chosen as $\mathbf{B}_c^H$, the vector $\mathbf{b}^u$ can be estimated using any of the three described algorithms with $\mathbf{b}$ and K being replaced by $\mathbf{b}^u$ and $K - K_1$. Computational saving can be gained in solving this *constrained* IQML problem because of smaller dimensionality of $\mathbf{b}^u$, compared with that of $\mathbf{b}$. More importantly, the performance of the constrained IQML algorithm is generally better than that of the unconstrained algorithm.

In studying the performance improvement obtained by the constrained IQML algorithm, it is useful to know what is the performance bound under this condition. Based on the compact CRLB formula derived in [9], which is formulated under the assumption that $\mathbf{s}(t)$ is a deterministic unknown parameter vector, the CRLB of an unbiased estimate of the unknown parameter vector $\boldsymbol{\theta}^u = [\theta_{z,1}^u \ \theta_{z,2}^u \ \dots \ \theta_{z,K-K_1}^u]^T$, given the known parameter vector $\boldsymbol{\theta}^c = [\theta_{z,1}^c \ \theta_{z,2}^c \ \dots \ \theta_{z,K_1}^c]^T$, is

$$\text{CRLB}(\hat{\theta}^u) \Big|_{\theta^c} = \frac{1}{2L} \text{Re}^{-1} \{ (\bar{\mathbf{D}}_u^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{D}}_u) \odot \hat{\mathbf{R}}_{s_u}^T \} \quad (5.42)$$

where $\hat{\mathbf{R}}_{s_u} = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{s}_u(t) \mathbf{s}_u^H(t)$, and $\mathbf{s}_u(t)$ is composed of the signal waveforms associated with the unknown parameters. The matrix $\bar{\mathbf{D}}_u$ is given by

$$\bar{\mathbf{D}}_u = \begin{bmatrix} \left. \frac{\partial \mathbf{a}_{u;1}}{\partial \theta_{z;1}^u} \right|_{\bar{\theta}_{z;1}^u} & \left. \frac{\partial \mathbf{a}_{u;2}}{\partial \theta_{z;2}^u} \right|_{\bar{\theta}_{z;2}^u} & \cdots & \left. \frac{\partial \mathbf{a}_{u;K_u}}{\partial \theta_{z;K_u}^u} \right|_{\bar{\theta}_{z;K_u}^u} \end{bmatrix} \quad (5.43)$$

where $\mathbf{a}_{u;k}$, $k = 1, \dots, K_u$, are the response vectors corresponding to the unknown parameters $\theta_{z;k}^u$, $k = 1, \dots, K_u$, and the number of unknown parameters $K_u = K - K_1$. Note that $\bar{\mathbf{A}}$ is the response matrix corresponding to the true values of both the known and unknown parameters.

One remark to be made is the connection between the above constrained algorithm and beamspace processing [40-42]. The IQML algorithm with the constraint Eq. (5.40) can be seen as the unconstrained algorithm being applied to the beamspace data snapshots $\mathbf{T}_{\text{BS}}^H \mathbf{x}(t)$, $t = 0, \dots, L - 1$, where $\mathbf{T}_{\text{BS}}$ is $\mathbf{B}_c$ in Eq. (5.41) with $N - K_1$ column vectors. In [42], a similar constraint was proposed to be applied to the MUSIC method. However, it was seen by computer simulation that the performance of the constrained IQML algorithm is superior to that of the constrained minimum-norm method, which is a variant of the method in [42].

5.4 Two-dimensional parameter estimation based on the IQML algorithm

The two-dimensional parameter estimation problem has been considered, for example, in [62-66]. However, these methods are either inefficient in terms of performance [65], or limited in applicability [62-63]. In [64], Hua proposed the use of two 1-D frequency estimators, based on the matrix pencil principle, to estimate separately each of the two parameter components. Pairing between the two components

is carried out in a subsequent step. Despite computational saving, Hua's method is inefficient because it operates on the subdata set of reduced dimensionality. To improve the estimation performance, a subdata set with larger dimension is required [64]. With the increased size of the subdata set, however, the computational complexity of the method increases rapidly due to the need for performing matrix eigendecomposition (or SVD decomposition).

From the shortcomings of the methods as mentioned, we propose an alternative 2-D parameter estimator. In a manner similar to Hua's MEMP (Matrix Enhancement and Matrix Pencil) method, our method breaks the 2-D parameter estimation problem into two 1-D parameter estimation problems. Unlike the former method, however, the two 1-D estimation problems are solved by using the IQML algorithm instead of the subspace-based method which is employed in [64]. As can be seen later, the method achieves comparable or even superior performance to that of the MEMP method. In addition, the method can be implemented by means of parallel processing. Before considering the proposed 2-D method, however, we first provide a solution to the 2-D parameter pairing problem.

First, for ease of presentation, we assume in this section that $\mathbf{R}_{\text{ns}} = \mathbf{I}_{MN}$. Next, let $\mathbf{A}_s(\mathbf{y}, \mathbf{z})$ be composed of p row vectors in $\mathbf{A}(\mathbf{y}, \mathbf{z})$. In addition, let $\bar{\mathbf{A}}_s$ be $\mathbf{A}_s^\dagger$ which corresponds to the true values of the K 2-D parameter pairs $(\bar{\mathbf{y}}_k, \bar{\mathbf{z}}_k) \Leftrightarrow (\bar{\boldsymbol{\theta}}_{\mathbf{y};k}, \bar{\boldsymbol{\theta}}_{\mathbf{z};k})$, $k = 1, \dots, K$. If the column vectors of $\bar{\mathbf{A}}_s$ span a unique K -dimensional subspace for a given set of K distinct 2-D parameter pairs, there exists a $p \times (p - K)$ matrix $\mathbf{V}$ such that

$$\bar{\mathbf{A}}_s^H \mathbf{V} = \mathbf{0} \quad (5.44)$$

It is seen that, the matrix $\mathbf{V}$, whose column vectors span the null space of $\bar{\mathbf{A}}_s^H$, provides sufficient information for pairing the two 1-D parameter components $\bar{\mathbf{y}}_k$ and $\bar{\mathbf{z}}_k$. The

[†] From this point, the explicit expression of $\mathbf{A}_s$ as a function of $(\mathbf{y}, \mathbf{z})$ will, for most of the time, be omitted for notational convenience.

column vectors of V can be obtained, for example, from the $p - K$ noise-subspace vectors of the average data covariance matrix

$$\mathbf{R}_{\text{av}} = \frac{1}{(M - q + 1)(N - v + 1)} \sum_{k=1}^{M-q+1} \sum_{l=1}^{N-v+1} \mathbf{R}_{k,l} \quad (5.45)$$

where $\mathbf{R}_{k,l} = \mathbb{E}\{\mathbf{x}_{k,l}(t)\mathbf{x}_{k,l}^H(t)\}$. The column vector $\mathbf{x}_{k,l}(t)$ of length $p = q.v$ is obtained by stacking the column vectors of $\mathbf{X}_{k,l}^T(t)$, i.e., $\mathbf{x}_{k,l}(t) = \text{vec}(\mathbf{X}_{k,l}^T(t))$, where the $q \times v$ subdata matrix $\mathbf{X}_{k,l}(t)$ is given by

$$\mathbf{X}_{k,l}(t) = \begin{bmatrix} x_{k,l}(t) & x_{k,l+1}(t) & \cdots & x_{k,l+v-1}(t) \\ x_{k+1,l}(t) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{k+q-1,l}(t) & \cdots & \cdots & x_{k+q-1,l+v-1}(t) \end{bmatrix} \quad (5.46)$$

If $q, v \geq K$, $M \geq K + q - 1$, and $N \geq K + v - 1$, the noise subspace vectors of $\mathbf{R}_{\text{av}}$, i.e., the $p - K$ eigenvectors of $\mathbf{R}_{\text{av}}$ which correspond to the $p - K$ smallest eigenvalues, form the null space of $\bar{\mathbf{A}}_s^H$. This choice of $\mathbf{x}_{k,l}(t)$, q , v , M , and N is adopted in [64]. The 2-D parameter pairing can be carried out by choosing K distinct pairs of the parameter estimates from the two dimensions, which are

$$(\hat{\mathbf{y}}, \hat{\mathbf{z}}) = \arg \min_{\substack{(\mathbf{y}_k, \mathbf{z}_k) \in \Xi \\ k=1, \dots, K}} \text{tr}\{\mathbf{A}_s^H \hat{\mathbf{V}} \hat{\mathbf{V}}^H \mathbf{A}_s\} \quad (5.47)$$

where $(\hat{\mathbf{y}}, \hat{\mathbf{z}}) \triangleq \{(\hat{\mathbf{y}}_k, \hat{\mathbf{z}}_k) \mid k = 1, \dots, K\}$ and Ξ is the set of all possible 2-D pairs of the given two sets of the 1-D parameter estimates. The vector $\hat{\mathbf{V}}$ is the estimate of V . It will be shown later how to obtain $\hat{\mathbf{V}}$. Given $\bar{\mathbf{y}}_k$ and $\bar{\mathbf{z}}_k$, if we let $\hat{\mathbf{V}} = V$, it is seen that $(\hat{\mathbf{y}}, \hat{\mathbf{z}})$, obtained by Eq. (5.47), is equal to $(\bar{\mathbf{y}}, \bar{\mathbf{z}}) \triangleq \{(\bar{\mathbf{y}}_k, \bar{\mathbf{z}}_k) \mid k = 1, \dots, K\}$. Although this choice of the subdata vector $\mathbf{x}_{k,l}(t)$ results in a consistent 2-D parameter pairing, the dimension of $\mathbf{x}_{k,l}(t)$ and $\mathbf{R}_{k,l}$ can be unnecessarily large when the number of the parameter pairs K increases. In fact, a subdata vector which gives a consistent pairing by Eq. (5.47) can be smaller than the described one. It can be seen that, given the distinct-

value parameters in one of the two dimensions, a consistent pairing can be done if A_s has the following properties:

- P.1 Given a set of K distinct 2-D parameter pairs, the K column vectors of A_s span a K -dimensional subspace.
- P.2 Let the $K_y \times 1$ vector y^o be composed of K_y distinct-value y-dimension parameters $y_k^o, k = 1, \dots, K_y$. Similarly, the $K_z \times 1$ vector z^o is composed of K_z distinct-value z-dimension parameters $z_k^o, k = 1, \dots, K_z$. Let $\{(y_{1,k}, z_{1,k}) \mid k = 1, \dots, K\}$ and $\{(y_{2,k}, z_{2,k}) \mid k = 1, \dots, K\}$ be two sets of K distinct 2-D parameter pairs. If y^o is given, it is assumed that for $i = 1, 2$, $y_{i,k} \in \{y_l^o \mid l = 1, \dots, K_y\}, k = 1, \dots, K$, with the number of distinct-value parameters K_y . Similarly, if z^o is given, it is assumed that for $i = 1, 2$, $z_{i,k} \in \{z_l^o \mid l = 1, \dots, K_z\}, k = 1, \dots, K$, with the number of distinct-value parameters K_z . Denote the given distinct-value parameter vector (either y^o or z^o) by w^o . In addition, let $y_i = [y_{i,1} \ y_{i,2} \ \dots \ y_{i,K}]^T$, and $z_i = [z_{i,1} \ z_{i,2} \ \dots \ z_{i,K}]^T$, for $i = 1, 2$. Given $A_s(y_1, z_1)|_{w^o}$ and $A_s(y_2, z_2)|_{w^o}$, each with the property P.1, there exists a full rank matrix C such that

$$A_s(y_1, z_1)|_{w^o} \cdot C = A_s(y_2, z_2)|_{w^o} \quad (5.48)$$

only if the 2-D parameter set $\{(y_{1,k}, z_{1,k}) \mid k = 1, \dots, K\}$ is equal to $\{(y_{2,k}, z_{2,k}) \mid k = 1, \dots, K\}$. The notation $A_s(y, z)|_{w^o}$ denotes the sub-response matrix as a function of (y, z) , given w^o .

The above requirement on A_s is noted to be slightly weaker than that of [64]. In the following theorem, we provide a fairly general rule for building $x_{k,l}(t)$ of which the corresponding sub-response matrix, denoted by $A_{s;k,l}$, has the two properties as described.

Theorem 5.1: Let z^0 be given. If, by selecting some elements in the subdata vector $\mathbf{x}_{k,l}(t)$, the following vectors can be constructed:

- a) $\mathbf{x}_\alpha(t) = [x_{m_1, n_1} \quad x_{m_1, n_1+1} \quad \cdots \quad x_{m_1, n_1+l_1-1}]^T, \quad l_1 \geq K_z,$
- b) $\mathbf{x}_\beta(t) = [x_{m_2, n_2} \quad x_{m_2, n_2+1} \quad \cdots \quad x_{m_2, n_2+l_2-1}]^T, \quad l_2 \geq \min(K - K_z + 1, K_z) \text{ and } m_2 \neq m_1,$
- c) $\mathbf{x}_\gamma(t) = [x_{m_3, n_3} \quad x_{m_3+1, n_3} \quad \cdots \quad x_{m_3+l_3-1, n_3}]^T, \quad l_3 \geq K - K_z + 2,$

then $\mathbf{A}_{s;k,l}$ has the properties P.1 and P.2 (with $\mathbf{w}^0 = \mathbf{z}^0$ in P.2). For $\mathbf{A}_{s;k,l}$ to have (at least) the property P.1, it is sufficient that $\mathbf{x}_\alpha(t)$ with $l_1 \geq K_z$ and $\mathbf{x}_\gamma(t)$ with $l_3 \geq K - K_z + 1$ can be constructed from $\mathbf{x}_{k,l}(t)$.

Proof: See Appendix G.

When the given distinct-value parameter vector $\mathbf{w}^0$ in P.2 is $\mathbf{y}^0$, a sufficient condition similar to the one given by Theorem 5.1 for the case $\mathbf{w}^0 = \mathbf{z}^0$ can be obtained. In addition, the subdata vector $\mathbf{x}_{k,l}(t)$ of which the sub-response matrix $\mathbf{A}_{s;k,l}$ has both properties P.1-2 can be used not only for 2-D parameter pairing, but also in estimating the unknown parameters in the other dimension. For instance, if $\mathbf{w}^0 = \mathbf{z}^0$ is given and $\mathbf{A}_{s;k,l}$ has the properties P.1-2, $\mathbf{y}$ can be estimated by

$$(\hat{\mathbf{y}}, \hat{\mathbf{z}}) = \arg \min_{(\mathbf{y}, \mathbf{z})} \text{tr}\{\mathbf{A}_s^H \hat{\mathbf{V}} \hat{\mathbf{V}}^H \mathbf{A}_s\} \quad (5.49)$$

subject to $z_k \in \{z_l^0 \mid l = 1, \dots, K_z\}, k = 1, \dots, K$. If $\hat{\mathbf{V}}$ is a consistent estimate of $\mathbf{V}$, the estimator Eq. (5.49) gives a consistent estimate of $(\mathbf{y}, \mathbf{z})$. Moreover, if $\mathbf{x}_{k,l}(t)$ is constructed according to Theorem 5.2 with either l_1 or $l_2 > K_z$, $\mathbf{A}_{s;k,l}$ span a unique K dimensional subspace for a given set of K distinct 2-D parameter pairs with known K_z . This implies that, with sufficient matrix averaging in Eq. (5.45), $\mathbf{R}_{av}$ contains sufficient information for the estimation of the K 2-D parameter pairs without prior information of either $\mathbf{y}^0$ or $\mathbf{z}^0$. It should be noted, however, that the condition as stated in Theorem 5.1

is sufficient but not necessary. Sometimes, $A_{s;k,l}$ can have the two properties P.1-2 with smaller $\mathbf{x}_{k,l}(t)$ than the one obtained by the theorem. This is the case when K_y and K_z are comparatively small, when compared with K . For instance, if the size of the vector $\mathbf{x}_{k,l}(t)$ which is constructed by Theorem 5.1 is larger than $(K_y K_z)$, $\mathbf{x}_{k,l}(t)$ may be chosen instead as $\mathbf{x}_{k,l}(t) = \text{vec}(X_{k,l}^T(t))$, where the $q \times v$ subdata matrix $X_{k,l}(t)$ is as given by Eq. (5.46) with $q \geq K_y$, $v \geq K_z$. This choice of $\mathbf{x}_{k,l}(t)$ (and $\mathbf{R}_{k,l}$) was suggested in [73]. When K_y and K_z are comparable to K , however, this alternative $\mathbf{x}_{k,l}(t)$ (and $\mathbf{R}_{k,l}$) generally has larger dimension than the one obtained according to Theorem 5.1.

Assuming that $A_{s;k,l}$ has at least the property P.1, and the signal covariance matrix $\mathbf{R}_s$ is positive semi-definite, the following theorem provides a sufficient condition in constructing $\mathbf{R}_{av}$ such that its noise subspace vectors span the null space of $A_{s;k,l}^H$. The condition stated in the theorem is noted to be more general than the condition $M \geq K + q - 1$, $N \geq K + v - 1$, used for the construction of $\mathbf{R}_{av}$ by Eq. (5.45).

Theorem 5.2: Assume that $\mathbf{R}_{av}$ is formed by averaging r subdata covariance matrices $\mathbf{R}_{k,l}$, i.e.,

$$\mathbf{R}_{av} = \frac{1}{r} \sum_{(k,l) \in \Psi} \mathbf{R}_{k,l} \quad (5.50)$$

where the $p \times p$ matrix $\mathbf{R}_{k,l} = E\{\mathbf{x}_{k,l}(t)\mathbf{x}_{k,l}^H(t)\}$ with $p > K$, and Ψ is the set of the data indices $\{(k_1, l_1), \dots, (k_r, l_r)\}$ whose members form the *centre* data vector

$$\mathbf{x}_c(t) = \begin{bmatrix} x_{k_1, l_1}(t) & x_{k_2, l_2}(t) & \cdots & x_{k_r, l_r}(t) \end{bmatrix}^T \quad (5.51)$$

If $\mathbf{x}_c(t)$ and $\mathbf{x}_{k,l}(t)$, $(k, l) \in \Psi$, are formed according to Theorem 5.1 such that the corresponding response matrices have at least the property P.1, then the noise subspace vectors of $\mathbf{R}_{av}$, which are the $p - K$ eigenvectors of $\mathbf{R}_{av}$ associated with the $p - K$ smallest eigenvalues, span the nullspace of $A_{s;k,l}^H$, for $(k, l) \in \Psi$.

Proof: See Appendix H.

From Theorem 5.2, V can be obtained as composed of the $p - K$ noise subspace vectors of $\mathbf{R}_{av}$. Regarding the formation of the subdata vectors $\mathbf{x}_{k,l}(t)$, $(k, l) \in \Psi$, such that their response matrices have the properties P.1 and P.2, it is seen that the formation may not be unique. Nevertheless, computational saving as well as performance enhancement can be gained if $\mathbf{A}_{s;k,l}$ can be decomposed as $\mathbf{A}_{s;k,l} = \mathbf{C}_{cs}\mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{C}_{cs}$ is column conjugate-symmetric, i.e., $\mathbf{C}_{cs} = \tilde{\mathbf{I}}\mathbf{C}_{cs}^*$, where $\tilde{\mathbf{I}}$ is assumed to have a suitable dimension. With this property, the eigendecomposition of $\mathbf{R}_{av}$ can be carried out with a real rather than complex matrix. In addition, performance improvement is gained because this allows for backward smoothing of $\mathbf{R}_{av}$, in addition to forward smoothing used in Eq. (5.50). Detail of the efficient methods in forming $\mathbf{R}_{av}$ by forward/backward smoothing can be found in [74].

With the development of Theorems 5.1 and 5.2, consistent 2-D parameter pairing can be carried out. We next provide the proposed 2-D method as follows:

The 1-D IQML-based 2-D parameter estimation method

1. From the data matrix $\mathbf{X}(t)$ of size $M \times N$ (Eq. (5.3)), construct two sets of data snapshots as follows:

$$\mathbf{x}_y(n, t) = \begin{bmatrix} x_{1,n}(t) & x_{2,n}(t) & \cdots & x_{M,n}(t) \end{bmatrix}^T, n = 1, \dots, N \quad (5.52)$$

$$\mathbf{x}_z(m, t) = \begin{bmatrix} x_{m,1}(t) & x_{m,2}(t) & \cdots & x_{m,N}(t) \end{bmatrix}^T, m = 1, \dots, M \quad (5.53)$$

Assuming that K_y and K_z are known, apply the IQML algorithm (any of the three versions described in Section 5.2.2) to $\mathbf{x}_z(m, t)$ to estimate the z-dimension parameters, and to $\mathbf{x}_y(n, t)$ to estimate the y-dimension parameters. The uniqueness of the parameter estimates is guaranteed by choosing $N \geq 2K_z$, and $M \geq 2K_y$ (see [50-51] for more details on the uniqueness problem).

2. Estimate the average covariance matrix $\mathbf{R}_{av}$ by

$$\hat{\mathbf{R}}_{av} = \frac{1}{rL} \sum_{(k,l) \in \Psi} \sum_{t=0}^{L-1} \mathbf{x}_{k,l}(t) \mathbf{x}_{k,l}^H(t) \quad (5.54)$$

with $\mathbf{x}_{k,l}(t)$ satisfying the condition stated in Theorem 5.1 and Ψ being chosen according to Theorem 5.2. Note that, if each of the sub-response matrices corresponding to $\mathbf{x}_{k,l}(t)$ can be decomposed as $\mathbf{A}_{s;k,l} = \mathbf{C}_{cs}\mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{C}_{cs}$ is column conjugate-symmetric, i.e., $\mathbf{C}_{cs} = \tilde{\mathbf{I}}\mathbf{C}_{cs}^*$, the forward/backward averaging matrix

$$\hat{\mathbf{R}}_{av}^{fb} = (\hat{\mathbf{R}}_{av} + \tilde{\mathbf{I}}\hat{\mathbf{R}}_{av}^*\tilde{\mathbf{I}})/2 \quad (5.55)$$

should be used instead of $\hat{\mathbf{R}}_{av}$. In addition, for the single data snapshot case, more subdata matrices than the ones stated in Theorem 5.2 may be required for sufficient noise averaging. From the estimates of the y- and z-dimension parameters, the parameter pairing is carried out using Eq. (5.47), with $\hat{\mathbf{V}}$ being composed of the $p - K$ eigenvectors of $\hat{\mathbf{R}}_{av}$ which are associated with the $p - K$ smallest eigenvalues.

We consider next the computational complexity of the proposed 2-D method. The main computational parts of the above 2-D method lie in the estimation of two 1-D parameters by the IQML algorithm, the formation of $\hat{\mathbf{R}}_{av}$ and its eigenvector decomposition. The computational complexity of the IQML algorithm for two 1-D parameter estimations can be obtained from the discussion on this topic in Section 5.2. Let $\hat{\mathbf{R}}_{av}$ be formed (through $\mathbf{x}_{k,l}(t)$) from $(M - q + 1)(N - v + 1)$ subdata matrices $\mathbf{X}_{k,l}(t)$ of dimension $q \times v$. For the single data snapshot case ($L = 1$), the required number of multiplications to form $\hat{\mathbf{R}}_{av}$ is

$$\begin{aligned} \text{Flop}_3 = & qv(qv + 1)/2 + (N - v)(qv - (q - 1)q/2) + \\ & (M - q)(v(qv) - (v - 1)v/2) + (M - q)(N - v)(qv) \end{aligned} \quad (5.56)$$

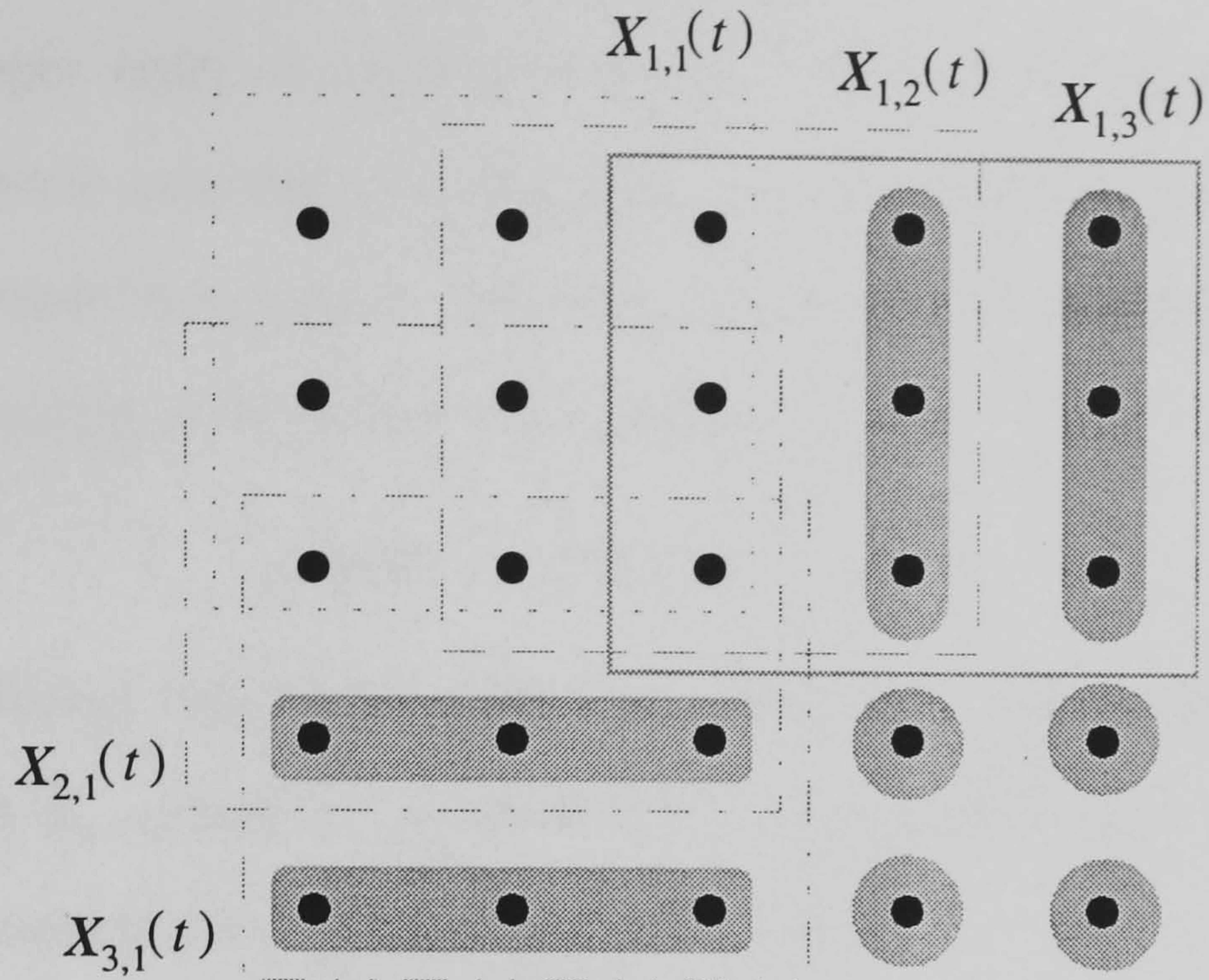


Figure 5.1: The 2-D 5×5 data matrix.

To see how this computational count is obtained, we consider the 2-D 5×5 data matrix as shown in Figure 5.1. From the figure, $q = v = 3$, and the total number of subdata matrices $r = 9$. Only the single data snapshot case is considered. Let $\mathbf{x}_{1,1}(t) = \text{vec}(X_{1,1}^T(t))$, where $X_{1,1}(t)$ is the first subdata matrix (see the figure and Eq. (5.46)). It is seen that, by taking into account the Hermitian structure, the number of multiplications required to form $\mathbf{x}_{1,1}(t)\mathbf{x}_{1,1}^H(t)$ is

$$\text{count\#1} = qv(qv + 1)/2 \quad (5.57)$$

Next, to compute $\mathbf{x}_{1,2}(t)\mathbf{x}_{1,2}^H(t)$, only the product of each of the three shaded data points in $X_{1,2}(t)$ to itself, as well as to other data points in $X_{1,2}(t)$, are needed to be computed. The same number of computations is required to compute $\mathbf{x}_{1,3}(t)\mathbf{x}_{1,3}^H(t)$. Thus, the total number of multiplications required at this stage is

$$\text{count\#2} = (N - v)(q(qv) - (q - 1)q/2) \quad (5.58)$$

In a similar manner to the case of $\mathbf{x}_{1,2}(t)\mathbf{x}_{1,2}^H(t)$ and $\mathbf{x}_{1,3}(t)\mathbf{x}_{1,3}^H(t)$, the total number of multiplications required to form $\mathbf{x}_{2,1}(t)\mathbf{x}_{2,1}^H(t)$ and $\mathbf{x}_{3,1}(t)\mathbf{x}_{3,1}^H(t)$ is

$$\text{count\#3} = (M - q)(v(qv) - (v - 1)v/2) \quad (5.59)$$

With proper order of computation, for each of the remaining outer product $\mathbf{x}_{k,l}(t)\mathbf{x}_{k,l}^H(t)$, there is only one new data point whose products with itself and the other points in $\mathbf{x}_{k,l}(t)$ must be computed. Thus, the total number of multiplications required to compute the remaining outer product terms $\mathbf{x}_{k,l}(t)\mathbf{x}_{k,l}^H(t)$ is

$$\text{count\#4} = (M - q)(N - v)(qv) \quad (5.60)$$

By combining Eqs. (5.57-5.60), we obtain the total number of operations required to form $\hat{\mathbf{R}}_{\text{av}}$, which is given by Eq. (5.56). In addition to this, the $O((qv)^3)$ computation is required for eigendecomposition of $\hat{\mathbf{R}}_{\text{av}}$.

Finally, to facilitate the performance study of 2-D methods in the next section, a compact CRLB for the 2-D *distinct-value* parameter vector $\boldsymbol{\theta}^o = [(\boldsymbol{\theta}_y^o)^T \ (\boldsymbol{\theta}_z^o)^T]^T$, is given by

$$\text{CRLB}(\hat{\boldsymbol{\theta}}^o) = \frac{1}{2L} \text{Re}^{-1} \{ \tilde{\mathbf{U}}_I^T \{ (\bar{\mathbf{D}}_{2d}^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2}\bar{\mathbf{A}}}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{D}}_{2d}) \odot (\hat{\mathbf{R}}_s)^T \} \tilde{\mathbf{U}}_I \} \quad (5.61)$$

where

$$\bar{\mathbf{D}}_{2d} = \begin{bmatrix} \left. \frac{\partial A}{\partial \theta_{y,1}^o} \right|_{\bar{\boldsymbol{\theta}}} & \left. \frac{\partial A}{\partial \theta_{y,2}^o} \right|_{\bar{\boldsymbol{\theta}}} & \cdots & \left. \frac{\partial A}{\partial \theta_{y,K_y}^o} \right|_{\bar{\boldsymbol{\theta}}} & \left. \frac{\partial A}{\partial \theta_{z,1}^o} \right|_{\bar{\boldsymbol{\theta}}} & \cdots & \cdots & \left. \frac{\partial A}{\partial \theta_{z,K_z}^o} \right|_{\bar{\boldsymbol{\theta}}} \end{bmatrix} \quad (5.62)$$

$$\tilde{\mathbf{U}}_I = \mathbf{I}_{K_{y+z}} \otimes \mathbf{I}_K \quad (5.63)$$

$$\hat{\mathbf{R}}_s = (\mathbf{I}_{K_{y+z}} \mathbf{I}_{K_{y+z}}^T) \otimes \hat{\mathbf{R}}_s \quad (5.64)$$

$$\hat{\mathbf{R}}_s = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{s}(t) \mathbf{s}^H(t) \quad (5.65)$$

The matrix $\mathbf{P}_{\mathbf{R}_n^{-1/2}\bar{\mathbf{A}}}^\perp$ is the null projection of the column space of $\mathbf{R}_n^{-1/2}\bar{\mathbf{A}}$, and $K_{y+z} = K_y + K_z$. As for $\mathbf{y}^o$ and $\mathbf{z}^o$, the vectors $\boldsymbol{\theta}_y^o$ and $\boldsymbol{\theta}_z^o$ are composed of, respectively, the K_y distinct-value y-dimension parameters $\theta_{y,k}^o, k = 1, \dots, K_y$, and the K_z distinct-value z-dimension parameters $\theta_{z,k}^o, k = 1, \dots, K_z$. When the number of distinct parameter values in each of the two dimensions is not assumed known, $K_y = K_z = K$, $\boldsymbol{\theta}^o = \boldsymbol{\theta}$, and

θ_y^o (θ_z^o) is treated as being composed of K distinct parameters. When this is the case, a somewhat simpler CRLB formula is given by

$$\text{CRLB}(\hat{\theta}) = \frac{1}{2L} \text{Re}^{-1} \{ (\bar{\tilde{D}}_{2d}^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{A}}^\perp \mathbf{R}_n^{-1/2} \bar{\tilde{D}}_{2d}) \odot ((\mathbf{I}_2 \mathbf{I}_2^T) \otimes \hat{\mathbf{R}}_s)^T \} \quad (5.66)$$

where

$$\bar{\tilde{D}}_{2d} = \begin{bmatrix} \left. \frac{\partial a_1}{\partial \theta_{y,1}} \right|_{\bar{\theta}_{y,1}} & \left. \frac{\partial a_2}{\partial \theta_{y,2}} \right|_{\bar{\theta}_{y,2}} & \cdots & \left. \frac{\partial a_K}{\partial \theta_{y,K}} \right|_{\bar{\theta}_{y,K}} & \left. \frac{\partial a_1}{\partial \theta_{z,1}} \right|_{\bar{\theta}_{z,1}} & \cdots & \left. \frac{\partial a_K}{\partial \theta_{z,K}} \right|_{\bar{\theta}_{z,K}} \end{bmatrix} \quad (5.67)$$

The formulas Eqs. (5.61, 5.66) can be derived from the results on the compact CRLB in [9]. A more general form of the bound Eq. (5.66) can be found in [75]. The proof for Eq. (5.61) is sketched in Appendix I. Note that, for large L , Eqs. (5.61, 5.66) can be approximated by replacing $\hat{\mathbf{R}}_s$ with $\mathbf{R}_s$ [9].

5.5 Numerical results

Some computer simulations were carried out to compare our methods with theoretical bounds and some existing methods. In the first six experiments, we demonstrate the performance improvement by the proposed consistent IQML algorithm employing the new nontriviality constraint, in comparison with the algorithms due to Bresler [22] and Nagesha [24]. The last two experiments are dedicated to the problem of 2-D parameter estimation. In all experiments, the additive noise is white and the noise covariance structure $\mathbf{R}_{ns} = \mathbf{I}_N$ ($= \mathbf{I}_{MN}$ for the last two experiments). Except for two constrained IQML methods in EXP 5.6, $\tilde{\mathbf{H}} = \tilde{\mathbf{I}}_{K+1} \tilde{\mathbf{J}}_{K+1}^T \tilde{\mathbf{I}}_{K+1}$, $N_1 = N$, and $\tilde{\mathbf{H}} = \mathbf{I}_N$ for the 1-D parameter estimation problem. For the IQML algorithm used in the 2-D parameter estimation problem, $\tilde{\mathbf{H}} = \tilde{\mathbf{I}}_{K_y+1} \tilde{\mathbf{J}}_{K_y+1}^T \tilde{\mathbf{I}}_{K_y+1}$, $N_1 = M$, and $\tilde{\mathbf{H}} = \mathbf{I}_M$ for the y-dimension parameter estimation, and $\tilde{\mathbf{H}} = \tilde{\mathbf{I}}_{K_z+1} \tilde{\mathbf{J}}_{K_z+1}^T \tilde{\mathbf{I}}_{K_z+1}$, $N_1 = N$, and $\tilde{\mathbf{H}} = \mathbf{I}_N$ for the

z-dimension parameter estimation. In the case of the two constrained IQML methods in EXP 5.6 with the known parameter z^c , $\tilde{\mathbf{H}} = \tilde{\mathbf{I}}_K \tilde{\mathbf{J}}_K^T \tilde{\mathbf{I}}_K \mathbf{B}_c$, where $\mathbf{B}_c$ is as given by Eq. (5.41) with $K_1 = 1$, $b_0^c = -1$, and $b_1^c = z^c$. In all experiments, the maximum number of iterations allowed for the algorithms is 50. The initial condition $\mathbf{F}_1$ is as given by Eq. (5.26) and, except for EXP 5.7-5.8, the threshold number ε is 0.5E-10. Note that, to allow for fair comparison, the estimate $\hat{\mathbf{b}}_i$ at the i th iteration, obtained by any of the three algorithms, was first normalised before being used to calculate $\|\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i-1}\|$. The performance of a parameter estimator is measured by the Root-Mean-Square Error (RMSE) of a parameter estimate (in degrees or radians). The RMSE_{av} is obtained as the Root-Mean-Square (RMS) value of the RMSEs due to all K parameter estimates. The RMS value of the CRLBs on the standard deviations of the K parameter estimates, denoted by CRLB_{av} , will be used as the theoretical bound.

In the following two experiments, EXP 5.1 and EXP 5.2, the problem of DOA estimation by a uniform linear array was considered. The IQML algorithms were performed with both the DML and WSF cost functions. The latter was applied with the optimal weight $\hat{\mathbf{W}}_{\text{wsf}} = \hat{\Lambda}^2 \hat{\Lambda}_s^{-1}$.

EXP 5.1: In this experiment, the uniform linear array of $N = 10$ sensors, with inter-sensor spacing of half the signal wavelength, was used. Three far-field narrow-band point sources came from the directions of 2.5, 0, and -30 degrees with respect to broadside. The waveforms of these signals were uncorrelated, complex Gaussian distributed, each with unity power. The complex additive white Gaussian noise was added at each sensor to obtain various values of $\text{SNR} \triangleq -10 \log_{10} \sigma^2$, where σ^2 is the noise power at each sensor. For each SNR value, five hundred simulations were carried out with $L = 100$ snapshots of the data vector being generated in each simulation.

From Figure 5.2, it is clearly seen that, based on the DML cost function, the performance of both Bresler's and Nagesha's IQML algorithms (BIQML and NIQML) is inferior to that of the proposed Consistent IQML algorithm (CIQML). The RMSE_{av} of

the estimates using the CIQML algorithm attains the corresponding CRLB_{av} for $\text{SNR} > 10$ dB while the RMSE_{av} of the estimates obtained by the RMNORM method never attains the bound. The RMNORM method is the root version of the minimum-norm procedure with single forward/backward averaging. The average bound CRLB_{av} shown in the figure is based on the large sample approximate formula derived in [9]. It should also be noted that the best result gained by the NIQML algorithm at $\text{SNR} < 0$ dB is immaterial. At this range of SNR, the estimation error of any of the methods performed here is too large for the result to be useful. In addition, the results shown in Figure 5.2, which reflect the performances of the three IQML algorithms and the RMNORM method, were based on the DOA estimates from all of five hundreds simulations, without excluding the simulations where the DOAs were not resolved. By dividing the array FOV (± 90 degrees) into three angular regions at 1.25 and -15 degrees, the DOAs are said to have been resolved by a DOA estimator if each of their estimates is located within each of the three angular regions. Nevertheless, for $\text{SNR} \geq 5$ dB, the results based only on the simulations where the DOAs were resolved are essentially the same as the ones shown in Figure 5.2. At $\text{SNR} \geq 11.25$ dB, the DOAs were resolved in all simulations by any of the methods performed here.

With the WSF cost function, however, the performance of the BIQML and NIQML algorithms, seen from Figure 5.3, was substantially improved. The NIQML algorithm, in particular, achieved the best result. Its performance is comparable to that of the proposed CIQML algorithm with the same cost function. The performance improvement of the BIQML and NIQML algorithms is due to the reduction of the noise covariance matrix term in the WSF data matrix $\hat{\mathbf{U}}_s \mathbf{W}_{\text{wsf}} \hat{\mathbf{U}}_s^H$, which is obtained by eigendecomposition of the sample data covariance matrix, as detailed in Chapter 2. Note that, the performance improvement of the BIQML algorithm, when applied with the WSF cost function, was also observed in [23]. However, the reason for the improvement was not detailed in [23]. From the same figure, the RMSE_{av} obtained by the CIQML

algorithm with the DML cost function is seen to be approximately the same as the RMSE_{av} of the estimates obtained by using the BIQML algorithm with the WSF cost function. In addition, for $\text{SNR} > 5$ dB, the CIQML with the DML cost function achieved performance comparable to that of the NIQML algorithm with the WSF cost function. For the DML cost function, eigendecomposition of the data covariance matrix is not required. Thus, for a large array where the decomposition becomes costly, the use of the CIQML algorithm with the DML cost function may be preferred in terms of performance-complexity.

In Figure 5.4, the average numbers of iterations for the three IQML algorithms using the DML and WSF cost functions, are shown with respect to SNR. It is seen from the figure that, among the methods based on the DML cost function, the proposed algorithm has the fastest convergence rate. When the WSF cost function was applied, however, the convergence rates of the BIQML and NIQML algorithms were greatly improved. With the WSF cost function, some improvement in the convergence rate of the proposed method can also be observed. At SNR above 5 dB, the average numbers of iterations of the CIQML algorithm (with any of the two cost functions), and the BIQML and NIQML algorithms with the WSF cost function, are between 5 and 10. Figure 5.5 shows the percentages of convergence of the three algorithms using the DML and WSF cost functions. An algorithm is said to have converged if the condition $\|\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i-1}\| < \epsilon$ is satisfied for the i th iteration number which is below 50. From the figure, down to $\text{SNR} = 2.5$ dB, the CIQML algorithm with the DML cost function converged in all of five hundred simulations while the percentages of convergence of the BIQML and NIQML with the DML cost function dropped sharply at SNR of 5 dB. When the WSF cost function was applied, however, all three algorithms had approximately the same percentage of convergence.

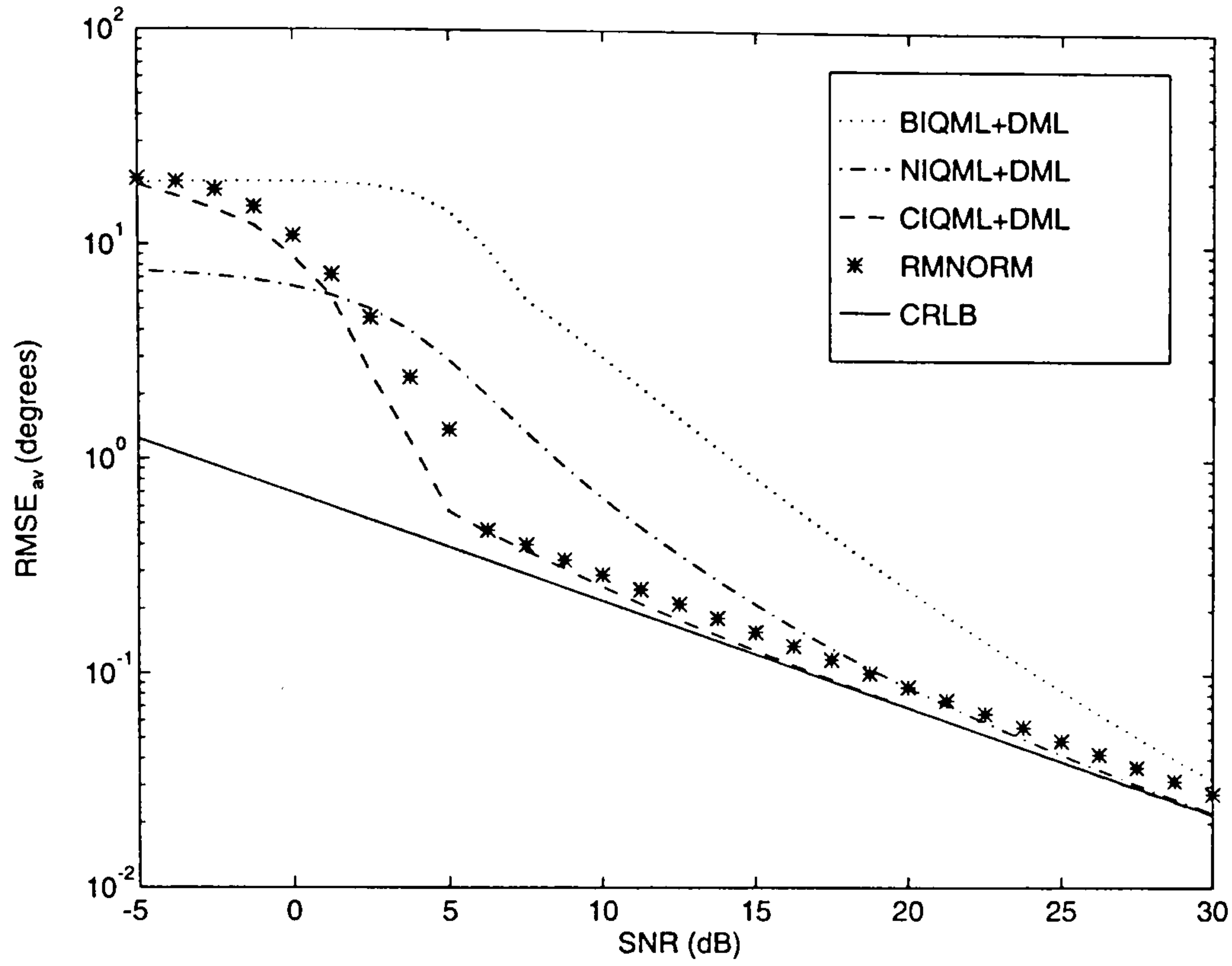


Figure 5.2: The estimated RMSE_{av} s of the DOA estimates in EXP 5.1 using three IQML algorithms with the DML cost function, and the RMNORM method, compared with the average CRLB on the std at various SNRs with a 1.25 dB step size.

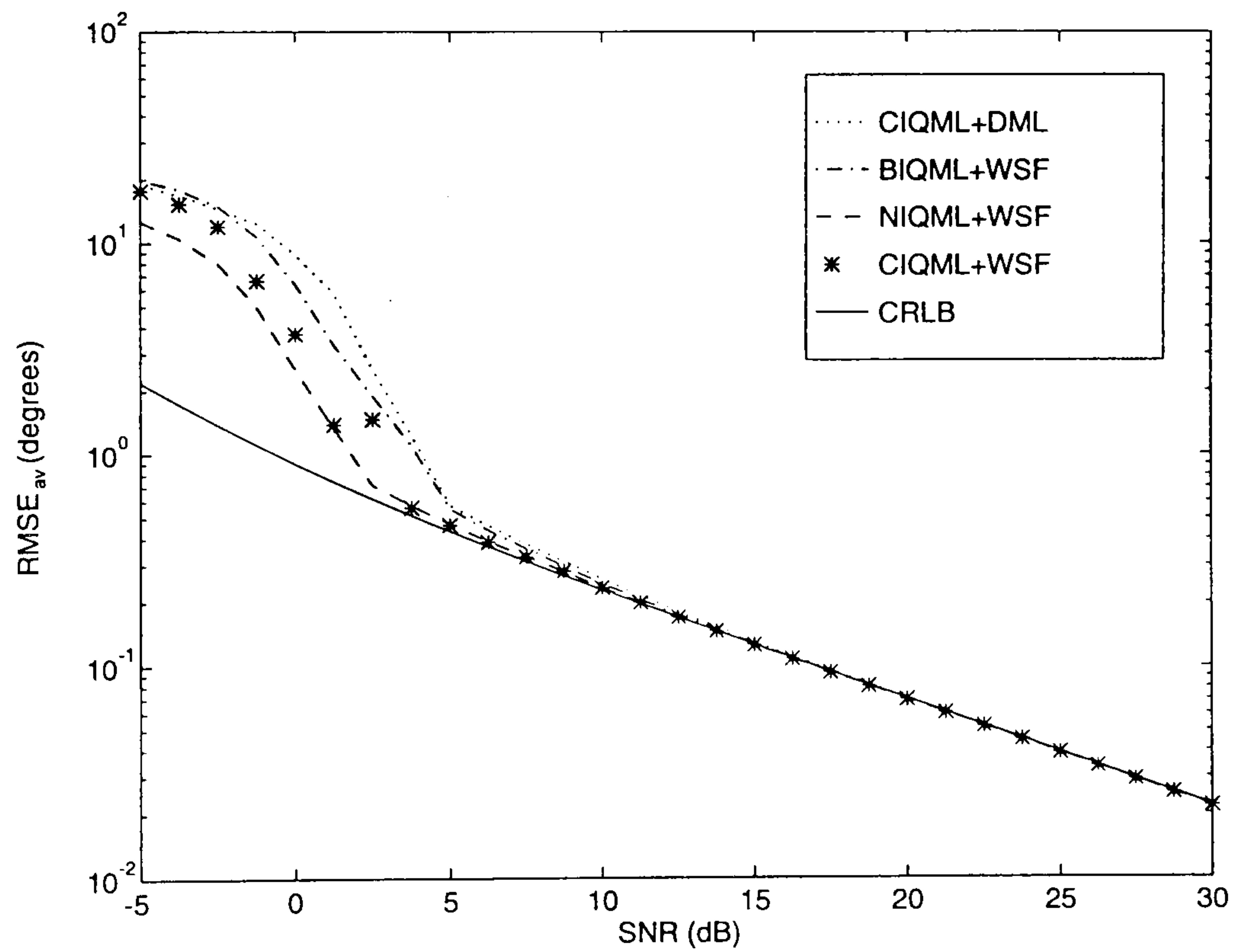


Figure 5.3: The estimated RMSE_{av} s of the DOA estimates in EXP 5.1 using three IQML algorithms with the DML and WSF cost functions, and the RMNORM method, compared with the average CRLB on the std at various SNRs with a 1.25 dB step size.

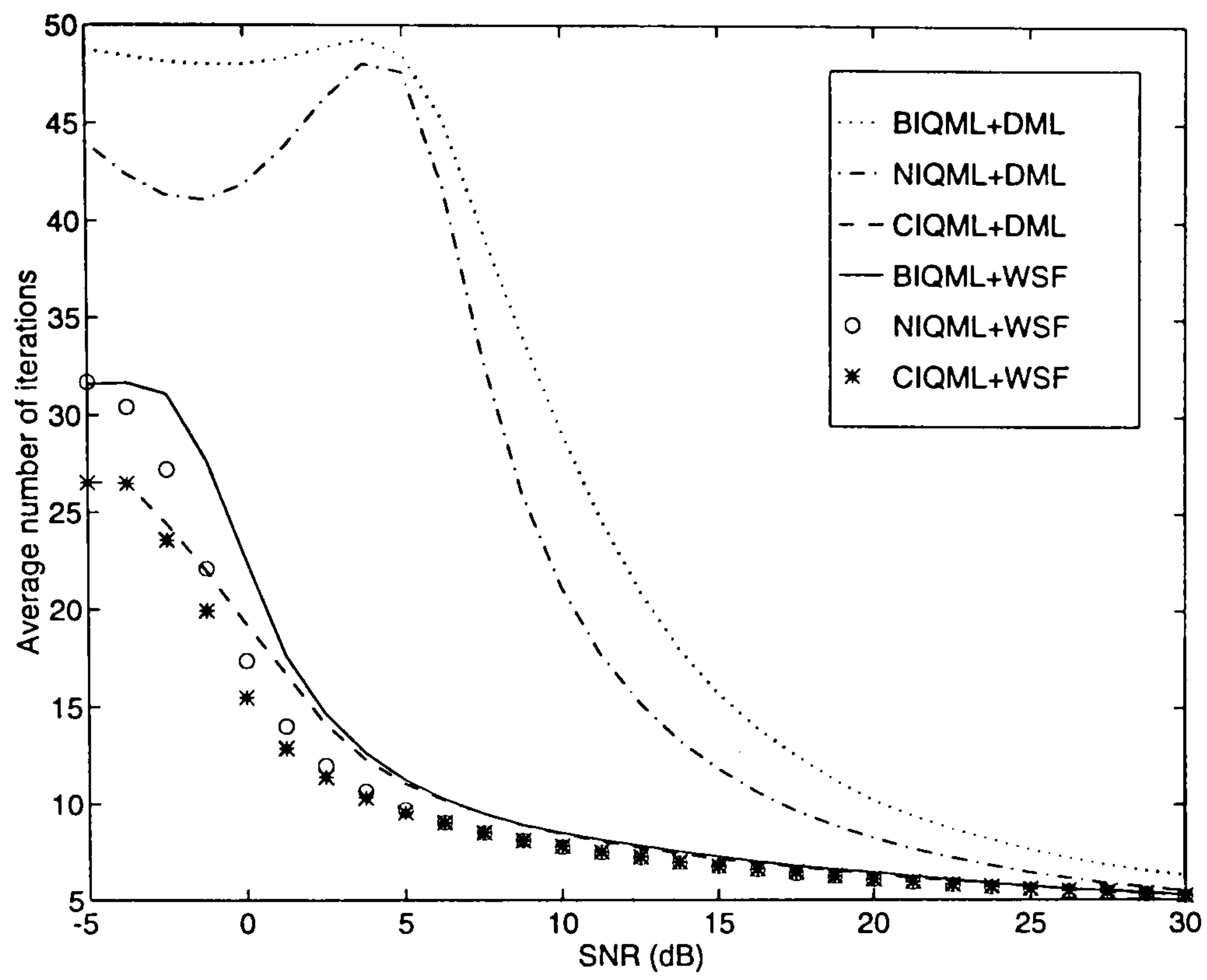


Figure 5.4: The average number of iterations of three IQML algorithms with the DML and WSF cost functions, at various SNRs with a step size of 1.25 dB.

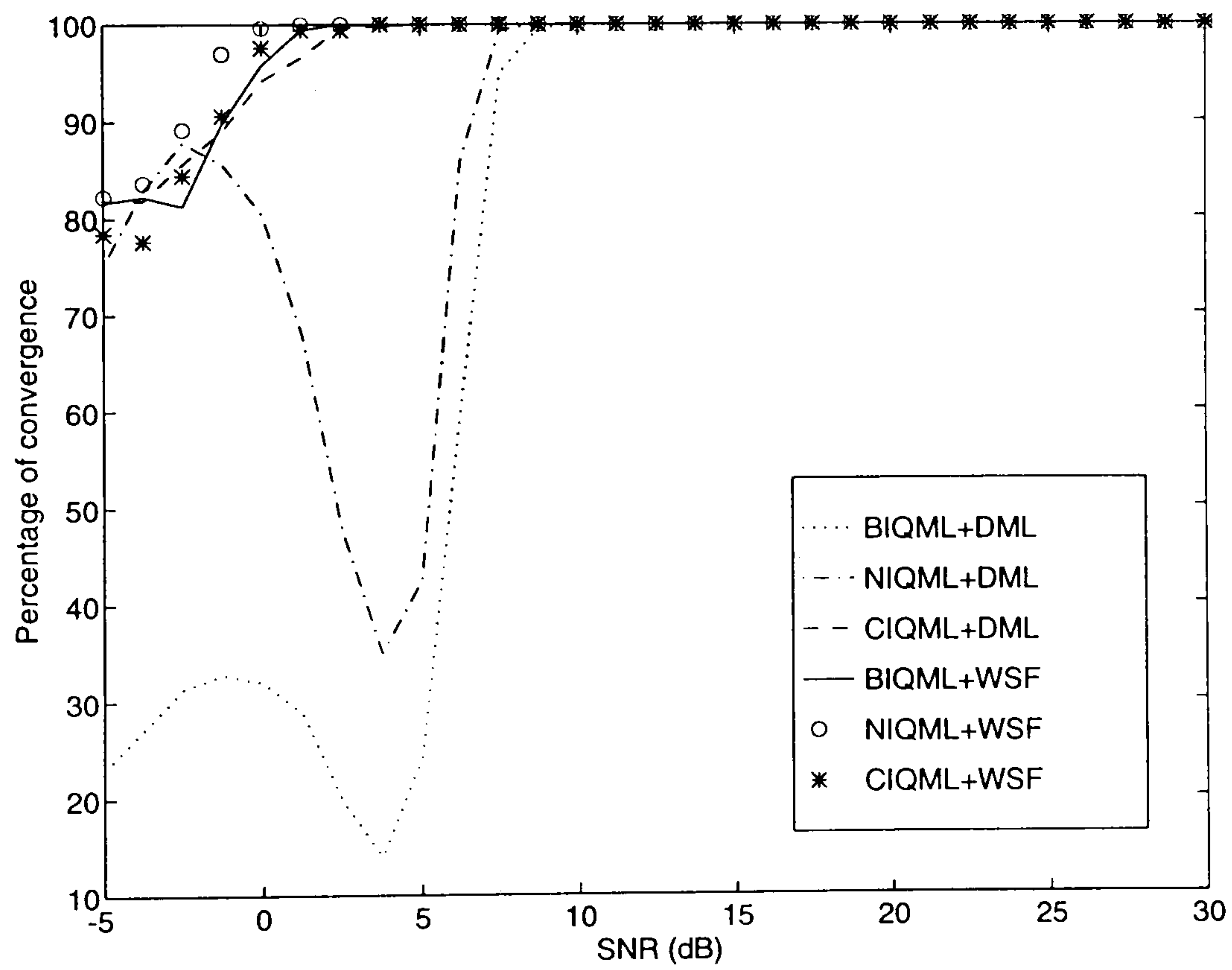


Figure 5.5: The percentages of convergence of three IQML algorithms with the DML and WSF cost functions, at various SNRs with a step size of 1.25 dB.

EXP 5.2: The three uncorrelated signals in EXP 5.1 were replaced by the correlated ones with the cross correlation between the i th and k th signals as given by $E\{s_i(t)s_k^*(t)\} = 0.99\{\exp(-j\pi(\alpha_i - \alpha_k))\}$. The signals $s_i(t)$, $i = 1, \dots, K = 3$, were measured at the middle point between the 5th and 6th sensors. Through study of the corresponding CRLB, the signal phases α_i were chosen as $\alpha_1 = \pi/8$, $\alpha_2 = \pi/4$, and $\alpha_3 = 3\pi/8$, to represent an average performance case.

From Figure 5.6, with the DML cost function, the inefficiency of the BIQML and NIQML algorithms is again seen. With the same cost function, the proposed CIQML algorithm yielded a superior performance to the other two algorithms. The SNR threshold of the proposed algorithm is also slightly higher than that of the RMNORM method. The RMNORM method employed forward/backward averaging of three overlapping subarrays of eight sensors each. As for the case of Figure 5.2 in EXP 5.1, the results shown in Figure 5.6 were based on all of five hundreds simulations. However, for $\text{SNR} \geq 13.75$ dB, the results based only on the simulations where the three DOAs were resolved are largely the same as the ones shown in Figure 5.6. At $\text{SNR} \geq 18.75$ dB, the three DOAs were resolved in all simulations by any of the methods performed here.

From Figure 5.7, with the WSF cost function being employed, the accuracy of the DOA estimates by all three IQML algorithms was improved. Among them, the CIQML and NIQML algorithms gave the best result. The SNR threshold of the CIQML and NIQML algorithms is approximately 7 dB lower than that of the RMNORM method. Note that, the CRLB_{av} shown in Figure 5.6 was based on the deterministic signal assumption while the one shown in Figure 5.7 was based on the stochastic signal assumption. The latter was obtained using the large sample approximate CRLB formula derived in [76]. In this case, the two bounds are nearly identical, with the bound corresponding to the deterministic signal assumption having slightly smaller value at low SNR.

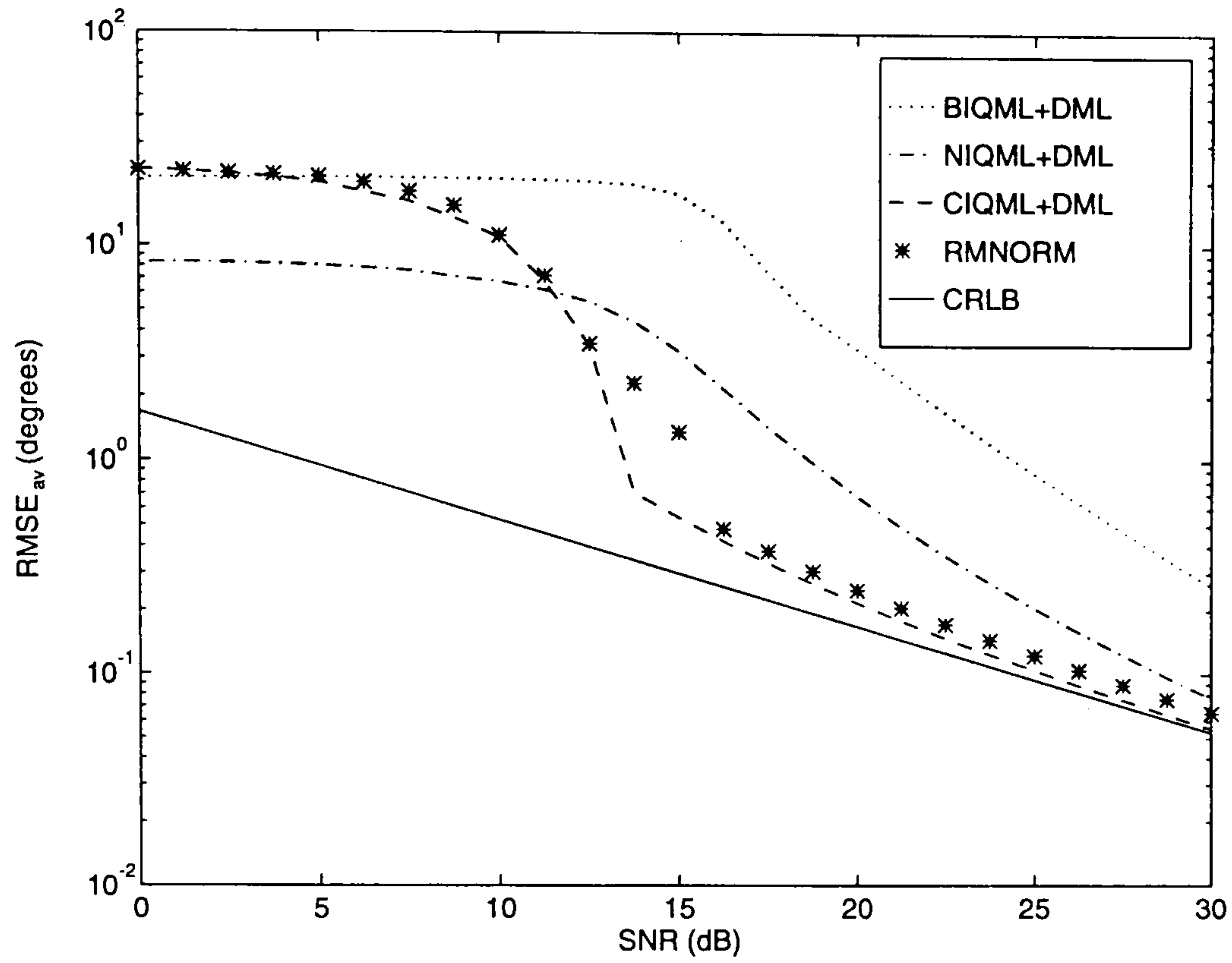


Figure 5.6: From EXP 5.2, the estimated RMSE_{av} s of the correlated source DOA estimates using three IQML algorithms with the DML cost function and the RMNORM method, compared with the average CRLB at various SNRs with a 1.25 dB step size.

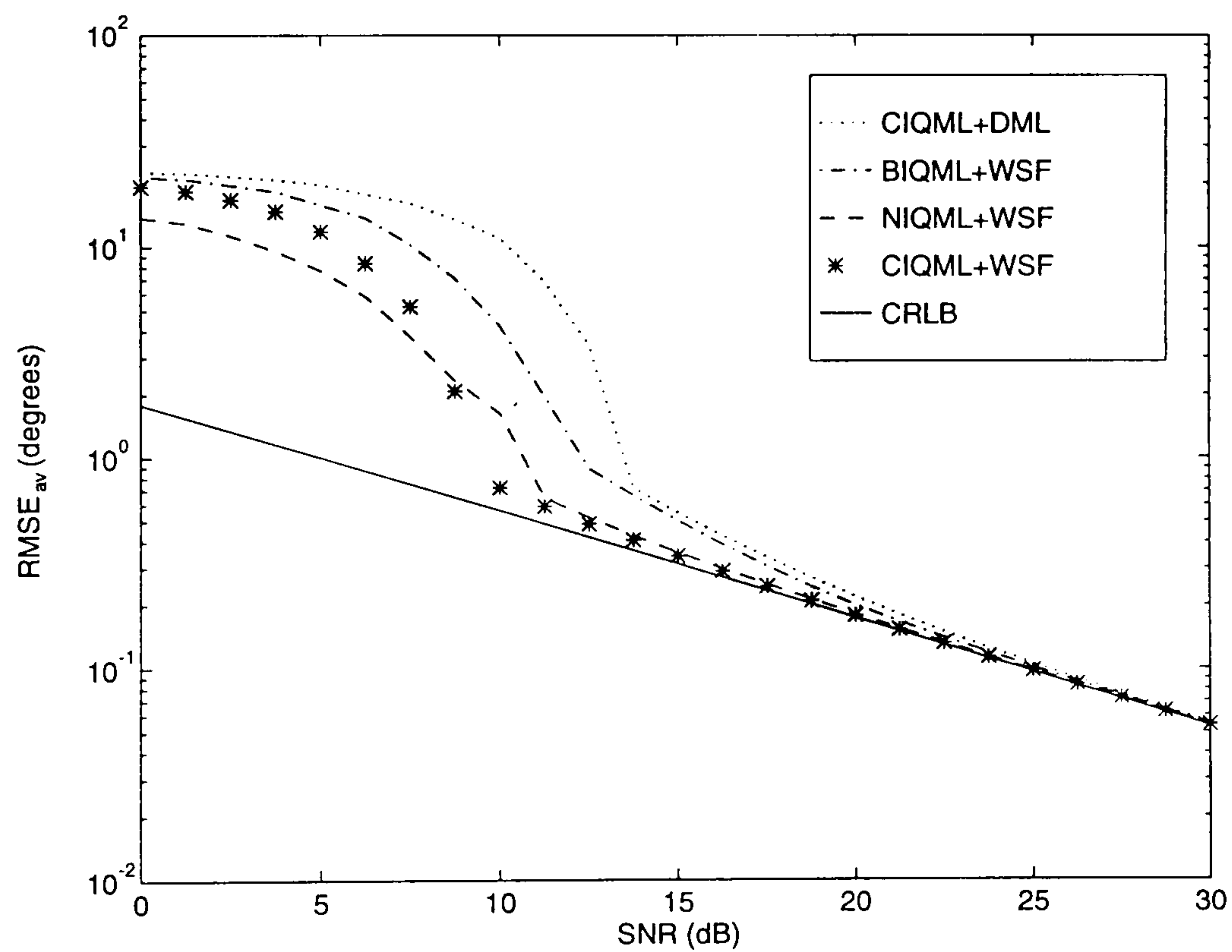


Figure 5.7: From EXP 5.2, the RMSE_{av} s of the correlated source DOA estimates using three IQML algorithms with the WSF cost function and the RMNORM method, compared with the average CRLB at various SNRs with a 1.25 dB step size.

In the next four experiments, the single data snapshot case was considered. The problem considered in these experiments is of multiple undamped sinusoidal frequency estimation. Nevertheless, due to mathematical similarity, it also represents the DOA estimation problem with a single array snapshot.

EXP 5.3: This example was taken from [24]. A single snapshot of $N = 25$ data points was generated from

$$x_k = \exp(j\alpha_1).\exp(j2\pi\theta_1(k-1)) + \exp(j\alpha_2).\exp(j2\pi\theta_2(k-1)) + n_k, \\ k = 1, \dots, 25 \quad (5.68)$$

where $\theta_1 = 0.35$, $\theta_2 = 0.37$, $\alpha_1 = 0$, and $\alpha_2 = \pi/4$.

The result shown in Figure 5.8 suggests that both the NIQML and CIQML algorithms yield slightly improved performance over that of the BIQML algorithm. At high SNR, both method gave slightly better result than that of the MFBLP (modified forward-backward linear prediction) method [32] (see also Chapter 2), which was performed with the prediction order of 18. The average number of iterations required by each of the three IQML algorithms is shown in Figure 5.9.

EXP 5.4: In this experiment, the effect of the signal phases on the estimation performance was investigated. The two sinusoidal signal model of EXP 5.3 was again considered. However, the phase reference point was moved from the first data point, assumed in the last example, to the middle of the data set. With this, the observed data can be described by

$$x_k = \exp(j\alpha_1).\exp(j2\pi\theta_1(k-c_o)) + \exp(j\alpha_2)\exp(j2\pi\theta_2(k-c_o)) + n_k, \\ k = 1, 2, \dots, N \quad (5.69)$$

where $\theta_1 = 0.35$, $\theta_2 = 0.37$, and $c_o = (N + 1)/2$. Note that by choosing the phase reference point as described, the effect of the signal phases on the estimation performance is decoupled with that due to the signal frequency values (see [77], for

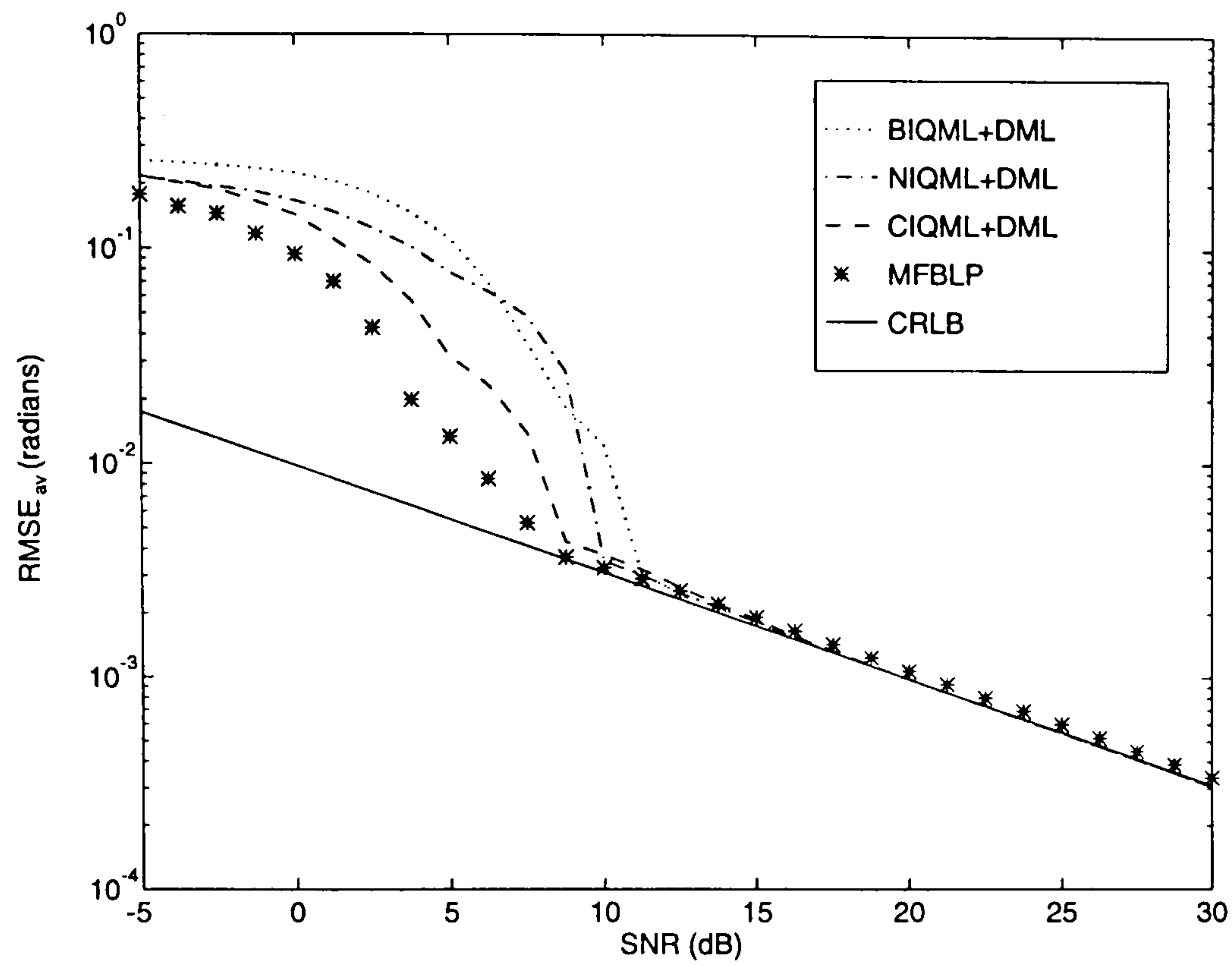


Figure 5.8: From EXP 5.3, the estimated RMSE_{av} s of the two sinusoidal frequency estimates using three IQML algorithms and MFBLP method, compared with the average CRLB on the std at various SNRs with a 1.25 dB step size.

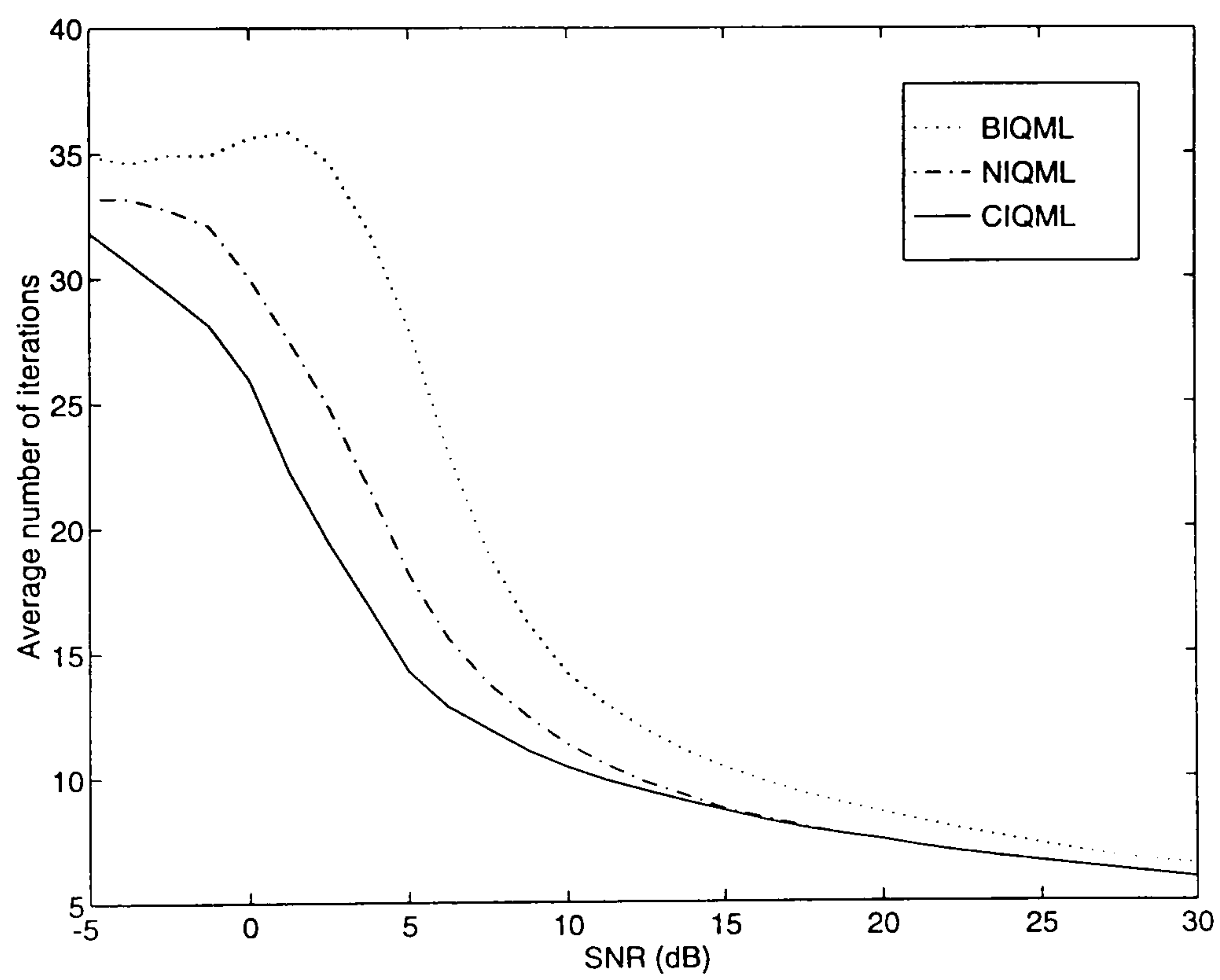
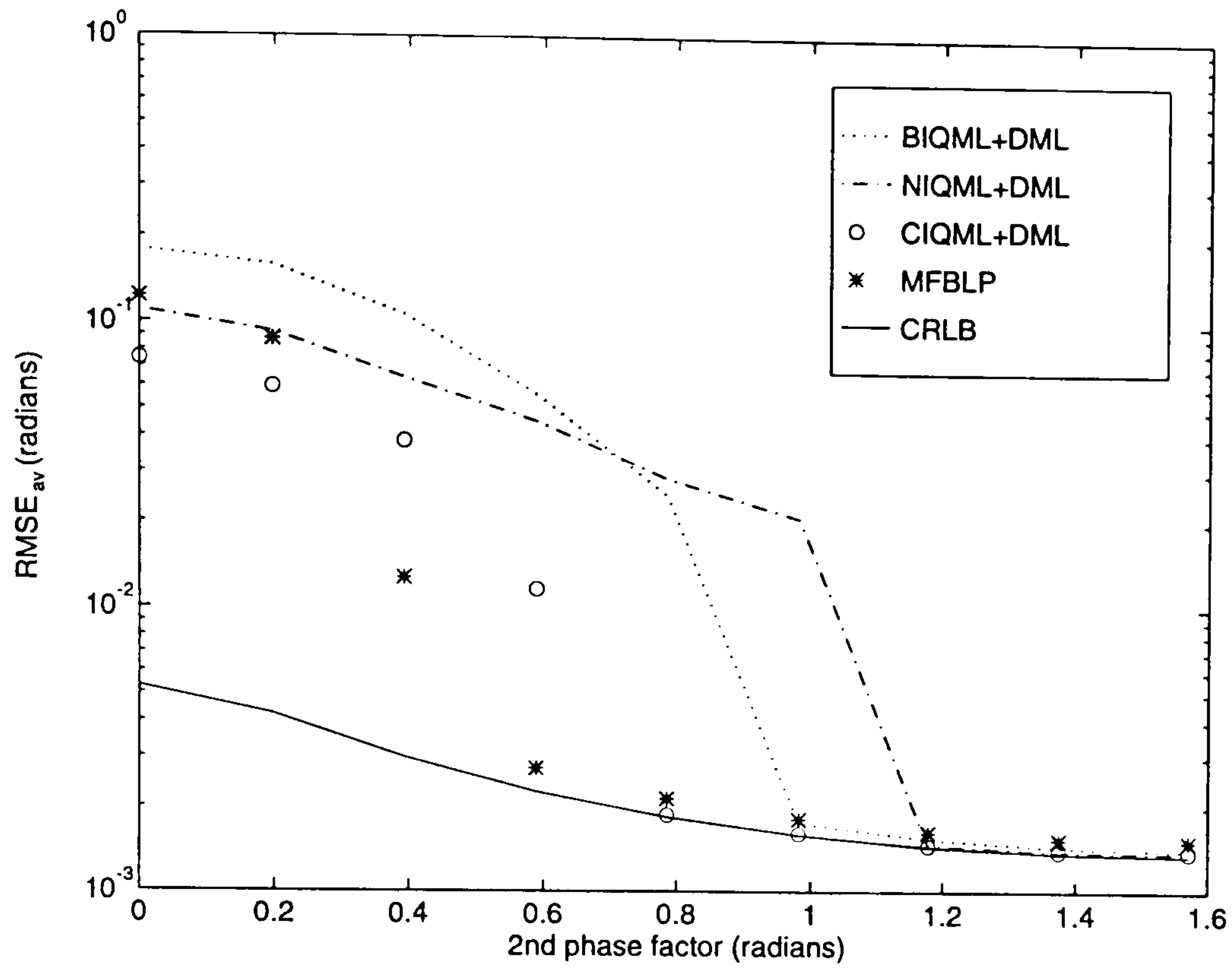


Figure 5.9: From EXP 5.3, the average number of iterations of three IQML algorithms, at various SNRs with a step size of 1.25 dB.

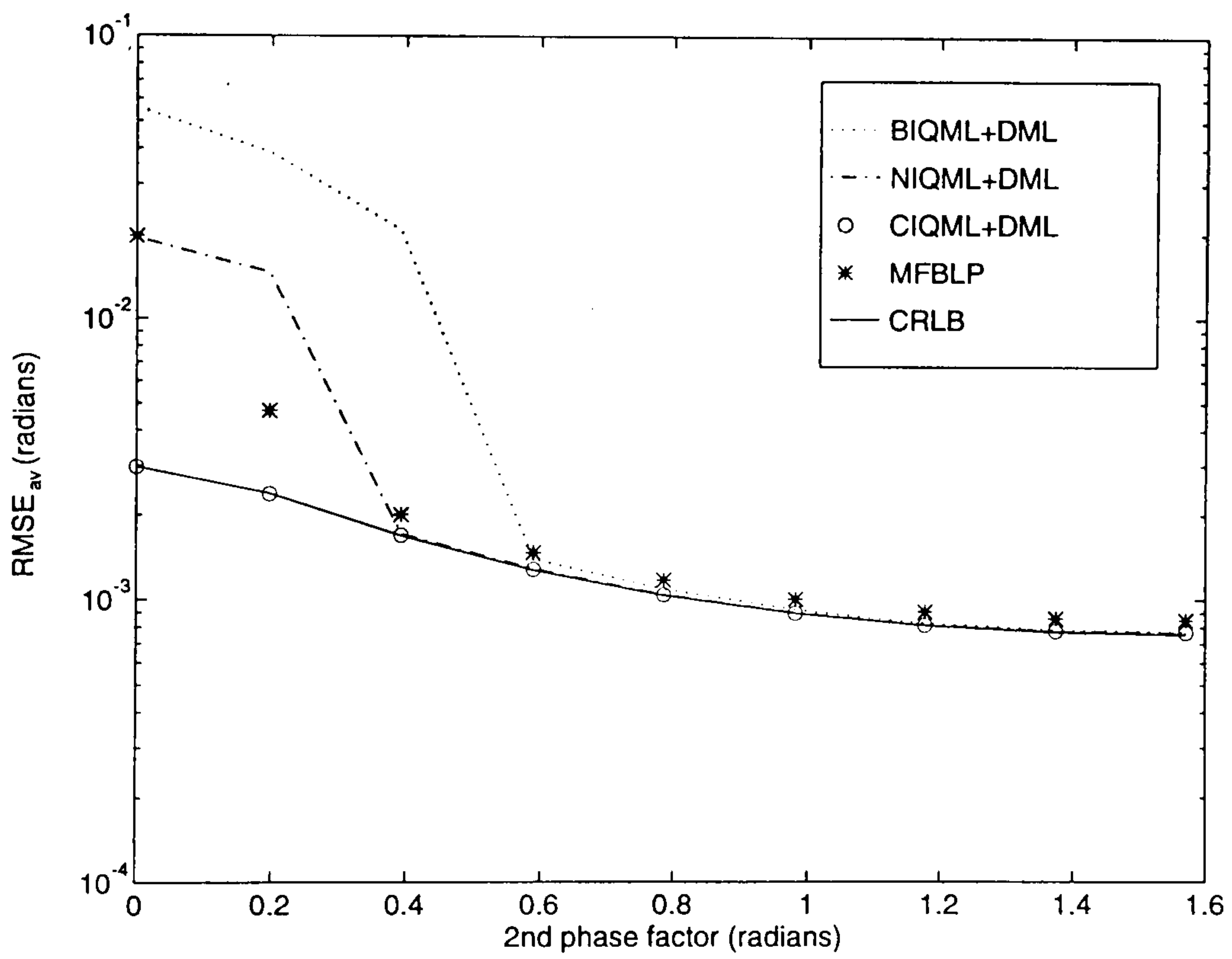
more detail on this issue). In this experiment, the data length was $N = 25$. Estimation of the two signal frequencies was carried out using the DML cost function with the three algorithms and the MFBLP method with the prediction order of 18, all at SNRs of 15 and 20 dB. At each SNR value, the phase factor of the first signal was fixed at $\alpha_1 = 0$, while the second phase factor α_2 was varied to obtain 9 different phase values as follows:

$$\alpha_2 = m\pi / 16, \quad m = 0, 1, \dots, 8 \quad (5.70)$$

The estimated RMSE_{av} s of the frequency estimates, obtained by the three IQML algorithms at SNRs of 15 and 20 dB are shown in Figure 5.10. From Figure 5.10a, the MFBLP method gave the best result as the second phase factor α_2 approached $\alpha_1 (= 0$ radian). However, for $\alpha_2 > 0.8$ radians, the proposed CIQML algorithm yielded the estimates whose RMSE_{av} attained the CRLB_{av} , while the RMSE_{av} due to the MFBLP method did not. It is also seen that the performance of the BIQML and NIQML algorithms is inferior to that of the CIQML algorithm. At $\text{SNR} = 20$ dB, the RMSE_{av} of the estimates using the CIQML algorithm attained the CRLB_{av} for all of the nine different phase factors. The performance of the BIQML and NIQML algorithms was greatly deteriorated as α_2 approached zero. Note that, while the RMSE_{av} of the ML estimates using either of the three IQML algorithms attained the bound CRLB_{av} as α_2 approached $\pi/2$, the RMSE_{av} of the MFBLP estimates never attained the bound.



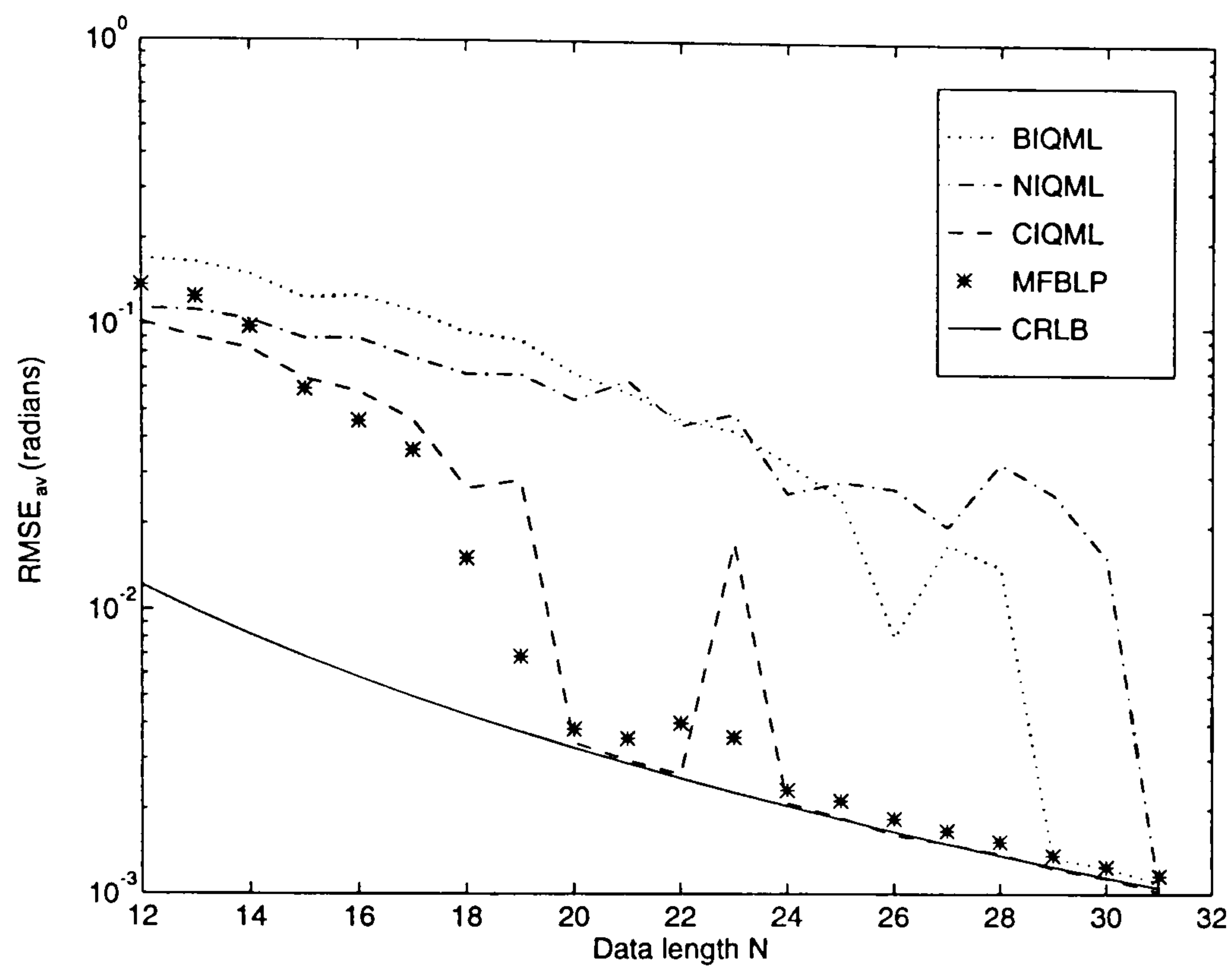
(a)



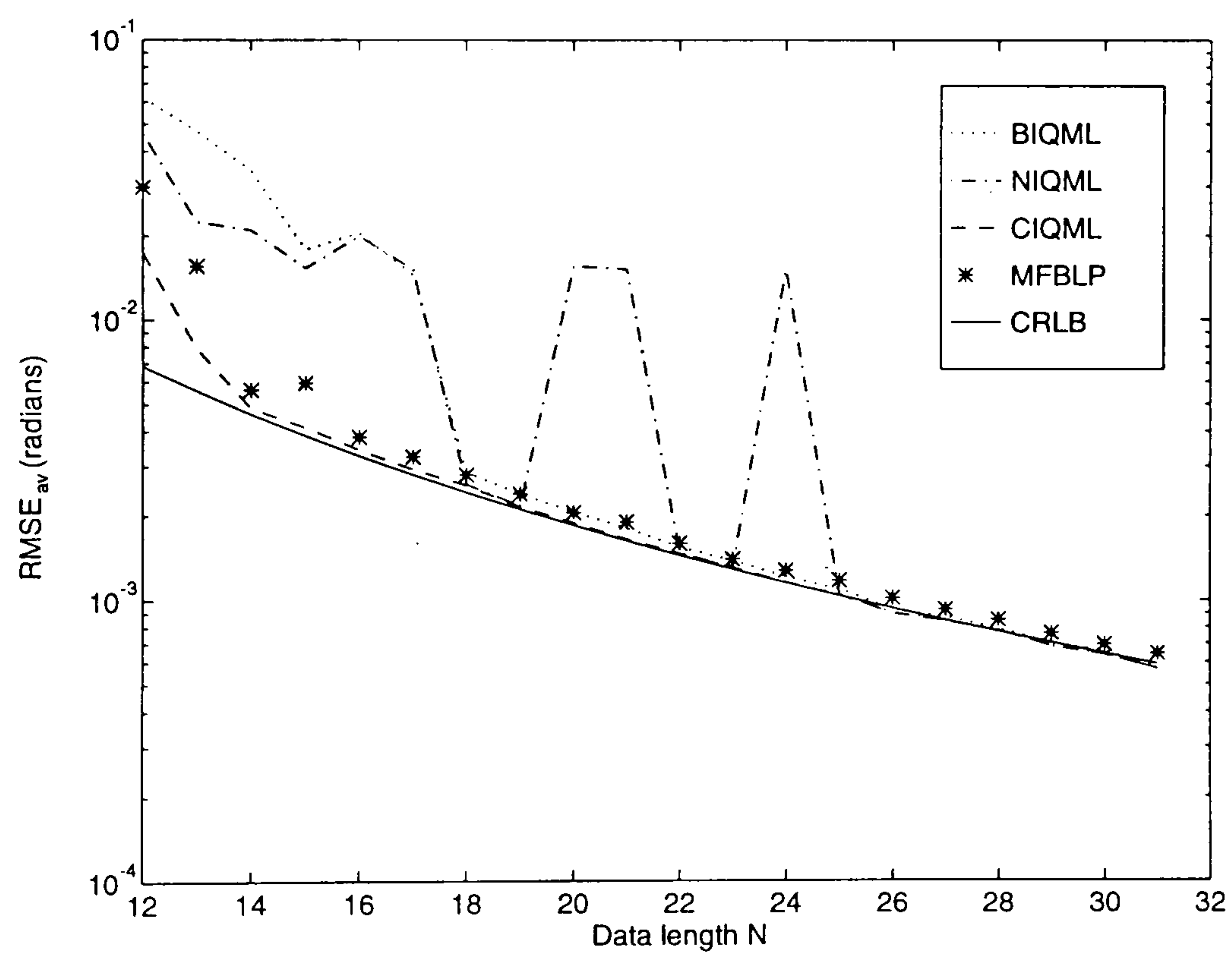
(b)

Figure 5.10: From EXP 5.4, the estimated RMSE_{av} s of two sinusoidal frequency estimates using three IQML algorithms and MFBLP method, at nine different values of the second phase factor α_2 . (a) SNR = 15 dB; (b) SNR = 20 dB.

EXP 5.5: The frequency model Eq. (5.69) was again applied. The data length N was, however, varied from 12 to 31 with a single step size. The experiment was repeated with two different SNRs of 15 and 20 dB. The two phase factors were chosen as $\alpha_1 = 0$, and $\alpha_2 = \pi/4$. The estimated RMSE_{av} s of the frequency estimates using the DML estimator with the three IQML algorithms and the MFBLP method at the two SNRs are shown in Figure 5.11. For the MFBLP method, the prediction order at each value of N was chosen as the smaller of the two closest integer numbers to $3N/4$. This prediction order selection was recommended in [32]. From the figure, it is seen that the CIQML algorithm offered better performance at a short data length N , than those of the BIQML and NIQML algorithms. In particular, at a SNR of 15 dB, the latter two algorithms performed badly for most of the chosen 20 values of N . In comparing between the MFBLP method and the DML method with the CIQML algorithm, the former yielded only slightly better results at $N < 20$. For $N > 20$, however, the DML method based on the CIQML algorithm yielded the estimates of which the RMSE_{av} attained the CRLB_{av} , while the RMSE_{av} obtained by the MFBLP method never reached the bound. Note that, at $N = 23$, the performance of the DML method using the CIQML algorithm sharply dropped. This is because, at one of five hundred simulations, the two frequency parameters could not be resolved and appeared as a single parameter. As a result, a spurious frequency estimate was picked as a real frequency estimate, giving the unusually large error. The same problem was also observed in the NIQML algorithm at the SNR of 20 dB (see Figure 5.11b). This problem may be remedied by using additional information of the signals. For instance, if the power of each signal, $E\{|s_k|^2\}$, is known to be well above some threshold level $\xi > \sigma^2$, the criterion $\frac{1}{L} \sum_{t=0}^{L-1} |\hat{s}_k(t)|^2 < \xi$, where $\hat{s}_k(t)$ is the estimate of $s_k(t)$, may be applied to detect a spurious frequency estimate.



(a)



(b)

Figure 5.11: From EXP 5.5, the estimated RMSE_{av} s of two sinusoidal frequency estimates using three IQML algorithms and the MFBLP method, compared with the average CRLB on the std for 20 values of the data length N . (a) $\text{SNR} = 15$ dB; (b) $\text{SNR} = 20$ dB.

EXP 5.6: With the data model Eq. (5.69), the third frequency $\theta_3 = 0.41$ was added. The frequency values of the other two signals were the same as in EXP 5.4. The chosen phase factors were $\alpha_1 = 0$, $\alpha_2 = \pi/8$, and $\alpha_3 = \pi/4$. From Figure 5.12, the proposed CIQML algorithm achieved the best performance. It also has the lowest SNR threshold (approximately 5 dB lower than that of the NIQML algorithm).

With the same scenario, two constrained IQML algorithms using Nagesha's and the proposed nontriviality constraints (constrained NIQML, and constrained CIQML), were performed using the constraint Eq. (5.40) under the assumption that θ_3 was known a priori. In addition, with θ_3 being assumed known, the constrained RMNORM method, which is a variant of the Constrained MUSIC method in [42], was performed with forward/backward averaging of eight 18×18 subdata covariance matrices. In Figure 5.13 the performances of the constrained methods are compared with those of the unconstrained ones. The estimated RMSE_{av} of the frequency estimates using each of the unconstrained methods shown in Figure 5.13 is the RMS value of the RMSEs of the first two parameter estimates $\hat{\theta}_1$ and $\hat{\theta}_2$. From the figure, the two constrained IQML algorithms greatly improved the performance of their unconstrained counterparts. Among all three constrained methods, the constrained CIQML method again yielded the best result. The SNR threshold of the method is more than 3 dB lower than that of the constrained NIQML method. The CRLB_{av} shown in the figure was based on Eq. (5.42), with θ_3 as the known parameter. The average number of iterations for each of five IQML methods is shown in Figure 5.14. From the figure, the proposed algorithm required a smaller number of iterations than those needed by the NIQML algorithm. In addition, the constrained algorithms needed fewer iterations than the unconstrained counterparts.

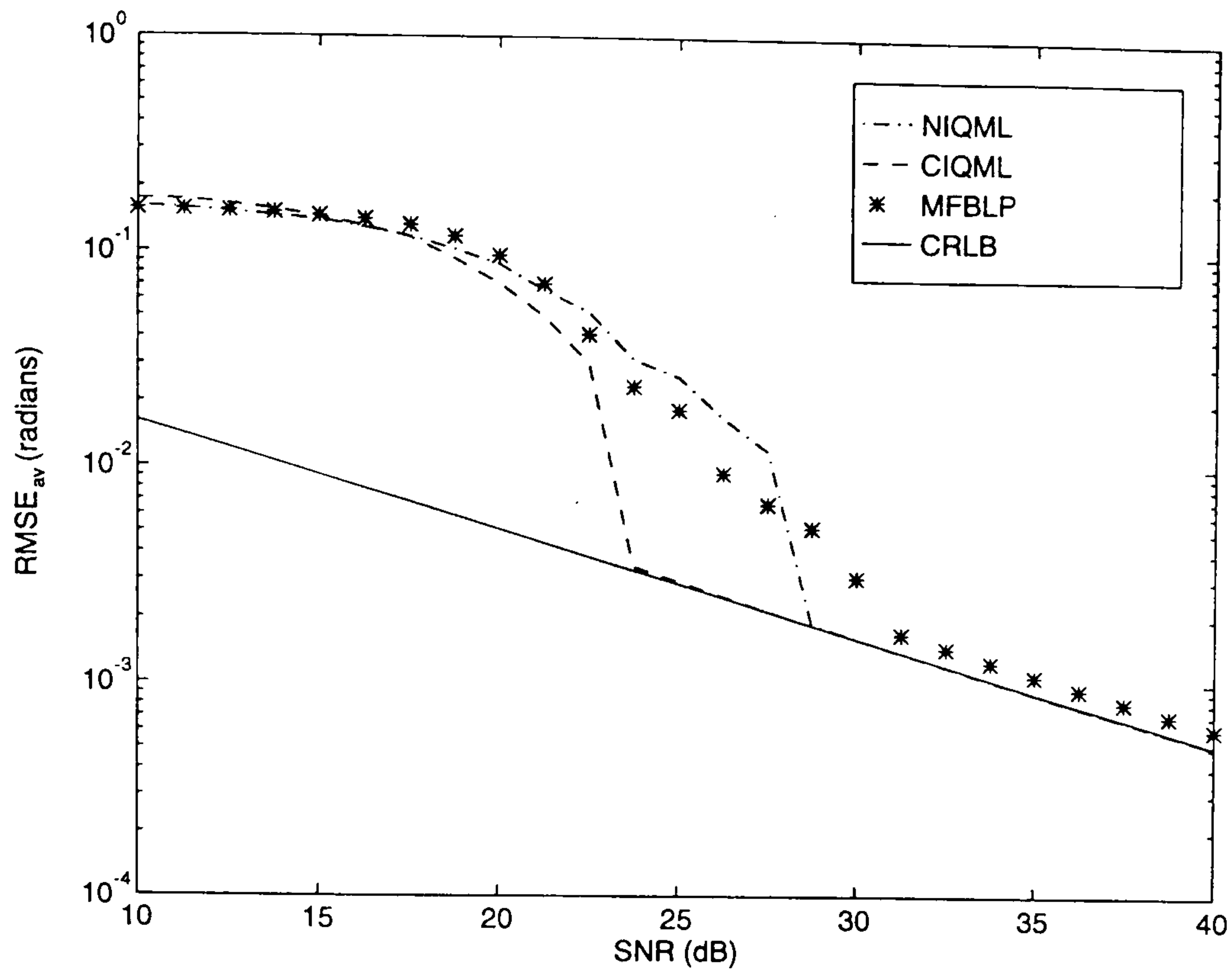


Figure 5.12: From EXP 5.6, the estimated RMSE_{av} s of three frequency estimates using two unconstrained IQML algorithms and the MFBLP method, compared with the average CRLB on the std at $N = 25$ and various SNRs with a step size of 1.25 dB.

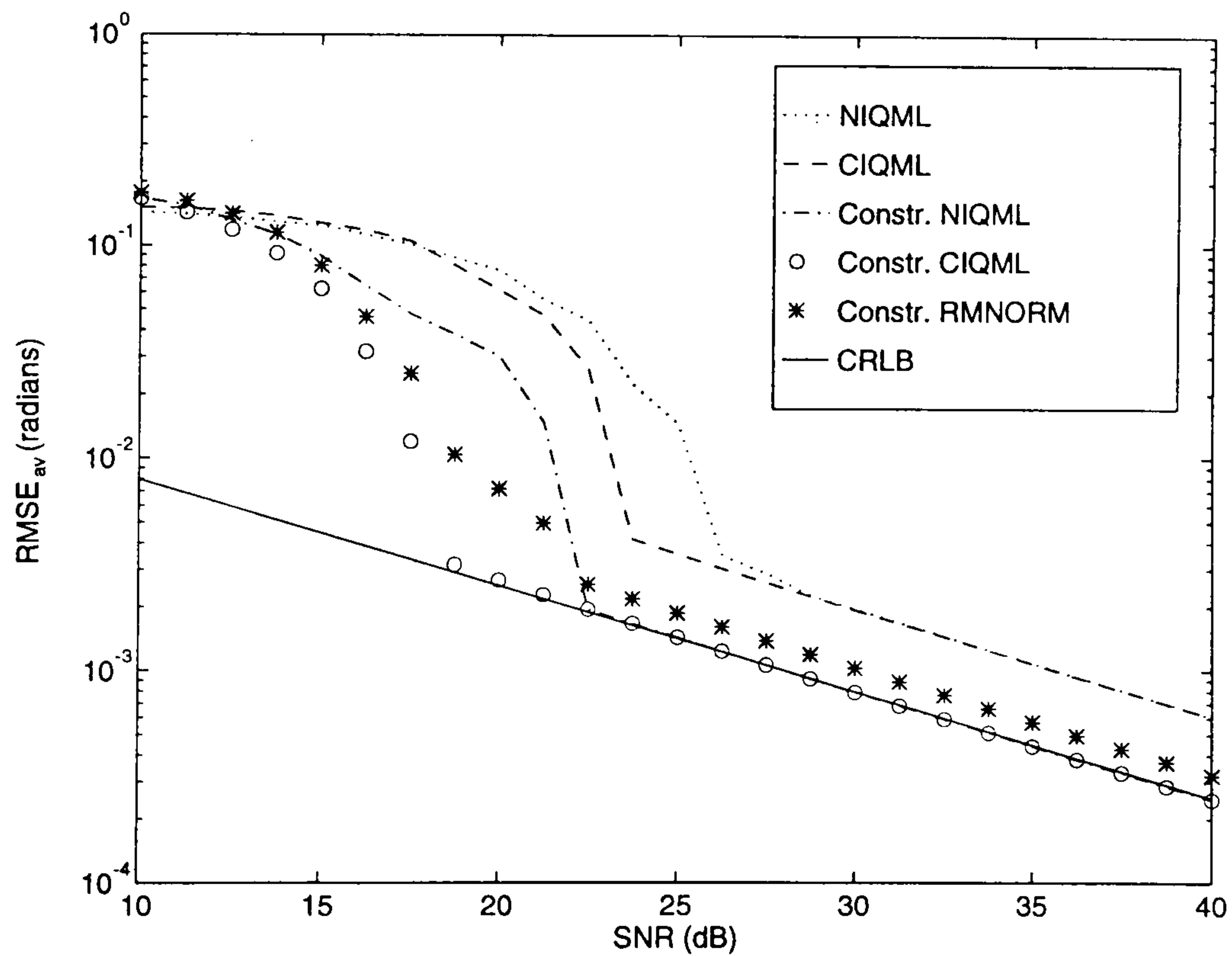


Figure 5.13: From EXP 5.6, the estimated RMSE_{av} s of two frequency estimates using four unconstrained and constrained IQML algorithms, and the constrained RMNORM method, at $N = 25$ and various SNRs with a step size of 1.25 dB.

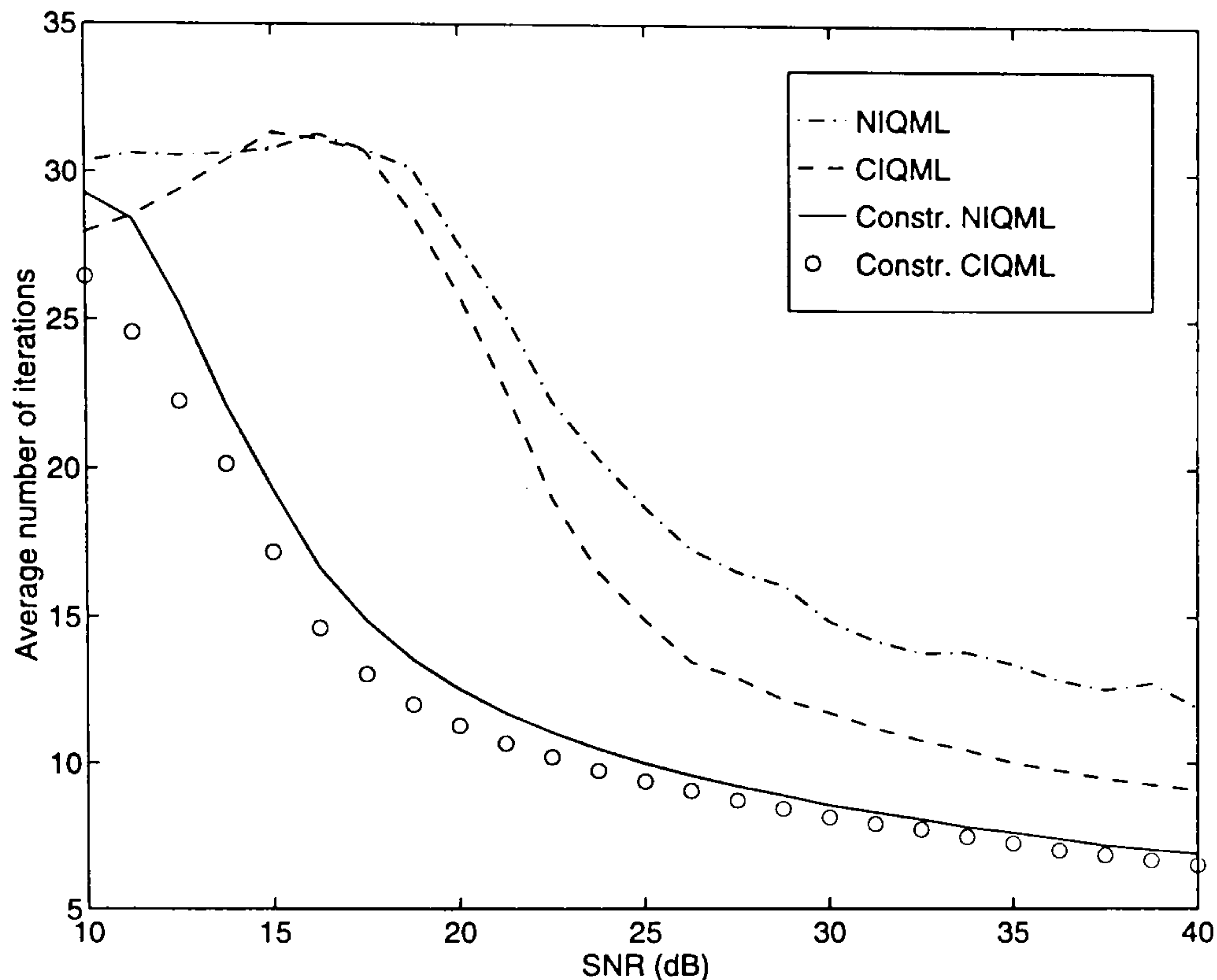


Figure 5.14: The average numbers of iterations required by two unconstrained and two constrained IQML algorithms.

In the last two experiments, we consider the 2-D frequency estimation problem. The Hua's MEMP method, the proposed IQML-based and RMNORM-based methods were performed. The RMNORM-based method is based on the proposed 1-D IQML-based method, with the root version of the minimum norm procedure being used in place of the IQML algorithm in estimating the y- and z-dimension parameters. In each of the following two examples, a single snapshot of a uniform rectangular data $\mathbf{X}$ of size 10×10 was generated. The structure of $\mathbf{X}$ is as shown in Eq. (5.3). From $\mathbf{X}$, two sets of 10 data snapshots, $\mathbf{x}_y(n)$ and $\mathbf{x}_z(m)$, were constructed by Eqs. (5.52-5.53). The proposed IQML-based method was applied to $\mathbf{x}_y(n)$, $n = 1, \dots, 10$, to estimate the y-dimension parameters, and to $\mathbf{x}_z(m)$, $m = 1, \dots, 10$, to estimate the z-dimension parameters. The proposed version of the algorithm (CIQML) was employed with the threshold number $\varepsilon = 0.5\text{E-}06$. Pairing of the two parameter components was carried out using the forward/backward average matrix constructed by Eqs. (5.54-5.55), and

$$\mathbf{x}_{k,l} = \begin{bmatrix} x_{k,l} & x_{k,l+1} & \cdots & x_{k,l+N_0-1} & x_{k+1,l} & \cdots & \cdots & x_{k+M_0,l+N_0-1} \end{bmatrix}^T, \\ k = 1, \dots, M - M_0 + 1, l = 1, \dots, N - N_0 + 1 \quad (5.71)$$

Two pairing procedures were performed in the proposed IQML-based method. They were carried out using two average pairing matrices $\hat{\mathbf{R}}_{\text{av};1}^{\text{fb}}$ and $\hat{\mathbf{R}}_{\text{av};2}^{\text{fb}}$, based on the subdata vectors Eq. (5.71) with $M_0 = 3$, $N_0 = 2$ for $\hat{\mathbf{R}}_{\text{av};1}^{\text{fb}}$, and $M_0 = 3$, $N_0 = 3$ for $\hat{\mathbf{R}}_{\text{av};2}^{\text{fb}}$. Only the pairing based on $\hat{\mathbf{R}}_{\text{av};2}^{\text{fb}}$ was performed for the RMNORM-based method. For the MEMP method, thirty-six 5×5 subdata matrices of the form described by Eq. (5.46) were constructed. Based on these subdata matrices, the 5×5 $\hat{\mathbf{R}}_{\text{av}}^{\text{fb}}$ was constructed by Eqs. (5.54-5.55). The resulting $\hat{\mathbf{R}}_{\text{av}}^{\text{fb}}$ was used to obtain the y- and z-dimension frequency estimates by the method described in [64]. The 2-D parameter pairing was carried out by Step 2 of the proposed 2-D procedure by using the described 5×5 average matrix $\hat{\mathbf{R}}_{\text{av}}^{\text{fb}}$.

EXP 5.7: In this example, a single snapshot of the data matrix $\mathbf{X}$ of size 10×10 was generated according to the following model:

$$x_{k,l} = \sum_{i=1}^3 \exp(j\alpha_i) \cdot \exp(j2\pi\theta_{y;i}(k - c_0)) \cdot \exp(j2\pi\theta_{z;i}(l - c_0)) + n_{k,l}, \\ k = 1, \dots, M, l = 1, \dots, N \quad (5.72)$$

where $M = N = 10$, $\alpha_1 = \pi/8$, $\alpha_2 = \pi/4$, $\alpha_3 = 3\pi/8$, and $c_0 = 11/2$. The 2-D parameter pairs $(\theta_{y;1}, \theta_{z;1}) = (0.40, 0.35)$, $(\theta_{y;2}, \theta_{z;2}) = (0.35, 0.35)$, and $(\theta_{y;3}, \theta_{z;3}) = (0.35, 0.40)$. In each of five different SNRs, three hundred computer simulations were run. The RMNORM-based method was performed with single forward/backward averaging of the sample covariance matrix, formed by $\mathbf{x}_y(n)$ for the estimation of the y-dimension parameters, and by $\mathbf{x}_z(m)$ for the estimation of the z-dimension parameters. The previously defined RMSE_{av} of the frequency estimates in each of the two dimensions, for the three methods, is shown in Table 5.1. It is seen that the proposed 2-D method with the 3×3 pairing matrix $\hat{\mathbf{R}}_{\text{av};2}^{\text{fb}}$ yielded the best result. The two IQML-

based methods, as well as the RMNORM-based method, performed much better than the MEMP method. It should be noted, however, that in this case the MEMP method and the proposed IQML-based and RMNORM-based methods are based on different prior information. While the MEMP method requires only the knowledge of the number of 2-D parameter pairs K , the proposed methods require also knowledge of K_y and K_z . Two theoretical bounds are included for the 2-D parameter estimates. Each of the two bounds was obtained as the RMS value of the 2-D CRLB on the standard deviations of the three parameter estimates in each dimension. For the $\text{CRLB}_{\text{av};1}$, the bound was based on Eq. (5.66), which assumes that only the number of 2-D parameter pairs K is known. The second bound, $\text{CRLB}_{\text{av};2}$, was based on Eq. (5.61), by assuming that K_y , K_z , and K are known. Note from the table that the second bound is smaller than the first one, a result of additional prior information of K_y and K_z . In addition, the RMSE_{av} obtained by the proposed 2-D method, with $\hat{\mathbf{R}}_{\text{av};1}^{\text{fb}}$ or $\hat{\mathbf{R}}_{\text{av};2}^{\text{fb}}$, shows a surprising result of attaining the bound $\text{CRLB}_{\text{av};2}$ at high SNR despite being based on 1-D parameter estimation. Figures 5.15-5.16 show the plots of the three 2-D frequency estimates over 300 simulations at $\text{SNR} = 20$ dB, obtained by, respectively, the MEMP method and the proposed IQML-based method with the pairing matrix $\hat{\mathbf{R}}_{\text{av};1}^{\text{fb}}$. The faulty pairing was observed in the proposed method. This type of error disappeared when either $\text{SNR} > 20$ dB or the pairing matrix $\hat{\mathbf{R}}_{\text{av};2}^{\text{fb}}$ was applied instead.

Method	SNR (dB)				
	10	15	20	25	30
IQML-1 [†] $\times 10^{-2}$	0.41	0.62	0.30	0.06	0.03
IQML-2 [‡] $\times 10^{-2}$	0.38	0.20	0.11	0.06	0.03
RMNORM $\times 10^{-2}$	0.46	0.25	0.14	0.08	0.04
MEMP	0.10	0.94×10^{-2}	0.40×10^{-2}	0.22×10^{-2}	0.12×10^{-2}
CRLB _{av;1} $\times 10^{-2}$	0.59	0.33	0.19	0.10	0.06
CRLB _{av;2} $\times 10^{-2}$	0.30	0.17	0.10	0.05	0.03

(a)

Method	SNR (dB)				
	10	15	20	25	30
IQML-1 $\times 10^{-2}$	0.47	0.62	0.31	0.06	0.03
IQML-2 $\times 10^{-2}$	0.36	0.19	0.10	0.06	0.03
RMNORM $\times 10^{-2}$	0.44	0.24	0.13	0.07	0.04
MEMP	0.10	0.87×10^{-2}	0.39×10^{-2}	0.22×10^{-2}	0.12×10^{-2}
CRLB _{av;1} $\times 10^{-2}$	0.59	0.33	0.19	0.10	0.06
CRLB _{av;2} $\times 10^{-2}$	0.30	0.17	0.10	0.05	0.03

(b)

[†] Proposed method with the pairing matrix $\hat{\mathbf{R}}_{av;1}^{fb}$. [‡] With the pairing matrix $\hat{\mathbf{R}}_{av;2}^{fb}$.

Table-5.1: The RMSE_{av}s of the frequency estimates by four estimation methods, together with the CRLB_{av;1} and CRLB_{av;2}. (a) The y-dimension estimates; (b) The z-dimension estimates.

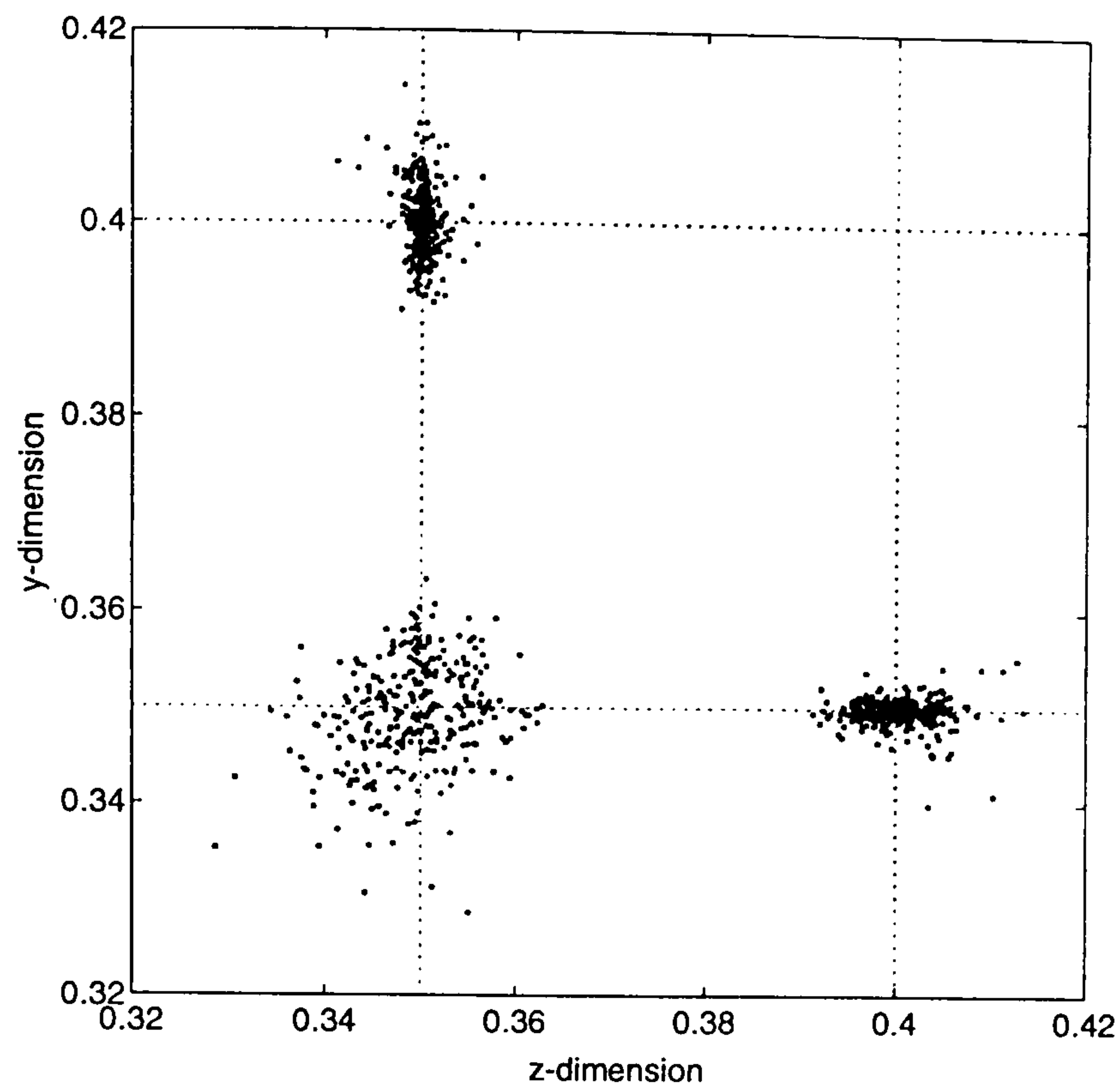


Figure 5.15: Estimates of three 2-D parameters by the MEMP method, based on the subdata matrix of size 5×5 at SNR of 20 dB.

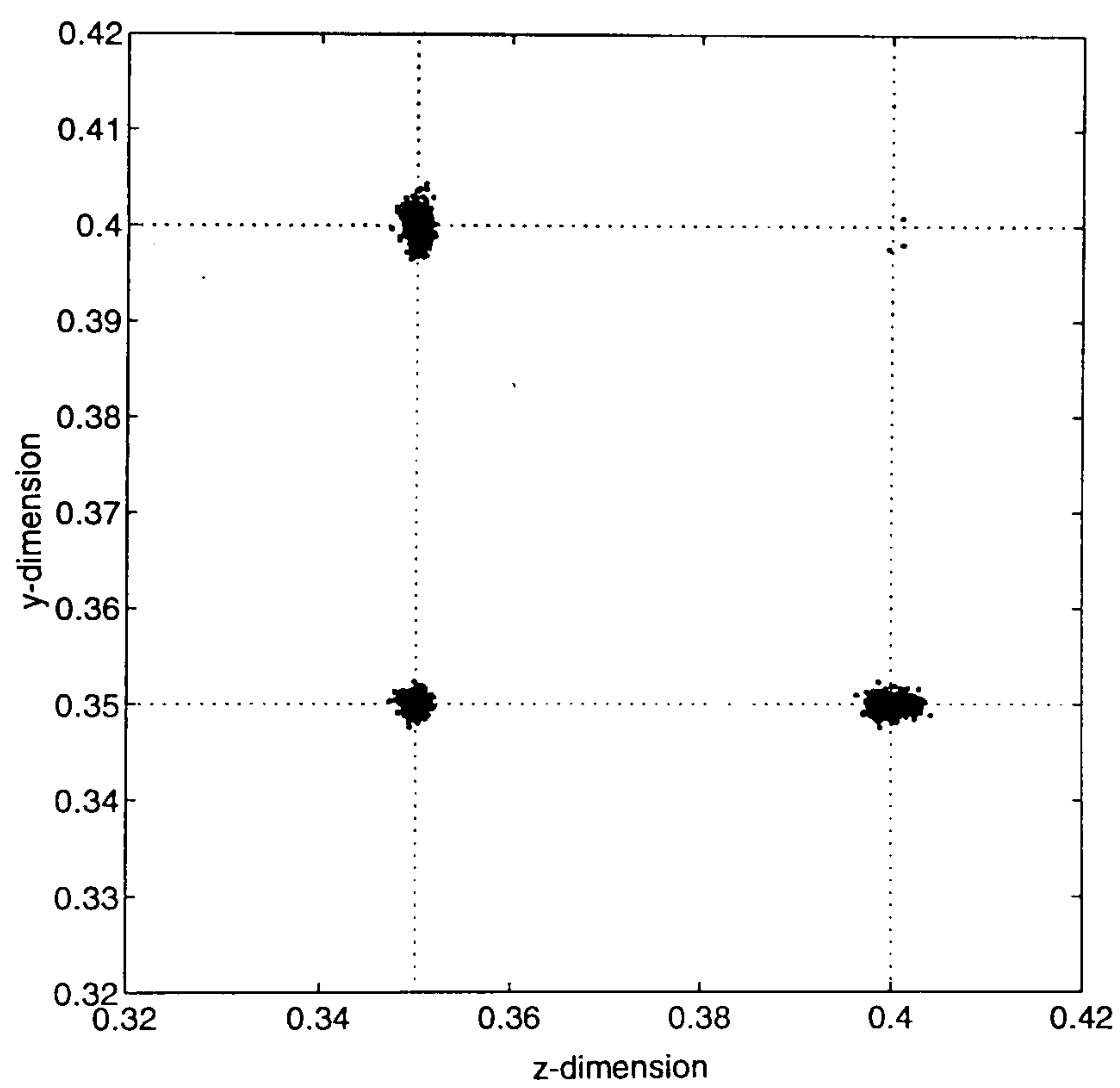


Figure 5.16: Estimates of three 2-D parameters by the proposed method with the pairing matrix $\hat{\mathbf{R}}_{av;1}^{fb}$ at SNR of 20 dB.

EXP 5.8: The 2-D data were generated by the same signal model Eq. (5.72) as in EXP 5.7. Other parameters were the same as in the former experiment. However, the three 2-D frequency pairs were chosen as $(\theta_{y;1}, \theta_{z;1}) = (0.45, 0.35)$, $(\theta_{y;2}, \theta_{z;2}) = (0.40, 0.40)$, and $(\theta_{y;3}, \theta_{z;3}) = (0.35, 0.45)$. The RMNORM-based method was performed with forward/backward averaging of two sample covariance matrices, formed by snapshots of two 9×1 subdata vectors, overlapped by one data point, of $\mathbf{x}_y(n)$ for the y-dimension parameter estimation, and of $\mathbf{x}_z(m)$ for the z-dimension parameter estimation.

From Table 5.2, it is seen that the IQML-based method with two different pairing matrices performed slightly better than the MEMP and RMNORM-based methods at SNR below 20 dB. No faulty pairing was observed for the IQML-based method when either $\hat{\mathbf{R}}_{av;1}^{fb}$ or $\hat{\mathbf{R}}_{av;2}^{fb}$ was used for the purpose. At higher SNR (≥ 20 dB), however the MEMP method gave slightly better results than the IQML-based and RMNORM-based methods. We also note that, none of the methods performed here yielded the RMSE_{av} which attained the average CRLB (on the standard deviation), even at high SNR. For the case of the IQML-based and RMNORM-based methods, this is because the methods are not exact 2-D parameter estimators. For the MEMP method, the frequency estimation is performed on subdata matrices with reduced dimensionality. The estimator is thus sub-optimal.

Method	SNR (dB)				
	10	15	20	25	30
IQML-1	0.11	0.02	0.49×10^{-2}	0.26×10^{-2}	0.14×10^{-2}
IQML-2	0.11	0.02	0.49×10^{-2}	0.26×10^{-2}	0.14×10^{-2}
RMNORM	0.19	0.08	0.48×10^{-2}	0.26×10^{-2}	0.14×10^{-2}
MEMP	0.15	0.04	0.44×10^{-2}	0.23×10^{-2}	0.12×10^{-2}
$\text{CRLB}_{\text{av};1} \times 10^{-2}$	0.69	0.39	0.22	0.12	0.07

(a)

Method	SNR (dB)				
	10	15	20	25	30
IQML-1	0.11	0.02	0.44×10^{-2}	0.23×10^{-2}	0.13×10^{-2}
IQML-2	0.11	0.01	0.44×10^{-2}	0.23×10^{-2}	0.13×10^{-2}
RMNORM	0.19	0.08	0.48×10^{-2}	0.26×10^{-2}	0.14×10^{-2}
MEMP	0.15	0.04	0.44×10^{-2}	0.23×10^{-2}	0.12×10^{-2}
$\text{CRLB}_{\text{av};1} \times 10^{-2}$	0.69	0.39	0.22	0.12	0.07

(b)

Table-5.2: The RMSE_{av} s of the frequency estimates using four estimation methods, together with the $\text{CRLB}_{\text{av};1}$ (on the standard deviation). (a) The y-dimension estimates; (b) The z-dimension estimates.

5.6 Conclusions

In this chapter, we first presented the generalised form of the IQML algorithm. With this form, additional constraints as well as some pre-processing are allowed to be included into the algorithm. In addition, the constrained IQML algorithm has been formulated, based on the knowledge of some parameter values. The compact CRLB formula under this type of constraint has also been provided. The improved IQML algorithm, employing the new consistent nontriviality constraint, has been proposed. Its performance has been shown to be superior to those of Bresler's and Nagesha's IQML algorithms. For the multiple data snapshot case, significant improvement of the proposed algorithm over the existing ones can be obtained when the DML cost function is employed. When the optimal WSF cost function is applied, the performance of the

Bresler's and Nagesha's IQML algorithms are improved. The reason for this performance improvement has been addressed.

The use of the WSF cost function in the correlated source scenario has been shown to yield superior performance to that of the DML cost function for all three IQML algorithms. In the uncorrelated source case with sufficiently high SNR, however, the performance of the proposed algorithm, when applied to the DML cost function, is comparable to that achieved by any of the three IQML algorithms using the WSF cost function.

In addition to the multiple data snapshot case, the single data snapshot case has also been considered. It has been shown that the performance of the proposed algorithm is superior to those of Bresler's and Nagesha's algorithms, especially in difficult situations such as short data length N , and certain range of the signal phase factors. Moreover, the constrained IQML algorithm has been shown to offer improved performance over the unconstrained one.

Based on the proposed algorithm, the IQML-based 2-D parameter estimation method has also been proposed. The parameter pairing problem has been considered. The sufficient condition for consistent parameter pairing has been provided. The proposed 2-D method has been shown to offer comparable, and sometimes superior performance to the MEMP method.

Chapter 6

Performance Study of the PFML DOA Estimators

In this chapter, the performance of the PFML DOA estimators is considered. The study concentrates upon the beamforming-based PFML methods, detailed in Chapter 4. Their performances are compared with those due to the DEML method, and the two PFML methods which represent the best-case PFML-2 and PFML-3 estimators. Detailed study of the IV-based DOA estimators, which are related to the PFML-3 method as shown in Section 3.2.3, can be found in [18-20]. The study of the DEML method for the special case of uncorrelated sources was reported in [16-17].

Results from computer simulation and analytical studies will be used to assess the performance of these PFML methods. In this chapter, a uniform linear array of eight sensors, with inter-sensor spacing of half of the signal wavelength, is assumed. In addition to the assumptions A. 2.1-2.6, listed in Chapter 2, it is assumed that both the signal vector $s(t)$ and the additive noise $n(t)$ are circular complex multivariate Gaussian distributed. In addition, $s(t)$ and $n(t)$ are assumed to be statistically independent. Only the problem of DOA estimation is considered. The number of the signal sources whose directions are of interest, is assumed known. Methods for estimating the number of sources when it is not known a priori, have been studied and documented, for instance, in [47, 78-81].

6.1 Simulation and analytical results

The analytical results reported here are based on the asymptotic analysis as provided in Chapters 3 and 4. For the computer simulation results, the proposed CIQML algorithm in Chapter 5 was applied to obtain the DOA estimates for all ML-based DOA estimators. The conjugate-symmetric constraint [22] was applied in the algorithm. The matrix $\tilde{\mathbf{H}}$ in Eq. (5.16) was chosen as $\mathbf{I}_N$, with $N_1 = N$. The maximum number of iterations allowed for the algorithm was 50. The threshold number ϵ was 0.5E-10. The initial condition matrix $\mathbf{F}_1$ was chosen as given by Eq. (5.26), with $\mathbf{R}_{\text{ns}}$ in Eq. (5.26) being replaced by $\hat{\mathbf{R}}_{\text{n}}$ in Eq. (3.27) for the PFML-3 estimator, and by $\hat{\mathbf{R}}_{\text{n},k}$ in Eq. (3.58) for the PFML-2 estimator. The performance of a DOA estimator is measured by the RMSE (root-mean-square error) of a DOA estimate (in degrees). In addition, sometimes the RMS value of the RMSEs of all source direction estimates, denoted by RMSE_{av} , will also be used. In all experiments, the signal vector $\mathbf{s}(t)$ was assumed to have the form described by

$$\mathbf{s}(t) = \mathbf{\Gamma}_{\text{cs}} \mathbf{s}_{\text{nc}}(t) \quad (6.1)$$

The signal covariance matrix was assumed to have the structure given by

$$\mathbf{R}_{\text{s}} = \mathbf{\Gamma}_{\text{cs}} \mathbf{R}_{\text{nc}} \mathbf{\Gamma}_{\text{cs}}^H \quad (6.2)$$

where $\mathbf{R}_{\text{nc}}$ is the covariance matrix of the $d_{\text{s}} \times 1$ vector $\mathbf{s}_{\text{nc}}(t)$. The vector $\mathbf{s}_{\text{nc}}(t)$ was assumed to be composed of d_{s} uncorrelated signals. Thus, $\mathbf{R}_{\text{nc}}$ is a diagonal matrix. The $K \times d_{\text{s}}$ correlated signal structure matrix $\mathbf{\Gamma}_{\text{cs}}$ takes the form

$$\mathbf{\Gamma}_{\text{cs}} = \begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,d_{\text{s}}} \\ \gamma_{2,1} & \gamma_{2,2} & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma_{K_{\text{c}},1} & \cdots & \cdots & \gamma_{K_{\text{c}},d_{\text{s}}} \end{bmatrix} \quad (6.3)$$

where $\gamma_{i,k}$, $i = 1, \dots, K_{\text{c}}$, $k = 1, \dots, d_{\text{s}}$, are the $K_{\text{cs}} \times 1$ complex vectors. This signal model can be used to describe a wide range of source scenarios. However, the form was

designed mainly to describe the mixed correlated-uncorrelated cluster source scenario, as mentioned in Chapter 4. As before, K is the total number of sources with distinct directions and K_c is the number of source clusters. In this study, the number of sources belonging to each cluster was restricted to be the same for all K_c clusters, and was equal to K_{cs} . Many DOA estimators were involved in the experiments. Thus, for ease of reference and presentation, some information which relates to the methods used is provided:

1. The DML method was performed with the total number of signal sources K being known a priori. Except for EXP 6.8, it was assumed that the noise covariance structure was known.
2. The WSF method was based on the optimal weight $\hat{\mathbf{W}}_{\text{wsf}} = \hat{\hat{\Lambda}}^2 \hat{\Lambda}_s^{-1}$ (see Chapter 2 for more detail). In addition to the total number of signal sources K , the rank d_s of the signal covariance matrix was known. Except for EXP 6.8, the noise covariance structure was also assumed known.
3. The DEML method is the estimator described in Section 3.2.1. The method was originally proposed in [16-17]. However, in the simulation study performed here, the method was applied under a more general signal covariance structure Eq. (6.2), instead of the one corresponding to the uncorrelated source scenario, considered in [16-17]. Except for one of the two DEML methods in EXP 6.6, the known pre-filtering reference signal was $s_{\text{nc}}(t)$. The noise covariance structure and the number of sources in each cluster were assumed known.
4. The PFML-2 method was performed with $s_{\text{nc}}(t)$ as the pre-filtering reference signal, which is $\mathbf{r}_2(t)$ in Eq. (3.50) with $\tilde{\mathbf{n}}(t) = \mathbf{0}$. As shown in Chapter 3, the performance of the method with this pre-filtering reference signal is the best achievable by the estimator with any $\mathbf{r}_2(t)$. The method was applied with no knowledge of the noise

covariance structure. However, the number of clusters and the number of signal sources in each cluster were known a priori.

5. The PFML-3 method was performed with $s_{nc}(t)$ as the pre-filtering reference signal, which is $r_3(t)$ in Eq. (3.67) with $\tilde{A} = I_{d_s}$ and $\tilde{n}(t) = 0$. As noted in Chapter 3, the performance of this method with $\tilde{n}(t) = 0$ is the best achievable by the method with any $r_3(t)$. Except for EXP 6.8, the additive noise covariance structure R_{ns} was assumed known. In the experiments where R_{ns} was known, $\hat{R}_n$ in Eq. (3.71) was replaced by R_{ns} . In addition, the total number of signal sources K was assumed known.

6. The BPFML-2 method is the beamforming-based PFML-2 method as described in Chapter 4. The method was applied with the type-4 pre-filtering reference signal $r_4(t) = s_c(t)$, where $s_c(t)$ was obtained from Eq. (4.13) by the proposed beamforming procedure of Section 4.2.2. For the k th source cluster, the estimates of the left and right cluster widths, $\hat{\Delta}_{l;k}$ and $\hat{\Delta}_{r;k}$, were obtained by the first step of the described beamforming procedure. The power factor β in Eqs. (4.10, 4.12) was chosen to be $\beta = 2$. This value of β is not optimised. However, from empirical study, it was observed to yield a reasonable result. In addition, the weighting function $W(\theta)$ in Eqs. (4.10, 4.12) was chosen to be uniform, i.e., its value is constant and independent of θ . The number of source clusters and the number of signal sources in each cluster were known. However, the noise covariance structure was not known.

7. The BPFML-2a is similar to the BPFML-2 method except that it was applied with the estimates of the left and right cluster widths, $\hat{\Delta}_{l;k}$ and $\hat{\Delta}_{r;k}$, being chosen as $\tilde{\Delta}_c/2$, where the approximate cluster angular width $\tilde{\Delta}_c$ (in degrees) is equivalent to 1.6 times the array beamwidth ($\tilde{\Delta}_c \sim 22$ degrees when calculated at broadside). Other parameter values used by the method are the same as those for the BPFML-2 method.

8. The BPFML-3 method is the beamforming-based PFML-3 method in Chapter 4. The type-4 pre-filtering reference signal, the estimates of the left and right cluster

widths, $\hat{\Delta}_{l;k}$ and $\hat{\Delta}_{r;k}$, as well as other parameters such as the weighting function $W(\theta)$, were obtained as described for the case of the BPFML-2 method. The noise covariance structure, the number of source clusters, and the total number of signal sources were assumed known.

9. The BPFML-4 method is the beamforming-based PFML-4 method as described in Section 4.2.3. The type-4 pre-filtering reference signal required by the method was obtained from Eq. (4.22) of the improved beamforming procedure in Section 4.2.3. Other parameters were the same as for the BPFML-2 method. The number of source clusters and the number of signal sources in each cluster were known. The additive noise covariance structure was not known.

10. The BPFML-2+bc, BPFML-3+bc, and BPFML-4+bc are, respectively, the bias corrected versions of the BPFML-2, BPFML-3, and BPFML-4 methods. The bias correction term for each method was obtained as described in Section 4.3, under the assumption that the noise covariance structure and the rank of the signal covariance matrix are known.

11. The RMNORM method is the root version of the minimum norm procedure, described in Chapter 2. The method was applied with forward/backward averaging of, say J , identical uniform linear subarrays which are overlapped by one sensor.

In all experiments, $s(t)$ was assumed to be temporally white. Except for EXP 6.8, $n(t)$ was spatial/temporal white with the noise power σ^2 at each sensor. In EXP 6.8, the noise covariance matrix $\mathbf{R}_n = \sigma^2 \mathbf{R}_{ns}$, where the diagonal elements of the noise covariance structure $\mathbf{R}_{ns}$ are 1's. Note that, each component of $s(t)$ was measured at the array centre which was defined as the middle point between the fourth and fifth sensors. Here the SNR is defined as $\text{SNR} \triangleq -10 \log_{10} \sigma^2$. The results for each signal and noise configuration were obtained from five hundred simulations using different pseudo noise sequences. Detail of the experiments and results are given as follows.

EXP 6.1: In the first experiment, two coherent signals arrived from the directions of 20 and 15 degrees, with respect to broadside. The first signal source is defined as the one with the DOA of 20 degrees. Both signals had unity power with phase factors of $\alpha_1 = 0$ and $\alpha_2 = \pi/4$. This signal configuration corresponds to the scenario of the single source cluster, consisting of two coherent sources. In this case, the matrix Γ_{cs} in Eq. (6.3) reduces to be the 2×1 vector $\gamma = [1 \ \exp(j\pi/4)]^T$. Thus, $s_{nc}(t)$ in Eq. (6.1) becomes the scalar quantity corresponding to the first signal source. In addition to the two coherent signals, the additive noise was added at each sensor to obtain eight different SNR values. Five DOA estimators were applied to the same data set. From Figure 6.1, it is seen that the (optimal) WSF method, the PFML-2, BPFML-2, and BPFML-2+bc methods achieved approximately the same performance level. The performance of the DML method is, however, clearly inferior to the other methods. The estimated RMSE_{av} of the DOA estimates using the BPFML-2 method (without bias correction) is essentially the same as that due to the BPFML-2+bc method (with bias correction). The RMS value of the CRLBs on the standard deviations, shown in the figure, was obtained by Eq. (3.45), with the first signal as the known waveform. Note that, in this source scenario, the CRLB under this known signal waveform assumption coincides with the CRLB under the unknown deterministic signal assumption. Figure 6.2 shows separately the estimation errors of the two source direction estimates obtained by the PFML-2 and BPFML-2 methods. From the figure, it is seen that the estimated RMSEs of the two source directions are essentially the same for either of the two methods. The CRLBs on the standard deviations were attained by the RMSEs of the estimates obtained by each of the two methods, within a wide range of the SNR value.

Next, the biases of the two source DOA estimates for three PFML methods are shown in Figure 6.3. The first DOA is at 20 degrees and the second DOA is at 15 degrees. The two source DOA estimates obtained by the PFML-2, BPFML-2, and BPFML-2+bc methods were found to be virtually unbiased at high SNR, i.e., the bias is

small in magnitude when compared with the standard deviation. The BPFML-2+bc and PFML-3 methods yielded DOA estimates with the same bias level. Note that, despite comparable RMSE value, at low SNR, the estimates obtained by the BPFML-2 method exhibited larger bias than that of those obtained by the BPFML-2+bc method.

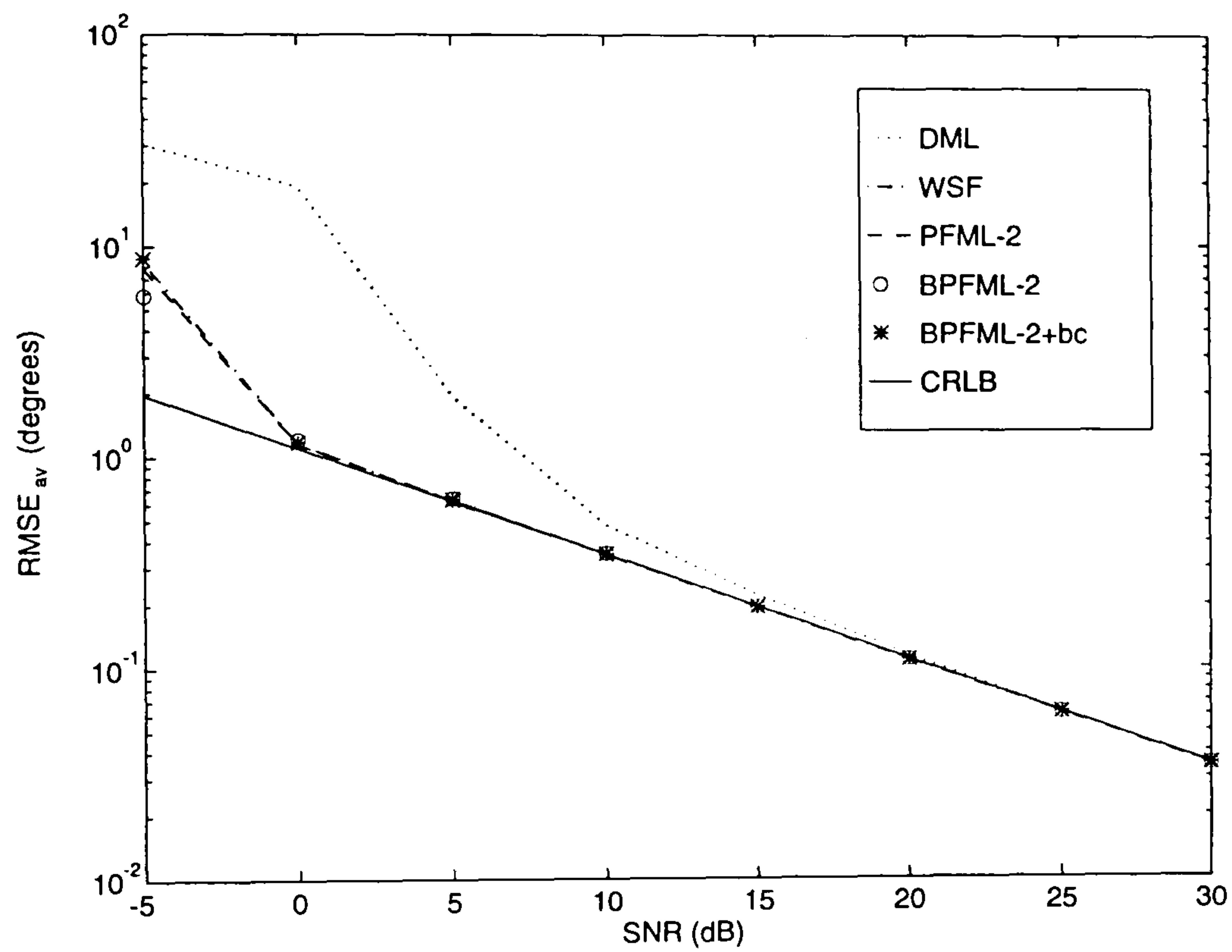


Figure 6.1: The estimated RMSE_{av} of two coherent-source DOA estimates obtained by each of five estimators, shown with the RMS value of the CRLBs on the standard deviations of the estimates at eight SNRs.

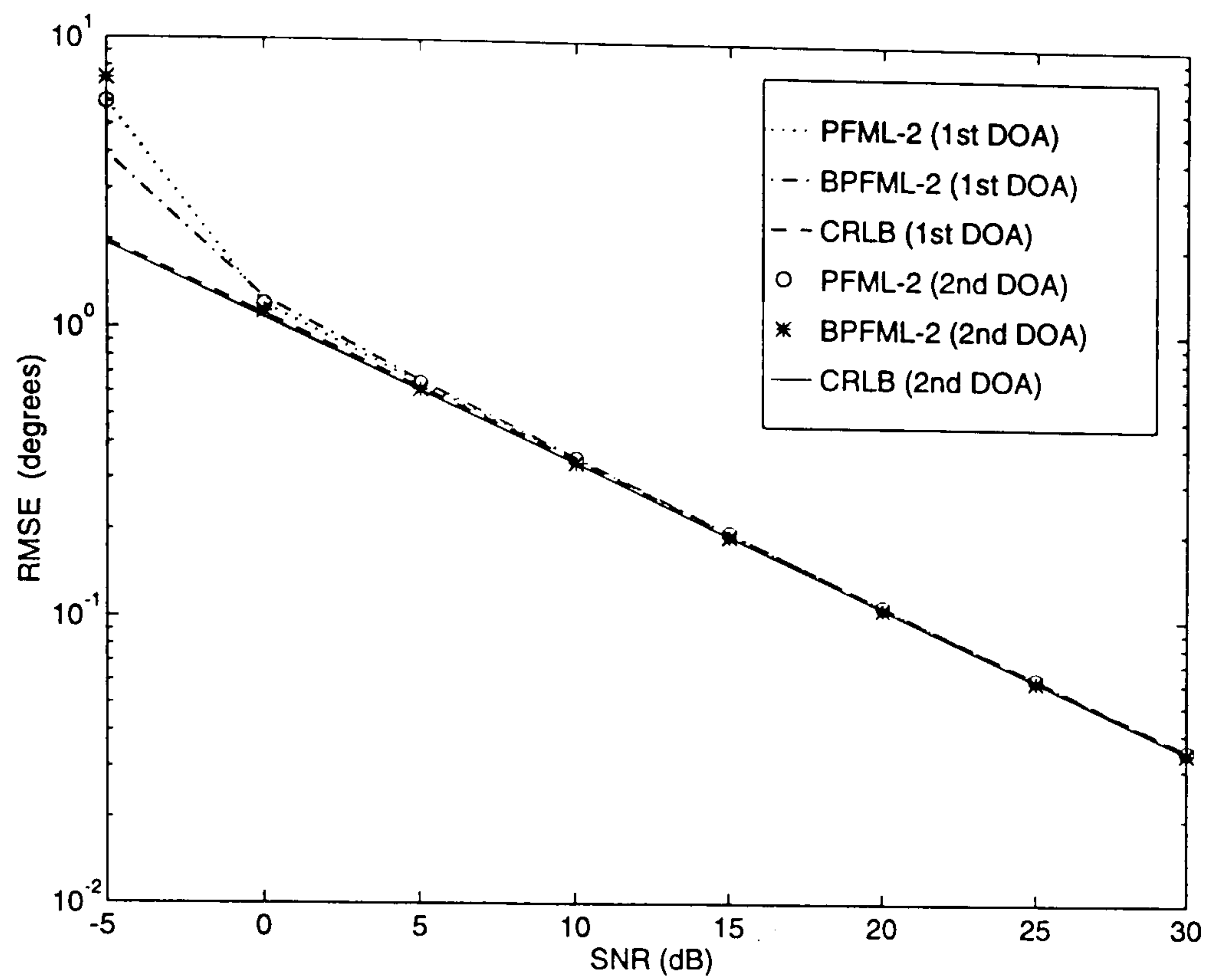


Figure 6.2: The estimated RMSEs of two coherent-source DOA estimates obtained by the PFML-2 and BPFML-2 methods, shown with the CRLBs on the standard deviations of the estimates at eight SNRs.

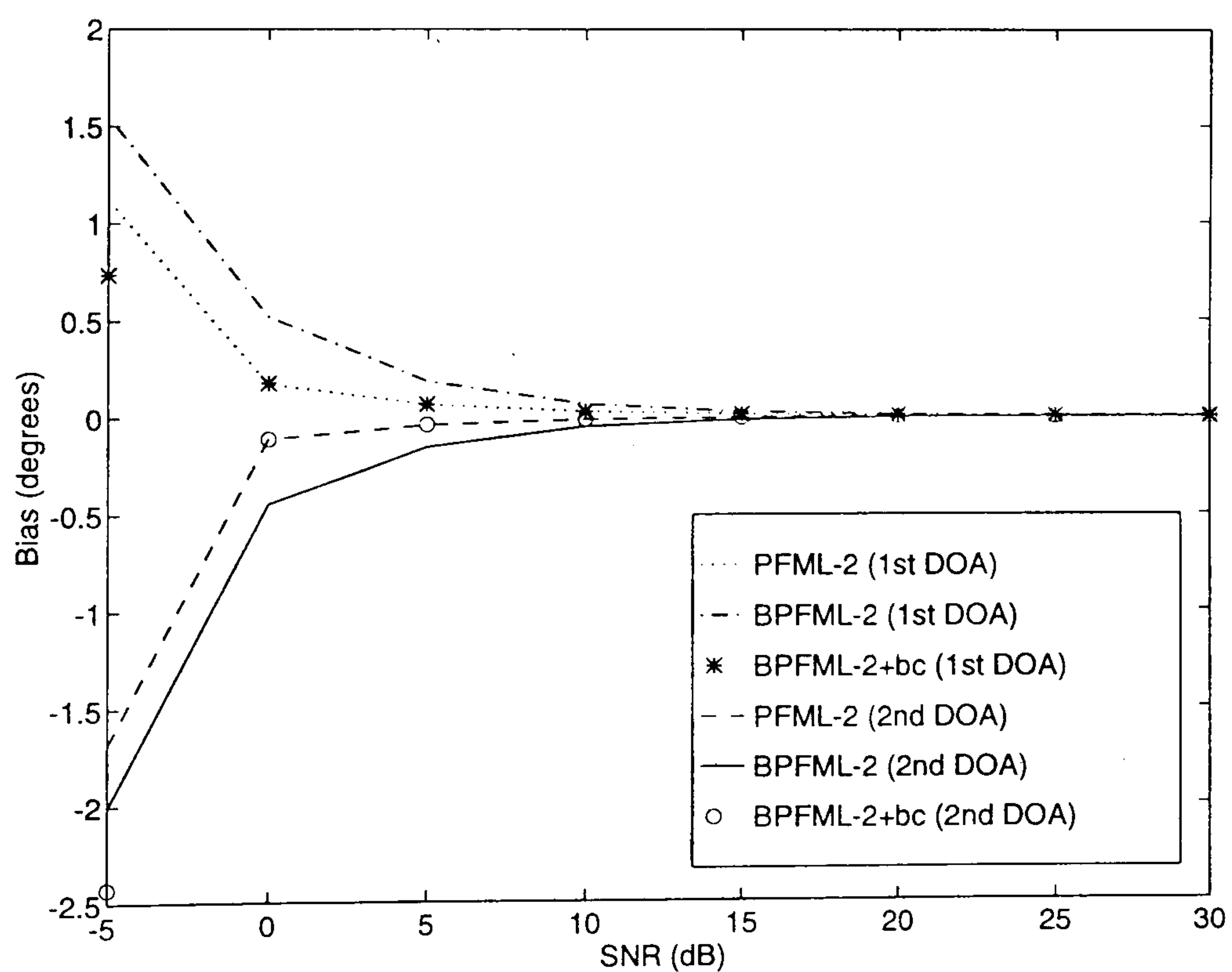


Figure 6.3: The biases of two coherent-source DOA estimates obtained by the PFML-2, BPFML-2, and BPFML-2+bc methods, at eight SNRs.

EXP 6.2: In this experiment, in addition to the first two coherent signals from the directions of 20 and 15 degrees as in EXP 6.1, another two coherent signals arrived from the directions of 5 and 0 degrees. Let the first two coherent signals belong to the first source cluster and be uncorrelated with the latter two signals of the second cluster. All signals had unity power with the phase difference between any two coherent signals equal to $\pi/4$. In this scenario, according to the signal model Eq. (6.1), the matrix

$$\Gamma_{cs} = \begin{bmatrix} 1 & \exp(j\pi/4) & 0 & 0 \\ 0 & 0 & 1 & \exp(j\pi/4) \end{bmatrix}^T \quad (6.4)$$

and $d_s = K_{cs} = K_c = 2$, $K = 4$. The vector $s_{nc}(t)$ was composed of the first and the third signal waveforms. The additive noise was added at each sensor to obtain eight different SNRs. The ensemble average of the array output spatial spectrum $P_{ss}(\theta)$ (see Eq. (4.9)) is shown in Figure 6.4. For the beamforming-based methods performed in the experiment, two highest peak locations in the spatial spectrum $P_{ss}(\theta)$ were chosen as the estimates of the two cluster nominal angular locations. Of the two peak locations, the one which was closer to the direction of 17 degrees was used as the location estimate of the first cluster. The array beam pattern corresponding to the asymptotic beamforming weight vector $w_{as;1}$, which points at the first cluster nominal direction and is calculated by Eq. (4.18), is also shown in Figure 6.4. Only the directions of the two sources which belong to the first cluster (DOAs of 20 and 15 degrees) are considered here. A similar result was also found for the case of the other two coherent source directions. The estimated RMSEs of the two 1st-cluster source DOA estimates using various DOA estimators are shown in Figure 6.5. From the figure, the DEML method has the minimum RMSE for both estimates. The performance of the PFML-2 method is slightly inferior to that of the DEML method, but better than those due to the other methods. The PFML-3 and WSF methods achieved comparable estimation accuracy. The DML method yielded inferior result to that of the WSF method. However, it still

performed better than the RMNORM method. The RMNORM method was applied with forward/backward averaging of two identical overlapping subarrays of seven sensors each. In comparing Figure 6.5a with Figure 6.5b, for the DML and WSF methods, the estimate of the first source direction (DOA = 20 degrees) is slightly more accurate than the estimate of the second source direction (DOA = 15 degrees). The RMSEs of the two estimates are slightly different for the case of the DEML and PFML-2 methods. For the PFML-2 method, the estimate of the first signal direction is slightly more accurate than the other. In the DEML method, the second DOA estimate is in turn more accurate. For the latter case, the better result (lower RMSE) for the second estimate is due to the closer DOA of the second source to the array broadside. It is a general property of a uniform linear array that, the closer the source direction is to the array broadside, the more accurate it can be estimated. Note that, for the case of the DEML and PFML-2 methods, at most SNR values, the estimated RMSEs by computer simulation of the two DOA estimates are in good agreement with the asymptotic analytical RMSEs of the same method. These analytical RMSEs were obtained as the square-root values of the diagonal elements of the error covariance matrix, which is given by Eq. (3.43) for the DEML method and by Eq. (3.59) for the PFML-2 method. The difference between the simulation and analytical results for both methods becomes noticeable at SNR below 5 dB. It is worth mentioning that each of the asymptotic error covariance matrices of the DEML and PFML-2 DOA estimates (Eq. (3.43) and Eq. (3.59)) coincides with the corresponding CRLB. It can thus be said that the best performance achievable by each of the two methods was attained for $\text{SNR} \geq 5$ dB.

Figure 6.6 compares the estimated RMSEs of the two 1st-cluster DOA estimates using the PFML-2 method with those obtained by the BPFML-2 and BPFML-2a methods. It is seen that the BPFML-2a method achieved comparable performance to that of the PFML-2 method. For SNR above 5 dB, the estimated RMSEs of the two estimates using both methods are closely matched with the asymptotic analytical

RMSEs of the PFML-2 estimates. The asymptotic analytical RMSEs for the BPFML-2a method were also found to be essentially the same as those of the PFML-2 method. In contrast to the BPFML-2a method, the BPFML-2 method exhibited performance irregularity and inconsistency, as seen from Figure 6.6. For instance, as SNR increased from 25 to 30 dB, the RMSEs of the estimates using the BPFML-2 method increased rather than decreased. This unusual result can be understood by examining the estimates of the 1st-cluster angular location and the corresponding left and right cluster-edge locations, shown in Figure 6.7 for all of the 500 simulations at the SNR of 30 dB. From the figure, in 2 of 500 simulations (simulation nos. 12 and 341), the values of the left 1st-cluster-edge location estimate ($\hat{\theta}_{1,1}$) differed significantly from the values of the same parameter estimate in other simulations. This is because, in these two simulations, the beamforming spectral peaks corresponding to the two source clusters were merged together, creating an unresolvable situation. As a result, in the two simulations, the beamformer which was constructed in a subsequent step (Eq. (4.11)) of the procedure in Section 4.2.2, was unable to suppress the two coherent signals which belong to the second cluster. The presence of these signals at the beamforming output, which was used as the pre-filtering reference signal in the BPFML-2 estimator, caused the unusually large estimation error as seen in Figure 6.6.

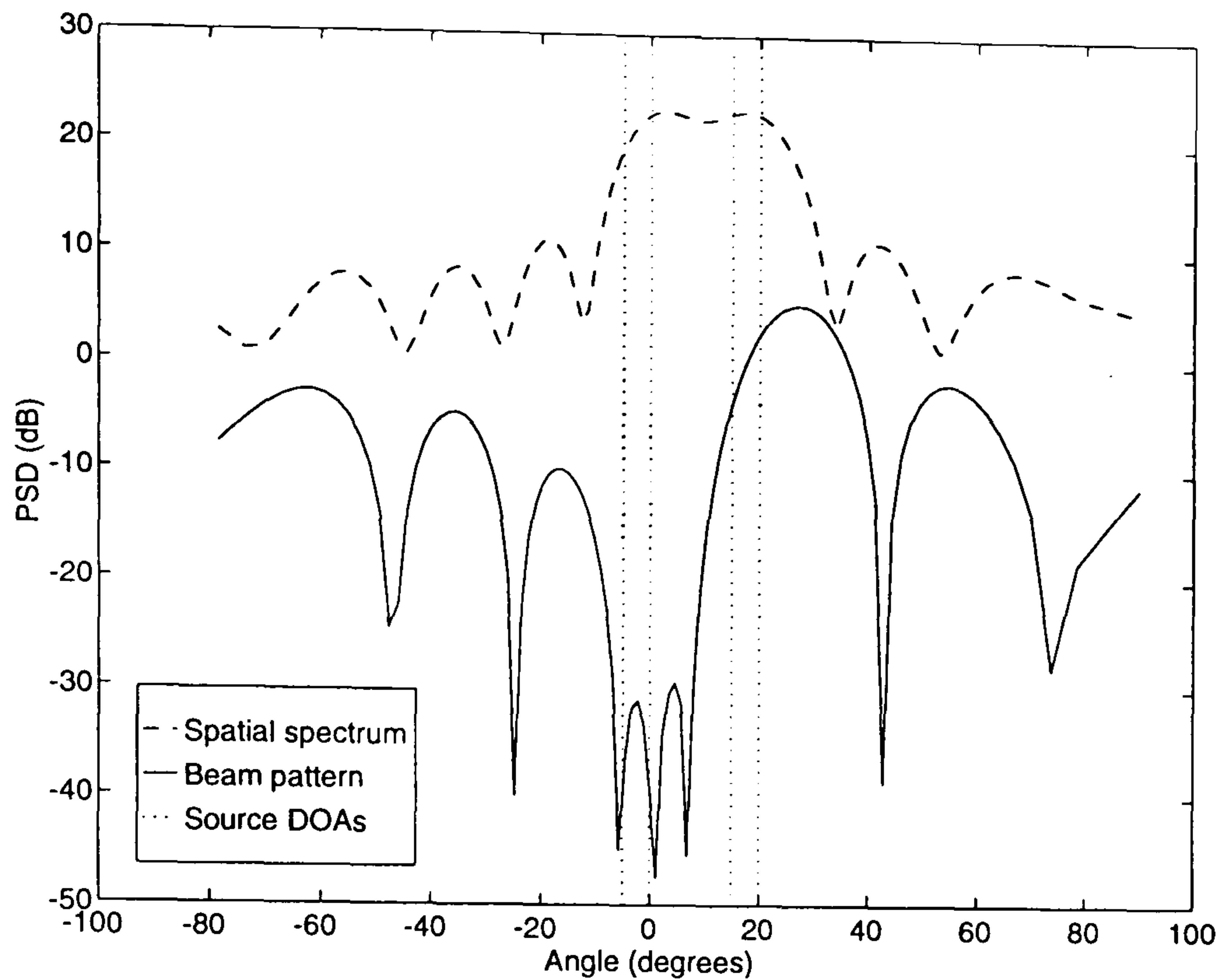
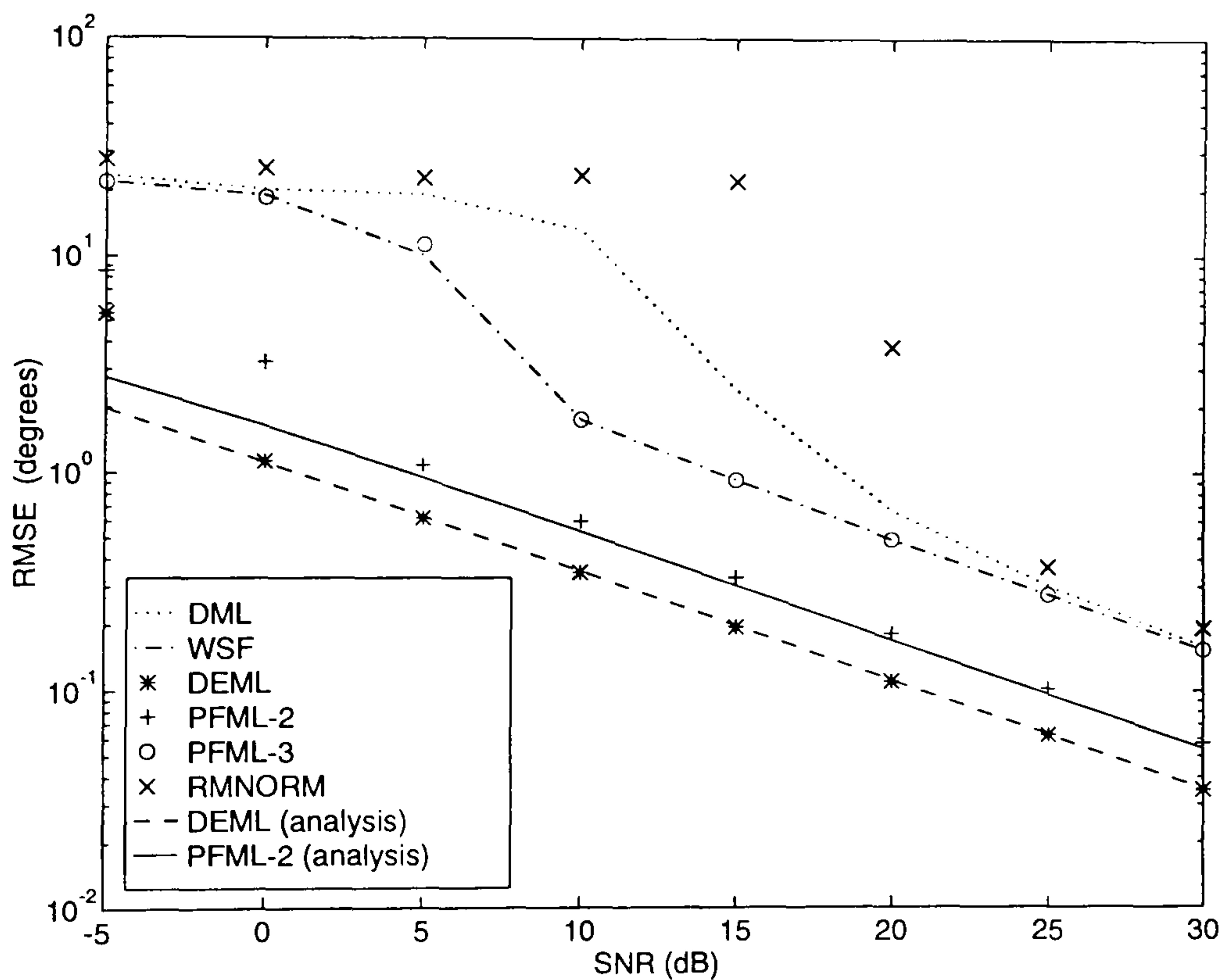
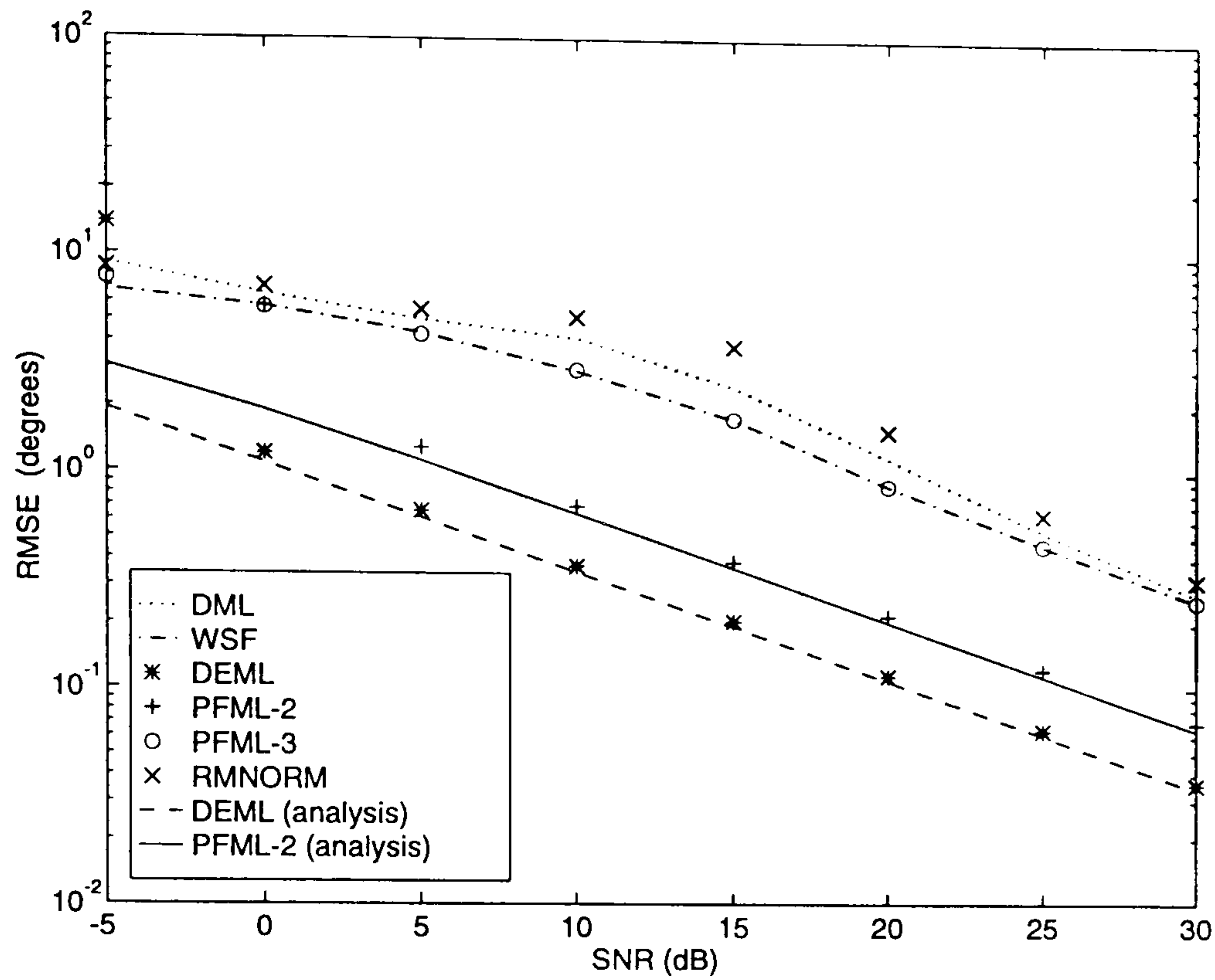


Figure 6.4: The ensemble average of the array output spatial spectrum, and the asymptotic beam pattern of the beamformer used by the BPFML-2 method in EXP 6.2, at the SNR of 10 dB.



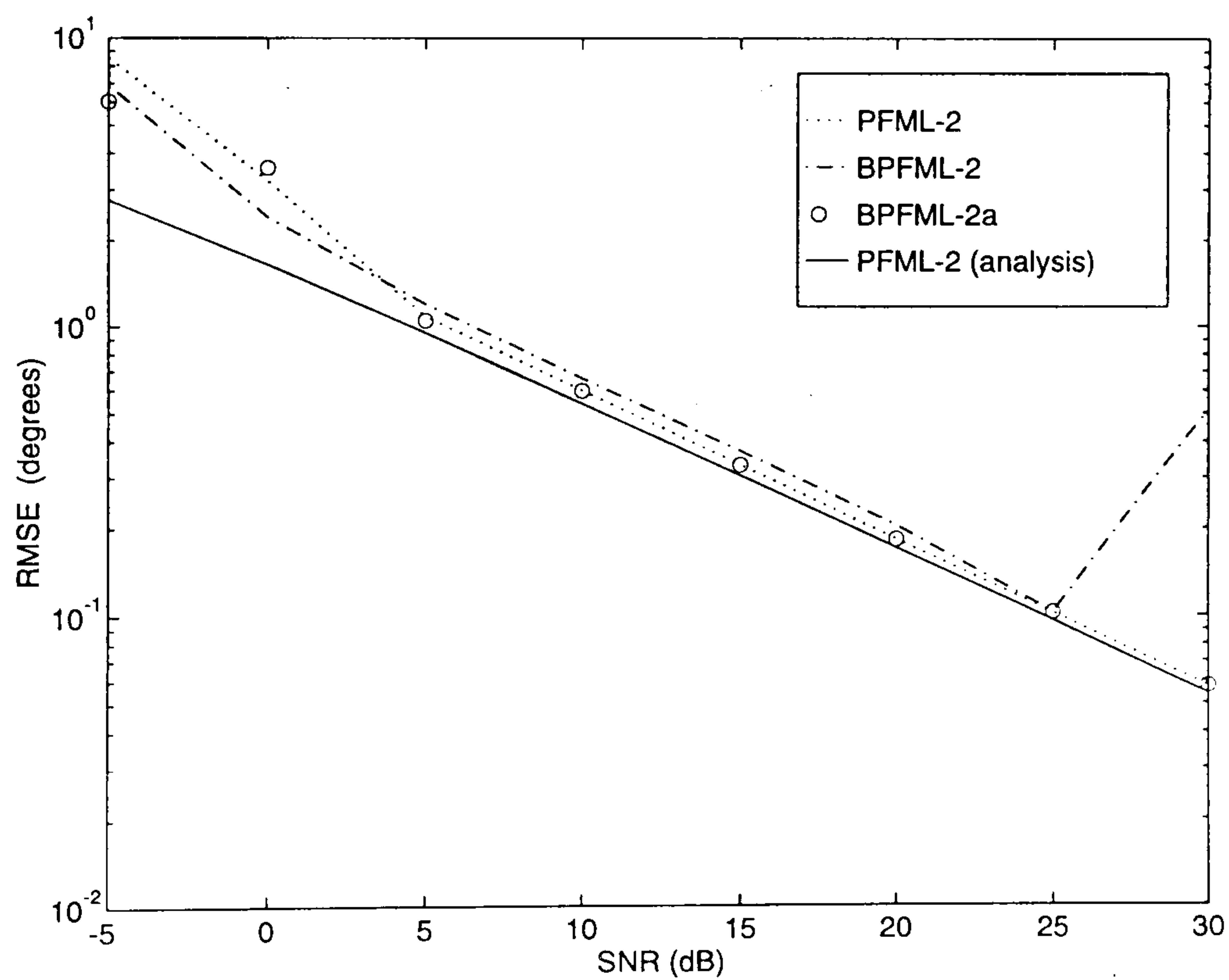
(a)

Figure 6.5: The estimated RMSEs of two 1st-cluster DOA estimates obtained by five ML-based methods and the RMNORM method, under the two-cluster four-source scenario of EXP 6.2, shown with the analytical RMSEs of the DEML and PFML-2 estimates at eight SNRs. (a) DOA = 20 degrees; (b) DOA = 15 degrees.



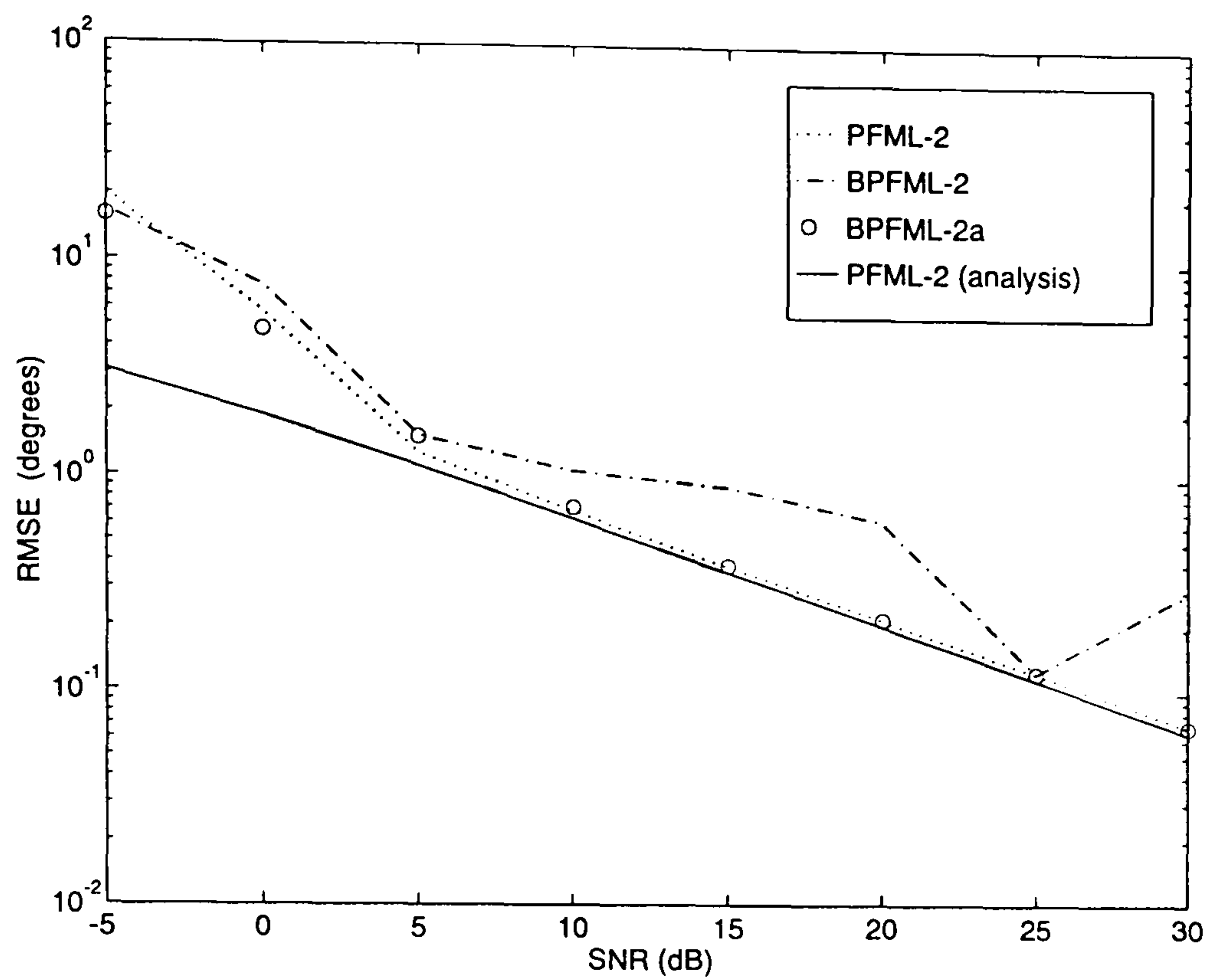
(b)

Figure 6.5: (Continued)



(a)

Figure 6.6: The RMSEs of two 1st-cluster DOA estimates obtained by the PFML-2, BPFML-2, and BPFML-2a methods, under the two-cluster four-source scenario of EXP 6.2, shown with the analytical RMSEs of the PFML-2 estimates at eight SNRs. (a) DOA = 20 degrees; (b) DOA = 15 degrees.



(b)

Figure 6.6: (Continued)

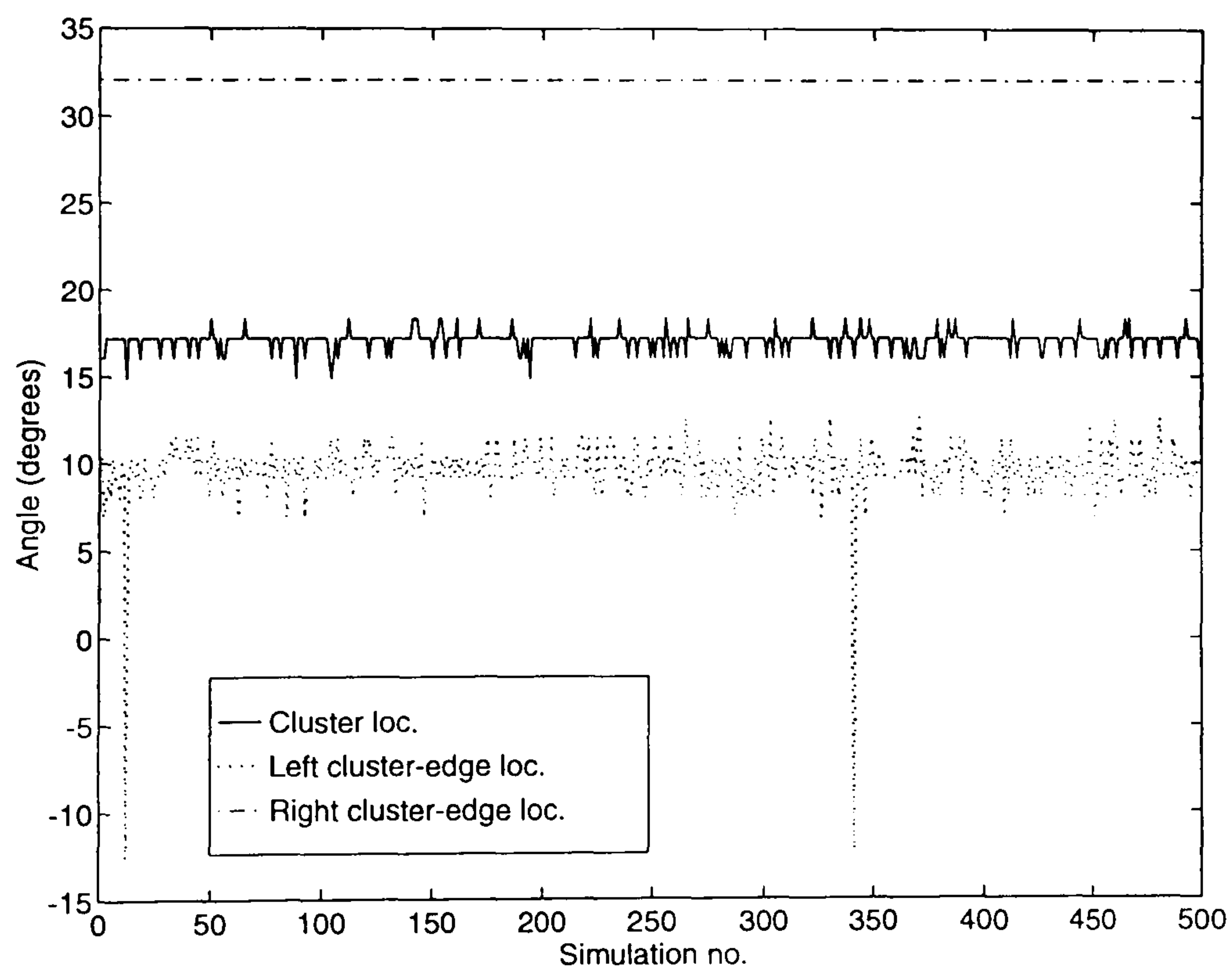


Figure 6.7: The estimates of the 1st-cluster angular location, and the corresponding left and right cluster-edge locations, under the two-cluster four-source scenario of EXP 6.2 with the SNR of 30 dB.

EXP 6.3: The two-cluster scenario of EXP 6.2 was again used. In this experiment, however, the directions of the 3rd and 4th coherent signals moved, respectively, to 0 and -5 degrees. As in EXP 6.2, only the directions of the first two coherent sources (20 and 15 degrees), which belong to the first cluster, are considered. For all of the beamforming-based methods performed in the experiment, two highest peak locations in the array output spatial spectrum $P_{ss}(\theta)$ were chosen as the estimates of the two cluster nominal angular locations. Of the two peak locations, the one which was closer to the direction of 17 degrees was used as the location estimate of the first cluster.

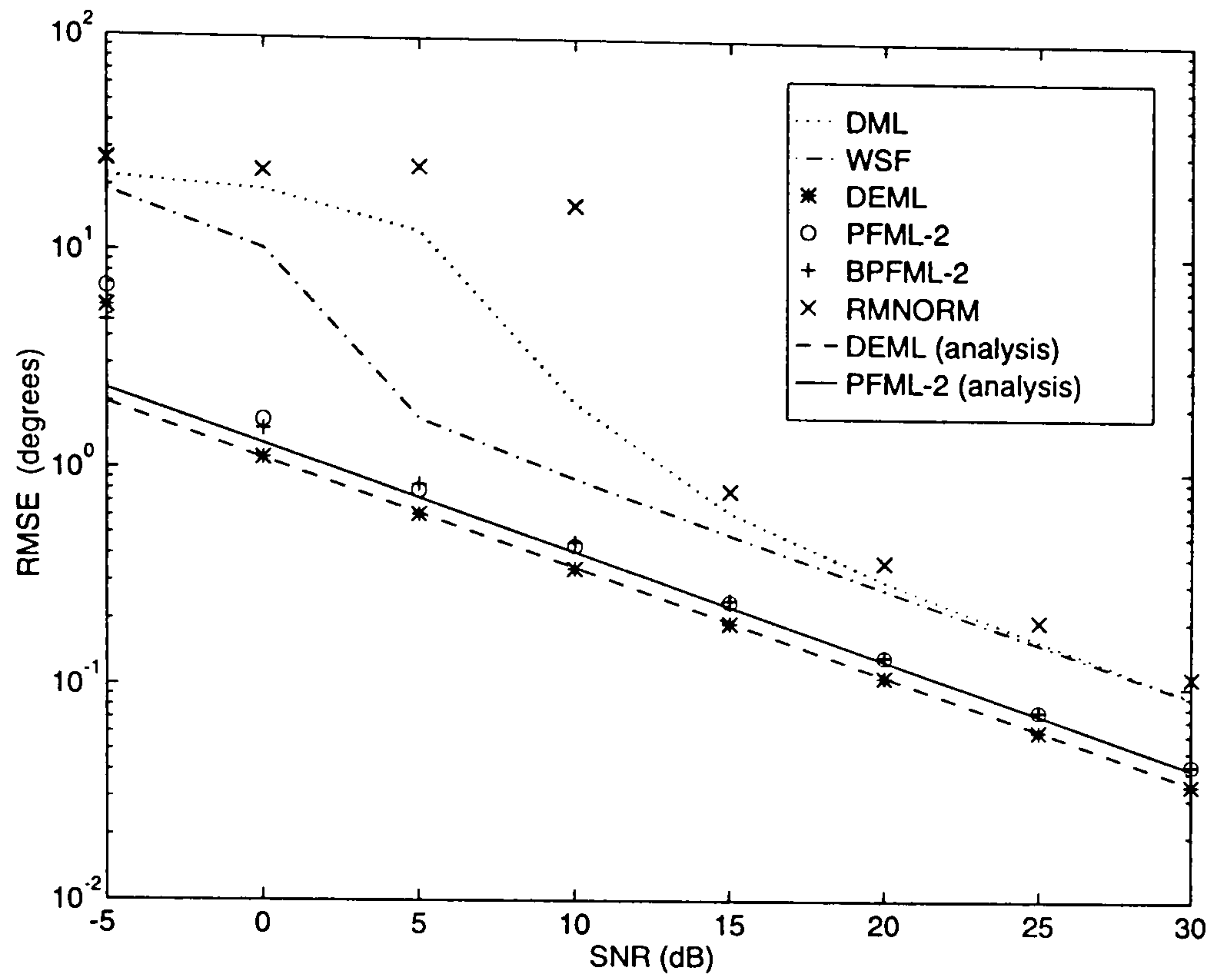
Figure 6.8 shows that, except for the DEML method, the RMSEs of the two DOA estimates are smaller than the ones of the same method in EXP 6.2. As addressed in Section 3.2.1, the performance of the DEML method is largely independent of the angular spacing between two uncorrelated sources. In our case, the performance of the DEML method remained unchanged as the directions of the third and fourth signals moved from 5 and 0 degrees in EXP 6.2 to 0 and -5 degrees in this experiment. In comparing Figure 6.8 with Figure 6.5, it is seen that the performance difference between the PFML-2 method and the WSF method in this experiment is less than in EXP 6.2. The improvement gained by using the PFML-2 and BPFML-2 methods instead of the DML, WSF, and RMNORM methods, remains significant. As in EXP 6.2, the RMNORM method is the root version of the minimum norm procedure with forward/backward averaging of two overlapping subarrays, each consisting of seven sensors. In addition, significant improvement can be seen from Figure 6.8 for the BPFML-2 method. The performance of the method is comparable to that of the PFML-2 method at most SNRs. This is because, in this experiment, the angular spacing between the two source clusters was sufficiently large such that the classical beamforming spectral estimator, used in Step 1 of the beamforming procedure of Section 4.2.2, correctly resolved the two cluster locations in all of the 500 simulations. The ensemble average of the array output spatial spectrum, and the array beam pattern corresponding

to the asymptotic weight vector $\mathbf{w}_{\text{as},1}$, calculated by Eq. (4.18), are shown in Figure 6.9. It is seen from the figure that, the spectral peaks corresponding to the two source clusters are well separated. Note that, in this experiment, the performance of the BPFML-2a method (not shown here) was found to be essentially the same as that of the BPFML-2 method. Figure 6.10 shows that the BPFML-2 method achieved superior performance to those of the BPFML-3 and WSF method. In addition, the BPFML-3 method achieved comparable performance to that of the WSF method. The BPFML-2 and BPFML-2+bc methods yielded approximately the same estimated RMSEs. At most SNRs, the estimated and analytical RMSEs of the BPFML-2 method are seen to be closely matched. The analytical RMSEs of the method are based on the asymptotic analysis as given in Chapters 3 and 4. Each of them is a combination of the error due to the finite data sample effect and the one due to the bias term $\mathbf{E}_b$ (see Sections 3.2.4 and 4.2.2 for more detail).

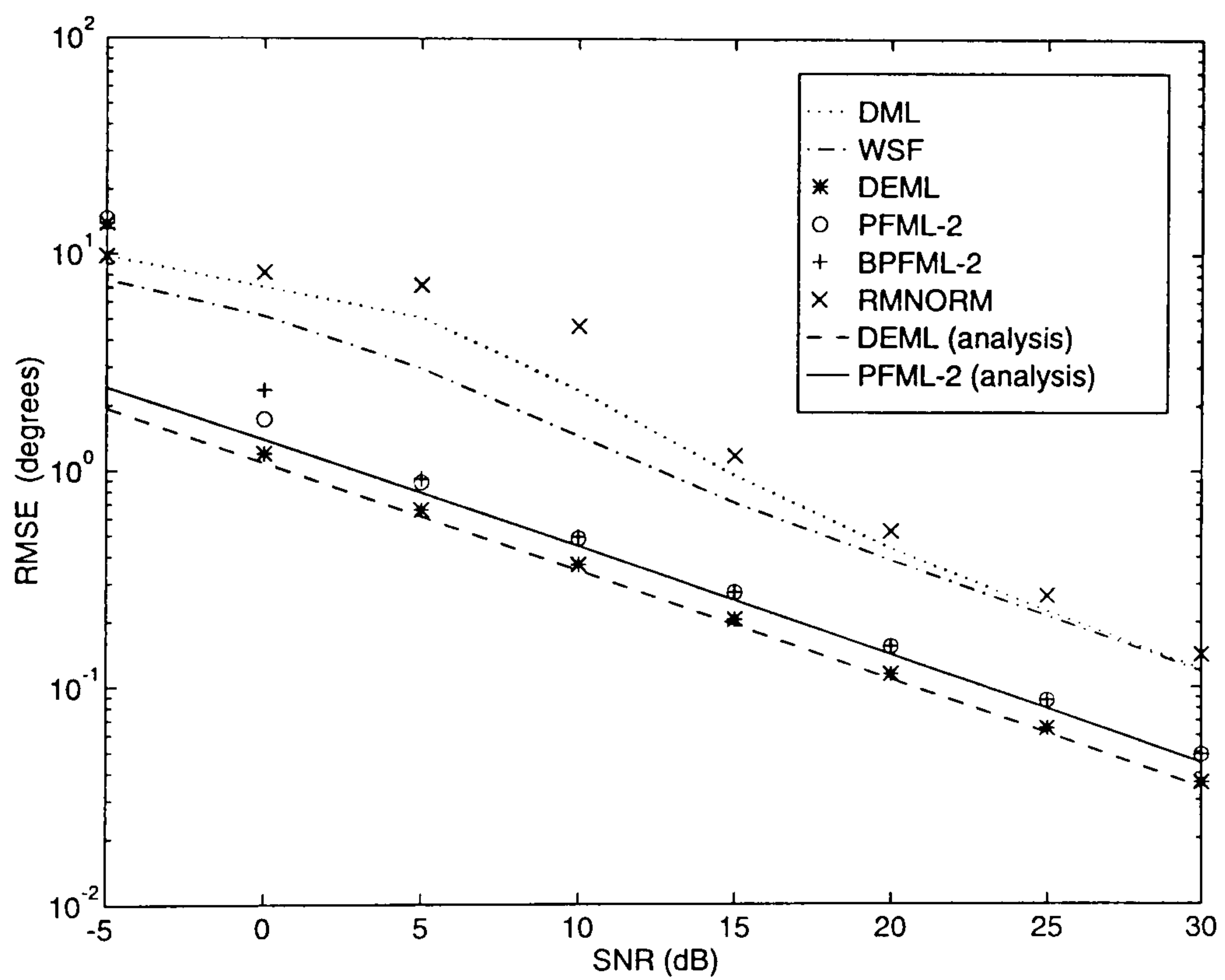
The bias in the DOA estimates is next considered. From Figure 6.11, at low SNR, the BPFML-2 method yielded estimates with larger bias than those of the PFML-2 method. The BPFML-2+bc method, however, is seen to have the same estimation bias level as that of the PFML-2 method. At high SNR (> 10 dB), the three estimators yielded virtually unbiased estimates.

So far the results from the PFML-2 and BPFML-2 methods were based on the use of the estimated noise covariance matrix $\hat{\mathbf{R}}_{n,1}$, given by Eq. (3.58). However, as mentioned in Chapter 3, except for the PFML estimators which are based on the type-4 pre-filtering reference signal, $\hat{\mathbf{R}}_n$ as given by Eq. (3.27) (or $\hat{\mathbf{R}}_{n,k}$ in Eq. (3.58)) can be replaced by the data covariance matrix $\mathbf{R}$ (or its consistent estimate) without affecting the asymptotic properties of the PFML estimators. This replacement can also be applied for the BPFML-2 method without significant change in its asymptotic properties, provided that the bias term $\mathbf{E}_b$ is sufficiently small. By applying the PFML-2 and

BPFML-2 methods with $\hat{\tilde{\mathbf{R}}}_{n;l}$ being replaced by $\hat{\mathbf{R}}$, we found that, for most of the eight SNR values, the properties (bias and RMSE) of the resulting DOA estimates were essentially the same as those obtained by using $\hat{\tilde{\mathbf{R}}}_{n;l}$. Only at low SNR (< 5 dB) did the methods using $\hat{\tilde{\mathbf{R}}}_{n;l}$ yield slightly lower RMSEs of the DOA estimates. Despite this performance similarity, the number of iterations required by the PFML-2 and BPFML-2 methods using $\hat{\tilde{\mathbf{R}}}_{n;l}$ was found to be smaller than in the methods using $\hat{\mathbf{R}}$ instead. To show this, the average numbers of iterations, required by the CIQML algorithm over 500 simulations, of the PFML-2 and BPFML-2 methods using $\hat{\mathbf{R}}$ and $\hat{\tilde{\mathbf{R}}}_{n;l}$, were plotted in Figure 6.12 at eight SNRs. Also included in the figure is the average number of iterations required by the DEML method. From the figure, at $\text{SNR} > 0$ dB, the PFML-2 and BPFML-2 methods based on $\hat{\tilde{\mathbf{R}}}_{n;l}$ required less iterations (converged faster) than the methods which were based on $\hat{\mathbf{R}}$. At $\text{SNR} = 30$ dB, the first two methods required approximately 3 iterations less than those needed by the latter two. Among all of the methods shown in the figure, the DEML method required the smallest number of iterations.



(a)



(b)

Figure 6.8: The estimated RMSEs of two 1st-cluster DOA estimates obtained by six estimators under the two-cluster four-source scenario of EXP 6.3, shown with the asymptotic analytical RMSEs due to the DEML and PFML-2 methods at eight SNRs. (a) DOA = 20 degrees; (b) DOA = 15 degrees.

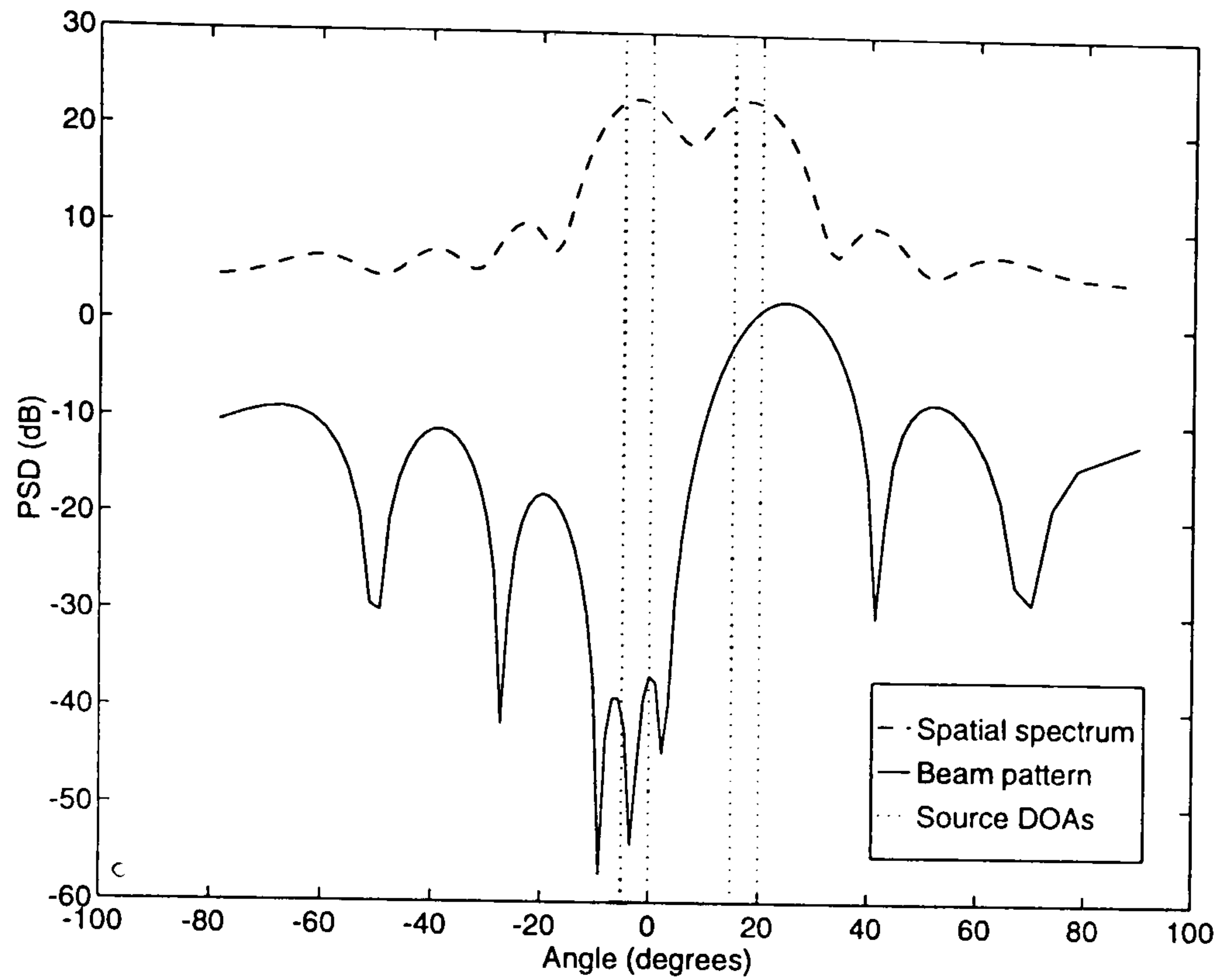
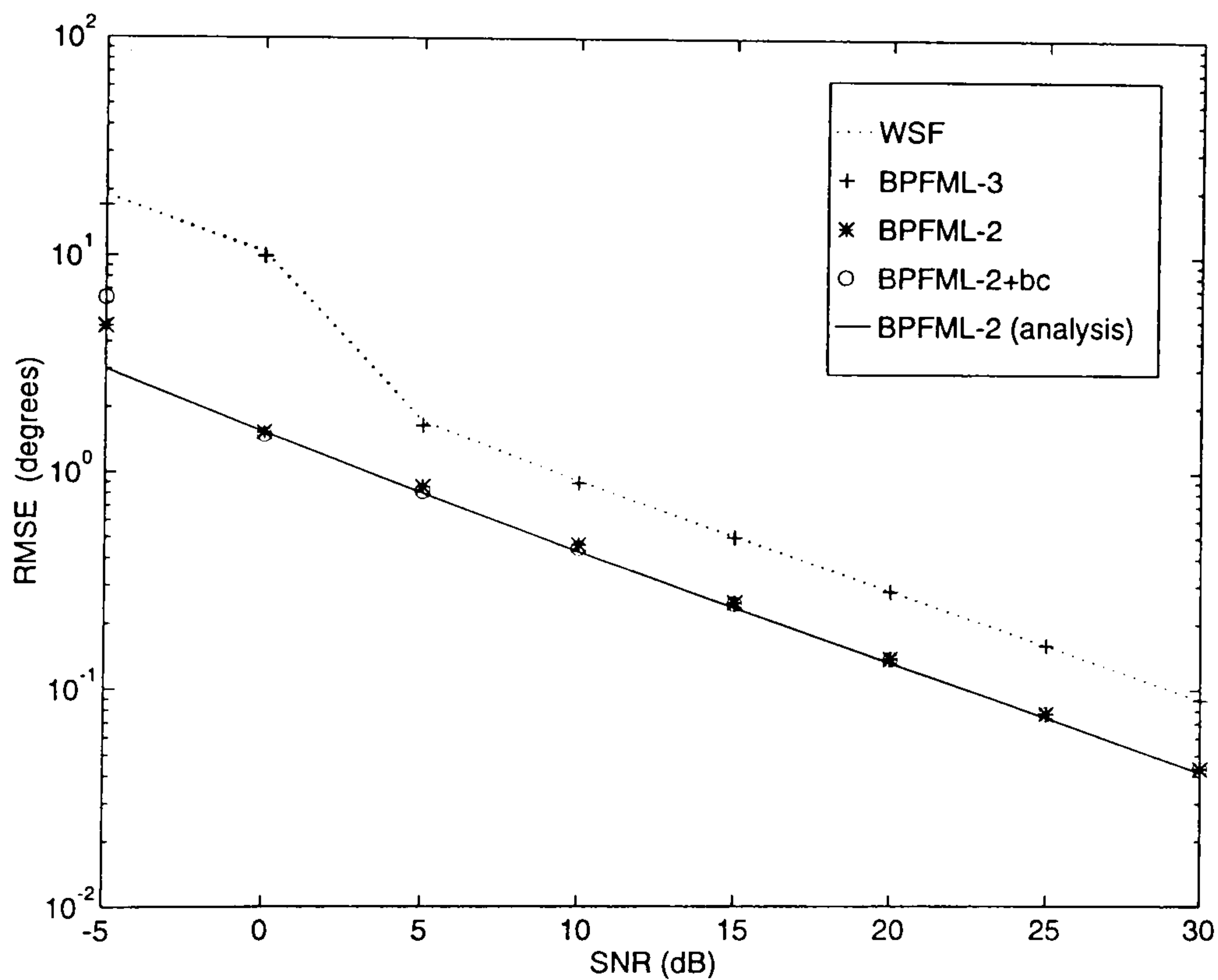
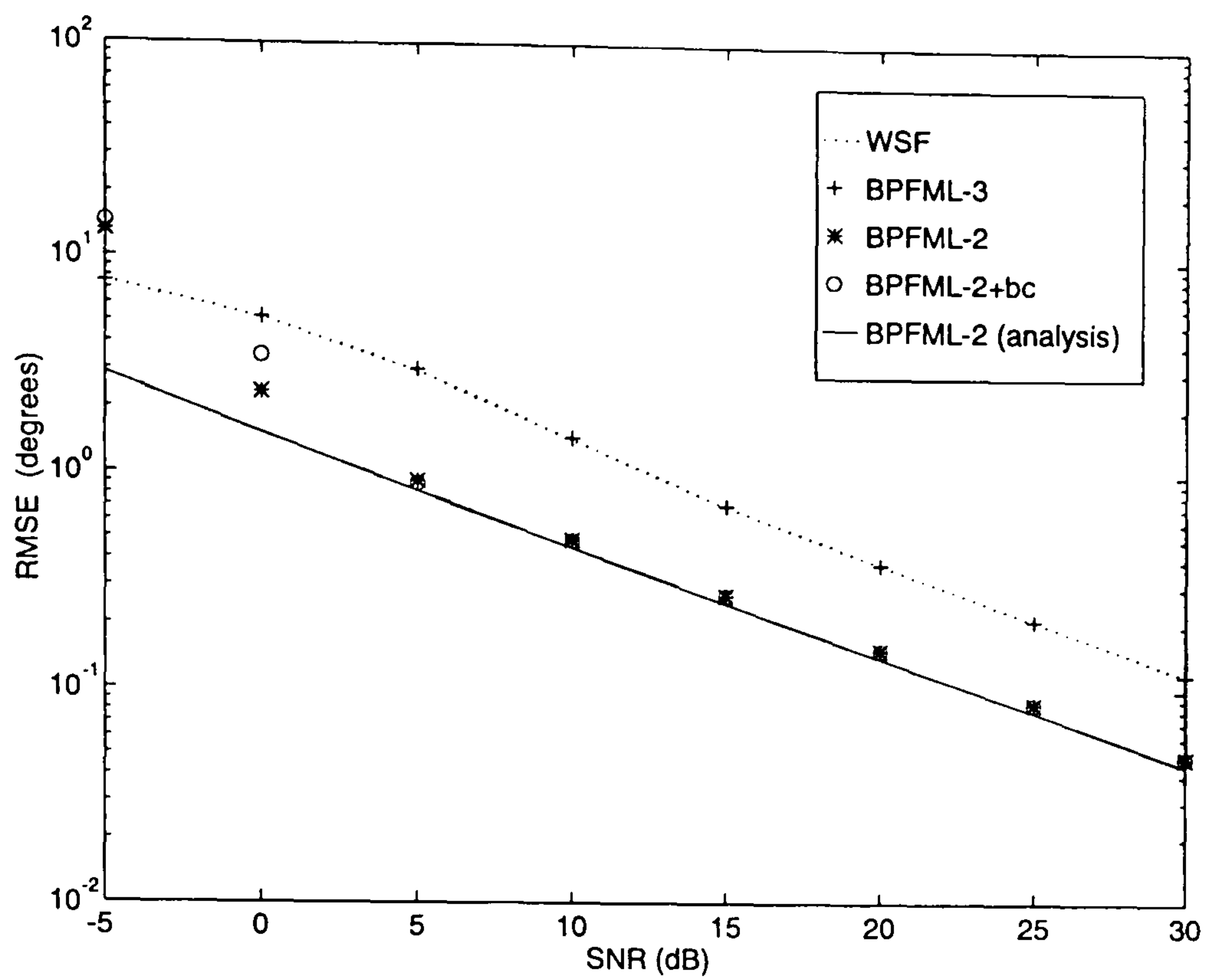


Figure 6.9: The ensemble average of the array output spatial spectrum and the asymptotic beam pattern of the beamformer used by the BPFML-2 and BPFML-2+bc methods, under the two-cluster four-source scenario of EXP 6.3 and the SNR of 10 dB.



(a)

Figure 6.10: The estimated RMSEs of two 1st-cluster DOA estimates obtained by the WSF method and three BPFML methods under the two-cluster four-source scenario of EXP 6.3, shown with the asymptotic analytical RMSEs due to the BPFML-2 method at eight SNRs. (a) DOA = 20 degrees; (b) DOA = 15 degrees.



(b)

Figure 6.10: (Continued)

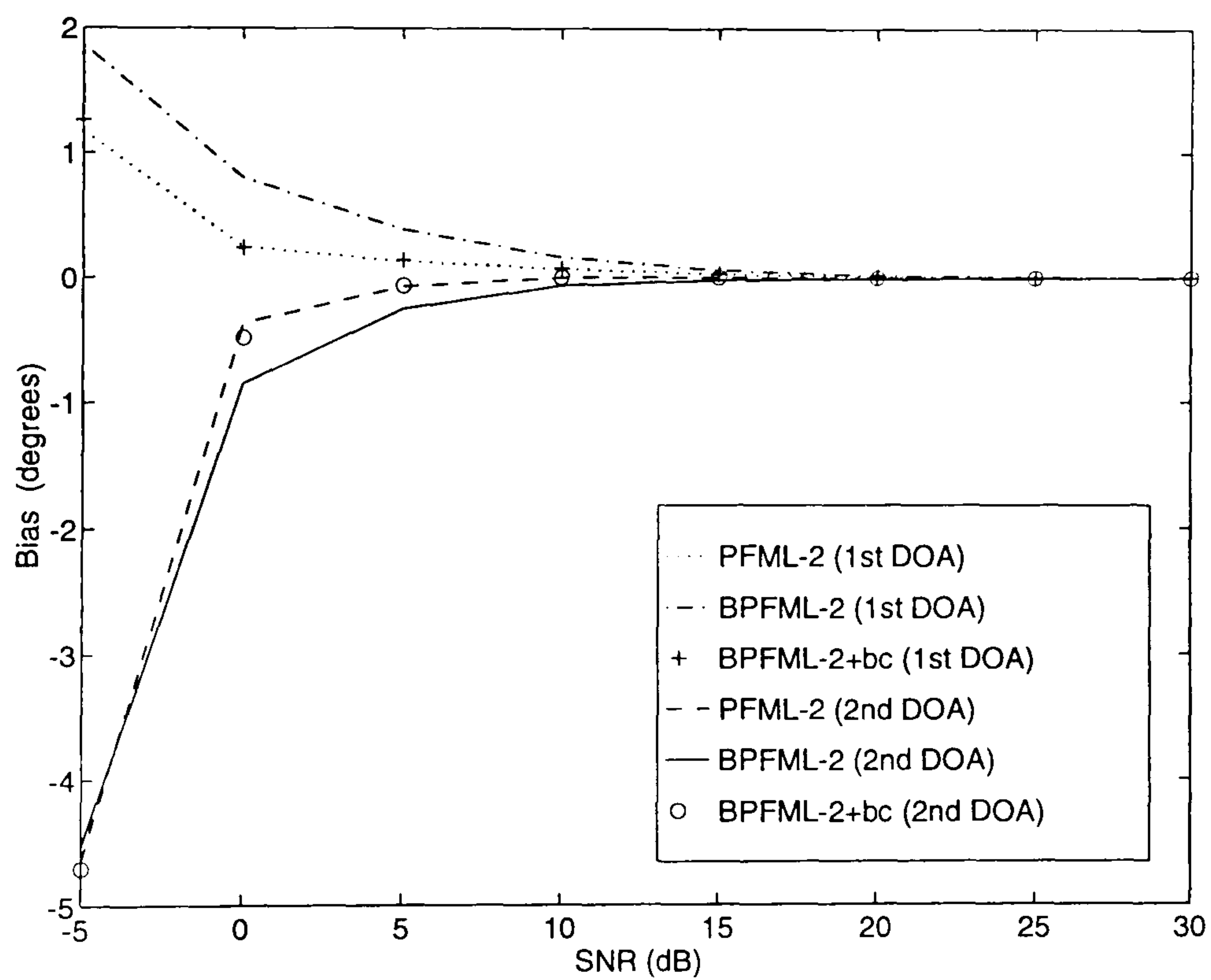


Figure 6.11: The biases of the first and second DOA (20 and 15 degrees) estimates, obtained by the PFML-2 and two BPFML methods, under the two-cluster four-source scenario of EXP 6.3 at eight SNRs.

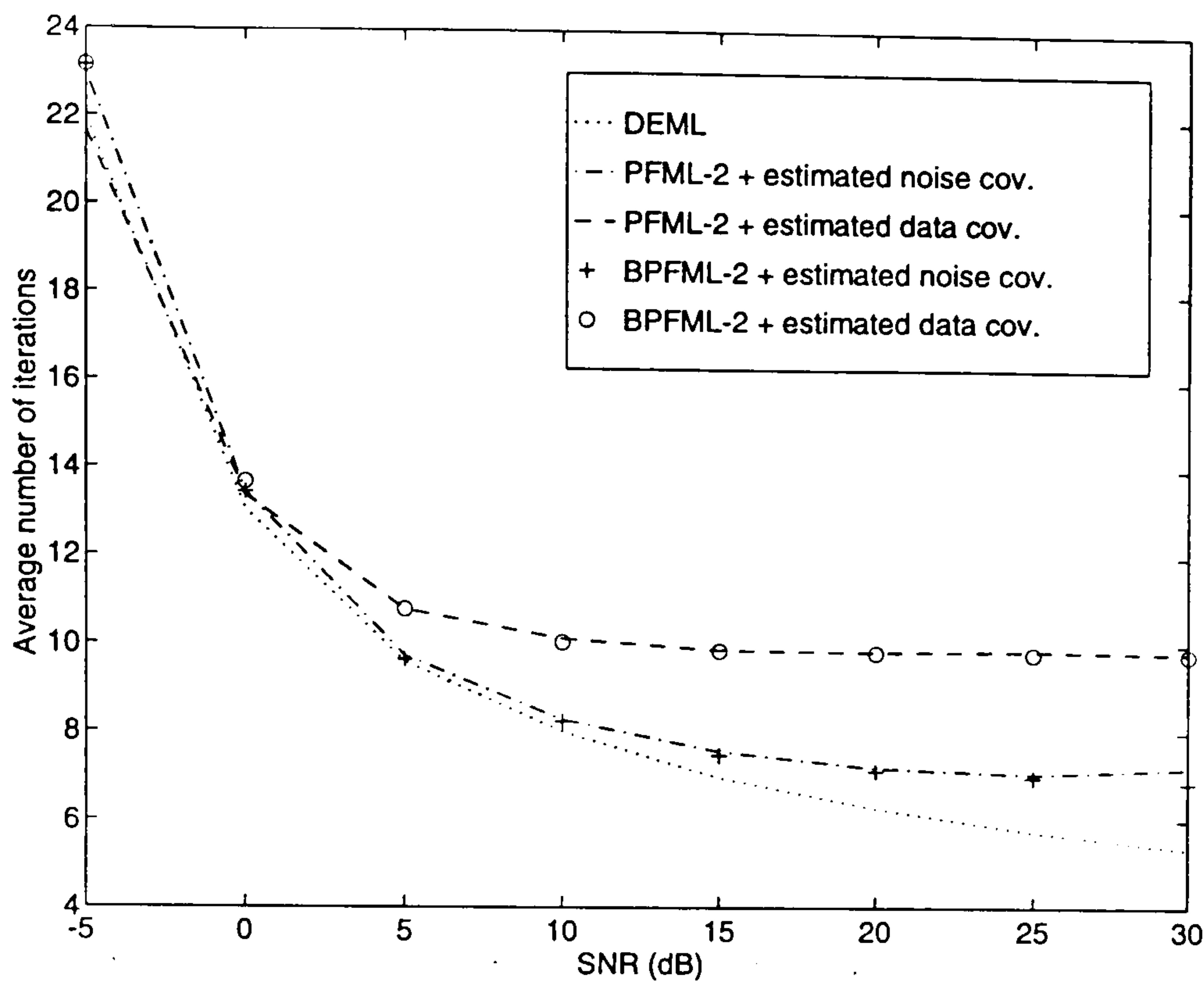
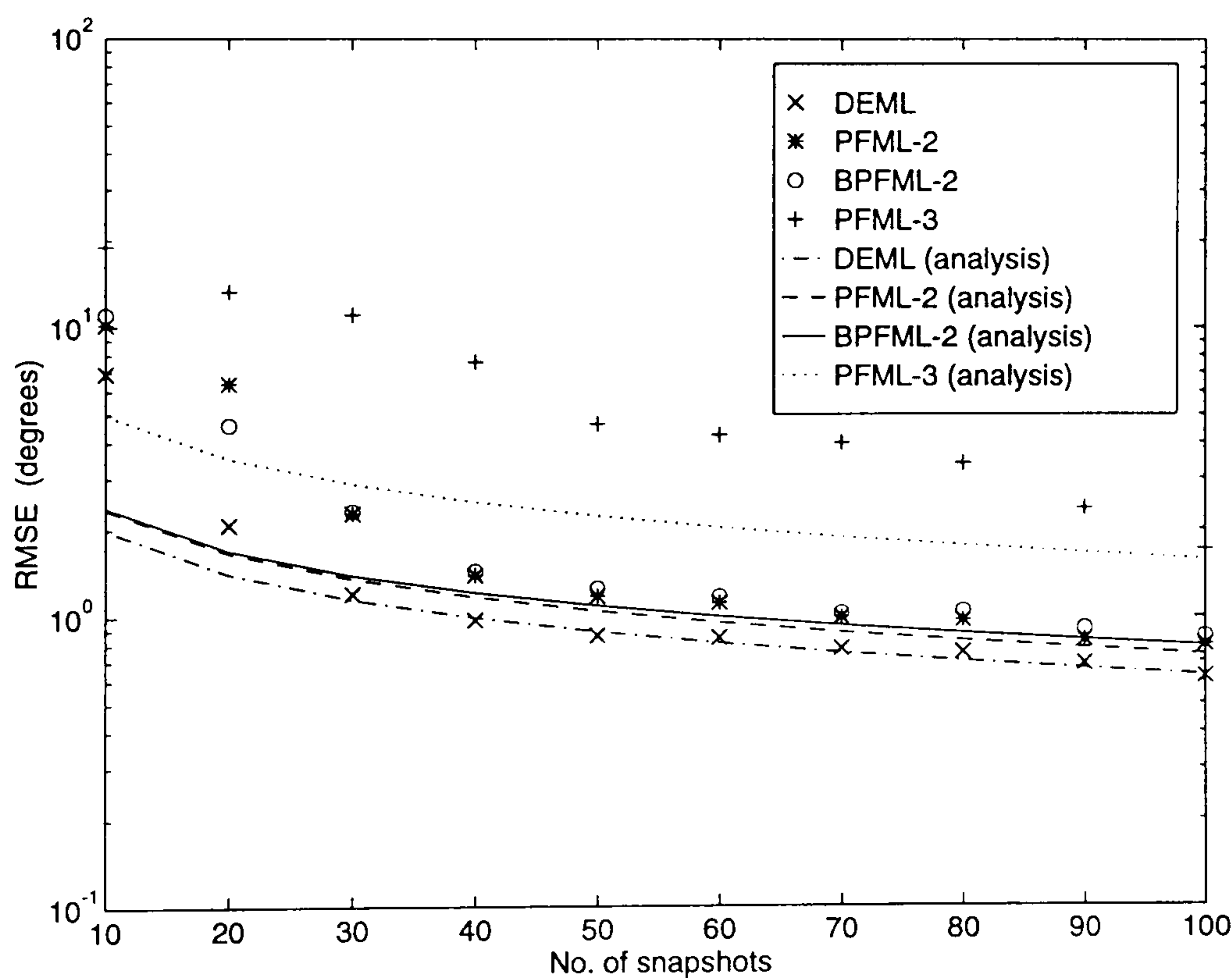


Figure 6.12: The average numbers of iterations required by five PFML estimators, performed in EXP 6.3 at eight SNRs.

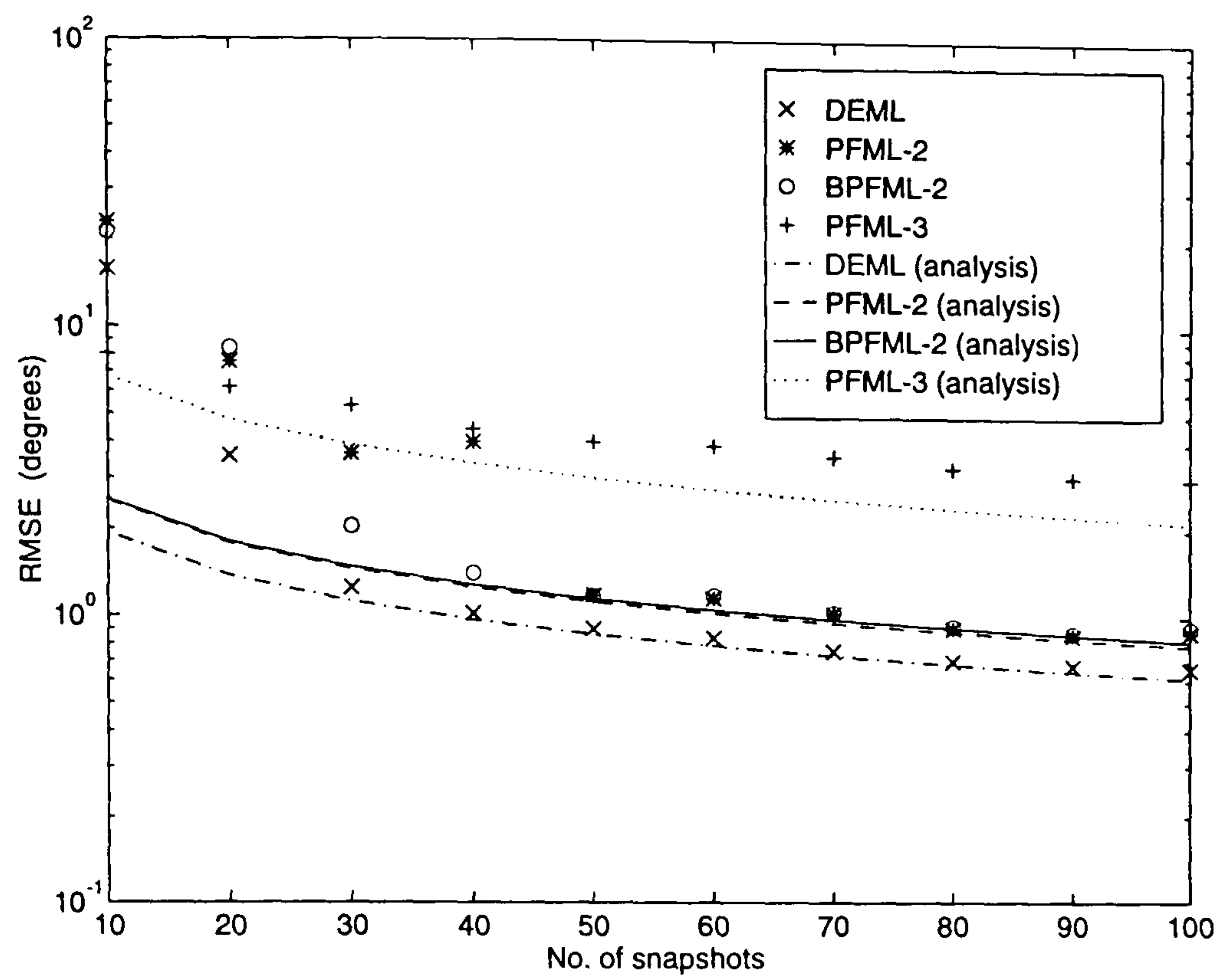
EXP 6.4: The signal and noise condition in this experiment was based on that of EXP 6.3. Here, however, the number of data snapshots L was varied to investigate its effect on the estimation performance. Results from the experiment will also be used to indicate how large the number of data snapshots is required for the asymptotic analytical result to be valid. Figure 6.13 shows the estimated RMSEs at the SNR of 5 dB for the first and second DOA estimates. From the figure, the PFML-2 and BPFML-2 methods yielded similar performance. The performance of these methods was inferior to that of the DEML method, particularly for $L < 40$. The PFML-3 method produced unreliable results for all values of L . The RMSEs of the estimates obtained by the method are of the same order or greater than the angular separation between the two signals. At $L = 10$, the results from all methods are poor. When comparing the estimated RMSEs due the DEML method with that obtained by the asymptotic analysis, it is seen that both results are closely matched for $L \geq 30$. For the case of the PFML-2 method, good agreement between the estimated and analytical RMSEs is seen for $L \geq 40$. The same remark can

also be said of the BPFML-2 method. As SNR was increased to 10 dB, it is seen from Figure 6.14 that the estimated RMSEs due to each of the DEML, PFML-2, and the BPFML-2 methods are close to the corresponding asymptotic analytical RMSEs for L as low as 20. The performance of the PFML-3 method was also improved. We note that, based on results from previous experiments and the fact that the error of the PFML-3 estimator asymptotically attains the CRLB for the deterministic signal model (see Chapter 3), the performance of the PFML-3 method represents the best performance which may be achievable by the DML method, the WSF method, and the IV-based DOA estimation methods. In addition, the RMSEs of the estimates using the BPFML-2+bc method (not shown here) was found to be comparable to those of the estimates obtained by the BPFML-2 method.



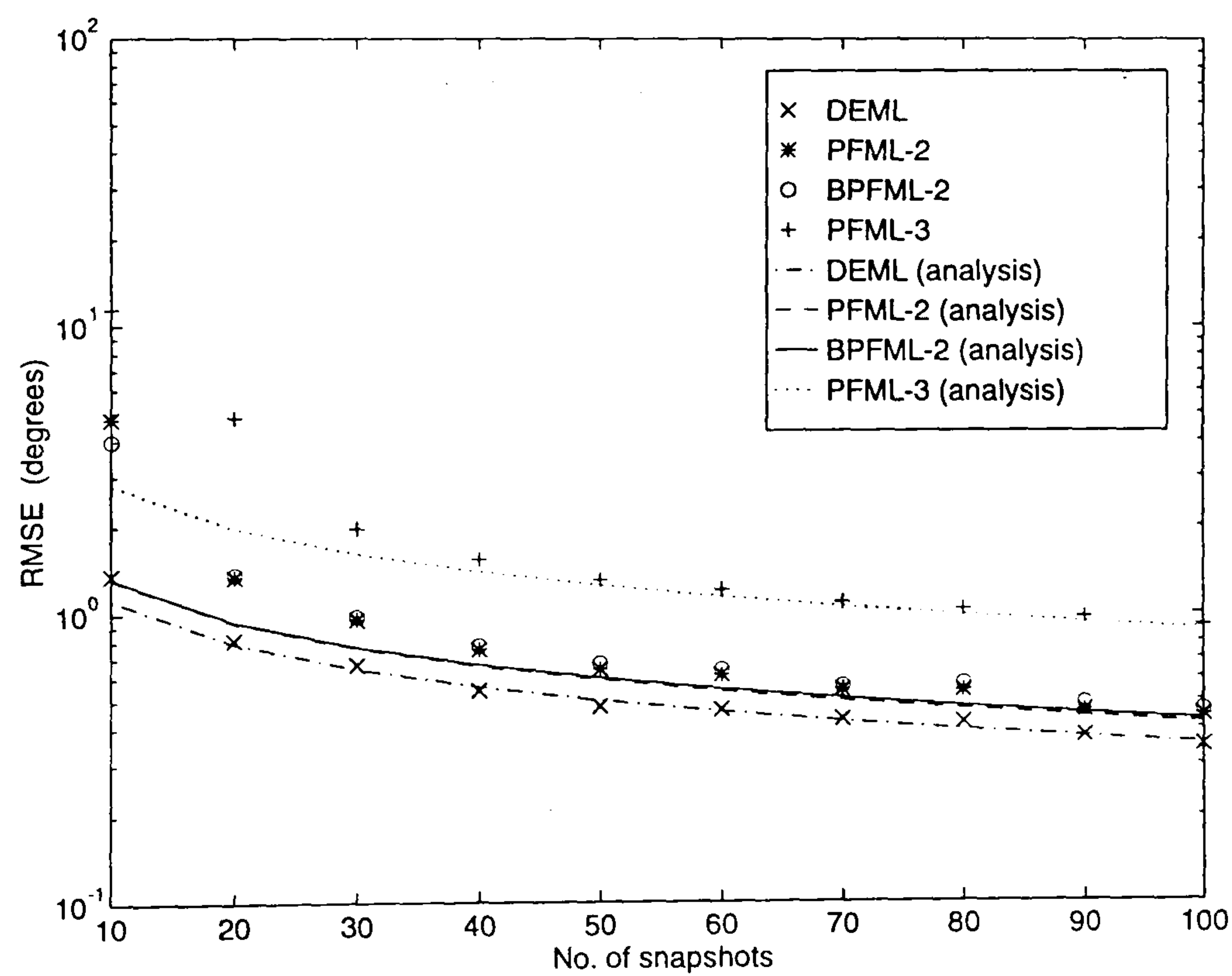
(a)

Figure 6.13: The estimated and asymptotic analytical RMSEs of two 1st-cluster DOA estimates obtained by four PFML methods under the signal scenario of EXP 6.4, at ten values of L with the SNR of 5 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.



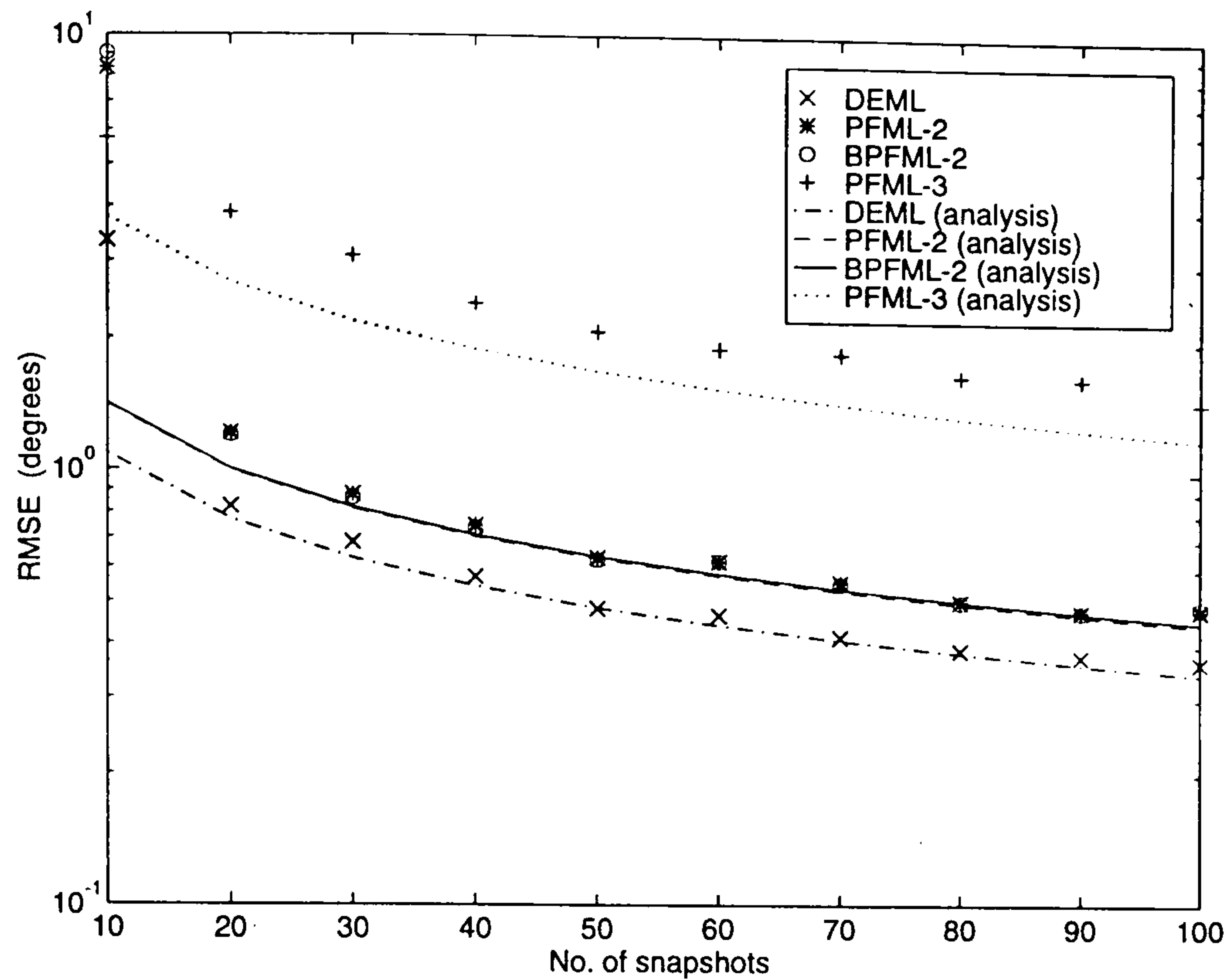
(b)

Figure 6.13: (Continued)



(a)

Figure 6.14: The estimated and asymptotic analytical RMSEs of two 1st-cluster DOA estimates obtained by four PFML methods under the signal scenario of EXP 6.4, at ten values of L with the SNR of 10 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.



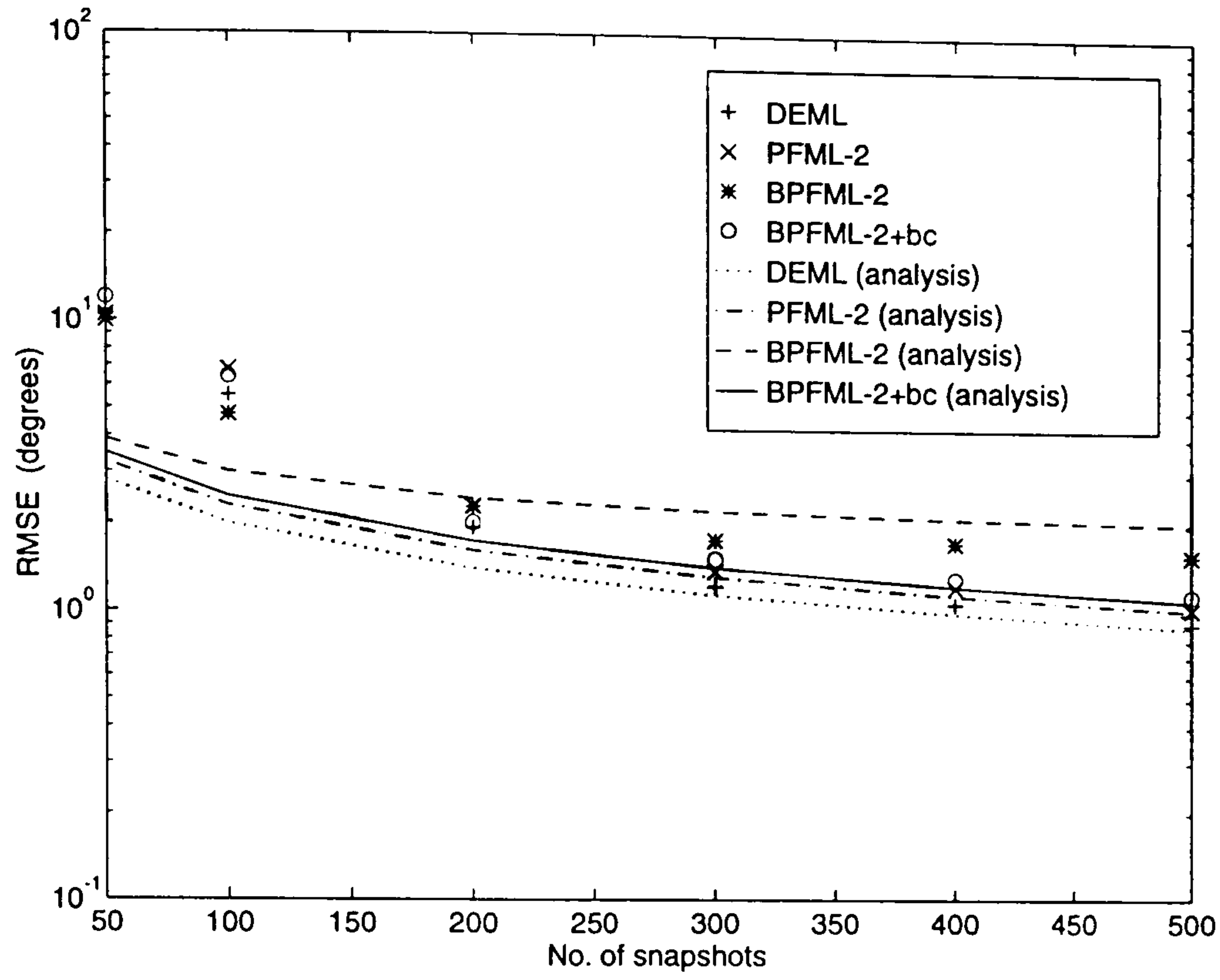
(b)

Figure 6.14: (Continued)

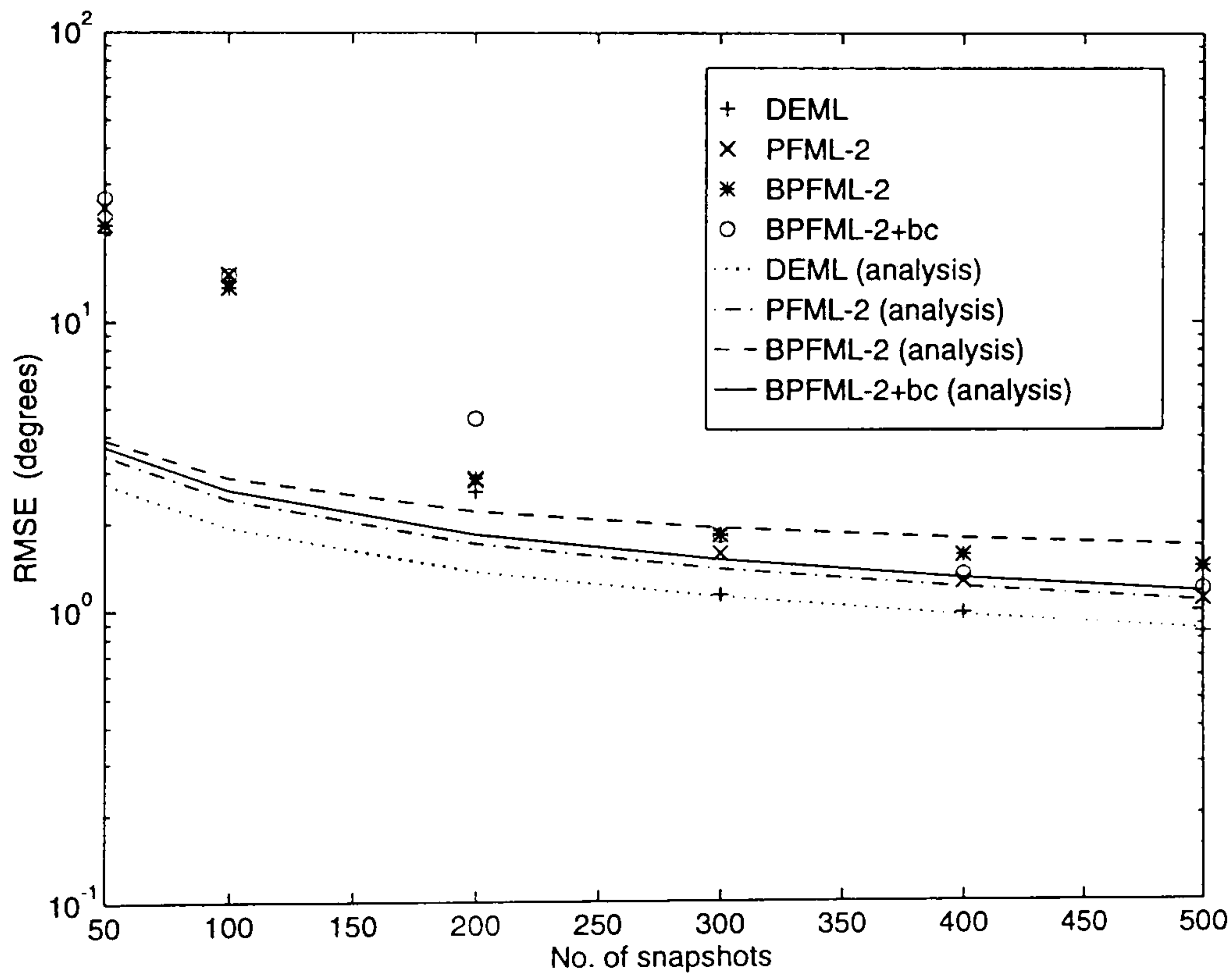
EXP 6.5: In this experiment, the biases and RMSEs of the DOA estimates using the two BPFML-2 methods, with and without bias correction, were examined at low SNR. The experimental results will also be compared with the corresponding asymptotic analytical RMSE and bias as given in Chapters 3 and 4. The experimental setup was based on the two-cluster model of four signal sources as described in EXP 6.3. The experiment was performed at SNRs of -5 and 0 dB, each with six different values of L .

Figures 6.15-6.16 show the estimated and asymptotic analytical RMSEs of the first two DOA estimates (20 and 15 degrees) using four PFML estimators, at two SNRs of -5 and 0 dB. From the figures, the estimated RMSEs of the PFML-2 estimates are seen to be close to those obtained by the corresponding asymptotic RMSE analysis for $L \geq 300$ at SNR = -5 dB, and $L \geq 200$ at SNR = 0 dB. At these ranges of L for the two SNRs, the RMSEs of the two DOA estimates using the BPFML-2+bc method are comparable to those of the PFML-2 method. We note that, for large L (≥ 300), the RMSEs of the BPFML-2 estimates did not decrease, but had an approximately constant

value. This is the result of the bias term in the filtered data used by the method. To see this, the estimation biases are shown in Figures 6.17 at the SNRs of -5 and 0 dB. From the figure, the biases of the two estimates using the BPFML-2 method remain unchanged as L increases beyond 200, while those due to the PFML-2 and BPFML-2+bc methods decrease. The biases of the estimates using the BPFML-2+bc are close to those due to the PFML-2 method. For the BPFML-2 method, with large L , the result from the bias analysis in Section 3.2.4 is seen to be in good agreement with the simulation counterpart at SNR = 0 dB. However, at the SNR of -5 dB, the difference between the estimated and analytical results is rather noticeable for the bias of the first DOA estimate. This is due to the violation at low SNR of the assumption used in the analysis that the bias term $[E_b]_1$ is sufficiently small. The bias term $[E_b]_1$ in the BPFML-2 method is noted to be the combined effect of the additive noise and the two 2nd-cluster signals (the 3rd and 4th signals). Nevertheless, at the two SNR values (-5 and 0 dB), the bias term due to the 2nd-cluster signals was found to be insignificant when compared with the one due to the additive noise (approximately 10 times and 6 times smaller in the 2-norm sense for, respectively, the SNRs of -5 and 0 dB). This explains why, for large L where the estimation error is due largely to $[E_b]_1$ rather than the finite data sample effect, the performance of the BPFML-2+bc method, of which the bias due to the additive noise is corrected, is close to that of the PFML-2 method, but different from that of the BPFML-2 method.

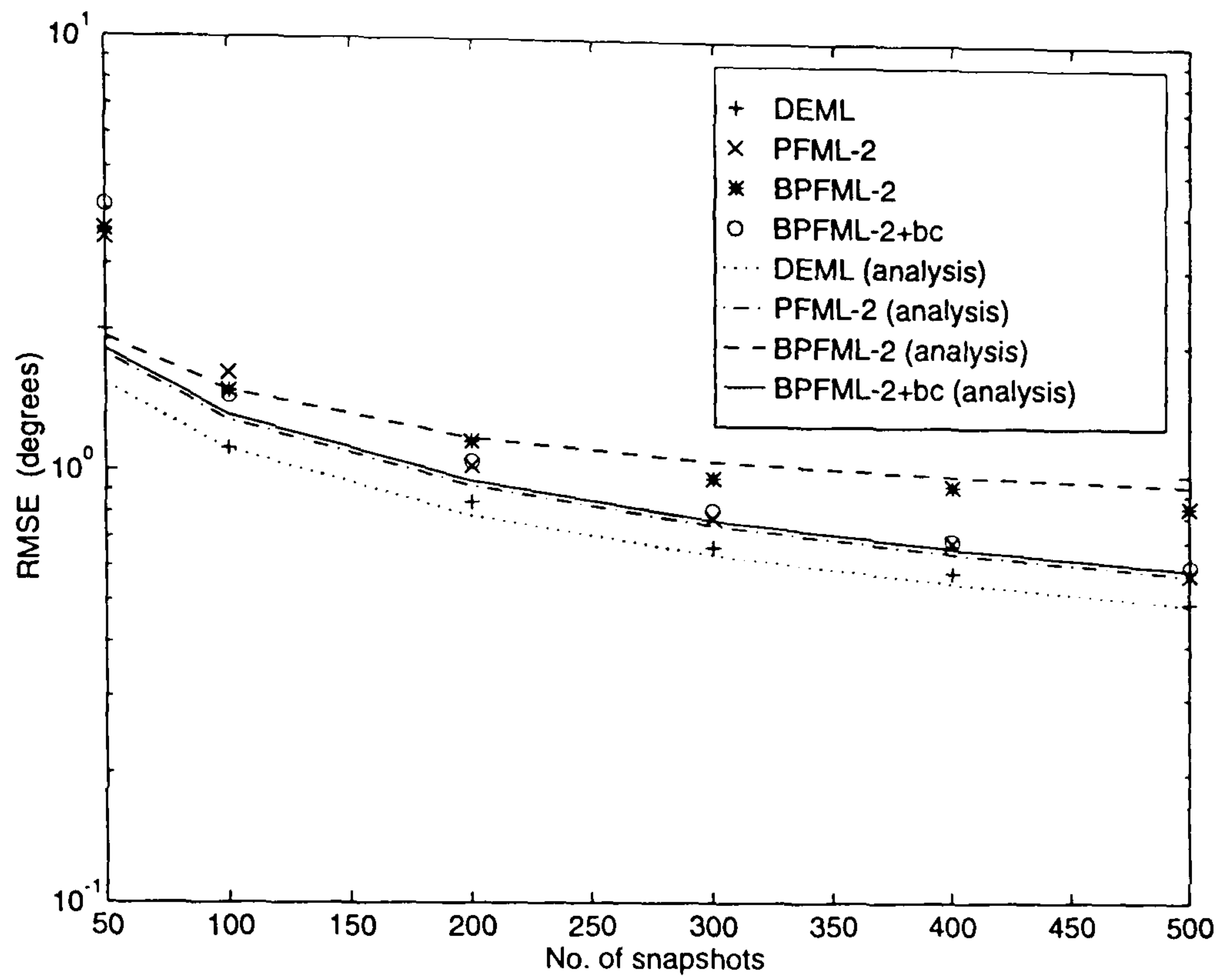


(a)

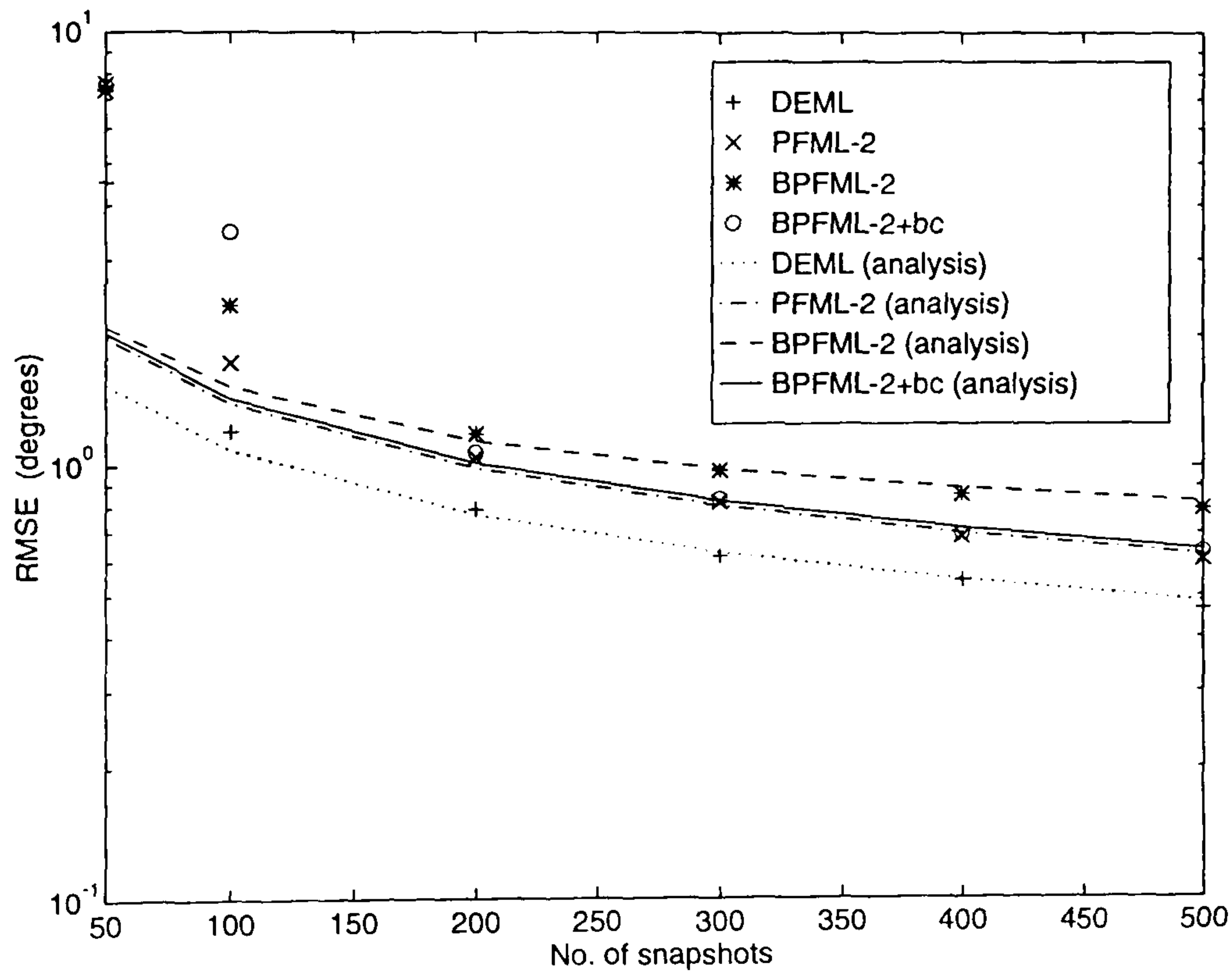


(b)

Figure 6.15: The estimated and asymptotic analytical RMSEs of two 1st-cluster DOA estimates obtained by the DEML, PFML-2, BPFML-2, and BPFML-2+bc methods, under the signal scenario of EXP 6.5, with six values of L and the SNR of -5 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.

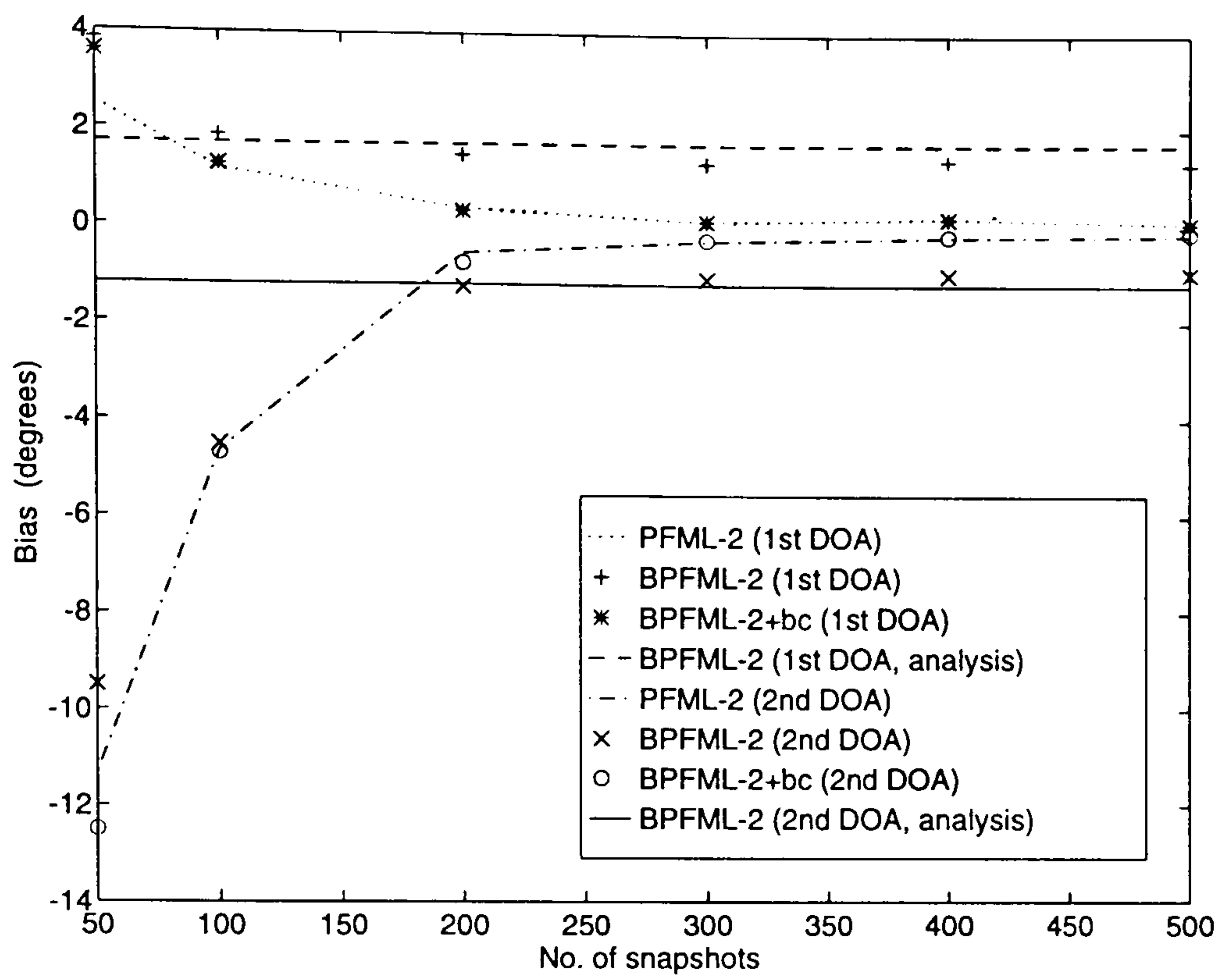


(a)

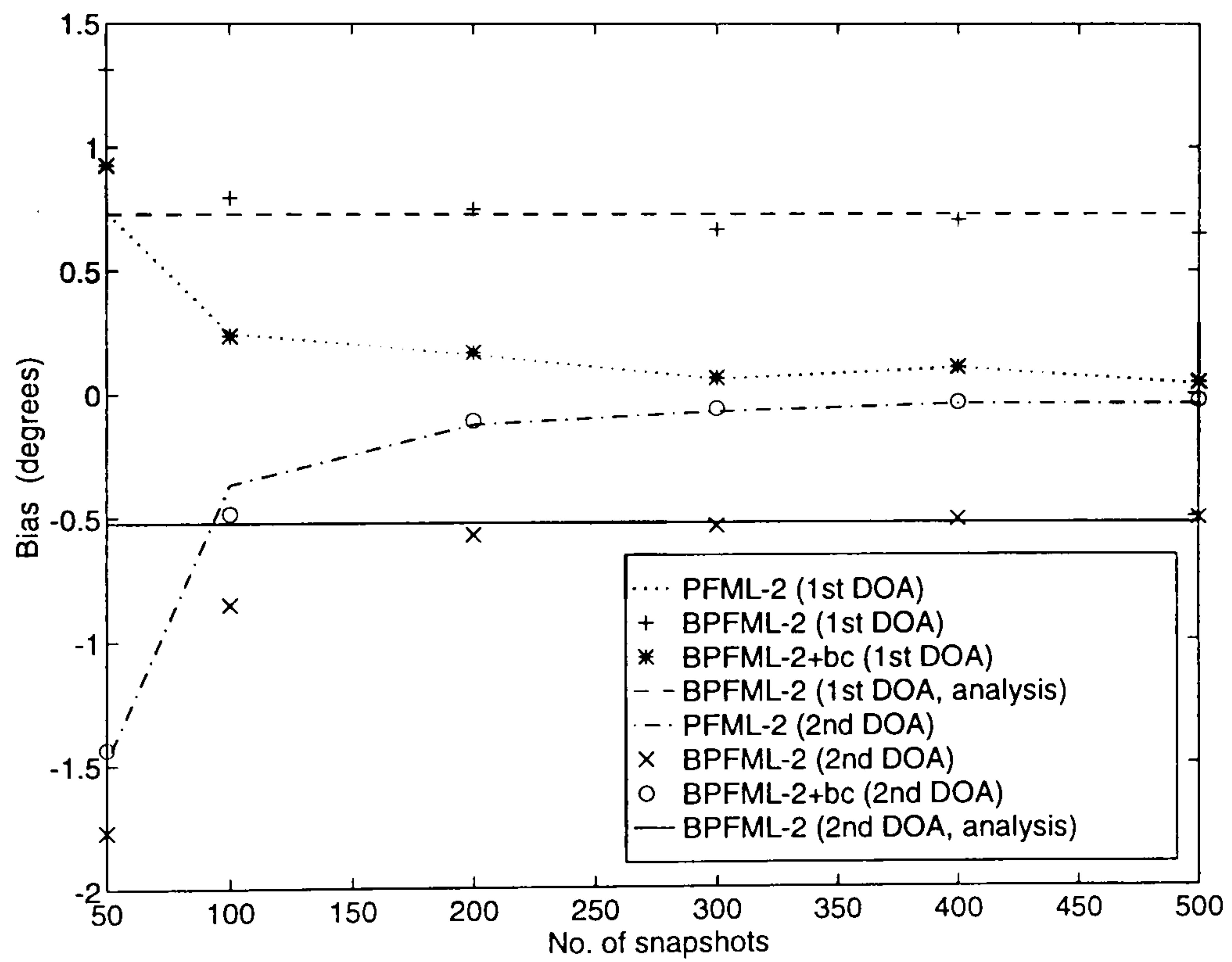


(b)

Figure 6.16: The estimated and asymptotic analytical RMSEs of two 1st-cluster DOA estimates obtained by the DEML, PFML-2, BPFML-2, and BPFML-2+bc methods, under the signal scenario of EXP 6.5, with six values of L and the SNR of 0 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.



(a)



(b)

Figure 6.17: The biases of the first and second DOA (20 and 15 degrees) estimates obtained by the PFML-2, BPFML-2, and BPFML-2+bc methods, under the signal scenario of EXP 6.5, with six different values of L . (a) SNR = -5 dB; (b) SNR = 0 dB.

EXP 6.6: The experiment was based on the two-cluster model of EXP 6.3. The main purpose is to investigate the estimation performance as a function of the correlation of the two signals in each cluster. The correlation of the first two signals, which form the first source cluster, and that of the last two signals in the second cluster, were varied simultaneously to obtain 9 different values of CORR, defined by

$$\text{CORR} \triangleq \left| \mathbb{E}\{s_i(t)s_k^*(t)\} \right| \cdot [\mathbb{E}\{s_i(t)s_i^*(t)\}\mathbb{E}\{s_k(t)s_k^*(t)\}]^{-1/2} \quad (6.5)$$

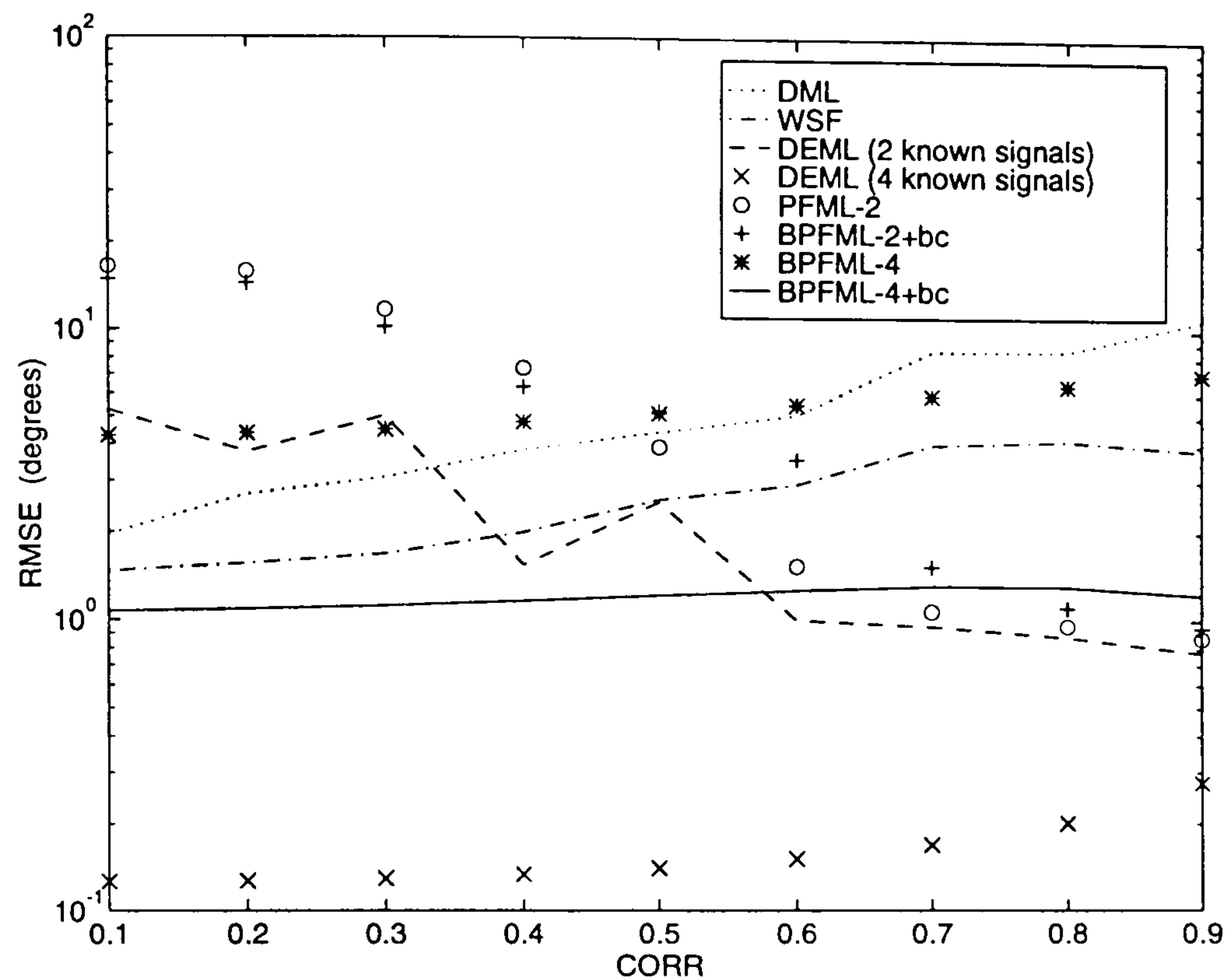
where $s_i(t)$ and $s_k(t)$ are the two correlated signals in each of the two clusters. The signal model can be written in the form described by Eqs. (6.1-6.3) with $s_{nc}(t) = \Gamma_{cs}^{-1} s(t)$, and

$$\Gamma_{cs} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \exp(j\pi/4) \cdot \text{CORR} & \sqrt{1 - (\text{CORR})^2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \exp(j\pi/4) \cdot \text{CORR} & \sqrt{1 - (\text{CORR})^2} \end{bmatrix} \quad (6.6)$$

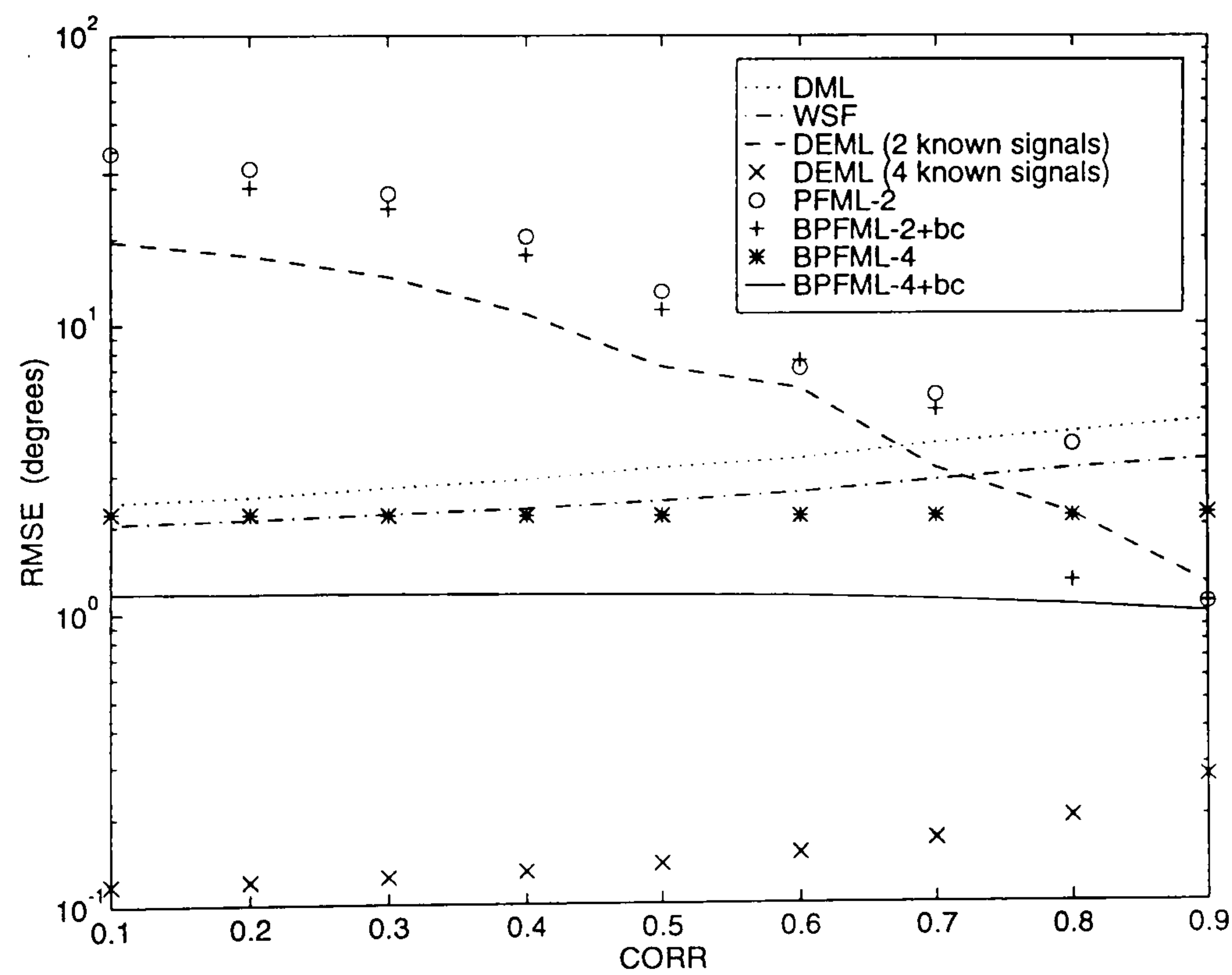
As in previous experiments, the first two source directions are of interest. Two DEML methods in this experiment were based on different known signal assumptions. The first DEML method (DEML with 2 known signals) assumed that the first and third signal waveforms were known. The second DEML method (DEML with 4 known signals) assumed that all four signal waveforms were known. The noise covariance structure assumed known by both methods was a scaled identity matrix. For the BPFML-4 and BPFML-4+bc methods, two beamformers were calculated from Eq. (4.19) by the improved beamforming procedure. The look directions of the two beamformers were obtained by deflecting the estimate of the first cluster nominal angular location by, respectively, +0.5 and -0.5 times the array beamwidth (approximately ± 6 degrees). In each beamformer, a single null constraint was imposed on the angular location which was the look direction of the other beamformer. Note that, as in EXP 6.3, two highest peak locations in the spatial spectrum $P_{ss}(\theta)$ were chosen as the

estimates of the two cluster nominal angular locations. Of the two peak locations, the one which was closer to the direction of 17 degrees was used as the location estimate of the first source cluster. The estimated RMSEs of the first two source DOA estimates by eight ML methods at the SNR of 5 dB are shown in Figure 6.18. From the figure, it is seen that as CORR decreases, the increase in the RMSEs of the DOA estimates is obtained for the BPFML-2+bc method, and the DEML method with 2 known signals. In contrast, the RMSEs of the estimates obtained by the DML method, the WSF method, and the DEML method with 4 known signals, increase as CORR increases. Unlike the methods already mentioned, the RMSEs of the estimates obtained by the BPFML-4 and BPFML-4+bc methods change moderately as a function of CORR. The marked improvement in estimation accuracy gained by the BPFML-4+bc method, compared with the BPFML-4 method, is clearly seen in the figure. In addition, only the DEML method with 4 known signals and the BPFML-4+bc method achieved superior performance to the DML and WSF methods for all of the nine values of CORR. The PFML-2 and BPFML-2+bc methods, as well as the DEML method with 2 known signal waveforms, began to give a better result than the WSF method at $\text{CORR} > 0.7$. For $\text{CORR} < 0.7$, however, the performance of these methods is clearly inferior to those of the DML and WSF methods. Figure 6.19 shows the results for the SNR of 10 dB. It is seen from the figure that the BPFML-4+bc still yielded a better result than the other methods except for the DEML method with 4 known signal waveforms. However, the performance difference between the BPFML-4+bc method and the DML and WSF methods is less than in the case of the SNR of 5 dB. Figure 6.20 shows the ensemble average of the array output spatial spectrum, and the two array beam patterns corresponding to the asymptotic beamforming weight vectors of the beamformers used by the BPFML-4 and BPFML-4+bc methods. It should be noted from the figure that, the nulling locations in both beam patterns are approximately 2-3 degrees away from the first-cluster source directions of 20 and 15 degrees. This suggests that only approxi-

mations of source directions are required by the BPFML-4+bc method to obtain improved estimation performance over the BPFML-2, DML, and WSF methods.

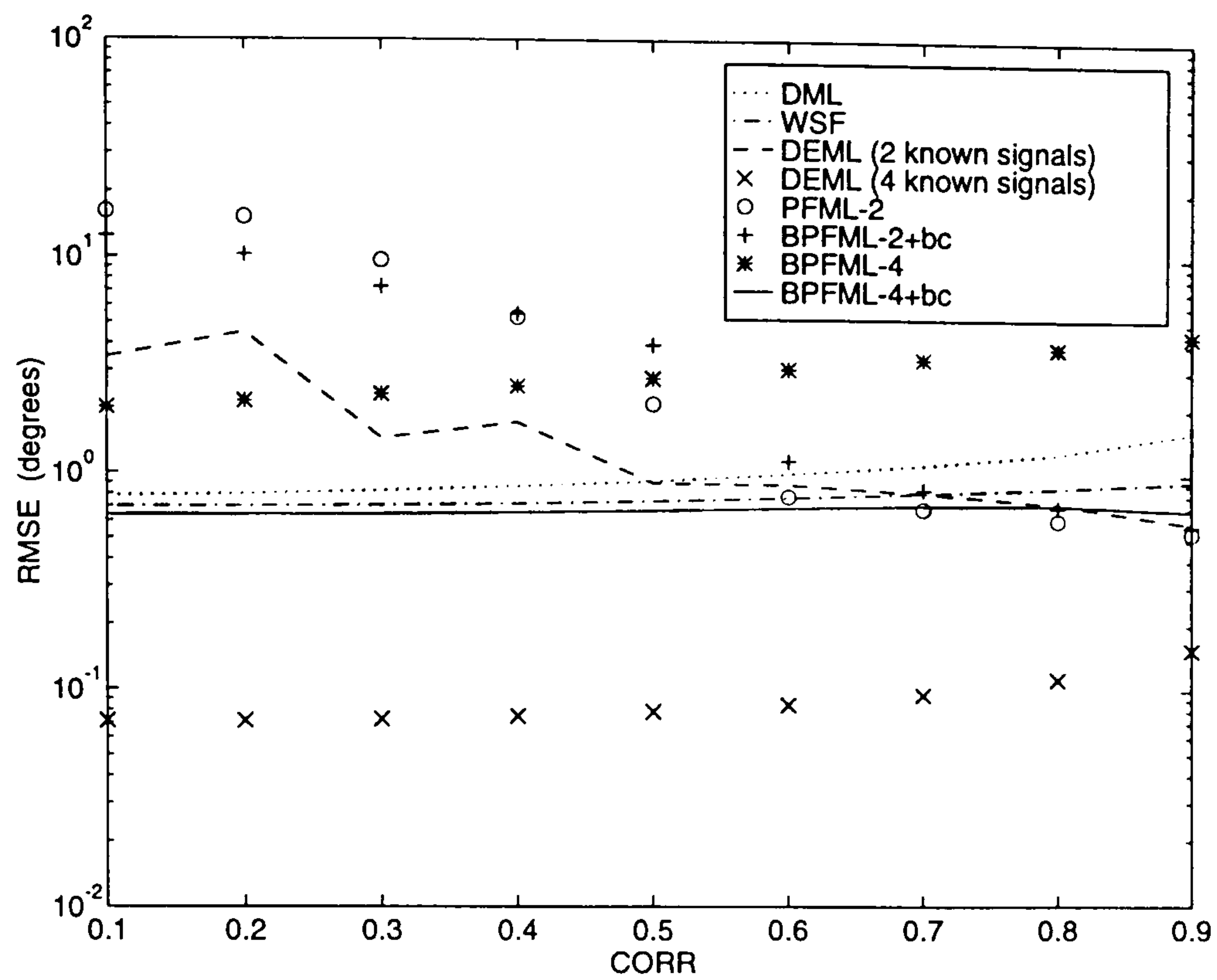


(a)

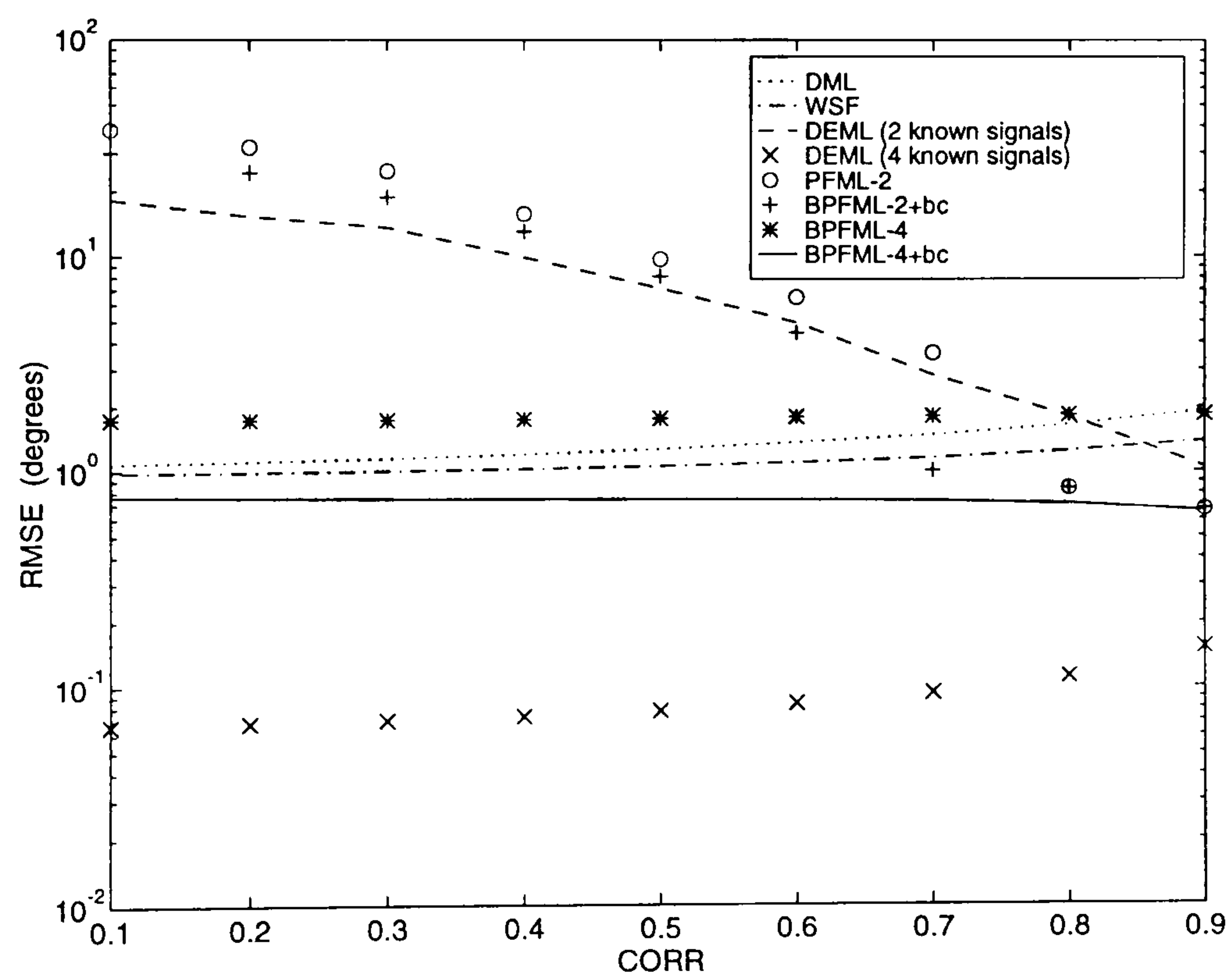


(b)

Figure 6.18: The estimated RMSEs of two 1st-cluster DOA estimates obtained by eight ML-based methods, under the signal scenario of EXP 6.6, at nine different values of CORR and the SNR of 5 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.



(a)



(b)

Figure 6.19: The estimated RMSEs of two 1st-cluster DOA estimates obtained by eight ML-based methods, under the signal scenario of EXP 6.6, at nine different values of CORR and the SNR of 10 dB. (a) DOA = 20 degrees; (b) DOA = 15 degrees.

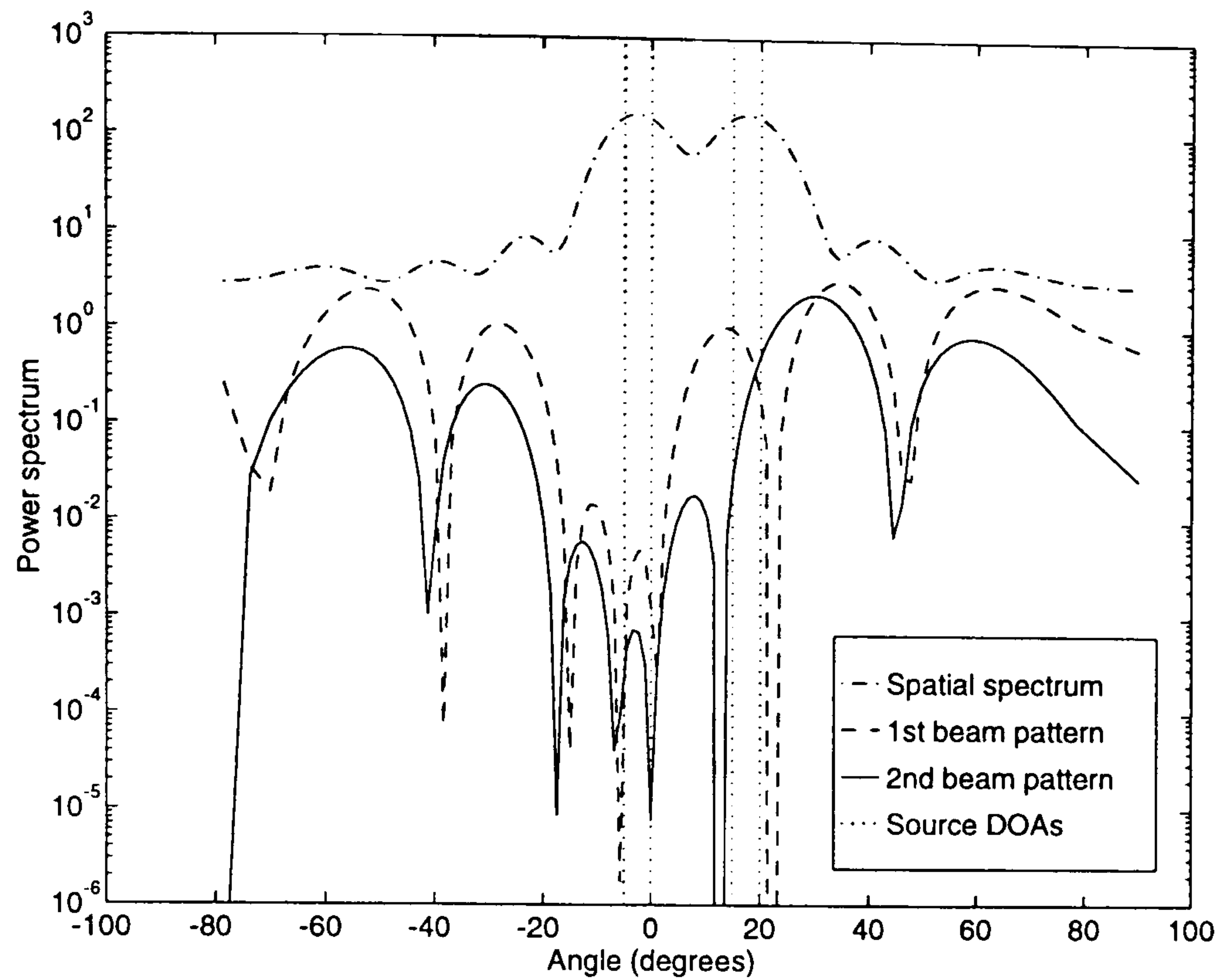


Figure 6.20: The ensemble average of the array output spatial spectrum, and two asymptotic beam patterns corresponding to the beamformers used by the BPFML-4 and BPFML-4+bc methods in EXP 6.6 at the SNR of 10 dB.

EXP 6.7: A scenario where the number of sources is more than the number of sensors was considered in this experiment. Ten complex Gaussian distributed signals, each with unity power, impinged upon the array from distinct directions as listed in Table 6.1. They were grouped into five clusters, each consisting of two coherent signals. Any two signals from different clusters were uncorrelated. Let the first and second signals belong to the first cluster, the third and fourth signals belong to the second cluster, and so on. The phase difference between any two coherent signals was $\pi/4$. The described scenario can be written by the model Eq. (6.1), with $s_{\text{nc}}(t) = [s_1(t) \ s_3(t) \ s_5(t) \ s_7(t) \ s_9(t)]^T$, and

$$\Gamma_{\text{cs}} = \begin{bmatrix} \gamma_{1,1} & & & & 0 \\ & \gamma_{2,2} & & & \\ & & \gamma_{3,3} & & \\ & & & \gamma_{4,4} & \\ 0 & & & & \gamma_{5,5} \end{bmatrix} \quad (6.7)$$

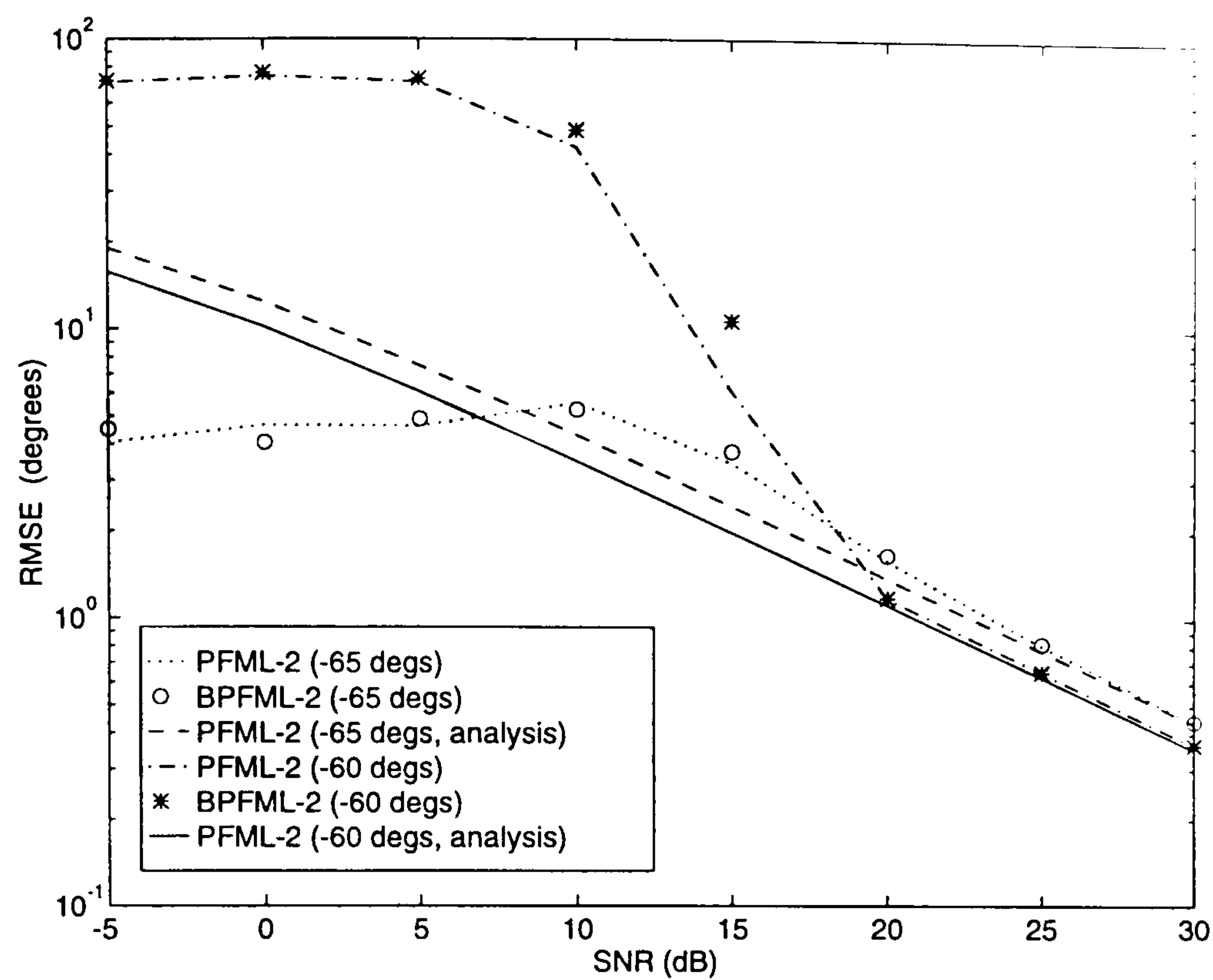
where

$$\gamma_{k,k} = [1 \quad \exp(j\pi/4)]^T, \quad k = 1, \dots, 5 \quad (6.8)$$

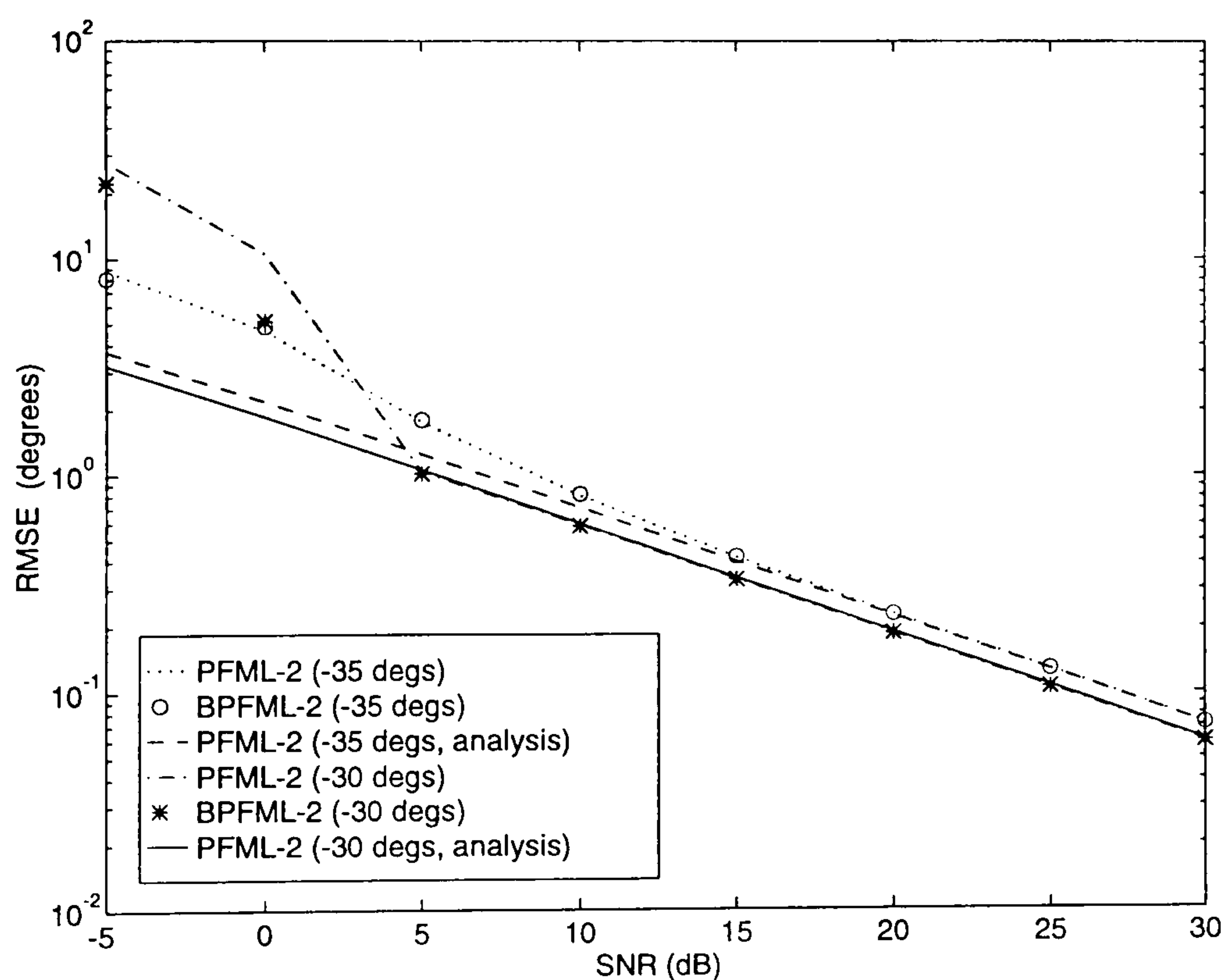
The estimated RMSEs of ten DOA estimates obtained by the PFML-2 and BPFML-2 methods are shown in coherent pairs by Figure 6.21. In the figure, the simulation results are compared with the asymptotic analytical RMSEs of the PFML-2 estimates. Note that, these theoretical RMSEs are also the CRLBs on the standard deviations of the estimates, under the signal and noise condition assumed by the PFML-2 method. From the figure, it is seen that for sufficiently high SNR, the two methods accurately estimated the ten signal directions. For $\text{SNR} \geq 25$ dB, the estimated RMSEs of all DOA estimates using either of the two methods attain the asymptotic analytical RMSEs. Notice from these figures that, for both methods, the estimate of the source direction which is closer to the array broadside, is more accurate than that of the direction further from broadside. This is to be expected from the resolution property of a uniform linear array. Except for the source directions of -65, -60, 55 and 60 degrees, all other source directions were reliably estimated for SNR as low as 5 dB. Note also that, the estimates using the BPFML-2+bc method (not shown here) were found to yield comparable estimated RMSEs to those of the BPFML-2 method, although the estimates obtained by the BPFML-2+bc method were slightly less biased at low SNR.

Signal no.	1	2	3	4	5	6	7	8	9	10
DOA (degrees)	-65	-60	-35	-30	-5	0	20	25	55	60

Table 6.1: The directions of arrival (in degrees) of ten signals, with respect to the array broadside.

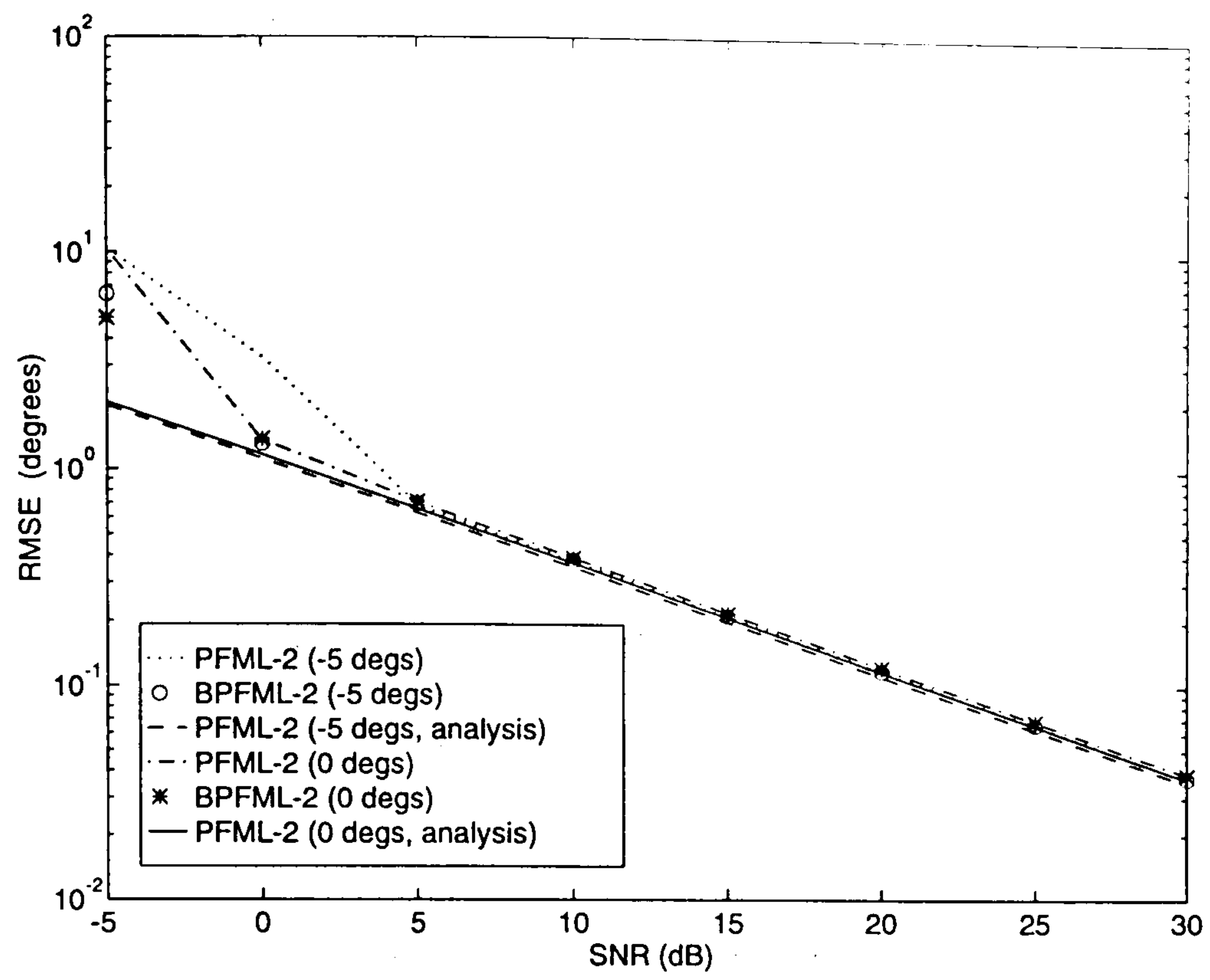


(a)

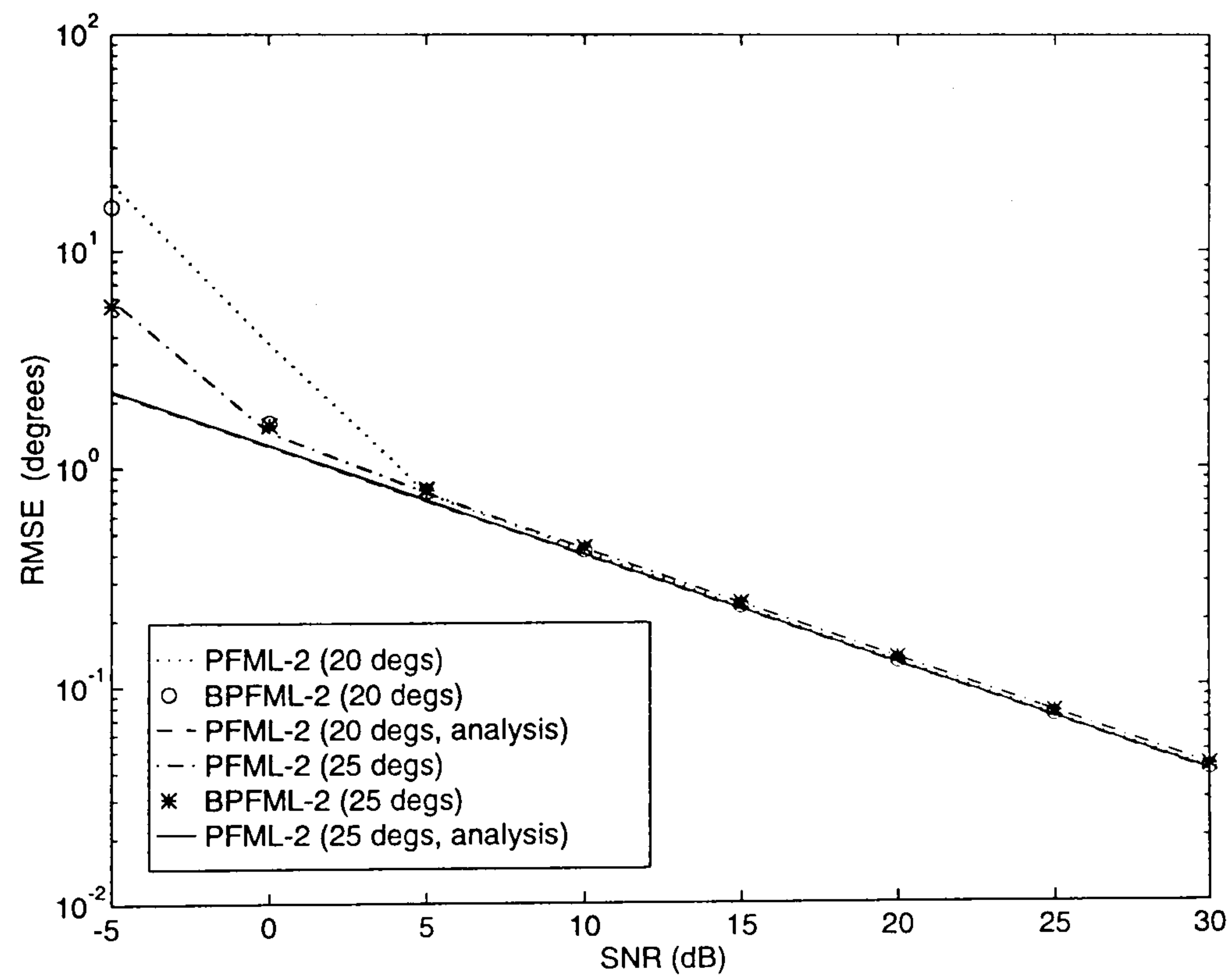


(b)

Figure 6.21: The estimated RMSEs of ten DOA estimates using the PFML-2 and BPFML-2 methods, under the ten-source scenario of EXP 6.7, compared with the analytical RMSEs of the PFML-2 estimates at eight SNRs. (a) DOAs = -65 and -60 degrees; (b) DOAs = -35 and -30 degrees; (c) DOAs = -5 and 0 degrees; (d) DOAs = 20 and 25 degrees; (e) DOAs = 55 and 60 degrees.

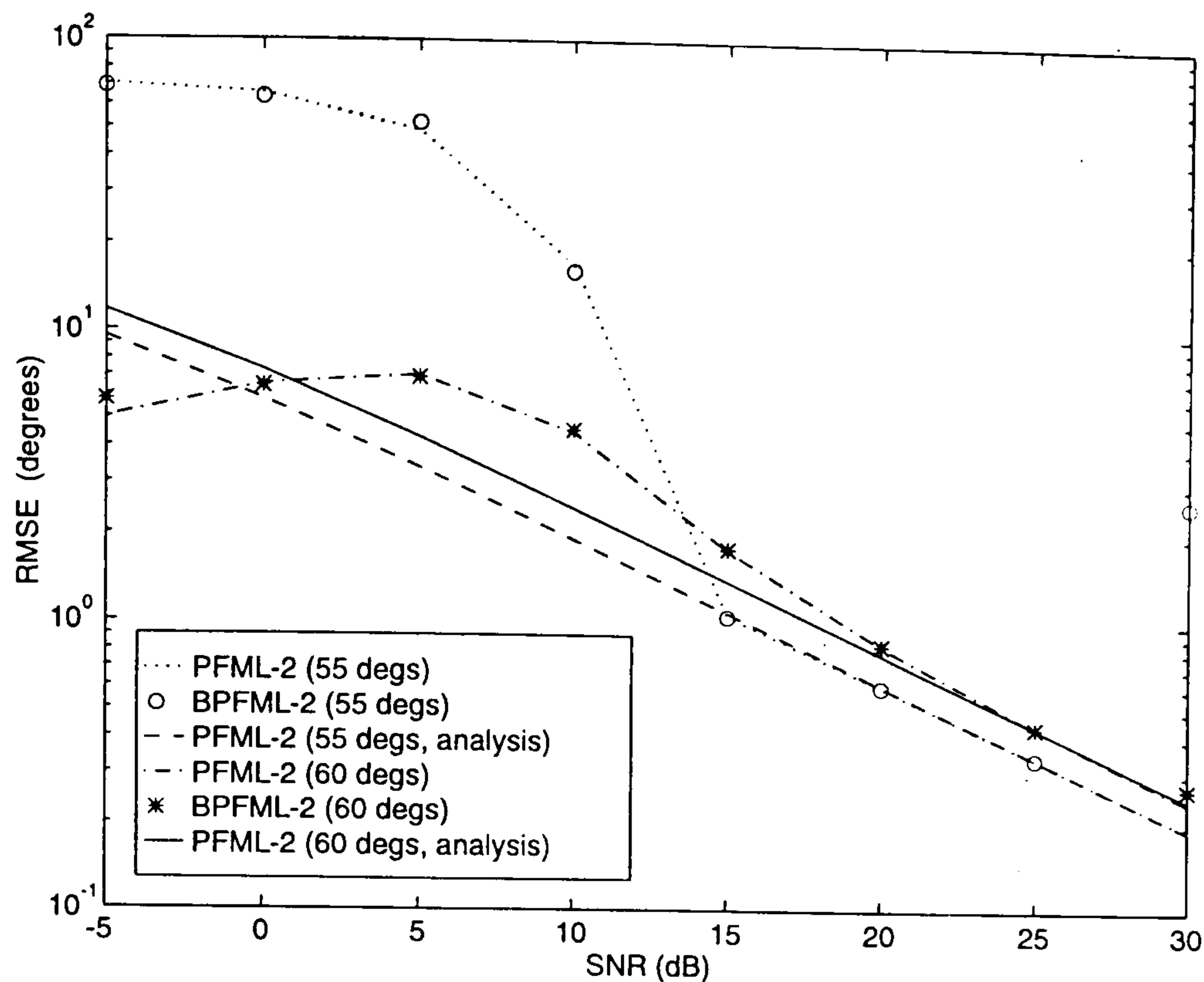


(c)



(d)

Figure 6.21: (Continued)



(e)

Figure 6.21: (Continued)

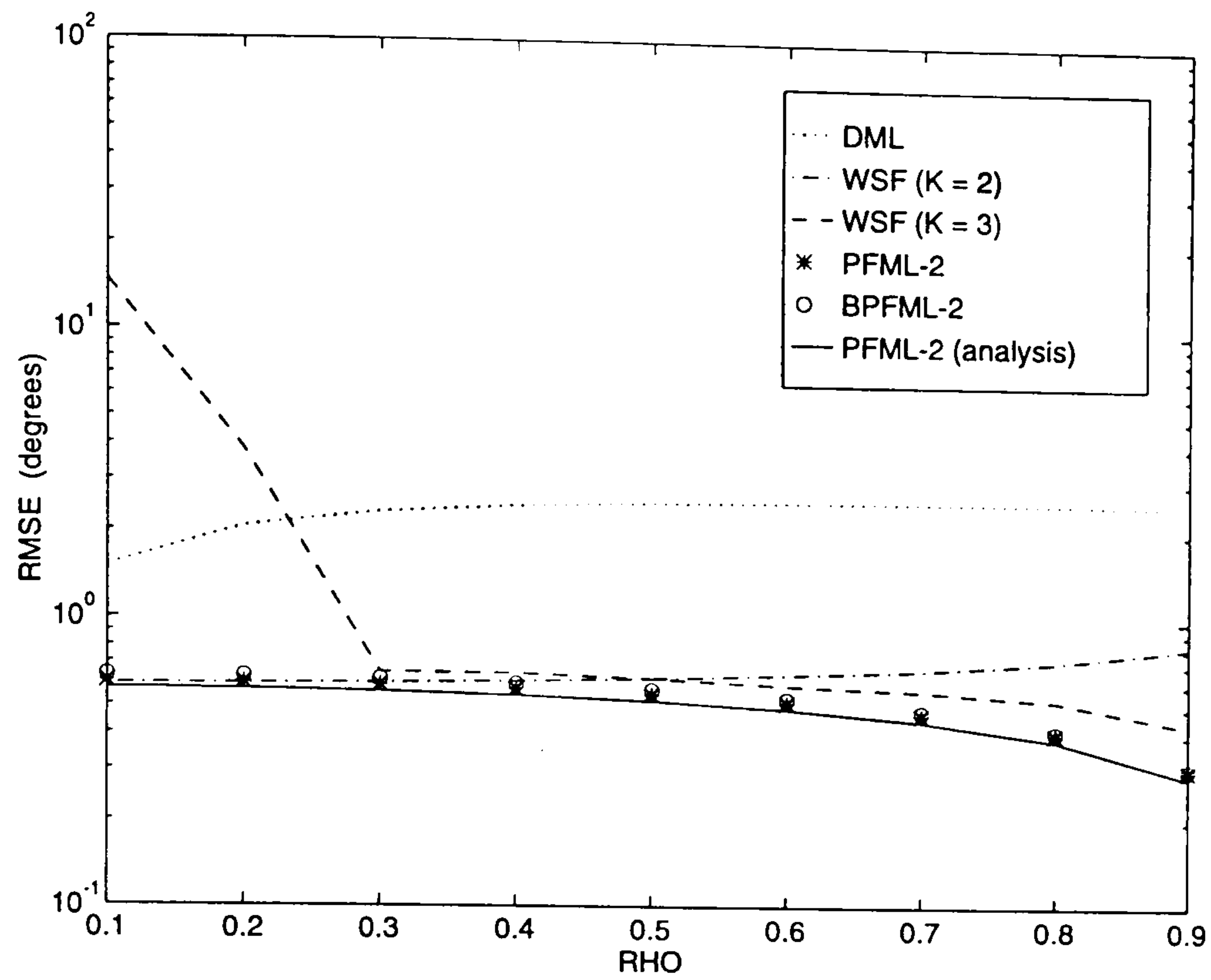
EXP 6.8: In this experiment, the case of an unknown spatially correlated noise field was considered. Two coherent signal sources, each with unity power, arrived from the directions of 5 and 10 degrees. The phase difference between the two signals was $\pi/4$. In the experiment, the noise covariance matrix was given by

$$[R_n]_{i,k} = \sigma^2 \cdot \{ \rho^{|i-k|} \exp(-j\pi(i-k)\sin(\theta_n)) \} \quad (6.9)$$

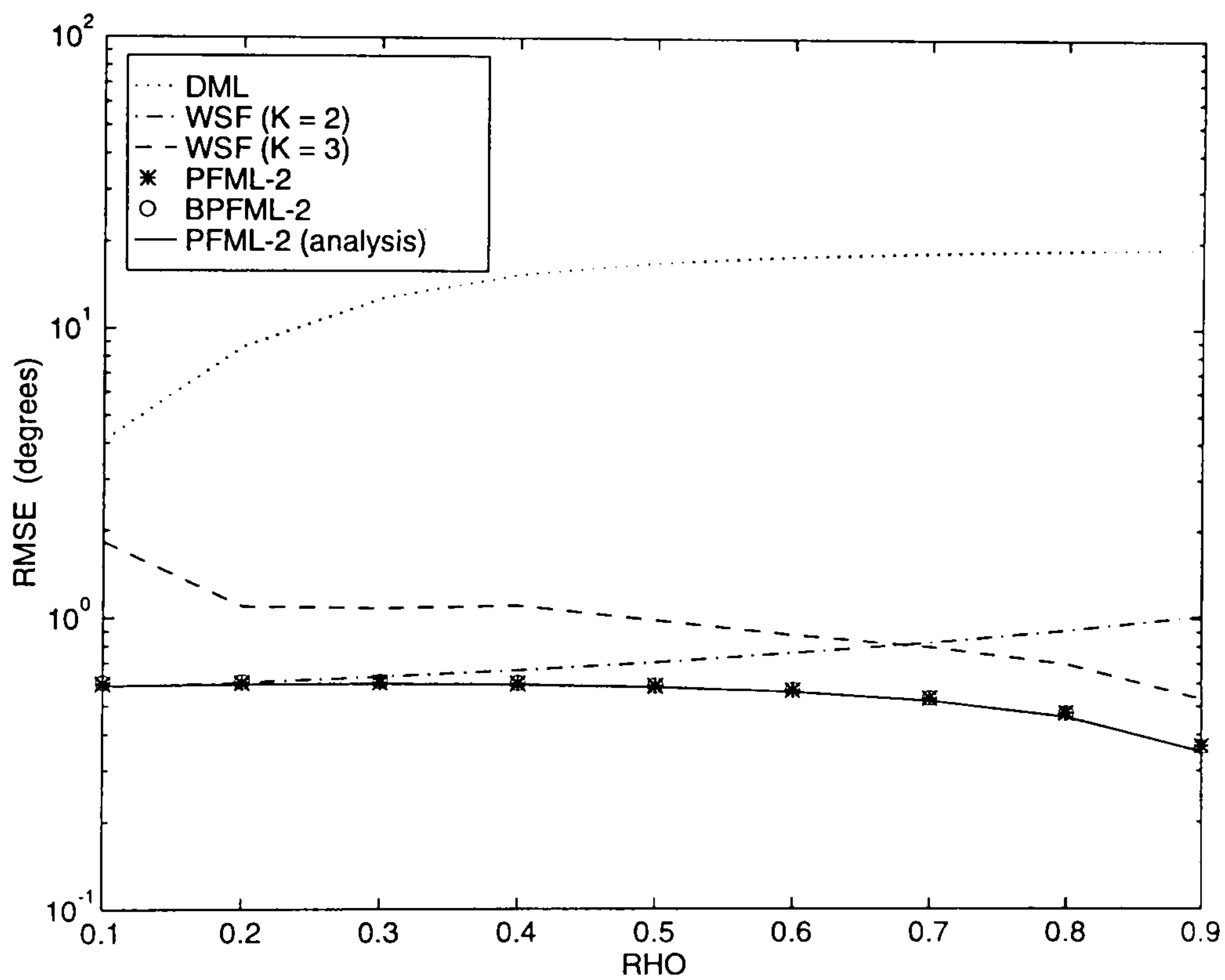
where σ^2 is the noise power at each sensor. The parameter ρ , whose value lies between 0 and 1, controls the degree of spatial whiteness. For $\rho \approx 0$, the additive noise is approximately spatially white. The parameter θ_n is the angular location where the noise spatial spectrum is at its peak.

Figure 6.22 shows the estimated RMSEs of the two DOA estimates using five estimators at nine different values of ρ , with $\theta_n = 30$ degrees and SNR = 5 dB. The DML method and the two optimal WSF methods were performed under the (wrong) assumption that the additive noise is spatially white. The WSF method with $K = 2$ was

applied under the correct number of the signal sources. The second WSF method was performed with $K = 3$. In the latter method, the peak in the noise spatial spectrum was modelled as the third signal source. From the figures, poor performance can be seen for the DML method. The performance of the WSF method with $K = 2$ deteriorates as ρ increases. Improved performance at high ρ is observed for the WSF method with $K = 3$. However, at low ρ , the performance of the WSF method with $K = 3$ is inferior to that of the method with $K = 2$. The performance of the PFML-2 and BPFML-2 methods is seen to be comparable to that of the WSF method with $K = 2$ at low ρ . As ρ increases, the PFML-2 and BPFML-2 methods have superior performance to those of the two WSF methods. Note that, the asymptotic analytical RMSEs of the PFML-2 estimates, shown in Figures 6.22, are also the CRLBs on the standard deviations of the DOA estimates under the same signal and noise condition assumed by the PFML-2 method. The noise spatial spectrum is shown in Figure 6.23 for three different values of ρ . The ensemble average of the array output spatial spectrum and the asymptotic beam pattern of the beamformer used by the BPFML-2 method, for three different values of ρ , are shown, respectively, in Figures 6.24 and 6.25.



(a)



(b)

Figure 6.22: The estimated RMSEs of two coherent-source DOA estimates obtained by four ML-based methods, under the unknown correlated noise scenario of EXP 6.8, shown with the analytical RMSEs of the PFML-2 estimates at nine different values of ρ (RHO) and the SNR of 5 dB. (a) DOA = 5 degrees; (b) DOA = 10 degrees.

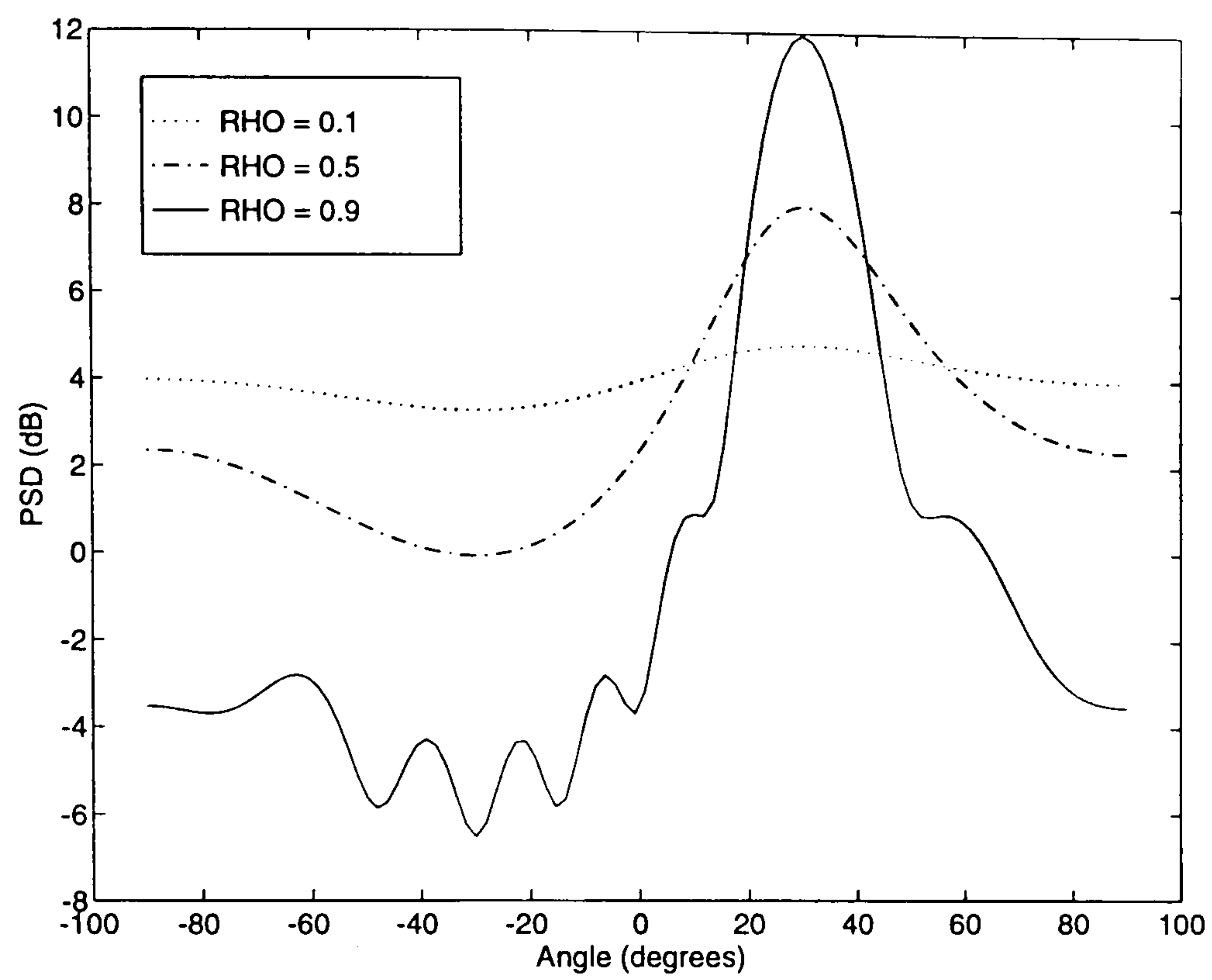


Figure 6.23: The noise spatial spectrum at three different values of ρ (RHO).

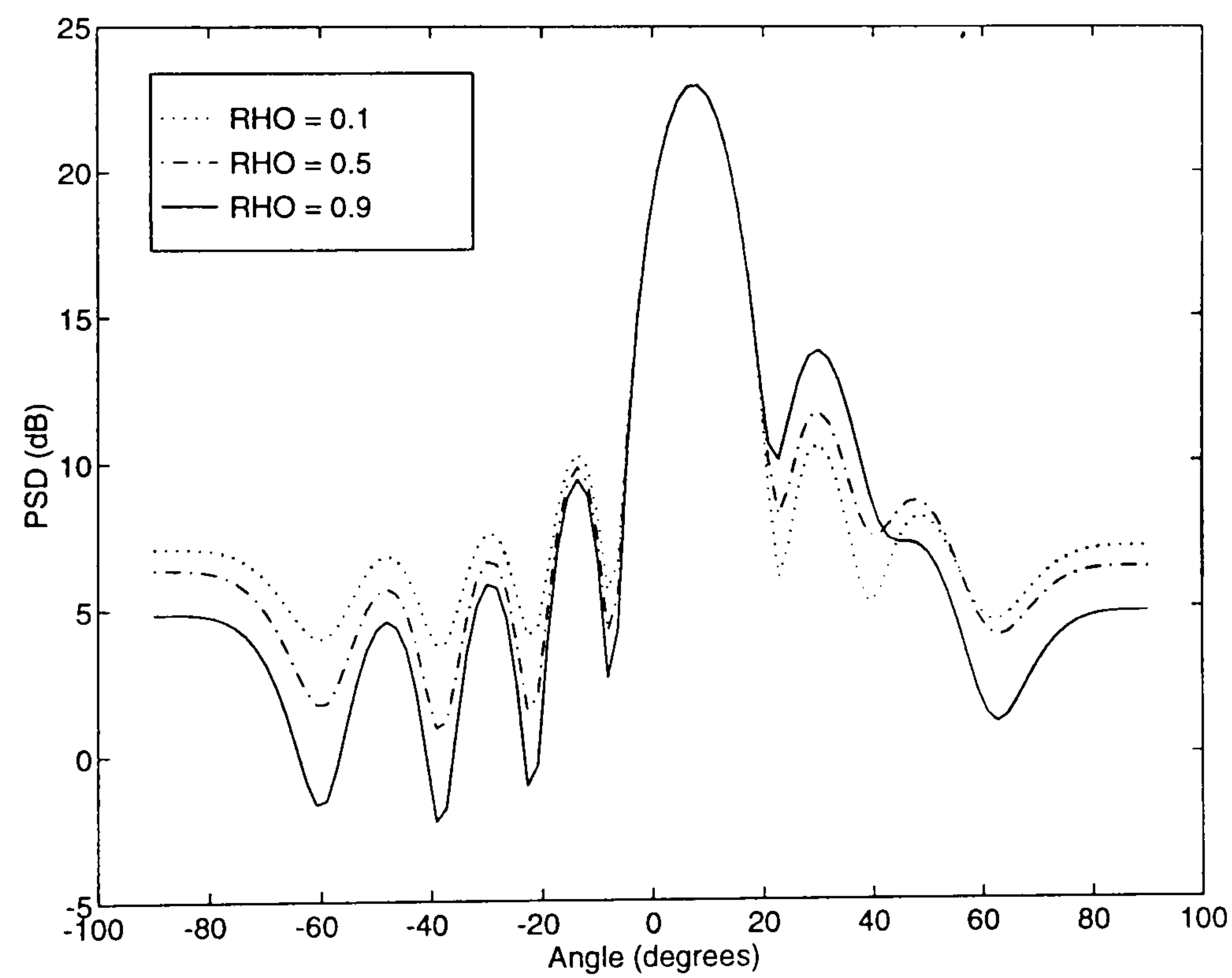


Figure 6.24: The ensemble average of the array output spatial spectrum at three different values of ρ (RHO).

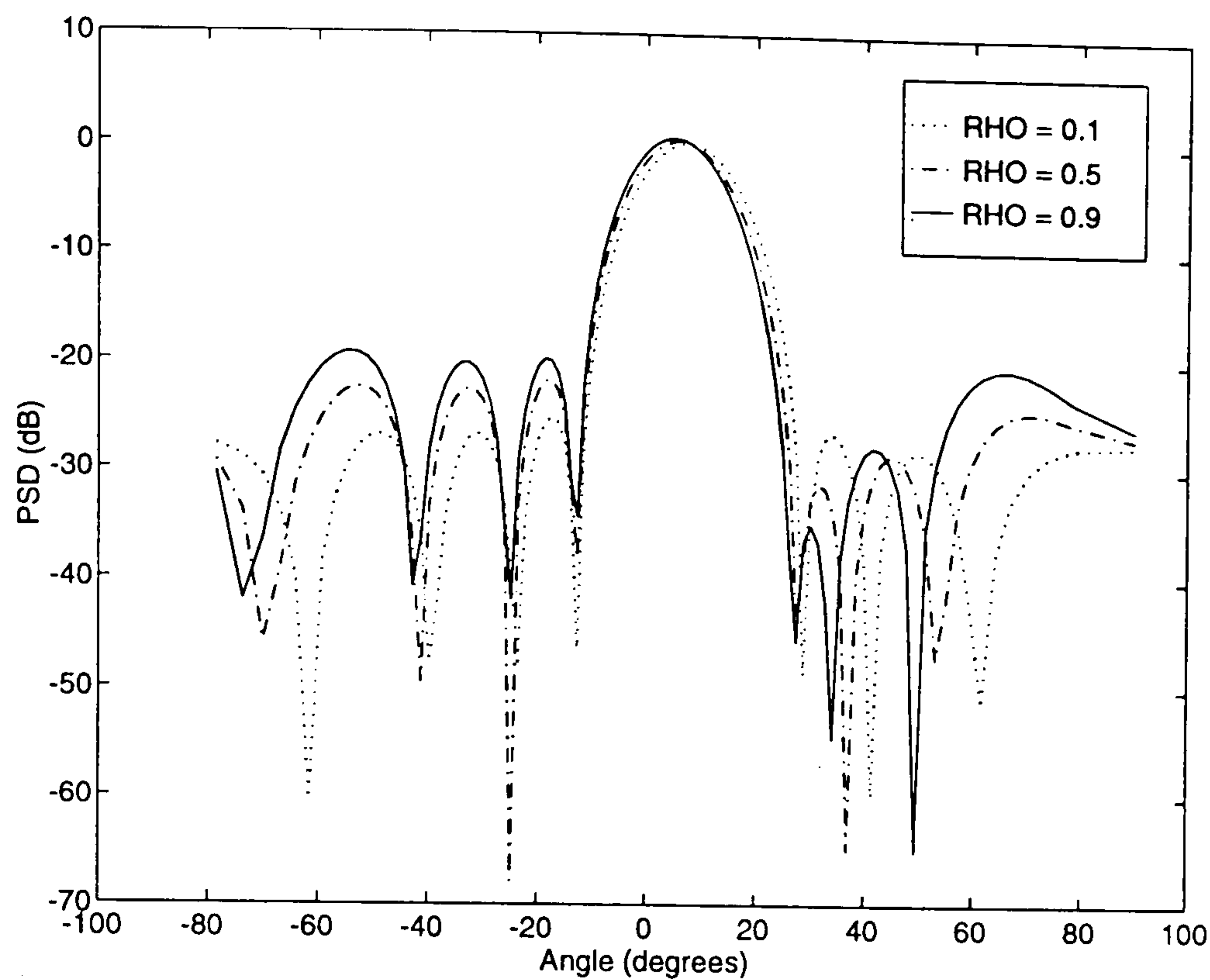


Figure 6.25: The asymptotic beam pattern of the beamformer used by the BPFML-2 method, at three different values of ρ (RHO).

6.2 Summary of the results and further discussion

From the performance study in the last section, which includes computer simulation as well as the evaluation of the asymptotic analytical formulas derived in Chapters 3 and 4, we summarise some main results from the study, and provide additional discussion as follows:

1. For the single source cluster scenario, considered in EXP 6.1, the BPFML-2 method (with or without bias correction) achieved comparable performance to the best-case PFML-2 method (with the type-1 pre-filtering reference signal) and the (optimal) WSF method.
2. For the two cluster source scenario with two coherent signals in each cluster, as considered in EXP 6.2-6.5, the BPFML-2 method, with or without bias correction, offered performance comparable to the best-case PFML-2 method. Its performance is superior to the RMNORM, DML, and WSF methods. As the angular separation between

the two clusters increases, the difference in performance between the BPFML-2 method and the DEML and WSF methods is reduced. For small angular separation between the two clusters, the BPFML-2 method can sometimes exhibit unusually large estimation error. This is due to the unresolvable situation in some simulations by the classical beamforming spectral estimator used in the beamforming procedure of Section 4.2.2. This problem was not found in the BPFML-2a method. Note that, it may be advantageous in this closely spaced cluster situation if a suitable non-uniform weighting function $W(\theta)$ is applied in Eqs. (4.10, 4.12) instead of the uniform weighting function used here.

3. The RMSEs of the DOA estimates obtained by the BPFML-2 method was found to be essentially the same as those due to the BPFML-2+bc method in most experiments. It was observed, however, that the BPFML-2+bc method yielded less biased estimates. In addition, for the case of low SNR with sufficiently large data snapshots, the estimates obtained by the BPFML-2+bc method were found to be more accurate and less biased than the ones obtained by the method without bias correction. Moreover, as seen in EXP 6.6, the use of bias correction was found to be essential in the BPFML-4 method. In EXP 6.6, it was found that the BPFML-4+bc method has superior performance to the BPFML-4 method. The reason for this performance difference is that the two beamforming weight vectors used in the BPFML-4 method (with or without bias correction) generally have larger vector norms than that of the weight vector used in the BPFML-2 or BPFML-3 method. Thus, even at low noise situation, the part of the bias term which is due to the additive noise ($= \sigma^2 \mathbf{w}$ in this white noise case) can be large (in the 2-norm sense). As a result, by correcting the bias term due to the additive noise, the performance of the BPFML-4 method can be greatly improved.

4. The BPFML-3 method was found to offer performance comparable to that of the optimal WSF method. This provides an alternative to the WSF method for the highly correlated source scenario. In addition, when the classical beamforming spatial spectrum

$P_{ss}(\theta)$ (see Eq. (4.9)) is already available or can be effectively computed (for instance, by using the FFT algorithm [82] for the case of a uniform linear array), the BPFML-3 method is more computationally attractive because, unlike in the WSF method, the eigendecomposition of the sample data covariance matrix is not required.

5. The BPFML-2 method (with or without bias correction), as well as the PFML-2 method and the DEML method with incomplete knowledge of all signal waveforms, was found to suffer severe performance degradation as the correlation between two closely spaced signals decreases, as demonstrated in EXP 6.6. The solution to this problem is to use the BPFML-4+bc method instead. The performance of the BPFML-4+bc method was found to be superior to the DML and WSF methods. For highly correlated sources, its performance is seen from the experiment to be comparable to those of the PFML-2 method and the BPFML-2 method (with or without bias correction).

6. In EXP 6.7, it was demonstrated that by using either the BPFML-2 method (with or without bias correction) or the PFML-2 method, the number of source directions can be greater than the number of array sensors (ten source directions using eight sensors, in the experiment). The DOA estimates using either of the two methods were found to be virtually unbiased for sufficiently high SNR. It should be noted, however, that for the BPFML-2 method to be applicable, it is important that the angular separation of any two clusters is sufficiently large to be resolvable by the classical beamforming spectral estimator.

7. In EXP 6.8, the problem of DOA estimation in an unknown spatially correlated noise field was considered. From the experiment, without prior knowledge of the noise covariance structure, the performance of the PFML-2 and BPFML-2 estimators for highly spatially correlated noise was found to be superior to those of the DML and WSF methods. However, the BPFML-2 method should be applied in this correlated noise case

only if the noise spatial spectrum is sufficiently smoothed in the vicinity of the source directions. The BPFML-2 method is expected to yield a poor result if the noise spectral peak location θ_n lies close to the source directions.

8. In all experiments, the DEML method with complete knowledge of all signal waveforms was found to achieve the lowest RMSEs and biases of the DOA estimates. When the complete knowledge of all signal waveforms is not available, however, the method can perform badly, as observed in EXP 6.6. A scenario where the complete knowledge of the signal waveforms may not be available is in the case of two (or more) multipath signals. Either when the delay time between the two signals is too long, or the bandwidth of the signals is too large, the two signals are only partially correlated. Thus, knowing only one of the two signal waveforms provides only partial knowledge of the two signals. Incomplete knowledge of the signal waveforms can also arise even in the single source case. If the known signal and the impinging signal are not well synchronised, the two signals may not be fully (or not even highly) correlated.

9. In the experiments, the estimated RMSEs by computer simulation have been compared with the analytical RMSEs obtained from the asymptotic analytical study, reported in Chapters 3 and 4. It was found that, in various signal and noise scenarios, the simulation and analytical results were in good agreement for a wide range of the number of data snapshots (L), under a comparatively low SNR condition (-5 to 10 dB).

10. As noted in EXP 6.3, the use of the estimated data covariance matrix $\hat{\mathbf{R}}$ in place of the estimated noise covariance matrix $\hat{\mathbf{R}}_{n;l}$ (see Eq. (3.58)), in the PFML-2 and BPFML-2 method (with or without bias correction) did not significantly affect the estimation error. However, it was shown also in the experiment that, either of the two methods which used $\hat{\mathbf{R}}$ converged slower than the one which used $\hat{\mathbf{R}}_{n;l}$ instead. This suggests that, at least when the CIQML algorithm is applied, improved performance in

terms of convergence rate can be obtained for the IV-based DOA estimators [18-20], by using $\hat{\mathbf{R}}_n$ as given by Eq. (3.27) instead of $\hat{\mathbf{R}}$.

11. It is interesting to compare the performance of the proposed BPFML methods with that of the beamspace-based DOA estimators [40-42]. For the BPFML methods to be efficiently applied, it is assumed that any two source clusters are sufficiently spaced angularly (more than one array beamwidth). Under this condition, at least approximations of the interfering signal directions can be obtained by using, for instance, the classical beamforming spectral estimator. Although no beamspace-based method was performed in the experiment, the CRLB under the assumption that some of the source directions are known a priori, as provided in Chapter 5, can be used to compare the performances of the two approaches. The bound as described sets the best performance which may be achievable (for unbiased estimates) by a beamspace-based method which is based on the exact knowledge of some source directions. In the case of the two-cluster four-source scenarios in EXP 6.2 and EXP 6.3, for fair comparison with the BPFML methods, the known signal directions for a beamspace-based method should be those of the 2nd-cluster signals (the 3rd and 4th signals). Because in the experiments these signals are uncorrelated with the first two signals, by comparing the constrained CRLB in Eq. (5.42) with the unconstrained bound in [9] for large L , it can be seen that the bounds on the asymptotic analytical RMSEs of the DOA estimates of the first two signal source directions remain unchanged whether the knowledge of the 2nd-cluster source directions is available or not. As a consequence, with the knowledge of the 3rd and 4th signal directions, the asymptotic performance of the DML and WSF methods under the scenarios used in EXP 6.2 and EXP 6.3 are still expected to be inferior to those of the PFML-2 and BPFML-2 methods. In addition, it is worth noting that, while the beamspace pre-processing is a spatial filtering scheme, the proposed pre-filtering approach is a temporal filtering scheme.

6.3 Conclusions

In this chapter, we have presented a detailed performance study of the PFML DOA estimators and comparison with some existing methods. The study was based on both computer simulation and analytical evaluation of the asymptotic formulas, derived in Chapters 3 and 4. The analytical study has been validated by simulation results for various signal and noise scenarios. The performance of the beamforming-based methods have been shown to be superior to the standard DML and WSF methods. The BPFML-4+bc method has been shown to offer performance improvement over the BPFML-2 method in the correlated source scenario, as demonstrated in EXP 6.6. Under the same scenario, the method has also yielded superior performance to the best-case PFML-2 method employing the type-1 pre-filtering reference signal, and the DEML method with incomplete knowledge of the signal waveforms. In addition, it has been demonstrated that the BPFML-2 method (with or without bias correction), as well as the PFML-2 method, is capable of estimating more source directions than those which can be estimated by most existing methods. Finally, provided that the noise spatial spectrum is sufficiently smoothed in the vicinity of the source directions, it has been shown that the performance of the BPFML-2 method is superior to those of the DML and WSF methods, when these methods are applied in an unknown spatially correlated noise field.

Chapter 7

Source Direction Estimation with Interpolated Arrays

7.1 Introduction

The study of DOA estimation in this thesis has so far been based on the assumption that exact knowledge of the array response vector is available. In addition, to allow for the application of the IQML algorithm, an ideal uniform linear (or rectangular) array, whose response vector is of the form described by Eqs. (5.6-5.9), was assumed in Chapters 5-6. In practice, however, a real uniform linear array has its response vector different from the ideal one due to various sources of perturbation. For instance, the array perturbation may arise from sensor placement error, non-identical sensor characteristics, and the presence of mutual coupling between sensors [83-86]. In this chapter, this realistic modelling error will be taken into consideration for the problem of DOA estimation using the IQML algorithm. The array interpolation approach will be used to improve the performance of the algorithm, when used with a perturbed uniform linear array. The least squares linear interpolation design procedure, due to Friedlander [25], is considered. Modification of the procedure in [25] is proposed. The bias of the DOA estimate using the interpolated array is next studied. The bias analysis for the PFML methods, detailed in Chapter 3, is applied to obtain the analytical estimation bias due to the array interpolation error. The result from this bias analysis is validated by the direct bias calculation using the asymptotic DML and WSF cost functions. Both the direct and analytical results are based on the data set of real array response vectors.

A problem associated with DOA estimation by the IQML algorithm with the interpolated array is next addressed. It is shown by means of computer simulation that, based on the IQML algorithm with the interpolated array, the WSF method suffers performance degradation in the presence of the interfering signal(s) whose direction lies outside the Interpolated Sector Of View (ISOV)[†]. The cause of this problem will be discussed. When the interfering signal(s) is uncorrelated with the signals whose DOAs lie inside the ISOV, the BPFML-2 method is shown to offer a solution to the problem.

7.2 The weighted least squares approach to array interpolation

In Chapters 5-6, the DOA estimation problem has been solved by using the IQML algorithm, under the assumption of an ideal uniform linear array. Doubt may be raised, however, on the applicability of the algorithm to a real uniform linear array with perturbation. Indeed, use of the IQML algorithm to a non-ideal uniform linear array results in degraded performance. This array irregularity problem can also degrade the performance of the Root-MUSIC method. In addition, highly biased DOA estimates may be obtained in applying spatial smoothing pre-processing to a perturbed uniform linear array. To alleviate these problems, the array linear interpolation approach has been considered by some researchers [25, 61]. Generally, the objective of the array interpolation is, through suitable choice of the interpolation matrix, to obtain the synthesised array of which the response vector is close to a desired one according to some criterion. In our case, the desired array response vector is that of the uniform linear array as described by Eq. (2.5). Before considering detailed discussion of the interpolated array design, however, it should be noted that the problem considered here has a different nature from another array perturbation problem, known as the auto-calibration (or self-calibration) problem [86-89]. In the latter problem, the perturbation

[†] The ISOV is the region of angular locations of which the corresponding array response vectors are required to be close to the desired ones.

is an unknown quantity, though some of its statistical properties are generally assumed known. The methods dealing with this problem involve joint estimation of source directions and array perturbation parameters. On the other hand, in the array interpolation problem, the array perturbation is assumed known, through real array response measurement, for instance. In addition, correction of the array perturbation is generally performed separately from the DOA estimation. Of course, difference between the true perturbed array response vector and the known nominal value is inevitable in practice. Nevertheless, due to the averaging nature of the interpolation processing, small random deviation of the array response vector from the known nominal value may be tolerable [25, 70].

Detail of the interpolated array design, based on the least squares approach in [25] with some modification, is provided in the following subsection.

7.2.1 The weighted least squares interpolated array design

In this subsection, we propose the interpolated array design procedure which is based on the least squares procedure of [25]. Modification is made to the method in [25], however, by adding the weighting function to the least squares formulation. In addition, control over the out-of-sector array response is introduced. Detail of the procedure is given as follows:

The interpolated array design procedure

1. Decide on the values of the following parameters:
 - The geometry of the interpolated array
 - The number of sensors N_i in the interpolated array
 - The interpolated array inter-sensor spacing d_i
 - The location in space of the interpolated array

- Two sets of angular locations. The first one, denoted by Θ_1 , is composed of the directions which form the ISOV. The second set of angular locations, Θ_2 , is composed of the directions which lie outside the ISOV.
- The weighting function $w_i(\theta)$

In this study, the uniform linear array geometry is chosen for the interpolated array. The two DOA sets Θ_1 and Θ_2 may be chosen by first dividing the entire array FOV into two angular regions. The first one belongs to the ISOV. The second one is the out-of-sector region, i.e., the region outside the ISOV. The members of Θ_1 are chosen as, say M_1 , angular locations which are uniformly distributed across the ISOV. The M_2 , say, members of Θ_2 can be obtained in a similar manner. The number of evaluated angular locations M_1 and M_2 should be chosen according to how smooth each real-array sensor directivity gain is as a function of the DOA. For instance, if the directivity gain of each sensor is sufficiently flat over the specified angular regions, only a small value of M_1 and M_2 is required.

The remaining parameter to be discussed is the weighting function $w_i(\theta)$. Generally, $w_i(\theta)$ should be selected to emphasise the ISOV, to obtain close approximation within the ISOV of the interpolated array response vector to the desired one. For instance, $w_i(\theta)$ may be chosen to have a (discrete) rectangular shape as described by

$$\begin{aligned} w_i(\theta) &= \kappa / M_1, \quad \theta \in \Theta_1 \\ w_i(\theta) &= 1 / M_2, \quad \theta \in \Theta_2 \\ w_i(\theta) &= 0, \quad \theta \notin \Theta_1 \cup \Theta_2 \end{aligned} \tag{7.1}$$

where the factor κ has positive value.

2. Obtain the $N \times N_i$ interpolation matrix $\mathbf{G}$ as a least squares solution to the following problem.

$$\mathbf{W}_i^{1/2} \tilde{\mathbf{A}}_r^H \mathbf{G} = \mathbf{W}_i^{1/2} \tilde{\mathbf{A}}_d^H + \mathbf{E} \tag{7.2}$$

where

$$\tilde{\mathbf{A}}_r = \begin{bmatrix} \mathbf{a}_r(\theta_1) & \mathbf{a}_r(\theta_2) & \cdots & \mathbf{a}_r(\theta_{M_{1+2}}) \end{bmatrix} \quad (7.3)$$

$$\tilde{\mathbf{A}}_d = \begin{bmatrix} \mathbf{a}_d(\theta_1) & \mathbf{a}_d(\theta_2) & \cdots & \mathbf{a}_d(\theta_{M_{1+2}}) \end{bmatrix} \quad (7.4)$$

$$\mathbf{W}_i = \begin{bmatrix} w_i(\theta_1) & & & 0 \\ & w_i(\theta_2) & & \\ & & \ddots & \\ 0 & & & w_i(\theta_{M_{1+2}}) \end{bmatrix} \quad (7.5)$$

The angular locations θ_k , $k = 1, \dots, M_1$, are the M_1 members of Θ_1 , while θ_k , $k = M_1+1, \dots, M_{1+2}$ ($M_{1+2} = M_1 + M_2$), are the M_2 members of Θ_2 . In addition, $\mathbf{a}_r(\theta)$ and $\mathbf{a}_d(\theta)$ are, respectively, the real and desired array response vectors at θ . The least squares solution to Eq. (7.2), which minimises the Frobenius norm of the error matrix $\mathbf{E}$, is given by

$$\mathbf{G} = (\tilde{\mathbf{A}}_r \mathbf{W}_i \tilde{\mathbf{A}}_r^H)^{-1} \tilde{\mathbf{A}}_r \mathbf{W}_i \tilde{\mathbf{A}}_d^H \quad (7.6)$$

3. The above two steps may be repeated several times with different values of the parameters, listed in Step 1, such as the interpolated array location. From the resulting interpolation matrices, the best one may be chosen by first calculating, for each $\mathbf{G}$, the $M_1 \times 1$ interpolation error vector $\mathbf{e}_i$ whose k th elements are given by

$$e_{i;k} = \|\tilde{\mathbf{e}}_k\|^2 / \|\mathbf{a}_r(\theta_k)\|^2, \quad k = 1, \dots, M_1 \quad (7.7)$$

where $\tilde{\mathbf{e}}_k$ is the k th column vector of

$$\tilde{\mathbf{E}}_i = \mathbf{G}^H \tilde{\mathbf{A}}_r - \tilde{\mathbf{A}}_d \quad (7.8)$$

Each of the elements $e_{i;k}$ in $\mathbf{e}_i$ is the interpolation error corresponding to each of the M_1 angular locations inside the ISOV. It is proposed that the best interpolation matrix $\mathbf{G}$ is chosen as the one with the minimum $\|\mathbf{e}_i\|_\infty$. From the chosen $\mathbf{G}$, the interpolated array output vector is given by

$$\mathbf{x}_i(t) = \mathbf{G}^H \mathbf{x}(t) \quad (7.9)$$

Further reduction of $\|\mathbf{e}_i\|_\infty$ can still be achieved by continuing the design by the following optional step.

4. With the power factor $\delta > 0$, the M_1 elements of $\tilde{\mathbf{e}}_i = \kappa \mathbf{e}_i^\delta / (\mathbf{I}_{M_1}^T \mathbf{e}_i^\delta)$ are used as the first M_1 diagonal elements of $\mathbf{W}_i$. Other diagonal elements of $\mathbf{W}_i$ remain unchanged. The newly formed $\mathbf{W}_i$ is subsequently used to find a new interpolation matrix $\mathbf{G}$ from Eq. (7.6). This step may be repeated several times with different values of δ . Among the interpolation matrices obtained by different values of δ , the one with the minimum $\|\mathbf{e}_i\|_\infty$ is again selected.

Before considering a design example using the above procedure, the difference between the proposed procedure and the one described in [25] should be mentioned. First, in addition to the least squares criterion applied also in [25], the ∞ -norm ($\|\mathbf{e}_i\|_\infty$) criterion is used in the 3rd step and the optional 4th step of the proposed procedure. The second difference between the two procedures is in the way in which $\tilde{\mathbf{A}}_r$ and $\tilde{\mathbf{A}}_d$ are formed. For the procedure in [25], $\tilde{\mathbf{A}}_r$ and $\tilde{\mathbf{A}}_d$ are composed of, respectively, the real and desired response vectors corresponding to the angular locations within the ISOV. The method does not impose control of the interpolated array directivity gain[†] in the out-of-sector region. In contrast, the proposed design procedure imposes mild directivity gain constraint in the out-of-sector region, through the inclusion in $\tilde{\mathbf{A}}_r$ and $\tilde{\mathbf{A}}_d$ of the real and desired response vectors corresponding to the out-of-sector angular locations. Without such control, the resulting interpolated array may give poor DOA estimation despite its lower interpolation error than that of the one with the out-of-sector gain control. This is because, as well as the interpolation error, the estimation bias is also affected by the interpolation matrix $\mathbf{G}$ through the interpolated array noise covariance structure, which is $\mathbf{G}^H \mathbf{R}_{ns} \mathbf{G}$ for the DML and WSF estimators. Note that, $\mathbf{R}_{ns}$ is the noise covariance

[†] The array directivity gain is defined here as the 2-norm of the array response vector.

structure of the real array. It will be shown in the next subsection that, despite lower interpolation error, the interpolated array based on the design procedure in [25] can yield larger estimation bias than that due to the interpolated array obtained by the proposed procedure.

7.2.2 Design example and the analysis of estimation bias

In this subsection, an example of the array interpolation design is presented. The design was based on a data set of 720 real array response vectors corresponding to the angular locations equally spaced by 0.25 degrees over the array FOV of ± 90 degrees[†]. The measurements correspond to a real uniform linear array of eight sensors. Each sensor is a T-shaped vertically polarised half wavelength dipole, in the form of a double sided printed circuit board with a single RF output. The inter-sensor spacing is 0.574 times the operating radio wavelength. This data set was part of the SCARP project, carried out by ERA Technology. Additional information of the array can be found in [90]. The number of sensors and their spacing for the interpolated array were chosen to be the same as in the real array. The ISOV was between -11 and +6 degrees, which is approximately one array beamwidth. The members of Θ_1 were $M_1 = 69$ angular locations which are uniformly distributed across the ISOV. Similarly, $M_2 = 20$ angular locations, which lie uniformly across the angular region outside the ISOV, were chosen as the members of Θ_2 . The weighting factor κ was 100. These values of M_1 , M_2 , κ , as well as the chosen location of the interpolated array, are the ones which yield, among a few other parameter choices, the minimum $\|e_i\|_\infty$. Based on this optimal parameter choice, Step 4 of the design procedure was carried out with various values of the power factor δ . The factor $\delta = 0.5$ was found, among five different values of δ between 0.25 and 2, to achieve the minimum $\|e_i\|_\infty$. The interpolation errors at M_1 angular locations

[†] The set of the real array response vectors was in fact obtained after linear interpolating of the array response measurements taken at 5 degree interval. This preliminary interpolation should not be confused with the interpolation performed by G in Eq. (7.9).

within the ISOV, previously defined as the elements $e_{i;k}$ of $\mathbf{e}_i$, are shown in Figure 7.1 for the array designed without Step 4 (Design #1), and the one designed with Step 4 (Design #2). The reduction of the maximum interpolation error ($\|\mathbf{e}_i\|_\infty$) by Design #2 is noted.

To compare the proposed design procedure with that of [25], the third interpolated array was designed with $\tilde{\mathbf{A}}_r$ and $\tilde{\mathbf{A}}_d$ composed of the real and desired response vectors corresponding to θ_k , $k = 1, \dots, M_1$. Other parameter values as listed in the 1st step of the proposed design procedure, are the same as the optimal ones chosen in the previous two designs. The interpolation error by this design (Design #3) is compared with that of Design #2 in Figure 7.2. From the figure, it may be suggested that this third design should be preferred due to its lower interpolation error. However, as will be seen later, it is the second design which results in less biased DOA estimate(s).

Next, we investigate the DOA estimation bias caused by the specific-case interpolation error matrix

$$\mathbf{E}_i = \mathbf{G}^H \mathbf{A}_r - \mathbf{A}_d \quad (7.10)$$

where $\mathbf{A}_r$ and $\mathbf{A}_d$ are the real and desired response matrices corresponding to a particular source direction scenario. Notice that the estimation bias due to $\mathbf{E}_i$ and the one due to $\mathbf{E}_b$ in the filtered data using the type-4 pre-filtering reference signal, as detailed in Chapter 3, are mathematically equivalent. Thus, the bias due to $\mathbf{E}_i$ can be analytically obtained by one of the asymptotic bias formulas Eqs. (3.120, 3.122, 3.125). With the PFML-3 DOA estimator, the estimation bias due to $\mathbf{E}_i$ for large L and sufficiently small $\mathbf{E}_i$ is given by

$$\tilde{\boldsymbol{\theta}} = \text{Re}^{-1} \{ (\bar{\mathbf{D}}^H \tilde{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\tilde{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^\perp \tilde{\mathbf{R}}_n^{-1/2} \bar{\mathbf{D}}) \odot (\bar{\mathbf{T}} \bar{\mathbf{T}}^H)^T \} \cdot \text{Re} \{ \{ (\bar{\mathbf{D}}^H \tilde{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\tilde{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^\perp \tilde{\mathbf{R}}_n^{-1/2}) \odot (\mathbf{E}_i \bar{\mathbf{T}} \bar{\mathbf{T}}^H)^T \} \mathbf{I}_N \} \quad (7.11)$$

where $\tilde{\boldsymbol{\theta}} = \text{E}\{\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}\}$, $\tilde{\mathbf{R}}_n = \mathbf{G}^H \bar{\mathbf{R}}_n \mathbf{G}$.

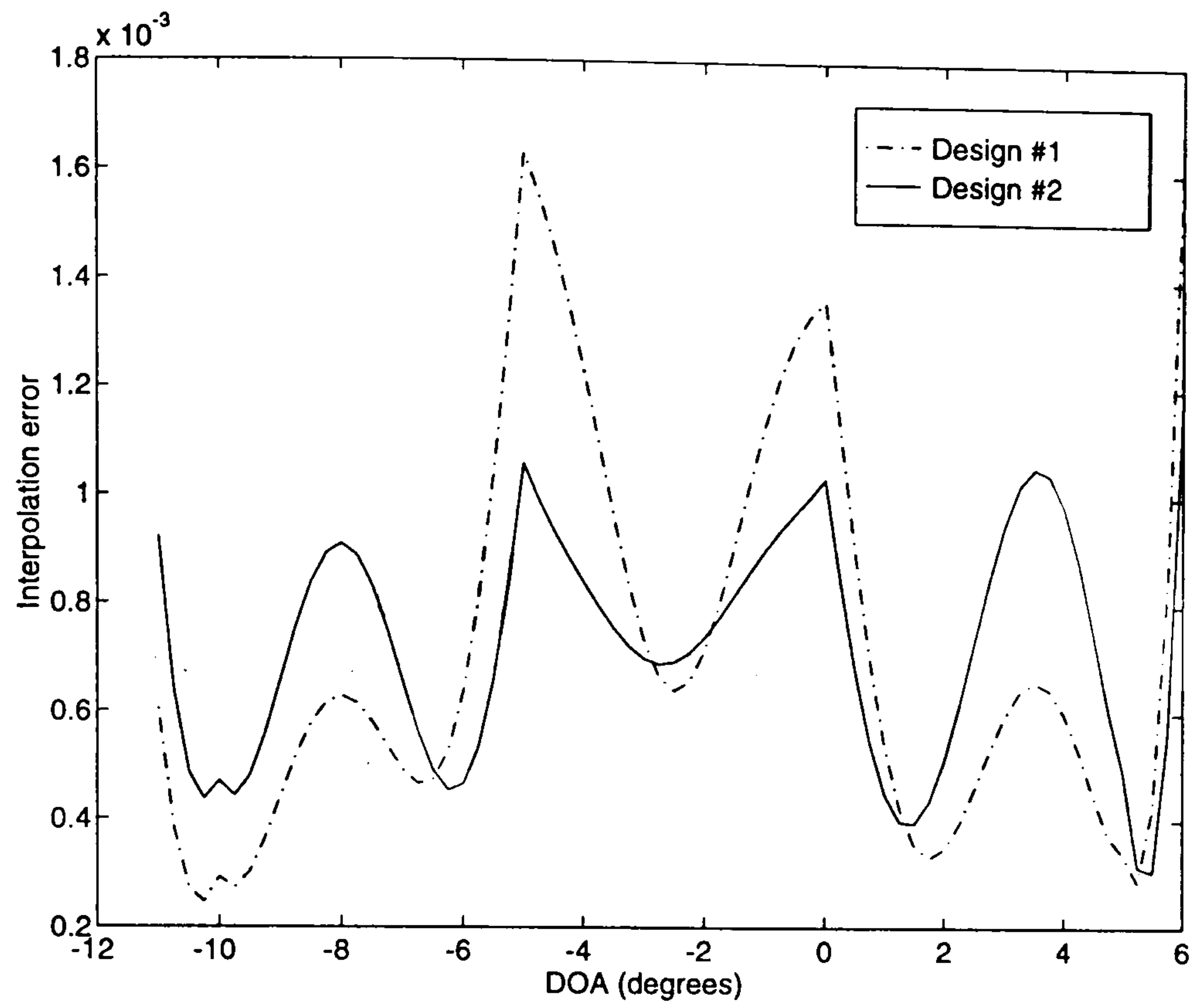


Figure 7.1: Array interpolation error ($e_{i,k}$) corresponding to each of two interpolated array designs, Designs #1 and #2, as a function of the angular location within the ISOV.

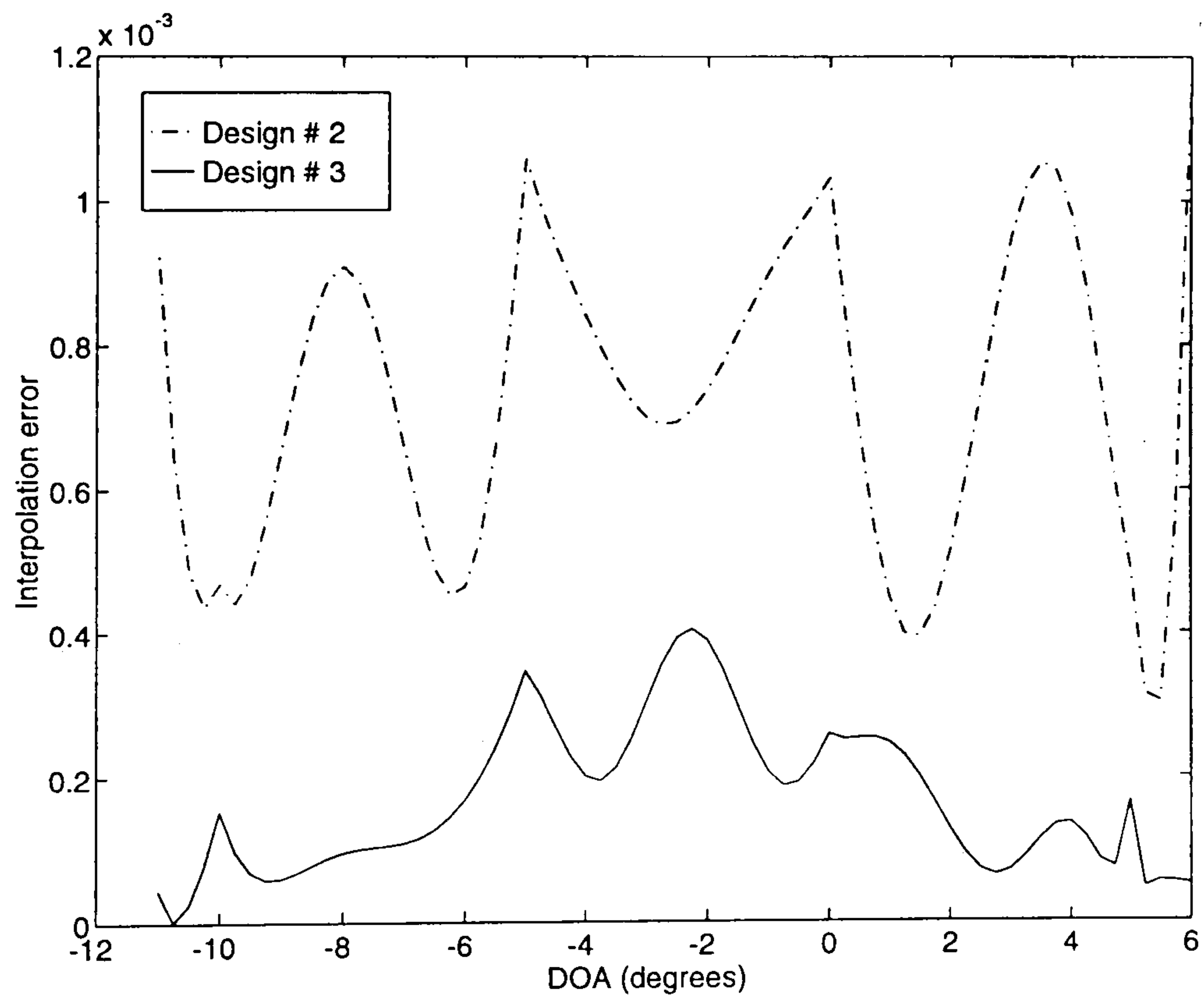


Figure 7.2: Array interpolation error ($e_{i,k}$) corresponding to each of two interpolated array designs, Designs #2 and #3, as a function of the angular location within the ISOV.

The matrix $\bar{\mathbf{T}}\bar{\mathbf{T}}^H$ in Eq. (7.11) is as given in Theorem 3.5. The asymptotic formula Eq. (7.11) can also be applied for the cases of the DML and optimal WSF methods. If the DML DOA estimator is applied, $\bar{\mathbf{T}}\bar{\mathbf{T}}^H = \mathbf{R}_s$. For the optimal WSF method, this matrix term is equal to $(\mathbf{R}_s \mathbf{A}_d^H \mathbf{R}^{-1} \mathbf{A}_d \mathbf{R}_s)$. The analytical bias formula Eq. (7.11) is next validated using the data set of real array response vectors with two interpolated arrays based on Designs #1 and #2. Under the large snapshot condition where the finite sample effect is negligible, two experiments were carried out for the purpose. The comparison between the three array designs (Designs #1-3) will also be given.

EXP 7.1: The single source scenario is first considered. In the simulation, the signal direction was varied among sixty-one values. In each of sixty-one DOA values, the exact array covariance matrix was generated using the set of real array response vectors with each of the two interpolation matrices by Designs #1 and #2. The identity matrix with a positive scaling factor was added into the array covariance matrix, to simulate the additive white noise covariance matrix. Note that, in this single source case, the SNR value has no effect on the estimation bias. The DOA estimate was calculated for each source direction configuration by using the CIQML algorithm with the asymptotic DML cost function, which is Eq. (2.24) with $\hat{\mathbf{R}}$ being replaced by $\mathbf{R}$. The algorithm was applied with the conjugate-symmetric constraint $\tilde{\mathbf{H}} = \tilde{\mathbf{I}}_{K+1} \tilde{\mathbf{J}}_{K+1}^T \tilde{\mathbf{I}}_{K+1}$, and $\tilde{\mathbf{H}} = \mathbf{G}^H$. The maximum number of iterations allowed for the algorithms was 50. The threshold number ϵ was 0.5E-10 and the chosen initial condition $\mathbf{F}_1$ is as given by Eq. (5.26). The estimation bias, obtained by this direct evaluation, compared with that of the analytical formula Eq. (7.11) for sixty-one DOA configurations, is shown in Figure 7.3. From the figure, it is seen that the analytical result is in good agreement with the result by direct evaluation. In comparing the estimation bias for each of the two interpolated arrays, the one from the interpolated array using Design #2 achieved the lower maximum bias. Figure 7.4 compares two analytical biases for two interpolated arrays obtained by Designs #2 and #3. The result shows that, despite lower interpolation error seen in

Figure 7.2, the array based on Design #3 does not have an improvement in terms of the estimation bias over the one obtained by Design #2.

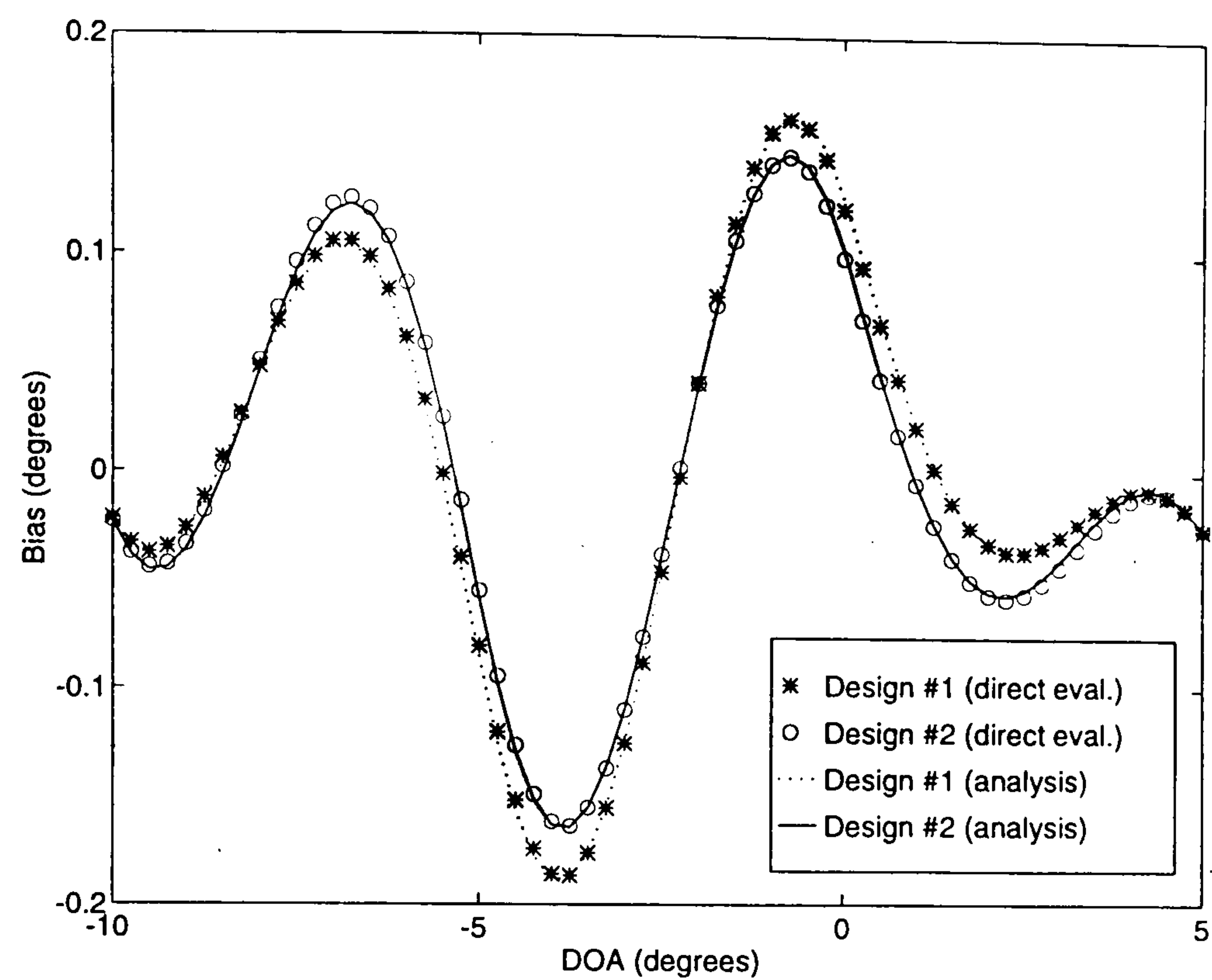


Figure 7.3: The estimated and analytical biases of the DOA estimate as a function of the single source direction using two interpolated arrays obtained by Designs #1 and #2.

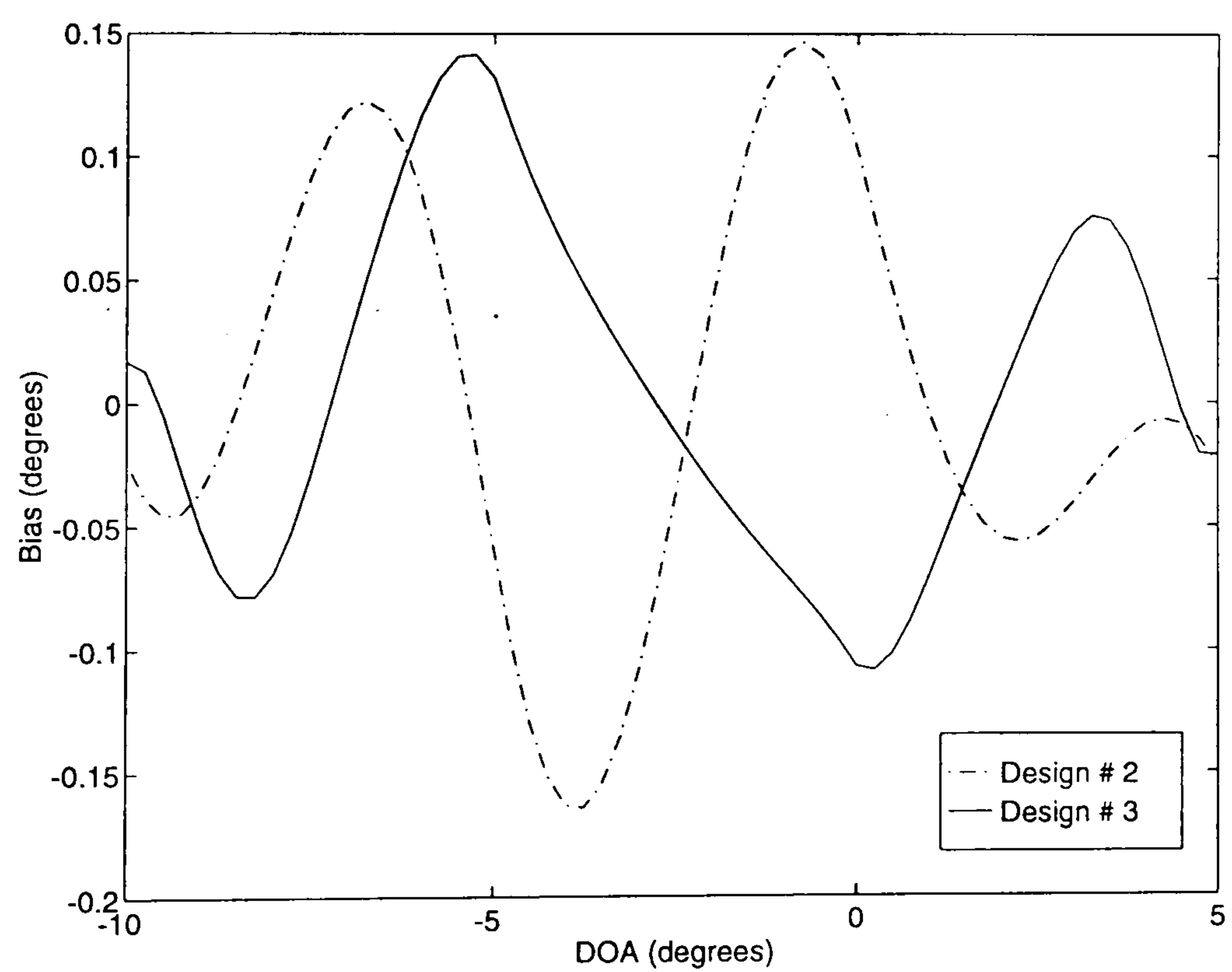
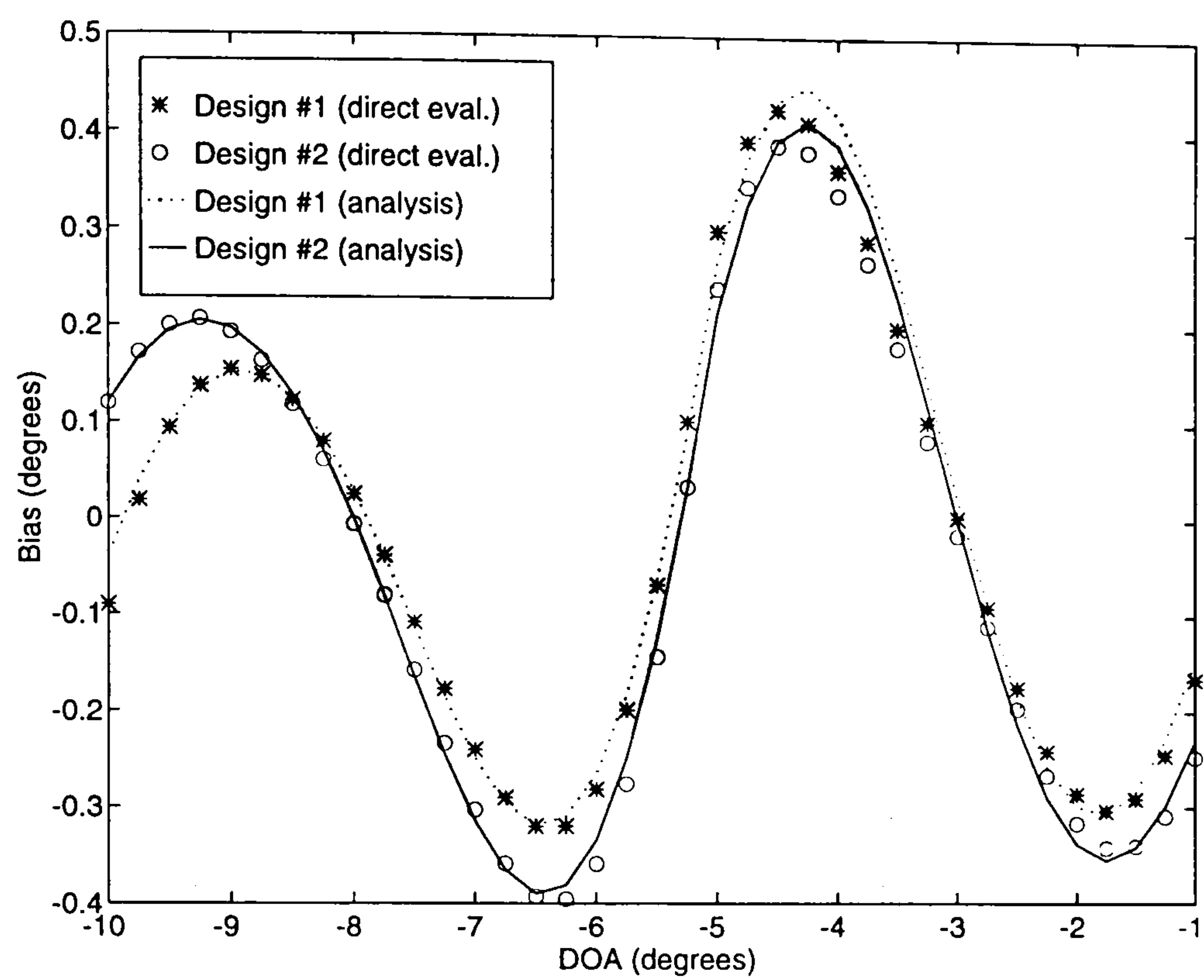


Figure 7.4: The analytical bias of the DOA estimate as a function of the single source direction using two interpolated arrays obtained by Designs #2 and #3.

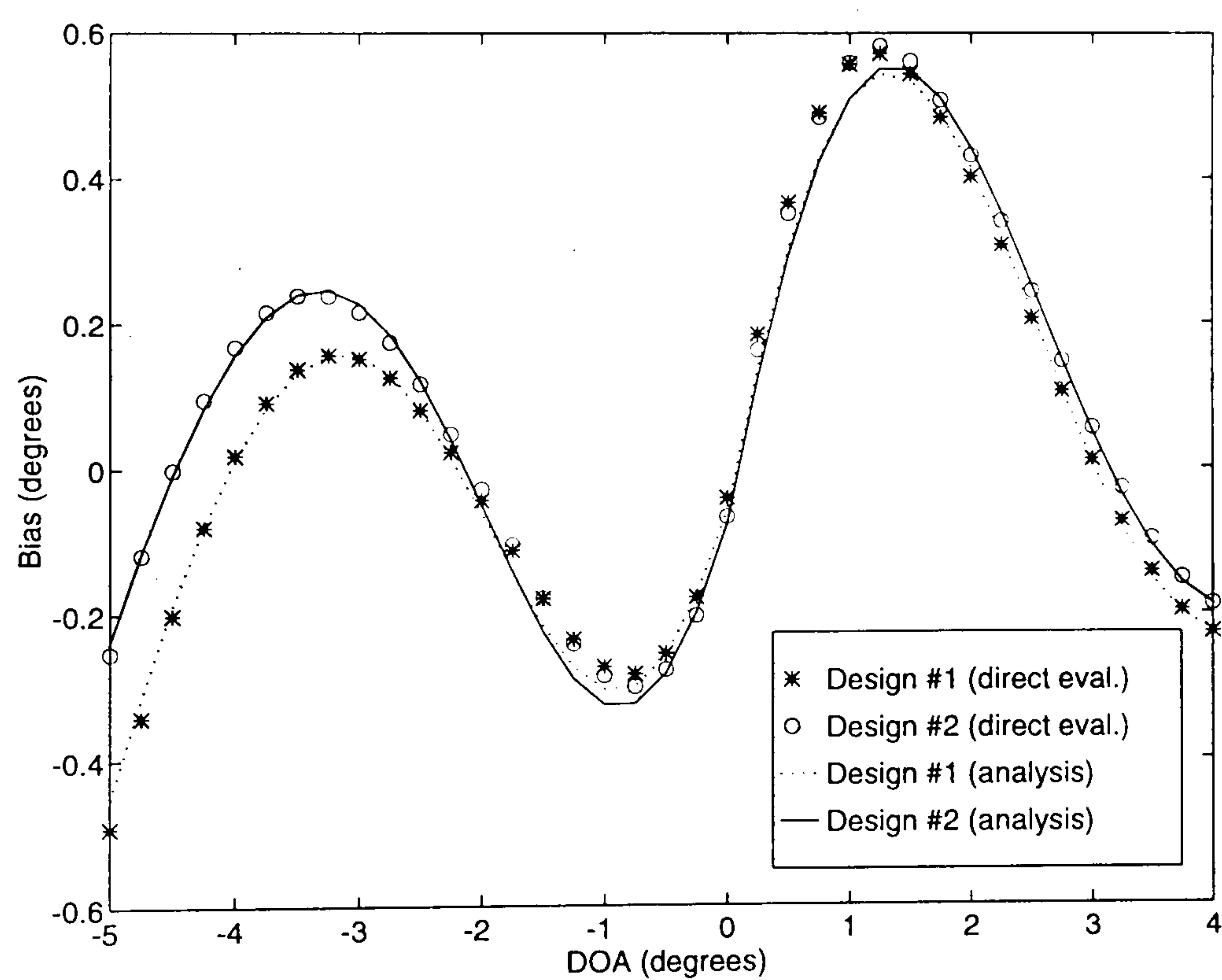
EXP 7.2: Next, a two coherent signal scenario is considered. In this experiment, two coherent signals with unity power each arrived at the array with the angular spacing of 5 degrees. The phase difference between the two signals was $\pi/4$. The signal phase was measured at the middle point between the fourth and fifth sensors of the real array. The directions of the two signals were varied to obtain thirty-seven DOA configurations. For each DOA configuration, the exact array covariance matrix was generated by using the real array response vectors with each of the two interpolation matrices, obtained by Designs #1 and #2. The scaled identity matrix was added, as the noise covariance matrix, into the array covariance matrix. The DOA estimates were obtained by the CIQML algorithm with the asymptotic optimal WSF cost function, which is based on the eigendecomposition of $\mathbf{R}$ instead of $\hat{\mathbf{R}}$. The algorithm parameters such as the threshold number ε were the same as in the EXP 7.1. By direct evaluation, the bias of each of the two source DOA estimates from 37 DOA configurations, compared with those obtained by Eq. (7.11), are shown in Figure 7.5. The analytical result is seen to be closely matched with the result from direct evaluation of the WSF cost function. However, unlike in EXP 7.1, the performance difference between the two interpolated arrays is less clear in this two coherent-source case. This seems to be because, in this case, the estimation bias for each source direction depends on the combination of the two column vectors in $\mathbf{E}_i$. The desirable property of the interpolated array obtained by Design #2 is its lower maximum interpolation error, when compared with that of the array obtained by Design #1 (see Figure 7.1). The averaging of the two error vectors in $\mathbf{E}_i$ makes the resulting averaged error vector less fluctuated. This appears to be the reason for the similar estimation performance by using the two arrays.

The analytical biases of the two DOA estimates using two interpolated arrays obtained by Designs #2 and #3 are shown in Figure 7.6. The biases of the two estimates using the array obtained by Design #3 are seen to fluctuate more, when compared with

those of the array obtained by Design #2. Thus, the interpolated array based on Design #2 should be preferred in terms of the estimation bias property.



(a)



(b)

Figure 7.5: The biases of the two DOA estimates in EXP 7.2 for various angular locations using two interpolated arrays obtained by Designs #1 and #2. (a) 1st source DOA; (b) 2nd source DOA.

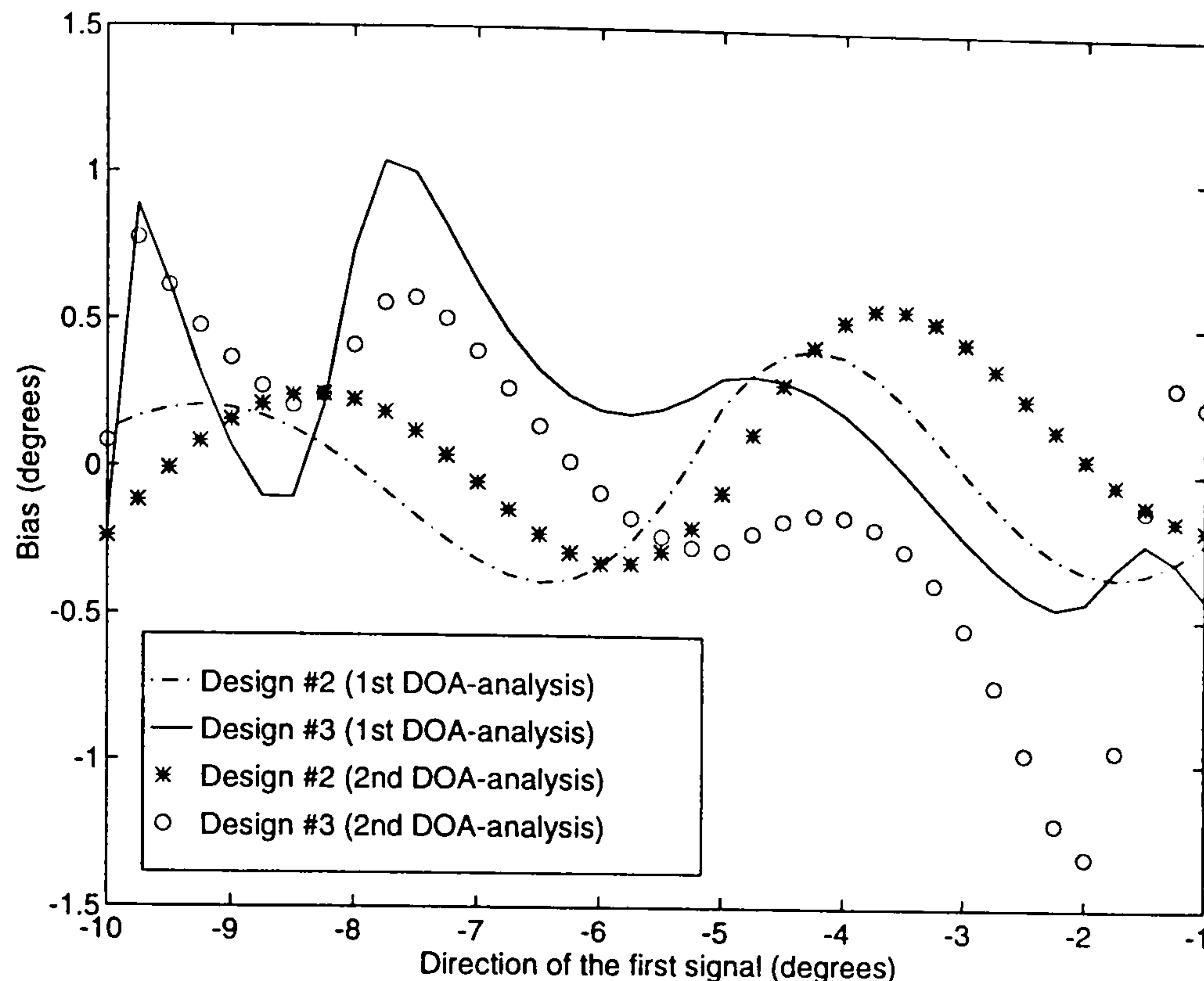


Figure 7.6: The analytical biases of two coherent source DOA estimates as a function of the first source direction using two interpolated arrays obtained by Designs #2 and #3.

7.3 Source direction estimation by the IQML algorithm with the interpolated uniform linear array

In Section 7.2, we investigated the DOA estimation bias using the interpolated array under the asymptotic condition, where the finite sample effect is neglected. In this section, the finite sample case will be considered. To study the performance of the CIQML algorithm with the interpolated array, two simulation experiments were carried out based on the set of real array response vectors, previously described, and the interpolation matrix obtained by Design #2. Details of the experiments and results are provided below.

EXP 7.3: In this simulation experiment, two coherent complex Gaussian distributed signals with unity power and phase difference of $\pi/4$ were assumed to impinge upon the real array with the angular spacing of 5 degrees. The directions of both signals were simultaneously changed with the step size of 1 degrees to obtain 11 different source

DOA configurations shown in Table 7.1. In each configuration, the additive white Gaussian noise was added at each sensor to obtain the SNR of 20 dB. Different signal and noise sequences were used in each of 50 simulations for each source direction configuration. The two source directions were estimated using the optimal WSF method with the CIQML algorithm. The algorithm parameters such as the threshold number ϵ were the same as in EXP 7.1. The estimated RMSEs of the two DOA estimates using the interpolated array and the real array are shown Figure 7.7. The improvement gained by using the interpolated array is seen from the figure. The maximum RMSE among the total of 22 DOA estimates is 0.560 degrees for the method with the interpolated array, while the maximum RMSE for the estimates with the real array is 4.581 degrees. The RMS value of the RMSEs of the 22 DOA estimates is equal to 0.266 degrees for the method with the interpolated array, and is equal to 2.565 degrees for the method with the real array.

Figure 7.8 shows the estimated biases of the DOA estimates using the real and interpolated arrays. By comparing the bias with the RMSE, it is seen that the estimation error is due largely to the bias for both cases of the real and interpolated arrays. For the estimates using the interpolated array, the ratio between the absolute value of the estimation bias and the RMSE is greater than 0.8 in 17 of 22 DOA estimates.

Subexp no.	1	2	3	4	5	6	7	8	9	10	11
1st DOA (degrees)	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
2nd DOA (degrees)	-5	-4	-3	-2	-1	0	1	2	3	4	5

Table 7.1: List of 11 source DOA configurations corresponding to two coherent signals in EXP 7.3.

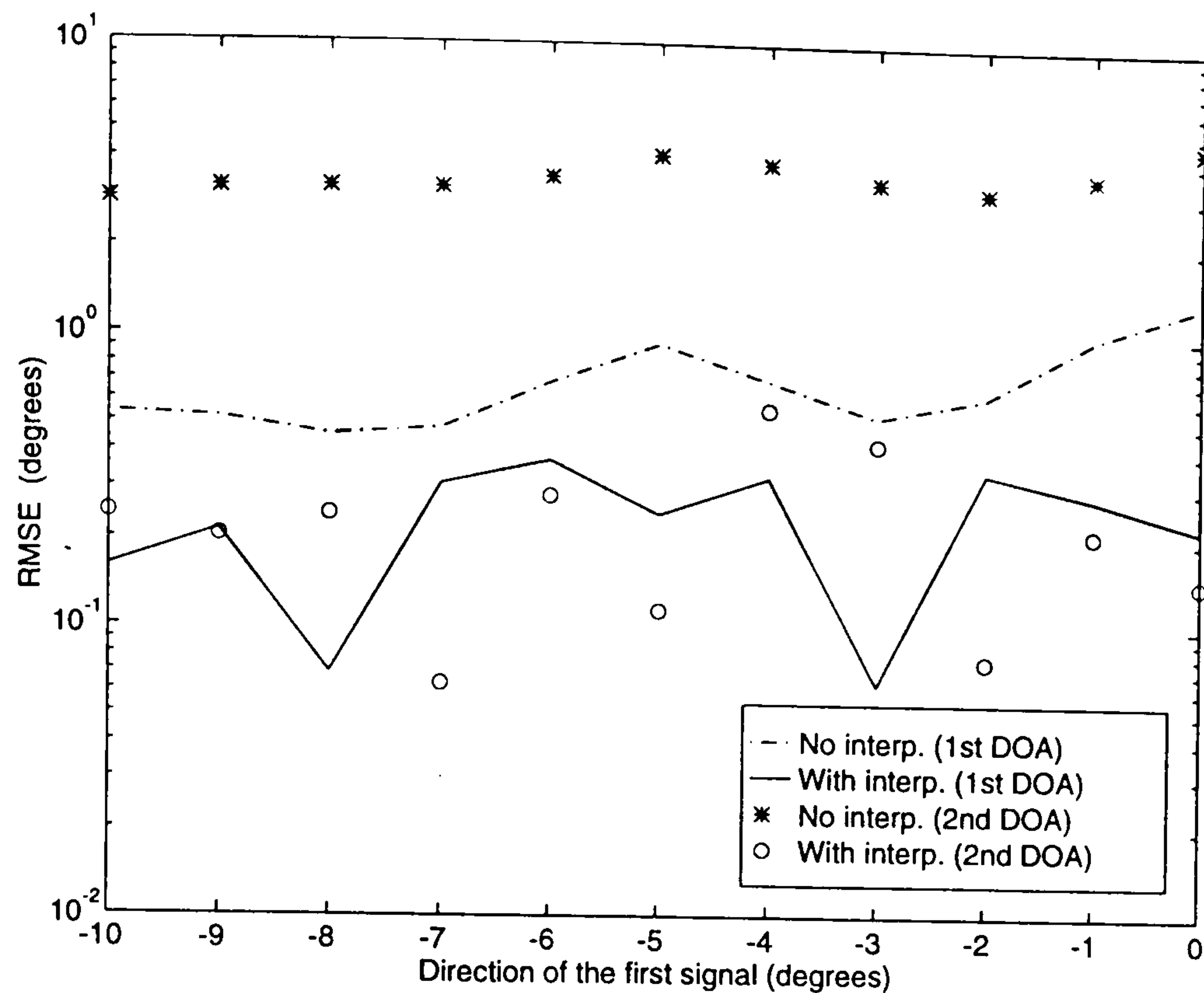


Figure 7.7: The estimated RMSEs of two coherent source DOA estimates in EXP 7.3 using the optimal WSF cost function with the CIQML algorithm. Two sets of the DOA estimates were obtained based on the real and interpolated arrays.

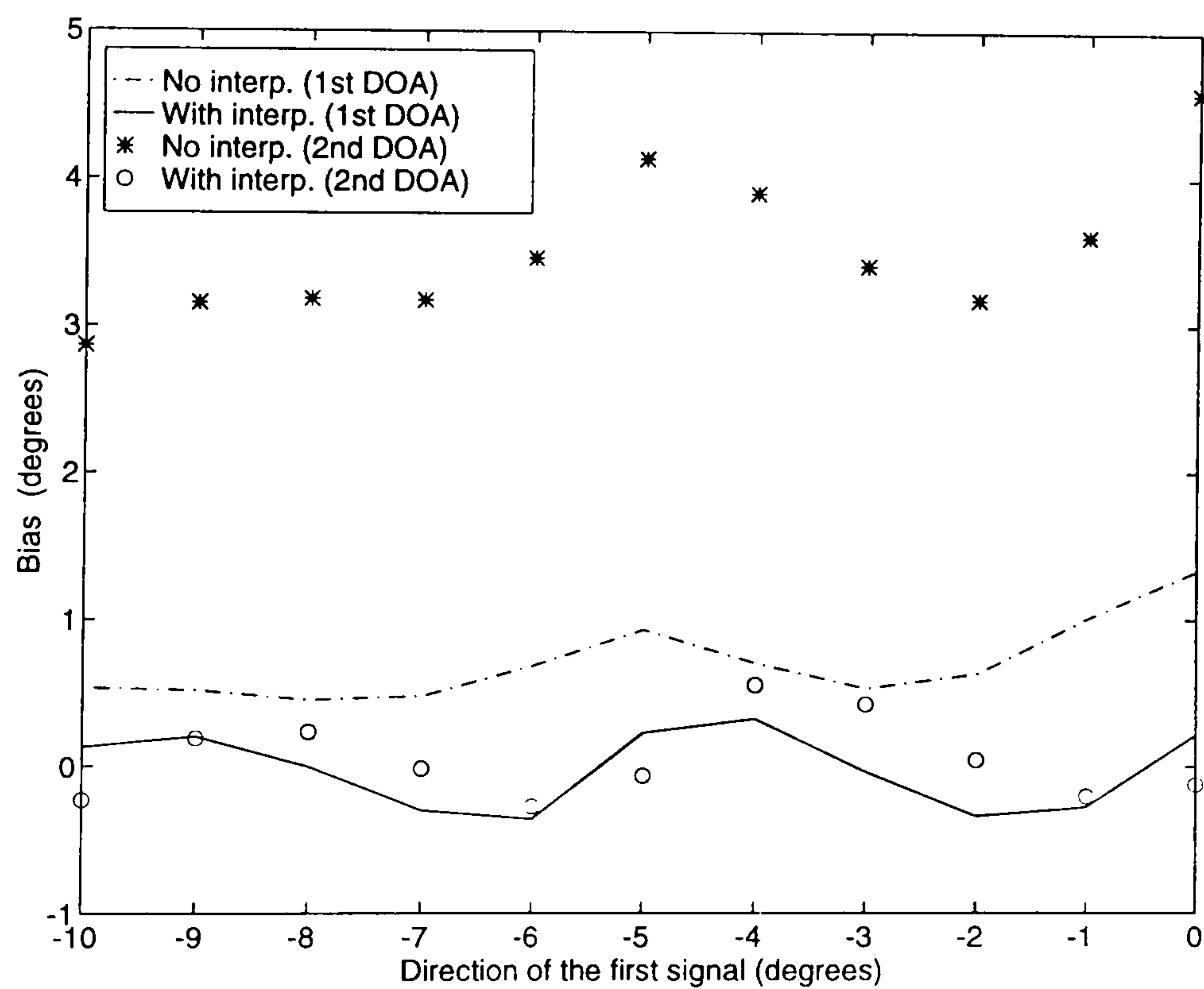


Figure 7.8: The estimated biases of two coherent source DOA estimates in EXP 7.3 using the optimal WSF cost function with the CIQML algorithm. Two sets of the DOA estimates were obtained based on the real and interpolated arrays.

EXP 7.4: The two coherent source scenario in EXP 7.3 was again assumed. In this experiment, however, the directions of the two sources were fixed at -5 and 0 degrees. In addition, the third uncorrelated signal was added with its direction being varied among four different values. The third signal has the same unity power as the two coherent signals. The additive noise was added to obtain the SNR of 20 dB. Direction estimation was carried out using three different methods. The first two methods were based on the optimal WSF cost function. The first one was applied to the data without the interpolation matrix. The second method was applied using the interpolation matrix which was obtained by Design #2. The same interpolation matrix was used in the third BPFML-2 method, described in Chapters 4 and 6. For the BPFML-2 method, only the two DOAs of the coherent signals were estimated. In terms of the source cluster model, described in Chapter 4, the two coherent signals form the first source cluster, while the third uncorrelated signal belongs to the second cluster. In all three methods, the CIQML algorithm was used to obtain the DOA estimates. The algorithm parameters such as the initial condition are the same as in EXP 7.1. From 100 simulations, the estimated RMSEs of the two coherent source DOA estimates by the three methods for each of four different DOAs of the third signal, are shown in Table 7.2. From the table, the BPFML-2 method outperforms the WSF methods with the real and interpolated arrays. The reason for poor performance of the WSF method with the real array is obvious. The case of the method with the interpolated array, however, needs detailed explanation. It may appear at first that the superior performance of the BPFML-2 method over the WSF method, both based on the interpolated array, is due to the general performance order among the two methods as demonstrated in Chapter 6. However, by comparing the performance of the WSF method using the interpolated array with that of the same method for the two-signal case in EXP 7.2, severe performance degradation can be observed for the result in this experiment even if the angular location of the third signal is far from those of the first two coherent signals. This suggests that the problem is not

simply due to the presence of more signal sources. Notice that the angular location of the third source lies outside the ISOV. Thus, its response vector is not well approximated by the uniform linear array response structure as given by Eq. (2.5). As a consequence, the IQML algorithm suffers a modelling error problem. This imposes a serious limitation to the application of the interpolated array to DOA estimation using the WSF cost function with the IQML algorithm. This out-of-sector-signal problem does not seem to have been addressed before. Although a method similar to the WSF method with the IQML algorithm was applied to the interpolated uniform linear array in [70], the simulation study provided in [70] considered only the case where all signal directions lie within the ISOV.

In addition to the proposed solution using the BPFML-2 method[†], an approach based on spatial filtering may be suggested as a possible solution to the above problem. It may appear that the beamspace pre-processing [40-42] can be applied to suppress the out-of-sector signals, provided that their locations are known at least approximately. Unfortunately, the response vectors of the out-of-sector signals may not have the structure Eq. (2.5). Thus, it is generally not possible to design the beamspace matrix so as to suppress the out-of-sector signals while preserving the (approximate) structure Eq. (2.5) of the interpolated response vectors. It is possible, however, to first place the beamspace matrix just after the real array output. The interpolation matrix is then applied to the resulting beamspace data vector where the out-of-sector signals are already suppressed. Nevertheless, in doing so, the interpolation matrix needs to be designed for a particular beamspace matrix, which in turn depends on the angular locations of the out-of-sector signals. This latter approach then turns out to be equivalent to the interpolated array design procedure of Bronez, detailed in [61]. The goal of this alternative design procedure is to match the interpolated array response vectors at some

[†] Similar performance is also expected by using the BPFML-2a method. In addition, when the signals whose directions lie within the ISOV are only partially correlated, the BPFML-4 method may be used instead.

angular locations within the ISOV to the desired ones, while attempting to minimise the array gain in the directions outside the ISOV. To allow for small variation in the out-of-sector signal directions without re-designing the interpolation matrix, more degrees of freedom are needed to provide sufficient signal suppression for a given angular region (outside the ISOV). Thus, the effective number of array sensors left for DOA estimation is reduced. As a consequence, DOA estimation performance has to be sacrificed. For the out-of-sector signal(s) which is uncorrelated with the signals whose angular locations lie within the ISOV, the proposed BPFML-2 method has been shown to offer a satisfactory solution. Not only in that it solves the problem without degrading estimation performance, the method generally yields better performance than the DML and WSF methods, as shown in Chapter 6. However, the presence of the out-of-sector signal(s) which is correlated with the signals within the ISOV remains as an unsolved problem.

The 3rd DOA (degrees)	20	30	40	50
WSF + real array	0.826	0.621	0.504	0.508
WSF + interp. array	0.298	1.195	0.759	2.752
BPFML-2 + interp. array	0.187	0.209	0.194	0.199

(a)

The 3rd DOA (degrees)	20	30	40	50
WSF + real array	4.031	3.811	2.562	3.030
WSF + interp. array	0.121	2.155	0.872	2.267
BPFML-2 + interp. array	0.169	0.185	0.181	0.181

(b)

Table 7.2: The estimated RMSEs (in degrees) of two coherent source DOA estimates in EXP 7.4, obtained by three ML methods at four DOAs of the 3rd uncorrelated signal with the SNR of 20 dB. (a) 1st DOA = -5 degrees; (b) 2nd DOA = 0 degrees.

7.4 Conclusions

In this chapter, the problem of source direction estimation using the IQML algorithm with the interpolated array has been considered. The weighted least squares procedure for the interpolated array design has been presented. The procedure is based on the least squares method of [25]. Modification has been made to the method in [25] by including the weighting function into the design equations. The mild constraint on the interpolated array directivity gain in the out-of-sector region has been imposed. It has been shown that, without this out-of-sector gain constraint, the resulting interpolated array can exhibit large estimation bias despite its lower interpolation error (than that of the array designed with the constraint).

The analysis of the estimation bias due to the interpolation error has also been made. The analytical bias, obtained by the adaptation of the asymptotic bias formula in Chapter 3, has been shown to be closely matched with the result obtained by direct evaluation of the asymptotic DML and WSF cost functions. The reduction of the maximum interpolation error by continuing the array design in Step 4 of the proposed procedure has been observed in the single source case. However, the performance difference between the two interpolated array designs, with and without Step 4, has so far been inconclusive for the more complicated case of two coherent sources.

The problem of source DOA estimation by the IQML algorithm, using the interpolated array with a finite number of snapshots, has been investigated. It has been shown, by means of computer simulation, that the use of the interpolated array greatly improves the estimation performance of the IQML algorithm. The out-of-sector-signal problem has been addressed. It has been shown that, with the WSF cost function, the estimation performance is degraded in the presence of the signal(s) whose angular location lies outside the ISOV. The cause of this problem has been pointed out. For the case of the uncorrelated out-of-sector signal(s), the use of the BPFML-2 method has been proposed as a solution to this problem. Its superior estimation performance over

the WSF method in the presence of the uncorrelated out-of-sector signal has been demonstrated by computer simulation.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

In this thesis, the problem of narrow-band source direction estimation from an array of sensors has been considered. The pre-filtering maximum likelihood (PFML) approach to DOA estimation has been proposed. This approach has been shown to provide an alternative and unified treatment of the ML method with known signal waveforms [16-17], and the IV-based DOA estimators [18-20]. The pre-filtering approach is based on certain temporal pre-processing of the raw data obtained from the outputs of the array sensors. The resulting filtered data snapshots are used to formulate a DOA estimator. Four filter types have been proposed for use in the pre-filtering step of the PFML framework. The coefficients of these filters are obtained from the pre-filtering reference signal, which should be highly correlated with the signal waveforms. Four types of the pre-filtering reference signals have been considered. Based on the statistical properties of the data filtered by each of the four filters, four PFML DOA estimators have been formulated. Analysis of the asymptotic properties of these PFML DOA estimators has been provided.

In Chapter 4, based on the described PFML framework, the new beamforming-based BPFML methods for DOA estimation have been proposed. The methods are the realisations of the PFML methods which are based on the type-4 pre-filtering reference signal. The combined classical and LCMV beamforming procedure has been proposed to obtain the type-4 pre-filtering reference signal from the observed data. This beamforming procedure is well suited for the scenario consisting of well-separated

clusters of highly or fully correlated signals. When the signals within the same cluster are only partially correlated, the improved beamforming procedure, also proposed in Chapter 4, should be used instead.

Next, it has been shown how the bias inherent in the BPFML methods can be corrected in practice. Based on the analysis of the PFML methods in Chapter 3, the asymptotic properties of the BPFML methods have been analysed. In addition, apart from its main purpose for DOA estimation, the combined classical and LCMV beamforming procedure has been shown to provide adaptive beamforming which is robust to pointing error, as well as to the presence of correlated sources.

In Chapter 5, the IQML algorithm for efficient computation of the ML parameter estimates has been considered. The generalised form of the algorithm has been provided. The improved version of the algorithm has been proposed. By employing the consistent nontriviality constraint, the proposed algorithm (CIQML) has been shown by computer simulations to outperform two existing IQML algorithms due to Bresler & Macovski, and Nagesha & Kay (BIQML and NIQML). With the DML cost function, the proposed algorithm yields superior performance to the existing ones for the multiple data snapshot case. For the single data snapshot case, the improved performance is gained by the proposed algorithm for a short data length N , or under the worst signal phase factor case [77]. In addition to the improvement in estimation accuracy, the proposed algorithm has been found to converge faster than the two existing methods. It has also been demonstrated that, the use of the optimal WSF cost function instead of the DML cost function can greatly improve the performance of the existing BIQML and NIQML algorithms. The reason for the improvement has been pointed out.

Some of the parameter values (to be estimated) may sometimes be known a priori. The formulation of the IQML algorithm which takes advantage of this prior information has been provided. Because this prior knowledge imposes constraint on some of the parameter values, the resulting algorithm has been referred to as the

constrained IQML algorithm, also in analogy with the related constrained MUSIC method [42]. Under the deterministic signal assumption, with the described prior information of the parameter values, the CRLB of the unknown parameter estimate(s) has been provided. It has been found that the constrained IQML algorithm outperforms the unconstrained one as well as the constrained root-minimum norm method. The latter is a variant of the constrained MUSIC method.

In addition to the 1-D problem, the problem of 2-D parameter estimation has also been considered. The problem of pairing two sets of 1-D parameters has been addressed. The sufficient condition to guarantee the unique 2-D parameter pairing has been derived. Based on this result, the 2-D parameter estimation method based on the 1-D IQML algorithm has been proposed for the case of uniform-sampled observed data. The method has been demonstrated through simulation to outperform the MEMP method, when there is the multiplication of the parameter values in each of the two dimensions so that K_y and $K_z < K$. For the case of distinct parameter values in each of the two dimensions ($K_y = K_z = K$), the proposed procedure has been found to outperform the MEMP method at low SNR, with slightly inferior performance to the latter at high SNR. Under the deterministic signal assumption, the CRLB of the 2-D parameter estimate(s) for each of the two cases has also been provided.

In Chapter 6, performance study of the PFML methods has been carried out. The computer simulation under various representative cases has been performed. The DEML method with complete knowledge of the signal waveforms has been found to achieve the best result in all experiment. It has been shown that the proposed BPFML-2 method outperforms the DML and WSF methods under the scenario consisting of multiple clusters of coherent signals, provided that the angular spacing between any two clusters is larger than one array beamwidth. For the case where the signals within the same cluster are only partially correlated, the BPFML-4+bc method has been found to achieve better performance than the DML and WSF methods. In addition, it has been

found that the BPFML-2 method achieves comparable performance to the best-case PFML-2 method, which is based on the type-1 pre-filtering reference signal. The same conclusion can be said of the BPFML-3 method and the best-case PFML-3 method. Although the BPFML methods with the bias correction and the no bias-correction counterparts have been observed to achieve comparable estimation accuracy, the use of the bias correction reduces the estimation bias to a level comparable to that of the best-case PFML methods. At low SNR with a large number of data snapshots available, the method with the bias correction yields more accurate DOA estimates than the method without the bias correction. In addition, the bias correction has been found to be essential for the BPFML-4 method.

In addition to its superior performance to the standard DML and WSF methods, the BPFML-2 method has been shown to be capable of estimating more source directions than those which can be estimated by the DML and WSF methods. In addition, it is capable of estimating source directions in an unknown spatially correlated noise field. These properties are also shared by the BPFML-4 method (with or without bias correction). In addition, because both the BPFML-2 and BPFML-4 methods (with or without bias correction) separately estimate the directions of two (or more) sources from different clusters, the two methods can be implemented by means of parallel processing.

Validation of the analytical result, which was based on the asymptotic analytical study in Chapters 3 and 4, has also been carried out in Chapter 6 using computer simulation. Both analytical and simulation results have been found to be closely matched in most studied scenarios. The asymptotic analytical and estimated RMSEs of the DOA estimates are in good agreement for the number of snapshots as low as 20 at the SNR range of 5 - 10 dB.

In Chapter 7, the problem of source direction estimation using an irregular uniform linear array has been considered. The array interpolation approach has been

taken to circumvent the problem. The new interpolated array design procedure has been proposed. The procedure is the modification of the least squares procedure of Friedlander [25]. Unlike the approach in [25], however, the proposed procedure introduces a weighting function into a design equation to reduce the maximum interpolation error. In addition, for the proposed procedure, a mild constraint is imposed onto the out-of-sector directivity gain of the interpolated array. It has been shown that, despite higher interpolation error, the proposed procedure results in lower estimation bias, when compared with the result based on the procedure in [25]. In addition, the estimation bias due to the interpolation error has been calculated for two interpolated arrays, with and without Step 4 of the proposed design procedure. It has been found that, among the two arrays, the interpolated array obtained by including Step 4 into the design procedure has lower maximum absolute bias value for the single source case. For a more complicated case of two coherent sources, however, it has not yet been conclusively shown which one of the two arrays should be preferred. In all of the described bias study, the result obtained by direct evaluation of the asymptotic DML and optimal WSF cost functions has been compared with the result obtained by the asymptotic analytical bias formula. The latter has been adapted from the bias formula derived in Chapter 3 for the type-4 PFML methods. It has been found that both results are in good agreement for all of the study cases.

The performance of source direction estimation using the IQML algorithm with the interpolated uniform linear array has been investigated. It has been found that, by using the IQML algorithm, the DOA estimates based on the interpolated array are more accurate than the ones with the real irregular array. In addition, the out-of-sector-signal problem has been addressed. It has been demonstrated that, by using the IQML algorithm with the interpolated uniform linear array, the WSF method suffers performance degradation in the presence of the signal(s) whose angular location lies outside the ISOV. The reason for this performance degradation has been explained. For

the case where the out-of-sector signal(s) is uncorrelated with the signals whose directions lie within the ISOV, this problem can be avoided by using the BPFML-2 method instead. Through computer simulation study, the latter method has been found to outperform the WSF method in estimating source directions within the ISOV, with the presence of the uncorrelated out-of-sector signal.

8.2 Future work

The following remaining issues are suggested for future work.

- The detection problem of estimating the number of sources has been entirely omitted in this thesis. As well as the DOA estimates, the BPFML methods also give the estimate of the noise covariance matrix. This suggests the possibility of estimating the number of sources in an unknown spatially correlated noise field. In fact, for some special type-2 PFML methods, the detectors based on the IV-based DOA estimators have already been available for estimating the number of sources in the unknown noise field condition (see Chapter 6 of [47]). It remains to be seen, however, which performance level can be achieved by a similar detector based instead on the BPFML methods.
- The proposed pre-filtering approach has so far been applied with the DOA estimators which are based on the ML principle. From a computational point of view, it may be advantageous to apply instead a sub-optimal high resolution DOA estimator such as MUSIC for the DOA estimator part of the pre-filtering framework (see Figure 3.1).
- The performance analysis of the CIQML has not yet been considered. Although the convergence analysis of the related Steiglitz-McBride algorithm has been considered in [91-92], similar study for the IQML algorithm is yet to be done. The availability of the convergence analysis for the three algorithms will allow

for a more thorough comparison, and may lead to further insight and possible improvement.

- The CIQML algorithm has been available in a batch form. The parameter estimation is carried out using a set of data snapshots under the assumption that the observed signal is stationary over the observation period. The availability of the algorithm in a recursive form, where the parameter estimation is carried out for each new available data snapshot, is of interest in two aspects. First, this may lead to a more computationally efficient algorithm for real time implementation. Secondly, it may improve the estimation performance of the algorithm, when applied to the parameter tracking problem under a highly non-stationary environment.
- As observed in one of the two experiments relating to the 2-D parameter estimation problem, detailed in Chapter 5, the RMSEs of the parameter estimates obtained by the proposed 1-D based 2-D parameter estimator did not attain the CRLB even at comparatively high SNR. This is due to the 1-D nature of the method. It is thus suggested that improvement may be obtained by continuing the parameter estimation using an exact 2-D estimation method, based on the initial parameter values obtained from the proposed 1-D based method.
- A possible improvement to the proposed interpolated array design procedure should be considered. As seen in Chapter 7, the interpolation error as defined in the context may not be a good criterion in assessing the quality of the interpolated array. A natural criterion should be to choose the interpolation matrix so as to minimise the DOA estimation bias. The interpolated array design based on this criterion may be possible through the use of the asymptotic analytical bias formula Eq. (7.11). Note, however, that the straightforward use of Eq. (7.11) in the interpolated array design, e.g., by minimising the 2-norm of the

bias vector given by Eq. (7.11) with respect to the interpolation matrix $\mathbf{G}$, leads to a highly complicated optimisation problem. Thus, future work should concentrate on a simplification to this direct approach.

- From both simulation and analytical studies, the proposed BPFML-2 and BPFML-4 methods have been shown to yield superior performance to the existing DML and WSF methods. The performance of the methods in a real environment, where various modelling errors generally exist, remains to be assessed. However, due to the applicability of the methods under an unknown spatially correlated noise field, and in the case where the number of sources exceeds the number of sensors, the methods are expected to be less susceptible to the related modelling errors than the DML and WSF methods.

Appendices

A Partitioned complex Wishart matrices

Let $L\hat{\mathbf{W}}$ be an $(M + N) \times (M + N)$ random matrix with the complex Wishart distribution, described by $L\hat{\mathbf{W}} \sim CW(L, M + N, \mathbf{W})$. Partition $L\hat{\mathbf{W}}$ and $\mathbf{W}$ by the M th row and column as follows:

$$L\hat{\mathbf{W}} = \begin{bmatrix} L\hat{\mathbf{W}}_{11} & L\hat{\mathbf{W}}_{21}^H \\ L\hat{\mathbf{W}}_{21} & L\hat{\mathbf{W}}_{22} \end{bmatrix} \quad (\text{A.1})$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{21}^H \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix} \quad (\text{A.2})$$

Some results on the distributions of the partitioned matrices in Eq. (A.1) are given below [39]:

i) The $N \times N$ Schur complement

$$L\hat{\tilde{\mathbf{W}}} = L(\hat{\mathbf{W}}_{22} - \hat{\mathbf{W}}_{21}\hat{\mathbf{W}}_{11}^{-1}\hat{\mathbf{W}}_{21}^H) \quad (\text{A.3})$$

has the complex Wishart distribution $CW(L - M, N, \tilde{\mathbf{W}})$, i.e.,

$$p(L\hat{\tilde{\mathbf{W}}}) = \frac{|\hat{\tilde{\mathbf{W}}}|^{L-M-N}}{\Gamma_N^f(L-M)|\tilde{\mathbf{W}}|^{L-M}} \exp\{-L \cdot \text{tr}(\hat{\tilde{\mathbf{W}}}\tilde{\mathbf{W}}^{-1})\} \quad (\text{A.4})$$

where

$$\tilde{\mathbf{W}} = \mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{21}^H \quad (\text{A.5})$$

$$\Gamma_N^f(L-M) = \pi^{N(N-1)/2} \Gamma(L-M) \Gamma(L-M-1) \cdots \Gamma(L-M-N+1) \quad (\text{A.6})$$

and $\Gamma(\cdot)$ is the gamma function.

ii) Given $L\hat{\mathbf{W}}_{11}$, $\hat{\mathbf{W}}_{21}\hat{\mathbf{W}}_{11}^{-1}$ has the Gaussian distribution as expressed by

$$p(\hat{\mathbf{W}}_{21}\hat{\mathbf{W}}_{11}^{-1} | L\hat{\mathbf{W}}_{11}) = \frac{|L\hat{\mathbf{W}}_{11}|^N}{\pi^{MN}|\tilde{\mathbf{W}}|^M} \exp\{-L \cdot \text{tr}\{(\hat{\mathbf{W}}_{21}\hat{\mathbf{W}}_{11}^{-1} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1})\hat{\mathbf{W}}_{11}(\hat{\mathbf{W}}_{21}\hat{\mathbf{W}}_{11}^{-1} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1})^H \tilde{\mathbf{W}}^{-1}\}\} \quad (\text{A.7})$$

iii) $L\hat{\tilde{\mathbf{W}}}$ is statistically independent of $\hat{\mathbf{W}}_{21}\hat{\mathbf{W}}_{11}^{-1}$ and $L\hat{\mathbf{W}}_{11}$.

B Proof of Theorem 3.1

Proof of the theorem can be directly obtained from the results on partitioned complex Wishart matrices, summarised in Appendix A. Let $M = d_s$, $\hat{W}_{11} = \hat{R}_r = \frac{1}{L} \mathbf{M}_r^H \mathbf{M}_r$, $\hat{W}_{21} \hat{W}_{11}^{-1} = Y W_{\text{pf}}^{-1} \hat{R}_r^{-1}$, $\hat{W}_{22} = \hat{R}$, and $\hat{\tilde{W}} = \hat{\tilde{R}}_n = \hat{R} - Y W_{\text{pf}}^{-1} \hat{R}_r^{-1} W_{\text{pf}}^{-1} Y^H$. Thus, from i), ii), and iii) in Appendix A, the joint PDF of $Y W_{\text{pf}}^{-1} \hat{R}_r^{-1}$ and $L \hat{\tilde{R}}$, given $\hat{R}_r$, is

$$\begin{aligned}
 p(L \hat{\tilde{R}}_n, Y W_{\text{pf}}^{-1} \hat{R}_r^{-1} | \hat{R}_r) &= p(L \hat{\tilde{R}}_n) p(Y W_{\text{pf}}^{-1} \hat{R}_r^{-1} | \hat{R}_r) \\
 &= \frac{|\hat{L} \hat{R}_r|^N |\hat{L} \hat{\tilde{R}}_n|^{L-N-d_s}}{\pi^{d_s N} \Gamma_N^f(L-d_s) |\tilde{R}_n|^L} \exp\{-L \cdot \{ \text{tr}\{(Y-AT) W_{\text{pf}}^{-1} \hat{R}_r^{-1} W_{\text{pf}}^{-1} (Y-AT)^H \tilde{R}_n^{-1}\} + \text{tr}\{\hat{\tilde{R}}_n \tilde{R}_n^{-1}\} \}\} \\
 &= \frac{|\hat{L} \hat{R}_r|^N |\hat{L} \hat{\tilde{R}}_n|^{L-N-d_s}}{\pi^{d_s N} \Gamma_N^f(L-d_s) |\tilde{R}_n|^L} \exp\{-\text{tr}\{ \{ (Y-AT) W_{\text{pf}}^{-1} \hat{R}_r^{-1} W_{\text{pf}}^{-1} (Y-AT)^H + \hat{\tilde{R}}_n \} \cdot L \tilde{R}_n^{-1} \}\} \\
 & \tag{B.1}
 \end{aligned}$$

where

$$T = \mathbf{R}_{\text{sr}} \mathbf{R}_r^{-1} \hat{R}_r W_{\text{pf}} \tag{B.2}$$

and $\mathbf{R}_{\text{sr}} = E\{s(t) \mathbf{r}^H(t)\}$. From Eq. (B.1), the likelihood function Eq. (3.19) follows. In addition, by taking logarithms of both sides of Eq. (3.19) and neglecting unrelated terms, we arrive at

$$\begin{aligned}
 \ell(Y, \hat{\tilde{R}}_n, \hat{R}_r, W_{\text{pf}}; \theta, \tilde{R}_n, T) &= \\
 &= -\ln|\tilde{R}_n| - \text{tr}\{ \{ (Y-AT) W_{\text{pf}}^{-1} \hat{R}_r^{-1} W_{\text{pf}}^{-1} (Y-AT)^H + \hat{\tilde{R}}_n \} \tilde{R}_n^{-1} \} \tag{B.3}
 \end{aligned}$$

The function $\ell(\cdot)$ is maximised with respect to $\tilde{R}_n$ when [39]

$$\tilde{R}_n = (Y-AT) W_{\text{pf}}^{-1} \hat{R}_r^{-1} W_{\text{pf}}^{-1} (Y-AT)^H + \hat{\tilde{R}}_n \tag{B.4}$$

By substituting Eq. (B.4) into Eq. (B.3), we arrived at Eq. (3.23).

C Proof of Theorem 3.2

The proof is based on the ideas of related proofs in [9, 16], and uses some results on the derivatives of the projection matrix in [33]. First, consider the DEML cost function[†]

$$f_{1;k}(\theta_k) = \text{tr} \{ \mathbf{P}_{\hat{\tilde{\mathbf{R}}}_n^{-1/2} \mathbf{A}_k}^\perp \hat{\tilde{\mathbf{R}}}_n^{-1/2} \mathbf{y}_k \mathbf{y}_k^H \hat{\tilde{\mathbf{R}}}_n^{-1/2} \}, \quad k = 1, \dots, d_s \quad (\text{C.1})$$

where $\hat{\tilde{\mathbf{R}}}_n$ is as given by Eq. (3.41). Due to the fact that $\hat{\tilde{\mathbf{R}}}_n$ is a consistent estimate of $\tilde{\mathbf{R}}_n$ ($\tilde{\mathbf{R}}_n = \mathbf{R}_n$ in this case), and the assumptions A. 2.4 and A. 2.6, the estimate $\hat{\theta}_k$, which is obtained by minimising Eq. (C.1), converges with probability 1 to the true parameter value $\bar{\theta}_k$ as $L \rightarrow \infty$. In addition, because the estimate $\hat{\theta}_k$ is obtained by minimising Eq. (C.1) with respect to θ_k , the first derivative of Eq. (C.1), when evaluated at $\hat{\theta}_k$, is $\dot{f}_{1;k}(\hat{\theta}_k) \triangleq \nabla_{\theta_k} f_{1;k}(\theta_k) \Big|_{\hat{\theta}_k} = \mathbf{0}$. Because $\hat{\theta}_k$ is a consistent estimate of θ_k , for large L , the first-order Taylor series expansion of $\dot{f}_{1;k}(\theta_k)$ around the true parameter value $\bar{\theta}_k$, when evaluated at $\hat{\theta}_k$, results in

$$\mathbf{0} \simeq \dot{f}_{1;k}(\bar{\theta}_k) + \ddot{F}_{1;k}(\bar{\theta}_k)(\hat{\theta}_k - \bar{\theta}_k) \quad (\text{C.2})$$

where $\ddot{F}_{1;k}(\bar{\theta}_k) \triangleq \nabla_{\theta_k}^2 f_{1;k}(\theta_k) \Big|_{\bar{\theta}_k}$, and the higher order terms are neglected. To obtain the distribution of $(\hat{\theta}_k - \bar{\theta}_k)$, the terms $\dot{f}_{1;k}(\bar{\theta}_k)$ and $\ddot{F}_{1;k}(\bar{\theta}_k)$ need to be evaluated as follows.

Let $\theta_{k,\eta}$ be the η th element of θ_k . The first derivative of $f_{1;k}(\theta_k)$ with respect to $\theta_{k,\eta}$, when evaluated at $\bar{\theta}_k$, can be expressed as

$$\begin{aligned} [\dot{f}_{1;k}(\bar{\theta}_k)]_\eta &= \text{tr} \{ \dot{\mathbf{P}}_{\hat{\tilde{\mathbf{R}}}_n^{-1/2} \mathbf{A}_k}^\perp (\theta_{k,\eta}, \bar{\theta}_k) \hat{\tilde{\mathbf{R}}}_n^{-1/2} \hat{\mathbf{y}}_k \hat{\mathbf{y}}_k^H \hat{\tilde{\mathbf{R}}}_n^{-1/2} \} \\ &= -2 \cdot \text{Re} \{ \hat{\mathbf{y}}_k^H \hat{\tilde{\mathbf{R}}}_n^{-1/2} \dot{\mathbf{P}}_{\hat{\tilde{\mathbf{R}}}_n^{-1/2} \mathbf{A}_k}^\perp \hat{\tilde{\mathbf{R}}}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} (\bar{\mathbf{A}}_k^H \hat{\tilde{\mathbf{R}}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \hat{\tilde{\mathbf{R}}}_n^{-1} \hat{\mathbf{y}}_k \} \end{aligned} \quad (\text{C.3})$$

where

[†] The cost function Eq. (C.1) uses $\hat{\tilde{\mathbf{R}}}_n$ instead of $\mathbf{R}_{ns}$ in Eq. (3.39). However, as will be seen, both cost functions yield the same asymptotic results.

$$\begin{aligned}
\dot{\mathbf{P}}_{\hat{\mathbf{R}}_n^{-1/2} \mathbf{A}_k}^\perp(\theta_{k,\eta}, \bar{\boldsymbol{\theta}}_k) &= \left. \frac{\partial \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2} \mathbf{A}_k}^\perp}{\partial \theta_{k,\eta}} \right|_{\bar{\boldsymbol{\theta}}_k} \\
&= - \{ \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1/2} + \\
&\quad (\mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1/2})^H \} \quad (\text{C.4})
\end{aligned}$$

and $\bar{\mathbf{A}}_{k,\eta} = \left. \frac{\partial \mathbf{A}_k}{\partial \theta_{k,\eta}} \right|_{\bar{\boldsymbol{\theta}}_k}$. Next, observe that $\text{vec}(\hat{\mathbf{Y}})$ is Gaussian distributed, with mean $\text{vec}(\bar{\mathbf{A}} \bar{\boldsymbol{\Gamma}}_s)$ and the covariance matrix $(\hat{\mathbf{R}}_{\text{nc}}^{-T} \otimes \mathbf{R}_n)/L$. Because $\mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k = \mathbf{0}$, and $\hat{\mathbf{R}}_n = \mathbf{R}_n + O_p(1/\sqrt{L})^\dagger$, we obtain

$$\begin{aligned}
[\dot{f}_{1;k}(\bar{\boldsymbol{\theta}}_k)]_\eta &= -2 \cdot \text{Re}\{(\hat{\mathbf{y}}_k - \bar{\mathbf{A}}_k \bar{\boldsymbol{\gamma}}_k)^H \hat{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k \bar{\boldsymbol{\gamma}}_k\} \\
&\quad + O_p(1/L) \\
&= -2 \cdot \text{Re}\{(\hat{\mathbf{y}}_k - \bar{\mathbf{A}}_k \bar{\boldsymbol{\gamma}}_k)^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} \bar{\boldsymbol{\gamma}}_k\} + O_p(1/L) \\
&\simeq -2 \cdot \text{Re}\{(\hat{\mathbf{y}}_k - \bar{\mathbf{A}}_k \bar{\boldsymbol{\gamma}}_k)^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} \bar{\boldsymbol{\gamma}}_k\} \quad (\text{C.5})
\end{aligned}$$

where, in the last line, we neglect the term with the order $O_p(1/L)$.

The second derivative matrix $\ddot{F}_{1;k}(\bar{\boldsymbol{\theta}}_k)$, evaluated at $\bar{\boldsymbol{\theta}}_k$, is next considered. Its $\eta\xi^{\text{th}}$ element is given by

$$\begin{aligned}
[\ddot{F}_{1;k}(\bar{\boldsymbol{\theta}}_k)]_{\eta,\xi} &= \text{tr}\{ \ddot{\mathbf{P}}_{\hat{\mathbf{R}}_n^{-1/2} \mathbf{A}_k}^\perp(\theta_{k,\eta}, \theta_{k,\xi}, \bar{\boldsymbol{\theta}}_k) \hat{\mathbf{R}}_n^{-1/2} \hat{\mathbf{y}}_k \hat{\mathbf{y}}_k^H \hat{\mathbf{R}}_n^{-1/2} \} \\
&= 2 \cdot \text{Re}\{ \text{tr}\{ \bar{\mathbf{A}}_{k,\xi}^H \hat{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \hat{\mathbf{y}}_k \hat{\mathbf{y}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \} \\
&\quad + \text{tr}\{ \mathbf{Q}_r \hat{\mathbf{R}}_n^{-1/2} \hat{\mathbf{y}}_k \hat{\mathbf{y}}_k^H \hat{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \} \} \quad (\text{C.6})
\end{aligned}$$

$$\text{where } \ddot{\mathbf{P}}_{\hat{\mathbf{R}}_n^{-1/2} \mathbf{A}_k}^\perp(\theta_{k,\eta}, \theta_{k,\xi}, \bar{\boldsymbol{\theta}}_k) = \left. \frac{\partial^2 \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2} \mathbf{A}_k}^\perp}{\partial \theta_{k,\eta} \partial \theta_{k,\xi}} \right|_{\bar{\boldsymbol{\theta}}_k},$$

[†] See [44] for discussion of the order symbols $O_p(\cdot)$ and $o_p(\cdot)$.

$$\begin{aligned}
Q_r = & \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{k,\xi} (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_{k,\eta} (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1/2} - \\
& \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta,\xi} (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1/2} - \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_{k,\xi}^H \hat{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp + \\
& \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_{k,\xi} (\bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \hat{\mathbf{R}}_n^{-1/2}
\end{aligned} \tag{C.7}$$

and $\bar{\mathbf{A}}_{k,\eta,\xi} = \left. \frac{\partial^2 \mathbf{A}_k}{\partial \theta_{k,\eta} \partial \theta_{k,\xi}} \right|_{\bar{\boldsymbol{\theta}}_k}$. Because $(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k) \rightarrow \mathbf{0}$ as $L \rightarrow \infty$, and $[\dot{\mathbf{f}}_{1;k}(\bar{\boldsymbol{\theta}}_k)]_\eta$ is of the order $O_p(1/\sqrt{L})$,

$$\begin{aligned}
[\ddot{\mathbf{F}}_{1;k}(\bar{\boldsymbol{\theta}}_k)]_{\eta,\xi} & \simeq 2 \cdot \text{Re} \{ \text{tr} \{ \bar{\mathbf{A}}_{k,\xi}^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} (\bar{\mathbf{A}}_k^H \mathbf{R}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \bar{\mathbf{A}}_k^H \mathbf{R}_n^{-1} \bar{\mathbf{A}}_k \bar{\boldsymbol{\gamma}}_k \bar{\boldsymbol{\gamma}}_k^H \bar{\mathbf{A}}_k^H \mathbf{R}_n^{-1} \bar{\mathbf{A}}_k (\bar{\mathbf{A}}_k^H \mathbf{R}_n^{-1} \bar{\mathbf{A}}_k)^{-1} \} \} \\
& = 2 \cdot \text{Re} \{ \text{tr} \{ \bar{\mathbf{A}}_{k,\xi}^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} \bar{\boldsymbol{\gamma}}_k \bar{\boldsymbol{\gamma}}_k^H \} \}
\end{aligned} \tag{C.8}$$

In Eq. (C.8), the term of the order $O_p(1/\sqrt{L})$ is neglected. From Eq. (C.8),

$$\ddot{\mathbf{F}}_{1,k}(\bar{\boldsymbol{\theta}}_k) = 2 \cdot \text{Re} \{ (\bar{\mathbf{D}}_k^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{D}}_k) \odot (\bar{\boldsymbol{\gamma}}_k \bar{\boldsymbol{\gamma}}_k^H)^T \} \tag{C.9}$$

where

$$\bar{\mathbf{D}}_k = \begin{bmatrix} \left. \frac{\partial \mathbf{a}_{k;1}}{\partial \theta_{k;1}} \right|_{\bar{\boldsymbol{\theta}}_{k;1}} & \left. \frac{\partial \mathbf{a}_{k;2}}{\partial \theta_{k;2}} \right|_{\bar{\boldsymbol{\theta}}_{k;2}} & \dots & \left. \frac{\partial \mathbf{a}_{k;K_k}}{\partial \theta_{k;K_k}} \right|_{\bar{\boldsymbol{\theta}}_{k;K_k}} \end{bmatrix} \tag{C.10}$$

From Eq. (C.2), we obtain

$$\mathbf{E} \{ (\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k)(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k)^T \} = \ddot{\mathbf{F}}_{1;k}^{-1}(\bar{\boldsymbol{\theta}}_k) \dot{\mathbf{F}}_{1;k,k}(\bar{\boldsymbol{\theta}}_k) \ddot{\mathbf{F}}_{1;k}^{-1}(\bar{\boldsymbol{\theta}}_k) \tag{C.11}$$

where $[\dot{\mathbf{F}}_{1;k,k}(\bar{\boldsymbol{\theta}}_k)]_{\eta,\xi} = \mathbf{E} \{ [\dot{\mathbf{f}}_{1;k}(\bar{\boldsymbol{\theta}}_k)]_\eta [\dot{\mathbf{f}}_{1;k}(\bar{\boldsymbol{\theta}}_k)]_\xi^T \}$. From Eq. (C.5), we obtain

$$\begin{aligned}
[\dot{\mathbf{F}}_{1;k,k}(\bar{\boldsymbol{\theta}}_k)]_{\eta,\xi} & = \\
2 \cdot \text{Re} \{ \mathbf{E} \{ \bar{\boldsymbol{\gamma}}_k^H \bar{\mathbf{A}}_{k,\xi}^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} (\mathbf{y}_k - \bar{\mathbf{A}}_k \bar{\boldsymbol{\gamma}}_k)(\mathbf{y}_k - \bar{\mathbf{A}}_k \bar{\boldsymbol{\gamma}}_k)^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} \bar{\boldsymbol{\gamma}}_k \} \} \\
& = \frac{2}{L} \cdot \text{Re} \{ [\hat{\mathbf{R}}_{\text{nc}}^{-1}]_{k,k} \bar{\boldsymbol{\gamma}}_k^H \bar{\mathbf{A}}_{k,\xi}^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_{k,\eta} \bar{\boldsymbol{\gamma}}_k \}
\end{aligned} \tag{C.12}$$

where we have used the fact that $E\{(\mathbf{y}_k - \bar{\mathbf{A}}_k \bar{\boldsymbol{\gamma}}_k)(\mathbf{y}_k - \bar{\mathbf{A}}_k \bar{\boldsymbol{\gamma}}_k)^T\} = \mathbf{0}$ and the identity $\text{Re}(z_1)\text{Re}(z_2) = \text{Re}(z_1 z_2^* + z_1 z_2)/2$ for any two complex numbers z_1 and z_2 . From Eq. (C.12), we obtain

$$\dot{\mathbf{F}}_{1;k,k}(\bar{\boldsymbol{\theta}}_k) = \frac{2[\hat{\mathbf{R}}_{\text{nc}}^{-1}]_{k,k}}{L} \text{Re}\{(\bar{\mathbf{D}}_k^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{D}}_k) \odot (\bar{\boldsymbol{\gamma}}_k \bar{\boldsymbol{\gamma}}_k^H)^T\} \quad (\text{C.13})$$

With the fact that $\mathbf{y}_k$ is Gaussian distributed, and from Eqs. (C.2, C.5, C.9, C.11, C.13), the estimation error $(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k)$ is asymptotically Gaussian distributed with zero mean and the error covariance matrix

$$E\{(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k)(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k)^T\} = \frac{[\mathbf{R}_{\text{nc}}^{-1}]_{k,k}}{2L} \text{Re}^{-1}\{(\bar{\mathbf{D}}_k^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \bar{\mathbf{A}}_k}^\perp \mathbf{R}_n^{-1/2} \bar{\mathbf{D}}_k) \odot (\bar{\boldsymbol{\gamma}}_k \bar{\boldsymbol{\gamma}}_k^H)^T\} \quad (\text{C.14})$$

where $[\hat{\mathbf{R}}_{\text{nc}}^{-1}]_{k,k}$ can be replaced by $[\mathbf{R}_{\text{nc}}^{-1}]_{k,k}$ without affecting the asymptotic results because $\hat{\mathbf{R}}_{\text{nc}}^{-1} = \mathbf{R}_{\text{nc}}^{-1} + O_p(1/\sqrt{L})$.

D Proof of Theorem 3.3

The PFML-2 cost function is given by

$$f_{2;k}(\theta_k) = \text{tr}\{P_{\hat{\tilde{R}}_{n;k}^{-1/2}A_k}^\perp \hat{\tilde{R}}_{n;k}^{-1/2} y_k y_k^H \hat{\tilde{R}}_{n;k}^{-1/2}\} \quad (\text{D.1})$$

By substituting $f_{2;k}(\theta_k)$ for $f_{1;k}(\theta_k)$ in Proof of Theorem 3.2 and following the same arguments in that proof, we obtain

$$\begin{aligned} [\dot{f}_{2;k}(\bar{\theta}_k)]_\eta &= -2 \cdot \text{Re}\{(\hat{y}_k - \bar{A}_k \bar{t}_k)^H \hat{\tilde{R}}_{n;k}^{-1/2} P_{\hat{\tilde{R}}_{n;k}^{-1/2} \bar{A}_k}^\perp \hat{\tilde{R}}_{n;k}^{-1/2} \bar{A}_{k,\eta} (\bar{A}_k^H \hat{\tilde{R}}_{n;k}^{-1} \bar{A}_k)^{-1} \bar{A}_k^H \hat{\tilde{R}}_{n;k}^{-1} \bar{A}_k \bar{t}_k\} + \\ &\quad O_p(1/L) \\ &= -2 \cdot \text{Re}\{(\hat{y}_k - \bar{A}_k \bar{t}_k)^H \bar{\tilde{R}}_{n;k}^{-1/2} P_{\bar{\tilde{R}}_{n;k}^{-1/2} \bar{A}_k}^\perp \bar{\tilde{R}}_{n;k}^{-1/2} \bar{A}_{k,\eta} \bar{t}_k\} + O_p(1/L) \\ &\simeq -2 \cdot \text{Re}\{(\hat{y}_k - \bar{A}_k \bar{t}_k)^H \bar{\tilde{R}}_{n;k}^{-1/2} P_{\bar{\tilde{R}}_{n;k}^{-1/2} \bar{A}_k}^\perp \bar{\tilde{R}}_{n;k}^{-1/2} \bar{A}_{k,\eta} \bar{t}_k\} \end{aligned} \quad (\text{D.2})$$

and

$$\ddot{F}_{2;k}(\bar{\theta}_k) \simeq 2 \cdot \text{Re}\{(\bar{D}_k^H \bar{\tilde{R}}_{n;k}^{-1/2} P_{\bar{\tilde{R}}_{n;k}^{-1/2} \bar{A}_k}^\perp \bar{\tilde{R}}_{n;k}^{-1/2} \bar{D}_k) \odot (\bar{t}_k \bar{t}_k^H)^T\} \quad (\text{D.3})$$

$$\bar{t}_k = \bar{\gamma}_k [R_{nc}]_{k,k} [R_{r_2}]_{k,k}^{-1} [\hat{R}_{r_2}]_{k,k}^{1/2} \quad (\text{D.4})$$

where $\dot{f}_{2;k}(\bar{\theta}_k) \triangleq \nabla_{\theta_k} f_{2;k}(\theta_k) \Big|_{\bar{\theta}_k}$, $\ddot{F}_{2;k}(\bar{\theta}_k) \triangleq \nabla_{\theta_k}^2 f_{2;k}(\theta_k) \Big|_{\bar{\theta}_k}$; $\bar{D}_k$ and $\bar{\tilde{R}}_{n;k}$ are given, respectively, by Eq. (C.10) and Eq. (3.54). By following Proof of Theorem 3.2, it is seen that $(\hat{\theta}_k - \bar{\theta}_k)$ is zero mean Gaussian distributed with the covariance matrix

$$\text{E}\{(\hat{\theta}_k - \bar{\theta}_k)(\hat{\theta}_k - \bar{\theta}_k)^T\} = \frac{1}{2L} \text{Re}^{-1}\{(\bar{D}_k^H \bar{\tilde{R}}_{n;k}^{-1/2} P_{\bar{\tilde{R}}_{n;k}^{-1/2} \bar{A}_k}^\perp \bar{\tilde{R}}_{n;k}^{-1/2} \bar{D}_k) \odot (\bar{t}_k \bar{t}_k^H)^T\} \quad (\text{D.5})$$

Because, for large L , $\hat{R}_{r_2}$ can be replaced by R_{r_2} ,

$$\begin{aligned} \bar{t}_k &= \bar{\gamma}_k [R_{nc}]_{k,k} [R_{r_2}]_{k,k}^{-1/2} \\ &= \bar{\gamma}_k [R_{nc}]_{k,k}^{1/2} \text{corr}_{r_2;k} \end{aligned} \quad (\text{D.6})$$

where

$$\begin{aligned}
\text{corr}_{\mathbf{r}_2;k} &= \left| E\{s_{\text{nc};k}(t)r_{2;k}^*(t)\} \right| [\mathbf{R}_{\text{nc}}]_{k,k}^{-1/2} [\mathbf{R}_{\mathbf{r}_2}]_{k,k}^{-1/2} \\
&= [\mathbf{R}_{\text{nc}}]_{k,k}^{1/2} [\mathbf{R}_{\mathbf{r}_2}]_{k,k}^{-1/2}
\end{aligned} \tag{D.7}$$

From Eqs. (D.5-D.6), we arrive at Eq. (3.59).

E Proof of Theorem 3.4

The PFML-3 cost function is given by

$$f_3(\boldsymbol{\theta}) = \text{tr}\{ \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2}\mathbf{A}}^\perp \hat{\mathbf{R}}_n^{-1/2} \mathbf{Y} \mathbf{Y}^H \hat{\mathbf{R}}_n^{-1/2} \} \quad (\text{E.1})$$

Because the estimate $\hat{\boldsymbol{\theta}}$, obtained by minimising Eq. (E.1), converges to a true parameter value $\bar{\boldsymbol{\theta}}$ as $L \rightarrow \infty$, and $\dot{f}_3(\hat{\boldsymbol{\theta}}) \triangleq \nabla_{\boldsymbol{\theta}} f_3(\boldsymbol{\theta}) \Big|_{\hat{\boldsymbol{\theta}}} = \mathbf{0}$, the first-order Taylor series expansion of $\dot{f}_3(\boldsymbol{\theta})$ around a true parameter value $\bar{\boldsymbol{\theta}}$, and evaluated at $\hat{\boldsymbol{\theta}}$, results in

$$\mathbf{0} \simeq \dot{f}_3(\bar{\boldsymbol{\theta}}) + \ddot{F}_3(\bar{\boldsymbol{\theta}}) \cdot (\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) \quad (\text{E.2})$$

where $\ddot{F}_3(\bar{\boldsymbol{\theta}}) \triangleq \nabla_{\boldsymbol{\theta}}^2 f_3(\boldsymbol{\theta}) \Big|_{\bar{\boldsymbol{\theta}}}$. With respect to θ_η , the first derivative of $f_3(\boldsymbol{\theta})$, evaluated at $\bar{\boldsymbol{\theta}}$, can be expressed as

$$\left[\dot{f}_3(\bar{\boldsymbol{\theta}}) \right]_\eta = -2 \cdot \text{Re}\{ \text{tr}\{ \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2}\bar{\mathbf{A}}}^\perp \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_\eta (\bar{\mathbf{A}}^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}})^{-1} \bar{\mathbf{A}}^H \hat{\mathbf{R}}_n^{-1} \mathbf{Y} \mathbf{Y}^H \hat{\mathbf{R}}_n^{-1/2} \} \} \quad (\text{E.3})$$

where $\bar{\mathbf{A}}_\eta \triangleq \frac{\partial \mathbf{A}}{\partial \theta_\eta} \Big|_{\bar{\boldsymbol{\theta}}}$. By following similar arguments used in Proof of Theorem 3.2 (Eqs. (C.5-C.9)), we obtain

$$\left[\dot{f}_3(\bar{\boldsymbol{\theta}}) \right]_\eta \simeq -2 \cdot \text{Re}\{ \text{tr}\{ (\mathbf{Y} - \bar{\mathbf{A}} \bar{\mathbf{T}})^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2}\bar{\mathbf{A}}}^\perp \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_\eta \bar{\mathbf{T}} \} \} \quad (\text{E.4})$$

and

$$\left[\ddot{F}_3(\bar{\boldsymbol{\theta}}) \right]_{\eta, \xi} \simeq 2 \cdot \text{Re}\{ \text{tr}\{ \bar{\mathbf{A}}_\xi^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2}\bar{\mathbf{A}}}^\perp \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_\eta \bar{\mathbf{T}} \bar{\mathbf{T}}^H \} \} \quad (\text{E.5})$$

where

$$\bar{\mathbf{T}} = \bar{\Gamma}_s \mathbf{R}_{\text{nc}} \tilde{\mathbf{A}}^H \mathbf{R}_{r_3}^{-1/2} \quad (\text{E.6})$$

$$\bar{\mathbf{R}}_n = \mathbf{R} - (\bar{\mathbf{A}} \bar{\mathbf{T}})(\bar{\mathbf{A}} \bar{\mathbf{T}})^H \quad (\text{E.7})$$

As a result,

$$\ddot{F}_3(\bar{\boldsymbol{\theta}}) = 2 \cdot \text{Re}\{ (\bar{\mathbf{D}}^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2}\bar{\mathbf{A}}}^\perp \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{D}}) \odot (\bar{\mathbf{T}} \bar{\mathbf{T}}^H)^T \} \quad (\text{E.8})$$

where

$$\bar{\mathbf{D}} = \begin{bmatrix} \left. \frac{\partial a_1}{\partial \theta_1} \right|_{\bar{\theta}_1} & \left. \frac{\partial a_2}{\partial \theta_2} \right|_{\bar{\theta}_2} & \cdots & \left. \frac{\partial a_K}{\partial \theta_K} \right|_{\bar{\theta}_K} \end{bmatrix} \quad (\text{E.9})$$

In addition, from

$$\begin{aligned} [\dot{\mathbf{F}}_3(\bar{\boldsymbol{\theta}})]_{\eta, \xi} &= \text{E}\{[\dot{\mathbf{f}}_3(\bar{\boldsymbol{\theta}})]_{\eta}[\dot{\mathbf{f}}_3(\bar{\boldsymbol{\theta}})]_{\xi}\} \\ &= 2 \cdot \text{Re}\{ \text{E}\{ \text{tr}\{ \bar{\mathbf{T}}^H \bar{\mathbf{A}}_{\xi}^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^{\perp} \bar{\mathbf{R}}_n^{-1/2} (\mathbf{Y} - \bar{\mathbf{A}} \bar{\mathbf{T}}) \} \cdot \\ &\quad \text{tr}\{ (\mathbf{Y} - \bar{\mathbf{A}} \bar{\mathbf{T}})^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^{\perp} \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{\eta} \bar{\mathbf{T}} \} \} \} \\ &= 2 \cdot \text{Re}\{ \text{E}\{ \{ \sum_{k=1}^{d_s} [\bar{\mathbf{T}}]_k^H \bar{\mathbf{A}}_{\xi}^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^{\perp} \bar{\mathbf{R}}_n^{-1/2} (\mathbf{y}_k - \bar{\mathbf{A}} [\bar{\mathbf{T}}]_k) \} \cdot \\ &\quad \{ \sum_{k=1}^{d_s} (\mathbf{y}_k - \bar{\mathbf{A}} [\bar{\mathbf{T}}]_k)^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^{\perp} \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{\eta} [\bar{\mathbf{T}}]_k \} \} \} \\ &= \frac{2}{L} \cdot \text{Re}\{ \text{tr}\{ \bar{\mathbf{T}}^H \bar{\mathbf{A}}_{\xi}^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^{\perp} \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{R}}_n \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^{\perp} \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{\eta} \bar{\mathbf{T}} \} \} \\ &= \frac{2}{L} \cdot \text{Re}\{ \text{tr}\{ \bar{\mathbf{A}}_{\xi}^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^{\perp} \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_{\eta} \bar{\mathbf{T}} \bar{\mathbf{T}}^H \} \} \end{aligned} \quad (\text{E.10})$$

we obtain

$$\dot{\mathbf{F}}_3(\bar{\boldsymbol{\theta}}) = \frac{2}{L} \cdot \text{Re}\{ (\bar{\mathbf{D}}^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^{\perp} \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{D}}) \odot (\bar{\mathbf{T}} \bar{\mathbf{T}}^H)^T \} \quad (\text{E.11})$$

From Eqs. (E.2, E.4, E.8, E.11), and that $\text{vec}(\mathbf{Y})$ is Gaussian distributed, the estimation error $(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})$ is asymptotically Gaussian distributed, with zero mean and the error covariance matrix

$$\begin{aligned} \text{E}\{ (\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^T \} &= \ddot{\mathbf{F}}_3^{-1}(\bar{\boldsymbol{\theta}}) \dot{\mathbf{F}}_3(\bar{\boldsymbol{\theta}}) \ddot{\mathbf{F}}_3^{-1}(\bar{\boldsymbol{\theta}}) \\ &= \frac{1}{2L} \text{Re}^{-1}\{ (\bar{\mathbf{D}}^H \bar{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}}^{\perp} \bar{\mathbf{R}}_n^{-1/2} \bar{\mathbf{D}}) \odot (\bar{\mathbf{T}} \bar{\mathbf{T}}^H)^T \} \end{aligned} \quad (\text{E.12})$$

F Proof of Theorem 3.5

The PFML-3 cost function is given by

$$f_3(\theta) = \text{tr}(\mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2}\mathbf{A}}^\perp \hat{\mathbf{R}}_n^{-1/2} \mathbf{Y} \mathbf{Y}^H \hat{\mathbf{R}}_n^{-1/2}) \quad (\text{F.1})$$

Let $\tilde{\theta} = (\hat{\theta} - \bar{\theta})$, where $\hat{\theta}$ minimises Eq. (F.1), the additive noise term (in $\mathbf{Y}$) $\mathbf{E}_n = \tilde{\mathbf{E}}_n / \sqrt{L}$, and the bias term $\mathbf{E}_b = \tilde{\mathbf{E}}_b / \sqrt{\tilde{L}}$, where $\tilde{\mathbf{E}}_n$ and $\tilde{\mathbf{E}}_b$ are independent of L and $\tilde{L}$. The analysis is carried out under the situation where $\tilde{L} \rightarrow \infty$ with L approaching infinity faster than $\tilde{L}$. Because $\tilde{\theta} \rightarrow \mathbf{0}$ under the described condition, the first-order Taylor series expansion of $\dot{f}_3(\theta)$ around a true parameter $\bar{\theta}$, when evaluated at $\hat{\theta}$, is given by

$$\begin{aligned} \mathbf{0} &= \dot{f}_3(\hat{\theta}) \\ &\simeq \dot{f}_3(\bar{\theta}) + \ddot{f}_3(\bar{\theta})(\hat{\theta} - \bar{\theta}) \end{aligned} \quad (\text{F.2})$$

With respect to θ_η , the first derivative of $f_3(\theta)$ at $\bar{\theta}$ can be expressed as

$$\left[\dot{f}_3(\bar{\theta}) \right]_\eta = -2 \cdot \text{Re} \{ \text{tr} \{ \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2}\bar{\mathbf{A}}}^\perp \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_\eta (\bar{\mathbf{A}}^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}})^{-1} \bar{\mathbf{A}}^H \hat{\mathbf{R}}_n^{-1} \mathbf{Y} \mathbf{Y}^H \hat{\mathbf{R}}_n^{-1/2} \} \} \quad (\text{F.3})$$

Observe that, for large L , $\text{vec}(\mathbf{Y})$ is Gaussian distributed, with mean $\text{vec}(\bar{\mathbf{A}}\bar{\mathbf{T}} + \bar{\mathbf{E}}_b) + O_p(1/\sqrt{L})$ and the covariance matrix $(\mathbf{I}_{d_s} \otimes \bar{\tilde{\mathbf{R}}}_n)/L$, where $\bar{\mathbf{T}} = \bar{\Gamma}_s \mathbf{R}_{\text{nc}} \tilde{\mathbf{A}}^H \mathbf{R}_{r_4}^{-1/2}$, $\bar{\mathbf{E}}_b = \mathbf{R}_n \mathbf{Q}_n^H \mathbf{R}_{r_4}^{-1/2}$, and $\bar{\tilde{\mathbf{R}}}_n = \mathbf{R} - \bar{\mathbf{Y}}\bar{\mathbf{Y}}^H$ with $\bar{\mathbf{Y}} = \bar{\mathbf{A}}\bar{\mathbf{T}} + \bar{\mathbf{E}}_b$. Because $L \rightarrow \infty$ at a faster rate than $\tilde{L}$, and that $\mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2}\bar{\mathbf{A}}}^\perp \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}} = \mathbf{0}$, we obtain the asymptotic expression of $\left[\dot{f}_3(\bar{\theta}) \right]_\eta$ as given by

$$\begin{aligned} \left[\dot{f}_3(\bar{\theta}) \right]_\eta &= -2 \cdot \text{Re} \{ \text{tr} \{ (\mathbf{Y} - \bar{\mathbf{A}}\bar{\mathbf{T}})^H \hat{\mathbf{R}}_n^{-1/2} \mathbf{P}_{\hat{\mathbf{R}}_n^{-1/2}\bar{\mathbf{A}}}^\perp \hat{\mathbf{R}}_n^{-1/2} \bar{\mathbf{A}}_\eta (\bar{\mathbf{A}}^H \hat{\mathbf{R}}_n^{-1} \bar{\mathbf{A}})^{-1} \bar{\mathbf{A}}^H \hat{\mathbf{R}}_n^{-1} \mathbf{Y} \} \} \\ &\simeq -2 \cdot \text{Re} \{ \text{tr} \{ \bar{\mathbf{E}}_b^H \bar{\tilde{\mathbf{R}}}_n^{-1/2} \mathbf{P}_{\bar{\tilde{\mathbf{R}}}_n^{-1/2}\bar{\mathbf{A}}}^\perp \bar{\tilde{\mathbf{R}}}_n^{-1/2} \bar{\mathbf{A}}_\eta \bar{\mathbf{T}} \} \} \end{aligned} \quad (\text{F.4})$$

where we have neglected terms of the order $o(1/\sqrt{\tilde{L}})$ ($o_p(1/\sqrt{\tilde{L}})$ for the terms which involve a stochastic signal). From Eq. (F.4), it can be shown that

$$\dot{f}_3(\bar{\theta}) = -2 \cdot \text{Re} \{ \{ (\bar{\mathbf{D}}^H \bar{\tilde{\mathbf{R}}}_n^{-1/2} \mathbf{P}_{\bar{\tilde{\mathbf{R}}}_n^{-1/2}\bar{\mathbf{A}}}^\perp \bar{\tilde{\mathbf{R}}}_n^{-1/2}) \odot (\bar{\mathbf{E}}_b \bar{\mathbf{T}}^H)^T \} \mathbf{I}_N \} \quad (\text{F.5})$$

Next, the second derivative matrix $\ddot{F}_3(\bar{\theta}) (= \nabla_{\theta}^2 f_3(\theta) \Big|_{\bar{\theta}})$ is given by

$$\ddot{F}_3(\bar{\theta}) \simeq 2 \cdot \text{Re}\{ (\bar{D}^H \bar{\tilde{R}}_n^{-1/2} P_{\bar{\tilde{R}}_n^{-1/2} \bar{A}}^{\perp} \bar{\tilde{R}}_n^{-1/2} \bar{D}) \odot (\bar{T} \bar{T}^H)^T \} \quad (\text{F.6})$$

From Eqs. (F.2, F.5, F.6), the asymptotic bias is given by

$$\begin{aligned} \tilde{\theta} &= -\ddot{F}_3^{-1}(\bar{\theta}) \dot{f}_3(\bar{\theta}) \\ &= \text{Re}^{-1} \{ (\bar{D}^H \bar{\tilde{R}}_n^{-1/2} P_{\bar{\tilde{R}}_n^{-1/2} \bar{A}}^{\perp} \bar{\tilde{R}}_n^{-1/2} \bar{D}) \odot (\bar{T} \bar{T}^H)^T \} \cdot \text{Re} \{ \{ (\bar{D}^H \bar{\tilde{R}}_n^{-1/2} P_{\bar{\tilde{R}}_n^{-1/2} \bar{A}}^{\perp} \bar{\tilde{R}}_n^{-1/2}) \odot \\ &\quad (\bar{E}_b \bar{T}^H)^T \} I_N \} \quad (\text{F.7}) \end{aligned}$$

G Proof of Theorem 5.1

Consider the data vector

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{x}_\alpha^T(t) & \mathbf{x}_\beta^T(t) & \mathbf{x}_\gamma^T(t) \end{bmatrix}^T \quad (\text{G.1})$$

where $\mathbf{x}_\alpha(t)$, $\mathbf{x}_\beta(t)$, and $\mathbf{x}_\gamma(t)$ are as defined in the main context (Section 5.4). It is sufficient to consider only the case where $l_1 \geq K_z$, $l_2 \geq \min(K - K_z + 1, K_z)$, $l_3 \geq K - K_z + 2$ ($\geq K - K_z + 1$ for the proof of the full-rank part related to P.1), $m_1 = m_3 = n_1 = n_2 = n_3 = 1$, and $m_2 = 2$. With the data vector $\mathbf{x}(t)$ given by Eq. (G.1) and the parameters as specified, the k th response vector is expressed as

$$\mathbf{a}_k = \begin{bmatrix} \mathbf{a}_{k,\alpha}^T & \mathbf{a}_{k,\beta}^T & \mathbf{a}_{k,\gamma}^T \end{bmatrix}^T \quad (\text{G.2})$$

where

$$\mathbf{a}_{k,\alpha} = \begin{bmatrix} 1 & z_k & \cdots & z_k^{l_1-1} \end{bmatrix}^T \quad (\text{G.3})$$

$$\mathbf{a}_{k,\beta} = y_k \cdot \begin{bmatrix} 1 & z_k & \cdots & z_k^{l_2-1} \end{bmatrix}^T \quad (\text{G.4})$$

$$\mathbf{a}_{k,\gamma} = \begin{bmatrix} 1 & y_k & \cdots & y_k^{l_3-1} \end{bmatrix}^T \quad (\text{G.5})$$

Let the response matrix $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_K]$ correspond to the K distinct 2-D parameter pairs (y_k, z_k) , $k = 1, \dots, K$, with $z_k \in \{z_l^0, l = 1, \dots, K_z\}$, $k = 1, \dots, K$. Note that, z_l^0 , $l = 1, \dots, K_z$, are K_z distinct-value z-dimension parameters. First, we will show that $\mathbf{A}$ has the property P.1, and is thus of full rank K , by showing that there is *no* solution to the equation

$$\begin{bmatrix} \mathbf{a}_{k_1} & \mathbf{a}_{k_2} & \cdots & \mathbf{a}_{k_{K-1}} \end{bmatrix} \cdot \mathbf{p} = \mathbf{a}_{k_0}, \quad k_l \in \{1, \dots, K\}, l = 0, \dots, K-1 \quad (\text{G.6})$$

for any vector $\mathbf{p}$.

CASE-1: The z-dimension parameter z_{k_0} of the k_0 th 2-D parameter pair is unique among the K z-dimension parameters z_{k_l} , $l = 0, \dots, K-1$.

With $l_1 \geq K_z$ and the fact that $\mathbf{a}_{k_0, \alpha}$ is independent of the other $(K_z - 1)$ independent vectors $\mathbf{a}_{k_l, \alpha}$ s, it is thus seen that the solution to Eq. (G.6) does not exist.

CASE-2: The k_0 th 2-D parameter pair shares the same z-dimension parameter value (z_{k_0}) as those of other, say, $(K_0 - 1)$ 2-D parameter pairs.

Observe that if there exists a solution to Eq. (G.6), there must also be a solution to

$$\begin{bmatrix} \mathbf{a}_{k_1} & \mathbf{a}_{k_2} & \cdots & \mathbf{a}_{k_{K_0-1}} \end{bmatrix} \cdot \mathbf{p}' = \mathbf{a}_{k_0} \quad (\text{G.7})$$

where $\mathbf{a}_{k_l}$, $l = 1, \dots, K_0 - 1$, correspond to the $(K_0 - 1)$ 2-D parameter pairs which share the same z-dimension parameter value as that of the k_0 th 2-D parameter pair. However, the matrix $\begin{bmatrix} \mathbf{a}_{k_0, \gamma} & \mathbf{a}_{k_1, \gamma} & \cdots & \mathbf{a}_{k_{K_0-1}, \gamma} \end{bmatrix}$ has full rank K_0 due to its Vandermonde structure and the fact that (in this case) $l_3 \geq K - K_z + 1 \geq K_0$. Thus, a solution to Eq. (G.7) and to Eq. (G.6) does not exist.

By combining the two cases, it is concluded that $\mathbf{A}$ has the property P.1. The full-rank part of the Theorem is thus proved. It remains to prove that $\mathbf{A}$ has also the property P.2. In the other word, it will be shown that given the distinct-value z-dimension parameters $z_k^0, k = 1, \dots, K_z$, there are K unique and distinct pairs (y_k, z_k) , $z_k \in \{z_l^0, l = 1, \dots, K_z\}$, of which the corresponding response matrix $\mathbf{A}$ span a unique K dimensional subspace. To show this, we define $\mathbf{a}_{K+1}$ as the response vector corresponding to the $(K + 1)$ th 2-D parameter pair, which shares the same z-dimension parameter as that of the k_0 th 2-D parameter pair (z_{k_0}) . However, the y-dimension parameter value of the $(K + 1)$ th 2-D parameter pair, denoted by y_{K+1} , is different from those of the k_0 th and those of other 2-D parameter pairs which have the same z-dimension parameter value as z_{k_0} . Again, we will prove this part of the theorem by showing that there is no solution to

$$\begin{aligned}
\mathbf{A}\mathbf{p} &= \mathbf{a}_{K+1} \\
&= \begin{bmatrix} \mathbf{a}_{k_0,\alpha}^T & \delta \cdot \mathbf{a}_{k_0,\beta}^T & \mathbf{a}_{K+1,\gamma}^T \end{bmatrix}^T
\end{aligned} \tag{G.8}$$

where $\mathbf{A}$ is of full rank K and $\delta = y_{K+1}/y_{k_0}$.

CASE-1: The z -dimension parameter value z_{k_0} of the k_0 th 2-D parameter pair is unique among those of the first K 2-D parameter pairs.

In this case, given that the solution to Eq. (G.8) does exist, there must also be a solution to

$$p_0 \begin{bmatrix} \mathbf{a}_{k_0,\alpha} \\ \mathbf{a}_{k_0,\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_{k_1,\alpha} & \mathbf{a}_{k_2,\alpha} & \cdots & \mathbf{a}_{k_{K-1},\alpha} \\ \mathbf{a}_{k_1,\beta} & \mathbf{a}_{k_2,\beta} & \cdots & \mathbf{a}_{k_{K-1},\beta} \end{bmatrix} \cdot \mathbf{p}' = \begin{bmatrix} \mathbf{a}_{k_0,\alpha} \\ \delta \cdot \mathbf{a}_{k_0,\beta} \end{bmatrix} \tag{G.9}$$

Because $l_1 \geq K_z$ and z_{k_0} is unique, $\mathbf{a}_{k_0,\alpha}$ is independent of the other $(K-1)$ vectors $\mathbf{a}_{k_l,\alpha}$, $l \in \{1, \dots, K-1\}$. Thus, from Eq. (G.9), $p_0 = 1$ and

$$\begin{bmatrix} \mathbf{a}_{k_1,\alpha} & \mathbf{a}_{k_2,\alpha} & \cdots & \mathbf{a}_{k_{K-1},\alpha} \end{bmatrix} \cdot \mathbf{p}' = \mathbf{0} \tag{G.10}$$

However, notice that, among the $(K-1)$ vectors, only $\mathbf{a}_{k_l,\alpha}$, $l \in \{1, \dots, K-1\}$, which have at least one replica, can be combined to form a zero vector for Eq. (G.10) to have a nontrivial solution. There are $u \leq \min(K - K_z, K_z - 1)$ groups of such vectors. Let $\mathbf{a}_{k_l,\alpha}$, $l = 1, \dots, u$, be u independent vectors. In addition, let $\mathbf{a}_{k_l,\alpha}$, $l = u+1, \dots, v$, be the response vectors which span the same subspace as that spanned by $\mathbf{a}_{k_l,\alpha}$, $l = 1, \dots, u$. A solution of Eq. (G.9) exists only if

$$\begin{bmatrix} \mathbf{a}_{k_1,\alpha} & \mathbf{a}_{k_2,\alpha} & \cdots & \mathbf{a}_{k_u,\alpha} \\ \mathbf{a}_{k_1,\beta} & \mathbf{a}_{k_2,\beta} & \cdots & \mathbf{a}_{k_u,\beta} \end{bmatrix} \cdot \mathbf{p}'' + \begin{bmatrix} \mathbf{a}_{k_{u+1},\alpha} & \mathbf{a}_{k_{u+2},\alpha} & \cdots & \mathbf{a}_{k_v,\alpha} \\ \mathbf{a}_{k_{u+1},\beta} & \mathbf{a}_{k_{u+2},\beta} & \cdots & \mathbf{a}_{k_v,\beta} \end{bmatrix} \cdot \mathbf{p}''' = \begin{bmatrix} \mathbf{0} \\ \delta_2 \cdot \mathbf{a}_{k_0,\beta} \end{bmatrix} \tag{G.11}$$

where $\delta_2 = \delta - 1$. Due to the fact that $l_2 \geq \min(K - K_z + 1, K_z) > u$, the vector $\mathbf{a}_{k_0,\beta}$ does not lie within the u -dimensional subspace spanned by the column vectors of $\begin{bmatrix} \mathbf{a}_{k_1,\beta} & \mathbf{a}_{k_2,\beta} & \cdots & \mathbf{a}_{k_u,\beta} & \mathbf{a}_{k_{u+1},\beta} & \cdots & \mathbf{a}_{k_v,\beta} \end{bmatrix}$. Because $y_{K+1} \neq y_{k_0}$, $\delta_2 \neq 0$ and there is no

solution to Eq. (G.11). As a result, a solution to Eq. (G.9) and to Eq. (G.8) does not exist.

CASE-2: The k_0 th 2-D parameter pair shares the same z-dimension parameter value as those of other $(K_0 - 1)$ 2-D parameter pairs.

Let the 2-D parameter pairs $(y_{k_l}, z_{k_l}), l = 1, \dots, K_0 - 1$, share the same z-dimension parameter value as that of the k_0 th 2-D parameter pair. From the proof for CASE-1, the response vectors corresponding to the 2-D parameter pairs whose z-dimension parameter values are different from the z-dimension parameter value of the k_0 th 2-D parameter pair, need not be considered. Thus, if the solution to Eq. (G.8) exists, there must also be a solution to

$$\begin{bmatrix} \mathbf{a}_{k_0, \alpha} & \mathbf{a}_{k_0, \alpha} & \cdots & \mathbf{a}_{k_0, \alpha} \\ \mathbf{a}_{k_0, \gamma} & \mathbf{a}_{k_1, \gamma} & \cdots & \mathbf{a}_{k_{K_0-1}, \gamma} \end{bmatrix} \cdot \mathbf{p}' = \begin{bmatrix} \mathbf{a}_{k_0, \alpha} \\ \mathbf{a}_{K+1, \gamma} \end{bmatrix} \quad (\text{G.12})$$

However, the $l_3 \times (K_0 + 1)$ matrix $\begin{bmatrix} \mathbf{a}_{k_0, \gamma} & \mathbf{a}_{k_1, \gamma} & \cdots & \mathbf{a}_{k_{K_0-1}, \gamma} & \mathbf{a}_{K+1, \gamma} \end{bmatrix}$ is of full rank $(K_0 + 1)$ due to its Vandermonde structure and the fact that $(K_0 + 1) \leq (K - K_z + 2) \leq l_3$. Thus, there is no solution to Eq. (G.12), and neither to Eq. (G.8).

By combining the two cases, it is proved that $\mathbf{A}$ has the property P.2.

H Proof of Theorem 5.2

The $p \times p$ average covariance matrix $\mathbf{R}_{\text{av}}$ can be written as

$$\begin{aligned} \mathbf{R}_{\text{av}} &= \frac{1}{r} \sum_{(k,l) \in \Psi} \mathbf{R}_{k,l} \\ &= \mathbf{A}_{s;k_1,l_1} \left\{ \frac{1}{r} \sum_{(k,l) \in \Psi} \mathbf{D}_{k,l} \mathbf{R}_s \mathbf{D}_{k,l}^H \right\} \mathbf{A}_{s;k_1,l_1}^H + \sigma^2 \mathbf{I}_p \end{aligned} \quad (\text{H.1})$$

where the full rank- K matrix $\mathbf{A}_{s;k_1,l_1}$ is the response matrix corresponding to $\mathbf{R}_{k_1,l_1}$, and

$$\mathbf{D}_{k,l} = \begin{bmatrix} y_1^{k-k_1} z_1^{l-l_1} & & & \mathbf{0} \\ & y_2^{k-k_1} z_2^{l-l_1} & & \\ & & \ddots & \\ \mathbf{0} & & & y_K^{k-k_1} z_K^{l-l_1} \end{bmatrix} \quad (\text{H.2})$$

Because the signal covariance matrix $\mathbf{R}_s$ is positive semi-definite with at least one non-zero eigenvalue, it can be decomposed as

$$\mathbf{R}_s = \lambda_1 \mathbf{s}_1 \mathbf{s}_1^H + \mathbf{S}' \quad (\text{H.3})$$

where $\mathbf{s}_1 = [s_{1,1} \ s_{1,2} \ \cdots \ s_{1,K}]^T$ is the eigenvector of $\mathbf{R}_s$ which corresponds to a non-zero eigenvalue λ_1 , and a positive semi-definite matrix $\mathbf{S}'$ is the remaining term. Based on this decomposition, Eq. (H.1) can be expressed as

$$\begin{aligned} \mathbf{R}_{\text{av}} &= \mathbf{A}_{s;k_1,l_1} \left\{ \mathbf{D}_s \left(\frac{\lambda_1}{r} \sum_{(k,l) \in \Psi} \mathbf{a}_{s;k,l} \mathbf{a}_{s;k,l}^H \right) \mathbf{D}_s^H \right\} \mathbf{A}_{s;k_1,l_1}^H + \mathbf{A}_{s;k_1,l_1} \left\{ \frac{1}{r} \sum_{(k,l) \in \Psi} \mathbf{D}_{k,l} \mathbf{S}' \mathbf{D}_{k,l}^H \right\} \mathbf{A}_{s;k_1,l_1}^H + \\ &\quad \sigma^2 \mathbf{I}_p \end{aligned} \quad (\text{H.4})$$

where

$$\mathbf{D}_s = \begin{bmatrix} s_{1,1} (y_1^{-k_1} z_1^{-l_1}) & & & \mathbf{0} \\ & s_{1,2} (y_2^{-k_1} z_2^{-l_1}) & & \\ & & \ddots & \\ \mathbf{0} & & & s_{1,K} (y_K^{-k_1} z_K^{-l_1}) \end{bmatrix} \quad (\text{H.5})$$

$$\mathbf{a}_{s;k,l} = [y_1^k z_1^l \ y_2^k z_2^l \ \cdots \ y_K^k z_K^l]^T \quad (\text{H.6})$$

Notice that the term $\sum_{(k,l) \in \Psi} \mathbf{a}_{s;k,l} \mathbf{a}_{s;k,l}^H$ in Eq. (H.4) is equal to $\mathbf{A}_c^T \mathbf{A}_c^*$, where $\mathbf{A}_c$ is the response matrix corresponding to the centre vector $\mathbf{x}_c(t)$, described by Eq. (5.51). Because $\mathbf{x}_c(t)$ satisfies the condition stated in Theorem 5.1 such that $\mathbf{A}_c$ has full rank K , the square matrix $\sum_{(k,l) \in \Psi} \mathbf{a}_{s;k,l} \mathbf{a}_{s;k,l}^H$ is also of full rank K . As a result, the first matrix term in Eq. (H.4) is composed of the column vectors which span the column space of $\mathbf{A}_{s;k_1,l_1}$. It is also seen that the subspace spanned by the column vectors of the second matrix term in Eq. (H.4) must lie within the column space of $\mathbf{A}_{s;k_1,l_1}$. Because $\mathbf{A}_{s;k_1,l_1}$ is assumed to have at least the property P.1, and $\mathbf{R}_{av}$ has the structure shown in Eq. (H.1) with $p > K$, the $p - K$ eigenvectors of $\mathbf{R}_{av}$ corresponding to the $p - K$ smallest eigenvalues form the nullspace of $\mathbf{A}_{s;k_1,l_1}^H$.

I Proof of the 2-D CRLB formula Eq. (5.61)

Assume that $\mathbf{x}(t)$ is modelled by Eq. (2.2), $\mathbf{n}(t)$ is complex Gaussian distributed, and $s(t)$ is an unknown parameter. Given L snapshots of $\mathbf{x}(t)$, $t = 0, 1, \dots, L-1$, the log-likelihood function is given by

$$\ell(\mathbf{x}(0), \dots, \mathbf{x}(L-1); \boldsymbol{\theta}^o, s(0), \dots, s(L-1)) = \text{const.} - L \ln |\mathbf{R}_n| - \sum_{t=0}^{L-1} (\mathbf{x}(t) - \mathbf{A}s(t))^H \mathbf{R}_n^{-1} (\mathbf{x}(t) - \mathbf{A}s(t)) \quad (\text{I.1})$$

where the 2-D *distinct-value* parameter vector

$$\begin{aligned} \boldsymbol{\theta}^o &= [\theta_1^o \quad \theta_2^o \quad \dots \quad \theta_{K_{y+z}}^o]^T \\ &\triangleq [(\boldsymbol{\theta}_y^o)^T \quad (\boldsymbol{\theta}_z^o)^T]^T \end{aligned} \quad (\text{I.2})$$

The vectors $\boldsymbol{\theta}_y^o$ and $\boldsymbol{\theta}_z^o$ are composed of, respectively, the K_y distinct-value y-dimension parameters and the K_z distinct-value z-dimension parameters. In addition, it is noted that $K_{y+z} = K_y + K_z$. In deriving the CRLB of an unbiased estimate of $\boldsymbol{\theta}^o$, the first derivatives of the log-likelihood Eq. (I.1) with respect to $\boldsymbol{\theta}^o$ and $s(t)$, $t = 0, \dots, L-1$, are required. By following the derivation of the compact CRLB formula in Appendix E of [9],

$$\begin{aligned} [\dot{\bar{\mathbf{g}}}]_i &\triangleq \left. \frac{\partial \ell}{\partial \theta_i^o} \right|_{\bar{\boldsymbol{\theta}}} \\ &= 2 \sum_{t=0}^{L-1} \text{Re}\{s^H(t) \bar{\mathbf{A}}_i^H \mathbf{R}_n^{-1} \mathbf{n}(t)\}, i = 1, \dots, K_{y+z} \end{aligned} \quad (\text{I.3})$$

where $\bar{\mathbf{A}}_i \triangleq \left. \frac{\partial \mathbf{A}}{\partial \theta_i^o} \right|_{\bar{\boldsymbol{\theta}}}$. From Eq. (I.3),

$$\bar{\mathbf{g}} = 2 \sum_{t=0}^{L-1} \text{Re}\{(\mathbf{R}_n^{-1/2} \bar{\mathbf{D}}_{2d} \mathbf{S}_D(t))^H \mathbf{R}_n^{-1/2} \mathbf{n}(t)\} \quad (\text{I.4})$$

where

$$\bar{\mathbf{D}}_{2d} = \begin{bmatrix} \left. \frac{\partial \mathbf{A}}{\partial \theta_{y;1}^o} \right|_{\bar{\boldsymbol{\theta}}} & \left. \frac{\partial \mathbf{A}}{\partial \theta_{y;2}^o} \right|_{\bar{\boldsymbol{\theta}}} & \dots & \left. \frac{\partial \mathbf{A}}{\partial \theta_{y;K_y}^o} \right|_{\bar{\boldsymbol{\theta}}} & \left. \frac{\partial \mathbf{A}}{\partial \theta_{z;1}^o} \right|_{\bar{\boldsymbol{\theta}}} & \dots & \dots & \left. \frac{\partial \mathbf{A}}{\partial \theta_{z;K_z}^o} \right|_{\bar{\boldsymbol{\theta}}} \end{bmatrix} \quad (\text{I.5})$$

and

$$\mathbf{S}_D(t) = \mathbf{I}_{K_{y+z}} \otimes \mathbf{s}(t) \quad (\text{I.6})$$

Next, notice that the derivative of Eq. (I.1) with respect to $\mathbf{s}(t)$ is the same for both 1-D and 2-D parameter cases. With the fact that the derivation of the compact CRLB in [9] involves the matrix product term $\mathbf{R}_n^{-1/2} \overline{\mathbf{D}}_{2d} \mathbf{S}_D(t)$, not the individual matrix components in it, a compact CRLB of an unbiased estimate of the 2-D parameter $\boldsymbol{\theta}^o$, given K_y , K_z , and K , can be obtained by following the derivation in [9], with the term $D\mathbf{X}(t)$ appeared in the mentioned reference being replaced by $\mathbf{R}_n^{-1/2} \overline{\mathbf{D}}_{2d} \mathbf{S}_D(t)$. A compact formula for the CRLB on the unbiased estimate $\hat{\boldsymbol{\theta}}^o$ is thus given by

$$\begin{aligned} \text{CRLB}(\hat{\boldsymbol{\theta}}^o) &= \\ & \frac{1}{2} \text{Re}^{-1} \left\{ \sum_{t=0}^L \mathbf{S}_D^H(t) \overline{\mathbf{D}}_{2d}^H \mathbf{R}_n^{-1/2} (\mathbf{I} - \mathbf{R}_n^{-1/2} \overline{\mathbf{A}} (\overline{\mathbf{A}}^H \mathbf{R}_n^{-1} \overline{\mathbf{A}})^{-1} \overline{\mathbf{A}}^H \mathbf{R}_n^{-1/2}) \mathbf{R}_n^{-1/2} \overline{\mathbf{D}}_{2d} \mathbf{S}_D(t) \right\} \\ &= \frac{1}{2L} \text{Re}^{-1} \left\{ \tilde{\mathbf{U}}_I^T \left\{ (\overline{\mathbf{D}}_{2d}^H \mathbf{R}_n^{-1/2} \mathbf{P}_{\mathbf{R}_n^{-1/2} \overline{\mathbf{A}}}^\perp \mathbf{R}_n^{-1/2} \overline{\mathbf{D}}_{2d}) \odot \hat{\mathbf{R}}_s^T \right\} \tilde{\mathbf{U}}_I \right\} \end{aligned} \quad (\text{I.7})$$

where

$$\hat{\mathbf{R}}_s = (\mathbf{I}_{K_{y+z}} \mathbf{I}_{K_{y+z}}^T) \otimes \hat{\mathbf{R}}_s \quad (\text{I.8})$$

$$\tilde{\mathbf{U}}_I = \mathbf{I}_{K_{y+z}} \otimes \mathbf{I}_K \quad (\text{I.9})$$

and $\hat{\mathbf{R}}_s$ is the sample covariance matrix of $\mathbf{s}(t)$, i.e., $\hat{\mathbf{R}}_s = \frac{1}{L} \sum_{t=0}^{L-1} \mathbf{s}(t) \mathbf{s}^H(t)$.

J Published papers

- 1) N. Chotikakamthorn, and J. A. Chambers, "Source Direction Estimation for Digital Mobile Communications," *2nd International Workshop on Mobile Multi-Media Communications*, University of Bristol, April 1995.
- 2) N. Chotikakamthorn, and J. A. Chambers, "Optimum Beamforming for Correlated Sources Based on the Subspace Decomposition Approach," *Proc. IEEE/IEE Int. Workshop on Signal Processing Methods in Multipath Environments*, University of Strathclyde, pp. 16-22, April 1995.
- 3) N. Chotikakamthorn, and J. A. Chambers, "On the IQML Algorithm for Multiple Signal Parameter Estimation," submitted to *IEE Proceeding Pt. F*.

References

- [1] E. Nicolau, and D. Zaharia, *Adaptive Array*, Elsevier Science Publishing Co. Inc., Romania, 1989.
- [2] J. E. Hudson, *Adaptive Array Principles*, Peter Peregrinus Ltd., London, 1981.
- [3] R. T. Compton, *Adaptive Antennas: Concepts and Performance*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.
- [4] B. Widrow, P. E. Mante, L. J. Griffiths, and B. B. Goode, "Adaptive antenna systems," *Proc. IEEE*, Vol. 55, pp. 2143-2159, December 1967.
- [5] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. R. astr. Soc.*, Vol. 33, pp. 347-366, 1973.
- [6] D. G. Childers, ed., *Modern Spectrum Analysis*, IEEE Press, New York, 1978.
- [7] S. Haykin ed., *Nonlinear Methods of Spectral Analysis, 2 ed.*, Springer-Verlag, Berlin Heidelberg, 1983.
- [8] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. Antenna Propagation*, Vol. 34, No. 3, pp. 276-280, March 1986.
- [9] 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.
- [10] J. E. Evans, J. R. Johnson, and D. F. Sun, "Application of advanced signal processing techniques to angle of arrival estimation in ATC navigation and surveillance systems," *MIT Lincoln Lab, Tech. Rep. 582*, June 1982.

- [11] T. J. Shan, M. Wax, and T. Kailath, "On Spatial Smoothing for Direction-of-Arrival Estimation of Coherent Signals," *IEEE Trans. Acoust. Speech Signal Processing*, Vol. 33, pp. 806-811, August 1985.
- [12] H. Wang, and M. Kaveh, "Coherent Signal-Subspace Processing for the Detection and Estimation of Angles of Arrival of Multiple Wideband Sources," *IEEE Trans. Acoust. Speech Signal Processing*, Vol. 33, pp. 823-831, August 1985.
- [13] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, 1991.
- [14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Englewood Cliffs, New Jersey, 1993.
- [15] J. Li and R. T. Compton Jr., "Maximum Likelihood Angle Estimation for Signals with Known Waveforms," *IEEE Trans. Signal Processing*, Vol. 41, No. 9, pp. 2850-2861, September 1993.
- [16] J. Li, B. Halder, P. Stoica, M. Viberg, and T. Kailath, "Decoupled Maximum Likelihood Angle Estimation for Signals with Known Waveforms," *Technical Report CTH-TE-8*, Chalmers University of Technology, 10 February, 1994.
- [17] J. Li, B. Halder, P. Stoica, and M. Viberg, "Computationally Efficient Angle Estimation for Signals with Known Waveforms," *IEEE Trans. Signal Processing*, Vol. 43, pp. 2154-2163, September 1995.
- [18] Q. Wu, and K. M. Wong, "UN-MUSIC and UN-CLE: An Application of Generalised Canonical Correlation Analysis to the Estimation of the Directions of Arrival," *IEEE Trans. Signal Processing*, Vol. 42, No. 10, pp. 2331-2343, September 1994.
- [19] M. Viberg, P. Stoica, and B. Ottersten, "Array Processing in Correlated Noise Fields Based on Instrumental Variables and Subspace Fitting," *IEEE Trans. Signal Processing*, Vol. 43, No. 5, pp. 1187-1199, May 1995.

- [20] P. Stoica, M. Viberg, M. Wong, and Q. Wu, "Optimal IV-SSF Approach to Array Signal Processing in Colored Noise Fields," *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, pp. 2088-2091, 1995.
- [21] B. Widrow, K. M. Duvall, R. P. Gooch, and W. C. Newman, "Signal cancellation phenomena in adaptive arrays: causes and cures, " *IEEE Trans. Antenna Propagation*, Vol. 30, pp. 469-478, May 1982.
- [22] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust. Speech Signal Processing*, Vol. 34, pp. 1081-1089, October 1986.
- [23] P. Stoica and K. C. Sharman, "Novel eigenanalysis method for direction estimation," *IEE Proceedings Pt. F*, Vol. 137, No. 1, pp. 19-26, February 1990.
- [24] V. Nagesha, and S. Kay, "On Frequency Estimation with the IQML algorithm," *IEEE Trans. Signal Processing*, Vol. 42, No. 9, pp. 2509-2513, September 1994.
- [25] B. Friedlander, "The root-MUSIC algorithm for direction finding with interpolated arrays," *Signal Processing*, Vol. 30, No. 1, pp. 15-29, January 1993.
- [26] S. Haykin, ed., *Array Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1985.
- [27] B. D. Van Veen, and K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," *IEEE ASSP Magazine*, pp. 4-24, April 1988.
- [28] K. M. Buckley, "Spatial/Spectral Filtering with Linearly Constrained Minimum Variance Beamformers," *IEEE Trans. Acoust. Speech Signal Processing*, Vol. 35, pp. 249-266, March 1987.
- [29] C. W. Therrian, *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1992.
- [30] A. J. Barabell, "Improving the resolution performance of eigenstructure based direction finding algorithm," *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, pp. 336-339, 1983.

- [31] B. D. Rao, and K. V. S. Hari, "Performance Analysis of Root-Music," *IEEE Trans. Acoust. Speech. Signal Processing*, Vol. 37, pp. 1939-1949, December 1989.
- [32] D. W. Tufts and R. Kumaresan, "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood," *Proc. IEEE*, Vol. 70, No. 9, pp. 975-989, September 1982.
- [33] M. Viberg, and B. Ottersten, "Sensor Array Processing Based on Subspace Fitting," *IEEE Trans. Signal Processing*, Vol. 39, No. 5, pp. 1110-1120, May 1991.
- [34] S. Haykin, J. Litva, and T. J. Shepherd (Eds.), *Radar Array Processing*, Springer-Verlag, 1993.
- [35] P. Stoica and K. C. Sharman, "Maximum Likelihood Methods for Direction-of-Arrival Estimation," *IEEE Trans. Acoust. Speech. Signal Processing*, Vol. 38, No. 7, pp.1132-1143, July 1990.
- [36] B. D. Rao, and K. V. S. Hari, "Statistical performance analysis of the minimum-norm method," *IEE Proceedings Pt. F*, Vol. 136, pp. 125-134, June 1989.
- [37] P. Stoica, and A. Nehorai, "MUSIC, Maximum Likelihood, and Cramer-Rao Bound: Further Results and Comparisons," *IEEE Trans. Acoust. Speech. Signal Processing*, Vol. 38, No. 12, pp. 2140-2150, December 1990.
- [38] C. Vaidyanathan, and K. Buckley, "Performance Analysis of the MVDR Spatial Spectrum Estimator," *IEEE Trans. Signal Processing*, Vol. 43, No. 6, pp. 1427-1437, June 1995.
- [39] S. U. Pillai, *Array Signal Processing*, Springer-Verlag, New York, 1989.
- [40] M. D. Zoltowski, G. Kautz, and S. Silverstein, "Beamspace root-MUSIC method," *IEEE Trans. Signal Processing*, Vol. 41, pp. 344-364, January 1993.

- [41] M. Zoltowski, "High Resolution Sensor Array Signal Processing in the Beamspace Domain: Novel Techniques Based on the Poor Resolution of Fourier Beamforming," *Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling*, pp. 350-355, August 1988.
- [42] R. D. DeGroat, E. M. Dowling, and D. A. Linebarger, "The Constrained MUSIC Problem," *IEEE Trans. Signal Processing*, Vol. 41, pp. 1445-1449, March 1993.
- [43] G. Su, and M. Morf, "Modal Decomposition Signal Subspace Algorithms," *IEEE Trans. Acoust. Speech Signal Processing*, Vol. 34, pp. 585-602, June 1986.
- [44] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979.
- [45] M. Viberg and A. Swindlehurst, "Analysis of the Combined Effects of Finite Samples and Model Errors on Array Processing Performance," *IEEE Trans. Signal Processing*, Vol. 42, No. 11, pp. 3073-3082, November 1994.
- [46] J. C. Mosher, P. S. Lewis, and R. M. Leahy, "Multiple Dipole Modeling and Localization from Spatio-Temporal MEG Data," *IEEE Trans. Biomedical Engineering*, pp. 541-557, Vol. 39, June 1992.
- [47] S. Haykin, ed., *Advances in Spectrum Estimation and Array Processing*, Vol. 3, Prentice-Hall, Englewood Cliffs, New Jersey, 1995.
- [48] E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, Kluwer Academic Publishers, Boston, 1988.
- [49] J. H. Winters, "Optimum Combining in Digital Mobile Radio with Cochannel Interference," *IEEE Trans. Vehicular Technology*, Vol. VT-33, No. 3, pp. 144-155, August 1984.
- [50] Y. Bresler, and A. Macovski, "On the Number of Signals Resolvable by a Uniform Linear Array," *IEEE Trans. Acoust. Speech Signal Processing*, Vol. 34, No. 6, pp. 1361-1375, December 1986.

- [51] M. Wax, and I. Ziskind, "On unique localization of multiple sources in passive sensor arrays," *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. 37, pp. 996-1000, July 1989.
- [52] W. C. Y. Lee, *Mobile Communications Engineering*, McGraw-Hill, New York, 1982.
- [53] S. Anderson, M. Millnert, M. Viberg, and B. Wahlberg, "An Adaptive Array for Mobile Communication Systems," *IEEE Trans. Vehicular Technology*, Vol. 40, No. 1, pp. 230-236, February 1991.
- [54] B. Ottersten, "Spatial Division Multiple Access (SDMA) in Wireless Communications," *Technical Report No. IR-S3-SB-9507*, Royal Institute of Technology, Stockholm, 1995.
- [55] T. Trump, and B. Ottersten, "Maximum Likelihood Estimation of Nominal Direction of Arrival and Angular Spread Using an Array of Sensors," *Technical Report No. IR-S3-SB-9501*, Royal Institute of Technology, Stockholm, January 1995.
- [56] N. Chotikakamthorn, and J. A. Chambers, "Source Direction Estimation for Digital Mobile Communications," *2nd International Workshop on Mobile Multi-Media Communications*, April 1995.
- [57] I. Ziskind and M. Wax, "Maximum Likelihood Localization of Multiple Sources by Alternating Projection," *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. 36, No. 10, pp. 1553-1560, October 1988.
- [58] M. I. Miller and D. R. Fuhrmann, "Maximum-Likelihood Narrow-Band Direction Finding and the EM Algorithm," *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. 38, No. 9, pp. 1560-1577, September 1990.
- [59] K. C. Sharman, "Maximum likelihood parameter estimation by simulated annealing," *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, pp. 2741-2744, April 1988.

- [60] J. H. McClellan and D. Lee, "Exact equivalent of the Steiglitz-McBride iteration and IQML," *IEEE Trans. Signal Processing*, Vol. 39, pp. 509-512, February 1991.
- [61] T. P. Bronez, "Sector Interpolation of Non-uniform Arrays for Efficient High Resolution Bearing Estimation," *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, Vol. 5, pp. 2885-2888, April 1988.
- [62] M. P. Clark and L. L. Scharf, "Two-Dimensional Modal Analysis Based on Maximum Likelihood," *IEEE Trans. Signal Processing*, Vol. 42, No. 6, pp. 1443-1451, June 1994.
- [63] A. J. van der Veen, P. B. Ober, and E. F. Deprettere, "Azimuth and Elevation Computation in High Resolution DOA Estimation," *IEEE Trans. Signal Processing*, Vol. 40, No. 7, pp. 1828-1832, July 1992.
- [64] Y. Hua, "Estimating Two-Dimensional Frequencies by Matrix Enhancement and Matrix Pencil," *IEEE Trans. Signal Processing*, Vol. 40, No. 9, pp. 2267-2280, September 1992.
- [65] J. J. Sacchini, W. L. Steedly, and R. L. Moses, "2-D Prony modeling and parameter estimation," *IEEE Trans. Signal Processing*, Vol. 41, No. 11, November 1993.
- [66] M. Haardt, M. D. Zoltowski, C. P. Mathews, and J. A. Nossek, "2D Unitary ESPRIT for Efficient 2D Parameter Estimation," *Proc. Int. Conf. Acoust. Speech, Signal Processing*, pp. 2096-2099, 1995.
- [67] Y. Bresler, V. U. Reddy, and T. Kailath, "Optimum beamforming for coherent signal and interferences," *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. 36, pp. 833-843, June 1988.
- [68] B. Ottersten, R. Roy, and T. Kailath, "Signal Waveform Estimation in Sensor Array Processing," *Proc. 23th Asilomar Conference on Signals, Systems, and Computers*, pp. 787-791, November 1989.

- [69] N. Chotikakamthorn, and J. A. Chambers, "Optimum Beamforming for Correlated Sources Based on the Subspace Decomposition Approach," *Proc. IEEE/IEE Int. Workshop on Signal Processing Methods in Multipath Environments*, pp. 16-22, April 1995.
- [70] A. J. Weiss, B. Friedlander, and P. Stoica, "Direction-of-Arrival Estimation Using MODE with Interpolated Arrays," *IEEE Trans. Signal Processing*, Vol. 43, No. 1, pp. 296-300, January 1995.
- [71] Y. Hua, "The Most Efficient Implementation of the IQML Algorithm," *IEEE Trans. Signal Processing*, Vol. 38, pp. 2203-2204, August 1994.
- [72] G. H. Golub, and C. F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, 1989.
- [73] Y. Hua, "High resolution imaging of continuously moving object using stepped frequency radar," *Signal Processing*, Vol. 35, pp. 33-40, January 1994.
- [74] D. A. Linebarger and R. D. DeGroate, "Efficient Direction-Finding Methods Employing Forward/Backward Averaging," *IEEE Trans. Signal Processing*, Vol. 42, No. 8, pp. 2136-2145, August 1994.
- [75] S. F. Yau, and Y. Bresler, "A Compact Cramér-Rao Bound Expression for Parametric Estimation of Superimposed Signals," *IEEE Trans. Signal Processing*, Vol. 40, pp. 1226-1230, May 1992.
- [76] 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, October 1990.
- [77] S. F. Yau, and Y. Bresler, "Worst Case Cramér-Rao Bounds for Parametric Estimation of Superimposed Signals with Applications," *IEEE Trans. Signal Processing*, Vol. 40, pp. 2973-2986, December 1992.
- [78] M. Wax, and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. 33, No. 2, pp. 387-392, April 1985.

- [79] M. Wax, and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 37, pp. 1190-1196, August 1989.
- [80] C. Cho, and P. M. Djuric, "Detection and Estimation of DOA's of Signals via Bayesian Predictive Densities," *IEEE Trans. Signal Processing*, Vol. 42, pp. 3051-3060, November 1994.
- [81] M. Viberg, B. Ottersten, and T. Kailath, "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting," *IEEE Trans. Signal Processing*, Vol. 39, No. 11, pp. 3436-3448, November 1991.
- [82] A. V. Oppenheim, and R. Schafer, *Discrete-Time Signal processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1989.
- [83] A. Swindlehurst, and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors-Part I: The MUSIC algorithm," *IEEE Trans. Signal Processing*, Vol. 40, No. 7, pp. 1758-1774, July 1992.
- [84] A. Swindlehurst, and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors-Part II: Multidimensional algorithms," *IEEE Trans. Signal Processing*, Vol. 41, No. 9, pp. 1277-1308, September 1993.
- [85] B. Friedlander, "Sensitivity analysis of the maximum likelihood direction-finding algorithm," *IEEE Trans. Aerospace Electron. Syst.*, Vol. 26, No. 6, pp. 953-968, November 1990.
- [86] A. Paulraj, and T. Kailath, "Direction of arrival estimation by eigenstructure methods with unknown sensor gain and phase," *Proc. Int. Conf. Acoust. Speech, Signal Processing*, pp. 640-643, August 1985.
- [87] A. J. Weiss, and B. Friedlander, "Array shape calibration using sources in unknown locations-A maximum likelihood approach," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 37, No. 12, pp. 1958-1966, December 1989.

- [88] M. P. Wylie, S. Roy, and H. Messer, "Joint DOA Estimation and Phase Calibration of Linear Equispaced (LES) Arrays," *IEEE Trans. Signal Processing*, Vol. 42, No. 12, pp. 3449-3459, December 1994.
- [89] M. Viberg, and A. L. Swindlehurst, "A Bayesian Approach to Auto-Calibration for Parametric Array Signal Processing," *IEEE Trans. Signal Processing*, Vol. 42, No. 12, pp. 3495-3507, December 1994.
- [90] R. Arnott, "Array Diversity Combining using Radial Basis Function Neural Networks," *MPhil/PhD Transfer Report*, Dept. of Electrical and Electronic Eng., Imperial College, London, April 1994.
- [91] P. Stoica, and T. Söderström, "The Steiglitz-McBride Identification Algorithm Revisited-Convergence Analysis and Accuracy Aspects," *IEEE Trans. Automatic Control*, Vol. 26, No. 3, pp. 712-717, June 1981.
- [92] H. Fan and M. Doroslovački, "On 'Global Convergence' of Steiglitz-McBride Adaptive Algorithm," *IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 40. No.2, pp. 73-87, February 1993.

