IT Licentiate Theses
2000-001

Low-Angle Estimation: Models, Methods and Bounds

KATARINA BOMAN

UPPSALA UNIVERSITY
Department of Information Technology

Low angle estimation: models, methods and bounds

Katarina Boman

January 19, 2000

Abstract

In this work we study the performance of elevation estimators and lower bounds on the estimation error variance for a low angle target in a smooth sea scenario using an array antenna. The article is structured around some key assumptions on multipath knowledge, signal parameterization and noise covariance, giving the reader a framework in which Maximum-Likelihood estimators exploiting different á priori information can be found.

The crucial factor that determines the estimator accuracy is the multipath modeling, and there are three alternative levels of knowledge that can be used: 1) two unknown target locations 2) the target and its corresponding sea-reflection are related via simple geometry 3) the sea reflection coefficient is known as a function of grazing angle.

A compact expression for the Cramér-Rao lower bound is derived, including all special cases of the key assumptions. We prove that the Cramér-Rao bound is highly dependent on the multipath model, while it is the same for the different signal parameterizations and that it is independent of the noise covariance. However, the Cramér-Rao bound is sometimes too optimistic and not achievable. The tighter Barankin bound is derived to predict the threshold behavior seen at low SNR. At high SNR the Barankin bound coincides with the Cramér-Rao bound. Simulations show that the Maximum Likelihood methods are statistically efficient and achieve the theoretical lower bound on error variance, in case of high enough SNR.

The bounds are also useful tools to design an improved array structure that can give better performance than the standard uniform linear array structure. The influence of the number of sensors and the number of snapshots on the error variariance is also studied, showing the rate of improvement with more sensors or snapshots. Finally we discuss the use of multiple frequencies, which is mainly a tool for suppressing ambiguities. We show for which signal models it provides improved performance.

Acknowledgements

I would like to thank Professor Petre Stoica for recruiting me to the group and for his supervision, guidance and support of my graduate studies. I am also very grateful for the financial support and interest of Celsius Tech Electronics, and a special thanks to Bert-Eric Tullsson who has been invaluable as a discussion partner. It has been a pleasure to work at the department of Systems and Control. Thanks to all of you who contribute to our creative environment with all our social activities.

Contents

1 Introduction

Tracking low angle targets over sea is known to be a difficult problem since the direct signal and its mirror image are closely spaced and coherent. In a smooth sea scenario the received signal is mainly a sum of two signals, the direct signal and the specular signal scattered by the sea surface.

If the transmitted signal is narrowband, the signals are coherent since they are delayed versions of the same transmitted signal. Barton [1] reviews the work on low angle tracking up to 1974 and he is still cited as a standard reference. In [1] a detailed model of the surface reflection is given, which is used in this thesis. By using this surface reflection model we are able to simulate the low angle scenario with properties corresponding to a smooth sea scenario.

The general problem of estimating the DOA (direction of arrival) using radar array processing is discussed in [2, 3]. However, many results are not valid or have to be modified to be applicable also in the low-angle scenario. As we will see there are several possible combinations of assumptions on parameterization of the signal and noise model. The systematic approach to derive an estimator for a given signal model is to derive the MLE (Maximum-Likelihood estimator). If the noise is Gaussian distributed, the MLE is known to be at least asymptotically (for large data records) efficient. The concept of Maximum Likelihood and statistical efficiency is thoroughly described in [4].

There are several more or less successful superresolution algorithms derived to separate the target from its image. The signal coherency resulted in the breakdown of the classical subspace technique MUSIC (multiple signal classification) which has been widely used for angle estimation. Yet there are subspace techniques that can handle coherent sources. Two good examples are MODE (method of direction estimation) and WSF (weighted subspace fitting) [2]. But since the mirror image also carries information on the target location, more accurate estimates can be found. The signal wavefront received at the antenna array can be parametrized as a function of elevation, and there are three possible levels of knowledge that can be used: 1) two unknown target locations 2) the target and its image are related via simple geometry 3) the sea reflection coefficient is known as a function of grazing angle. The signal propagation in time is determined by the transmitted signal, modulation schemes and the target movement and can be considered as unknown in some cases and modeled as a function of doppler frequency in other cases. Traditionally the noise is assumed to be spatially and temporally white, but here we will also derive methods that deal with spatially correlated noise. Spatially colored noise can arise partly because sea clutter mainly arrive from sea level, which may give a dominant noise direction.

There is also a number of publications on how to estimate the DOA in some of the special cases [5–8]. This work will present combinations of special cases of different multipath knowledge, signal parameterization and noise assumptions in a unifying framework.

The relevant problem is estimation of the parameters in the following model

$$\mathbf{y}(t) = \begin{bmatrix}\mathbf{a}(\theta_1) & \mathbf{a}(\theta_2)\end{bmatrix}\begin{bmatrix}b_1 & b_2\end{bmatrix}^T x(t) + \varepsilon(t) \qquad t = 1, \ldots, N \tag{1.1}$$

where $\{\mathbf{y}(t)\} \in C^{m\times 1}$ are the observed data in m sensors, $\{x(t)\} \in C^{1\times 1}$ is the signal wave form and $\{\varepsilon(t)\}$ is an additive noise. The vectors $\mathbf{a}(\theta_{1,2}) \in C^{m\times 1}$ are the standard single path array responses for the DOAs $\theta_{1,2}$ and $\begin{bmatrix}b_1 & b_2\end{bmatrix}^T$ are proportional to the power received from the two directions.

For many parameter estimation problems researchers continue to look for estimators with better accuracy until it is proved that an estimator achieve the lower bound on error variance. A lower bound on the error variance is very useful in practice. It proves that it is impossible to find any unbiased estimator whose variance is less than the bound. It can also be used to determine how different model assumptions and array design affect the bound, which helps optimizing the conditions for the estimator that achieve the bound. However, there are many existing bounds and only the tightest (highest) lower bound can possibly be achievable. The Cramér-Rao bound (CRB) [4] is one of the easiest to determine and is by far the most used. But if the problem studied suffers from ambiguities the CRB will be too optimistic. A tighter bound that can take ambiguities into account is the Barankin bound (BB) [9–11], which turns out to predict the threshold behavior seen in simulations. With the help of the CRB and the BB we will be able to show that the MLE approaches the lower bound on error variance, and no method can do significantly better.

2 Contributions

The work in this thesis is based on one review article on low angle target radar, and its corresponding models, methods and bounds. The review paper has been submitted to the review journal Digital Signal Processing. Parts of the material presented in this thesis can be found in the conference proceedings [12, 13]. My research has been formulated around a basic signal model which can be refined to fit different assumptions on multipath geometry, signal waveform and noise properties. The aim has been to cover all combinations of assumptions, providing Maximum Likelihood estimators and theoretical bounds on error variance, to have a tool to compare how different assumptions and designer choices affect the estimation error variance. For some special cases the MLE and the CRB have already been published by others, but a unifying framework has been missing.

3 Basic assumptions, signal model and special cases

In this section the signal model for a low angle multipath scenario will be formulated and several special cases in terms of array response parameterization, signal waveform and noise properties follow. Let us start with some basic assumptions that will hold throughout the paper.

3.1 Basic assumptions:

B1 The transmitted signal is narrow-band.

B2 The direct and specular signals are coherent.

B3 The noise $\{\boldsymbol{\varepsilon}(t)\}$ is a stationary, temporally white random process with zero mean.

Assumption B1 is standard and B2 follows from the definition of specular reflection. The noise $\{\boldsymbol{\varepsilon}(t)\}$ represents interference due to receiver noise, clutter, diffuse reflections, and jamming. Hence, the assumption B3 is a somewhat simplistic assumption. Depending on the pulse shape of the transmitted signal, the sampling rate, modulation schemes and the actual surroundings this assumption will be more or less accurate. But as long as the temporal correlation of the noise is not too large it is not of any crucial importance for the estimates.

3.2 Signal model and notation

The center of the antenna array is located at height h and the target at height H above sea level. The received target signal is a sum of the direct and specular signals arriving from the two directions θ_1' and θ_2'. The path lengths of the direct and specular paths are R_1 and R_2. In Figure 1 the geometry of the multipath scenario and the above notation is displayed.

The general signal model (1.1) is restricted to the physical constraint of equal angles of incidence and reflection for the specular reflection. Assuming low angles (θ_1' small) the following expressions for the angles of arrival $\theta_{1,2}'$ are easily derived using a flat earth model and simple geometry

$$\theta_1 \stackrel{\triangle}{=} \sin\theta_1' = \frac{H-h}{\sqrt{x^2+(H-h)^2}} \approxeq \frac{H-h}{R} \tag{3.1}$$

$$\theta_2 \stackrel{\triangle}{=} \sin\theta_2' = -\frac{H+h}{\sqrt{x^2+(H+h)^2}} \approxeq \frac{-H-h}{R} \tag{3.2}$$

where

$$R = \sqrt{x^2+H^2+h^2} \tag{3.3}$$

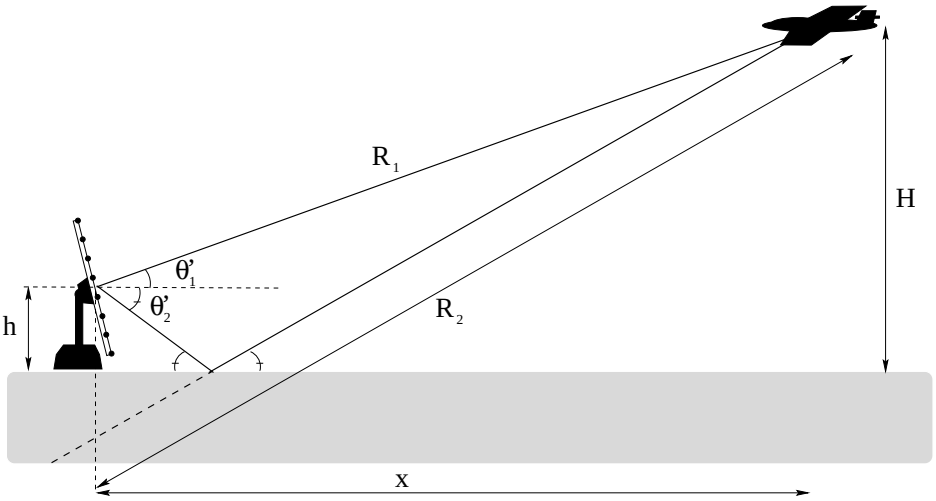

Figure 1. Multipath scenario.

It has been established [5–7] that a two-way radar model and a one-way beacon model will lead to essentially the same signal model

$$\mathbf{y}(t) = \begin{bmatrix}\mathbf{a}(\theta_1) & \mathbf{a}(\theta_2)\end{bmatrix}\begin{bmatrix}1 & G\end{bmatrix}^T x(t) + \boldsymbol{\varepsilon}(t) \qquad t = 1, \ldots, N \tag{3.4}$$

where

$$G = , \; e^{-j\kappa 2Hh/R} \tag{3.5}$$

$$\mathbf{a}_k(\theta_{1,2}) = \exp\bigl(-j\kappa \sin\theta'_{1,2} d_k\bigr) = \exp\bigl(-j\kappa \frac{\pm H - h}{R} d_k\bigr) \tag{3.6}$$

$\kappa = 2\pi/\lambda$ is the wave number and d_k is the vertical distance from sensor k to the center of the antenna. The only difference between the beacon model and the two-way radar model is the expression for the received signal $x(t)$, which is of interest when we wish to parameterize $x(t)$ in terms of the doppler frequency. If the sea conditions are known, which is not always the case, the specular reflection coefficient , is known as a function of grazing angle and hence G can be expressed in terms of elevations H, h and range R [7]. The array responses $\mathbf{a}_k(\theta_{1,2})$ in Equation (3.6) correspond to a vertical array. If the array is tilted relative to the horizon (3.6) has to be compensated for the known tilt.

This model assumes flat earth but nothing essential is changed when that assumption is violated since corrections to the model are deterministic and arise from geometry. In [5–7] the beacon model and the two-way radar models are given with a possibility to compensate for earth curvature and atmospheric refraction. For simplicity let us use the flat earth model (3.4–3.6).

In the following section we will see that, depending on which parameters are known, it is possible to rewrite the model into

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{b}x(t) + \boldsymbol{\varepsilon}(t) \qquad t = 1, \ldots, N \tag{3.7}$$

where the noise is characterized by

$$\mathrm{E}\bigl(\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}(s)^*\bigr) = \mathbf{Q}\delta_{t,s} \tag{3.8}$$

$$\mathrm{E}\bigl(\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}(s)^T\bigr) = \mathbf{0} \tag{3.9}$$

where $\mathbf{A}(\boldsymbol{\theta}) \in C^{m\times n}$ is the array response matrix for the unknown elevation parameters $\boldsymbol{\theta}$.

Additional notation: Tr (.) denotes the trace operation, |(.)| denotes the determinant operation, ||(.)|| denotes the euclidian norm, $(.)^*$ denotes a conjugate transpose and $\mathrm{Re}(.) = \overline{(.)}$ and $\mathrm{Im}(.) = \widetilde{(.)}$ denote real and imaginary parts.

$$\mathbf{Y} = \begin{bmatrix}\mathbf{y}(1) & \dots & \mathbf{y}(N)\end{bmatrix} \tag{3.10}$$

$$\mathbf{f}_\nu = \mathbf{f}(\nu) = \frac{1}{\sqrt{N}}\begin{bmatrix}1 & e^{j\nu} & \dots & e^{j\nu N}\end{bmatrix}^* \tag{3.11}$$

$$\mathbf{d}_k = \frac{\partial \mathbf{A}(\boldsymbol{\theta})\mathbf{b}}{\partial \theta_k} \tag{3.12}$$

$$\mathbf{D} = \begin{bmatrix}\mathbf{d}_1 & \dots & \mathbf{d}_p\end{bmatrix} \tag{3.13}$$

$$\hat{\mathbf{R}}_{yy} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{y}(t)\mathbf{y}(t)^* \tag{3.14}$$

$$\hat{\mathbf{R}}_{yx} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{y}(t)\mathbf{x}(t)^* \tag{3.15}$$

$$\hat{\mathbf{R}}_{xx} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{x}(t)\mathbf{x}(t)^* \tag{3.16}$$

$$\mathbf{Q} = \mathrm{E}\big(\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}(t)^*\big) \tag{3.17}$$

$$\tag{3.18}$$

The signal to noise ratio (SNR) will be frequently used, and we define it as follows:

$$\mathrm{SNR} = \frac{\mathrm{E}\big(x(t)^*x(t)\big)\,\|\mathbf{A}(\boldsymbol{\theta})\mathbf{b}\|^2}{\mathrm{E}\big(\boldsymbol{\varepsilon}(t)^*\boldsymbol{\varepsilon}(t)\big)} \tag{3.19}$$

In simulations the RMSE of various estimators will be used to display estimator accuracy. When there are two source DOA's estimated $\hat{\theta}_{1,2}$ in each realization, the RMSE value of K realizations is computed as follows

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{k=1}^{K}(\tilde{\theta}_1(k)-\theta_1)^2 + (\tilde{\theta}_2(k)-\theta_2)^2}{2K}} \tag{3.20}$$

where

$$\hat{\theta}_1(k) = \min\{\hat{\theta}_1, \hat{\theta}_2\} \tag{3.21}$$

$$\hat{\theta}_2(k) = \max\{\hat{\theta}_1, \hat{\theta}_2\} \tag{3.22}$$

Default radar setup: In simulations there will be a default radar setup used if nothing else is specified. This standard setup is chosen to resemble a radar built at CelsiusTech Electronics. This radar is a ULA (Uniform Linear Array) and has 8 sensors which are directional elements, recieving signals uniformly from a sector of 60 degrees in elevation. Due to the 60 degree sector instead of 180 degrees the sensor spacing d can be larger than the standard limit of half the wavelength without aliasing.

$$\begin{aligned} \lambda &= 0.1\text{m} \\ d &= 0.08\text{m} \\ m &= 8 \\ N &= 64 \\ h &= 5\text{m} \\ R &= 10^4\text{m} \end{aligned} \tag{3.23}$$

3.3 Special cases

The signal model (3.7) can be structured in a few special cases. In most papers there is usually only one specific combination of assumptions on array response parameterization, signal waveform and noise properties. Here we bring them all together in a joint framework.

C1 **Optional assumptions on multipaths**

a The two elevation parameters $\theta_{1,2}$ are independent and unknown. This occurs if the radar height and the range to the target are unknown. Use $\theta_1 = (H-h)/R$ and $\theta_2 = (-H-h)/R$,

$$\begin{aligned} \mathbf{A}(\theta_{1,2}) &= \begin{bmatrix} \mathbf{a}(\theta_1) & \mathbf{a}(\theta_2) \end{bmatrix} \\ \mathbf{b} &= \begin{bmatrix} b_1 & b_2 \end{bmatrix}^T, \qquad \|\mathbf{b}\| = 1 \end{aligned} \tag{3.24}$$

b Range R to the target and radar height h are known, i.e. $\theta_2 = f(\theta_1)$. Usually the radar location is known and the range could be determined if the radar uses some sort of frequency modulation technique and preprocessing prior to the elevation estimation. Use $\theta = H/R$ and $\theta_h = h/R$,

$$\begin{aligned} \mathbf{A}(\theta) &= \begin{bmatrix} \mathbf{a}(\theta - \theta_h) & \mathbf{a}(-\theta - \theta_h) \end{bmatrix} \\ \mathbf{b} &= \begin{bmatrix} b_1 & b_2 \end{bmatrix}^T, \qquad \|\mathbf{b}\| = 1 \end{aligned} \tag{3.25}$$

c The reflection coefficient G is a known function of elevation $G = G(\theta, h, R)$, with $\theta = H/R$. This implicitly requires that h, R are known. An interpretation of this case is that there is one signal source and one array response is formed as the sum of the two wavefronts impinging on the array,

$$\begin{aligned} \mathbf{A}(\theta) &= \mathbf{a}(\theta - \theta_h) + G(\theta)\mathbf{a}(-\theta - \theta_h) \\ \mathbf{b} &= 1 \end{aligned} \tag{3.26}$$

C2 **Optional assumptions on signal waveform**

a $\{x(t)\}_{t=1}^{N}$ are unknown deterministic complex variables

b $x(t) = x_0 e^{j\nu t}, \quad t = 1, \ldots, N$, where $x_0 \in C$ is constant and ν is the doppler frequency. Both x_0 and ν are unknown and have to be estimated, but if $N > 2$ this parameterization give less unknown parameters than case C2a above.

C3 **Optional noise assumptions**

a Spatially white noise of equal power in all sensors, $\mathbf{Q} = \mathrm{E}\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}^*(t) = \sigma\mathbf{I}$, where σ is unknown.

b Spatially colored noise, hence $\mathbf{Q}$ above is arbitrary and unknown.

4 Maximum-Likelihood Estimators

The chief systematic approach to most parameter estimation problems is the MLE. This technique requires a probabilistic setup of the problem at hand, i.e. the probability density function (pdf) of the observations, conditioned on the unknown parameters η. The conditioned pdf is also called the likelihood function. The ML estimate is obtained by maximizing the likelihood, which can be interpreted as finding the parameter set $\hat{\eta}$ that is the most likely set to have produced the observed data. This section begins with a brief review of some interesting papers where ML estimators of relevant DOA estimation problems have been studied. We see that the MLE for some of the cases we wish to study have been derived, but there is a need for a comprehensive survey. To do that we derive the MLE of all cases related to the present low-angle scenario in a unified framework. The derivations are given in Appendix A.1–A.4 and the resulting estimators are given by Equation (4.1), (4.3) and (4.5).

Since the MLE is a general method, providing different estimators for different signal assumptions, it is of vital importance to study the assumptions made when using an MLE derived by someone else. There are a few articles published that have derived MLE's for a low-angle multipath scenario [6, 7]. However, they assume that the noise power is different and unknown at different snapshots to allow the use of multiple frequencies.

A case studied thoroughly [2, 14–16] is the case of finding the DOA of n sources, where the k'th source signal $\{x_k(t)\}_{t=1}^N$, $k = 1, \ldots, n$ is a sequence of random, deterministic or stochastic, unknown parameters. In these papers the correlation between the source signal is arbitrary and unknown (coherent signals is a special case included) and the noise is spatially white. In [17, 18] a surprising fact is discovered: the MLE not exploiting the fact that the direct and specular reflections are coherent gives the same asymptotic (for large data sets) error variance as the MLE assuming coherent signals.

The spatially white noise assumption is often violated. The noise power in different sensors may not be identical and the noise in different sensors can easily be correlated if the equipment is not properly isolated. It is also a fact that clutter from the surroundings mainly arrives from ground level and it is hence by definition spatially correlated in elevation. An intuitive guess is that the MLE based on a correlated noise assumption is more robust to the presence of diffuse reflections from the sea. The problem of arbitrary spatial noise correlation requires some parameterization of the source signal to reduce the number of unknowns. In [19] there are n sources and their corresponding signals $\{x_k(t)\}_{t=1}^N$, $k = 1, \ldots, n$ are parameterized as a sum of known basis functions. In [20] one single source is assumed and $\{x(t)\}_{t=1}^N$ is parameterized by an unknown doppler frequency.

So the MLE for some of the cases we wish to study have been derived, however the similarities and differences between MLE's for various assumptions are easily lost when scattered in different papers. For completeness and comparison we provide the derivations of all the various MLE's in Appendix A.1-A.4.

The MLE for all three assumptions on multipath knowledge C1a–C1c can be expressed in one equation. This is easily done since the general signal model (3.7) can be used as is, without exploiting any specific structure to derive the ML estimator. Other papers tend to treat these cases separately, perhaps due to the fact that the implementation of the ML criterion will require different search algorithms and there is a significant improvement in performance as the knowledge about the generating system is increased.

MLE(C1a–c,C2a,C3a). The MLE corresponding to C2a (coherent, deterministic signals $\{x(t)\}_{t=1}^{N}$) and C3a (spatially white noise) is given by, see Appendix A.1 for a derivation

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \lambda(\boldsymbol{\theta}) \tag{4.1}$$

where

$$\lambda_1(\boldsymbol{\theta}) = \lambda_{\max}\left[(\mathbf{A}^*\mathbf{A})^{-1/2}\mathbf{A}^*\hat{\mathbf{R}}_{yy}\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1/2}\right] \tag{4.2}$$

where $\lambda_{\max}[\hat{\mathbf{R}}_{yy}]$ denotes the maximum eigenvalue of the data covariance matrix $\hat{\mathbf{R}}_{yy}$. Multipath assumptions C1a–c enter in the parameterization of $\mathbf{A}(\boldsymbol{\theta})$.

MLE(C1a–c,C2b,C3a). The MLE corresponding to waveform assumption C2b ($x(t) = x_0 e^{j\nu t}$) and noise assumption C3a is given by, see Appendix A.2 for a derivation

$$\hat{\boldsymbol{\theta}}, \hat{\nu} = \arg\max_{\boldsymbol{\theta},\nu} V(\boldsymbol{\theta}, \nu) \tag{4.3}$$

where, with definitions (3.10) and (3.11)

$$V(\boldsymbol{\theta}, \nu) = \mathbf{f}^*\mathbf{Y}^*\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\hat{\mathbf{Y}}\mathbf{f} \tag{4.4}$$

MLE(C1a–c,C2b,C3b). For the noise assumption C3b ($\mathbf{Q}$ is arbitrary and unknown) it is almost necessary to have a parametrized model for $x(t)$ for identifiability. We do not consider a case where $\mathbf{Q}$ and $x(t)$ are unknown simultaneously. Under the assumptions C2b and C3b the ML estimate of $\boldsymbol{\theta}$ and ν are given by, see Appendix A.3 for a derivation

$$\{\hat{\boldsymbol{\theta}}, \hat{\nu}\} = \arg\max_{\boldsymbol{\theta},\nu} V_R(\boldsymbol{\theta}, \nu) \tag{4.5}$$

where

$$V_R(\boldsymbol{\theta}, \nu) = \frac{\mathbf{f}^*\mathbf{Y}^*\hat{\mathbf{R}}_{yy}^{-1}\mathbf{A}(\mathbf{A}^*\hat{\mathbf{R}}_{yy}^{-1}\mathbf{A})^{-1}\mathbf{A}^*\hat{\mathbf{R}}_{yy}^{-1}\mathbf{Y}\mathbf{f}}{1 - \mathbf{f}^*\mathbf{Y}^*\hat{\mathbf{R}}_{yy}^{-1}\mathbf{Y}\mathbf{f}} \tag{4.6}$$

In Appendix A.4 the MLE for multiple frequencies is derived, which is based on the MLE expressions above. However, the multiple frequency MLE is discussed separately in Section 12.

5 Bounds

Bounds on the error variance are tools of great value when evaluating the expected impact of different assumptions and the performance of various estimators. If an unbiased estimator that attains the lower bound is found, it is not physically possible to find another unbiased estimator whose variance is less. A bound which is achievable can be useful when designing the array structure that gives optimal conditions for the corresponding estimator. However, there are many existing bounds and only the tightest (highest) lower bound can possibly be achievable. The CRB [4] is one of the easiest to determine and is by far the most used. But if the problem studied suffers from ambiguities the CRB will be too optimistic. A tighter bound that can take ambiguities into account is the BB [9, 10].

The CRB has been derived earlier for different special cases of DOA estimation and low angle scenarios [15, 20, 21], but here we show that the CRB for $\boldsymbol{\theta}$ can be expressed in one compact equation, including all combinations of model assumptions, which is easily evaluated. See Appendix B for derivations.

$$\mathrm{CRB}(\boldsymbol{\theta}) = \frac{1}{2N\hat{\mathbf{R}}_{xx}}\left(\mathbf{D}^*\mathbf{Q}^{-1}\mathbf{D} - \mathbf{D}^*\mathbf{Q}^{-1}\mathbf{A}\left(\mathbf{A}^*\mathbf{Q}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^*\mathbf{Q}^{-1}\mathbf{D}\right)^{-1} \tag{5.1}$$

where $\mathbf{D}$ contains the derivatives with respect to elevation according to definition (3.13). The array response $\mathbf{A}$ is parameterized by the p elevation parameters $\theta_1, \ldots, \theta_p$. For C1a $p = 2$, while for C1b and C1c $p = 1$. It turns out that the CRB for elevation is the same for an unknown signal waveform and a signal parameterized by the doppler frequency, C2a and C2b respectively. Furthermore it is seen that if the noise is white, assumption C3a and C3b will give the same CRB.

The BB was originally derived by Barankin [9], but in [10] a simpler form of the BB for the error variances of unbiased estimation of vector parameters has been given. It was shown that the BB can be written as a sum of two terms where the first term is the CRB, which involves the signal structure near the true parameter values, while the second term takes into account system performance at other parameter values. The pdf for the observations $\mathbf{Y}$ conditioned on the vector of unknown parameters $\boldsymbol{\eta}$ is given by

$$p(\{\mathbf{y}(t)\}|\eta_i) = \frac{1}{\pi^{mN}\left|\mathbf{Q}_i\right|^N}\exp\sum_{t=1}^{N}\left(\mathbf{y}(t) - \boldsymbol{\mu}(t)\right)^*\mathbf{Q}_i^{-1}\left(\mathbf{y}(t) - \boldsymbol{\mu}(t)\right) \tag{5.2}$$

where

$$\boldsymbol{\mu}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{b}x(t) \tag{5.3}$$

Using (5.2) the BB can be written as

$$\mathrm{E}[(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^*] \geq \mathbf{F}^{-1} + (\mathbf{T} - \mathbf{F}^{-1}\mathbf{U})\boldsymbol{\Delta}^{-1}(\mathbf{T} - \mathbf{F}^{-1}\mathbf{U})^* \tag{5.4}$$

where

$$\mathbf{\Delta} = \mathbf{B} - \mathbf{U}^*\mathbf{F}^{-1}\mathbf{U} \tag{5.5}$$

$$\begin{aligned}\mathbf{F}_{i,j} &= \int_{C^{mN}} \frac{\partial \ln p(\mathbf{Y}|\boldsymbol{\eta})}{\partial \eta_i} \frac{\partial \ln p(\mathbf{Y}|\boldsymbol{\eta})}{\partial \eta_j} \cdot p(\mathbf{Y}|\boldsymbol{\eta}) d\mathbf{Y} = \\ &= 2\mathrm{Re} \sum_{t=1}^{N} \left\{ \left(\frac{\partial \boldsymbol{\mu}(t)}{\partial \eta_i} \right)^* \mathbf{Q}^{-1} \left(\frac{\partial \boldsymbol{\mu}(t)}{\partial \eta_j} \right) \right\} \qquad i,j = 1, \ldots, n_\eta\end{aligned} \tag{5.6}$$

$$\begin{aligned}\mathbf{U} &= \int_{C^{mN}} \frac{\partial \ln p(\mathbf{Y}|\boldsymbol{\eta})}{\partial \eta_i} \cdot p(\mathbf{Y}|\boldsymbol{\eta}_k) d\mathbf{Y} = \\ &= 2\mathrm{Re} \sum_{t=1}^{N} \left\{ \left(\frac{\partial \boldsymbol{\mu}(t)}{\partial \eta_i} \right)^* \mathbf{Q}^{-1} \Big(\boldsymbol{\mu}_k(t) - \boldsymbol{\mu}(t) \Big) \right\} \qquad i = 1, \ldots, n_\eta; \quad k = 1, \ldots, n_t\end{aligned} \tag{5.7}$$

$$\mathbf{B} = \int_{C^{mN}} \frac{p(\mathbf{Y}|\boldsymbol{\eta}_k)}{p(\mathbf{Y}|\boldsymbol{\eta})} \frac{p(\mathbf{Y}|\boldsymbol{\eta}_p)}{p(\mathbf{Y}|\boldsymbol{\eta})} \cdot p(\mathbf{Y}|\boldsymbol{\eta}) d\mathbf{Y} \qquad k,p = 1, \ldots, n_t \tag{5.8}$$

$$\mathbf{T} = \begin{bmatrix} \boldsymbol{\eta}_1 & \cdots & \boldsymbol{\eta}_{n_t} \end{bmatrix} \tag{5.9}$$

Notice that the η_i, $i = 1, \ldots, n_\eta$ refers to the ith component of the true parameter vector $\boldsymbol{\eta}$, while $\boldsymbol{\eta}_k$, $k = 1, \ldots, n_t$ refers to an arbitrary value of the parameter vector $\boldsymbol{\eta}$ other than its true value. These vectors are called test-vectors and are to be chosen by the user. The number of test-vectors, n_t, is also arbitrary. The matrix $\mathbf{F}$ is recognized as the Fisher information matrix for unbiased estimation. If we choose $n_t = 0$, the BB equals the CRB. A more precise expression for $\mathbf{F}$ is given in the Appendix B along with the CRB for elevation and the derivatives needed to compute $\mathbf{U}$. The $\mathbf{B}$ matrix, evaluated for the n_t test-vectors is derived in Appendix C and can be formulated as

$$\begin{aligned}B(k,p) = \left(\frac{|\mathbf{Q}| \left| \tilde{\mathbf{Q}}_{kp} \right|}{|\mathbf{Q}_k| \, |\mathbf{Q}_p|} \right)^N \exp \sum_{t=1}^{N} \Big\{ &- \left\| \mathbf{Q}_k^{-1/2} \boldsymbol{\mu}_k(t) \right\|^2 - \left\| \mathbf{Q}_p^{-1/2} \boldsymbol{\mu}_p(t) \right\|^2 + \left\| \mathbf{Q}^{-1/2} \boldsymbol{\mu}(t) \right\|^2 \\ &+ \left\| \tilde{\mathbf{Q}}_{kp}^{1/2} \left(\mathbf{Q}_k^{-1} \boldsymbol{\mu}_k(t) + \mathbf{Q}_p^{-1} \boldsymbol{\mu}_p(t) - \mathbf{Q}^{-1} \boldsymbol{\mu}(t) \right) \right\|^2 \Big\} \qquad k,p = 1, \ldots, n_t\end{aligned} \tag{5.10}$$

where

$$\tilde{\mathbf{Q}}_{kp} = [\mathbf{Q}_k^{-1} + \mathbf{Q}_p^{-1} - \mathbf{Q}^{-1}]^{-1} \tag{5.11}$$

When choosing the test-vectors there are a few facts worth noticing about the Barankin bound. First, the addition of one or more test-vectors never reduce the bound since the term $(\mathbf{T} - \mathbf{F}^{-1}\mathbf{U})\mathbf{\Delta}^{-1}(\mathbf{T} - \mathbf{F}^{-1}\mathbf{U})^*$ is at least positive semidefinite [10]. However, if several test-vectors are used, care must be taken to avoid numerical errors when computing the inverse $\mathbf{\Delta}^{-1}$ which easily becomes ill conditioned. A bad choice is a test-vector that brings no additional information. So in practice there is a limit of how many test-vectors that are feasible. Moreover, it is possible in most practical examples to get a good estimate of the maximum achievable Barankin bound by using one properly chosen test point. Good test-vectors are usually points which are nearly ambiguous with respect to the true parameters.

Since ambiguous points are found for parameters giving a local maximum of the likelihood function we can choose points in a way that resemble the MLE. For a given test-parameter θ_k the corresponding choice of $\mathbf{b}_k$, and $\{x_k(t)\}_{t=1}^N$ are parameter values that maximize the function

$$V(\eta_k|\eta_0) = \sum_{t=1}^{N} \Big(\mathbf{A}_0\mathbf{b}_0 x_0(t) - \mathbf{A}_k\mathbf{b}_k x_k(t) \Big)^* \mathbf{Q}^{-1} \Big(\mathbf{A}_0\mathbf{b}_0 x_0(t) - \mathbf{A}_k\mathbf{b}_k x_k(t) \Big) \tag{5.12}$$

giving

$$\mathbf{b}_k = (\mathbf{A}_k\mathbf{Q}^{-1}\mathbf{A}_k)^{-1}\mathbf{A}_k^*\mathbf{A}_0\mathbf{b}_0 \tag{5.13}$$

$$\tag{5.14}$$

To compute the BB end up in finding some candidates θ_k that give a local maximum of $V(\eta|\eta_0)$ and to compute the BB, Equation (5.4), for these candidates and determine the maximum BB out of these ones. However, this is only one way of choosing the testvectors and there may be better choices. But to combine two or more of these testvectors will lead to numerical problems and does not seem to increase the bound compared to only using the best one of them.

In Figure 2 and 3 the CRB for assumptions C1a–c and the BB for C1c are plotted for different elevations and SNR respectively. The bounds indicate that the error variance will fluctuate as the target elevation changes, i.e. as the relative phase shift between direct and specular path fluctuates. CRB(C1a) > CRB(C1b) >>CRB(C1c) which is natural as the knowledge increases. The fact that CRB(C1b) >>CRB(C1c) indicates that the reflection coefficient in C1c carries a lot of information about the target location. However, since the CRB is a measure of the curvature of the global minimum of the negative log-likelihood function, and CRB cannot not consider nearly ambiguous minima. The BB can take ambiguous minima into account and is always tighter than the CRB. For assumptions C1a and C1b there are no ambiguities to consider and the BB would equal the CRB, while for C1c there is a significant difference for low SNR, see Figure 3. The BB(C1c) and the CRB(C1c) coincide at high SNR. These bounds are only lower bounds, showing that there is no unbiased estimator that for the given assumptions could give an RMSE below the tightest bound. But there might be tighter bounds than the ones given. However, in the threshold region the BB(C1c) proves that the CRB(C1c) is not achievable.

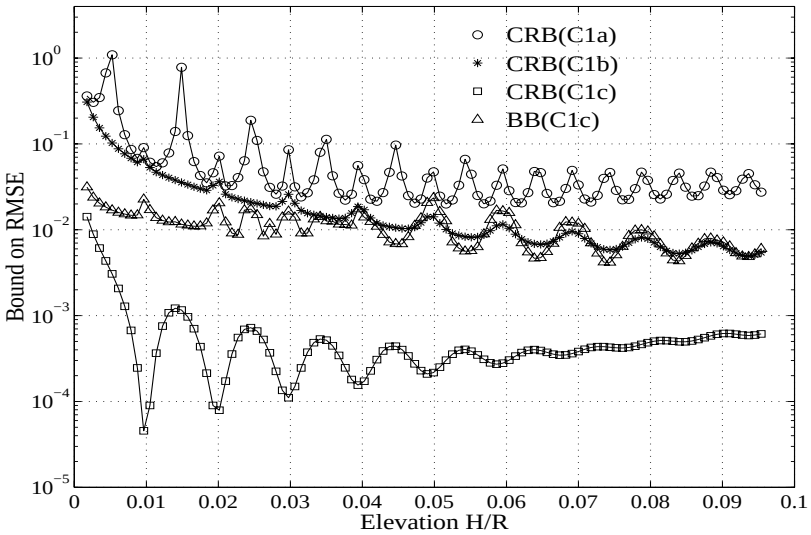

Figure 2. CRB and BB at different target elevations. Model assumptions C1a-b, SNR=-5dB.

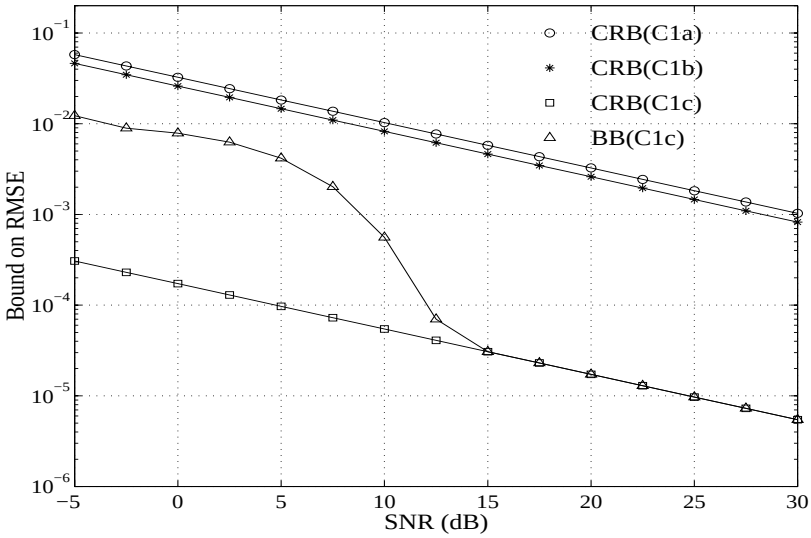

Figure 3. CRB and BB at different SNR. Model assumptions C1a-c $H_0/R = 0.021$.

6 Alternative estimators

To compute the exact MLE involves a search to find the parameters that minimize or maximize the nonlinear MLE criterion. This search can be computationally very demanding and quicker solutions that give approximately the same performance are of great interest. In this section we give some suggestions of alternative methods for various combinations of signal assumptions.

6.1 Subspace methods

Various subspace methods for DOA estimation have been derived. Assumptions common to most subspace methods are that the noise is spatially white (C3a) and that the signals $\{x(t)\}_{t=1}^{N}$ are unknown stochastic variables. Usually the different sources are also assumed to be noncoherent. However, the case of coherent deterministic signals (C2a) which we are interested in is also studied. In [17, 18] it was established that the deterministic coherent signal MLE and the stochastic coherent signal MLE are asymptotically identical. It was also shown that these two methods are asymptotically identical to the stochastic MLE not exploiting the coherency. This gives us the motivation of studying the MODE and WSF subspace estimators which are large sample realizations of the stochastic MLE. The frequently used MUSIC estimator (also a subspace method) will however break down if the sources are coherent and give severely biased estimates. This illustrates the importance of using methods derived under appropriate assumptions. A good review of all these methods is found in [2].

The MODE/WSF is a subspace method based on the eigendecomposition of the sample covariance matrix $\hat{\mathbf{R}}_{yy}$ of $\mathbf{y}(t)$, into the signal and noise subspace. If there are n signal sources the eigendecomposition is written as

$$\hat{\mathbf{R}}_{yy} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{y}(t)\mathbf{y}^*(t) = \mathbf{S}\Lambda_s\mathbf{S}^* + \mathbf{E}\Lambda_e\mathbf{E}^* \tag{6.1}$$

where the n columns of $\mathbf{S}$ are the signal eigenvectors corresponding to the n largest eigenvalues. The columns of $\mathbf{E}$ are the noise eigenvectors corresponding to the $m-n$ smallest eigenvalues. The diagonal matrices Λ_s, Λ_e contain the eigenvalues. The WSF estimator is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\left(Tr(\mathbf{I} - \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*)\mathbf{S}\mathbf{W}\mathbf{S}^* \right) \tag{6.2}$$

where

$$\mathbf{W} = (\Lambda_s - \hat{\sigma}\mathbf{I})^2\Lambda_s^{-1} \tag{6.3}$$

and $\hat{\sigma} = \frac{1}{m-n}Tr\Lambda_e$ is a consistent estimate of σ, In the case of two coherent signal sources, we know that the best estimate of $\mathbf{W}$ is a singular matrix and only the principal eigenvector $\mathbf{s}_1$ of $\hat{\mathbf{R}}$ is relevant to the estimate (6.2). The resulting WSF estimator is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\left(\mathbf{s}_1^*(\mathbf{I} - \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*)\mathbf{s}_1 \right) \tag{6.4}$$

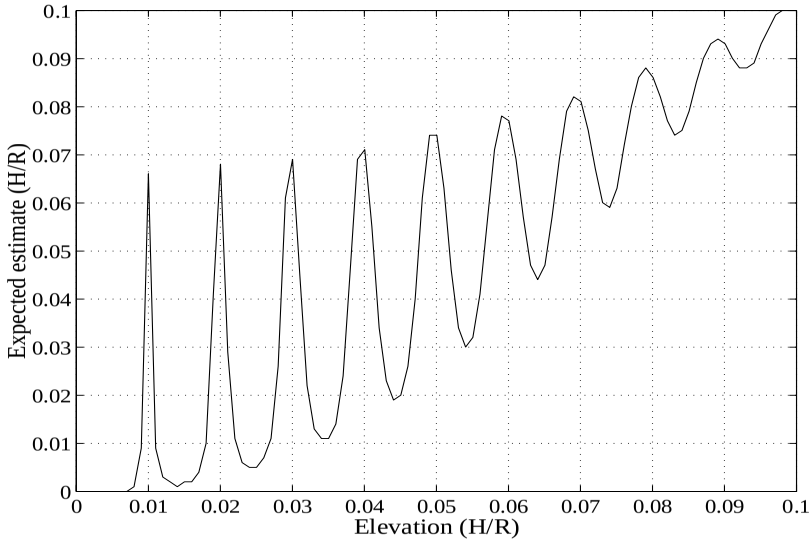

Figure 4. Biased MUSIC estimate for noise free data of the default radar setup, (3.23).

All previous methods involve a search to find the elevation estimate, and a closed form estimator would be of great interest. If the array is uniform and linear and C1a is assumed, the WSF can be rewritten into the equivalent MODE algorithm [2, 14]. It is also easy to modify the derivation to exploit the assumption C1b, but the nonlinear array response of C1c cannot be used in an efficient MODE algorithm. Running the MODE algorithm for case C1a or C1b to find the estimate only involves a simple iterative search, but since the derivation is quite lengthy it has been placed in the Appendix D.

Finally, it is also important to stress that MUSIC, a subspace method frequently used for DOA estimation, fails when the signals are coherent. It has been shown [15] that in case of uncorrelated sources MUSIC is a large sample ML estimator, but for correlated sources MUSIC is not statistically efficient and for coherent sources MUSIC breaks down. If the standard MUSIC estimator were implemented to try to estimate the elevation of a low angle target flying over sea, a severely biased estimate would be found. This is due to the fact that two coherent signals sum up to one wavefront that cannot be described by either the direct or indirect signal. MUSIC can only detect one signal source and it tries to fit the received array output to one source. Depending on whether the two signal paths arrive in or out of phase the estimate bias will fluctuate in a deterministic pattern as the target moves. The MUSIC estimator is given by the following equation and the result showin its bias is plotted in Figure (4).

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(\mathbf{a}^*(\theta)(\mathbf{I} - \mathbf{s}_1\mathbf{s}_1)\mathbf{a}(\theta) \right) \tag{6.5}$$

6.2 MODE in two dimensions

Let us assume spatially white noise (C3a) and that the signal $\{x(t)\}_{t=1}^{N}$ is parameterized by an unknown doppler frequency, case C2b. Finding the exat ML estimate exploiting that knowledge (4.3) involves a computationally demanding search in elevation and doppler frequency. Nevertheless, there are quick and accurate estimators available. Given a doppler

estimate $\hat{\nu}$ the elevation estimate can be solved for explicitly in case of C1a and C1b, using the method of direction estimation (MODE) [2, 19, 20]. A requirement for MODE is that the data has been collected with a uniform linear array (ULA). The elevation estimates are very robust to errors in $\hat{\nu}$, which reasonable since the CRB was unaffected by the introduction of the refined parameterization, and hence a simple DFT to find the initial doppler estimate will work very well. Given the elevation estimate, the refined doppler estimate can be found using a bank of DFT's or it can also be solved for explicitly using MODE.

The maximization (4.3) can be rewritten in two alternative ways

$$\hat{\boldsymbol{\theta}}, \hat{\nu} = \arg\min_{\boldsymbol{\theta},\nu} \Big(-V(\boldsymbol{\theta}, \nu) \Big) = \tag{6.6}$$

$$= \arg\min_{\boldsymbol{\theta},\nu} \Big(\mathrm{Tr}\,(\mathbf{I} - \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*)\mathbf{Y}\mathbf{f}\mathbf{f}^*\mathbf{Y}^* - \mathbf{f}^*\mathbf{Y}^*\mathbf{Y}\mathbf{f} \Big) \tag{6.7}$$

$$= \arg\min_{\boldsymbol{\theta},\nu} \Big(\mathrm{Tr}\,(\mathbf{I} - \mathbf{f}\mathbf{f}^*)\mathbf{Y}^*\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{Y} - \mathrm{Tr}\,\mathbf{Y}^*\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{Y} \Big) \tag{6.8}$$

Finding the minimum with respect to $\boldsymbol{\theta}$ or ν, given a fixed estimate of the other parameter, reduces to finding the minimum of the first term in (6.7) or (6.8) respectively. The final estimate is given by reiterations with one parameter fixed at the time. The non-search MODE solution is applicable to both cases if the columns of $\mathbf{A}$ and $\mathbf{f}$ have the ULA structure with spatial frequencies $\omega_1, \ldots, \omega_{na}$ and doppler frequencies $\nu_1, \ldots, \nu_{nf}$. In other words, the non search MODE solution is applicable if $\mathbf{A}$ and $\mathbf{f}$ have a Vandermonde structure, see Appendix D

In order to start the MODE iteration we need a consistent initial estimate of $\boldsymbol{\theta}$ and ν. A suggestion is the simpler separated estimator, based on the extended invariance principle [20], where the doppler estimate is found using an unstructured model estimating $\tilde{\mathbf{a}} = \mathbf{A}\mathbf{b}$ instead of $\boldsymbol{\theta}, \mathbf{b}$.

$$\hat{\nu} = \arg\max_{\nu} \Big(\mathbf{f}^*\mathbf{Y}^*\mathbf{Y}\mathbf{f} \Big) \tag{6.9}$$

$$\hat{\boldsymbol{\theta}}(\hat{\nu}) = \arg\max_{\boldsymbol{\theta}} \Big(\hat{\mathbf{f}}^*\mathbf{Y}^*\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}\mathbf{Y}\hat{\mathbf{f}} \Big) \tag{6.10}$$

In [20] it was shown that this estimator gives an error variance comparable to MLE and this estimate could hence be good enough without reiterations over $\boldsymbol{\theta}, \nu$.

6.3 Separated estimator

If the spatial noise correlation is arbitrary and unknown (C3b) and the doppler parameterization (C2b) is valid a simplified separated estimator similar to the one above can be obtained by using the extended invariance principle [20]. The estimator is written as follows

$$\hat{\nu} = \arg\max_{\nu} \left(\mathbf{f}^* \mathbf{Y}^* \hat{\mathbf{R}}_{yy}^{-1} \mathbf{Y} \mathbf{f} \right) \tag{6.11}$$

$$\hat{\boldsymbol{\theta}}(\hat{\nu}) = \arg\max_{\boldsymbol{\theta}} \left(\hat{\mathbf{f}}^* \mathbf{Y}^* \hat{\mathbf{R}}_{yy}^{-1} \mathbf{A} (\mathbf{A}^* \hat{\mathbf{R}}_{yy}^{-1} \mathbf{A})^{-1} \mathbf{A}^* \hat{\mathbf{R}}_{yy}^{-1} \mathbf{Y} \hat{\mathbf{f}} \right) \tag{6.12}$$

which can be solved for explicitly with the MODE solution if the array is ULA and if case C1a and C1b is assumed. For C1c a search in elevation is necessary.

7 Multipath knowledge

In this section we will discuss the ML estimator for the different assumptions on multipath knowledge C1a–C1c, given unknown signal waveform $\{x(t)\}_{t=1}^{N}$ (case C2a) and spatially white noise (case C3a). The multipath knowledge has a very large impact on how the estimates are distributed and it will become obvious that much can be gained when fully exploiting the signal model.

7.1 The ML criterion

The ML estimator for all three assumptions on multipath knowledge C1a–C1c have been expressed in one equation, (4.1). If the reflection coefficient G is unknown (case C1a or C1b), computing the MLE involves an eigenvalue decomposition of a 2×2 matrix. But if G is a known function of elevation (case C1c), the MLE does not require any eigendecomposition. On the other hand, the ML criterion corresponding to the latter signal model is a multi modal criterion function with several very sharp, almost ambiguous maxima. We can clearly see the problem if we plot the ML criterion, Equation (4.1), for the tree different multipath assumptions (C1a–c) for a noise free case, Figure 5 and 6. The ML criterion plotted is derived under assumption C2a and C3a and the radar setup used is the default setup in Section 3. The noise free signal is generated to simulate a target at elevation $H/R = 0.026$.

- C1a: The ML criterion when no geometry relations are used is plotted as the surface in Figure 5. The maximum is found for the elevations $\theta_{1,2} = \pm H/R - h/R = (\pm 0.026 - 5 * 10^{-4})$. The maximum is very flat. The sharpest line through the maximum point corresponds to assumption C1b. Also notice that the criterion is close to maximum when θ_1 and/or θ_2 is close to elevation zero. Hence, it is difficult to find two parameter values corresponding to a distinct maximum.

- C1b: When exploiting the geometry of the multipath scenario, we know that $\theta_{1,2} = \pm H/R - h/R$ which corresponds to a line close to the diagonal in Figure 5 and the smooth line in Figure 6.

- C1c: If the function for the reflection coefficient is known the ML criterion shows several closely spaced, very sharp but almost ambiguous maxima. We hence expect to find estimates distributed in a few almost discrete levels corresponding to the ambiguous maximum points.

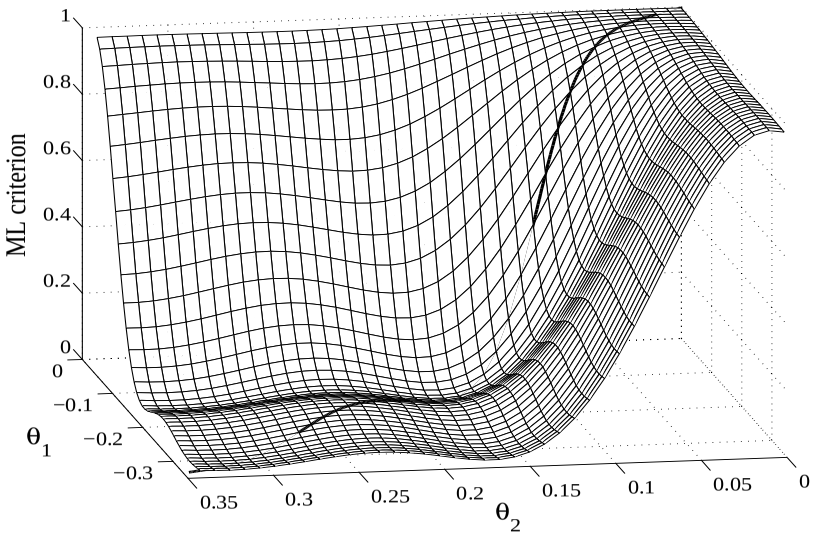

Figure 5. ML criterion for model assumptions C1a and C1b in a noise free case. The signal is generated for elevation $H/R = 0.026$.

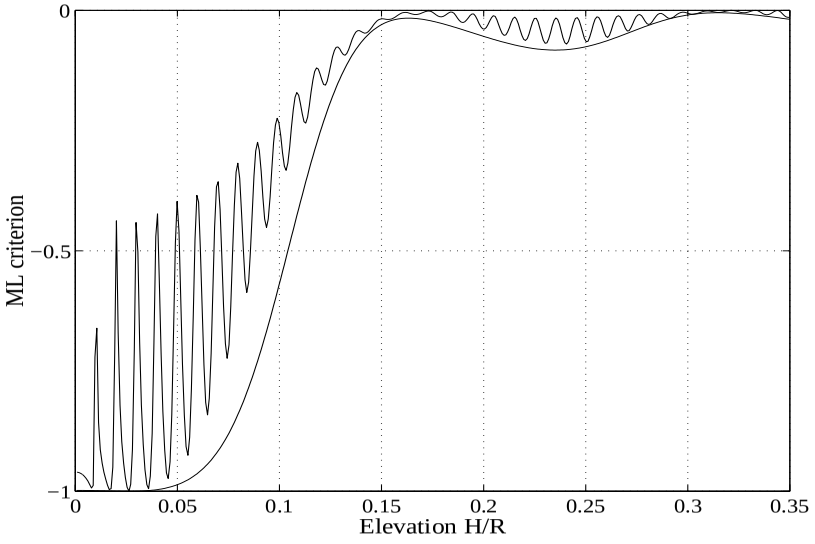

Figure 6. ML criterion for the model assumptions C1b and C1c in a noise free case, $H_/R = 0.021$.

7.2 Distribution of estimates

The shape of the ML criterion can explain the distribution of estimates which is seen in simulations. Figure 7–9 show how the estimates from Monte Carlo simulations are distributed by plotting the histogram of 200 realizations. The default radar setup is used and the signal is generated to simulate a target at elevation $H/R = 0.021$ with $x(t) = 1$, $t = 1, \ldots, N$ and spatially white noise corresponding to SNR = 0dB. The MLE's are based on assumptions C2a and C3a. The true elevation H/R (or $\pm H/R - h/R$) corresponds to the vertical dashed line.

- C1a: Two source DOA are to be estimated. The result tends to be either two resolved unbiased source DOA or a DOA corresponding to the signal center (close to elevation zero) and a second DOA corresponding to a strong noise direction. Without prior knowledge of the geometry there is no way of telling what type of estimates that are obtained.

- C1b: If we use the geometry constraints on the reflection and search for positive elevations only, the estimates will cluster around the true elevation. But a large percentage of the estimates equal zero, a phenomenon which can be interpreted as if the sources are not separable.

- C1c: If the reflection coefficient is known as a function of elevation, only elevations giving rise to multipaths with appropriate phase difference are possible maximum points of the MLE criterion. The estimates will be distributed over a few distinct elevations, which all give almost the same response at the antenna array.

The different types of distributions of the estimates for the three optional assumptions on the multipath will appear for both known and unknown signal waveforms and regardless of the noise assumption.

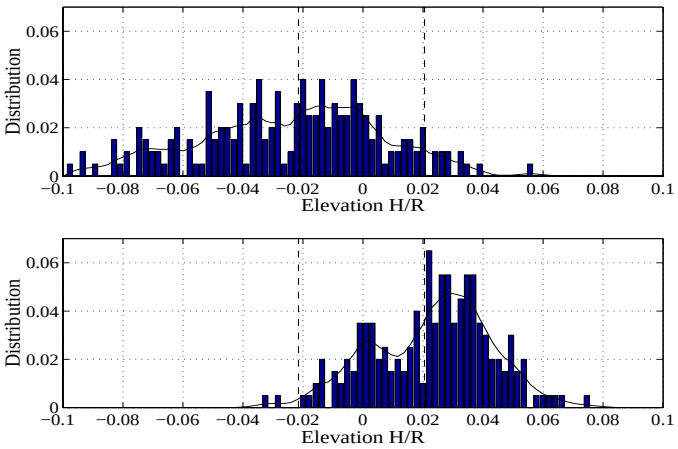

Figure 7. Distribution of ML estimates given assumptions C1a, C2a, C3a. SNR = 0dB, $H/R = 0.021$.

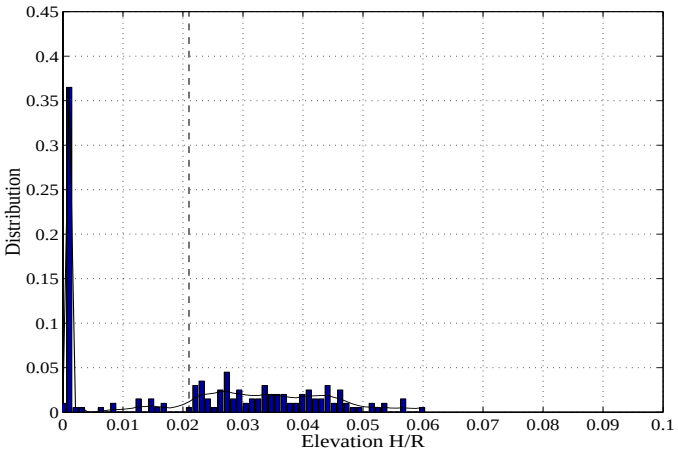

Figure 8. Distribution of ML estimates given assumptions C1b, C2a, C3a. SNR = 0dB, $H/R = 0.021$.

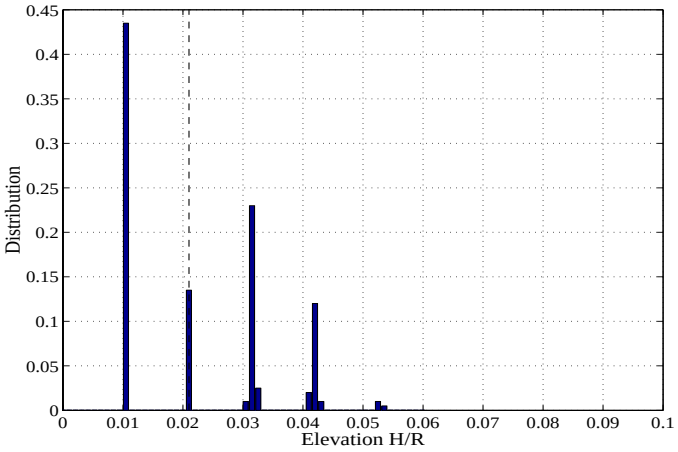

Figure 9. Distribution of ML estimates given assumptions C1c, C2a, C3a. SNR = 0dB, $H/R = 0.021$.

7.3 RMSE of the MLE compared with bounds

In Figure 10–15 we compare the RMSE of the MLE's corresponding to assumptions C1a–c to their respective lower bound on error variance. We do this for varying SNR and elevations. The MLE's are based on assumptions C2a and C3a. The default radar setup is used and each RMSE value in the plots is estimated from 200 Monte Carlo realizations. In Figure 10, 11 and 12 the observations are generated to simulate a target at elevation $H/R = 0.021$ with $x(t) = 1$, $t = 1, \ldots, N$ and spatially white noise of varying SNR. In Figure 13, 14 and 15 the target elevation H/R varies while SNR = 0dB is fixed. For assumption C1a there are two elevations to be estimated and the RMSE is computed according to Equation (3.20).

- C1a: The MLE achieves the CRB for high SNR, Figure 10, and is hence asymptotically efficient. In Figure 13 we see MLE fluctuates along with the CRB. The SNR = 0dB is rather low and we see that for higher elevations the MLE is poor compared to the bound, while at low elevations the MLE has an error variance smaller than the CRB. But the distribution of estimates revealed that the estimates are biased when the two DOA's in C1a are not separated, which happens for low SNR. The CRB and BB are bounds for bias free estimators only.

- C1b: When exploiting the symmetry properties of C1b the estimates naturally become more accurate. The MLE achieves the CRB at high SNR, Figure 17. At low SNR the MLE even gives lower variance than the CRB, due to a large concentration of zero estimates when the sources are inseparable.

- C1c: Here we encounter a threshold behavior. For high SNR the MLE always finds the maxima corresponding to the true elevation and achieves the BB, which in fact also equals the CRB at high SNR values. Below the threshold SNR (about 25 dB) the estimates suffer from ambiguities. Without the BB a user may believe that the estimators are poor since they do not come close to achieve the CRB. But the BB predicts the threshold, and no method can do better except perhaps in the transition region between very high and very low SNR. There is room for some improvement of the estimators in the transition region, but it is also possible that there are other test-points for the BB that could give an even tighter bound, see Section 5.

We now have shown that these methods achieve or at least approach the bounds for the different assumptions C1a–c. Since we also have shown that the CRB is the same for a signal waveform parameterized by the doppler frequency C2b, (see Appendix B) we can not possibly hope to improve the high SNR accuracy by exploiting the extra information in C2b (see Section 9). The fact that the ML estimators achieve the bounds makes the comparision between the bounds in Figure 2 and 3 more valuable as a tool to relate the expected improvement in performance when exploiting more knowledge about the multipath geometry.

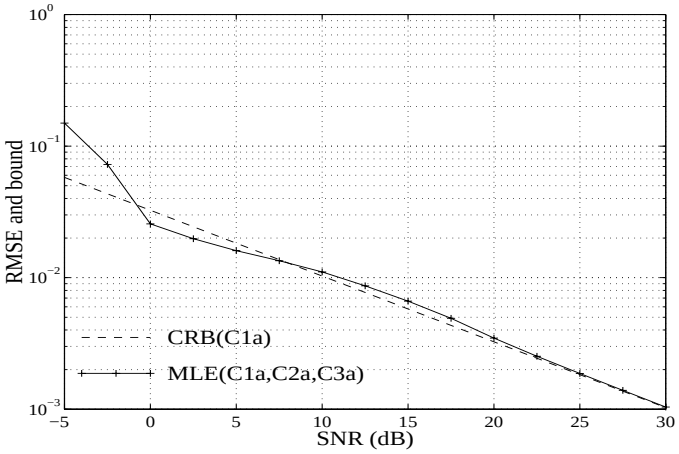

Figure 10. RMSE of MLE(C1a,C2a,C3a) and the corresponding CRB for varying SNR, $H_0/R = 0.021$.

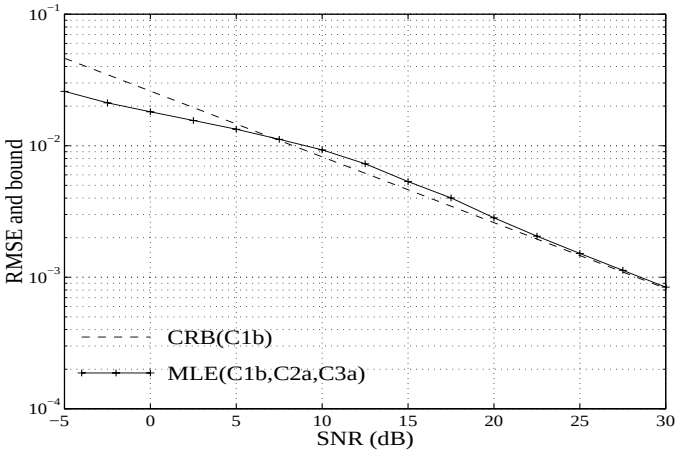

Figure 11. RMSE of MLE(C1b,C2a,C3a) and the corresponding CRB for varying SNR, $H_0/R = 0.021$.

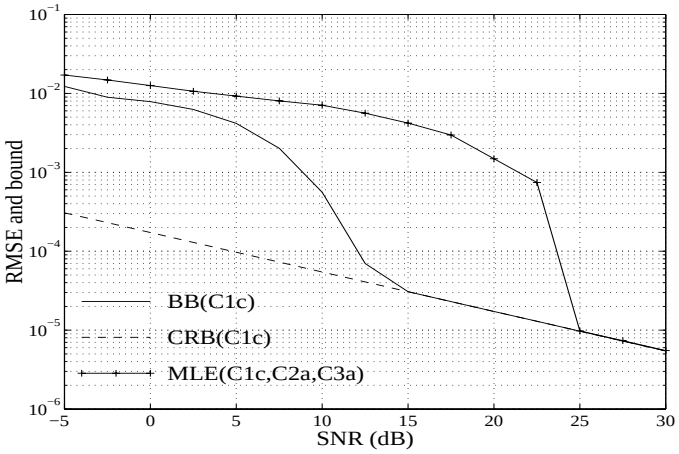

Figure 12. RMSE of MLE(C1c,C2a,C3a) and the corresponding CRB for varying SNR, $H_0/R = 0.021$.

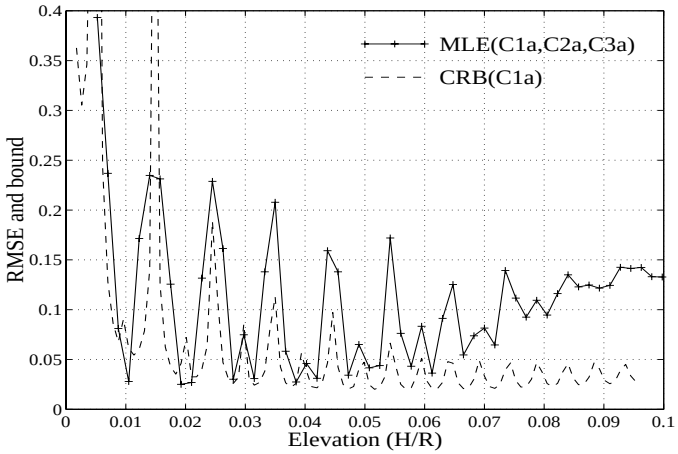

Figure 13. RMSE of MLE(C1a,C2a,C3a) and the corresponding CRB for varying elevations H/R, SNR = 0dB.

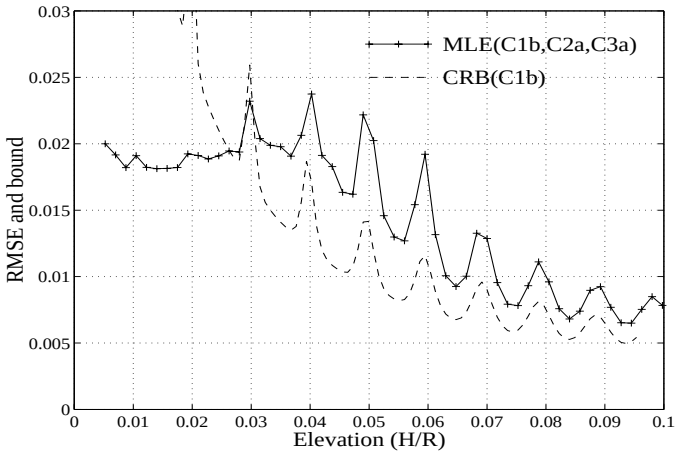

Figure 14. RMSE of MLE(C1b,C2a,C3a) and the corresponding CRB for varying elevations H/R, SNR = 0dB.

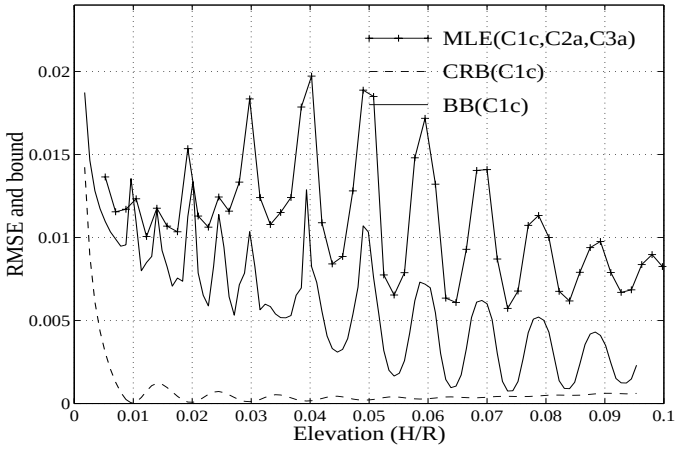

Figure 15. RMSE of MLE(C1c,C2a,C3a) and the corresponding CRB for varying elevations H/R, SNR = 0dB.

8 Methods, a performance comparision

The subspace techniques WSF and MODE are compared with the MLE in simulations, and are shown to give similar performance. It is only at low SNR that the methods differ some in estimation error variance. To achieve good results it is much more important to use the best model possible than to fine tune an ML or subspace method.

The same signal scenario as in previous study of the MLE is used. In simulations, Figure 16–21, we plot the RMSE for the different methods MLE, WSF and MODE and the corresponding bounds for varying SNR and varying elevations. All estimators are based are based on assumptions C2a and C3a. The default radar setup is used and each RMSE value in the plots is estimated from 200 Monte Carlo realizations. In Figures 16, 17 and 18 the observations are generated to simulate a target at elevation $H/R = 0.021$ with $x(t) = 1$, $t = 1, \ldots, N$ and spatially white noise of varying SNR. In Figures 19, 20 and 21 the target elevation H/R varies while SNR = 0dB is fixed. For assumption C1a there are two elevations to be estimated and the RMSE is computed according to Equation 3.20.

- C1a: The MLE, WSF and MODE all achieve the CRB for high SNR, Figure 16, and are hence asymptotically efficient estimators. For low SNR there is a difference in robustness between the different methods, which is clearly seen in Figure 19 where the RMSE is plotted for varying elevations at low SNR (0 dB). First of all note that the error variance fluctuates as the relative phase between direct path and specular path varies. For C1a, where we try to separate the direct and specular path, MODE has superior performance compared to MLE and WSF, which verifies the results of [21].

- C1b: In Figure 17 and 20 we see no significant difference in performance between the methods except at low SNR and very low elevations where WSF is more robust than the other two methods.

- C1c: The MLE and WSF give practically the same threshold behaviour, Figure 18, and the same fluctuation in performance with elevation, Figure 21.

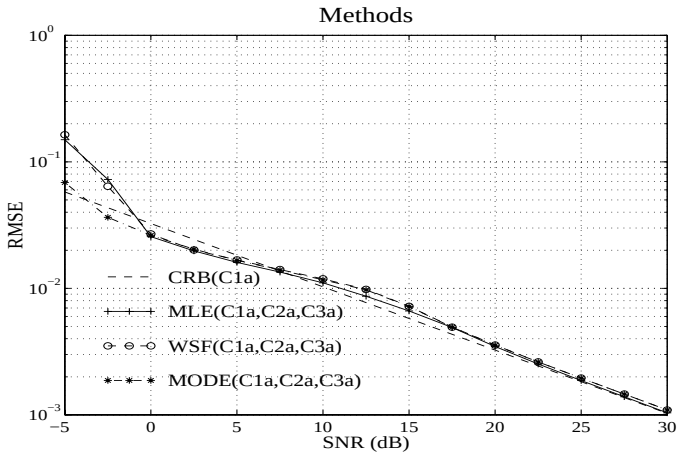

Figure 16. RMSE for MLE, WSF and MODE. Assumption C1a, C2a, C3a $H_0/R = 0.021$.

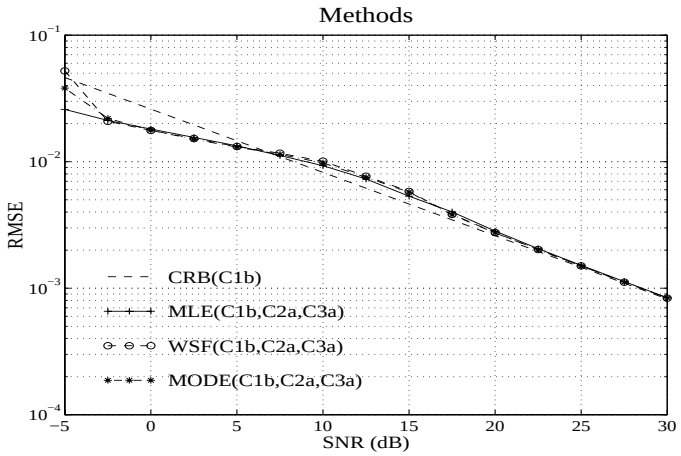

Figure 17. RMSE for MLE, WSF and MODE. Assumption C1b, C2a, C3a $H_0/R = 0.021$.

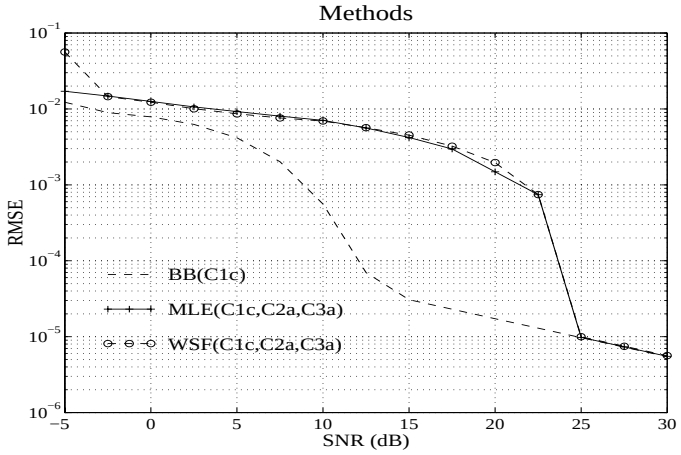

Figure 18. RMSE for MLE, WSF and MODE. Assumption C1c, C2a, C3a $H_0/R = 0.021$.

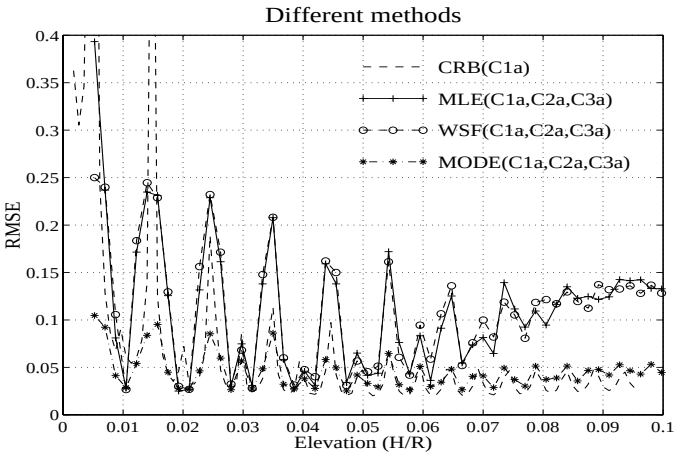

Figure 19. RMSE for MLE, WSF and MODE. Assumption C1a, C2a, C3a, SNR = 0dB.

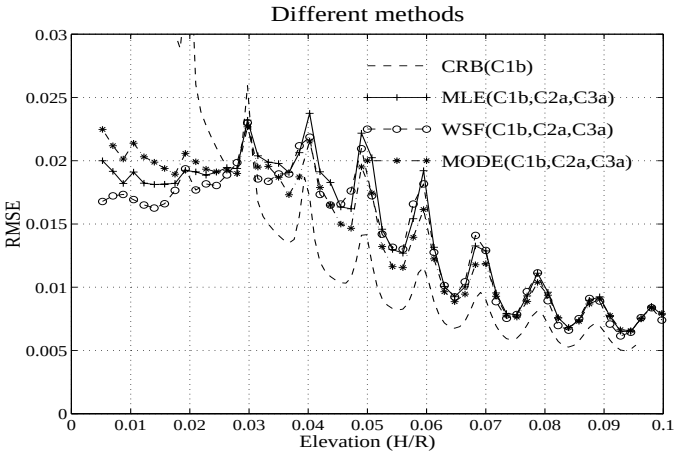

Figure 20. RMSE for MLE, WSF and MODE. Assumption C1b, C2a, C3a, SNR = 0dB.

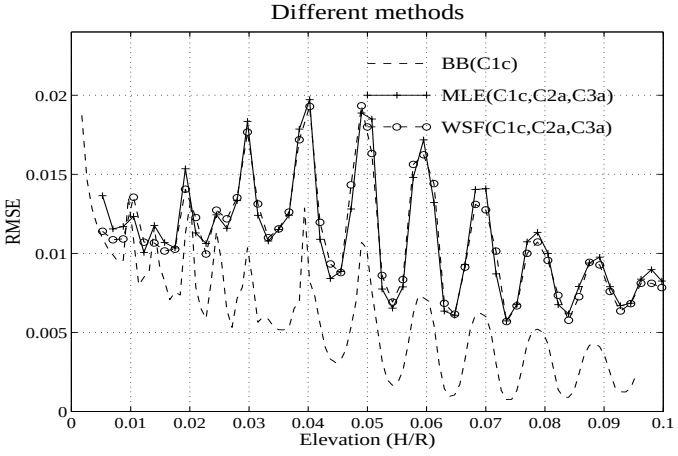

Figure 21. RMSE for MLE, WSF and MODE. Assumption C1c, C2a, C3a, SNR = 0dB.

9 Waveform and noise assumptions

In this section we study how the signal wave form parameterization (C2a or C2b) and the noise assumption (C3a or C3b) affect the estimator accuracy. If the signal wave form, $x(t)$, is parameterized by a single unknown doppler frequency (C2a) we might expect the estimates to be significantly improved compared to case C2a, since there are less parameters to estimate. But since we have proven that the CRB of C2b equals the CRB for C2a, Section 5, and that there are methods that achieve or approach the bounds for case C2a, Section 7, no significant improvement should be possible.

Moreover, in Section 5 we also showed that if the true noise covariance matrix is $\mathbf{Q} = \sigma^2\mathbf{I}$, the more general assumption C3b gives the same bound as the more precise assumption C3a. But if the noise is truly colored the C3b assumption is likely to give more accurate results than C3a, since C3a is incorrect.

In simulations the default radar setup is used and each RMSE value in the plots is estimated from 200 Monte Carlo realizations. In Figure 22, 23 and 24 the observations are generated to simulate a target at elevation $H/R = 0.021$ with $x(t) = 1$, $t = 1, \ldots, N$ and spatially white noise of varying SNR. The doppler frequency is hence $\nu = 0$. In Figure 25, 26 and 27 the target elevation H/R varies while SNR = 0dB is fixed. For assumption C1a there are two elevations to be estimated and the RMSE is computed according to Equation (3.20).

The performance of methods with different noise and signal wave form assumptions (C2a,C3a), (C2b,C3a), (C2b,C3b) are studied for different levels of knowledge on the geometry C1a–c. The estimators chosen are the different quick MODE implementations of Section 6 for multipath assumptions C1a and C1b. MODE for case (C2a,C3a) is described in Section 6.1, case (C2b,C3a) is described in Section 6.2 and case (C2b,C3b) is described in Section 6.3. But for C1c, Figure 24, MLE/MODE is used since the quick MODE algorithm cannot be implemented to find the elevations. Instead the reiterations are implemented with MLE for $\hat{\theta}$ given $\hat{\nu}$ and MODE for $\hat{\nu}$ given $\hat{\theta}$.

In Figure 22, 23 and 24 we see that there is no difference between the results at high SNR, all methods coincide and achieve the bound. By exploiting the doppler frequency the low SNR robustness is slightly improved but the noise assumption is of little importance (if the true noise covariance matrix $\mathbf{Q} = \sigma^2\mathbf{I}$ as in these simulations). But if the noise truly is colored the C3b assumption is likely to give more accurate results than C3a. Figure 25, 26 and 27 verify that these estimators give similar performance also for other elevations.

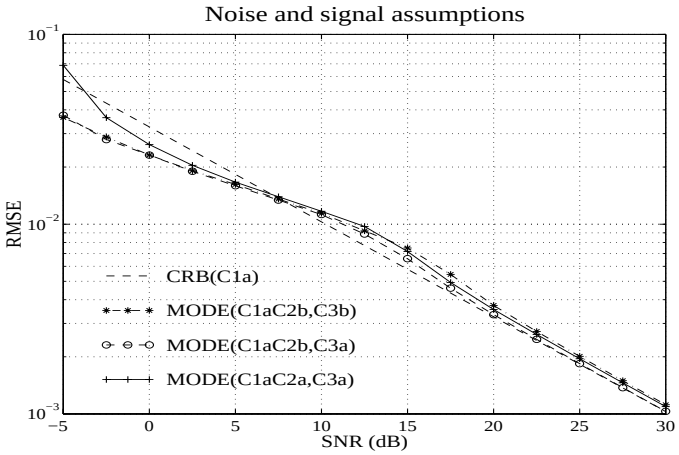

Figure 22. RMSE of algorithms at different waveform and noise assumptions. Geometry assumption C1a, $H/R = 0.021$.

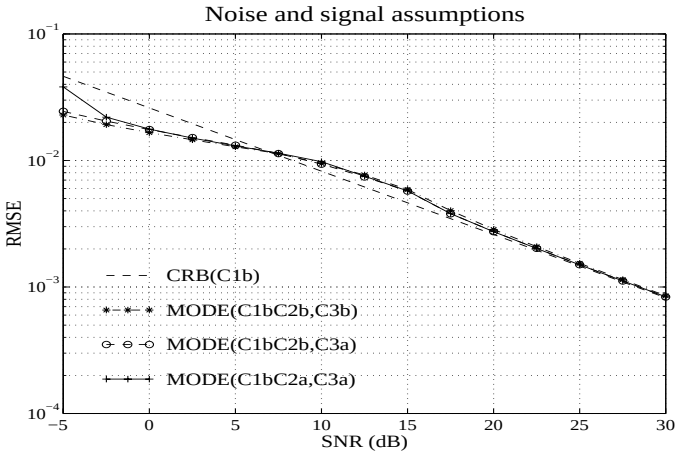

Figure 23. RMSE of algorithms at different waveform and noise assumptions.Geometry assumption C1b, $H/R = 0.021$.

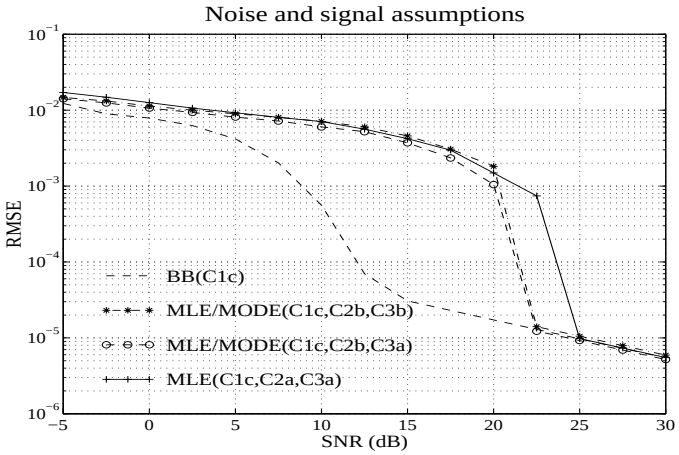

Figure 24. RMSE of algorithms at different waveform and noise assumptions. Geometry assumption C1c, $H/R = 0.021$.

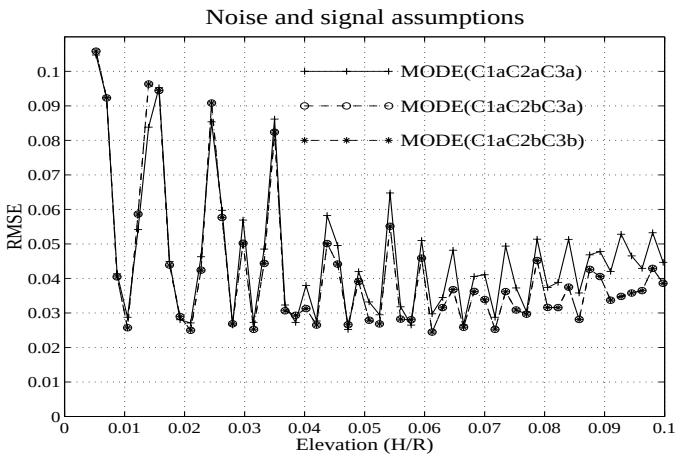

Figure 25. RMSE of algorithms at different waveform and noise assumptions. Geometry assumption C1a, SNR=0dB.

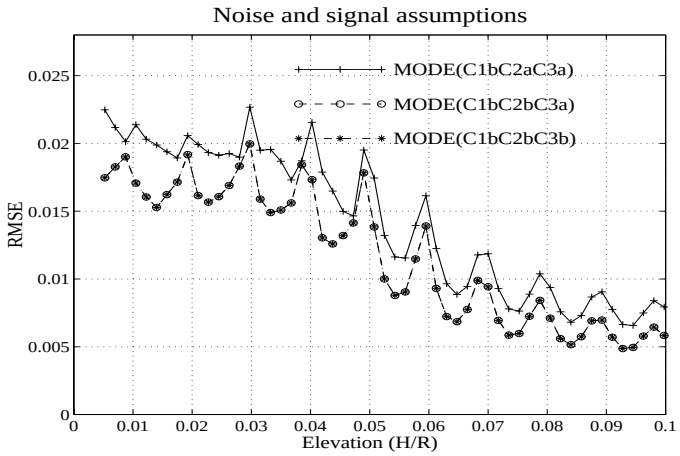

Figure 26. RMSE of algorithms at different waveform and noise assumptions.Geometry assumption C1b, SNR=0dB.

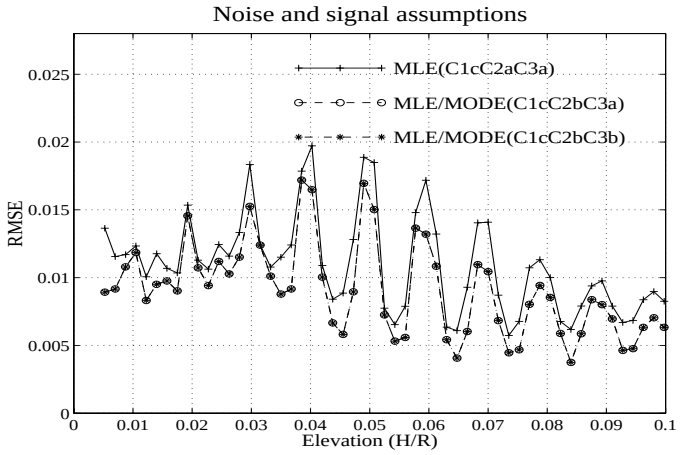

Figure 27. RMSE of algorithms at different waveform and noise assumptions. Geometry assumption C1c, SNR=0dB.

10 Array shape design

It is well known that a larger aperture results in better resolution capabilities. But it is also a well known fact that a ULA will need a sensor spacing $d \leq \lambda/2$ to ensure that there is no aliasing for $DOA \in [-\pi/2, \pi/2]$ [22]. This section will give two examples of how the Barankin bound can be used to derive an array structure which give better estimator accuracy for assumptions C1c, C2a and C3a.

An m-element ULA can uniquely resolve $m-1$ signal sources if $d \leq \lambda/2$. A potential way of solving the ambiguity problem and still use a larger aperture is to use a subarray (ULA) of m_1 elements with $d_1 \leq \lambda/2$ to avoid aliasing and the ramaining m_2 elements placed in a clever way that gives the best tradeoff between a sharp main lobe and low sidelobe levels [3]. However, this only guarantees that the array can uniquely resolve $m_1 - 1$ sources. These results are derived for single path sources, but if there is no aliasing for C1a, it means that adding the information of assumptions C1b or C1c cannot introduce aliasing. But the fine grating pattern for C1c with nearly ambiguous minima cannot be avoided.

Since the RMS error of the ML methods have been found to approach the Barankin bound, the Barankin bound is very useful in designing the optimal array. Remember that the Barankin bound can take ambiguities into account, and hence the sidelobes are included. To evaluate the ML estimators via Monte Carlo simulations for varying apertures, element configurations, elevations and SNR would be very time consuming. Instead the array structure that is expected to minimize the RMS error is found by computing the Barankin bound for a range of elevations and at the worst case SNR.

Example 1, varying aperture Let the ULA be split in two equal parts with 4 sensors each which are separated to increase the total aperture. This array configuration is denoted 4-0-4. The sensor spacing in the two uniform linear subarrays is kept at the default value $d = 0.08m$, and the wavelength is $0.1m$. As mentioned when the default radar setup was given, see Section 3, the radar we have studied has directional elements recieving signals from 60 degrees only. So there was no aliasing within this sector even though $d < \lambda/2$. The observations are generated to simulate a target at elevation $H/R = 0.021$ with $x(t) = 1$, $t = 1, \ldots, N$ and spatially white noise of varying SNR. The RMSE of the MLE corresponding to assumptions C1c, C2a and C3a for varying apertures and SNR is displayed in Figure 28. Each RMSE value in the plot is estimated from 200 Monte Carlo realizations. The Barankin bound computed for the same conditions is given in Figure 29. The two figures verifies that the estimates are improved by incresing the aperture up to a point where outliers appear due to large sidelobes. Even though the Barankin bound does not coincide with the RMSE in the threshold region, it is close enough to take account of the sidelobe ambiguities and it is proved to be a good tool when designing the array.

Example 2, varying element spacing Previous example showed that a large aperture can be beneficial if we manage to avoide sidelobes. Let us use the aperture $L = 1.5m$, which had signs of sidelobe ambiguities, and modify the element spacings of the second linear uniform subarray. In Figure 30 the Barankin bound is plotted for different element spacings of the second linear uniform subarray. We see that a large spread of the second ULA elements is will lower the bound, except when the SNR is very low. Traditionally one looks only at the CRB which is minimized for $d_2 \to 0$, which is natural since only the mainlobe is considered when deriving the CRB.

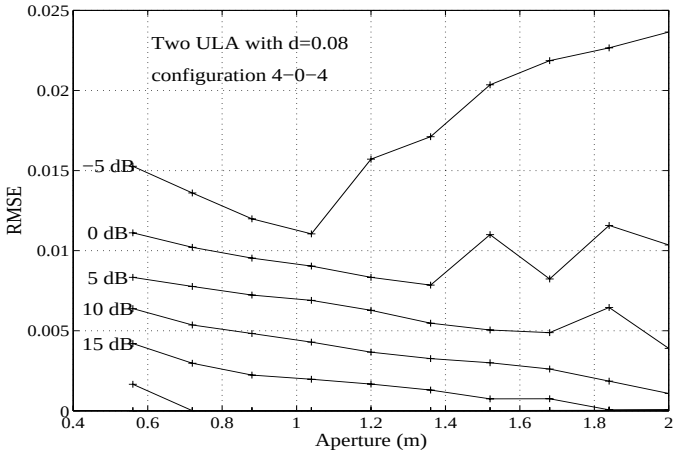

Figure 28. Example 1, RMSE of MLE(C1c,C2b,C3a) for varying aperture and SNR.

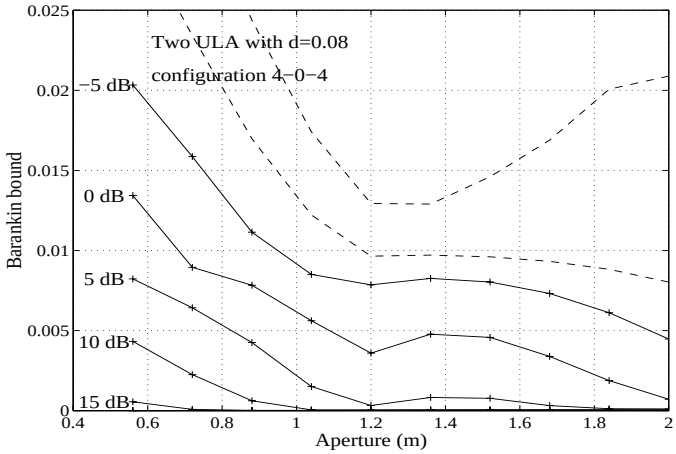

Figure 29. Example 1, BB(C1c) for varying aperture and SNR.

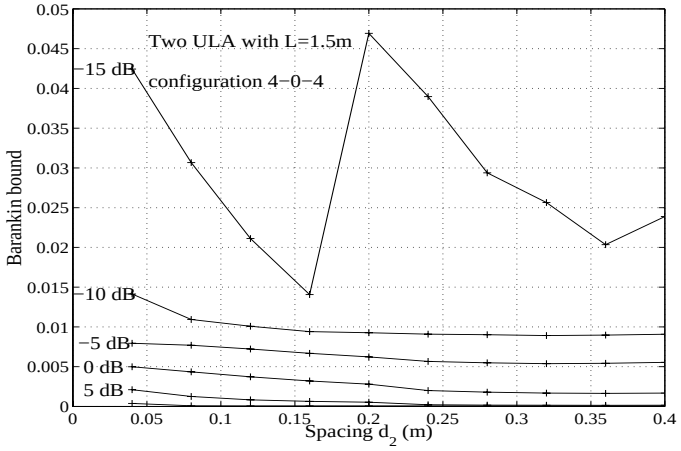

Figure 30. Example 2, BB(C1c) for varying element spacing and SNR.

11 Sensors and snapshots

Intuitively we know that more sensors, more snapshots and better SNR will improve the estimates. The question is how much improvement to expect, since the radar design often is a tradeoff between performance and cost or computation load. It is then of vital importance to relate the expected RMSE to the design parameters. In [15] it was shown that the asymptotic CRB is proportional to $1/(Nm^3SNR)$. However, in the low angle case with two closely spaced incoming rays, the asymptotic result is not applicable unless the array is very large. More reasonably sized arrays are of greater interest for most real applications, and hence we must compute the actual bounds instead of using the asymptotic results.

In Figure 31–33 we vary the number of sensors m and plot the RMSE of the MLE for assumptions C1a–c and C2b and C2a and corresponding bounds. The snapshot SNR and N are kept fixed and the 200 Monte Carlo realizations are used to compute every RMSE value. It turns out that for $m < 60$ assumptions C1a and C1b give a decay of the CRB approximately proportional to $1/(Nm^kSNR)$, where $4 < k < 5$. This means that the MSE achieved when doubling the number of sensors could also approximately be obtained by improving the SNR by 12-15dB or using 16-32 times as many snapshots.

The RMSE is slightly smaller than the bounds for C1a and C1b. It has already been explained that for very low elevations the nonsymmetric distribution of estimates around the true value causes this effect, see Section 7.

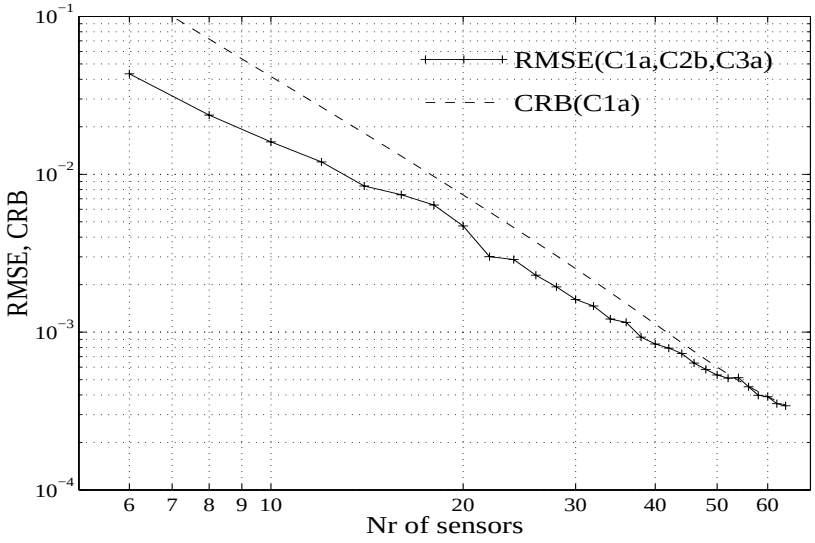

Figure 31. Varying number of sensors, C1a: RMSE of MODE(C1a,C2b,C3a) and corresponding CRB. SNR = 0dB, $H_0/R = 0.021$ and $N = 64$.

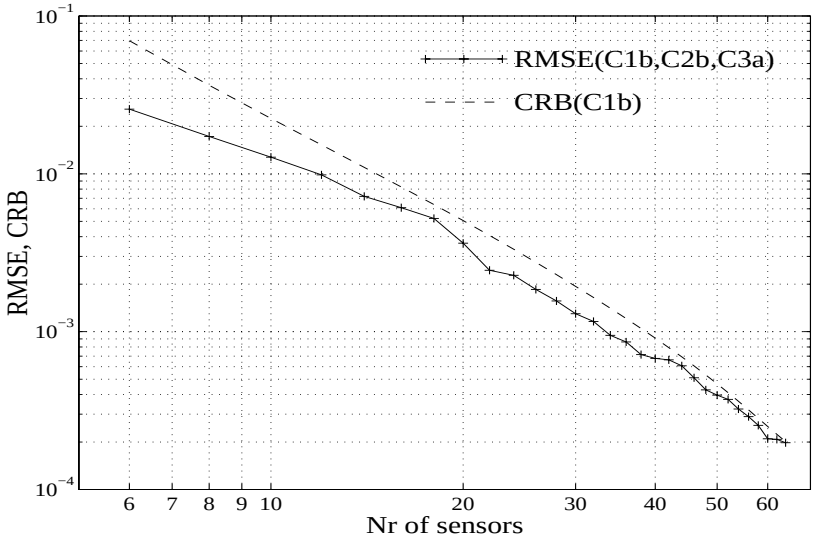

Figure 32. Varying number of sensors, C1b: RMSE of MODE(C1b,C2b,C3a) and corresponding CRB. SNR = 0dB, $H_0/R = 0.021$ and $N = 64$.

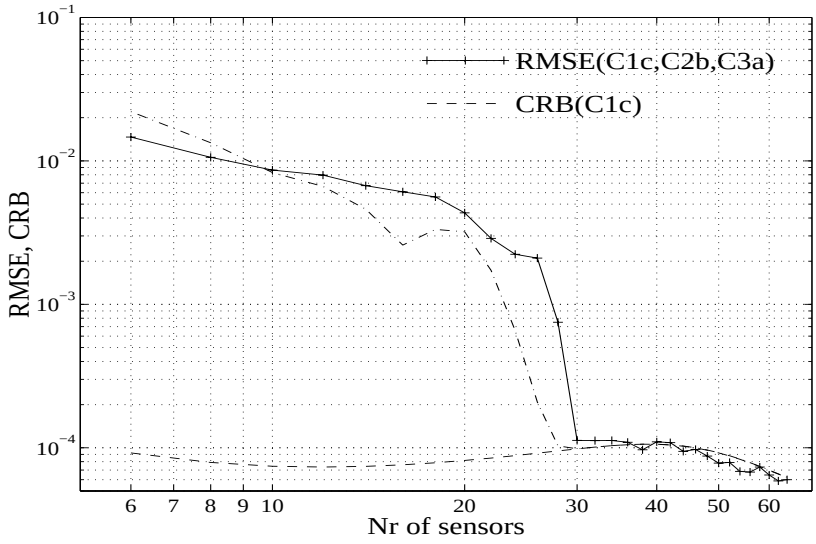

Figure 33. Varying number of sensors, C1c: RMSE of MLE(C1c,C2b,C3a) and corresponding CRB and BB. SNR = 0dB, $H_0/R = 0.021$ and $N = 64$.

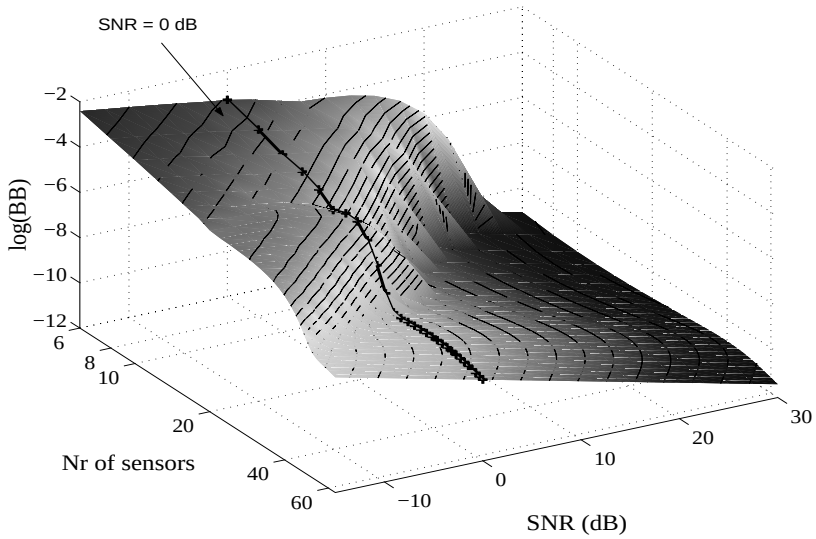

Figure 34. Sensors and SNR, C1c: RMSE of MLE(C1c,C2b,C3a) and corresponding CRB and BB. $H_0/R = 0.021$ and $N = 64$.

12 Multiple frequencies

Let us assume that the frequencies used in the N different snapshots are not all equal. So far, only the single frequency case has been considered. The signal model is then rewritten as

$$\mathbf{y}_q(t) = \mathbf{A}_q(\theta)\mathbf{b}_q x_q(t) + \varepsilon_q(t) \qquad t = 1, \ldots, N_q \quad q = 1, \ldots, K \tag{12.1}$$

where the subindex q corresponds to the K different frequencies $\{f_q\}_{q=1}^K$ or wavelengths $\{\lambda_q\}_{q=1}^K$. The total number of snapshots $N = N_1 + N_2 + \ldots + N_K$ is a fixed number, so that the introduction of more frequencies is not confused with using more snapshots.

Multiple frequencies have been used to suppress the closely spaced ambiguities which occur when the multipath model C1c is assumed [6], i.e when the reflection coefficient is included in the model. These multiple ambiguities correspond to elevations with the same phase on the relative reflection coefficient. According to Equation (3.5) the total reflection coefficient is given by

$$G = , \ (\lambda, H, h, R)e^{-j\frac{2\pi}{\lambda}\frac{2Hh}{R}} \tag{12.2}$$

The phase of , is varying slowly with frequency ($f \sim 1/\lambda$) and elevation H/R, compared to the quick phase shift due to changes in $H/R\lambda$. The phase is hence approximately given by

$$phase(G(\theta, \lambda)) = phase(G(\theta_0, \lambda_0)) + 4\pi h(\frac{\theta_0}{\lambda_0} - \frac{\theta}{\lambda}) \tag{12.3}$$

The multiple ambiguities corresponding to assumption C1c are hence approximately located at elevations θ_k, where

$$\theta_k = k\frac{\lambda}{2h}, \qquad k = 0, \pm 1, \ldots \tag{12.4}$$

The MLE criterion, derived in Appendix A.4 will have a global minimum at the true elevation regardless of the frequency, but the side ambiguities are located at different elevations for different frequencies. Since the local minima are very sharp, even small frequency shifts, a few percent, can suppress the sidelobes significantly. This is the reason why the use of multiple frequencies is so efficient in suppressing the sidelobes in case of C1c.

Is the use of multiple frequencies equally successful when we have to rely on a less refined signal model, C1a or C1b? The answer is no. The limiting problem for estimates based on C1a or C1b is not caused by ambiguous sidelobes, but rather by the sharpness of the mainlobe. Remember that the CRB describes the sharpness of the mainlobe. The multiple frequency CRB is given by

$$CRB(\theta) = \big(\sum_{q=1}^{K} CRB_q(\theta)^{-1}\big)^{-1} > \frac{1}{K} min(CRB_q(\theta)) \tag{12.5}$$

This shows that for one specific elevation the best frequency combination, giving the smallest CRB value, is to use a single frequency f_q, corresponding to the smallest CRB_q. The big drawback, that makes it impossible to use this result, is that the optimal frequency is

different for different elevations and the elevation is unknown in advance. In Figure 35 the single-frequency CRB of assumption C1b is plotted for different frequencies, at one specific elevation.

From earlier discussion we know that the single frequency CRB fluctuates with changing elevations as the relative phase between direct and specular paths is changing, see Figure 2. By using multiple frequencies these fluctuations in performance can be smoothed out to some extent, so that the worst case CRB is improved at the price of a larger bound for some other elevations. An example with two frequencies is given in Figure 36.

If a single frequency is used, aliasing will appear when the sensor spacing is larger than $\lambda/2$. But remember that a larger frequency also means better resolution, since increasing the frequency (decreasing the wavelength) will mean that the array sensor spacing is larger in units of wavelength λ. By studying the CRB of two different elevations, Figure 35, we see a clear trend of a decreasing bound as the frequency is increased. When designing the array geometry we pointed out that as long as some sensors had the sensor spacing less than $\lambda/2$ the remaining sensors could be placed at arbitrary locations to improve the resolution. The equivalence in terms frequencies is to have some snapshots at a wavelength $\lambda > 2d$ to avoid ambiguities, and the remaining snapshots at some higher frequency/frequencies to get improved resolution. However, the CRB is a local measure only, and cannot show the aliasing problem.

To summarize, using multiple frequencies is a good tool to suppress the multiple ambiguities occurring when the relative reflection coefficient is modeled, C1c. Multiple frequencies can smooth out the otherwise fluctuating performance in case of the less refined multipath assumptions C1a and C1b. The resolution can be improved by using larger frequencies for some of the snapshots, but care must be taken to avoid ambiguities.

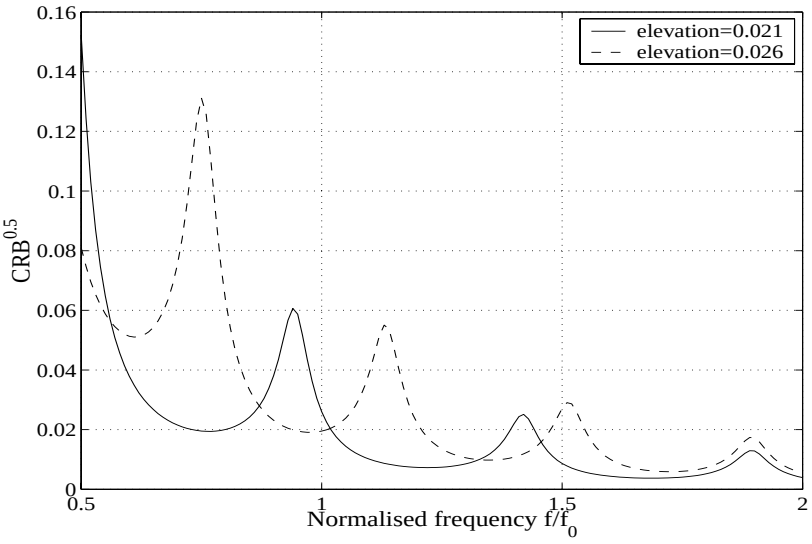

Figure 35. CRB for assumption C1b at varying frequencies. The number of snapshots N=64, SNR=0 dB. The CRB is calculated for two example elevations showing similar but perturbed fluctuating patterns.

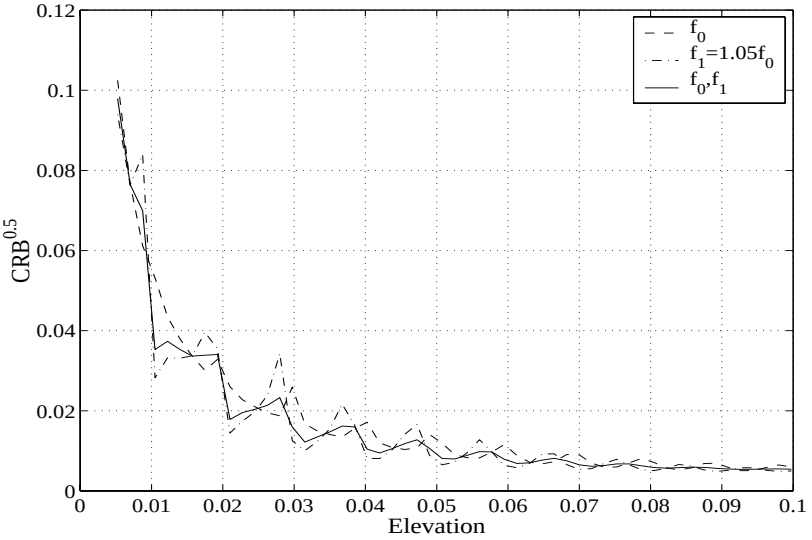

Figure 36. The plot compare single frequency CRB $f_0 = 3GHz$ and $f_1 = 1.05f_0$ with the two frequency CRB with equally many snapshots at frequency f_0 and f_1. The total number of snapshots N=64, SNR=0 dB. Assuming C1b, the CRB:s are calculated for varying elevations.

13 Conclusions

The crucial factor that determines the elevation estimator accuracy is the multipath modeling. To be able to separate the direct signal from the specular reflection, C1a, a high SNR is required. A significant improvement is seen if the geometry of reflections is included in the model, C1b. If the multipath model is refined further as in C1c where the reflection coefficient is known as a function of grazing angle, the estimates will be distributed over a few closely spaced and nearly discrete levels. At some threshold SNR the estimators based on C1c will find only the discrete level corresponding to the true maximum and an extremely good estimate is found.

Different signal parameterizations, C2a or C2b, give the same performance at high SNR, but the parameterization of the signal in terms of doppler frequency, C2b, improves the accuracy at low SNR.

White noise was used in all simulations. It turned out that the methods derived under a white noise assumption (C3a) or a general unknown noise covariance (C3b) gave the same RMSE. Hence, there was no degradation in performance by using C3b. In the opposite case where the noise is spatially colored the assumption C3a would be incorrect and the corresponding estimators cannot be expected to be efficient. Note however that such estimators are not necessarily biased. As a result assumption C3b may be preferred since it can handle colored noise better and white noise equally well, but it requires that C2b is also valid.

Simulations show that the all the derived MLE's are statistically efficient and achieve the theoretical lower bound on error variance for large SNR. Another interesting result was that the performance of all suggested alternative methods for a given set of assumptions practically coincided with the performance of the corresponding MLE. The choice of a preferred method will depend on low SNR performance and computational cost. A good choice can be the MODE algorithm based on assumptions C1b in combination with (C2a,C3a) or (C2b,C3b) as it is a quick and fairly accurate estimator. For C1c estimators require a time consuming search and at low SNR the RMSE is not much better than for C1b. But on the other hand they can provide estimates distributed over a few nearly discrete levels in elevation and at high SNR excellent estimates are found when the ambiguous levels do no longer interfere.

The CRB($\boldsymbol{\theta}$) shows an improvement when exploiting multipath information C1a-c. But the CRB corresponding to C1c is too optimistic in the threshold region, since it is a local bound on the error variance and cannot take ambiguous minima into account. The tighter Barankin bound was derived and found to predict the threshold behavior at low SNR. We showed that the CRB for all combinations of assumptions can be expressed in one equation, from which it is easily seen that CRB for elevation is the same for the different signal parameterizations and independently of the assumption on noise covariance. The Barankin bound was found to be useful to design an improved array structure that can give better performance than the standard ULA structure.

A Derivation of MLE

A.1 MLE for unknown deterministic source signal

The negative log-likelihood function for N samples of data from the parametrized model, Equation (3.24)-(3.26), is within a multiplicative constant

$$V(\boldsymbol{\eta}) = \log(m\sigma) + \text{Tr}\left\{\frac{1}{\sigma}\mathbf{C}(\boldsymbol{\eta})\right\} \tag{A.1}$$

where

$$\mathbf{C}(\boldsymbol{\eta}) = \frac{1}{N}\sum_{t=1}^{N}\left(\mathbf{y}(t) - \mathbf{A}\mathbf{b}x(t)\right)\left(\mathbf{y}(t) - \mathbf{A}\mathbf{b}x(t)\right)^* \tag{A.2}$$

$\boldsymbol{\eta}$ denotes the vector of unknown parameters, i.e $\boldsymbol{\theta}$ and real and imaginary parts of $\mathbf{b}$ and $\{x(t)\}$. The minimization of (A.1) with respect to σ^2, $\{x(t)\}$ and $\mathbf{b}$ can be evaluated explicitly. Standard matrix algebra will give the derivative of the criterion with respect to σ as

$$\frac{\partial V(\boldsymbol{\eta})}{\partial \sigma} = \frac{1}{\sigma} - \frac{1}{\sigma^2}\text{Tr}\,\mathbf{C}(\boldsymbol{\eta}) \tag{A.3}$$

and hence the ML estimate of σ is given by

$$\hat{\sigma} = \text{Tr}\,\mathbf{C}(\boldsymbol{\eta}) \tag{A.4}$$

Inserting (A.4) into (A.1) yields the concentrated negative log-likeliood function

$$V(\boldsymbol{\eta}|\sigma = \hat{\sigma}) = 1 + \log(m) + \log(\text{Tr}\,\mathbf{C}(\boldsymbol{\eta})) \tag{A.5}$$

Since minimizing $V(\boldsymbol{\eta}|\sigma = \hat{\sigma})$ is the same as minimizing Tr $\mathbf{C}(\boldsymbol{\eta})$, we rewrite the trace expression in order to find the minimizing sequence $\{x(t)\}$.

$$\begin{aligned}
\text{Tr}\,\mathbf{C}(\boldsymbol{\eta}) &= \frac{1}{N}\sum_{t=1}^{N}\big(\mathbf{y}(t) - \mathbf{A}\mathbf{b}x(t)\big)^*\big(\mathbf{y}(t) - \mathbf{A}\mathbf{b}x(t)\big) = \\
&= \frac{1}{N}\sum_{t=1}^{N}\Big(\mathbf{y}(t)^*\mathbf{y}(t) - 2\text{Re}\{\mathbf{y}(t)^*\mathbf{A}\mathbf{b}x(t)\} + \mathbf{b}^*\mathbf{A}^*\mathbf{A}\mathbf{b}x(t)x(t)^*\Big) \\
&= \frac{1}{N}\sum_{t=1}^{N}\left(\left\|x(t) - \frac{\mathbf{b}^*\mathbf{A}^*\mathbf{y}(t)}{\|\mathbf{A}\mathbf{b}\|^2}\right\|^2 \|\mathbf{A}\mathbf{b}\|^2 + \mathbf{y}(t)^*\mathbf{y}(t) - \frac{\|\mathbf{y}^*(t)\mathbf{A}\mathbf{b}\|^2}{\|\mathbf{A}\mathbf{b}\|^2}\right)
\end{aligned} \tag{A.6}$$

Since only the first term of (A.6) depends on $\{x(t)\}$ it is obvious that the minimizing $\{x(t)\}$ is given by

$$\hat{x}(t) = \frac{\mathbf{b}^*\mathbf{A}^*\mathbf{y}(t)}{\|\mathbf{A}\mathbf{b}\|^2} \qquad t = 1, \ldots, N \tag{A.7}$$

Substituting (A.7) into (A.6) the trace expression can be rewritten as

$$\text{Tr}\,\mathbf{C}(\boldsymbol{\theta}, \mathbf{b}) = \frac{1}{N}\sum_{t=1}^{N}\left(\mathbf{y}^*(t)\mathbf{y}(t) - \frac{\|\mathbf{y}^*(t)\mathbf{A}\mathbf{b}\|^2}{\|\mathbf{A}\mathbf{b}\|^2}\right) \tag{A.8}$$

Use the trace relation Tr $(\mathbf{AB})$ = Tr $(\mathbf{BA})$ and normalise the equations using $\mathbf{b} = (\mathbf{A}^*\mathbf{A})^{-1/2}\boldsymbol{\beta}$.

$$\text{Tr}\,\mathbf{C}(\boldsymbol{\theta},\boldsymbol{\beta}) = \text{Tr}\,\hat{\mathbf{R}}_{yy} - \boldsymbol{\beta}^*(\mathbf{A}^*\mathbf{A})^{-1/2}\mathbf{A}^*\hat{\mathbf{R}}_{yy}\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1/2}\boldsymbol{\beta} \tag{A.9}$$

The minimizing $\boldsymbol{\beta}$ is given by the normalised eigenvector corresponding to the largest eigenvalue, $\lambda_1(\boldsymbol{\theta})$ of $(\mathbf{A}^*\mathbf{A})^{-1/2}\mathbf{A}^*\hat{\mathbf{R}}_{yy}\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1/2}$ It follows that the minimum of (A.9) can be expressed as

$$\text{Tr}\,\mathbf{C}(\boldsymbol{\theta}) = \text{Tr}\,\hat{\mathbf{R}}_{yy} - \lambda_1(\boldsymbol{\theta}) \tag{A.10}$$

where

$$\lambda_1(\boldsymbol{\theta}) = \lambda_{max}[(\mathbf{A}^*\mathbf{A})^{-1/2}\mathbf{A}^*\hat{\mathbf{R}}_{yy}\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1/2}] \tag{A.11}$$

and the ML estimator is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\left(\hat{\mathbf{R}}_{yy} - \lambda_1(\boldsymbol{\theta})\right) \tag{A.12}$$

A.2 MLE for parameterized waveform

If the white noise assumption is still used the minimization with respect to σ will be the same as in previous section, and hence we can start the derivations from equations (A.5),(A.6). If $\{x(t)\}$ is parameterized by $\boldsymbol{\alpha}$ we can rewrite (A.6) and get the ML estimate $\hat{\mathbf{b}}(\boldsymbol{\theta},\boldsymbol{\alpha})$. This time we choose to normalize the equations using $\hat{R}_{xx} = \frac{1}{N}\sum_{t=1}^{N} x(t)x^*(t) = 1$. Let $\boldsymbol{\eta}$ denote the vector of unknown parameters, i.e $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$ and real and imaginary parts of $\mathbf{b}$. Equation (A.6) is now rewritten

$$\begin{aligned}\text{Tr}\,\mathbf{C}(\boldsymbol{\eta}) &= \hat{\mathbf{R}}_{yy} - \hat{\mathbf{R}}_{yx}^*\mathbf{A}\mathbf{b} - \mathbf{b}*\mathbf{A}^*\hat{\mathbf{R}}_{yx} + \mathbf{b}^*\mathbf{A}^*\mathbf{A}\mathbf{b}\hat{R}_{xx} \\ &= \left(\mathbf{b} - (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\hat{\mathbf{R}}_{yx}\right)^*\mathbf{A}^*\mathbf{A}\left(\mathbf{b} - (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\hat{\mathbf{R}}_{yx}\right) + \hat{\mathbf{R}}_{yy} - \hat{\mathbf{R}}_{yx}^*\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\hat{\mathbf{R}}_{yx}\end{aligned} \tag{A.13}$$

The minimizing $\mathbf{b}$ is consequently given by

$$\mathbf{b} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\hat{\mathbf{R}}_{yx} \tag{A.14}$$

Substituting (A.14) into (A.13) the trace can be rewritten as

$$\text{Tr}\,\mathbf{C}(\boldsymbol{\theta},\boldsymbol{\alpha}) = \text{Tr}\,\hat{\mathbf{R}}_{yy} - \hat{\mathbf{R}}_{yx}^*\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\hat{\mathbf{R}}_{yx} \tag{A.15}$$

and the MLE is given by

$$\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\theta},\boldsymbol{\alpha}}\hat{\mathbf{R}}_{yx}^*\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\hat{\mathbf{R}}_{yx} \tag{A.16}$$

A.3 MLE for spatially correlated noise

The negative log-likelihood function for N samples of data from the parametrized model is, within a multiplicative constant

$$V(\boldsymbol{\eta}) = \log|\mathbf{Q}| + \mathrm{Tr}\left\{\mathbf{Q}^{-1}\mathbf{C}(\boldsymbol{\eta})\right\} \tag{A.17}$$

where

$$\mathbf{C}(\boldsymbol{\eta}) = \frac{1}{N}\sum_{t=1}^{N}\left(\mathbf{y}(t) - \mathbf{A}\mathbf{b}x(t)\right)\left(\mathbf{y}(t) - \mathbf{A}\mathbf{b}x(t)\right)^{*} \tag{A.18}$$

The minimization of (A.17) with respect to $\mathbf{Q}$ is easily seen to give

$$\hat{\mathbf{Q}} = \mathbf{C}(\boldsymbol{\eta}) \tag{A.19}$$

Inserting (A.19) into (A.17) yields the concentrated negative log-likeliood function

$$V(\boldsymbol{\eta}) = m + \log|\mathbf{C}(\boldsymbol{\eta})| \tag{A.20}$$

To avoid confusion, remember that $\mathbf{Q}$ denotes the true spatial noise covariance and $\hat{\mathbf{Q}}$ the corresponding ML estimate. The following definition of $\tilde{\mathbf{Q}}$ is just a definition, but it is also an unstructured estimate of $\mathbf{Q}$

$$\tilde{\mathbf{Q}} = \hat{\mathbf{R}}_{yy} - \hat{\mathbf{R}}_{yx}\hat{R}_{xx}^{-1}\hat{\mathbf{R}}_{yx}^{*} \tag{A.21}$$

Now, using the definitions of $\hat{\mathbf{R}}_{yy}$, $\hat{\mathbf{R}}_{yx}$ and that $\hat{R}_{xx} = 1$ we can rewrite the equation for $\mathbf{C}(\boldsymbol{\eta})$ as

$$\begin{aligned}\mathbf{C}(\boldsymbol{\eta}) &= \sum_{t=1}^{N}\left(\mathbf{y}(t) - \mathbf{A}\mathbf{b}x(t)\right)\left(\mathbf{y}(t) - \mathbf{A}\mathbf{b}x(t)\right)^{*} = \\ &= \hat{R}_{yy} - \hat{\mathbf{R}}_{yx}\hat{R}_{xx}^{-1}\hat{\mathbf{R}}_{yx}^{*} + (\frac{\hat{\mathbf{R}}_{yx}}{\hat{R}_{xx}} - \mathbf{A}\mathbf{b})\hat{R}_{xx}(\frac{\hat{\mathbf{R}}_{yx}}{\hat{R}_{xx}} - \mathbf{A}\mathbf{b})^{*} = \\ &= \tilde{\mathbf{Q}}(\mathbf{I} + \tilde{\mathbf{Q}}^{-1}(\hat{\mathbf{R}}_{yx} - \mathbf{A}\mathbf{b})(\hat{\mathbf{R}}_{yx} - \mathbf{A}\mathbf{b})^{*})\end{aligned} \tag{A.22}$$

where the argument $\boldsymbol{\theta}$ has been left out for notational convenience. Using the determinant rule, $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$, and the fact that $|\mathbf{A}\mathbf{B}| = |\mathbf{A}|\,|\mathbf{B}|$, the concentrated negative log-likelihood can be shown to equal

$$V(\boldsymbol{\theta}, \boldsymbol{\alpha}) = m + \log\left|\tilde{\mathbf{Q}}\right| + \log\left|1 + (\hat{\mathbf{R}}_{yx} - \mathbf{A}\mathbf{b})^{*}\tilde{\mathbf{Q}}^{-1}(\hat{\mathbf{R}}_{yx} - \mathbf{A}\mathbf{b})\right| \tag{A.23}$$

The equation (A.7) is minimized with respect to $\mathbf{b}$ if the following expression is minimized

$$\begin{aligned}\left(\hat{\mathbf{R}}_{yx} - \mathbf{A}\mathbf{b}\right)^{*}\tilde{\mathbf{Q}}^{-1}\left(\hat{\mathbf{R}}_{yx} - \mathbf{A}\mathbf{b}\right) &= \hat{\mathbf{R}}_{yx}^{*}\tilde{\mathbf{Q}}\hat{\mathbf{R}}_{yx} - \hat{\mathbf{R}}_{yx}^{*}\tilde{\mathbf{Q}}\mathbf{A}(\mathbf{A}^{*}\tilde{\mathbf{Q}}\mathbf{A})^{-1}\mathbf{A}^{*}\tilde{\mathbf{Q}}\hat{\mathbf{R}}_{yx} \\ &+ \left(\mathbf{b} - (\mathbf{A}^{*}\tilde{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^{*}\tilde{\mathbf{Q}}\hat{\mathbf{R}}_{yx}\right)^{*}\mathbf{A}^{*}\tilde{\mathbf{Q}}^{-1}\mathbf{A}\left(\mathbf{b} - (\mathbf{A}^{*}\tilde{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^{*}\tilde{\mathbf{Q}}\hat{\mathbf{R}}_{yx}\right)\end{aligned} \tag{A.24}$$

Since only the last term of (A.24) depends on $\mathbf{b}$ it is obvious that the minimizing $\mathbf{b}$ is given by

$$\hat{\mathbf{b}} = (\mathbf{A}^*\tilde{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^*\tilde{\mathbf{Q}}\hat{\mathbf{R}}_{yx} \tag{A.25}$$

Substituting (A.25) into (A.24) the concentrated negative log-likelihood function can be rewritten as

$$V(\boldsymbol{\theta},\boldsymbol{\alpha}) = m + \log\left|\tilde{\mathbf{Q}}\right| + \log\left(1 + \hat{\mathbf{R}}_{yx}^*\tilde{\mathbf{Q}}^{-1}\hat{\mathbf{R}}_{yx} - \hat{\mathbf{R}}_{yx}^*\tilde{\mathbf{Q}}^{-1}\mathbf{A}(\mathbf{A}^*\tilde{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^*\tilde{\mathbf{Q}}^{-1}\hat{\mathbf{R}}_{yx}\right) \tag{A.26}$$

Using the definition for $\tilde{\mathbf{Q}}$ in (A.21), the matrix inversion lemma $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{B}^*\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}\mathbf{A}^{-1}$ and the determinant rule, the equation (A.26) can be shown to equal

$$V(\boldsymbol{\theta},\boldsymbol{\alpha}) = m + \log\left|\hat{\mathbf{R}}_{yy}\right| + \log\left(\frac{1}{1+V_Q(\boldsymbol{\theta},\boldsymbol{\alpha})}\right) \tag{A.27}$$

where

$$V_Q(\boldsymbol{\theta},\boldsymbol{\alpha}) = \frac{\hat{\mathbf{R}}_{yx}^*\hat{\mathbf{R}}_{yy}^{-1}\mathbf{A}(\mathbf{A}^*\hat{\mathbf{R}}_{yy}^{-1}\mathbf{A})^{-1}\mathbf{A}^*\hat{\mathbf{R}}_{yy}^{-1}\hat{\mathbf{R}}_{yx}}{1 - \hat{\mathbf{R}}_{yx}^*\hat{\mathbf{R}}_{yy}^{-1}\hat{\mathbf{R}}_{yx}} \tag{A.28}$$

Minimizing (A.27) is the same as maximizing $V_Q(\boldsymbol{\theta},\boldsymbol{\alpha})$. Hence the ML estimate of $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ is given by

$$\{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}\} = \arg\max_{\boldsymbol{\theta},\boldsymbol{\alpha}} V_Q(\boldsymbol{\theta},\boldsymbol{\alpha}) \tag{A.29}$$

$$\tag{A.30}$$

Where $V_Q(\boldsymbol{\theta},\boldsymbol{\alpha})$ is given by (A.28).

A.4 MLE for multiple frequencies

$$V(\boldsymbol{\eta}) = \frac{1}{N}\sum_{q=1}^{K} N_q\big(\log|\mathbf{Q}_q| + \mathrm{Tr}\left\{\mathbf{Q}_q^{-1}\mathbf{C}_q(\boldsymbol{\eta})\right\}\big) \tag{A.31}$$

where

$$\mathbf{C}_q(\boldsymbol{\eta}) = \frac{1}{N_q}\sum_{t=1}^{N_q}\left(\mathbf{y}(t) - \mathbf{A}_q(\boldsymbol{\theta})\mathbf{b}_q x_q(t)\right)\left(\mathbf{y}(t) - \mathbf{A}_q(\boldsymbol{\theta})\mathbf{b}_q x_q(t)\right)^* \tag{A.32}$$

which means that when minimizing (A.31) with respect to σ_q (if $\mathbf{Q}_q = \sigma_q\mathbf{I}$) or $\mathbf{Q}_q$ will lead to the concentrated negative likelihood functions

$$V(\boldsymbol{\theta}|C2a, C3a) = 1 + \log(m) + \frac{1}{N}\sum_{q=1}^{K} N_q \log(\mathrm{Tr}\,\hat{\mathbf{R}}_{yy(q)} - \lambda_q(\boldsymbol{\theta})) \tag{A.33}$$

corresponding to (A.5),(A.11), and

$$V(\boldsymbol{\theta}, \boldsymbol{\alpha}|C2b, C3a) = 1 + \log(m) + \frac{1}{N}\sum_{q=1}^{K} N_q \log(\mathrm{Tr}\,\hat{\mathbf{R}}_{yy(q)} - \hat{\mathbf{R}}^*_{yx(q)}\mathbf{A}_q(\mathbf{A}_q^*\mathbf{A}_q)^{-1}\mathbf{A}_q^*\hat{\mathbf{R}}_{yx(q)}) \tag{A.34}$$

corresponding to (A.5),(A.15), and

$$V(\boldsymbol{\theta}, \boldsymbol{\alpha}|C2b, C3a) = m + \frac{1}{N}\sum_{q=1}^{K} N_q \log|\hat{\mathbf{R}}_{yy(q)}| + \frac{1}{N}\sum_{q=1}^{K} N_q \log(\mathrm{Tr}\,\hat{\mathbf{R}}_{yy(q)} - \lambda_q(\boldsymbol{\theta})) \tag{A.35}$$

corresponding to (A.27),(A.28).

B Cramér-Rao Bound

The CRB has been derived for many separate cases which are closely related to one or more of the cases discussed in this work. But here we derive a general CRB that covers all in one compact expression. However, to do that we need to derive the $CRB(\boldsymbol{\theta})$ twice, once to give the $CRB(\boldsymbol{\theta})$ for case (C1a–c, C2a, C3a–b) and once to give the $CRB(\boldsymbol{\theta})$ for case (C1a–c, C2b, C3a–b). It will turn out that the expressions are identical under mild conditions on $x(t)$.

The observations $\mathbf{y}(t)$, $t = 1, \ldots, N$ are independent circular Gaussian random variables with covariance $\mathbf{Q}$ and mean $\boldsymbol{\mu}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{b}x(t)$. Remember that $x(t)$ is unknown for assumption C2a, and $x(t) = e^{j\nu t}$ for C2b. Hence, the unknown signal parameters are

$$\boldsymbol{\eta} = \begin{bmatrix} \mathrm{Re}\{x(t)\}_{t=2}^{N} & \mathrm{Im}\{x(t)\}_{t=2}^{N} & \mathrm{Re}\{\mathbf{b}\}^T & \mathrm{Im}\{\mathbf{b}\}^T & \boldsymbol{\theta}^T \end{bmatrix} \tag{B.1}$$

for (C1a–c, C2a, C3a–b) and

$$\boldsymbol{\eta} = \begin{bmatrix} \mathrm{Re}\{\mathbf{b}\}^T & \mathrm{Im}\{\mathbf{b}\}^T & \nu & \boldsymbol{\theta}^T \end{bmatrix} \tag{B.2}$$

for (C1a–c, C2b, C3a–b).

Since $\mathbf{Q}$ does not depend on the signal parameters we can determine the CRB for the signal parameters by evaluating the Fisher information matrix [4, 23]

$$\mathbf{FIM}_{k,l}(\boldsymbol{\eta}) = 2\mathrm{Re}\sum_{t=1}^{N}\left\{ \left(\frac{\partial \boldsymbol{\mu}(t)}{\partial \boldsymbol{\eta}_k}\right)^* \mathbf{Q}^{-1} \left(\frac{\partial \boldsymbol{\mu}(t)}{\partial \boldsymbol{\eta}_l}\right) \right\} \tag{B.3}$$

The partial derivatives needed to compute (B.3) are given by

$$\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\{b_k\}} = \mathbf{A}_k(\boldsymbol{\theta})x(t) \tag{B.4}$$

$$\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\{b_k\}} = j\mathbf{A}_k(\boldsymbol{\theta})x(t) \tag{B.5}$$

$$\frac{\partial \boldsymbol{\mu}(t)}{\partial \theta_k} = \mathbf{d}_k x(t) \tag{B.6}$$

$$C2a: \frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\{x(k)\}} = \mathbf{A}(\boldsymbol{\theta})\mathbf{b}\delta_{tk} \tag{B.7}$$

$$C2a: \frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\{x(k)\}} = j\mathbf{A}(\boldsymbol{\theta})\mathbf{b}\delta_{tk} \tag{B.8}$$

$$C2b: \frac{\partial \boldsymbol{\mu}(t)}{\partial \nu} = jt\mathbf{A}(\boldsymbol{\theta})\mathbf{b}x(t) \tag{B.9}$$

$\mathbf{D}$ was defined in Equation (3.13) and contains the derivatives with respect to $\boldsymbol{\theta}$. To simplify the derivations the following notation and definitions will be used

$$\mathbf{H} = \mathbf{A}^*\mathbf{Q}^{-1}\mathbf{A} \tag{B.10}$$

$$\mathbf{L} = \mathbf{D}^*\mathbf{Q}^{-1}\mathbf{D} \tag{B.11}$$

$$\mathbf{G} = \mathbf{A}^*\mathbf{Q}^{-1}\mathbf{D} \tag{B.12}$$

where

$$\mathbf{\Omega} = 2\chi_{1,N}\left(\mathbf{H} - \frac{\chi_{2,N}}{\chi_{1,N}}\frac{\mathbf{Hbb}^*\mathbf{H}}{\mathbf{b}^*\mathbf{Hb}}\right) \tag{B.13}$$

$$\mathbf{\Delta} = 2\chi_{1,N}\left(\mathbf{G} - \frac{\chi_{2,N}}{\chi_{1,N}}\frac{\mathbf{Hbb}^*\mathbf{G}}{\mathbf{b}^*\mathbf{Hb}}\right) \tag{B.14}$$

$$\mathbf{,} = 2\chi_{1,N}\left(\mathbf{L} - \frac{\chi_{2,N}}{\chi_{1,N}}\frac{\mathbf{G}^*\mathbf{bb}^*\mathbf{G}}{\mathbf{b}^*\mathbf{Hb}}\right) \tag{B.15}$$

$$\beta_N = \frac{1}{N}\sum_{t=1}^{N} t = \frac{N+1}{2} \tag{B.16}$$

$$\gamma_N = \frac{1}{N}\sum_{t=1}^{N} t^2 = \frac{(N+1)(N+1/2)}{3} \tag{B.17}$$

$$\chi_{1,N} = \sum_{t=1}^{N} x(t)^* x(t) \tag{B.18}$$

$$\chi_{2,N} = \sum_{t=2}^{N} x(t)^* x(t) \tag{B.19}$$

CRB for assumptions (C1a–c, C2a, C3a–b): To derive the $CRB(\boldsymbol{\theta})$ for (C1a–c, C2a, C3a–b) we substitute the expressions (B.4–B.8) into (B.3), whith the unknown parameters

$$\boldsymbol{\eta} = \left[\overline{x(2)} \quad \widetilde{x(2)} \quad \ldots \quad \overline{x(N)} \quad \widetilde{x(N)} \quad \vdots \quad \overline{\mathbf{b}}^T \quad \widetilde{\mathbf{b}}^T \quad \boldsymbol{\theta}^T\right] \tag{B.20}$$

giving

$$CRB(\boldsymbol{\eta}) = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{12}^T & \mathbf{F}_{22} \end{bmatrix}^{-1} \tag{B.21}$$

where

$$\mathbf{F}_{11} = 2\begin{bmatrix} \overline{\mathbf{b}^*\mathbf{Hb}} & -\widetilde{\mathbf{b}^*\mathbf{Hb}} & & 0 \\ \widetilde{\mathbf{b}^*\mathbf{Hb}} & \overline{\mathbf{b}^*\mathbf{Hb}} & & \\ & & \ddots & \\ & & & \overline{\mathbf{b}^*\mathbf{Hb}} \quad -\widetilde{\mathbf{b}^*\mathbf{Hb}} \\ & 0 & & \widetilde{\mathbf{b}^*\mathbf{Hb}} \quad \overline{\mathbf{b}^*\mathbf{Hb}} \end{bmatrix} \tag{B.22}$$

$$\mathbf{F}_{12} = \begin{bmatrix} \overline{\mathbf{b}^*\mathbf{H}x(2)} & -\widetilde{\mathbf{b}^*\mathbf{H}x(2)} & \overline{\mathbf{b}^*\mathbf{G}x(2)} \\ \widetilde{\mathbf{b}^*\mathbf{H}x(2)} & \overline{\mathbf{b}^*\mathbf{H}x(2)} & \widetilde{\mathbf{b}^*\mathbf{G}x(2)} \\ & \vdots & \\ \overline{\mathbf{b}^*\mathbf{H}x(N)} & -\widetilde{\mathbf{b}^*\mathbf{H}x(N)} & \overline{\mathbf{b}^*\mathbf{G}x(N)} \\ \widetilde{\mathbf{b}^*\mathbf{H}x(N)} & \overline{\mathbf{b}^*\mathbf{H}x(N)} & \widetilde{\mathbf{b}^*\mathbf{G}x(N)} \end{bmatrix} \tag{B.23}$$

$$\mathbf{F}_{22} = 2\chi_{1,N}\begin{bmatrix} \overline{\mathbf{H}} & -\widetilde{\mathbf{H}} & \overline{\mathbf{G}} \\ \widetilde{\mathbf{H}} & \overline{\mathbf{H}} & \widetilde{\mathbf{G}} \\ (\overline{\mathbf{G}})^T & (\widetilde{\mathbf{G}})^T & \overline{\mathbf{L}} \end{bmatrix} \tag{B.24}$$

Observe that $\overline{\mathbf{b}^*\mathbf{H}\mathbf{b}} = \mathbf{b}^*\mathbf{H}\mathbf{b}$, which is a realvalued scalar and $\widetilde{\mathbf{b}^*\mathbf{H}\mathbf{b}} = 0$. Partitioning the matrix inverse, the lower right hand block of the CRB matrix corresponding to $[\overline{\mathbf{b}}^T, \widetilde{\mathbf{b}}^T, \boldsymbol{\theta}]$ is given by

$$CRB(\overline{\mathbf{b}}^T, \widetilde{\mathbf{b}}^T, \boldsymbol{\theta})^{-1} = \mathbf{F}_{22} - \mathbf{F}_{12}^T \mathbf{F}_{11}^{-1} \mathbf{F}_{12} \tag{B.25}$$

where, using Equation (B.23) and (B.24)

$$\mathbf{F}_{12}^T \mathbf{F}_{11}^{-1} \mathbf{F}_{12} = \sum_{t=2}^{N} \frac{2}{\mathbf{b}^*\mathbf{H}\mathbf{b}} \begin{bmatrix} (\overline{\mathbf{b}^*\mathbf{H}x(t)})^T & (\widetilde{\mathbf{b}^*\mathbf{H}x(t)})^T \\ -(\widetilde{\mathbf{b}^*\mathbf{H}x(t)})^T & (\overline{\mathbf{b}^*\mathbf{H}x(t)})^T \\ (\overline{\mathbf{b}^*\mathbf{G}x(t)})^T & (\widetilde{\mathbf{b}^*\mathbf{G}x(t)})^T \end{bmatrix} \cdot \tag{B.26}$$

$$\begin{bmatrix} \overline{\mathbf{b}^*\mathbf{H}x(t)} & -\widetilde{\mathbf{b}^*\mathbf{H}x(t)} & \overline{\mathbf{b}^*\mathbf{G}x(t)} \\ \widetilde{\mathbf{b}^*\mathbf{H}x(t)} & \overline{\mathbf{b}^*\mathbf{H}x(t)} & \widetilde{\mathbf{b}^*\mathbf{G}x(t)} \end{bmatrix} =$$

$$= \frac{2\chi_{2,N}}{\mathbf{b}^*\mathbf{H}\mathbf{b}} \begin{bmatrix} \overline{\mathbf{H}\mathbf{b}\mathbf{b}^*\mathbf{H}} & -\widetilde{\mathbf{H}\mathbf{b}\mathbf{b}^*\mathbf{H}} & \overline{\mathbf{H}\mathbf{b}\mathbf{b}^*\mathbf{G}} \\ \widetilde{\mathbf{H}\mathbf{b}\mathbf{b}^*\mathbf{H}} & \overline{\mathbf{H}\mathbf{b}\mathbf{b}^*\mathbf{H}} & \widetilde{\mathbf{H}\mathbf{b}\mathbf{b}^*\mathbf{G}} \\ \overline{\mathbf{G}^*\mathbf{b}\mathbf{b}^*\mathbf{H}} & -\widetilde{\mathbf{G}^*\mathbf{b}\mathbf{b}^*\mathbf{H}} & \overline{\mathbf{G}^*\mathbf{b}\mathbf{b}^*\mathbf{G}} \end{bmatrix} \tag{B.27}$$

Equation (B.24) and (B.27) inserted in (B.25) gives

$$CRB(\overline{\mathbf{b}}^T, \widetilde{\mathbf{b}}^T, \boldsymbol{\theta})^{-1} = \begin{bmatrix} \overline{\boldsymbol{\Omega}} & -\widetilde{\boldsymbol{\Omega}} & \overline{\boldsymbol{\Delta}} \\ \widetilde{\boldsymbol{\Omega}} & \overline{\boldsymbol{\Omega}} & \widetilde{\boldsymbol{\Delta}} \\ \overline{\boldsymbol{\Delta}}^T & \widetilde{\boldsymbol{\Delta}}^T & \overline{\boldsymbol{,}} \end{bmatrix} \tag{B.28}$$

where the matrixes $\boldsymbol{\Omega}$, $\boldsymbol{\Delta}$ and $\boldsymbol{,}$ were defined in Equation (B.13)-(B.15). Partitioning the matrix inverse, Equation (B.28), the $CRB(\boldsymbol{\theta})$ is given by

$$CRB(\boldsymbol{\theta})^{-1} = \overline{\boldsymbol{,}} - \overline{\boldsymbol{\Delta}^*\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}} \tag{B.29}$$

where

$$\boldsymbol{\Omega}^{-1} = \frac{1}{2\chi_{1,N}} \left(\mathbf{H}^{-1} - \frac{\mathbf{b}\mathbf{b}^*}{(1 - \chi_{2,N}/\chi_{1,N})\mathbf{b}^*\mathbf{H}\mathbf{b}} \right) \tag{B.30}$$

$$\boldsymbol{\Delta}^*\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta} = 2\chi_{1,N}\mathbf{G}^* \left(\mathbf{H}^{-1} - \frac{\chi_{2,N}}{\chi_{1,N}} \frac{\mathbf{b}\mathbf{b}^*}{\mathbf{b}^*\mathbf{H}\mathbf{b}} \right) \mathbf{G} \tag{B.31}$$

The $CRB(\boldsymbol{\theta})$ for (C1a–c, C2a, C3a–b) is then finally given by inserting the definition of $\boldsymbol{,}$ (B.15), and the expression for $\boldsymbol{\Omega}^{-1}$ (B.31) into (B.29)

$$CRB(\boldsymbol{\theta}) = \frac{1}{2\chi_{1,N}} \mathrm{Re} \left(\overline{\mathbf{L}} - \frac{\chi_{2,N}}{\chi_{1,N}} \frac{\overline{\mathbf{G}^*\mathbf{b}\mathbf{b}^*\mathbf{G}}}{\mathbf{b}^*\mathbf{H}\mathbf{b}} - \overline{\mathbf{G}^* \left(\mathbf{H}^{-1} - \frac{\chi_{2,N}}{\chi_{1,N}} \frac{\mathbf{b}\mathbf{b}^*}{\mathbf{b}^*\mathbf{H}\mathbf{b}} \right) \mathbf{G}} \right)^{-1} = \tag{B.32}$$

$$= \frac{1}{2\chi_{1,N}} \left(\overline{\mathbf{L}} - \overline{\mathbf{G}^*\mathbf{H}^{-1}\mathbf{G}} \right)^{-1} \tag{B.33}$$

CRB for assumptions (C1a–c, C2b, C3a–b): Now we will derive the second CRB expression and show that the $CRB(\boldsymbol{\theta})$ for (C1a–c, C2b, C3a–b) equals (B.33). We substitute the expressions (B.4–B.8) into (B.3), whith the unknown parameters

$$\boldsymbol{\eta} = \begin{bmatrix} \overline{\mathbf{b}}^T & \widetilde{\mathbf{b}}^T & \vdots & \nu & \boldsymbol{\theta}^T \end{bmatrix} \tag{B.34}$$

giving

$$CRB(\boldsymbol{\eta}) = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{12}^T & \mathbf{F}_{22} \end{bmatrix}^{-1} \tag{B.35}$$

where

$$\mathbf{F}_{11} = 2N \begin{bmatrix} \overline{\mathbf{H}} & -\widetilde{\mathbf{H}} \\ \widetilde{\mathbf{H}} & \overline{\mathbf{H}} \end{bmatrix} \tag{B.36}$$

$$\mathbf{F}_{12} = 2N \begin{bmatrix} -\beta\widetilde{\mathbf{H}\mathbf{b}} & \overline{\mathbf{G}} \\ \beta\overline{\mathbf{H}\mathbf{b}} & \widetilde{\mathbf{G}} \end{bmatrix} \tag{B.37}$$

$$\mathbf{F}_{22} = 2N \begin{bmatrix} \gamma\overline{\mathbf{b}^*\mathbf{H}\mathbf{b}} & \beta\widetilde{\mathbf{b}^*\mathbf{G}} \\ -\beta\widetilde{\mathbf{G}^*\mathbf{b}} & \overline{\mathbf{L}} \end{bmatrix} \tag{B.38}$$

Since $\mathbf{H}$ is hermitian, the inverse $\mathbf{F}_{11}^{-1}$ is given by

$$\mathbf{F}_{11}^{-1} = 2N \begin{bmatrix} \overline{\mathbf{H}^{-1}} & -\widetilde{\mathbf{H}^{-1}} \\ \widetilde{\mathbf{H}^{-1}} & \overline{\mathbf{H}^{-1}} \end{bmatrix} \tag{B.39}$$

wich is easily prooved by observing

$$\overline{\mathbf{H}}\overline{\mathbf{H}^{-1}} - \widetilde{\mathbf{H}}\widetilde{\mathbf{H}^{-1}} = \mathbf{I} \tag{B.40}$$

$$\overline{\mathbf{H}}\widetilde{\mathbf{H}^{-1}} + \widetilde{\mathbf{H}}\overline{\mathbf{H}^{-1}} = \mathbf{0} \tag{B.41}$$

Partitioning the matrix inverse, the lower right hand block of the CRB matrix corresponding to $[\nu, \boldsymbol{\theta}]$ is given by

$$CRB(\nu, \boldsymbol{\theta})^{-1} = \mathbf{F}_{22} - \mathbf{F}_{12}^T \mathbf{F}_{11}^{-1} \mathbf{F}_{12} \tag{B.42}$$

where, using (B.37)-(B.39)

$$\begin{aligned} \mathbf{F}_{12}^T \mathbf{F}_{11}^{-1} \mathbf{F}_{12} = 2N &\begin{bmatrix} \beta\widetilde{\mathbf{b}^*\mathbf{H}} & \beta\overline{\mathbf{b}^*\mathbf{H}} \\ \overline{\mathbf{G}^*} & -\widetilde{\mathbf{G}^*} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{H}^{-1}} & -\widetilde{\mathbf{H}^{-1}} \\ \widetilde{\mathbf{H}^{-1}} & \overline{\mathbf{H}^{-1}} \end{bmatrix} \cdot \\ &\begin{bmatrix} -\beta\widetilde{\mathbf{H}\mathbf{b}} & \overline{\mathbf{G}} \\ \beta\overline{\mathbf{H}\mathbf{b}} & \widetilde{\mathbf{G}} \end{bmatrix} = \end{aligned} \tag{B.43}$$

$$= 2N \begin{bmatrix} \beta^2\overline{\mathbf{b}^*\mathbf{H}\mathbf{b}} & \beta\widetilde{\mathbf{b}^*\mathbf{G}} \\ -\beta\widetilde{\mathbf{G}^*\mathbf{b}} & \overline{\mathbf{G}^*\mathbf{H}^{-1}\mathbf{G}} \end{bmatrix} \tag{B.44}$$

Substituting (B.38) and (B.44) into (B.42) gives $CRB(\nu, \boldsymbol{\theta})$ for (C1a–c, C2a, C3a–b)

$$CRB(\nu, \boldsymbol{\theta}) = \frac{1}{2N}\begin{bmatrix} (\gamma - \beta^2)\overline{\mathbf{b}^*\mathbf{H}\mathbf{b}} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{L}} - \overline{\mathbf{G}^*\mathbf{H}^{-1}\mathbf{G}} \end{bmatrix}^{-1} \tag{B.45}$$

and hence

$$CRB(\boldsymbol{\theta}) = \frac{1}{2N}\left(\overline{\mathbf{L}} - \overline{\mathbf{G}^*\mathbf{H}^{-1}\mathbf{G}}\right)^{-1} \tag{B.46}$$

The $CRB(\boldsymbol{\theta})$ of Equation (B.46) equals Equation (B.33) since $\chi_{1,N} = N$ when $x(t) = e^{j\nu t}$. We can thus conclude that Equation (B.33) gives the $CRB(\boldsymbol{\theta})$ for all combinations of assumptions C1a–c, C2a–b, C3a–b.

C Kernel matrix for the Barankin Bound

The scalar valued kernel $B(i,j)$ is defined by, see Section 5

$$B(i,j)\Delta \triangleq \mathrm{E}\left\{\frac{p(\{\mathbf{y}(t)\}|\boldsymbol{\eta}_i)}{p(\{\mathbf{y}(t)\}|\boldsymbol{\eta})}\frac{p(\{\mathbf{y}(t)\}|\boldsymbol{\eta}_j)}{p(\{\mathbf{y}(t)\}|\boldsymbol{\eta})}\right\} \tag{C.1}$$

where

$$p(\{\mathbf{y}(t)\}|\boldsymbol{\eta}_i) = \frac{1}{\pi^{mN} * det\mathbf{Q}_i^N}\exp\sum_{t=1}^{N}\left(\mathbf{y}(t)-\mathbf{A_i}\mathbf{b_i}e^{j\nu_i t}\right)^* \mathbf{Q}_i^{-1}\left(\mathbf{y}(t)-\mathbf{A_i}\mathbf{b_i}e^{j\nu_i t}\right) \tag{C.2}$$

Using the notation $\boldsymbol{\mu}_i(t) = \mathbf{A}(\boldsymbol{\theta}_i)\mathbf{b}_i e^{\nu_i t}$ the kernel expression is evaluated in (C.3-C.5)

$$\begin{aligned} B(i,j) = \prod_{t=1}^{N}\int_{\Omega_\mathbf{y}} \frac{|\mathbf{Q}|}{\pi^m|\mathbf{Q}_i||\mathbf{Q}_j|}\exp\Big\{ & -\mathbf{y}^*(t)[\mathbf{Q}_i^{-1}+\mathbf{Q}_j^{-1}-\mathbf{Q}^{-1}]\mathbf{y}(t) \\ & + 2\mathrm{Re}\,\mathbf{y}^*(t)[\mathbf{Q}_i^{-1}\boldsymbol{\mu}_i(t)+\mathbf{Q}_j^{-1}\boldsymbol{\mu}_j(t)-\mathbf{Q}^{-1}\boldsymbol{\mu}(t)] \\ & - \boldsymbol{\mu}_i^*(t)\mathbf{Q}_i^{-1}\boldsymbol{\mu}_i(t)-\boldsymbol{\mu}_j^*(t)\mathbf{Q}_j^{-1}\boldsymbol{\mu}_j(t)+\boldsymbol{\mu}^*(t)\mathbf{Q}^{-1}\boldsymbol{\mu}(t)\Big\}d\mathbf{y}(t) \end{aligned} \tag{C.3}$$

$$\begin{aligned} = & \left(\frac{|\mathbf{Q}|}{2\pi|\mathbf{Q}_i||\mathbf{Q}_j|}\right)^N \cdot \exp\sum_{t=1}^{N}\Big\{-\left\|\mathbf{Q}_i^{-1/2}\boldsymbol{\mu}_i(t)\right\|^2-\left\|\mathbf{Q}_j^{-1/2}\boldsymbol{\mu}_j(t)\right\|^2+\left\|\mathbf{Q}^{-1/2}\boldsymbol{\mu}(t)\right\|^2 \\ & + \left\|\tilde{\mathbf{Q}}_{ij}^{1/2}\left(\mathbf{Q}_i^{-1}\boldsymbol{\mu}_i(t)+\mathbf{Q}_j^{-1}\boldsymbol{\mu}_j(t)-\mathbf{Q}^{-1}\boldsymbol{\mu}(t)\right)\right\|^2\Big\}\cdot \\ & \cdot\int_{\Omega_\mathbf{y}}\exp\sum_{t=1}^{N}-\left\|\tilde{\mathbf{Q}}_{ij}^{-1/2}\left(\mathbf{y}(t)-\tilde{\mathbf{Q}}_{ij}\left(\mathbf{Q}_i^{-1}\boldsymbol{\mu}_i(t)+\mathbf{Q}_j^{-1}\boldsymbol{\mu}_j(t)-\mathbf{Q}^{-1}\boldsymbol{\mu}(t)\right)\right)\right\|^2 d\mathbf{y}(t) \end{aligned} \tag{C.4}$$

$$\begin{aligned} = & \left(\frac{|\mathbf{Q}|\left|\tilde{\mathbf{Q}}_{ij}\right|}{|\mathbf{Q}_i||\mathbf{Q}_j|}\right)^N \cdot \exp\sum_{t=1}^{N}\Big\{-\left\|\mathbf{Q}_i^{-1/2}\boldsymbol{\mu}_i(t)\right\|^2-\left\|\mathbf{Q}_j^{-1/2}\boldsymbol{\mu}_j(t)\right\|^2+\left\|\mathbf{Q}^{-1/2}\boldsymbol{\mu}(t)\right\|^2 \\ & + \left\|\tilde{\mathbf{Q}}_{ij}^{1/2}\left(\mathbf{Q}_i^{-1}\boldsymbol{\mu}_i(t)+\mathbf{Q}_j^{-1}\boldsymbol{\mu}_j(t)-\mathbf{Q}^{-1}\boldsymbol{\mu}(t)\right)\right\|^2\Big\} \end{aligned} \tag{C.5}$$

D Modified MODE algorithm

In this section we show how the MODE algorithm can be modified to be applicable for estimating the elevation parameters assuming C1a or C1b. For any Vandermonde matrix $\mathbf{H} \in C^{m\times n}$, $m > n$ it is possible to reparameterize Tr $(\mathbf{I} - \mathbf{H}(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*)\mathbf{S}\mathbf{S}^*$ by the coefficients of the following function

$$p(z) = p_0 z^n + p_1 z^{n-1} + \ldots + p_n = p_0 \prod_{k=1}^{n} (z - e^{-j\omega_k}) \tag{D.1}$$

where the ω_k's are frequencies of $\mathbf{H}$

$$\mathbf{H} = \begin{bmatrix} 1 & \ldots & 1 \\ e^{j\omega_1} & \ldots & e^{j\omega_n} \\ \vdots & \ddots & \vdots \\ e^{j(m-1)\omega_1} & \ldots & e^{j(m-1)\omega_n} \end{bmatrix} \tag{D.2}$$

Thus, the frequencies can easily be extracted from the roots of the polynomial $p(z)$. Introduce the matrix $\mathbf{B} \in C^{m\times(m-n)}$

$$\mathbf{B}^* = \begin{bmatrix} p_n & \ldots p_0 & & \mathbf{0} \\ & \ddots & \ddots & \\ \mathbf{0} & p_n & \ldots & p_0 \end{bmatrix} \tag{D.3}$$

and observe that

$$\mathbf{B}^*\mathbf{H} = 0 \tag{D.4}$$

Since $\mathbf{B}$ has rank $m - d$, it follows that $\mathbf{B}$ spans the nullspace of $\mathbf{H}^*$, i.e.

$$\mathbf{B}(\mathbf{B}^*\mathbf{B})^{-1}\mathbf{B}^* = (\mathbf{I} - \mathbf{H}(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*) \tag{D.5}$$

The reparameterization is then given as

$$\text{Tr}\,(\mathbf{I} - \mathbf{H}(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*)\mathbf{S}\mathbf{S}^* = \text{Tr}\,\mathbf{B}(\mathbf{B}^*\mathbf{B})^{-1}\mathbf{B}^*\mathbf{S}\mathbf{S}^* = \tag{D.6}$$

$$\text{Tr}\,(\mathbf{B}^*\mathbf{B})^{-1}\mathbf{B}^*\mathbf{S}\mathbf{S}^*\mathbf{B} \tag{D.7}$$

It has been shown [14], that the term $(\mathbf{B}^*\mathbf{B})^{-1}$ can be replaced by any consistent estimate without changing the asymptotic properties of the minimizing $\mathbf{B}$. The resulting function to minimize is then

$$\text{Tr}\,(\hat{\mathbf{B}}^*\hat{\mathbf{B}})^{-1}\mathbf{B}^*\mathbf{S}\mathbf{S}^*\mathbf{B} \tag{D.8}$$

which is quadratic in $\mathbf{B}$ formed by the unknown polynomial coefficients of $p(z)$. The quadratic function can be minimized explicitly with respect to the unknown real and imaginary parts of the p_k's, by solving a system of linear equations. But first we need to make sure that the roots are on the unit circle and that they satisfy any other constraint such as the multipath symmetry.

To find the minimizing $\boldsymbol{\theta}$ for the Vandermonde matrix $\mathbf{H} = \mathbf{A}(\boldsymbol{\theta})$ given an initial $\hat{\boldsymbol{\theta}}$ translates into minimizing the MODE function (D.8) with $p(z) = p_0(z - e^{j\omega_1})(z - e^{j\omega_2})$. A root on the unit circle is guaranteed if

$$p(z) = pz^2 + \mu z + p^* = |p|\left(e^{-j\frac{\omega_1+\omega_2}{2}}z^2 - 2cos(\frac{\omega_1 - \omega_2}{2})z + e^{j\frac{\omega_1+\omega_2}{2}}\right) \tag{D.9}$$

For C1a there is no further information relating ω_1, ω_2 and we use the scaling $p = 1 + jp_i$. The minimizing p_i and μ of the quadratic criterion will result in a standard LS solution. The spatial frequencies, $\omega_{1,2} = \kappa d\theta_{1,2}$, defined according to (3.6), are the roots of (D.9).

For C1b geometry leads to

$$\omega_1 + \omega_2 = 2 * \omega_h = 2\kappa dh/R \tag{D.10}$$

$$\omega_1 - \omega_2 = 2 * \omega = 2\kappa dH/R \tag{D.11}$$

allowing the polynomial (D.9) to be rewritten

$$p(z) = e^{-j\omega_h}z^2 - 2cos(\omega)z + e^{j\omega_h} \tag{D.12}$$

where $\omega_h = \kappa d\theta_h$ is known and $\omega = \kappa d\theta$. The minimization problem reduces to the simple problem of finding the single unknown $\mu = -2cos(\omega)$. But for ω close to zero, noise distortion can give $|\mu| > 2$. If that happens, the best estimate is $\hat{\omega} = \hat{\theta} = 0$.

For C1c we cannot exploit the additional information and keep the ULA structure of $\mathbf{A}(\theta)$. We have to use brute force search, i.e. a grid search, to find $\hat{\boldsymbol{\theta}}$.

References

[1] D. K. Barton. Low-angle radar tracking. IEEE Proc., 62(6):687–704, December 1974.

[2] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai. Radar Array Processing, pages 99–151. Springer-Verlag, New-York, USA, 1993.

[3] P. Stoica and R. Moses. Introduction to Spectral Analysis. Prentice-Hall, Inc., Upper Saddle River, NJ, 1997.

[4] S. Kay. Fundamentals of Statistical Signal Processing Estimation Theory. Prentice-Hall, Inc., Upper Saddle River, NJ, 1993.

[5] E. Bossé, R. Turner, and M. Lecours. Tracking swerling fluctuating targets at low altitude over the sea. IEEE Trans. Aerosp. Electron. Syst., 27(5):806–822, Sept. 1991.

[6] E. Bossé, R. Turner, and E. S. Riseborough. Model-based multifrequency array signal processing for low-angle tracking. IEEE Trans. Aerosp. Electron. Syst., 31(1):194–210, Jan. 1995.

[7] T. Lo and J. Litva. Use of a highly deterministic multipath signal model in low-angle tracking. IEE proc. -F, 138(2):163–171, Apr. 1991.

[8] T. S. Lee M. D. Zoltowski. Maximum likelihood based sensor array signal processing in the beamspace domain for low angle radar tracking. Signal Proc., 39(3):656–671, 3 1991.

[9] E. W. Barankin. Locally best unbiased estimates. Ann. Math. Statist., 20:477–501, 1949.

[10] R. J. McAulay and E. M. Hofstetter. Barankin bounds on parameter estimation. IEEE Trans. Inform. Theory, 17(6):669–676, Nov. 1971.

[11] I. Reuven and H. Messer. A barankin-type lower bound on the estimation of a hybrid parameter vector. IEEE Trans. Inform. Theory, 43(3):1084–1093, May 1997.

[12] K. Boman. The low-angle target problem: maximum likelihood, barankin bound and optimal arrays. In Proc. IEE EUREL meeting on radar and sonar signal processing, pages 16/1–16/2, Peebles, UK, July 1998.

[13] B. Andersson, M Persson, and K. Boman. Fmcw and superresolution techniques applied to an lpi short range air search radar. In Proc. IEE International Conference RADAR 97, pages 406–410, Edinburgh, UK, October 1997.

[14] P. Stoica and K. Sharman. Maximum likelihood methods for direction of arrival estimation. IEEE Trans. Acoust., Speech, Signal Proc., 38(7):1132–1143, Jul. 1990.

[15] P. Stoica and A. Nehorai. Music, maximum likelihood and cramér-rao bound. IEEE Trans. Acoust., Speech, Signal Proc., 37(5):720–741, May 1989.

[16] P. Stoica and A. Nehorai. Music, maximum likelihood and cramér-rao bound: further results and comparisions. IEEE Trans. Acoust., Speech, Signal Proc., 38(12):2140–2150, Dec. 1990.

[17] P. Stoica, B. Ottersten, M. Viberg, and R. L. Moses. Maximum likelihood array processing for stochastic coherent sources. IEEE Trans. Signal Proc., 44(1):96–105, Jan. 1996.

[18] J. Sheinvald, M. Wax, and A. J. Weiss. On maximum-likelihood localization of coherent signals. IEEE Trans on Signal Proc., 44(10):2475–2481, Oct. 1996.

[19] M. Viberg, P. Stoica, and B. Ottersten. Maximum likelihood array processing in spatially correlated noise fields using parametrized signals. IEEE Trans Signal Proc., 45(4):996–1004, Apr. 1997.

[20] L. Swindlehurst and P. Stoica. Maximum likelihood methods in radar array signal processing. Proceedings of the IEEE, 86(2):421–441, Feb. 1998.

[21] P. Stoica and A. Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. IEEE Trans. Acoust., Speech, Signal Proc., 38(10):1783–1795, Oct. 1990.

[22] M. H. Er K. C. Tan, G. L. Oh. A study of the uniqueness of steering vectors in array processing. Signal Proc., 34(3):245–256, December 1993.

[23] P. Stoica and R. Moses. Introduction to Spectral Analysis, pages 285–297. Prentice-Hall, Inc., Upper Saddle River, NJ, 1997.