ON FORWARD-BACKWARD MODE FOR DIRECTION OF ARRIVAL ESTIMATION

Magnus Jansson
Dept. of Signals, Sensors, & Systems
Royal Institute of Technology (KTH)
SE-100 44 Stockholm, SWEDEN
Tel: +46 8 790 7425, Fax: +46 8 790 7329
magnus.jansson@s3.kth.se

Petre Stoica
Systems & Control Group
Uppsala University
SE-751 03 Uppsala
SWEDEN

ABSTRACT

We apply the MODE (method of direction estimation) principle to the forward-backward (FB) covariance of the output vector of a sensor array to obtain what we call the FB-MODE procedure. The derivation of FB-MODE is an interesting exercise in matrix analysis, the outcome of which was somewhat unexpected: FB-MODE simply consists of applying the standard MODE approach to the eigenelements of the FB sample covariance matrix. By using an asymptotic expansion technique we also establish the surprising result that FB-MODE is outperformed, from a statistical standpoint, by the standard MODE applied to the forward-only sample covariance (F-MODE). We believe this to be an important result that shows that the FB approach, which proved quite useful for improving the performance of many suboptimal array processing methods, should *not* be used with a statistically optimal method such as F-MODE.

1 INTRODUCTION

MODE is a method of direction estimation by means of a sensor array, which is known to be statistically efficient in cases when either the number of data samples (N) or the signal-to-noise ratio (SNR) is sufficiently large [1, 2]. For uniform and linear arrays (ULAs) the implementation of MODE requires only standard matrix operations (such as an eigendecomposition) and is thus very straightforward. In fact in the ULA case MODE is a clear candidate for the best possible (i.e., computationally simple and statistically efficient) array processing method.

We stress that the aforementioned statistical efficiency of MODE holds asymptotically (in N or SNR). In short-sample or low-SNR cases other methods may provide better performance. The forward-backward (FB) approach is a well-known methodology that has been successfully used to enhance the performance of a number of array signal processing algorithms [3, 4]. Recently this approach has been used in conjunction with MODE probably in an attempt to obtain enhanced (finite sample or SNR) performance [5]. More exactly in the cited publications the standard MODE approach was applied to the eigenelements of the FB sample covariance matrix.

The first problem we deal with in this paper concerns the formal application of the MODE principle to the FB sample covariance. After an exercise in matrix analysis we prove the somewhat unexpected result that FB-MODE should indeed consist of applying the standard MODE to the eigenelements of the FB sample covariance matrix.

Then we go on to establish the asymptotic statistical performance of FB-MODE. Because the standard F-MODE is asymptotically statistically efficient (in the sense that it achieves the Cramér-Rao bound (CRB)) [1, 2], we cannot expect FB-MODE to perform better in asymptotic regimes. What one might expect is that FB-MODE outperforms F-MODE in non-asymptotic regimes and that the two methods have the same asymptotic performance. We show that the conjecture on identical asymptotic performances, even though quite natural, is *false*: FB-MODE's asymptotic performance is usually inferior to that of F-MODE. Regarding the comparison of the two methods in non-asymptotic regimes, because the finite-sample/SNR analysis is intractable, we resort to Monte-Carlo simulations to show that F-MODE typically outperforms FB-MODE also in such cases.

In conclusion, the FB approach – which has been successfully employed to enhance the performance of several array signal processing algorithms – cannot be recommended for use with MODE. This result also shows that the related subspace fitting or asymptotic maximum likelihood approaches of [1] and [2], which produced the statistically efficient MODE in the forward-only case, may fail to yield statistically optimal estimators in other cases.

2 DATA MODEL

Let $\mathbf{x}(t) \in \mathbb{C}^{m\times 1}$ denote the output vector of a ULA that comprises m elements. The variable t denotes the sampling point; hence, $t = 1, 2, \ldots, N$, where N is the number of available samples. Under the assumption that the signals impinging on the array are narrowband with the same center frequency, the array output can be described by the equation

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t). \tag{1}$$

Here, $\mathbf{s}(t) \in \mathbb{C}^{d\times 1}$ is the vector of the d signals (translated to baseband) that impinge on the array, $\mathbf{n}(t) \in \mathbb{C}^{m\times 1}$ is a noise term, and $\mathbf{A}$ is the Vandermonde matrix

$$\mathbf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ e^{i\omega_1} & & e^{i\omega_d} \\ \vdots & & \vdots \\ e^{i(m-1)\omega_1} & \cdots & e^{i(m-1)\omega_d} \end{bmatrix},$$

where $\{\omega_k\}_{k=1}^d$ are the so-called *spatial frequencies.* The following assumption, frequently used in the array processing literature, is also made here: the signal vector and the noise are temporally white, zero-mean, circular Gaussian random variables that are independent of one another; additionally, the noise is spatially white and has the same power in all sensors. Mathematically, this assumption implies that

$$\mathrm{E}\{\mathbf{s}(t)\mathbf{s}^*(s)\} = \mathbf{P}_0\delta_{t,s}\,, \qquad \mathrm{E}\{\mathbf{s}(t)\mathbf{s}^T(s)\} = \mathbf{0}, \tag{2}$$

$$\mathrm{E}\{\mathbf{n}(t)\mathbf{n}^*(s)\} = \sigma^2\mathbf{I}\delta_{t,s}\,, \qquad \mathrm{E}\{\mathbf{n}(t)\mathbf{n}^T(s)\} = \mathbf{0}, \tag{3}$$

where E is the statistical expectation operator, $\delta_{t,s}$ is the Kronecker delta, and the superscripts $*$ and T denote the conjugate transpose and the transpose, respectively.

A principal goal of array processing consists of estimating $\{\omega_k\}_{k=1}^d$ from $\{\mathbf{x}(t)\}_{t=1}^N$. Once $\{\omega_k\}$ are estimated, the directions (also called angles-of-arrival) of the d signals can be easily obtained.

It follows from (1), (2)-(3) that the array output $\mathbf{x}(t)$ is a temporally white, zero-mean, circular Gaussian random variable with covariance

$$\mathrm{E}\{\mathbf{x}(t)\mathbf{x}^*(t)\} = \mathbf{A}\mathbf{P}_0\mathbf{A}^* + \sigma^2\mathbf{I} \triangleq \mathbf{R}_0. \tag{4}$$

Let $\mathbf{J}$ denote the reversal matrix of dimension $m \times m$,

$$\mathbf{J} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & & 1 & 0 \\ \vdots & \iddots & & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}.$$

It is readily checked that

$$\mathbf{J}\mathbf{A}^{\mathrm{c}} = \mathbf{A}\mathbf{\Phi}, \tag{5}$$

where the superscript c stands for the complex conjugate, and

$$\mathbf{\Phi} = \mathrm{diag}\{e^{-i(m-1)\omega_1} \;\ldots\; e^{-i(m-1)\omega_d}\}.$$

Hence, we have

$$\mathbf{J}\mathbf{R}_0^{\mathrm{c}}\mathbf{J} = \mathbf{A}\mathbf{\Phi}\mathbf{P}_0^{\mathrm{c}}\mathbf{\Phi}^*\mathbf{A}^* + \sigma^2\mathbf{I}.$$

Let

$$\mathbf{R} = \frac{1}{2}\left(\mathbf{R}_0 + \mathbf{J}\mathbf{R}_0^{\mathrm{c}}\mathbf{J}\right) = \mathbf{A}\mathbf{P}\mathbf{A}^* + \sigma^2\mathbf{I}, \tag{6}$$

where $\mathbf{P} = \frac{1}{2}\left(\mathbf{P}_0 + \mathbf{\Phi}\mathbf{P}_0^{\mathrm{c}}\mathbf{\Phi}^*\right)$. In the equations above, $\mathbf{R}_0$ is the so-called "forward" covariance matrix, whereas $\mathbf{R}$ is called the "forward-backward" covariance. It will be useful for what follows to observe that $\mathbf{R}_0$ and $\mathbf{R}$ have the same structure: the only difference between the expressions in (4) and (6) for these two matrices is that $\mathbf{P}_0$ in $\mathbf{R}_0$ is replaced by $\mathbf{P}$ in $\mathbf{R}$.

The sample covariance matrices corresponding to $\mathbf{R}_0$ and $\mathbf{R}$ are defined as

$$\hat{R}_0 = \frac{1}{N}\sum_{t=1}^{N}\mathbf{x}(t)\mathbf{x}^*(t),$$

and

$$\hat{R} = \frac{1}{2}\left(\hat{R}_0 + \mathbf{J}\hat{R}_0^{\mathrm{c}}\mathbf{J}\right). \tag{7}$$

It is well known that, under the assumptions made, $\hat{R}_0$ and $\hat{R}$ converge (with probability one and in mean square) to $\mathbf{R}_0$ and, respectively, $\mathbf{R}$, as N tends to infinity. Hence $\hat{R}_0$ and $\hat{R}$ are consistent estimates of the corresponding theoretical covariance matrices. In fact $\hat{R}_0$ can be shown to be the (unstructured) maximum likelihood estimate (MLE) of $\mathbf{R}_0$. By the invariance principle of ML estimation, it then follows that $\hat{R}$ is the MLE of $\mathbf{R}$.

It is straightforward to prove that the condition number of $\mathbf{P}$ is less than or equal to the condition number of $\mathbf{P}_0$. Based on that observation we can say that the signals are less correlated in $\mathbf{R}$ than in $\mathbf{R}_0$. In particular, $\mathbf{P}$ may be nonsingular even though $\mathbf{P}_0$ is singular. This observation, along with the fact that the performance of many array processing algorithms deteriorates as the signal correlation increases, lies at the basis of the belief that the FB approach should outperform the F-only approach. This would really be true if one had two sets of independent data vectors with covariance matrices $\mathbf{R}_0$ and, respectively, $\mathbf{R}$. However, only $\mathbf{R}_0$ and $\hat{R}_0$ correspond to such a data set, whereas $\mathbf{R}$ and $\hat{R}$ do not. In fact, *the statistical properties of $\hat{R}$ are quite different from those of $\hat{R}_0$, and this may very well counterbalance the desirable property that $\mathbf{P}$ in $\mathbf{R}$ is better conditioned than $\mathbf{P}_0$ in $\mathbf{R}_0$.* Consequently, the FB approach may not lead to any performance enhancement. This discussion provides an intuitive motivation for the superiority of F-MODE over FB-MODE, which will be shown later in the paper.

Next, let us introduce the Toeplitz matrix

$$\mathbf{B}^* = \begin{bmatrix} b_0 & \cdots & b_d & & \mathbf{0} \\ & \ddots & & \ddots & \\ \mathbf{0} & & b_0 & \cdots & b_d \end{bmatrix} \qquad (m-d)\times m, \tag{8}$$

where the complex-valued coefficients $\{b_k\}$ are defined through

$$b_0 + b_1 z + \cdots + b_d z^d = b_d \textstyle\prod_{k=1}^d \left(z - e^{i\omega_k}\right). \tag{9}$$

Because the polynomial in (9) has all the zeros on the unit circle, it can be written such that its coefficients satisfy the so-called "conjugate symmetry constraint"

$$b_k = b_{d-k}^{\mathrm{c}} \qquad (\text{for } k = 0, 1, \ldots, d). \tag{10}$$

It follows easily from (8) and (9) that $\mathbf{B}^*\mathbf{A} = \mathbf{0}$, which, along with the fact that $\mathrm{rank}(\mathbf{A}) = d$ and $\mathrm{rank}(\mathbf{B}) =$

$m-d$, implies that $\mathcal{R}(\mathbf{B}) = \mathcal{N}(\mathbf{A}^*)$. Hereafter, $\mathcal{R}(\cdot)$ and $\mathcal{N}(\cdot)$ denote the range, respectively, the null space associated with the matrix in question.

To end up these notational preparations, let

$$\mathbf{R} = [\underbrace{\mathbf{E}_s}_{d} \quad \underbrace{\mathbf{E}_n}_{m-d}] \begin{bmatrix} \mathbf{\Lambda}_s & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{E}_s^* \\ \mathbf{E}_n^* \end{bmatrix} \tag{11}$$

denote the eigenvalue decomposition (EVD) of the matrix $\mathbf{R}$, where $\mathbf{\Lambda}_s = \text{diag}\{\lambda_1 \ \dots \ \lambda_d\}$. Here, $\{\lambda_k\}$ are the d largest eigenvalues of $\mathbf{R}$ arranged in a decreasing order: $\lambda_1 > \lambda_2 > \cdots > \lambda_d$. We assume that $\lambda_k \neq \lambda_p$ (for $k \neq p;\ k, p = 1, 2, \dots, d$), which is generically true. Note from (11) that the smallest $(m-d)$ eigenvalues of $\mathbf{R}$ are identical and equal to σ^2, a fact that follows easily from (6). The EVD of $\mathbf{R}_0$ is similarly defined, but with $\mathbf{E}_s$, $\mathbf{E}_n$, and $\mathbf{\Lambda}_s$ replaced by $\mathbf{E}_{s_0}$, $\mathbf{E}_{n_0}$, and $\mathbf{\Lambda}_{s_0}$. We let

$$\hat{R} = [\underbrace{\hat{\mathbf{E}}_s}_{d} \quad \underbrace{\hat{\mathbf{E}}_n}_{m-d}] \begin{bmatrix} \hat{\mathbf{\Lambda}}_s & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{\Lambda}}_n \end{bmatrix} \begin{bmatrix} \hat{\mathbf{E}}_s^* \\ \hat{\mathbf{E}}_n^* \end{bmatrix}$$

denote the EVD of $\hat{R}$, with the eigenvalues arranged in a decreasing order. The EVD of $\hat{R}_0$ is similarly defined, but with a subscript 0 attached to $\hat{\mathbf{E}}_s$, etc., to distinguish it from the EVD of $\hat{R}$.

3 BRIEF REVIEW OF F-MODE

The approach of MODE consists of applying the maximum likelihood (ML) principle to the sample eigenvectors of $\hat{R}_0$ [2]. More precisely, MODE provides the asymptotically best (in the mean square sense) consistent (ABC) estimate of the parameter vector $\boldsymbol{\omega} \triangleq [\omega_1 \ \dots \ \omega_d]^T$ based on the random vector

$$\mathbf{z} = \text{vec}(\mathbf{B}^* \hat{\mathbf{E}}_{s_0}), \tag{12}$$

where vec denotes column-wise vectorization. The ABC criterion corresponding to $\mathbf{z}$ above can be shown to be

$$\text{Tr}\{(\mathbf{B}^*\mathbf{B})^{-1} \mathbf{B}^* \hat{\mathbf{E}}_{s_0} \hat{\tilde{\mathbf{\Lambda}}}_0^2 \hat{\mathbf{\Lambda}}_{s_0}^{-1} \hat{\mathbf{E}}_{s_0}^* \mathbf{B}\}, \tag{13}$$

where Tr denotes the trace operator, $\hat{\tilde{\mathbf{\Lambda}}}_0 = \hat{\mathbf{\Lambda}}_{s_0} - \hat{\sigma}^2 \mathbf{I}$ and where $\hat{\sigma}^2$ is a consistent estimate of the noise power. The F-MODE estimate of $\boldsymbol{\omega}$ is obtained as an asymptotically valid approximation to the minimizer of (13), in the following steps [2]:

1. Obtain $\hat{\mathbf{E}}_{s_0}$, $\hat{\mathbf{\Lambda}}_{s_0}$, and $\hat{\tilde{\mathbf{\Lambda}}}_0$ from the EVD of $\hat{R}_0$. Derive an initial estimate of $\{b_k\}$ by minimizing the quadratic function
$$\text{Tr}\{\mathbf{B}^* \hat{\mathbf{E}}_{s_0} \hat{\tilde{\mathbf{\Lambda}}}_0^2 \hat{\mathbf{\Lambda}}_{s_0}^{-1} \hat{\mathbf{E}}_{s_0}^* \mathbf{B}\}.$$

2. Let $(\hat{\mathbf{B}}^*\hat{\mathbf{B}})$ be the estimate of $(\mathbf{B}^*\mathbf{B})$ made from the previously obtained estimates of $\{b_k\}$. Derive refined estimates of $\{b_k\}$ by minimizing the quadratic function
$$\text{Tr}\{(\hat{\mathbf{B}}^*\hat{\mathbf{B}})^{-1} \mathbf{B}^* \hat{\mathbf{E}}_{s_0} \hat{\tilde{\mathbf{\Lambda}}}_0^2 \hat{\mathbf{\Lambda}}_{s_0}^{-1} \hat{\mathbf{E}}_{s_0}^* \mathbf{B}\}.$$
Possibly repeat this operation once more, using the latest estimate of $\{b_k\}$ to obtain $(\hat{\mathbf{B}}^*\hat{\mathbf{B}})$. Finally, derive estimates of $\{\omega_k\}$ by rooting the polynomial $\sum_{k=0}^{d} \hat{b}_k z^k$ (see (9)).

The minimization in steps 1 and 2 above should be conducted under an appropriate constraint on $\{b_k\}$ (to prevent the trivial solution $\{b_k = 0\}$). Typically, one uses $\sum_{k=0}^{d} |b_k|^2 = 1$. Additionally the $\{b_k\}$ should satisfy (10).

Asymptotically (as N increases) the F-MODE estimate is Gaussian distributed with mean equal to the true parameter vector $\boldsymbol{\omega}$ and covariance matrix,

$$\mathbf{C}_{\text{F}} = \frac{\sigma^2}{2N} \left\{ \text{Re}\left[\mathbf{U} \odot \mathbf{Q}_0^T \right] \right\}^{-1}, \tag{14}$$

where $\odot$ denotes the Hadamard product (i.e., element-wise multiplication),

$$\mathbf{U} = \mathbf{D}^* \mathbf{\Pi}_{\mathbf{A}}^{\perp} \mathbf{D},$$
$$\mathbf{Q}_0 = \mathbf{P}_0 \mathbf{A}^* \mathbf{R}_0^{-1} \mathbf{A} \mathbf{P}_0,$$

and where

$$\mathbf{D} = [\mathbf{d}(\omega_1) \quad \dots \quad \mathbf{d}(\omega_d)]; \qquad \mathbf{d}(\omega) = \frac{d\mathbf{a}(\omega)}{d\omega}$$

and $\mathbf{\Pi}_{\mathbf{A}}^{\perp} = \mathbf{I} - \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$ is the orthogonal projector onto $\mathcal{N}(\mathbf{A}^*)$. Because (14) is also the CRB matrix for the estimation problem under consideration, it follows that F-MODE is asymptotically statistically efficient [1, 2].

4 DERIVATION OF FB-MODE

Similarly to (12), the FB-MODE is associated with the ABC estimate derived from

$$\mathbf{z} = \text{vec}(\mathbf{B}^* \hat{\mathbf{E}}_s)$$

where $\hat{\mathbf{E}}_s$ is obtained from the EVD of the FB sample covariance matrix. In [6], we show that the ABC criterion corresponding to the FB approach is given by

$$\text{Tr}\{(\mathbf{B}^*\mathbf{B})^{-1} \mathbf{B}^* \hat{\mathbf{E}}_s \hat{\tilde{\mathbf{\Lambda}}}^2 \hat{\mathbf{\Lambda}}_s^{-1} \hat{\mathbf{E}}_s^* \mathbf{B}\}. \tag{15}$$

Comparing (15) with (13) we see that *the FB-MODE criterion has exactly the same structure as the F-MODE criterion*, with the only difference that the former criterion depends on the eigenelements of the FB sample covariance matrix. Owing to this neat result, the FB-MODE estimate of $\boldsymbol{\omega}$ can be obtained by applying to $\hat{R}$ the two- (or three-) step algorithm outlined in Section 3.

5 ANALYSIS OF FB-MODE

The asymptotic statistical properties of the FB-MODE estimates can be derived in a rather standard way by using a Taylor expansion technique. In [6] it is shown that, asymptotically (as N increases), the FB-MODE estimate is Gaussian distributed with mean equal to the true parameter vector $\boldsymbol{\omega}$ and covariance matrix,

$$\mathbf{C}_{\mathrm{FB}} = \frac{\sigma^2}{2N}\left\{\mathrm{Re}\left[\mathbf{U}\odot\mathbf{Q}^T\right]\right\}^{-1}, \qquad (16)$$
$$\mathbf{Q} = \mathbf{P}\mathbf{A}^*\mathbf{R}^{-1}\mathbf{A}\mathbf{P}.$$

Comparing (16) and (14) we see that the only difference between $\mathbf{C}_{\mathrm{FB}}$ and $\mathbf{C}_{\mathrm{F}}$ is that $\mathbf{Q}_0$ in $\mathbf{C}_{\mathrm{F}}$ is replaced by $\mathbf{Q}$ in $\mathbf{C}_{\mathrm{FB}}$. If $\mathbf{P}_0$ *is diagonal* then $\mathbf{Q} = \mathbf{Q}_0$ and hence $\mathbf{C}_{\mathrm{FB}} = \mathbf{C}_{\mathrm{F}}$. However, if $\mathbf{P}_0$ *is non-diagonal* (i.e., the signals are correlated) then, in general, $\mathbf{Q} \neq \mathbf{Q}_0$. By the CRB inequality, we must have ($\mathbf{A} \geq \mathbf{B}$ below means that the matrix $\mathbf{A}-\mathbf{B}$ is positive semi-definite)

$$\mathbf{C}_{\mathrm{FB}} \geq \mathbf{C}_{\mathrm{F}}. \qquad (17)$$

As already indicated above, the inequality in (17) is usually "strict" in the sense that $\mathbf{C}_{\mathrm{FB}} \neq \mathbf{C}_{\mathrm{F}}$. Every time this happens, *the FB-MODE is asymptotically statistically less efficient than the F-MODE.* To study the difference $(\mathbf{C}_{\mathrm{FB}} - \mathbf{C}_{\mathrm{F}})$ quantitatively, as well as the extent to which the asymptotic results derived above hold in samples of practical lengths, we resort to numerical simulations (see the next section).

6 NUMERICAL EXAMPLES

Consider a ULA consisting of four omni-directional and identical sensors separated by half of the carrier's wavelength. The two signals impinge on the array from $\theta_1 = -7.5°$ and $\theta_2 = 7.5°$ relative to broadside. These directions correspond to the spatial (angular) frequencies: $\omega_i = \pi\sin(\theta_i)$ $(i = 1, 2)$. The signal covariance is

$$\mathbf{P}_0 = 10^{\mathrm{SNR}/10}\begin{bmatrix} 1 & 0.99e^{i\pi/4} \\ 0.99e^{-i\pi/4} & 1 \end{bmatrix},$$

where SNR is expressed in decibels (dB).

In the first example, SNR=0 dB. The mean-square errors (MSEs) for F-MODE and FB-MODE are compared for different sample lengths in Figure 1. (Only the MSE values for θ_1 are shown; the MSE plot corresponding to θ_2 is similar.) The sample MSEs are based on 1000 independent trials. The MSE predicted by the large sample analysis is also depicted in Figure 1. It can be seen that the theoretical and simulation results are similar even for quite small sample lengths. It is clearly not useful to use FB-MODE in this case, not even for small samples.

In the second example, the number of samples is fixed to $N = 100$ and the SNR is varied. All other parameters are as before. The MSEs (computed from 1000 trials) are shown in Figure 2. We see that F-MODE outperforms FB-MODE for all SNR values.

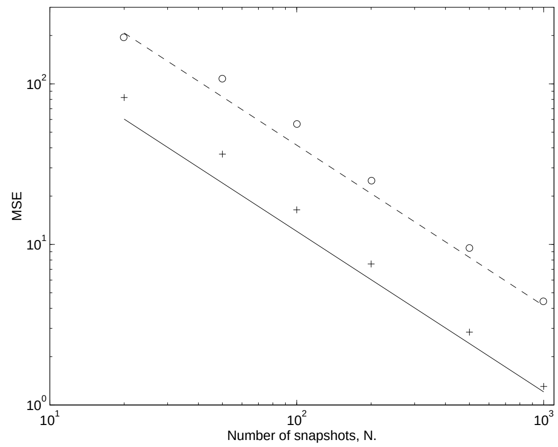

Figure 1: Mean square errors for θ_1 (in degrees^2) for F-MODE (+) and FB-MODE (o) versus the number of snapshots N. The solid line is the theoretical (asymptotic) MSE for F-MODE and the dashed line is the theoretical MSE for FB-MODE.

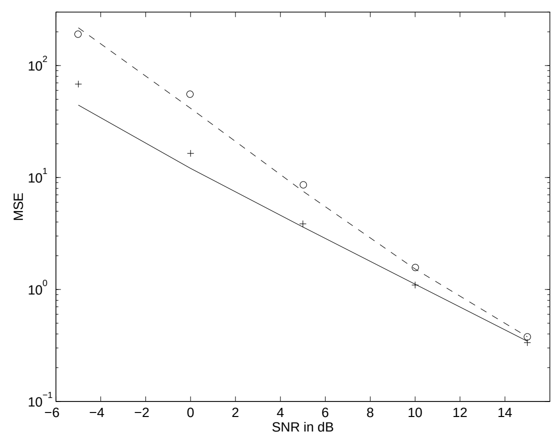

Figure 2: Mean square errors for θ_1 (in degrees^2) for F-MODE (+) and FB-MODE (o) versus the SNR. The solid line is the theoretical (asymptotic) MSE for F-MODE and the dashed line is the theoretical MSE for FB-MODE.

7 REFERENCES

[1] M. Viberg and B. Ottersten. "Sensor Array Processing Based on Subspace Fitting". *IEEE Trans. on Signal Processing*, SP-39(5):1110–1121, May 1991.

[2] P. Stoica and K. Sharman. "Maximum Likelihood Methods for Direction-of-Arrival Estimation". *IEEE Trans. ASSP*, ASSP-38:1132–1143, July 1990.

[3] B. D. Rao and K. V. S. Hari. "Weighted Subspace Methods and Spatial Smoothing: Analysis and Comparison". *IEEE Trans. ASSP*, 41(2):788–803, Feb. 1993.

[4] B. H. Kwon and S. U. Pillai. "A Self Inversive Array Processing Scheme for Angle-of-Arrival Estimation". *Signal Processing*, 17(3):259–277, Jul. 1989.

[5] M. Haardt. "Structured Least Squares to Improve the Performance of ESPRIT-Type Algorithms". *IEEE Trans. Signal Processing*, 45(3):792–799, Mar. 1997.

[6] P. Stoica and M. Jansson. "On Forward-Backward MODE for Array Signal Processing". *Digital Signal Processing – A Review Journal*, 7(4):239–252, Oct. 1997.