[36] B. Ottersten and L. Ljung, "Asymptotic Results for Sensor Array Processing", In *Proc. ICASSP 89*, pages 2266–2269, Glasgow, Scotland, May 1989.

[37] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis,* 2$^{\text{nd}}$ *edition*, John Wiley & Sons, New York, 1984.

[23] M. Wax and I. Ziskind, "On Unique Localization of Multiple Sources by Passive Sensor Arrays", *IEEE Trans. on ASSP*, **ASSP-37**(7):996–1000, July 1989.

[24] Y. Bresler, "Maximum Likelihood Estimation of Linearly Structured Covariance with Application to Antenna Array Processing", In *Proc. 4th ASSP Workshop on Spectrum Estimation and Modeling*, pages 172–175, Minneapolis, MN, Aug. 1988.

[25] A. G. Jaffer, "Maximum Likelihood Direction Finding of Stochastic Sources: A Separable Solution", In *Proc. ICASSP 88*, volume 5, pages 2893–2896, New York, New York, April 1988.

[26] T. W. Anderson, Multivariate linear relations, In *Proc. Second Int. Tampere Conf. in Stat.*, pages 9–36, Tampere, Finland, 1987.

[27] E.L. Lehmann, *Theory of Point Estimation*, John Wiley & Sons, New York, 1983.

[28] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[29] K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979.

[30] P. Billingsley, *Probability and Measure*, John Wiley & Sons, New York, 1985.

[31] D.R. Brillinger, *Time series: Data Analysis and Theory*, Expanded Edition, Holden-Day, 1980.

[32] W.F. Stout, *Almost Sure Convergence*, Academic Press, New York, 1974.

[33] R. P. Gupta, "Asymptotic Theory for Principal Component Analysis in the Complex Case", *Journal of the Indian Stat. Assoc.*, **Vol. 3**:97–106, 1965.

[34] A.W. Davis, "Asymptotic Theory for Principal Component Analysis: Non-Normal Case", *Austral. J. Statist.*, **19**(3):206–212, 1977.

[35] J.A. Cadzow, "A High Resolution Direction-of-Arrival Algorithm for Narrow-Band Coherent and Incoherent Sources", *IEEE Trans. ASSP*, **ASSP-36**:965–979, July 1988.

[12] R. Roy and T. Kailath, "ESPRIT – Estimation of Signal Parameters via Rotational Invariance Techniques", *IEEE Trans. on ASSP*, **ASSP-37**(7):984–995, July 1989.

[13] M. Viberg and B. Ottersten, "Sensor Array Processing Based on Subspace Fitting", *IEEE Trans. ASSP*, **ASSP-39**, May 1991.

[14] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood and Cramér-Rao Bound: Further Results and Comparisons", *IEEE Trans. ASSP*, **ASSP-38**:2140–2150, December 1990.

[15] M. Kaveh and A. J. Barabell, "The Statistical Performance of the MUSIC and the Minimum-Norm Algorithms in Resolving Plane Waves in Noise", *IEEE Trans. ASSP*, **ASSP-34**:331–341, April 1986.

[16] B. Porat and B. Friedlander, "Analysis of the Asymptotic Relative Efficiency of the MUSIC Algorithm", *IEEE Trans. ASSP*, **ASSP-36**:532–544, April 1988.

[17] T.W. Anderson, "Asymptotic Theory for Principal Component Analysis", *Ann. Math. Statist.*, **34**:122–148, 1963.

[18] P. Stoica and A. Nehorai, "Performance Study of Conditional and Unconditional Direction-of-Arrival Estimation", *IEEE Trans. ASSP*, **ASSP-38**:1783–1795, October 1990.

[19] H.L. Van Trees, *Detection, Estimation and Modulation Theory*, volume I, Wiley, New York, 1968.

[20] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria", *IEEE Trans. on ASSP*, **ASSP-33**(2):387–392, April 1985.

[21] M. Wax and I. Ziskind, "Detection of the Number of Coherent Signals by the MDL Principle", *IEEE Trans. on ASSP*, **ASSP-37**(8):1190–1196, Aug. 1989.

[22] M. Viberg, B. Ottersten, and T. Kailath, "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting", *IEEE Trans. ASSP*, **ASSP-39**, Nov. 1991.

References

[1] R. O. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*, PhD thesis, Stanford Univ., Stanford, CA, Nov. 1981.

[2] G. Bienvenu and L. Kopp, "Adaptivity to Background Noise Spatial Coherence for High Resolution Passive Methods", In *Proc. IEEE ICASSP*, pages 307–310, Denver, CO., 1980.

[3] M. Wax, *Detection and Estimation of Superimposed Signals*, PhD thesis, Stanford Univ., Stanford, CA, March 1985.

[4] W. J. Bangs, *Array Processing with Generalized Beamformers*, PhD thesis, Yale University, New Haven, CT, 1971.

[5] M. Wax and T. Kailath, "Optimal Localization of Multiple Sources by Passive Arrays", *IEEE Trans. on ASSP*, **ASSP-31**:1210–1218, Oct. 1983.

[6] J.F. Böhme, "Estimation of Source Parameters By Maximum Likelihood and Nonlinear Regression", In *Proc. ICASSP 84*, pages 7.3.1–7.3.4, 1984.

[7] J.F. Böhme, "Estimation of Spectral Parameters of Correlated Signals in Wavefields", *Signal Processing*, **10**:329–337, 1986.

[8] M. Feder and E. Weinstein, "Parameter Estimation of Superimposed Signals Using the EM Algorithm", *IEEE Trans. ASSP*, **ASSP-36**:477–489, April 1988.

[9] B. Ottersten and M. Viberg, "Asymptotic Results for Multidimensional Sensor Array Processing", In *Proc.* 22^{nd} *Asilomar Conf. Sig., Syst., Comput.*, pages 833–837, Monterey, CA, November 1988.

[10] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood and Cramér-Rao Bound", *IEEE Trans. ASSP*, **ASSP-37**:720–741, May 1989.

[11] B. Ottersten and M. Viberg, "Analysis of Subspace Fitting Based Methods for Sensor Array Processing", In *Proc. ICASSP 89*, pages 2807–2810, Glasgow, Scotland, May 1989.

Appendix C Proof of Theorem 6

Following the ideas in [17, 34, 33], we introduce a transformation of the sample covariance

$$\mathbf{T} = \mathbf{E}_N^* \hat{\mathbf{R}} \mathbf{E}_N \ , \tag{C.1}$$

where $\mathbf{E}_N$ is defined in (38). Observe that $\mathrm{E}[\mathbf{T}] = \mathbf{\Lambda}_N$. The second-order ergodicity of $\mathbf{s}(t)$ guarantees that the central limit theorem can be applied, see for instance [28], Lemma 9A.1. Hence, $\sqrt{N}(\mathbf{T} - \mathbf{\Lambda}_N)$ has a limiting zero-mean Gaussian distribution. The perturbation arguments in, e.g., the proof of Theorem 13.5.1. of [37], show that for λ_k being a simple eigenvalue, the i^{th} component of $\tilde{\mathbf{e}}_k$ is related to $\mathbf{T}$ by

$$\tilde{E}_{ik} = \sum_{p=1, p \neq k}^{m} \frac{E_{ip} T_{pk}}{\lambda_k - \lambda_p} + o_p(1/\sqrt{N}) \ , \tag{C.2}$$

where E_{ip} and T_{pk} denote the elements of $\mathbf{E}$ and $\mathbf{T}$ respectively. It follows that $\sqrt{N}\tilde{\mathbf{e}}$ also has a limiting zero-mean Gaussian distribution. The covariance of the asymptotic distribution is obtained from

$$N\mathrm{E}[\tilde{E}_{ik}\tilde{E}_{jl}^*] = \sum_{p=1, p\neq k}^{m} \sum_{q=1, q \neq l}^{m} \frac{1}{(\lambda_k - \lambda_p)(\lambda_l - \lambda_q)} E_{ip} E_{jq}^* \times \lim_{N \to \infty} N\mathrm{E}[T_{pk} T_{ql}^*] \ . \tag{C.3}$$

Denote $\tilde{\mathbf{R}} = \hat{\mathbf{R}} - \mathbf{R}_N$ and note that for $p \neq k$, $\{\mathbf{E}_N^* \hat{\mathbf{R}} \mathbf{E}_N\}_{pk} = \{\mathbf{E}_N^* \tilde{\mathbf{R}} \mathbf{E}_N\}_{pk}$. Thus,

$$\begin{aligned}
\lim_{N\to\infty} N\mathrm{E}[T_{pk}T_{ql}^*] &= \lim_{N\to\infty} N\mathrm{E}\left[\{\mathbf{E}_N^*\hat{\mathbf{R}}\mathbf{E}_N\}_{pk}\{\mathbf{E}_N^*\hat{\mathbf{R}}\mathbf{E}_N\}_{ql}^*\right] && \text{(C.4)}\\
&= \lim_{N\to\infty} N\mathrm{E}\left[\{\mathbf{E}^*\tilde{\mathbf{R}}\mathbf{E}\}_{pk}\{\mathbf{E}^*\tilde{\mathbf{R}}\mathbf{E}\}_{lq}\right] && \text{(C.5)}\\
&= \sum_{ijrs} E_{ip}^* E_{jk} E_{rl}^* E_{sq} \times \lim_{N\to\infty} N\mathrm{E}\left[\tilde{R}_{ij}\tilde{R}_{rs}\right] \ . && \text{(C.6)}
\end{aligned}$$

It is straightforward to show that

$$\lim_{N\to\infty} N\mathrm{E}[\tilde{R}_{ij}\tilde{R}_{rs}] = R_{rj}R_{is} - \{\mathbf{ASA}^*\}_{rj}\{\mathbf{ASA}^*\}_{is} \ . \tag{C.7}$$

Inserting this in (C.6) leads to

$$\begin{aligned}
\lim_{N\to\infty} N\mathrm{E}[T_{pk}T_{ql}^*] &= \sum_{ijrs} E_{rl}^* R_{rj} E_{jk} E_{ip}^* R_{is} E_{sq} - E_{rl}^*\{\mathbf{ASA}^*\}_{rj} E_{jk} E_{ip}^* \{\mathbf{ASA}^*\}_{is} E_{sq} && \text{(C.8)}\\
&= \mathbf{e}_l^* \mathbf{R} \mathbf{e}_k \mathbf{e}_p^* \mathbf{R} \mathbf{e}_q - \mathbf{e}_l^* \mathbf{ASA}^* \mathbf{e}_k \mathbf{e}_p^* \mathbf{ASA}^* \mathbf{e}_q && \text{(C.9)}\\
&= \delta_{kl}\delta_{pq}(\lambda_k\lambda_p - \tilde{\lambda}_k\tilde{\lambda}_p) \ . && \text{(C.10)}
\end{aligned}$$

Finally, inserting (C.10) into (C.3) gives the ij^{th} component of (40). The proof of (41) is similar.

The third term in Equation (B.15) can be expressed as

$$\begin{aligned}
\mathrm{Tr}\{\mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_i}\mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_j}\} &= \mathrm{Tr}\{\mathbf{R}^{-1}\mathbf{P}_{\theta_i}\mathbf{ASA}^*\mathbf{R}^{-1}\mathbf{P}_{\theta_j}\mathbf{ASA}^* \\
&\quad +\mathbf{R}^{-1}\mathbf{P}_{\theta_i}\mathbf{ASA}^*\mathbf{R}^{-1}\mathbf{ASA}^*\mathbf{P}_{\theta_j} + (\ldots)^*\} \\
&= \mathrm{Tr}\{\sigma^{-2}\mathbf{P}^{\perp}\mathbf{A}_{\theta_i}\mathbf{A}^{\dagger}\mathbf{ASA}^*\mathbf{R}^{-1}\mathbf{ASA}^*\mathbf{A}^{\dagger *}\mathbf{A}^*_{\theta_j}\mathbf{P}^{\perp} + (\ldots)^*\} \\
&= \sigma^{-2}\mathrm{Tr}\{\mathbf{A}^*_{\theta_j}\mathbf{P}^{\perp}\mathbf{A}_{\theta_i}\mathbf{SA}^*\mathbf{R}^{-1}\mathbf{AS} + (\ldots)^*\}\ . \qquad \text{(B.24)}
\end{aligned}$$

Collect the Equations (B.22), (B.23), and (B.24) and insert them in (B.15) to obtain

$$\begin{aligned}
\bar{J}_{\theta_i\theta_j} &= \mathrm{Tr}\{\mathbf{A}^*_{\theta_j}\mathbf{P}^{\perp}\mathbf{A}_{\theta_i}(\sigma^{-2}\mathbf{S} - \sigma^{-2}\mathbf{S} + 2\sigma^{-4}\mathbf{SA}^*\mathbf{AY} - \sigma^{-2}\mathbf{SA}^*\mathbf{R}^{-1}\mathbf{AS}) + (\ldots)^*\} \\
&= \mathrm{Tr}\{\mathbf{A}^*_{\theta_j}\mathbf{P}^{\perp}\mathbf{A}_{\theta_i}(2\sigma^{-4}\mathbf{SA}^*\mathbf{AY} - \sigma^{-2}\mathbf{SA}^*\mathbf{R}^{-1}\mathbf{AS}) + (\ldots)^*\}\ . \qquad \text{(B.25)}
\end{aligned}$$

Note that

$$\begin{aligned}
\mathbf{Y} &= \mathbf{S}(\mathbf{I} + \sigma^{-2}\mathbf{A}^*\mathbf{AS})^{-1} = \mathbf{S}(\mathbf{I} - \mathbf{A}^*(\mathbf{ASA}^* + \sigma^2\mathbf{I})^{-1}\mathbf{AS}) \\
&= \mathbf{S} - \mathbf{SA}^*\mathbf{R}^{-1}\mathbf{AS}\ . \qquad \text{(B.26)}
\end{aligned}$$

This gives

$$\begin{aligned}
\sigma^{-4}\mathbf{SA}^*\mathbf{AY} &= \sigma^{-4}\mathbf{SA}^*(\mathbf{I} - \mathbf{ASA}^*\mathbf{R}^{-1})\mathbf{AS} \\
&= \sigma^{-4}\mathbf{SA}^*(\mathbf{R} - \mathbf{ASA}^*)\mathbf{R}^{-1}\mathbf{AS} \\
&= \sigma^{-2}\mathbf{SA}^*\mathbf{R}^{-1}\mathbf{AS}\ . \qquad \text{(B.27)}
\end{aligned}$$

Use this in (B.25)

$$\begin{aligned}
\bar{J}_{\theta_i\theta_j} &= \sigma^{-2}\mathrm{Tr}\{\mathbf{A}^*_{\theta_j}\mathbf{P}^{\perp}\mathbf{A}_{\theta_i}\mathbf{SA}^*\mathbf{R}^{-1}\mathbf{AS} + (\ldots)^*\} \\
&= 2\sigma^{-2}\mathrm{Re}\left[\mathrm{Tr}\{\mathbf{A}^*_{\theta_j}\mathbf{P}^{\perp}\mathbf{A}_{\theta_i}\mathbf{SA}^*\mathbf{R}^{-1}\mathbf{AS}\}\right]\ , \qquad \text{(B.28)}
\end{aligned}$$

The above can be put in matrix form by the relation

$$\mathbf{A}_{\theta_i} = [0, \ldots, 0, \mathbf{d}(\theta_i), 0, \ldots, 0]\ . \qquad \text{(B.29)}$$

This leads to (17), and the proof of Theorem 1 is complete.

Thus, we have

$$\bar{J}_{\theta_i\theta_j} = -\sigma^{-2}\mathrm{Tr}\{\mathbf{P}_{\theta_i\theta_j}\mathbf{R}\} + \mathrm{Tr}\{\mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_i\theta_j} - \mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_i}\mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_j}\} \tag{B.15}$$

$$\bar{J}_{\theta_i\sigma^2} = \sigma^{-4}\mathrm{Tr}\{\mathbf{P}_{\theta_i}\mathbf{R}\} + \mathrm{Tr}\{\mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_i\sigma^2} - \mathbf{R}^{-1}\bar{\mathbf{X}}_{\sigma^2}\mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_i}\}\ . \tag{B.16}$$

Introduce the notation

$$\mathbf{Y} = \mathbf{S}(\mathbf{I} + \sigma^{-2}\mathbf{A}^*\mathbf{A}\mathbf{S})^{-1}\ , \tag{B.17}$$

and examine the terms in the expression for $\bar{J}_{\theta_i\sigma^2}$

$$\begin{aligned}\mathrm{Tr}\{\mathbf{P}_{\theta_i}\mathbf{R}\} &= \mathrm{Tr}\{(\mathbf{P}^\perp\mathbf{A}_{\theta_i}\mathbf{A}^\dagger + (\ldots)^*)(\mathbf{A}\mathbf{S}\mathbf{A}^* + \sigma^2\mathbf{I})\}\\ &= \sigma^2\mathrm{Tr}\{\mathbf{P}_{\theta_i}\} = 0 \end{aligned}\tag{B.18}$$

$$\begin{aligned}\mathrm{Tr}\{\mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_i\sigma^2}\} &= -\mathrm{Tr}\{\mathbf{R}^{-1}\mathbf{P}_{\theta_i}\} = -\mathrm{Tr}\{(\sigma^{-2}\mathbf{I} - \sigma^{-4}\mathbf{A}\mathbf{Y}\mathbf{A}^*)\mathbf{P}_{\theta_i}\}\\ &= -\sigma^{-2}\mathrm{Tr}\{\mathbf{P}_{\theta_i}\} = 0\end{aligned}\tag{B.19}$$

$$\begin{aligned}\mathrm{Tr}\{\mathbf{R}^{-1}\bar{\mathbf{X}}_{\sigma^2}\mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_i}\} &= \mathrm{Tr}\{\mathbf{R}^{-1}\mathbf{P}^\perp\mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_i}\} = \sigma^{-4}\mathrm{Tr}\{\mathbf{P}^\perp\bar{\mathbf{X}}_{\theta_i}\}\\ &= \sigma^{-4}\mathrm{Tr}\{\mathbf{P}^\perp\mathbf{P}_{\theta_i}\mathbf{A}\mathbf{S}\mathbf{A}^* + (\ldots)^*\} = 0\ .\end{aligned}\tag{B.20}$$

Hence, $\bar{J}_{\theta_i\sigma^2} = 0$, implying that $\boldsymbol{\theta}$ and σ^2 are uncorrelated (and thus independent) in the asymptotic distribution. Let us now show (18)

$$\begin{aligned}\bar{J}_{\sigma^2\sigma^2} &= 2\sigma^{-6}\mathrm{Tr}\{\mathbf{P}^\perp\mathbf{R}\} - \mathrm{Tr}\{\mathbf{R}^{-1}\mathbf{P}^\perp\mathbf{R}^{-1}\mathbf{P}^\perp\}\\ &= \frac{2(m-d)}{\sigma^4} - \sigma^{-4}\mathrm{Tr}\{\mathbf{P}^\perp\} = \frac{m-d}{\sigma^4}\ ,\end{aligned}\tag{B.21}$$

which by (B.1) yields (18). Consider next the evaluation of $\bar{J}_{\theta_i\theta_j}$. Examine the first term in Equation (B.15)

$$\begin{aligned}\sigma^{-2}\mathrm{Tr}\{\mathbf{P}_{\theta_i\theta_j}\mathbf{R}\} &= \sigma^{-2}\mathrm{Tr}\{\mathbf{P}_{\theta_i\theta_j}\mathbf{A}\mathbf{S}\mathbf{A}^*\} + \mathrm{Tr}\{\mathbf{P}_{\theta_i\theta_j}\}\\ &= -\sigma^{-2}\mathrm{Tr}\{\mathbf{A}^{\dagger *}\mathbf{A}^*_{\theta_j}\mathbf{P}^\perp\mathbf{A}_{\theta_i}\mathbf{A}^\dagger\mathbf{A}\mathbf{S}\mathbf{A}^* + (\ldots)^*\}\\ &= -\sigma^{-2}\mathrm{Tr}\{\mathbf{A}^*_{\theta_j}\mathbf{P}^\perp\mathbf{A}_{\theta_i}\mathbf{S} + (\ldots)^*\}\ .\end{aligned}\tag{B.22}$$

Examine the second term in (B.15)

$$\begin{aligned}\mathrm{Tr}\{\mathbf{R}^{-1}\bar{\mathbf{X}}_{\theta_i\theta_j}\} &= \mathrm{Tr}\{\mathbf{R}^{-1}(\mathbf{P}_{\theta_i\theta_j}\mathbf{A}\mathbf{S}\mathbf{A}^* + \mathbf{P}_{\theta_j}\mathbf{A}\mathbf{S}\mathbf{A}^*\mathbf{P}_{\theta_i}) + (\ldots)^*\}\\ &= \mathrm{Tr}\{(\sigma^{-2}\mathbf{I} - \sigma^{-4}\mathbf{A}\mathbf{Y}\mathbf{A}^*)(\mathbf{P}_{\theta_i\theta_j}\mathbf{A}\mathbf{S}\mathbf{A}^* + \mathbf{P}_{\theta_j}\mathbf{A}\mathbf{S}\mathbf{A}^*\mathbf{P}_{\theta_i}) + (\ldots)^*\}\\ &= \mathrm{Tr}\{-2\sigma^{-2}\mathbf{A}^{\dagger *}\mathbf{A}^*_{\theta_j}\mathbf{P}^\perp\mathbf{A}_{\theta_i}\mathbf{A}^\dagger\mathbf{A}\mathbf{S}\mathbf{A}^* + \sigma^{-2}\mathbf{P}^\perp\mathbf{A}_{\theta_i}\mathbf{A}^\dagger\mathbf{A}\mathbf{S}\mathbf{A}^*\mathbf{A}^{\dagger *}\mathbf{A}^*_{\theta_j}\mathbf{P}^\perp\\ &\quad + 2\sigma^{-4}\mathbf{A}\mathbf{Y}\mathbf{A}^*\mathbf{A}^{\dagger *}\mathbf{A}^*_{\theta_j}\mathbf{P}^\perp\mathbf{A}_{\theta_i}\mathbf{A}^\dagger\mathbf{A}\mathbf{S}\mathbf{A}^* + (\ldots)^*\}\\ &= \mathrm{Tr}\{\mathbf{A}^*_{\theta_j}\mathbf{P}^\perp\mathbf{A}_{\theta_i}(-\sigma^{-2}\mathbf{S} + 2\sigma^{-4}\mathbf{S}\mathbf{A}^*\mathbf{A}\mathbf{Y}) + (\ldots)^*\}\ .\end{aligned}\tag{B.23}$$

Appendix B Evaluation of H

Herein, the matrix $\mathbf{H}$, defined in (21) is calculated. For ease of notation, overbar is used to indicate the limit as N tends to infinity with the parameters evaluated at the true parameter values. Moreover, we use $\mathbf{P}$ and $\mathbf{P}^{\perp}$ for the projection matrices $\mathbf{P}_{\mathbf{A}}$ and $\mathbf{P}^{\perp}_{\mathbf{A}}$, and their derivatives with respect to $\boldsymbol{\theta}_i$ are denoted $\mathbf{P}_{\theta_i}$ and $\mathbf{P}^{\perp}_{\theta_i}$ respectively.

It will be shown that $\bar{J}_{\boldsymbol{\theta}\sigma^2} = 0$. Thus,

$$\mathbf{H}^{-1} = \begin{bmatrix} \bar{J}_{\boldsymbol{\theta\theta}} & 0 \\ 0 & \bar{J}_{\sigma^2\sigma^2} \end{bmatrix}^{-1} = \begin{bmatrix} \bar{J}^{-1}_{\boldsymbol{\theta\theta}} & 0 \\ 0 & \bar{J}^{-1}_{\sigma^2\sigma^2} \end{bmatrix}, \tag{B.1}$$

where $\partial J/\partial\boldsymbol{\theta} = J_{\boldsymbol{\theta}}$. We start by evaluating some derivatives

$$\begin{aligned}
J_{\theta_i} &= -\sigma^{-2}\mathrm{Tr}\{\mathbf{P}_{\theta_i}\hat{\mathbf{R}}\} - \mathrm{Tr}\{(\mathbf{X}^{-1})_{\theta_i}\mathbf{X}\} \\
&= -\sigma^{-2}\mathrm{Tr}\{\mathbf{P}_{\theta_i}\hat{\mathbf{R}}\} + \mathrm{Tr}\{\mathbf{X}^{-1}\mathbf{X}_{\theta_i}\} && \text{(B.2)} \\
J_{\sigma^2} &= -\sigma^{-4}\mathrm{Tr}\{\mathbf{P}^{\perp}\hat{\mathbf{R}}\} + \mathrm{Tr}\{\mathbf{X}^{-1}\mathbf{X}_{\sigma^2}\} && \text{(B.3)} \\
J_{\theta_i\theta_j} &= -\sigma^{-2}\mathrm{Tr}\{\mathbf{P}_{\theta_i\theta_j}\hat{\mathbf{R}}\} + \mathrm{Tr}\{\mathbf{X}^{-1}\mathbf{X}_{\theta_i\theta_j} - \mathbf{X}^{-1}\mathbf{X}_{\theta_i}\mathbf{X}^{-1}\mathbf{X}_{\theta_j}\} && \text{(B.4)} \\
J_{\theta_i\sigma^2} &= \sigma^{-4}\mathrm{Tr}\{\mathbf{P}_{\theta_i}\hat{\mathbf{R}}\} + \mathrm{Tr}\{\mathbf{X}^{-1}\mathbf{X}_{\theta_i\sigma^2} - \mathbf{X}^{-1}\mathbf{X}_{\sigma^2}\mathbf{X}^{-1}\mathbf{X}_{\theta_i}\} && \text{(B.5)} \\
J_{\sigma^2\sigma^2} &= 2\sigma^{-6}\mathrm{Tr}\{\mathbf{P}^{\perp}\hat{\mathbf{R}}\} - \mathrm{Tr}\{\mathbf{X}^{-1}\mathbf{X}_{\sigma^2}\mathbf{X}^{-1}\mathbf{X}_{\sigma^2}\} && \text{(B.6)} \\
\mathbf{X}_{\theta_i} &= \mathbf{P}_{\theta_i}\hat{\mathbf{R}}\mathbf{P} + \mathbf{P}\hat{\mathbf{R}}\mathbf{P}_{\theta_i} + \sigma^2\mathbf{P}^{\perp}_{\theta_i} = \mathbf{P}_{\theta_i}\hat{\mathbf{R}}\mathbf{P} + \mathbf{P}\hat{\mathbf{R}}\mathbf{P}_{\theta_i} - \sigma^2\mathbf{P}_{\theta_i} && \text{(B.7)} \\
\mathbf{X}_{\theta_i\theta_j} &= \mathbf{P}_{\theta_i\theta_j}\hat{\mathbf{R}}\mathbf{P} + \mathbf{P}_{\theta_j}\hat{\mathbf{R}}\mathbf{P}_{\theta_i} + \mathbf{P}_{\theta_i}\hat{\mathbf{R}}\mathbf{P}_{\theta_j} + \mathbf{P}\hat{\mathbf{R}}\mathbf{P}_{\theta_i\theta_j} - \sigma^2\mathbf{P}_{\theta_i\theta_j} && \text{(B.8)} \\
\mathbf{X}_{\sigma^2} &= \mathbf{P}^{\perp} && \text{(B.9)} \\
\mathbf{X}_{\theta_i\sigma^2} &= -\mathbf{P}_{\theta_i}\ . && \text{(B.10)}
\end{aligned}$$

The derivatives of the projection matrix $\mathbf{P}$, are given by [13] Appendix B

$$\begin{aligned}
&\mathbf{P}_{\theta_i} = \mathbf{P}^{\perp}\mathbf{A}_{\theta_i}\mathbf{A}^{\dagger} + (\ldots)^*, \\
&\mathbf{P}_{\theta_i\theta_j} = -\mathbf{P}^{\perp}\mathbf{A}_{\theta_j}\mathbf{A}^{\dagger}\mathbf{A}_{\theta_i}\mathbf{A}^{\dagger} - \mathbf{A}^{\dagger *}\mathbf{A}^*_{\theta_j}\mathbf{P}^{\perp}\mathbf{A}_{\theta_i}\mathbf{A}^{\dagger} + \mathbf{P}^{\perp}\mathbf{A}_{\theta_i\theta_j}\mathbf{A}^{\dagger} + \\
&\qquad + \mathbf{P}^{\perp}\mathbf{A}_{\theta_i}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*_{\theta_j}\mathbf{P}^{\perp} - \mathbf{P}^{\perp}\mathbf{A}_{\theta_i}\mathbf{A}^{\dagger}\mathbf{A}_{\theta_j}\mathbf{A}^{\dagger} + (\ldots)^*\ ,
\end{aligned} \tag{B.11}$$

where $(\ldots)^*$ indicates that the Hermitian transpose of the same expression appears again.

Using the relations, $\mathbf{P}^{\perp}\mathbf{A} = \mathbf{0}$ and $\mathrm{Tr}\{\mathbf{AB}\} = \mathrm{Tr}\{\mathbf{BA}\}$, some calculations lead to

$$\begin{aligned}
\bar{\mathbf{X}} &= \mathbf{PRP} + \sigma^2\mathbf{P}^{\perp}\mathbf{ASA}^* + \sigma^2\mathbf{P} + \sigma^2\mathbf{P}^{\perp} = \mathbf{R} && \text{(B.12)} \\
\bar{\mathbf{X}}_{\theta_i} &= \mathbf{P}_{\theta_i}\mathbf{ASA}^* + (\ldots)^* && \text{(B.13)} \\
\bar{\mathbf{X}}_{\theta_i\theta_j} &= \mathbf{P}_{\theta_i\theta_j}\mathbf{ASA}^* + \mathbf{P}_{\theta_j}\mathbf{ASA}^*\mathbf{P}_{\theta_i} + (\ldots)^*\ . && \text{(B.14)}
\end{aligned}$$

Appendix A Separable Likelihood Functions

The following result is required for the proof of Theorem 1. Since the result is of interest in itself, it is formulated as a lemma.

Lemma 1 *Let $l(\boldsymbol{\eta}, \boldsymbol{\xi})$ be a regular log likelihood function so that the ML estimates of the parameter vectors $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are consistent. Assume that $l(\boldsymbol{\eta}, \boldsymbol{\xi})$ is separable in $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. Denote the concentrated likelihood function $J(\boldsymbol{\xi}) = l(\hat{\boldsymbol{\eta}}(\boldsymbol{\xi}), \boldsymbol{\xi})$, where $\hat{\boldsymbol{\eta}}(\boldsymbol{\xi})$ is such that*

$$\left.\frac{\partial}{\partial \boldsymbol{\eta}} l(\boldsymbol{\eta}, \boldsymbol{\xi})\right|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}(\boldsymbol{\xi})} = 0 \ . \tag{A.1}$$

Denote the partial derivatives of J and l according to

$$J_i = \frac{\partial}{\partial \boldsymbol{\xi}_i} J(\boldsymbol{\xi}), \quad l_i = \frac{\partial}{\partial \boldsymbol{\xi}_i} l(\boldsymbol{\eta}, \boldsymbol{\xi}) \tag{A.2}$$

Then, the following identities hold

$$J_i = l_i|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}(\boldsymbol{\xi})} \tag{A.3}$$

$$\mathrm{E}[J_i J_j] = -\mathrm{E}[J_{ij}] \ . \tag{A.4}$$

Proof The chain rule implies

$$J_i = l_i|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}(\boldsymbol{\xi})} + \left.\frac{\partial}{\partial \boldsymbol{\eta}} l(\boldsymbol{\eta}, \boldsymbol{\xi})^T\right|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}(\boldsymbol{\xi})} \frac{\partial}{\partial \boldsymbol{\xi}_i} \hat{\boldsymbol{\eta}}(\boldsymbol{\xi}) \ . \tag{A.5}$$

By (A.1), the second term vanishes and (A.3) follows. Apply (A.3) to the left side of (A.4)

$$\mathrm{E}[J_i J_j] = \mathrm{E}[l_i J_j] \ . \tag{A.6}$$

Since l_i is the derivative of a regular log likelihood function, Theorem 4.1.1 of [29] can be applied to give

$$\mathrm{E}[l_i J_j] = \frac{\partial}{\partial \boldsymbol{\xi}_i} \mathrm{E}[J_j] - \mathrm{E}\left[\frac{\partial J_j}{\partial \boldsymbol{\xi}_i}\right] = -\mathrm{E}[J_{ij}] \ . \tag{A.7}$$

The last equality follows since $\mathrm{E}[J_j] = \mathrm{E}[l_j] = 0$. Clearly (A.6-A.7) implies (A.4) and the lemma is proved. $\square$

7 Conclusions

The asymptotic (for large amounts of data) properties of the ML and WSF estimation techniques are studied. It is shown that the likelihood function based on stochastic (Gaussian) modeling of the emitter signals satisfies the regularity conditions, ensuring an efficient ML estimate. The asymptotic distribution of the WSF method is compared to that of ML, and they are found to be the same. Hence, also WSF is asymptotically efficient and achieves the CRB for Gaussian signals. As a by-product, a compact matrix expression for the CRB on the signal parameters is obtained. The effects of non-Gaussian emitter signals on the statistical properties of the WSF and ML estimates are examined. Both methods are found to be asymptotically robust, i.e., the distribution derived under the Gaussian assumption is valid for general second-order ergodic emitter signals. This result has implications on the modeling aspects of the sensor array processing problem. The deterministic and stochastic modeling techniques are discussed and it is argued that the stochastic model is appropriate for the problem under consideration.

The corollary above is a more general version of a similar theorem from [36]. The first inequality provides strong justification for the stochastic model being appropriate for the sensor array problem. The second inequality says that the case of Gaussian signal waveforms can be interpreted as a *worst case*. If the signal waveforms have a known non-Gaussian distribution, it is usually possible to beat the stochastic ML and WSF methods. This would require the maximization of the appropriate likelihood function, which can be very complicated. Notice also that the resulting method cannot be asymptotically robust. Indeed, we have

Corollary 2 $\mathbf{C}_{\text{STO}} = \text{CRB}_{\text{STO}}$ *is a lower bound for the covariance of* $\sqrt{N}\tilde{\boldsymbol{\theta}}$ *for any (asymptotically) unbiased estimator which is also asymptotically robust.*

Proof Immediate. □

It should be noted that the results here refer to asymptotic in *the number of snapshots*, N. Although there is a strong connection between DOA estimation in sensor arrays and cisoid estimation in time series, the asymptotic statements are different. The dual statement to having a large number of data in time series, is that *the number of sensors*, m, is large. For large m, the deterministic ML method is efficient and the deterministic model is appropriate. Let us comment on the applicability of the results of this paper to more general array parametrizations. The array response matrix, $\mathbf{A}(\boldsymbol{\theta})$ is assumed to be parametrized by the DOAs only. However, the analysis here is not restricted to this case, except for the explicit construction of the matrix form of the covariance matrix (17). All results are therefore valid for a general parametrization, and can include bearing, elevation, range, unknown array response, etc., provided that all parameters are consistently estimated. The expression (B.28), see Appendix B, for the ij^{th} element of the asymptotic covariance and the CRB, is valid for a general parametrization of $\mathbf{A}$. The matrix form of the covariance depends, of course, upon which parametrization is used.

Although the stochastic ML and WSF techniques have identical asymptotic properties, the finite sample behavior may be different. There are other aspects which also need to be considered when comparing the methods, for example computational complexity and global minimization properties. These issues require further study.

6 Discussion

The asymptotic efficiency and robustness properties of the stochastic ML and WSF methods have some interesting implications on the modeling aspects of the sensor array problem. There are two competing models for the sensor array problem, namely *deterministic* versus *stochastic* modeling of the emitter signals. These lead to different ML criteria and CRBs. If the emitter signals are Gaussian, it is clear that the stochastic ML method is optimal as shown by Theorem 3. If, on the other hand, the signals have an unknown non-Gaussian distribution or are modeled as deterministic, the relation between the deterministic ML criterion and the Gaussian based ML method is unclear. It is also of interest to compare the CRB for the Gaussian case with the corresponding bounds for other signal distributions. The following corollary to Theorems 3, 5, and 7 establishes these relations.

Corollary 1 *Let* $\mathbf{C}_{\mathrm{STO}}$ *and* $\mathbf{C}_{\mathrm{DET}}$ *denote the asymptotic covariances of* $\sqrt{N}\hat{\boldsymbol{\theta}}$, *for the stochastic and deterministic ML estimates respectively. Furthermore, let* $\mathrm{CRB}_{\mathrm{STO}}$ *and* $\mathrm{CRB}_{\mathrm{ARB}}$ *be the stochastic CRB and the CRB for any other signal distribution or deterministic signals respectively. The following inequalities then hold*

$$\mathbf{C}_{\mathrm{DET}} \geq \mathbf{C}_{\mathrm{STO}} = \mathrm{CRB}_{\mathrm{STO}} \geq \mathrm{CRB}_{\mathrm{ARB}} \ . \tag{43}$$

Proof Consider the first inequality. Apply the deterministic ML method to Gaussian signals. The deterministic ML estimator provides asymptotically unbiased parameter estimates[6]. The CRB inequality then implies that the first inequality in (43) holds (see also (34)). This, together with the asymptotic robustness of the deterministic and stochastic ML methods implies that the first inequality holds for arbitrary signals.

To prove the second inequality, apply the stochastic ML method to non-Gaussian signals. The ML estimates are asymptotically unbiased and the asymptotic covariance of the stochastic ML estimates still equals $\mathbf{C}_{\mathrm{STO}} = \mathrm{CRB}_{\mathrm{STO}}$. Hence, also this result follows from the Cramér-Rao inequality. $\square$

[6] The signal waveform estimates are not consistent: $\hat{\mathbf{s}}(t) = \mathbf{A}^{\dagger}(\hat{\boldsymbol{\theta}})\mathbf{x}(t) \rightarrow \mathbf{s}(t) + \mathbf{A}^{\dagger}(\boldsymbol{\theta}_0)\mathbf{n}(t)$. However, it can be shown that they are indeed asymptotically unbiased.

the second-order moments of $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}_0)\hat{\mathbf{e}}_k$ is zero. Hence, the limiting distribution of these variables is the same under the deterministic and the Gaussian signal assumptions. $\square$

We remark that Theorem 7 applies to the subspace fitting criterion (15) for any choice of weighting matrix $\mathbf{W}$ and also for an arbitrary (identifiable) parametrization of $\mathbf{A}$. Hence, one can conclude that the deterministic ML method, MD-MUSIC (Cadzow's method, [35]), and TLS-ESPRIT are all asymptotically robust estimation techniques. See further [13] for details on how these methods relate to the subspace fitting criterion.

However, asymptotic robustness is only guaranteed for *multidimensional* subspace fitting methods. The qualification *multidimensional* here refers to the fact that $\Re(\mathbf{E}_s) \subseteq \Re(\mathbf{A}(\boldsymbol{\theta}_0))$. This is not true for the MUSIC algorithm, which, in the subspace fitting formulation fits a one dimensional subspace, $\mathbf{a}(\theta)$ to $\mathbf{E}_s$.

as $\tilde{\mathbf{e}}_k = \hat{\mathbf{e}}_k - \mathbf{e}_{N,k}$*. Then* $\sqrt{N}\tilde{\mathbf{e}}_k$ *has a limiting Gaussian distribution with first and second-order moments given by*

$$\mathrm{E}[\tilde{\mathbf{e}}_k] = 0 + O(N^{-1}) \tag{39}$$

$$\mathrm{E}\left[\tilde{\mathbf{e}}_k\tilde{\mathbf{e}}_l^*\right] = \frac{\delta_{kl}}{N}\sum_{i=1,i\neq k}^{m}\frac{\lambda_k\lambda_i - \tilde{\lambda}_k\tilde{\lambda}_i}{(\lambda_k-\lambda_i)^2}\mathbf{e}_i\mathbf{e}_i^* \tag{40}$$

$$\mathrm{E}\left[\tilde{\mathbf{e}}_k\tilde{\mathbf{e}}_l^T\right] = -(1-\delta_{kl})\frac{\lambda_k\lambda_l - \tilde{\lambda}_k\tilde{\lambda}_l}{N(\lambda_k-\lambda_l)^2}\mathbf{e}_l\mathbf{e}_k^T\ , \tag{41}$$

where $\tilde{\lambda}_k$ *is the* k^{th} *largest eigenvalue of* $\mathbf{ASA}^*$*, i.e.,* $\tilde{\lambda}_k = \lambda_k - \sigma^2$*.*

Proof The proof is given in Appendix C. □

Remark 1. It is interesting to compare (39-41) with the corresponding quantities in the case of Gaussian signals, see e.g. [15]. The deterministic signal assumption introduces a bias of order $O(1/\sqrt{N})$, whereas the second-order moments differ from the Gaussian case only by the extra term involving $\tilde{\lambda}_k$.

Notice also that the proof of Theorem 6 is valid for a general "deterministic signal in additive Gaussian noise" model, and does not depend on the assumption of spatially white noise. □

The main result of this section follows easily from the above:

Theorem 7 *Let the assumptions of Theorem 6 hold. Then the WSF estimation error has the same asymptotic properties as in the case of Gaussian signal waveforms, i.e.*

$$\sqrt{N}\tilde{\boldsymbol{\theta}} \in \mathrm{AsN}(0, \mathbf{C}), \tag{42}$$

where $\mathbf{C}$ *is given by (26).*

Proof It is clear from the proof of Theorem 2 of [13] that the asymptotic distribution of the estimation error only depends on the distribution of the variables $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}_0)\hat{\mathbf{e}}_k,\ k = 1, \ldots, d'$. It is easy to verify that $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}_0)\mathbf{e}_{N,k} = 0,\ k = 1, \ldots, d'$. Thus, (39) shows that $\mathrm{E}[\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}_0)\hat{\mathbf{e}}_k] = 0,\ k = 1, \ldots, d'$ as in the Gaussian case. Moreover, since $\tilde{\lambda}_k = 0,\ k = d'+1, \ldots, m$, the equations (40-39) show that the non-Gaussian contribution to

The following result demonstrates the asymptotic robustness of the stochastic ML estimator.

Theorem 5 *Let* $\hat{\boldsymbol{\theta}}$ *and* $\hat{\sigma}^2$ *be the minimizing arguments of (13). Then,* $\hat{\boldsymbol{\theta}}$ *and* $\hat{\sigma}^2$ *have the same asymptotic distribution for* $\mathbf{s}(t)$ *being a fixed second-order sequence as in the case of Gaussian signals, i.e., the expressions (17-18) are still valid.*

Proof See [26], Theorem 3.5[5] . □

5.2 Subspace Fitting Techniques

The asymptotic distribution of subspace fitting estimates is based on the distribution of the signal eigenvectors of the sample covariance matrix. In the case of real Gaussian observations, the distribution was established in [17]. These results were extended to the complex case in [33]. The case of real non-Gaussian stochastic observations was studied by Davis [34]. Using the techniques developed in these references, the extension to our case where the observations consist of a deterministic signal in additive Gaussian noise is not difficult.

The expected value of the sample covariance is denoted $\mathbf{R}_N$,

$$\mathbf{R}_N = \mathrm{E}[\hat{\mathbf{R}}] = \mathbf{A}\mathbf{S}_N\mathbf{A}^* + \sigma^2\mathbf{I} \tag{36}$$

$$\mathbf{S}_N = \frac{1}{N}\sum_{t=1}^{N} \mathbf{s}(t)\mathbf{s}^*(t) \tag{37}$$

Introduce the eigendecomposition of $\mathbf{R}_N$

$$\mathbf{R}_N = \mathbf{E}_N\boldsymbol{\Lambda}_N\mathbf{E}_N^* \ , \tag{38}$$

where $\mathbf{E}_N = [\mathbf{e}_{N,1}, \ldots, \mathbf{e}_{N,m}]$. The following theorem establishes the distribution of the eigenvectors.

Theorem 6 *Let* $\mathbf{s}(t)$ *be a fixed sequence of bounded and second-order ergodic quantities and let* λ_k *be the* k^{th} *largest non-repeated eigenvalue of* $\mathbf{R}$*. Define the estimation error*

[5] In Anderson's notation, $\boldsymbol{\Lambda}_0(\boldsymbol{\lambda})$ corresponds to $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{z}_\alpha^{(0)}$ corresponds to $\mathbf{s}(t)$.

5 Asymptotic Robustness

The asymptotic results presented in the previous section, are based on the assumption that the signal waveforms, $\mathbf{s}(t)$ and the additive noise, $\mathbf{n}(t)$ have Gaussian distributions. This assumption can be justified, for example, by assuming that a Discrete Fourier transform is applied to the array output. Results from Brillinger [31] then state asymptotic normality and independence of the output of the individual frequency bins. However, in many applications, filters other than the DFT may be applied and a non-Gaussian signal distribution may result. The present section examines the asymptotic distribution of the estimation error for stochastic ML and subspace fitting techniques for a general signal distribution. The noise is, however, still assumed to be Gaussian. It is found that the asymptotic distribution derived for Gaussian signals is preserved in the non-Gaussian case. This interesting property is known in the statistical literature as *asymptotic robustness.*

In the analysis we will assume that $\{\mathbf{s}(t)\}$ is an arbitrary deterministic (i.e., fixed), second-order ergodic sequence of vectors. The signal "covariance" is still denoted $\mathbf{S}$, but the definition (4) is replaced by the following deterministic limit

$$\mathbf{S} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathbf{s}(t)\mathbf{s}^*(t) \ . \tag{35}$$

Let us first point out that the strong consistency of the estimates immediately generalizes to the non-Gaussian case. For all methods discussed, consistency depends only on the convergence of the sample covariance matrix and if the signals are second-order ergodic, $\hat{\mathbf{R}}$ can easily be shown to converge w.p.1 to $\mathbf{R}$ using e.g. Theorems 3.2.2. and 4.2.4. of [32]. Notice, though, that $\mathbf{R}$ is not the "true" array covariance in the deterministic signal case.

5.1 Stochastic ML

There are strong connections between the parameter estimation problem considered here and factor analysis. In the statistical literature, the stochastic model presented here is often referred to as a linear *structural* relationship. The properties of the resulting ML estimator have been widely studied, see [26] and references therein.

Since the matrix $\mathbf{S}(\mathbf{I} + \mathbf{A}^*\sigma^{-2}\mathbf{A}\mathbf{S})^{-1}$ is Hermitian (by the equality above) and positive semi-definite, it follows that $\mathbf{S}\mathbf{A}^*\mathbf{R}^{-1}\mathbf{A}\mathbf{S} \leq \mathbf{S}$. Hence, application of [14] Lemma A.2, yields

$$\frac{\sigma^2}{2}\left[\mathrm{Re}\left\{(\mathbf{D}^*\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}) \odot (\mathbf{S}\mathbf{A}^*\mathbf{R}^{-1}\mathbf{A}\mathbf{S})^T\right\}\right]^{-1} \geq \frac{\sigma^2}{2}\left[\mathrm{Re}\left\{(\mathbf{D}^*\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}) \odot \mathbf{S}^T\right\}\right]^{-1} . \qquad (34)$$

If the matrices $\mathbf{D}^*\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}$ and $\mathbf{S}$ are both positive definite, the inequality is strict showing that the stochastic bound is in this case strictly tighter that the deterministic bound.

Proof Notice first that the derivative of the log-likelihood function has expected value zero, see [1]. Using (27), it is easily verified that the remaining conditions of Theorem 4.1 of [27] are satisfied. □

The following gives the ij^{th} element of the inverse of the CRB, [1, 4]

$$\{CRB_{\mathrm{STO}}^{-1}\}_{ij} = \mathrm{E}\left[\left.\frac{\partial^2 l(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}_i \partial \boldsymbol{\eta}_j}\right|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} \right] = \mathrm{Tr}\{\mathbf{R}^{-1}\mathbf{R}_i\mathbf{R}^{-1}\mathbf{R}_j\}, \tag{29}$$

where $\mathbf{R}_i$ represents the derivative of $\mathbf{R}$ with respect to the i^{th} component of the parameter vector (11). Although the above formula can easily be numerically evaluated, it does not give much insight into the performance. The reason is the many nuisance parameters involved in the stochastic ML formulation.

By combining the results from Theorems 1 and 3, a compact matrix expression for the CRB on the DOA and noise variance estimates is obtained.

Theorem 4 *For Gaussian signal waveforms, the Cramér-Rao inequality can be expressed as*

$$N\mathrm{E}[\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T] \geq \frac{\sigma^2}{2}\left[\mathrm{Re}\left\{(\mathbf{D}^*\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}) \odot (\mathbf{S}\mathbf{A}^*\mathbf{R}^{-1}\mathbf{A}\mathbf{S})^T\right\}\right]^{-1} \tag{30}$$

$$N\mathrm{E}[(\tilde{\sigma}^2)^2] \geq \frac{\sigma^4}{m-d}\ . \tag{31}$$

Proof Theorems 1 and 3, give the asymptotic matrix expressions for the CRB. Equation (29) gives a bound on $N\mathrm{E}[\tilde{\boldsymbol{\eta}}\tilde{\boldsymbol{\eta}}^T]$ that is valid for all N. Since the right hand sides of (29-31) are independent of N, it follows that the expressions (30-31) are also valid for all N □

The compact matrix expression (30) for the CRB on the DOA estimates has not been available earlier. It is interesting to compare (30) with the corresponding CRB for deterministic signal waveforms. For large N, the deterministic inequality reads, [10]

$$N\mathrm{E}[\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T] \geq \frac{\sigma^2}{2}\left[\mathrm{Re}\left\{(\mathbf{D}^*\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}) \odot \mathbf{S}^T\right\}\right]^{-1}\ . \tag{32}$$

Using the matrix inversion lemma we can relate the two bounds. Observe that

$$\begin{aligned}\mathbf{S}\mathbf{A}^*\mathbf{R}^{-1}\mathbf{A}\mathbf{S} &= \mathbf{S} - \mathbf{S}\left(\mathbf{I} - \mathbf{A}^*(\mathbf{A}\mathbf{S}\mathbf{A}^* + \sigma^2\mathbf{I})^{-1}\mathbf{A}\mathbf{S}\right) \\ &= \mathbf{S} - \mathbf{S}\left(\mathbf{I} + \mathbf{A}^*\sigma^{-2}\mathbf{A}\mathbf{S}\right)^{-1}\ .\end{aligned} \tag{33}$$

Since the sample covariance, $\hat{\mathbf{R}}$, converges to $\mathbf{R}$ w.p.1 and J is twice continuously differentiable, the dominated convergence theorem [30] can be used to show that

$$\lim_{N\to\infty} \frac{\partial^2 J}{\partial i \partial j} = \lim_{N\to\infty} \mathrm{E}\left[\frac{\partial^2 J}{\partial i \partial j}\right] . \tag{25}$$

Consequently, $\mathbf{Q} = \mathbf{H}$, and the covariance of the asymptotic distribution of the ML estimates is simply given by $\mathbf{H}^{-1}$. We defer the somewhat lengthy evaluation of the matrix $\mathbf{H}$ to Appendix B. $\square$

4.2 The WSF Method

The asymptotic properties of the WSF method are derived in [13]. Therein, it is shown that the normalized WSF estimation error, $\sqrt{N}\tilde{\boldsymbol{\theta}}$ has a limiting normal distribution with zero-mean and covariance matrix

$$N\mathrm{E}[\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T] = \frac{\sigma^2}{2}\left[\mathrm{Re}\left\{(\mathbf{D}^*\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}) \odot (\mathbf{S}\mathbf{A}^*\mathbf{R}^{-1}\mathbf{A}\mathbf{S})^T\right\}\right]^{-1} . \tag{26}$$

The following important result is obtained by comparing the expression above with (17).

Theorem 2 *The WSF estimate of $\boldsymbol{\theta}$ has the same asymptotic distribution as the stochastic ML estimate.* $\square$

4.3 Asymptotic Efficiency of the Estimates

The CRB is a lower bound on the estimation error variance for *any* unbiased estimator. An estimator that (asymptotically) achieves the CRB is said to be (asymptotically) *efficient*. It is well-known that the ML method is asymptotically efficient provided that the likelihood function is *sufficiently regular*. Under the assumptions of Section 2 the "general theory" of ML estimation applies.

Theorem 3 *Let $\hat{\boldsymbol{\eta}}$ be the ML estimate of the true parameter vector $\boldsymbol{\eta}_0$. Assume that*

$$\left|\frac{\partial^3}{\partial\theta^3}\mathbf{a}(\theta)\right| < const, \quad \theta \in \Theta . \tag{27}$$

Then, as N tends to infinity, the ML estimate has the following limiting distribution

$$\sqrt{N}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \in \mathrm{AsN}(\mathbf{0}, \mathrm{CRB}_{\mathrm{STO}}), \tag{28}$$

where $\mathrm{CRB}_{\mathrm{STO}}$ is the Cramér-Rao bound for a single observation ($N = 1$).

and

$$N\mathrm{E}[(\tilde{\sigma}^2)^2] \quad = \quad \frac{\sigma^4}{m-d} \tag{18}$$

respectively. Here,

$$\mathbf{D}(\boldsymbol{\theta}) = \left[\frac{d}{d\theta_1}\mathbf{a}(\theta_1), \ldots, \frac{d}{d\theta_d}\mathbf{a}(\theta_d)\right] \ . \tag{19}$$

The symbol $\odot$ denotes elementwise multiplication and all expressions above are evaluated at the true parameter values.

Proof The asymptotic normality and zero-mean of the ML estimation error is provided by Theorem 3 in Section 4.3. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ be the ML estimates obtained from the concentrated negative log likelihood function in Equation (13). Since the estimates are strongly consistent, a first-order Taylor expansion shows that the covariance of the parameter estimates, $\sqrt{N}\left[\hat{\boldsymbol{\theta}}^T, \hat{\sigma}^2\right]^T$ in the asymptotic distribution is

$$\mathbf{H}^{-1}\mathbf{Q}\mathbf{H}^{-1}, \tag{20}$$

where the ij^{th} element of the matrices $\mathbf{H}$ and $\mathbf{Q}$ are obtained from

$$\{\mathbf{H}\}_{ij} \quad = \quad \lim_{N\to\infty} \frac{\partial^2 J}{\partial i \partial j} \tag{21}$$

$$\{\mathbf{Q}\}_{ij} \quad = \quad \lim_{N\to\infty} N\mathrm{E}\left[\frac{\partial J}{\partial i}\frac{\partial J}{\partial j}\right] \ , \tag{22}$$

see, e.g. [13] or [28]. In the partial derivatives above, i and j denote the i^{th} and j^{th} components of $\left[\boldsymbol{\theta}^T, \sigma^2\right]^T$ and the derivatives are evaluated at the true parameter values. The derivation of $\mathbf{Q}$ is tedious. However, it is well-known that the normalized negative log likelihood function satisfies

$$\mathrm{E}\left[\frac{\partial^2 l}{\partial i \partial j}\right] = N\mathrm{E}\left[\frac{\partial l}{\partial i}\frac{\partial l}{\partial j}\right] \ , \tag{23}$$

see e.g., [29], page 98. It is then not surprising that a similar relation holds for the concentrated criterion function. Applying Lemma 1 of Appendix A yields[4]

$$\mathrm{E}\left[\frac{\partial^2 J}{\partial i \partial j}\right] = N\mathrm{E}\left[\frac{\partial J}{\partial i}\frac{\partial J}{\partial j}\right] \ . \tag{24}$$

[4]Since we are dealing with the normalized negative log likelihood function here, this expression differs in a sign and a factor N with the expression in the Appendix.

4 Asymptotic Properties for Gaussian Signals

In this section, the accuracy of the DOA estimates is investigated under the assumption of Gaussian signal waveforms. It is shown that the stochastic ML method and the WSF method have the same asymptotic properties and that they both achieve the Cramér-Rao Bound (CRB).

4.1 The Stochastic ML Method

Let us first establish strong consistency of the ML estimates. Since $\hat{\mathbf{R}}$ converges to $\mathbf{R}$ with probability one (w.p.1), and $\mathbf{R}$ is continuously differentiable w.r.t. all parameters, it is easy to show that the log likelihood function converges w.p.1, uniformly in the parameters, to the limiting function

$$\bar{l}(\boldsymbol{\eta}) = m \log \pi + \log |\mathbf{R}(\boldsymbol{\eta})| + \mathrm{Tr}\{\mathbf{R}^{-1}(\boldsymbol{\eta})\mathbf{R}\} \ . \tag{16}$$

Consequently, the estimate of $\boldsymbol{\eta}$ converges w.p.1 to the value that minimizes (16). In the proof of Theorem 2.1 of [26], it is shown that (16) is minimized if $\mathbf{R}(\boldsymbol{\eta}) = \mathbf{R}$. By assumption, the covariance uniquely determines the true parameters, $\boldsymbol{\eta}_0$. From this it follows that the estimate of $\boldsymbol{\eta}$ converges w.p.1 to the true value.

The "general theory" of ML estimation states, that under certain regularity conditions the ML estimate is asymptotically Gaussian distributed with covariance equal to the CRB, [27]. These regularity conditions are verified in Section 4.3. The CRB gives the covariance of the entire parameter vector $\boldsymbol{\eta}$, (11), whereas the WSF method only provides estimates of the DOAs, $\boldsymbol{\theta}$. In order to compare the asymptotic properties of ML and WSF, an expression for the "$\boldsymbol{\theta}$ corner" of the CRB must be found.

Extracting the covariance of $\hat{\boldsymbol{\theta}}$ from the CRB matrix directly is far from trivial. By exploiting the separability of the likelihood function, an easier path is obtained.

Theorem 1 *Let $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ minimize (13). Then as N tends to infinity, the normalized estimation errors $\sqrt{N}\tilde{\boldsymbol{\theta}} = \sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ and $\sqrt{N}\tilde{\sigma}^2 = \sqrt{N}(\hat{\sigma}^2 - \sigma^2)$ are asymptotically independent, and have limiting Gaussian distributions with zero-means and covariance matrices*

$$N\mathrm{E}[\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T] \;=\; \frac{\sigma^2}{2}\left[\mathrm{Re}\left\{(\mathbf{D}^*\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}) \odot (\mathbf{S}\mathbf{A}^*\mathbf{R}^{-1}\mathbf{A}\mathbf{S})^T\right\}\right]^{-1} \tag{17}$$

where $\mathbf{P_A} = \mathbf{AA}^\dagger$, $\mathbf{P_A^\perp} = \mathbf{I} - \mathbf{P_A}$, and $\mathbf{X} = \mathbf{P_A}\hat{\mathbf{R}}\mathbf{P_A} + \sigma^2\mathbf{P_A^\perp}$. The current parametrization of the emitter covariance, which allows the separation of the ML criterion, does not guarantee a positive semi-definite estimate of $\mathbf{S}$. See [24] for a discussion on this. Since the estimate is consistent, positive definiteness can be guaranteed if the true $\mathbf{S}$ is positive definite and if N is "sufficiently large". However, if $\mathbf{S}$ is singular this is not necessarily the case.

3.2 The Weighted Subspace Fitting Method

As observed in Section 2, the d' dimensional signal subspace is confined to a d dimensional subspace that is spanned by the array manifold vectors corresponding to the true signal parameters, $\boldsymbol{\theta}_0$. A natural estimation criterion is to find the *best least squares fit* of the two subspaces

$$[\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}] = \arg\min_{\boldsymbol{\theta}, \mathbf{T}} \|\hat{\mathbf{E}}_s \mathbf{W}^{1/2} - \mathbf{A}(\boldsymbol{\theta})\mathbf{T}\|_F^2, \tag{14}$$

where $\mathbf{W}$ is a positive definite weighting matrix and $\|\cdot\|_F$ denotes the Frobenius norm. By solving for $\mathbf{T}$, and substituting back in Equation (14), $\hat{\boldsymbol{\theta}}$ can equivalently be expressed as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathrm{Tr}\{\mathbf{P_A^\perp}(\boldsymbol{\theta})\hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^*\}. \tag{15}$$

As proved in [13], the weighting matrix that gives lowest asymptotic estimation error variance is given by, $\mathbf{W} = \tilde{\mathbf{\Lambda}}^2 \hat{\mathbf{\Lambda}}_s^{-1}$ where $\tilde{\mathbf{\Lambda}} = \hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2\mathbf{I}$, and $\hat{\sigma}^2$ is any consistent estimate of the noise variance. Herein, (15) with this optimal choice of $\mathbf{W}$, is referred to as the *Weighted Subspace Fitting* (WSF) method. Methods for solving the minimization problem are discussed in [22]. In [13], it is observed through numerical examples that the WSF method is asymptotically efficient for Gaussian signal waveforms. This empirical observation is proved in the next section.

3 Estimation Methods

Since the main focus in this paper is on the performance of the stochastic ML and the WSF methods, these techniques are briefly described below.

3.1 The Stochastic ML Method

Under the Gaussian signal waveform assumption, the array output constitutes a stationary, temporally white, zero-mean complex Gaussian random process with covariance matrix $\mathbf{R}$, given by (6). The normalized negative log likelihood function of the observations $\mathbf{x}(1), \ldots, \mathbf{x}(N)$, has the following form

$$l(\boldsymbol{\eta}) = m \log \pi + \log |\mathbf{R}(\boldsymbol{\eta})| + \mathrm{Tr}\{\mathbf{R}^{-1}(\boldsymbol{\eta})\hat{\mathbf{R}}\} \ , \tag{10}$$

where $\boldsymbol{\eta}$ represents the unknown parameters of the array covariance matrix. The ML estimate of $\boldsymbol{\eta}$ is obtained by minimizing $l(\boldsymbol{\eta})$. Herein, the unknown parameters are assumed to be $\boldsymbol{\theta}$, $\mathbf{S}$, and σ^2. Noting that $\mathbf{S}$ is a Hermitian matrix, $\boldsymbol{\eta}$ contains the following $d^2 + d + 1$ real parameters

$$\boldsymbol{\eta} = [\theta_1, \ldots, \theta_d, s_{11}, \ldots, s_{dd}, \mathrm{Re}\{s_{12}\}, \mathrm{Im}\{s_{12}\}, \ldots, \mathrm{Re}\{s_{d-1,d}\}, \mathrm{Im}\{s_{d-1,d}\}, \sigma^2] \ . \tag{11}$$

The stochastic ML method requires a non-linear, $(d^2 + d + 1)$-dimensional optimization. Although some methods have been reported for performing this search (e.g., [5, 8]) it is often unacceptably expensive.

As noted in [7], the log likelihood function (10) can be separated and thus the dimension of the optimization can be reduced. For fixed $\boldsymbol{\theta}$ and σ^2, the minimum of (10) with respect to an unrestricted, Hermitian emitter covariance matrix is given by [7, 24, 25]

$$\hat{\mathbf{S}} = \mathbf{A}^{\dagger}(\hat{\mathbf{R}} - \sigma^2\mathbf{I})\mathbf{A}^{\dagger *}, \tag{12}$$

where $\mathbf{A}^{\dagger} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$. When (12) is substituted into (10), the following concentrated negative log likelihood function is obtained ($\boldsymbol{\theta}$- and σ^2-independent terms are omitted),

$$\begin{aligned} J(\boldsymbol{\theta}, \sigma^2) &= -\sigma^{-2}\mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}\hat{\mathbf{R}}\} + \sigma^{-2}\mathrm{Tr}\{\hat{\mathbf{R}}\} + \log |\mathbf{P}_{\mathbf{A}}\hat{\mathbf{R}}\mathbf{P}_{\mathbf{A}} + \sigma^2\mathbf{P}_{\mathbf{A}}^{\perp}| \\ &= \sigma^{-2}\mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}^{\perp}\hat{\mathbf{R}}\} + \log |\mathbf{X}| \ , \end{aligned} \tag{13}$$

Introduce the eigendecomposition of the sample covariance matrix in a similar fashion as (7)

$$\hat{\mathbf{R}} = \hat{\mathbf{E}}\hat{\mathbf{\Lambda}}\hat{\mathbf{E}}^* = \hat{\mathbf{E}}_s\hat{\mathbf{\Lambda}}_s\hat{\mathbf{E}}_s^* + \hat{\mathbf{E}}_n\hat{\mathbf{\Lambda}}_n\hat{\mathbf{E}}_n^* \ . \tag{9}$$

Determination of the number of emitters, d, and the dimension of the signal subspace, d', is a crucial task for the methods described herein. This *detection* problem is discussed in, e.g., [20, 21, 22]. It is assumed here that both d and d' are known. To guarantee unique estimates of $\boldsymbol{\theta}_0$, we further assume that $d < (m + d')/2$. This implies that $\Re(\mathbf{E}_s) \subseteq \Re(\mathbf{A}(\boldsymbol{\theta}))$ holds if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, see [23].

the main interest for our purposes is in estimating the unknown DOAs. However, the parameter space usually includes other unknowns as well, see further Section 3.1.

The vector of signal waveforms, $\mathbf{s}(t)$, is assumed to be a stationary, temporally white[3], zero-mean complex Gaussian random process with second moments

$$\begin{aligned} \mathrm{E}[\mathbf{s}(t)\mathbf{s}^*(s)] &= \mathbf{S}\delta_{ts} && (4) \\ \mathrm{E}[\mathbf{s}(t)\mathbf{s}^T(s)] &= \mathbf{0} \ , && (5) \end{aligned}$$

where δ_{ts} is the Kronecker delta. In the analysis of Section 5, this assumption is replaced by a much weaker requirement on $\mathbf{s}(t)$.

The additive noise, $\mathbf{n}(t)$, is modeled as a stationary, temporally white, zero-mean complex Gaussian random process. For simplicity, we will also require $\mathbf{n}(t)$ to be *spatially white*, i.e., $\mathrm{E}[\mathbf{n}(t)\mathbf{n}^*(t)] = \sigma^2\mathbf{I}$. The assumption of spatially white noise is no restriction if the noise covariance is known (up to an unknown scalar σ^2). The noise is assumed to be uncorrelated with the signal waveforms. The covariance matrix of the array output has the following familiar structure

$$\mathbf{R} = \mathrm{E}[\mathbf{x}(t)\mathbf{x}^*(t)] = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{S}\mathbf{A}^*(\boldsymbol{\theta}_0) + \sigma^2\mathbf{I} \ . \tag{6}$$

The rank of the signal covariance matrix, $\mathbf{S}$, is denoted d'. Let

$$\mathbf{R} = \sum_{i=1}^{m} \lambda_i \mathbf{e}_i \mathbf{e}_i^* = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^* = \mathbf{E}_s\boldsymbol{\Lambda}_s\mathbf{E}_s^* + \mathbf{E}_n\boldsymbol{\Lambda}_n\mathbf{E}_n^* \tag{7}$$

be the eigendecomposition of $\mathbf{R}$, where $\lambda_1 > \lambda_2 > \cdots > \lambda_{d'+1} = \cdots = \lambda_m = \sigma^2$. The fact that the smallest eigenvalue of $\mathbf{R}$ has multiplicity $m - d'$ and is equal to the noise variance is well-known, see e.g., [1]. The columns of the matrix $\mathbf{E}_s$ are the normalized d' eigenvectors of $\mathbf{R}$ that correspond to the largest eigenvalues. The range space of $\mathbf{E}_s$ is often referred to as the *signal subspace*. Its orthogonal complement is the *noise subspace* and is spanned by the columns of $\mathbf{E}_n = [\mathbf{e}_{d'+1}, \ldots, \mathbf{e}_m]$. It is easy to see that the signal subspace is a subset of $\Re(\mathbf{A}(\boldsymbol{\theta}_0))$, i.e., $\Re(\mathbf{E}_s) \subseteq \Re(\mathbf{A}(\boldsymbol{\theta}_0))$. These subspaces coincide if and only if $d' = d$, in which case the signals are said to be non-coherent. The *sample covariance matrix*, $\hat{\mathbf{R}}$, is defined by

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}(t)\mathbf{x}^*(t) \ . \tag{8}$$

[3]In this context, a signal is temporally white if it has a flat spectrum over the (narrow) frequency band of interest, and if the sampling frequency is slow compared to the bandwidth, see further [19].

2 Assumptions and Notation

Assume that d narrow band plane waves impinge on an array of m sensors, where $m > d$. The measured array output is a weighted superposition of the wavefronts, corrupted by additive noise. The narrow band signal assumption allows us to model the time delays of the wavefronts at different sensors as phase shifts. The measured sensor outputs are also affected by the individual sensor gain and phase responses, modeled as a complex weighting of the wavefronts. The output of the i^{th} sensor is represented by

$$x_i(t) = \sum_{j=1}^{d} a_i(\theta_j) s_j(t) + n_i(t) \ , \tag{1}$$

where θ_j represents unknown signal parameters associated with the j^{th} wavefront, $a_i(\theta_j)$ is a complex scalar representing the propagation delay of the j^{th} emitter signal and the gain and phase adjustments by the i^{th} sensor. The j^{th} emitter signal is represented by $s_j(t)$ and the additive noise sequence is denoted $n_i(t)$. The sensor outputs are collected in the complex m-vector $\mathbf{x}(t)$, to form the array output

$$\mathbf{x}(t) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_d)]\mathbf{s}(t) + \mathbf{n}(t) = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{s}(t) + \mathbf{n}(t) \ , \tag{2}$$

where $\boldsymbol{\theta}_0$ is a parameter vector corresponding to the true signal parameters. The vector $\mathbf{a}(\theta_j) = [a_1(\theta_j) \cdots a_m(\theta_j)]^T$ contains the sensor responses to a unit wavefront having parameters θ_j. The collection of these vectors over the parameter space of interest

$$\mathcal{A} = \{\mathbf{a}(\theta_j) \mid \theta_j \in \Theta\} \tag{3}$$

is called the *array manifold*. It is assumed that the array manifold vectors have bounded third derivatives with respect to the parameters, and that for any collection of m distinct θ_j, the matrix $\mathbf{A}(\boldsymbol{\theta})$ has full rank. In general, each wavefront is parametrized by several signal parameters, such as bearing and elevation angle, polarization, range, center frequency, etc. The results presented here apply to a general parametrization. However, to avoid unnecessary notational complexity, we restrict the discussion to the one-parameter problem. Thus, θ_j is a real scalar, referred to as the direction-of-arrival (DOA), and $\boldsymbol{\theta}_0 = [\theta_1, \ldots, \theta_d]^T$ is a real d-dimensional vector of unknown parameters. Collect N independent observations, $\mathbf{x}(1), \ldots, \mathbf{x}(N)$, of the array output. Given these observations, coherence. Also, the papers differ in the methods of derivation. As a consequence we present results not found in [18], for example, the asymptotic distribution of the signal eigenvectors is derived under quite general conditions, the CRB on the noise variance is given, a result on separable likelihood functions is provided in Appendix A.

for the deterministic CRB was derived.

The Weighted Subspace Fitting (WSF) approach to the estimation problem, [9, 11], is a multidimensional signal subspace method and belongs to the same general class of *subspace fitting methods* as deterministic ML, Conventional Beamforming, MUSIC [1, 2], and ESPRIT [12]. The asymptotic properties of the estimates have been derived and WSF has been shown to yield the lowest estimation error variance in the class of subspace fitting methods, [13]. Herein, it is shown that the WSF method has the same asymptotic properties as the stochastic ML technique. Consequently, WSF is asymptotically efficient for Gaussian signal waveforms. This in turn results in a compact matrix expression for the stochastic CRB, which has not been available previously.

Although the analysis of the deterministic ML method, [9, 14], does not assume Gaussian signal waveforms, most other results do. For instance, the asymptotic properties of the stochastic ML method are known only for the Gaussian case. Also, the asymptotic analysis of eigenvector-based methods as reported in, for example [13, 15, 16], relies on the assumption of Gaussian signal waveforms. This is because the asymptotic analysis of the sample covariance eigenvectors, e.g. [17], is based on independent, zero-mean, Gaussian observations. Herein, the analysis is extended to deterministic signal waveforms. The asymptotic distribution of the signal eigenvectors is derived for a general "deterministic signal in additive Gaussian noise" model, and applied to the sensor array estimation problem. It is found that the stochastic ML technique and all multidimensional subspace fitting methods mentioned above preserve their asymptotic properties *regardless* of the actual signal waveforms.

During the review process, it has come to the authors' knowledge that many of the results presented herein parallel those of [18]. The compact expression for the stochastic CRB appears also in [18], and the asymptotic properties of the stochastic ML method are derived. Furthermore, a subspace-based estimation procedure termed MODE (Method Of Direction Estimation) is analyzed. The latter technique is asymptotically identical to the WSF method when the emitter covariance matrix has full rank. However, for coherent sources only WSF is efficient. An important difference between the analysis presented in [18] and in the present paper is the assumption of full rank emitter covariance matrix made in [18]. Our analysis is valid for arbitrary signal correlation, including full

1 Introduction

Sensor array processing has been an active research area for several years. The problems under consideration concern information extraction from measurements using spatially distributed sensors. The measured outputs are assumed to be noise-corrupted superpositions of narrow band plane waves. Given observations of the sensor outputs, the objective is to estimate unknown parameters associated with the wavefronts. These parameters can include bearings and/or elevation angles, signal waveforms, center frequencies, etc. Areas such as radar arrays, radio and microwave communication, acoustic sensor arrays in underwater applications and seismic exploration industry, are all concerned with estimating parameters from observations of a sensor array output.

A vast number of methods have been proposed in the literature for solving the estimation problem, see for example [1, 2, 3]. When formulated in an appropriate statistical framework, the Maximum Likelihood (ML) principle provides a systematic way to obtain an estimator. Under certain regularity conditions, the ML estimator is known to be asymptotically *efficient*, i.e., it achieves the Cramér-Rao Bound (CRB) on the estimation error variance. In this sense, ML has the best possible asymptotic properties.

The sensor noise is often regarded a superposition of several "error sources". Due to the central limit theorem, it is therefore natural to model the noise as a Gaussian random process. For the signal waveforms, two main models have appeared in the literature. One approach assumes that also the signal waveforms are Gaussian. The corresponding ML method and CRB have been formulated and studied in several papers, see e.g., [1, 4, 5, 6, 7, 8], and are referred to as the *stochastic* ML and CRB respectively. It is easily checked that the stochastic likelihood function is sufficiently regular, resulting in an asymptotically efficient ML method.

In many applications, the signal waveforms are not well-approximated by Gaussian random processes. It has then been proposed to model the signals as arbitrary deterministic sequences. The corresponding *deterministic* (or conditional) ML method is studied in, for instance, [3, 6, 9, 10]. The deterministic likelihood function does not meet the required regularity conditions, and the deterministic ML estimate does not achieve the corresponding CRB. This unusual fact was noted in [10], where a compact expression

Analysis of Subspace Fitting and ML Techniques for Parameter Estimation from Sensor Array Data[1]

B. Ottersten, MEMBER, IEEE, M. Viberg, MEMBER, IEEE,
AND T. Kailath[2] FELLOW, IEEE

Dept. of Electrical Engineering
Linköping University
S-581 83 Linköping, Sweden
46-13-282680

Revised Dec. 4, 1990

To appear in Transactions on Signal Processing

Abstract

Signal parameter estimation from sensor array data is a problem that is encountered in many engineering applications. Under the assumption of Gaussian distributed emitter signals, the so-called *stochastic* Maximum Likelihood (ML) technique is known to be statistically efficient, i.e., the estimation error covariance attains the Cramér-Rao Bound (CRB) asymptotically. Herein, it is shown that also the multidimensional signal subspace method, termed Weighted Subspace Fitting (WSF), is asymptotically efficient. This also results in a novel, compact matrix expression for the CRB on the estimation error variance. The asymptotic analysis of the ML and WSF methods is extended to deterministic emitter signals. The asymptotic properties of the estimates for this case are shown to be identical to the Gaussian emitter signal case, i.e., independent of the actual signal waveforms. Conclusions, concerning the modeling aspect of the sensor array problem are drawn.

[1]This work was supported in part by the SDI/IST Program managed by the Office of Naval Research, under Contract N00014-85-K-0550 and by the Joint Services Program at Stanford University (US Army, US Navy, US Air Force) under Contract DAAL03-88-C-0011

[2]Information Systems Laboratory, Stanford University