A SUBSPACE FITTING-LIKE METHOD FOR ALMOST LOW RANK MODELS

Mats Bengtsson
Signal Processing, S3, Royal Institute of Technology
100 44 Stockholm, SWEDEN
Tel: +46 8 790 84 63, fax: +46 8 790 73 29
Email: matsb@s3.kth.se

ABSTRACT

Subspace fitting methods have grown popular for parameter estimation in many different application, for example sensor array signal processing, blind channel identification and identification of linear state space systems. Here we show that similar procedures can be used even for data models where the noise free signal gives a full rank contribution to the covariance matrix. A general weighting is introduced and the optimal weight matrix is given together with the resulting asymptotic covariance of the parameter estimates. The method works well when the number of dominating eigenvalues still is fairly small. As an example, we study estimation of direction and spread angle of a source subject to local scattering, using a uniform linear array of sensors. As the algorithm is computationally expensive, the results are not primarily intended for practical implementations, rather they show the theoretical limit for any estimation procedure that uses a low rank approximation of the covariance matrix.

1 PRELIMINARIES

We describe first a fairly general data model followed by a couple of examples. Assume that $\mathbf{y}(t)$ is a real or complex vector valued signal given by

$$\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{n}(t) \tag{1}$$

where $\mathbf{n}(t)$ is spatially and temporally white Gaussian noise with $\mathrm{E}[\mathbf{n}(t_1)\mathbf{n}^*(t_2)] = \sigma_n^2\mathbf{I}\delta(t_1 - t_2)$. Suppose further that $\mathbf{x}(t)$ is a stochastic or deterministic quasi-stationary [5] signal with (for simplicity) $\overline{\mathrm{E}}[\mathbf{x}(t)] = 0$ and a covariance matrix $\overline{\mathrm{E}}[\mathbf{x}(t)\mathbf{x}^*(t)] = \mathbf{R}_x(\boldsymbol{\eta})$ which is a known function of $\boldsymbol{\eta}$, the vector of parameters of interest. The notation $\overline{\mathrm{E}}$ is defined by $\overline{\mathrm{E}}[f(t)] = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mathrm{E}[f(t)]$. From (1), $\mathbf{R}_y = \overline{\mathrm{E}}[\mathbf{y}(t)\mathbf{y}^*(t)] = \mathbf{R}_x(\boldsymbol{\eta}) + \sigma_n^2\mathbf{I}$.

In normal applications of subspace system identification, $\mathbf{R}_x(\boldsymbol{\eta})$ is low rank, i.e., $\mathrm{rank}[\mathbf{R}_x(\boldsymbol{\eta})] < m$, the dimension of $\mathbf{x}(t)$. Suppose instead that $\mathrm{rank}[\mathbf{R}_x(\boldsymbol{\eta})] = m$, but the matrix is almost low rank in the sense that at least one of the eigenvalues is significantly smaller than the others.

One application, which will be used in the examples below, is a model of the signal received at an antenna array from a transmitting source in an environment that introduces local scattering around the source, see [1, 2, 9]. Assuming that the scattering is uniformly distributed around the nominal direction of arrival and the antenna array is linear and uniform, it can be shown that the received baseband signal is [3]

$$\mathbf{y}(t) = s(t)\mathbf{v}(t) + \mathbf{n}(t) \tag{2}$$

where $s(t)$ is the transmitted waveform, $\mathbf{v}(t)$ is a complex random vector approximately $\mathbf{N}\big(0, \mathbf{R}_v(\theta, \sigma_\theta)\big)$ distributed, where

$$\big[\mathbf{R}_v(\theta,\sigma_\theta)\big]_{kl} \approx e^{j2\pi(k-l)\Delta\sin\theta}\,\mathrm{sinc}\big(\frac{2}{\sqrt{3}}(k-l)\Delta\,\sigma_\theta\cos\theta\big) \tag{3}$$

θ is the nominal direction of arrival, σ_θ is the standard deviation of the angular spread and Δ is the element separation in wavelengths. Thus, $\mathbf{x}(t) = s(t)\mathbf{v}(t)$, $\mathbf{R}_x(\boldsymbol{\eta}) = \sigma_s^2\mathbf{R}_v(\theta,\sigma_\theta)$ and the parameters to be estimated are $\boldsymbol{\eta} = [\theta, \sigma_\theta]^T$ (σ_s^2 can be estimated using linear least squares). An example with a deterministic signal is the estimation of a linear swept-frequency chirp signal, $\mathbf{x}(t) = [c(t), c(t+1), \ldots, c(t+m-1)]^T$, where $c(t) = A\cos(Bt^2 + Ct + D)$ and $\boldsymbol{\eta} = [B, C]^T$ are the parameters of interest (A and D can be found using linear least squares methods).

Common to these examples is that the number of significantly large eigenvalues changes with the parameters (σ_θ and A, respectively).

Since $\mathbf{R}_x(\boldsymbol{\eta})$ is full rank, it is impossible to define a true signal subspace. However, it is still possible to perform an eigenvalue decomposition of $\mathbf{R}_y = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^*$, select the d principal eigenvectors of $\mathbf{R}_y$ as a pseudo-signal subspace and write the covariance matrix as

$$\mathbf{R}_y = \mathbf{E}_s\boldsymbol{\Lambda}_s\mathbf{E}_s^* + \mathbf{E}_n\boldsymbol{\Lambda}_n\mathbf{E}_n^*$$

where

$$\boldsymbol{\Lambda}_s = \mathrm{diag}[\lambda_{s,1}, \ldots, \lambda_{s,d}]$$
$$\boldsymbol{\Lambda}_n = \mathrm{diag}[\lambda_{n,1}, \ldots, \lambda_{n,m-d}]$$

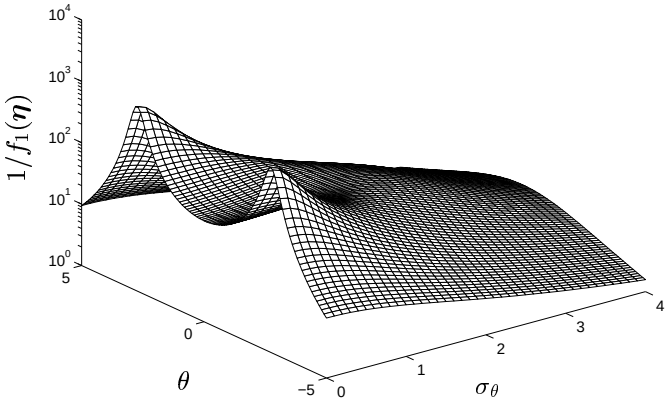

Figure 1: Cost function of the DSPE algorithm, $1/f_1(\boldsymbol{\eta})$ given in equation (5). $d = 2$, $\theta = 0°$, $\sigma_\theta = 2°$, 8 sensors, no noise, true covariance matrix.

and $\lambda_{s,1} \geq \cdots \geq \lambda_{s,d} > \lambda_{n,1} \geq \cdots \geq \lambda_{n,m-d}$. Here, d, the dimension of the pseudo-signal subspace is a parameter to be chosen by the user.

We assume that the following identifiability condition is fulfilled.

$$\text{span}[\mathbf{E}_s(\boldsymbol{\eta})] = \text{span}[\mathbf{E}_s(\boldsymbol{\xi})] \qquad \Longleftrightarrow \qquad \boldsymbol{\eta} = \boldsymbol{\xi} \tag{4}$$

Form the sample covariance matrix $\hat{\mathbf{R}}_y = \frac{1}{N}\sum_{t=1}^{N} \mathbf{y}(t)\mathbf{y}^*(t)$ and denote the corresponding estimated pseudo-signal and pseudo-noise eigenvectors with $\hat{\mathbf{E}}_s$ and $\hat{\mathbf{E}}_n$, respectively.

2 ALGORITHMS

2.1 Previous Results

Recently, a number of different algorithms have been suggested to estimate θ and σ_θ in the model for local scattering.

The maximum likelihood estimator is formulated in [9] together with a weighted covariance matching algorithm,

$$\hat{\boldsymbol{\eta}} = \arg \min_{\boldsymbol{\eta},\sigma_s^2,\sigma_n^2} \text{Tr}\left\{ \left((\mathbf{R}_y(\boldsymbol{\eta},\sigma_s^2,\sigma_n^2) - \hat{\mathbf{R}}_y)\mathbf{W} \right)^2 \right\}$$

where the optimal weighting $\mathbf{W} = \hat{\mathbf{R}}_y^{-1}$ gives asymptotically efficient estimates.

To generalize the idea of MUSIC, note that $\mathbf{E}_s^*\mathbf{R}_x \approx 0$ as long as the angular spread is small. This observation has led to two closely related algorithms,

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} f_k(\boldsymbol{\eta}) = \arg\min_{\boldsymbol{\eta}} \text{Tr}[\hat{\mathbf{E}}_n^* \mathbf{R}_x^k(\boldsymbol{\eta}) \hat{\mathbf{E}}_n] \tag{5}$$

where $k = 1$ gives the DSPE algorithm [10] and $k = 2$ gives the DISPARE algorithm [6]. However both methods in general fail to give consistent estimates. To see this, insert the true $\mathbf{R}_x$ decomposed into a pseudo-signal and a pseudo-noise part

$$\begin{aligned} f_k(\boldsymbol{\eta}) &= \text{Tr}[\mathbf{E}_n^* \mathbf{R}_x^k(\boldsymbol{\eta}) \mathbf{E}_n] \\ &= \text{Tr}\Big[\mathbf{E}_n^* \mathbf{E}_s(\boldsymbol{\eta}) \boldsymbol{\Lambda}_s^k(\boldsymbol{\eta}) \mathbf{E}_s^*(\boldsymbol{\eta}) \mathbf{E}_n \\ &\qquad + \mathbf{E}_n^* \mathbf{E}_n(\boldsymbol{\eta}) \boldsymbol{\Lambda}_n^k(\boldsymbol{\eta}) \mathbf{E}_n^*(\boldsymbol{\eta}) \mathbf{E}_n \Big] \end{aligned}$$

At the true parameter, the first term is zero and

$$f_k(\boldsymbol{\eta}_0) = \text{Tr}[\boldsymbol{\Lambda}_n^k(\boldsymbol{\eta}_0)]$$

but since both terms depend on the parameters, it will typically be possible to find some other parameter value that decreases the second term so much that the cost function is less than $f_k(\boldsymbol{\eta}_0)$ even though the first term is non-zero. This is illustrated in figure 1, where the global peaks are found at $[\theta, \sigma_\theta] = [\pm 2.7°, 0°]$, whereas the true source is located at $[\theta_0, \sigma_{\theta_0}] = [0°, 2°]$.

If the ordinary root-MUSIC algorithm is applied to $\mathbf{R}_y$ searching for two point source directions, it can be shown that within good approximation, the directions found are given by $\hat{\theta}_{1,2} = \theta_0 \pm \lambda(\sigma_{\theta_0})$ for some increasing function $\lambda(x)$. This observation can be exploited to obtain estimates of θ and σ_θ at a low computational cost, see [2, 3].

2.2 Subspace Fitting

Following the ideas of traditional subspace fitting methods, we would like to find an estimate $\hat{\boldsymbol{\eta}}$ that makes $\mathbf{E}_s(\hat{\boldsymbol{\eta}})$ as orthogonal as possible to $\hat{\mathbf{E}}_n$.

Two possible cost functions that solve the consistency problems of DSPE and DISPARE (5) are

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} f_{\text{NSF}}(\boldsymbol{\eta}) = \arg\min_{\boldsymbol{\eta}} \|\mathbf{E}_s^*(\boldsymbol{\eta}) \hat{\mathbf{E}}_n\| \tag{6}$$

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} f_{\text{SSF}}(\boldsymbol{\eta}) = \arg\min_{\boldsymbol{\eta}} \|\mathbf{E}_n^*(\boldsymbol{\eta}) \hat{\mathbf{E}}_s\| \tag{7}$$

representing pseudo-noise and pseudo-signal subspace fitting approaches, respectively. Since these cost functions can be computationally expensive (d or $m - d$ eigenvectors of $\mathbf{R}_y(\boldsymbol{\eta})$ must be calculated at each parameter value), an attractive alternative is

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} f'_{\text{NSF}}(\boldsymbol{\eta}) = \arg\min_{\boldsymbol{\eta}} \|\mathbf{S}^*(\boldsymbol{\eta}) \hat{\mathbf{E}}_n\| \tag{8}$$

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} f'_{\text{SSF}}(\boldsymbol{\eta}) = \arg\min_{\boldsymbol{\eta}} \|\mathbf{N}^*(\boldsymbol{\eta}) \hat{\mathbf{E}}_s\| \tag{9}$$

where $\mathbf{S}(\boldsymbol{\eta})$ and $\mathbf{N}(\boldsymbol{\eta})$ are matrix functions such that $\text{span}\{\mathbf{S}(\boldsymbol{\eta})\} = \text{span}\{\mathbf{E}_s(\boldsymbol{\eta})\}$, $\text{span}\{\mathbf{N}(\boldsymbol{\eta})\} = \text{span}\{\mathbf{E}_n(\boldsymbol{\eta})\}$ and where $\mathbf{S}(\boldsymbol{\eta})$, $\mathbf{N}(\boldsymbol{\eta})$ or good approximations thereof can be computed quickly. Any norm can be used in (6)–(9), for example a weighted Frobenius norm.

Note that the algorithms give consistent estimates when $N \to \infty$ as long as the identifiability condition (4) holds.

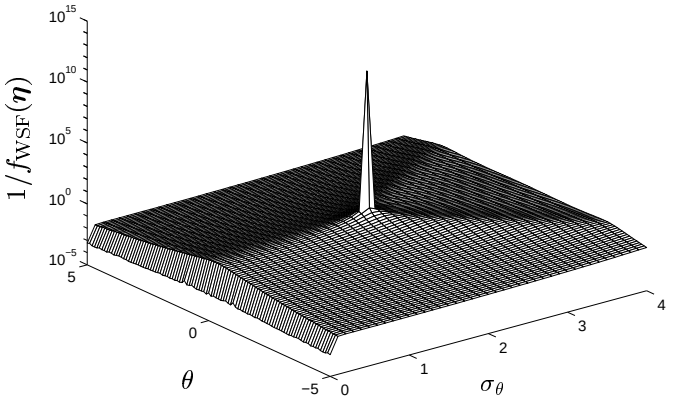

Figure 2: Cost function of the optimally weighted signal subspace fitting criterion, $1/f_{\mathrm{WSF}}(\boldsymbol{\eta})$. $d = 2$, $\theta = 0°$, $\sigma_\theta = 2°$, 8 sensors, no noise, true covariance matrix.

2.3 Optimal Weighting

The most general weighted squared norm that can be applied to the SSF cost function in equation (7) is

$$f_{\mathrm{SSF}}(\boldsymbol{\eta}) = (\mathrm{vec}[\mathbf{E}_n^*(\boldsymbol{\eta})\hat{\mathbf{E}}_s])^* \, \mathbf{W} \, \mathrm{vec}[\mathbf{E}_n^*(\boldsymbol{\eta})\hat{\mathbf{E}}_s]$$

Using the theory of Asymptotically Best Consistent (ABC) estimation [7] and a Taylor expansion of the eigendecomposition [2, 4] it can be shown that the optimal weighting $\mathbf{W}_{\mathrm{opt}}$ is

$$\mathbf{W}_{\mathrm{opt}}^{-1} = \mathrm{diag}[\mathrm{vec}\,\mathbf{M}_{ns}](\boldsymbol{\Lambda}_s \otimes \boldsymbol{\Lambda}_n)\,\mathrm{diag}[\mathrm{vec}\,\mathbf{M}_{ns}]$$

where

$$[\mathbf{M}_{ns}]_{kl} = \frac{1}{\lambda_{n,k} - \lambda_{s,l}}$$

After some simplifications, the resulting optimally weighted subspace fitting criterion is given by

$$f_{\mathrm{WSF}}(\boldsymbol{\eta}) = \|\boldsymbol{\Lambda}_n^{-1/2}(\boldsymbol{\Lambda}_n\mathbf{E}_n^*\hat{\mathbf{E}}_s - \mathbf{E}_n^*\hat{\mathbf{E}}_s\boldsymbol{\Lambda}_s)\boldsymbol{\Lambda}_s^{-1/2}\|_F^2$$

(note that all the entities except for $\hat{\mathbf{E}}_s$ depend on $\boldsymbol{\eta}$ and that in addition both $\boldsymbol{\Lambda}_s$ and $\boldsymbol{\Lambda}_n$ depend on the noise power σ_n^2.)

The optimal weighting matrix can be replaced by a consistent estimate without affecting the large sample performance which gives the following practical cost function

$$\begin{aligned} f_{\mathrm{WSF}}(\boldsymbol{\eta}) =& \|\hat{\boldsymbol{\Lambda}}_n^{-1/2}(\hat{\boldsymbol{\Lambda}}_n\mathbf{E}_n^*(\boldsymbol{\eta})\hat{\mathbf{E}}_s - \mathbf{E}_n^*(\boldsymbol{\eta})\hat{\mathbf{E}}_s\hat{\boldsymbol{\Lambda}}_s)\hat{\boldsymbol{\Lambda}}_s^{-1/2}\|_F^2 \\ =& \mathrm{Tr}[\hat{\mathbf{E}}_s\hat{\boldsymbol{\Lambda}}_s^{-1}\hat{\mathbf{E}}_s^*\mathbf{E}_n\hat{\boldsymbol{\Lambda}}_n\mathbf{E}_n^* + \hat{\mathbf{E}}_s\hat{\boldsymbol{\Lambda}}_s\hat{\mathbf{E}}_s^*\mathbf{E}_n\hat{\boldsymbol{\Lambda}}_n^{-1}\mathbf{E}_n^* \\ & - 2\hat{\mathbf{E}}_s\hat{\mathbf{E}}_s^*\mathbf{E}_n\mathbf{E}_n^*] \end{aligned}$$

($\|\mathbf{M}\|_F^2 = \mathrm{Tr}[\mathbf{M}^*\mathbf{M}]$, the Frobenius norm) Note that for the special case of direction estimation of point sources, $\boldsymbol{\Lambda}_n = \sigma_n^2\mathbf{I}$ and $f_{\mathrm{WSF}}(\boldsymbol{\eta})$ reduces to the familiar WSF/MODE criterion [8, 11]. One example of the WSF cost function is shown in Figure 2.

Assuming that $\mathbf{x}(t)$ is Gaussian and temporally white, it can be shown that the asymptotic large sample distribution of the parameter estimates is given by

$$\begin{aligned} &\sqrt{N}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \in \mathrm{AsN}(0, \mathbf{C}) \text{ , where} \\ &\mathbf{C} = \frac{1}{2}\,\mathrm{Re}\Big[\Big(\frac{\partial}{\partial\boldsymbol{\eta}}\,\mathrm{vec}[\mathbf{R}_y(\boldsymbol{\eta})]\Big)^*\Big([\mathbf{R}_y^{-1}]_s^c \otimes [\mathbf{R}_y^{-1}]_n\Big) \\ &\qquad \times \frac{\partial}{\partial\boldsymbol{\eta}}\,\mathrm{vec}[\mathbf{R}_y(\boldsymbol{\eta})]\Big]^{-1} \text{ and} \\ &[\mathbf{R}_y^{-1}]_s = \mathbf{E}_s\boldsymbol{\Lambda}_s^{-1}\mathbf{E}_s^* \\ &[\mathbf{R}_y^{-1}]_n = \mathbf{E}_n\boldsymbol{\Lambda}_n^{-1}\mathbf{E}_n^* \end{aligned}$$

Again, for the traditional point source model, it is easy to show that this expression reduces to the WSF results given in [8, 11].

Note however that in general the performance does not reach the Cramér-Rao lower bound, in contrast to the point source model where WSF indeed is asymptotically efficient. Still, for many problems, the performance is near optimal and the method could have computational advantages compared to ML or covariance matching techniques [9].

3 NUMERICAL EXAMPLE

Simulations have been performed on a scenario with an eight element uniform linear antenna array with half wavelength element separation, a single source located at broadside subject to uniformly distributed local scattering and SNR 10 dB. Each estimate was calculated from a data burst of $N = 100$ data samples. Figure 3 shows the theoretical and estimated RMS values of $\hat{\theta}$ and $\hat{\sigma}_\theta$ for different values of σ_θ, calculated from 500 trials for each test case. For comparison, the Cramér-Rao lower bound is included. A few (less than 1% in each experiment) outliers were removed in the simulation results.

Three different values of d, 1,2 and 3, were tested. As expected, the algorithm gives best performance when d is chosen as the number of eigenvalues of $\mathbf{R}_x$ that are significantly larger than the background noise, which can also be deduced from the theoretical performance expressions. In this application, $d = 2$ gives the best overall performance for the parameter range of interest. In general, d could be selected based on some a priori information on the parameter range or set adaptively using a numerically estimated rank of $\hat{\mathbf{R}}_y$.

4 CONCLUSIONS

The idea of weighted subspace fitting has been extended to full rank models. The resulting algorithm gives con-

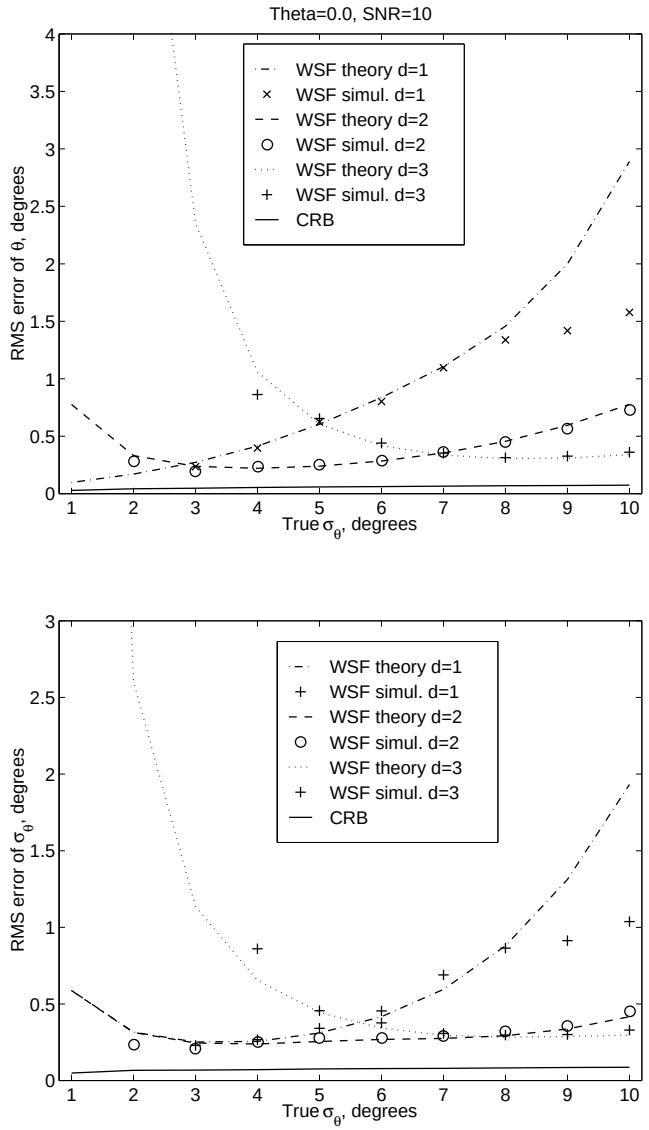

Figure 3: Theoretical and empirical performance of the weighted pseudo-signal subspace fitting algorithm for different values of d.

sistent estimates and asymptotic expressions for the estimation error have been derived.

Similar ideas have been presented previously [6, 10] but it was shown that those algorithms fail to give consistent estimates.

In general, the weighted pseudo-subspace fitting estimator is not asymptotically efficient and since both ML and weighted covariance matching [9] do provide asymptotically efficient estimates at a comparable or even lower computational cost, the WSF algorithm may seem to be of minor interest. However, the algorithm may retain the implicit noise filtering and computationally well-behaved cost function that is offered by traditional subspace methods. Furthermore, the results can give ideas for and provide a better understanding of other algorithms, such as the low complexity algorithm presented in [3], which can be seen as an approximation of (8) where $\mathbf{S}(\boldsymbol{\eta})$ is approximated by the array response of two closely separated point sources.

REFERENCES

[1] F. Adachi, M.T. Feeney, A.G. Williamson, and J.D. Parsons. Cross correlation between the envelopes of 900 MHz signals received at a mobile radio base station site. *IEE Proceedings, Pt. F*, 133(6):506–512, October 1986.

[2] Mats Bengtsson. Sensor array processing for scattered sources. Licentiate Thesis TRITA-S3-SB-9729, Signal Processing, Royal Institute of Technology, Stockholm, Sweden, December 1997.

[3] Mats Bengtsson and Björn Ottersten. Rooting techniques for estimation of angular spread with an antenna array. In *Proceedings of VTC'97*, pages 1158–1162, May 1997.

[4] R.P. Gupta. Asymptotic theory for principal component analysis in the complex case. *J. Indian Statist. Assoc.*, 3:97–106, 1965.

[5] Lennart Ljung. *System Identification: Theory for the User*. Prentice-Hall, Inc., Englewood Cliffs, N.J, 1987.

[6] Y. Meng, P. Stoica, and K.M. Wong. Estimation of the directions of arrival of spatially dispersed signals in array processing. *IEE Proceedings - Radar, Sonar and Navigation*, 143(1):1–9, February 1996.

[7] Torsten Söderström and Petre Stoica. *System Identification*. Prentice-Hall, Int., 1989.

[8] Petre Stoica and Kenneth C. Sharman. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38:1132–1142, July 1990.

[9] Tõnu Trump and Björn Ottersten. Estimation of nominal direction of arrival and angular spread using an array of sensors. *Signal Processing*, 50(1-2):57–69, April 1996.

[10] Shahrokh Valaee, Benoit Champagne, and Peter Kabal. Parametric localization of distributed sources. *IEEE Transactions on Signal Processing*, 43:2144–2153, September 1995.

[11] Mats Viberg and Björn Ottersten. Sensor array processing based on subspace fitting. *IEEE Transactions on Signal Processing*, 39(5):1110–1121, May 1991.