ON DOA ESTIMATION IN UNKNOWN COLORED NOISE-FIELDS USING AN IMPERFECT ESTIMATE OF THE NOISE COVARIANCE

Karl Werner and Magnus Jansson

Dept. of Signals, Sensors and Systems
KTH-Royal Institute of Technology
SE-100 44 Stockholm, Sweden
karl.werner@s3.kth.se

ABSTRACT

Most algorithms for direction-of-arrival (DOA) estimation require the noise covariance matrix to be known or to possess a known structure. In many cases the noise covariance is in fact estimated from separate measurements. This paper addresses the combined effects of finite sample sizes both in the estimated noise covariance matrix and in the data with signals present. It is assumed that a batch of signal-free samples is available in addition to the signal-containing samples. No assumption is made on the structure of the noise covariance.

In this work the asymptotical covariance of weighted subspace fitting (WSF) is derived for the case when the data are whitened using an estimated noise covariance. The obtained expression suggests an optimal weighting that improves performance compared to the standard choice. In addition, a new method based on covariance matching is proposed. The proposed method is by construction asymptotically efficient.

Monte Carlo simulations show promising small sample performance for the two new methods and confirm the asymptotical results. The CRB for the data model is included in the numerical evaluations and it is found to coincide with the WSF asymptotical covariance derived.

1. INTRODUCTION

The problem of estimating parameters of narrowband signals impinging on an array of sensors is important in a wide range of applications. Examples include radar, sonar, seismic exploration, microwave- and radio communications and biomedicine. Consider the signal model

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

where, at time t, $\mathbf{x}(t)$ is the $m-$vector of sensor outputs, $\mathbf{s}(t)$ is the $d-$vector of impinging signals and $\mathbf{n}(t)$ is an $m-$ vector of noise. The $m \times d$-matrix $\mathbf{A}(\boldsymbol{\theta})$ is composed of the steering vectors of the array, parameterized by the n_θ-vector $\boldsymbol{\theta}$. The direction-of-arrival (DOA) estimation problem is that of estimating $\boldsymbol{\theta}$ from the sequence $\{\mathbf{x}(t)\}_{t=0}^{N-1}$. In this work, the vectors $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are assumed to be circularly symmetric complex Gaussian vectors with covariance matrices $\mathbf{P}$ and $\mathbf{Q}$, respectively. A number of estimation techniques have been proposed for this data model in the literature [1, 2, 3, 4]. Most algorithms are based on the assumption that the noise is spatially white and uniform, i.e., $\mathbf{Q} = \sigma^2\mathbf{I}_m$. If this assumption cannot be made there are two main approaches:

The first one is to parameterize the covariance matrix $\mathbf{Q}$ and then perform a joint estimation of $\boldsymbol{\theta}$ and the noise parameters [5, 6, 7, 8]. Alternatively, some structure can be imposed on the signals [9]. Another possibility is to assume and exploit certain structures in the noise covariance matrix, see e.g. [10] and [11]. While the above approaches have the distinct advantage of not requiring prior knowledge of the elements of $\mathbf{Q}$, they do require prior knowledge of either the structure of the noise covariance or the signals. Fundamental identifiability issues pose a limit on how general the imposed structure can be.

The second standard approach is to estimate $\mathbf{Q}$ using a separate batch of signal-free samples and then use the so-obtained estimate to whiten the noise part of the signals $\mathbf{x}(t)$. While this approach is frequently suggested or implicitly assumed in the literature, there are few results published on the effects on the quality of the DOA estimates due to errors in the whitening. An exception to this is given in [12], where the effect of small deterministic deviations from the uniform spatially white noise assumption is analyzed for some common algorithms. In [13] and [14] the effect on subspace estimation algorithms of stochastic perturbations of the noise covariance and array manifold are investigated. The derivations in those two papers do however not assume that the noise covariance errors stem from an imperfectly estimated whitening matrix. Thus, they are not directly applicable to the problem at hand. Also, unlike the present work, [13] and [14] do not treat finite sample effects in the same framework as perturbations in the noise covariance matrix.

In this work a statistical approach is taken to assess the impact of estimation errors in the whitening matrix and to devise algorithms with improved performance. It is assumed that M samples $\{\mathbf{n}(t-M)\}_{t=0}^{M-1}$ drawn from the same distribution as $\mathbf{n}(t)$ are available in addition to the data $\{\mathbf{x}(t)\}_{t=0}^{N-1}$.

In Section 3, the asymptotical covariance of the *Weighted Subspace Fitting* (WSF) [4] estimate based on whitened data is derived. Moreover, an optimal weighting is introduced that improves performance compared to the standard weighting at no additional computational cost. With the new weighting, the asymptotical covariance for the WSF estimate coincides with the Cramér-Rao bound (CRB) for the data model. The expression for the asymptotical covariance allows for a useful and simple comparison to the standard (assuming known noise covariance) WSF asymptotical covariance derived in e.g. [15].

In Section 4, an asymptotically optimal algorithm based on covariance matching is derived for the estimation problem at hand. This new algorithm does not explicitly rely on pre-whitening of the data. The paper is concluded with some numerical results in Section 5.

In the following, $\mathbf{X}^{\dagger}$ denotes the Moore-Penrose pseudo inverse of the matrix $\mathbf{X}$. The projection matrices $\mathbf{\Pi}_{\mathbf{X}} = \mathbf{X}\mathbf{X}^{\dagger}$ and $\mathbf{\Pi}_{\mathbf{X}}^{\perp} = \mathbf{I} - \mathbf{\Pi}_{\mathbf{X}}$ will be used frequently. The vector $\mathrm{vec}\{\mathbf{X}\}$ is obtained by stacking the columns of $\mathbf{X}$. The Kronecker product is denoted by $\otimes$ and the direct (Schur) product by $\odot$ (see e.g. [16]). The superscript $\cdot^*$ denotes conjugate transpose and $\cdot^T$ denotes transpose. The statement $\mathbf{X}_1 \geq \mathbf{X}_2$ implies that $\mathbf{X}_1 - \mathbf{X}_2$ is positive semi definite (p.s.d.) and $\mathbf{X}_1 > \mathbf{X}_2$ implies that the difference is positive definite (p.d.).

2. DATA MODEL AND CRAMÉR-RAO BOUND

The estimation problem treated is that of estimating $\boldsymbol{\theta}$ from the data

$$\{\mathbf{n}(-M), \ldots, \mathbf{n}(-1), \mathbf{x}(0), \ldots, \mathbf{x}(N-1)\},$$

with

$$\begin{aligned}
&\mathrm{E}\left[\,\mathbf{n}(t)\mathbf{n}^*(t)\,\right] = \mathbf{Q},\\
&\mathrm{E}\left[\,\mathbf{x}(t)\mathbf{x}^*(t)\,\right] = \mathbf{A}(\boldsymbol{\theta})\mathbf{P}\mathbf{A}^*(\boldsymbol{\theta}) + \mathbf{Q} = \mathbf{R},
\end{aligned} \tag{2}$$

Both $\mathbf{n}(t)$ and $\mathbf{x}(t)$ are complex Gaussian circularly symmetric random vectors. The functional form of the matrix

$$\mathbf{A}(\boldsymbol{\theta}) = \left[\mathbf{a}([\boldsymbol{\theta}]_{1:l}),\ \ldots,\mathbf{a}([\boldsymbol{\theta}]_{(d-1)l+1:n_\theta})\right] \tag{3}$$

depends, among other things, on the array geometry and sensor characteristics. The vector $[\boldsymbol{\theta}]_{(i-1)l+1:li}$ is the subvector of $\boldsymbol{\theta}$ containing the parameters associated with source i. If not otherwise stated, the results of this work hold when the steering vectors, $\mathbf{a}(\boldsymbol{\theta})$, are parameterized by multiple parameters per source ($l > 1$). The steering vectors are assumed to be sufficiently smooth functions of $\boldsymbol{\theta}$ (see [4] for details). This assumption is necessary for the asymptotical analysis that will follow. The array must also be unambiguous, which means that $\mathbf{a}(\boldsymbol{\theta}_1)$ and $\mathbf{a}(\boldsymbol{\theta}_2)$ are linearly independent for any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$. It is assumed that $\mathbf{Q}$ is a p.d. matrix. Furthermore, identifiability requires that $\frac{(m+d')}{2} > d$, where d' is the rank of the signal correlation matrix, $\mathbf{P}$ (see [17] for details).

The stochastic CRB for the case of uniform white noise ($\mathbf{Q} = \sigma^2\mathbf{I}$) was derived in [18, 15]. It was generalized to unknown colored noise fields in [19]. The CRB that will be presented in this section is conceptually quite different due to the extra observed signal-free samples in the data model. It should be noted that the bound derived here is for an estimator with no knowledge of the rank d' of $\mathbf{P}$, see [20] for more details about this assumption. By defining the sample correlation matrices

$$\hat{\mathbf{R}}_1 = \frac{1}{M}\sum_{k=-M}^{-1}\mathbf{n}(k)\mathbf{n}^*(k)$$

and

$$\hat{\mathbf{R}}_2 = \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{x}(k)\mathbf{x}^*(k), \tag{4}$$

and by using the statistical independence of the observations, the negative log-likelihood function can be written (neglecting constant terms)

$$\begin{aligned}
l(\boldsymbol{\eta}) = M\log|\mathbf{Q}| + M\mathrm{tr}\{\mathbf{Q}^{-1}\hat{\mathbf{R}}_1\} + N\log|\mathbf{R}|\\
+ N\mathrm{tr}\{\mathbf{R}^{-1}\hat{\mathbf{R}}_2\},
\end{aligned} \tag{5}$$

where

$$\boldsymbol{\eta} = [\boldsymbol{\theta}^T,\ \boldsymbol{\phi}^T,\ \boldsymbol{\xi}^T]^T. \tag{6}$$

The d^2-vector $\boldsymbol{\phi}$ and the m^2-vector $\boldsymbol{\xi}$ contain the real parameters required to parameterize the hermitian matrices $\mathbf{P}$ and $\mathbf{Q}$, respectively. In the parameterization, the fact that the matrices are p.d. or p.s.d. is not exploited. This will not affect the final result (see e.g. [20] or [19]). The so-called Bang's formula (see e.g. [15]) gives the Fischer information matrix

$$\begin{aligned}
[\mathbf{I}(\boldsymbol{\eta})]_{ij} = \mathrm{E}\left[\frac{\delta^2 l(\boldsymbol{\eta})}{\delta[\boldsymbol{\eta}]_i\delta[\boldsymbol{\eta}]_j}\right] =& M\mathrm{tr}\{\mathbf{Q}^{-1}\mathbf{Q}_i\mathbf{Q}^{-1}\mathbf{Q}_j\}\\
&+ N\mathrm{tr}\{\mathbf{R}^{-1}\mathbf{R}_i\mathbf{R}^{-1}\mathbf{R}_j\},
\end{aligned} \tag{7}$$

where $\mathbf{X}_j$ denotes the element-wise derivative of the matrix $\mathbf{X}$ w.r.t. $[\boldsymbol{\eta}]_j$. The bound for the signal parameters (those in $\boldsymbol{\theta}$), $\mathrm{CRB}_{\boldsymbol{\theta}}$, is the upper left $n_\theta \times n_\theta$ submatrix of the inverse of the Fisher information matrix.

3. WHITENED WEIGHTED SUBSPACE FITTING

The asymptotical (in a sense made clear later on) covariance of the WSF estimate of $\boldsymbol{\theta}$ when whitened data are used will be derived in this section. The estimate is given by

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_W &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ V_W(\boldsymbol{\theta}),\\
V_W(\boldsymbol{\theta}) &= \mathrm{tr}\{\mathbf{\Pi}_{\hat{\mathbf{Z}}\mathbf{A}}^{\perp}\hat{\mathbf{E}}_s\mathbf{W}\hat{\mathbf{E}}_s^*\},\\
\hat{\mathbf{Z}}\hat{\mathbf{R}}_2\hat{\mathbf{Z}} &= \hat{\mathbf{E}}_s\hat{\mathbf{\Lambda}}_s\hat{\mathbf{E}}_s^* + \hat{\mathbf{E}}_n\hat{\mathbf{\Lambda}}_n\hat{\mathbf{E}}_n^*,\\
\hat{\mathbf{Z}}\hat{\mathbf{Z}} &= \hat{\mathbf{R}}_1^{-1}.
\end{aligned} \tag{8}$$

The matrix $\hat{\mathbf{\Lambda}}_s$ is diagonal and the diagonal elements are the d' largest eigenvalues of the whitened array covariance matrix, $\hat{\mathbf{Z}}\hat{\mathbf{R}}_2\hat{\mathbf{Z}}$. The columns of $\hat{\mathbf{E}}_s$ are the corresponding orthogonal eigenvectors. In the same way the matrix $\hat{\mathbf{\Lambda}}_n$ contains the remaining eigenvalues and the columns of $\hat{\mathbf{E}}_n$ are the corresponding eigenvectors. The weighting matrix $\mathbf{W}$ is an arbitrary hermitian p.d. matrix. This whitened version of WSF seems to be the standard approach in the literature for the problem at hand. The following theorem can be shown.

Theorem 1 *Let $\boldsymbol{\theta}_0$ be the true parameter values in the data model of Section 2. Then the quantity $\sqrt{N}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0)$ has a limiting (as $N \to \infty$ and $\frac{N}{M} \to \alpha$, α constant) zero-mean gaussian distribution*

$$\begin{aligned}
\sqrt{N}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) \sim \mathrm{AsN}(\mathbf{0}, \mathbf{C}(\mathbf{W})),\\
\mathbf{C}(\mathbf{W}) = \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1},
\end{aligned} \tag{9}$$

where

$$\begin{aligned}
[\mathbf{H}]_{ij} =& 2\mathrm{Re}\ \mathrm{tr}\{\mathbf{A}_j^*\mathbf{Z}\bar{\mathbf{\Pi}}^{\perp}\mathbf{Z}\mathbf{A}_i\bar{\mathbf{A}}^{\dagger}\mathbf{E}_s\mathbf{W}\mathbf{E}_s^*\bar{\mathbf{A}}^{\dagger *}\},\\
[\mathbf{G}]_{ij} =& 2\mathrm{Re}\ \mathrm{tr}\{\bar{\mathbf{\Pi}}^{\perp}\mathbf{Z}\mathbf{A}_i\bar{\mathbf{A}}^{\dagger}\mathbf{E}_s\mathbf{W}(\mathbf{\Lambda}_s - \mathbf{I})^{-2}\\
&\times(\mathbf{\Lambda}_s + \alpha\mathbf{I})\mathbf{W}\mathbf{E}_s^*\bar{\mathbf{A}}^{\dagger *}\mathbf{A}_j^*\mathbf{Z}\}.
\end{aligned} \tag{10}$$

For brevity of presentation, the definitions

$$\begin{aligned}
\mathbf{Z}\mathbf{Z} &= \mathbf{Q}^{-1}, \\
\bar{\mathbf{A}} &= \mathbf{Z}\mathbf{A}, \\
\bar{\mathbf{\Pi}} &= \mathbf{\Pi}_{\bar{\mathbf{A}}}, \\
\alpha &= \frac{N}{M}, \\
\bar{\mathbf{R}} &= \mathbf{Z}\mathbf{R}\mathbf{Z} \ \text{ and} \\
\bar{\mathbf{R}} &= \mathbf{E}_s\mathbf{\Lambda}_s\mathbf{E}_s^* + \mathbf{E}_n\mathbf{E}_n^* \qquad (11)
\end{aligned}$$

are used in the above theorem and in the following. The last equality is an eigen-decomposition of the whitened covariance matrix $\bar{\mathbf{R}}$. The diagonal matrix $\mathbf{\Lambda}_s$ contains the d' largest eigenvalues, and the corresponding eigenvectors are the columns of $\mathbf{E}_s$. The columns of $\mathbf{E}_n$ are the remaining eigenvectors. The proof of the theorem can be found in [21].

If the array steering matrix $\mathbf{A}$ is parameterized with one parameter per source, then

$$\mathbf{A}_i = \left[\mathbf{0}_{m\times(i-1)},\, \mathbf{a}_i,\, \mathbf{0}_{m\times(n_\theta-i)}\right]. \qquad (12)$$

By defining the matrix

$$\mathbf{\Delta}_{\mathbf{A}} = [\mathbf{a}_1,\, \ldots,\, \mathbf{a}_{n_\theta}] \qquad (13)$$

it is easy to show that (10) reduces to

$$\begin{aligned}
\mathbf{H} &= 2\mathrm{Re}\ \mathbf{\Delta}_{\mathbf{A}}^*\mathbf{Z}\bar{\mathbf{\Pi}}^{\perp}\mathbf{Z}\mathbf{\Delta}_{\mathbf{A}} \odot \bar{\mathbf{A}}^{\dagger}\mathbf{E}_s\mathbf{W}\mathbf{E}_s^*\bar{\mathbf{A}}^{\dagger *} \\
\mathbf{G} &= 2\mathrm{Re}\ \mathbf{\Delta}_{\mathbf{A}}^*\mathbf{Z}\bar{\mathbf{\Pi}}^{\perp}\mathbf{Z}\mathbf{\Delta}_{\mathbf{A}} \odot \bar{\mathbf{A}}^{\dagger}\mathbf{E}_s\mathbf{W} \\
&\quad \times (\mathbf{\Lambda}_s - \mathbf{I})^{-2}(\mathbf{\Lambda}_s + \alpha\mathbf{I})\mathbf{W}\mathbf{E}_s^*\bar{\mathbf{A}}^{\dagger *}. \qquad (14)
\end{aligned}$$

With a slight modification of the derivations in [4] and [22] the following theorem can be shown.

Theorem 2 *Let $\mathbf{C}(\mathbf{W})$ be given as in (9) and (14). Then*

$$\begin{aligned}
\mathbf{C}(\mathbf{W}) &\geq \mathbf{C}(\mathbf{W}_{\mathit{new}}), \\
\mathbf{W}_{\mathit{new}} &= (\mathbf{\Lambda}_s - \mathbf{I})^2(\mathbf{\Lambda}_s + \alpha\mathbf{I})^{-1}. \qquad (15)
\end{aligned}$$

In fact, as proved in [21], the weighting is optimal also for the case of multiple parameters per source. Furthermore, just as in standard WSF, the weighting matrix $\mathbf{W}_{\text{new}}$ can be replaced by a consistent estimate without affecting the asymptotical performance. This is also proven in [21]. It can furthermore be shown that [21]

$$\mathbf{C}(\mathbf{W}_{\text{new}}) = N\mathrm{CRB}_{\boldsymbol{\theta}}, \qquad (16)$$

where $\mathrm{CRB}_{\boldsymbol{\theta}}$ is the CRB for the signal parameters. This implies that WSF with the new, optimal weighting is asymptotically statistically efficient.

It is easy to see that the asymptotical covariance coincides with the standard expression for WSF (see e.g. [15]) as the ratio $\alpha = \frac{N}{M}$ tends to zero.

It is interesting to note that the standard method can be made optimal by a small modification of the weighting matrix. The change of weighting does not complicate the implementation of the algorithm in any way (note that the new weighting matrix is diagonal). Algorithms such as MODE [3] can be used with only minor modifications, avoiding a multidimensional search by exploiting certain array geometries.

4. A COVARIANCE MATCHING APPROACH

Although conceptually appealing, it is not obvious that explicit whitening of the data using an estimated whitening matrix is the best way to solve the estimation problem at hand. In this section, a covariance matching approach that is based on joint estimation of $\mathbf{Q}$, $\mathbf{P}$ and $\boldsymbol{\theta}$ will be pursued. The underlying framework for covariance matching is presented more closely in [23]. The proposed method is by construction asymptotically (in N and M in the same way as in the previous section) efficient. This follows from the *extended invariance principle* [24] and from [23]. In order to claim this, it is necessary to assume that the signal covariance matrix $\mathbf{P}$ is p.d. (see [20] for a discussion on this issue).

The criterion function to be minimized is (since $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ are uncorrelated)

$$\begin{aligned}
V_C(\boldsymbol{\theta}, \mathbf{P}, \mathbf{Q}) &= \mathrm{vec}^*\{\hat{\mathbf{R}}_1 - \mathbf{Q}\}\mathbf{W}_1\mathrm{vec}\{\hat{\mathbf{R}}_1 - \mathbf{Q}\} \\
&\quad + \mathrm{vec}^*\{\hat{\mathbf{R}}_2 - \mathbf{Q} - \mathbf{A}\mathbf{P}\mathbf{A}^*\}\mathbf{W}_2 \\
&\quad \times \mathrm{vec}\{\hat{\mathbf{R}}_2 - \mathbf{Q} - \mathbf{A}\mathbf{P}\mathbf{A}^*\}. \qquad (17)
\end{aligned}$$

The weighting matrices should be chosen as [24]

$$\begin{aligned}
\mathbf{W}_1 &= \mathrm{Cov}\left[\mathrm{vec}\{\hat{\mathbf{R}}_1\}\right]^{-1} = M(\mathbf{Q}^{-T} \otimes \mathbf{Q}^{-1}), \\
\mathbf{W}_2 &= \mathrm{Cov}\left[\mathrm{vec}\{\hat{\mathbf{R}}_2\}\right]^{-1} = N(\mathbf{R}^{-T} \otimes \mathbf{R}^{-1}). \qquad (18)
\end{aligned}$$

The matrices $\mathbf{Q}$ and $\mathbf{R}$ can be replaced by the consistent estimates $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$, respectively, without affecting the asymptotical properties of the estimate. The criterion function (17) can be concentrated w.r.t. $\mathbf{P}$ and $\mathbf{Q}$. To that end,

$$\begin{aligned}
\hat{\mathbf{P}}(\boldsymbol{\theta}, \mathbf{Q}) &= \underset{\mathbf{P}}{\mathrm{argmin}}\ V_C(\boldsymbol{\theta}, \mathbf{P}, \mathbf{Q}) \\
&= (\hat{\mathbf{R}}_2^{-\frac{1}{2}}\mathbf{A})^{\dagger}\hat{\mathbf{R}}_2^{-\frac{1}{2}}(\hat{\mathbf{R}}_2 - \mathbf{Q})\hat{\mathbf{R}}_2^{-\frac{1}{2}}(\hat{\mathbf{R}}_2^{-\frac{1}{2}}\mathbf{A})^{\dagger *}. \qquad (19)
\end{aligned}$$

In order to concentrate the criterion w.r.t. $\mathbf{Q}$, write

$$\begin{aligned}
&V_C(\boldsymbol{\theta}, \hat{\mathbf{P}}, \mathbf{Q}) = \\
&\quad M\mathrm{vec}^*\{\hat{\mathbf{R}}_1 - \mathbf{Q}\}(\hat{\mathbf{R}}_1^{-T} \otimes \hat{\mathbf{R}}_1^{-1})\mathrm{vec}\{\hat{\mathbf{R}}_1 - \mathbf{Q}\} \\
&\quad + \mathrm{vec}^*\{\hat{\mathbf{R}}_2 - \mathbf{Q}\}\hat{\mathbf{W}}_3\mathrm{vec}\{\hat{\mathbf{R}}_2 - \mathbf{Q}\}, \qquad (20)
\end{aligned}$$

where

$$\begin{aligned}
\hat{\mathbf{W}}_3 &= N\big(\hat{\mathbf{R}}_2^{-T} \otimes \hat{\mathbf{R}}_2^{-1} \\
&\quad - \hat{\mathbf{R}}_2^{-\frac{T}{2}}\mathbf{\Pi}^T_{\hat{\mathbf{R}}_2^{-\frac{1}{2}}\mathbf{A}}\hat{\mathbf{R}}_2^{-\frac{T}{2}} \otimes \hat{\mathbf{R}}_2^{-\frac{1}{2}}\mathbf{\Pi}_{\hat{\mathbf{R}}_2^{-\frac{1}{2}}\mathbf{A}}\hat{\mathbf{R}}_2^{-\frac{1}{2}}\big). \qquad (21)
\end{aligned}$$

As (20) is quadratic in $\mathbf{Q}$, its minimizer is easily found to be

$$\begin{aligned}
\mathrm{vec}\{\hat{\mathbf{Q}}\} &= (M(\hat{\mathbf{R}}_1^{-T} \otimes \hat{\mathbf{R}}_1^{-1}) + \hat{\mathbf{W}}_3)^{-1} \\
&\quad \times (M\mathrm{vec}\{\hat{\mathbf{R}}_1\} + \hat{\mathbf{W}}_3\mathrm{vec}\{\hat{\mathbf{R}}_2\}). \qquad (22)
\end{aligned}$$

It can be proved that $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$ are hermitian by construction. Furthermore, due to the consistency, both $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$ are p.d. for large enough N and M. These are in fact necessary conditions for optimality (efficiency) of the approach. Finally, insertion gives

$$\begin{aligned}\hat{\boldsymbol{\theta}}_C &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ V_C(\boldsymbol{\theta}),\\ V_C(\boldsymbol{\theta}) &= V_C(\boldsymbol{\theta}, \hat{\mathbf{P}}, \hat{\mathbf{Q}}) +'' \boldsymbol{\theta}\text{−indep.term}''\\ &= -\text{vec}^*\{M\hat{\mathbf{R}}_1^{-1} + N\hat{\mathbf{R}}_2^{-\frac{1}{2}}\boldsymbol{\Pi}^{\perp}_{\hat{\mathbf{R}}_2^{-\frac{1}{2}}\mathbf{A}}\hat{\mathbf{R}}_2^{-\frac{1}{2}}\}\\ &\quad\times\left(M(\hat{\mathbf{R}}_1^{-T}\otimes\hat{\mathbf{R}}_1^{-1}) + \hat{\mathbf{W}}_3\right)^{-1}\\ &\quad\times\text{vec}\{M\hat{\mathbf{R}}_1^{-1} + N\hat{\mathbf{R}}_2^{-\frac{1}{2}}\boldsymbol{\Pi}^{\perp}_{\hat{\mathbf{R}}_2^{-\frac{1}{2}}\mathbf{A}}\hat{\mathbf{R}}_2^{-\frac{1}{2}}\},\end{aligned} \tag{23}$$

since

$$\text{vec}^*\{\hat{\mathbf{R}}_2\}\hat{\mathbf{W}}_3\text{vec}\{\hat{\mathbf{R}}_2\} = N(m-d). \tag{24}$$

While having the same large-sample performance as the optimally weighted whitened WSF method, the two methods can be expected to behave differently in small samples. In Section 5, several numerical studies are presented that indicate that the covariance matching approach performs better for small sample sizes. However, existing techniques for avoiding the multidimensional minimization of the criterion function can be used for the WSF method while no such methods are known for the covariance matching technique.

5. SIMULATIONS

A number of computer simulations have been performed in order to illustrate the theoretical, asymptotical results derived in this work and to investigate small sample performance of the methods.

5.1. Simulation setup

The asymptotical covariance expression for whitened WSF as introduced in Theorem 1 is included in all plots. Two weighting matrices ($\mathbf{W}$ in (8)) were tested. The asymptotical covariance with $\mathbf{W}_{\text{new}} = (\boldsymbol{\Lambda}_s - \mathbf{I})^2(\boldsymbol{\Lambda}_s + \alpha\mathbf{I})^{-1}$, which is the optimal weighting for the problem, is shown as a solid line. The asymptotical covariance with $\mathbf{W}_{\text{WSF}} = (\boldsymbol{\Lambda}_s - \mathbf{I})^2\boldsymbol{\Lambda}_s^{-1}$, which would be the optimal weighting if $\mathbf{Q}$ (and thereby $\mathbf{Z}$) was known exactly, is shown as a dotted line.

In addition to the asymptotical covariances, three versions of the CRB are included. The CRB for the estimation problem in this work coincides with the asymptotical covariance for optimally weighted WSF (solid line). The CRB for the estimation problem if $\mathbf{Q}$ was known exactly is also included. It is shown as a dashed line in the plots. In one case the CRB with $N \gg M$ is included. It is shown as a dash-dotted line.

In addition to the theoretical quantities, a number of estimation methods were used and their performances were evaluated using Monte Carlo methods. Let $\hat{\boldsymbol{\theta}}_k^i$ be the estimate of the kth DOA in Monte Carlo trial i (using the method in question) and $\boldsymbol{\theta}_k$ be the true DOA. Then the performance measure used is the *RMS per source*

$$\sqrt{\frac{1}{d}\frac{1}{L}\sum_{i=1}^{L}\sum_{k=1}^{d}(\hat{\boldsymbol{\theta}}_k^i - \boldsymbol{\theta}_k)^2}, \tag{25}$$

where L is the number of Monte Carlo trials. In the simulations presented here, $L = 1000$.

Two versions of whitened WSF were tested. The results for the version based on the optimal $\mathbf{W}_{\text{new}}$ weighting are marked with stars in the plots. The squares are for the version that is based on $\mathbf{W}_{\text{WSF}}$.

The covariance matching approach of Section 4 was also tested. The results using this method are shown as circles in the plots.

In addition to the methods discussed in the present work, the well-known root-MUSIC method was also included in one of the plots (see Figure 2). It was implemented using whitening in the same way as the WSF method above. The results for the root-MUSIC method are shown as hollow stars.

In all simulations, the signal covariance matrix $\mathbf{P}$ was of the form

$$\mathbf{P} = \begin{pmatrix} 1 & \rho \\ \rho^* & 1 \end{pmatrix}, \quad 0 \le |\rho| < 1. \tag{26}$$

This makes it possible to use the parameter ρ to control the correlation of the signals. In the plots, ρ is real. The noise covariance matrix $\mathbf{Q}$ was designed to be reminiscent of a spatial AR(1)-process. The i, jth element of $\mathbf{Q}$ is

$$[\mathbf{Q}]_{ij} = \sigma^2 a^{|i-j|}, \tag{27}$$

where a is a parameter used to control the noise covariance. The array steering vectors correspond to those of a uniform linear array (ULA) with sensors separated by a half wavelength. The DOAs of the sources are measured from array broadside.

5.2. Results

Results for a case with sources at 0 and 10 degrees and five sensors are shown in Figure 1. The two sample sizes M and N have a fixed ratio of five in the experiment, and they are both varied. The noise was spatially correlated with parameters $a = 0.5$ and $\sigma^2 = 0.5$ and the signals were correlated with $\rho = 0.7$. It can be seen from the plot that the suggested optimal weighting gives a small performance gain compared to the standard choice. This is evident both in the asymptotical expressions and in the experimental evaluations of the different methods. It can also be noted that the covariance matching approach performs better than optimally weighted WSF for small M and N. Asymptotically in N and M the methods have the same error covariance.

In the next plot, Figure 2, the number of samples were kept constant at $N = 5000$ and $M = 1000$. Instead, the signal correlation, ρ, was varied. As could be suspected, the gain obtained by using $\mathbf{W}_{\text{new}}$ instead of $\mathbf{W}_{\text{WSF}}$ depends on ρ. This can be explained by the fact that $\boldsymbol{\Lambda}_s$ will have fewer dominating eigenvalues when the signals are correlated, making $\mathbf{W}_{\text{new}}$ differ more from $\mathbf{W}_{\text{WSF}}$. In the experiment, ρ was real. The phase of ρ is however not unimportant. Further investigations also indicate that the source spacing has a similar significance.

It is quite clear that the advantage of the improved methods developed here will be larger when M is small compared to N. This is illustrated in Figure 3. For the particular case studied there it can be seen that the effects of small M (poor whitening matrix) will dominate until $M > \frac{N}{3}$. Note that the new methods never will perform worse than the standard methods. The plot (and the other plots in the study) also indicates that the asymptotical expressions are valid for fairly small sample set sizes.

In Figure 4, the impact of the noise correlation is investigated. In terms of difference between the tested methods, the plot indicates that the noise correlation is of little importance. It is somewhat surprising to note that increasing noise correlation gives an

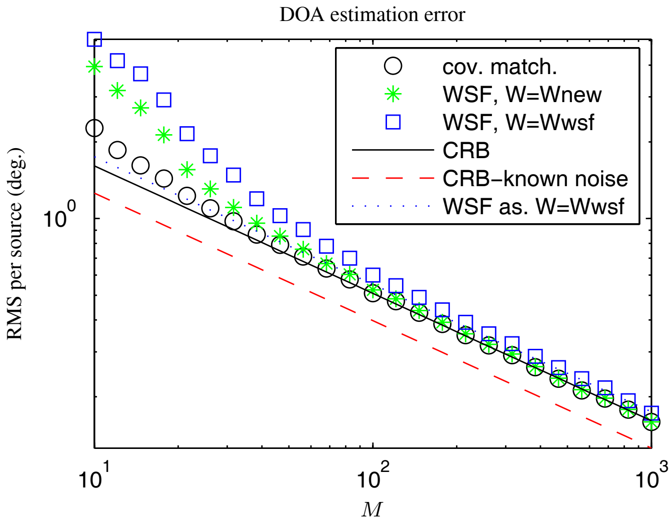

Fig. 1. RMS per source in estimation of direction of arrival with two sources at 0 and 10 degrees. A ULA with $m = 5$ elements was used. The sources were correlated with $\rho = 0.7$. The noise covariance matrix had parameters $a = 0.5$ and $\sigma^2 = 0.5$. The number of signal-free samples, M was varied and the number of signal-containing samples N was $5M$.

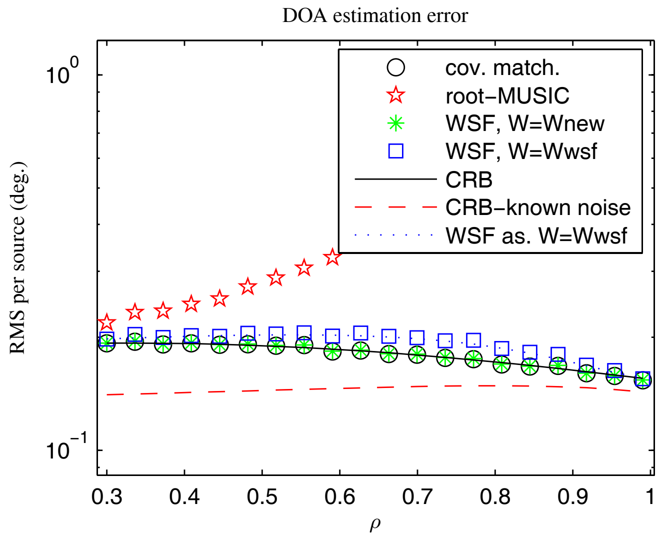

Fig. 2. RMS per source in estimation of direction of arrival with two sources at 0 and 10 degrees. A ULA with $m = 5$ elements was used. The sources were correlated with ρ varying from 0.3 to 0.99. The noise covariance matrix had parameters $a = 0$ and $\sigma^2 = 0.5$. The sample sizes were fixed at $M = 1000$, $N = 5000$.

improved performance. An explanation for this is that the noise is then partially filtered-out by the whitening transformation. Note that a performance gain can be obtained with the new methods also when the true noise covariance is identity, as is the case in Figure 2.

6. CONCLUSION

The problem of estimating parameters of narrowband signals impinging on an array of sensors has been studied. It is assumed that the sensor array output is the superposition of the response to the impinging signals and a spatially correlated noise (with an unknown correlation matrix). In addition to the samples containing both signals and noise, a batch of samples containing only noise was assumed to be available.

A standard approach is to use the noise-only samples to estimate a transformation that is used to whiten the signal containing samples. In the paper this procedure was used together with WSF. The asymptotical (in the sample sizes) covariance of the estimation error for the resulting estimate was derived for a general weighting matrix. It was further shown that a particular choice of weighting matrix minimized the asymptotical covariance and, in fact, made it coincide with the CRB for the problem. The optimal weighting matrix is diagonal and the new formulation does not increase the computational complexity compared to the standard choice.

A covariance matching approach to the estimation problem was also introduced, avoiding explicit whitening of the data. The method is asymptotically efficient.

Computer simulations indicated that the asymptotical results are useful also for relatively small sample sizes, and that the new, optimal WSF weighting gives a performance improvement. In the simulations, the covariance matching method had the best small sample behavior of the tested methods.

7. REFERENCES

[1] R. Schmidt, “Multiple emitter location and signal parameter estimation,” in *Proc. of RADC Spectrum Estimation Workshop*, 1979, pp. 243–258.

[2] R. Roy and T. Kailath, “ESPRIT-estimation of signal parameters via rotational invariance techniques,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[3] P. Stoica and K. Sharman, “Maximum likelihood methods for direction-of-arrival estimation,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 7, pp. 1132–1143, July 1990.

[4] M. Viberg and B. Ottersten, “Sensor array processing based on subspace fitting,” *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1110–1121, May 1991.

[5] B. Friedlander and A. J. Weiss, “Direction finding using noise covariance modeling,” *IEEE Transactions on Signal Processing*, vol. 43, no. 7, pp. 1557–1567, July 1995.

[6] B. Göransson and B. Ottersten, “Direction estimation in partially unknown noise fields,” *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2375–2385, Sept. 1999.

[7] M. Pesavento and A. B. Gershman, “Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise,” *IEEE Transactions on Signal Processing*, vol. 49, no. 7, pp. 1310–1324, July 2001.

[8] V. Nagesha and S. Kay, “Maximum likelihood estimation for array processing in colored noise,” *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 169–180, Feb. 1996.

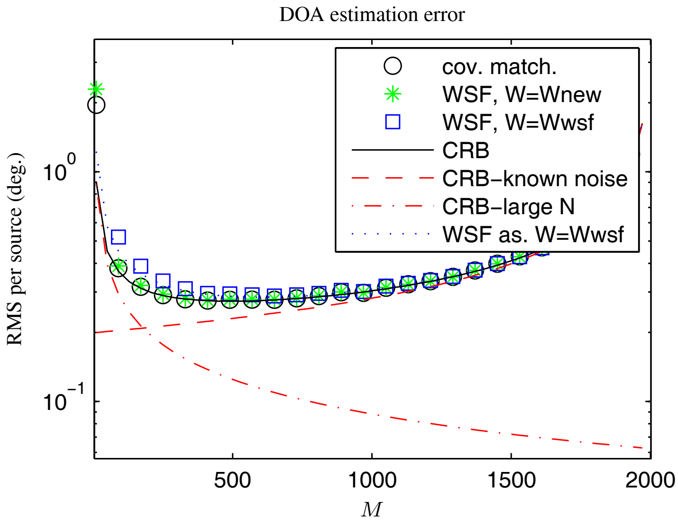

Fig. 3. RMS per source in estimation of direction of arrival with two sources at 0 and 10 degrees. A ULA with $m = 5$ elements was used. The sources were correlated with $\rho = 0.7$. The noise covariance matrix had parameters $a = 0.5$ and $\sigma^2 = 0.5$. The number of signal-free samples varied from 10 to 1980 with $N = 2000 - M$.

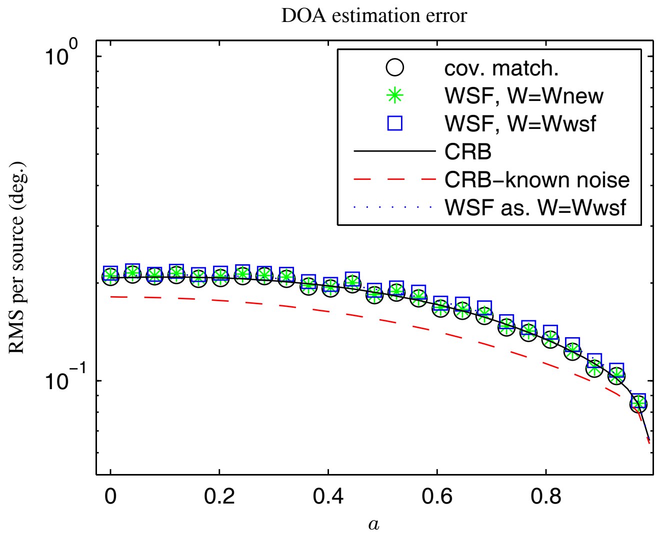

Fig. 4. RMS per source in estimation of direction of arrival with two sources at -8 and 10 degrees. A ULA with $m = 5$ elements was used. The sources were correlated with $\rho = 0.7$. The noise covariance matrix had parameters a varying from 0 to 0.99 and $\sigma^2 = 0.5$. The number of signal-free samples was $M = 200$ and $N = 800$.

[9] M. Viberg, P. Stoica, and B. Ottersten, "Maximum likelihood array processing in spatially correlated noise fields using parameterized signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 996–1004, April 1997.

[10] Q. T. Zhang and K. M. Wong, "Information theoretic criteria for the determination of the number of signals in spatially correlated noise," *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1652–1663, April 1993.

[11] P. Stoica, K. M. Wong, and Q. Wu, "On a nonparametric detection method for array signal processing in correlated noise fields," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 1030–1032, April 1996.

[12] M. Viberg, "Sensitivity of parametric direction finding to colored noise fields and undermodeling," *Signal Processing*, vol. 34, pp. 207–222, 1993.

[13] A. Swindlehurst and T.Kailath, "A performance analysis of subspace-based methods in the presence of model errors, part I: The MUSIC algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1758–1773, July 1992.

[14] A. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors: Part II - multidimensional algorithms," *IEEE Transactions on Signal Processing*, vol. 41, no. 9, pp. 2882–2890, Sept. 1993.

[15] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 590–600, March 1992.

[16] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[17] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 996–1000, July 1989.

[18] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[19] A. Gershman, P. Stoica, M. Pesavento, and E. Larsson, "Stochastic cramér-rao bound for direction estimation in unknown noise fields," *IEE Proc.-Radar Sonar Navig.*,, vol. 149, no. 1, pp. 2–8, Feb. 2002.

[20] P. Stoica, B. Ottersten, M. Viberg, and R. L. Moses, "Maximum likelihood array processing for stochastic coherent sources," *IEEE Transactions on Signal Processing*, vol. 44, no. 1, pp. 96–105, Jan. 1996.

[21] K. Werner and M. Jansson, "Optimal utilization of signal-free samples for array processing in unknown colored noise-fields," *IEEE Transactions on Signal Processing*, in preparation.

[22] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and cramér-rao bound: Further results and comparisons," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 38, no. 12, pp. 2140–2150, Dec 1990.

[23] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," in *Digital Signal Processing*, 1998, vol. 8, pp. 185–210.

[24] P. Stoica and T. Söderström, "On reparameterization of loss functions used in estimation and the invariance principle," *Signal Processing*, vol. 17, pp. 383–387, 1989.