08-229

ROBUST IMPLEMENTATION OF THE MUSIC ALGORITHM

Johan Xi Zhang[1], *Mads Græsbøll Christensen*[1], *Joachim Dahl*[1], *Søren Holdt Jensen*[1], *Marc Moonen*[2]

[1]Dept. of Electronic Systems (ES-MISP), Aalborg University, Aalborg, Denmark
[2]Dept. of Electrical Engineering (ESAT-SCD), Katholieke Universiteit Leuven, Leuven, Belgium
{jxz,mgc,joachim,shj}@es.aau.dk, marc.moonen@esat.kuleuven.be

ABSTRACT

The problem of estimating frequencies of sinusoids in noise has been studied intensively by the signal processing community during the last decades. Traditionally high resolution subspace-based techniques suffer from high computational complexity, and generally sensitive to the colored noise. We present here a frequency-domain based subspace parameter estimation algorithm termed frequency-selective MUltiple SIgnal Classification (F-MUSIC) that is based on the signal and noise subspace orthogonality property. The method is computationally efficient in providing estimates in the selected subband compared to the classic MUSIC. The performance of F-MUSIC is evaluated and compared to both MUSIC and Cramér-Rao lower bound (CRLB). In a low signal to noise ratio (SNR) with colored noise scenarios, F-MUSIC outperforms MUSIC.

Index Terms— Frequency estimation, subband, subspace orthogonality, colored noise.

1. INTRODUCTION

An extensively studied problem in signal processing is the estimation of parameters in complex exponentials embedded in the white noise. The model is defined as:

$$y(t) = \sum_{k=1}^{\bar{n}} \beta_k e^{j\omega_k t} + e(t), \qquad \beta_k = \alpha_k e^{j\theta_k}, \qquad (1)$$

for $t = 0, \ldots, N-1$, where α_k is the real amplitude of the complex exponentials, ω_k is the frequency parameters, θ_k is the phase of the harmonics, $\bar{n}$ is the number of complex exponentials, and $e(t)$ denotes the complex symmetric white Gaussian noise. The estimation problem associated with the real case can be cast as (1) by the use of the analytic signals, which is valid when there is little or no spectral content of interest near 0 and π.

The objective considered here is to estimate the parameters of $y(t)$ based on subspace techniques. This problem has been studied by many researchers during the past decades, and many algorithms have been suggested. One common approach of solving the problem is to first estimate the model parameters based on a sample covariance matrix, and in a second step either directly calculate the frequencies or find peaks on the constructed spectrum [1]. Parametric methods have a superior accuracy compared to non-parametric methods, but drawbacks are often the high computational complexity and the sensitivity to colored noise.

In many practical applications, harmonics of interest reside in a subband, while the disturbance signal lies outside. The filtering and the pre-whitening are standard methods to suppress interferences before frequency estimation. A serious drawback of the time-domain filtering on short finite sequences is that the transients from filtering may disturb the frequency estimates. While in the noise pre-whitening, the statistical parameters of the noise needs to be known in advance or to be estimated. Furthermore, the efficiency of the covariance matrix based subspace estimator will decrease with an increased length of the observed signal. Also when the number of sinusoids are too large compared to the harmonics of interest. Instead of deriving a parametric model from the time-domain sample covariance estimates, a new model based on a sub-set of data from the discrete-time Fourier transform (DFT) termed frequency-selective (FS) data model has been formulated in [2, 3]. The resulting algorithm is termed Frequency-selective ESPRIT (F-ESPRIT) which has the advantage to be computationally efficient compared to the standard ESPRIT.

In this paper we will further develop the concept used in the F-ESPRIT to propose a new Frequency-selective MUltiple SIgnal Classification (F-MUSIC) algorithm using ideas from the classical MUSIC based on the noise signal subspace orthogonality principle which is a general property. Recently, an algorithm on single and multi pitch estimator has been proposed [4, 5] where the essence of the algorithm is based on the subspace orthogonality property which makes this classical principle especially interesting. However, the proposed method F-MUSIC will be more computationally efficient and possibly more robust toward the colored noise than classical MUSIC. Even in the case of estimating the same number of frequency parameters F-MUSIC can still be computationally more efficient than MUSIC. By dividing the signal spectrum into Q equally spaced subbands and considering each band as an individual subproblem, we get a computationally more efficient approach than estimating the parameters from the covariance matrix model associated with MUSIC. The potential application areas using the proposed algorithm could be in speech/audio coding, musical instrument retrieval and time-scale modification.

The remaining part of the paper is organized as follows. In Section 2, the frequency-domain based model is described and the involved equations will be defined. In Section 3 the proposed method is described, and in Section 4 numerical examples demonstrating the estimation performance and noise robustness are presented. Finally, discussions and conclusions are given in Section 5 and 6, respectively.

The work of Johan Xi Zhang is supported by the Marie Curie EST-SIGNAL Fellowship (http://est-signal.i3s.unice.fr), contract no. MEST-CT-2005-021175.

The work of Mads Græsbøll Christensen is supported by the Parametric Audio Processing project, Danish Research Council for Technology and Production Sciences, grant no. 274060521.

2. FREQUENCY-SELECTIVE DATA MODEL

An FS data model can be formulated using the equations stated in [1] where samples from DFT are used as input data. Let us assume that the component of interests lie in a prespecified subband composed of the following Fourier frequencies:

$$\left\{ \ \tfrac{2\pi}{N}k_1 \quad \tfrac{2\pi}{N}k_2 \quad \dots \quad \tfrac{2\pi}{N}k_M \ \right\}, \tag{2}$$

where $\{k_1 \dots k_M\}$ are M given consecutive integers. The number of components n of (1) lying in the subband specified by (2) is assumed to be $n \leq \bar{n}$.

In the derivation of F-MUSIC, the following definitions will be used:

$$\begin{aligned}
w_k &= e^{j\frac{2\pi}{N}k}, \qquad k = 0, 1, \dots, N-1 && (3)\\
\mathbf{u}_k &= \left[\ w_k \quad \dots \quad w_k^m \ \right]^T && (4)\\
\mathbf{v}_k &= \left[\ 1 \quad w_k \quad \dots \quad w_k^{N-1} \ \right]^T && (5)\\
\mathbf{y} &= \left[\ y(0) \quad \dots \quad y(N-1) \ \right]^T && (6)\\
Y_k &= \mathbf{v}_k^* \mathbf{y}, \qquad k = 0, 1, \dots, N-1 && (7)\\
\mathbf{e} &= \left[\ e(0) \quad \dots \quad e(N-1) \ \right]^T && (8)\\
E_k &= \mathbf{v}_k^* \mathbf{e}, \qquad k = 0, 1, \dots, N-1, && (9)
\end{aligned}$$

where $\mathbf{u}_k$ is the phase shift vector, $\mathbf{v}_k$ is the Fourier vector, $\mathbf{y}$ is the signal vector, $\mathbf{e}$ is the noise vector, $*$ is the complex conjugate, transpose T is the operator of vector transpose, and m is a user parameter which is limited to $M > m > n$. Previous experience of MUSIC, and other similar approaches have shown that the user parameter m should be selected as large as possible in order to increase the linearly independent vectors of the noise subspace, but less than M in order to still achieve a correct estimate of the FS data model. Furthermore, to express the components of the signal, vectors $\mathbf{a}(\omega_k)$ and $\mathbf{b}(\omega_k)$ are introduced and denoted as:

$$\begin{aligned}
\mathbf{a}(\omega_k) &= \left[\ e^{j\omega_k} \quad \dots \quad e^{jm\omega_k} \ \right]^T && (10)\\
\mathbf{b}(\omega_k) &= \left[\ 1 \quad e^{j\omega_k} \quad \dots \quad e^{j(N-1)\omega_k} \ \right]^T. && (11)
\end{aligned}$$

The key equation for the FS data model involving the FFT sequence Y_k is denoted as:

$$\mathbf{u}_k Y_k = [\mathbf{a}(\omega_1) \dots \mathbf{a}(\omega_{\bar{n}})] \begin{bmatrix} \beta_1 \mathbf{v}_k^* \mathbf{b}(\omega_1) \\ \vdots \\ \beta_{\bar{n}} \mathbf{v}_k^* \mathbf{b}(\omega_{\bar{n}}) \end{bmatrix} + \mathbf{\Gamma} \mathbf{u}_k + \mathbf{u}_k E_k, \tag{12}$$

where $\mathbf{\Gamma} \in \mathbb{C}^{m \times m}$ is a known matrix. The matrix $\mathbf{\Gamma}$ will not be described because it has no importance for what follows. For details we refer to [1].

Let $\{w_k\}_{k=1}^n$ denote the frequencies of complex exponentials. To separate the terms corresponding to the component of interest from those associated with the interfering components in (12), we use the notation:

$$\begin{aligned}
\mathbf{A} &= \left[\ \mathbf{a}(\omega_1) \quad \dots \quad \mathbf{a}(\omega_n) \ \right] && (13)\\
\mathbf{x}_k &= \begin{bmatrix} \beta_1 \mathbf{v}_k^* \mathbf{b}(\omega_1) \\ \vdots \\ \beta_n \mathbf{v}_k^* \mathbf{b}(\omega_n) \end{bmatrix}, && (14)
\end{aligned}$$

for the component of interest, and similarly $\tilde{A}$ and $\tilde{x}_k$ for the other components which are the leakage signal in the subband. A compact matrix form of (12) for $\{k_1 \dots k_M\}$ is

$$\mathbf{Y} = \mathbf{AX} + \mathbf{\Gamma U} + \tilde{\mathbf{A}}\tilde{\mathbf{X}} + \mathbf{E}, \tag{15}$$

where matrices in (15) are defined as:

$$\begin{aligned}
\mathbf{Y} &= \left[\ \mathbf{u}_{k_1} Y_{k_1} \quad \dots \quad \mathbf{u}_{k_M} Y_{k_M} \ \right] && (16)\\
\mathbf{E} &= \left[\ \mathbf{u}_{k_1} E_{k_1} \quad \dots \quad \mathbf{u}_{k_M} E_{k_M} \ \right] && (17)\\
\mathbf{U} &= \left[\ \mathbf{u}_{k_1} \quad \dots \quad \mathbf{u}_{k_M} \ \right] && (18)\\
\mathbf{X} &= \left[\ \mathbf{x}_{k_1} \quad \dots \quad \mathbf{x}_{k_M} \ \right], && (19)
\end{aligned}$$

with $\mathbf{Y} \in \mathbb{C}^{m \times M}$. In (15) the second term is eliminated by postmultiplying with a projection matrix,

$$\mathbf{\Pi}_{\mathbf{U}}^{\perp} = \mathbf{I} - \mathbf{U}^*(\mathbf{U}\mathbf{U}^*)^{-1}\mathbf{U}, \tag{20}$$

which is the orthogonal projection matrix onto the null space of $\mathbf{U}$. The third and fourth terms in (15) are, respectively, the out-of-band components and the noise term.

3. PROPOSED METHOD F-MUSIC

The starting point for deriving the proposed estimation algorithm F-MUSIC is the following equation [1]:

$$\mathbf{Y}\mathbf{\Pi}_{\mathbf{U}}^{\perp} = \mathbf{AX}\mathbf{\Pi}_{\mathbf{U}}^{\perp} + \tilde{\mathbf{A}}\tilde{\mathbf{X}}\mathbf{\Pi}_{\mathbf{U}}^{\perp} + \mathbf{E}\mathbf{\Pi}_{\mathbf{U}}^{\perp}. \tag{21}$$

Let the matrix $\mathbf{Y}\mathbf{\Pi}_{\mathbf{U}}^{\perp}$ be decomposed into subspaces using either singular value decomposition (SVD) or eigenvalue decomposition (EVD), where the decomposed noise and signal subspaces are denoted as:

$$\mathbf{Y}\mathbf{\Pi}_{\mathbf{U}}^{\perp} = \left[\ \mathbf{S} \quad \mathbf{G} \ \right] \begin{bmatrix} \mathbf{\Sigma}_S & 0 \\ 0 & \mathbf{\Sigma}_G \end{bmatrix} \begin{bmatrix} \mathbf{V}_S^H \\ \mathbf{V}_G^H \end{bmatrix}. \tag{22}$$

Here, $\mathbf{\Sigma}_S$ denotes a diagonal matrix with the n largest singular values. Let $\mathbf{S}$ be the orthonormal signal subspace associated with n principal singular values, and $\mathbf{G}$ be the orthonormal noise singular vectors associated with $m - n$ singular values. The interference terms $\tilde{\mathbf{A}}\tilde{\mathbf{X}}$ and $\mathbf{E}$ will be attenuated by the projection matrix, but the least significant singular values will still provides us with a indication to the noise subspace. By definition, the spaces spanned by $\mathbf{S}$ and $\mathbf{G}$ are orthogonal. The vectors in the signal subspace can be modeled by the linearly independent Vandermonde vectors $\{\mathbf{a}(\omega_k)\}_{k=1}^n$.

The spectral F-MUSIC is formed by:

$$P(\omega) = \frac{1}{|\mathbf{a}^*(\omega)\mathbf{G}\mathbf{G}^*\mathbf{a}(\omega)|^2}, \tag{23}$$

where $\omega \in \left[\frac{2\pi}{N}k_1, \frac{2\pi}{N}k_M\right]$. In order to achieve high spectral resolution of the spectrogram, the frequency samples ω in (23) should be more closely sampled than the number of DFT points used in Y_k.

For the frequency estimation, the root F-MUSIC algorithm can be formed by determining the frequency estimates as the angular positions of the n roots of the equation

$$\mathbf{a}^T(z^{-1})\mathbf{G}\mathbf{G}^*\mathbf{a}(z) = 0, \tag{24}$$

which are located nearest to the unit circle. In (24), $\mathbf{a}(z)$ is given by

$$\mathbf{a}(z) = \left[\ z^{-1} \quad z^{-2} \quad \dots \quad z^{-m} \ \right]^T, \qquad z = e^{-j\omega}. \tag{25}$$

The algorithm can be summarized in the following steps:

1. Create $\mathbf{Y}\mathbf{\Pi}_{\mathbf{U}}^{\perp}$ from the observed data.
2. Calculate the SVD of $\mathbf{Y}\mathbf{\Pi}_{\mathbf{U}}^{\perp}$ to form $\mathbf{G}$ from singular vectors associated with the $m - n$ least significant singular values.

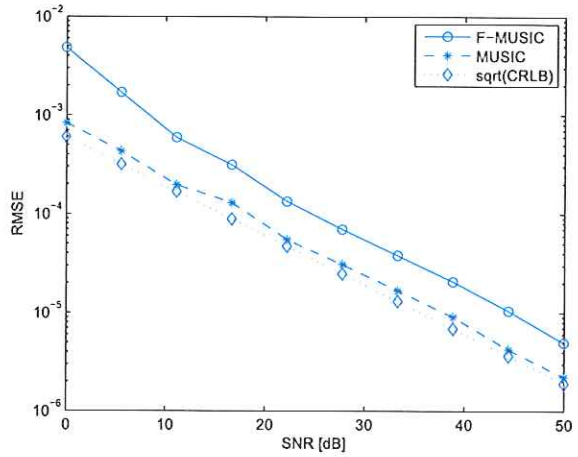

Fig. 1. *The calculated RMSE of F-MUSIC compared with MUSIC and CRLB versus different SNR with white noise.*

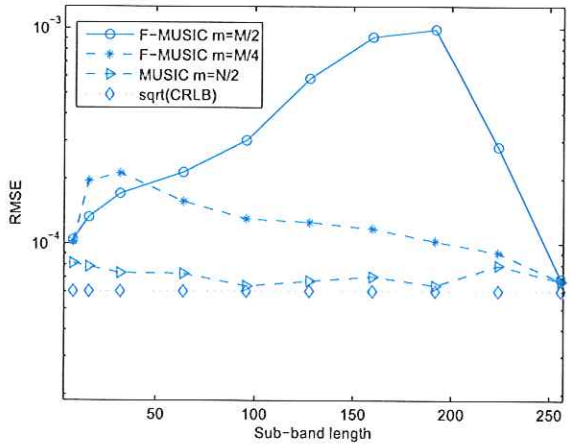

Fig. 2. *The calculated RMSE with different subband length, where the SNR with white noise is fixed to 20dB.*

3. a) Spectral F-MUSIC: Calculate (23) for $\omega \in \left[\frac{2\pi}{N}k_1, \frac{2\pi}{N}k_M\right]$.

   OR

   b) Root F-MUSIC: Determine frequency estimates as the angular positions of the n roots of (24).

4. NUMERICAL RESULTS

In this section we consider three numerical examples. The degradation of the estimation accuracy and the computational savings are demonstrated on the proposed algorithm in the case of signal embedded in white noise scenario. Selection of the user parameter of F-MUSIC for white noise scenario is investigated, and the robustness against colored noise will be shown.

The signal setup used in the following examples consists of one complex exponential embedded in noise with frequency $\omega_1 = 0.1$, and amplitude $\alpha_1 = 1$, and the observed data is a sequence of $N = 256$ samples. The calculated root mean square estimation errors (RMSE) of F-MUSIC are compared to both MUSIC and the asymptotic Cramér-Rao lower bound (CRLB). Signal-to-noise ratio (SNR) is calculated using the definition in [6], in our examples:

$$SNR = 10\log_{10}\frac{\alpha_1^2}{\phi(\omega_1)}, \tag{26}$$

with the function $\phi(\omega_1)$ being the power spectrum of the noise at frequency ω_1. The asymptotic CRLB defined in [1] in the case of one complex exponential embedded in noise is denoted as

$$CRLB = \frac{6\phi(\omega_1)}{N^3\alpha_1^2}. \tag{27}$$

To get a fair comparison between MUSIC and the proposed method, for each SNR, 100 Monte Carlo simulations were used with a uniform distributed phase between $[0, 2\pi]$ and a new realization of the noise is generated. The RMSE is computed for each simulation.

In the first example, a symmetric white Gaussian noise is used with the SNR calculated using (26); with $\phi(\omega_1)$ denoting the variance of the white noise. The subband of F-MUSIC was kept fixed to $k_1 = 0$ and $k_{32} = 31$, and by experience the user parameter is set to $m = M/2$ [1]. In MUSIC the model order of the covariance matrix is selected to be $N/2$. The calculated RMSE versus SNR is plotted in Fig. 1. It shows that the estimation performance of F-MUSIC is a bit worse than classical MUSIC in the case of white noise. However, for many applications, the degraded performance might still be attractive due to the computational savings. In F-MUSIC the computational complexity by selecting a subset of DFT samples with length M is of order $\mathcal{O}\left(M^3\right)$, where the complexity of MUSIC is of order $\mathcal{O}\left(N^3\right)$. Based on the selected subband length M, a relative amount of computational savings are achieved. In the case of full-band processing, F-MUSIC can still be computationally superior to the classical MUSIC. Instead of solving the estimation problem directly from the time-domain signal, the complexity can be reduced by splitting the estimation problem into solving subproblems on each subband. The total computational complexity of processing the entire signal will then be $\mathcal{O}\left(N^3/Q^2\right)$, where Q is the total number of equally spaced subbands. The trade off here is the estimation accuracies and the computational savings.

Following numerical example is used to demonstrate the selection of the user parameter evaluated on various subband length. Here SNR is fixed to 20dB and length of the subband is defined as $k_1 = 0$ and $k_M = M - 1$, where M varies from 8 to 256. The result is shown in Fig. 2. For many applications, only a course estimate of frequency is needed, in which the selection of M is not crucial. If, however, very accurate estimates are desired, the selection of m should be considered more carefully. Using the signal setup in this example, the selection of the user parameter m is dependent on the number of DFT points covered by the subband. In the case when all DFT samples are used in the estimation, the performance of F-MUSIC will then be the same as MUSIC.

In the last example, the robustness against colored noise will be demonstrated. The objective here is to show that by isolating the estimated signal in the subband, the interference from the colored noise will be reduced. The same signal setup is used except that the white noise is filtered with a second order AR process $1/(1 + 0.3z^{-1} + 0.8z^{-2})$ is selected to give a colored noise spectrum. One realization of the spectrum with SNR fixed at -3.6dB is shown in Fig. 3. SNR

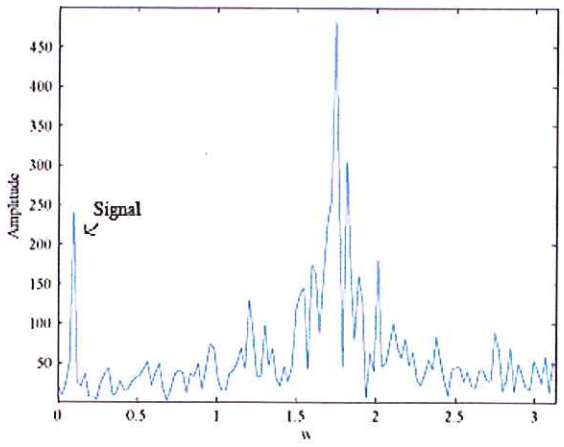

Fig. 3. *Magnitude spectrum of the signal in colored noise with SNR fixed at -3.6dB.*

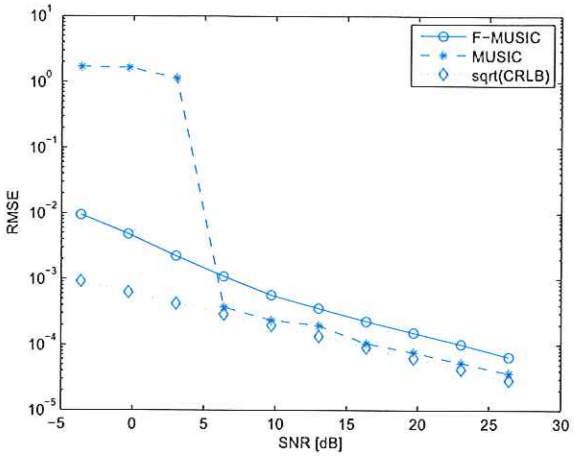

Fig. 4. *The calculated RMSE of F-MUSIC compared with MUSIC and CRLB, versus different SNR with colored noise.*

for the colored noise is calculated using (26) with function $\phi(\omega_1)$ being the power spectrum of the filtered noise at frequency ω_1. The subband is specified at $k_1 = 0$ to $k_{16} = 15$ which contains enough samples to estimate the embedded signal, while keeps the main part of the colored noise residing outside the subband. User parameters in F-MUSIC and the model order in MUSIC are selected to be $M/2$ and $N/2$, respectively. RMSE of F-MUSIC compared to MUSIC and the asymptotic colored CRLB are shown in Fig. 4. The graph shows that MUSIC algorithm fails to operate under low SNR with colored noise conditions while F-MUSIC can still achieve good estimates close to CRLB. The performance drop down of MUSIC can be explained by the increased probability off erroneous estimate of the signal subspace where part of the noise subspace is considered as the signal subspace. A problem sometimes referred to as subspace swapping; while F-MUSIC can avoid the problem by excluding interference signals from the FS data model. We believe that excluding the noise from the estimation problem as done here is a better approach than the noise pre-whitening. The F-MUSIC algorithm does not require knowledge of the noise statistic. Only frequency regions where the colored noise is concentrated on are of interest. This may be easier than directly estimate noise parameters from the observed signal, especially when the noise is non-stationary which is often the case for speech and audio signals.

5. CONCLUSIONS

In this paper, a subspace-based frequency estimator termed F-MUSIC has been proposed. This algorithm is a frequency-domain based frequency estimator uses subspaces decomposed from the FS data model, which is constructed from the observed DFT samples in the selected subband. The performance of F-MUSIC has been evaluated and compared to both MUSIC and CRLB. In low SNR with colored noise scenario, F-MUSIC outperforms MUSIC. From the simulations of F-MUSIC, we conclude that in general the price we pay for the reduced computational complexity and the increased robustness against colored noise is a slightly reduction in the estimation accuracy.

6. REFERENCES

[1] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, 2005.

[2] T. McKelvey and M. Viberg, "A robust frequency domain subspace algorithm for multi-component harmonic retrieval," in *Rec. of 34th Asilomar Conference on Signals, Systems and Computers*, 2001.

[3] J. Gunnarsson and T. McKelvey, "High SNR performance analysis of F-ESPRIT," in *Rec. of 38th Asilomar Conference on Signals, Systems and Computers*, 2004.

[4] M.G. Christensen, A. Jakobsson, and S.H. Jensen, "Joint high-resolution fundamental frequency and order estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 15, no. 5, pp. 1635–1644, 2007.

[5] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation," *Elsevier Signal Processing*, vol. 88(4), pp. 972–983, 2008.

[6] P. Stoica, A. Jakobsson, and J. Li, "Cisoid parameter estimation in the colored noise case: asymptotic cramer-rao bound, maximum likelihood, and nonlinear least-squares," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 45, no. 8, pp. 2048–2059, 1997.