of cumulants, we obtain the following expression as a function of the non-zero input cumulants and the unknown kernels

$$\begin{aligned} c_{y_{(1)}^{(1)}}(\tau) &= \sum_{i_1}\sum_{i_2} h_1^*(i_1)h_1(i_2)\,\mathrm{cum}\{z_{n-i_1}^*, z_{n-i_2+\tau}\} \\ &\quad + \mathrm{cum}\{\eta_n^*, \eta_{n+\tau}\} \\ &\quad + \sum_{i_1,j_1}\sum_{i_2,j_2} h_2^*(i_1,j_1)h_2(i_2,j_2) \\ &\quad \times \mathrm{cum}\{z_{n-i_1}^* z_{n-j_1}^*, z_{n-i_2+\tau}z_{n-j_2+\tau}\}. \end{aligned} \tag{36}$$

The first term of the above expression is simplified using the i.i.d. assumptions $\mathrm{cum}\{z_{n-i_1}^*, z_{n-i_2+\tau}\} = \gamma_{1,1}\delta(i_1 - i_2 + \tau)$ and $\mathrm{cum}\{\eta_n^*, \eta_{n+\tau}\} = c_{\eta_{(1)}^{(1)}}(\tau)\delta(\tau)$. The Leonov-Shiryaev theorem is used to manipulate cumulants involving products of random variables such as the second term of (36). Thus,

$$\begin{aligned} &\mathrm{cum}\{z_{n-i_1}^* z_{n-j_1}^*, z_{n-i_2+\tau}z_{n-j_2+\tau}\} = \gamma_{2,2} \\ &\delta(i_1 - j_1, i_1 - i_2 + \tau, i_1 - j_2 + \tau) \\ &\quad + 2\gamma_{1,1}\delta(i_1 - i_2 + \tau)\delta(j_1 - j_2 + \tau). \end{aligned}$$

If we substitute the above results in (36) we arrive at (31).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments.

REFERENCES

[1] Z. Ding and Y. Li, *Blind Equalization and Identification.*. Boca Raton, FL: CRC Press, 2001.
[2] J. Fang, A. Leyman, Y. Chew, and Y. Liang, "A cumulant interference subspace cancellation method for blind SISO channel estimation," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 784–790, Feb. 2006.
[3] S. Benedetto and S. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. Norwell, MA: Kluwer Academic, 1999.
[4] A. Stenger and R. Rabenstein, "Adaptive Volterra filters for nonlinear acoustic echo cancellation," in *IEEE Proc. NSIP*, 1999, pp. 679–683.
[5] L. Agarossi, S. Bellini, A. Canella, and P. Migliorati, "A Volterra model for the high density optical disc," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1998, pp. 1605–1608.
[6] R. Hermann, "Volterra modeling of digital magnetic saturation recording channels," *IEEE Trans. Magn.*, vol. 26, no. 5, pp. 2125–2127, Sep. 1990.
[7] M. Krob and M. Benidir, "Blind identification of a linear-quadratic model using higher order statistics," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1993, pp. 440–443.
[8] K. Abed-Meraim, A. Belouchiani, and Y. Hua, "Blind identification of a linear-quadratic mixture of independent components based on a joint diagonalization procedure," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1996, pp. 2718–2721.
[9] N. Petrochilos and P. Comon, "Blind identification of linear-quadratic channels with usual communication inputs," in *Proc. IEEE Workshop SSAP*, 2000, pp. 181–185.
[10] J. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, 1991.
[11] C. Spooner and W. Gardner, "The cumulant theory of cyclostationary time-series. II. Development and applications," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3409–3429, Dec. 1994.
[12] G. Giannakis and G. Zhou, "Harmonics in multiplicative and additive noise: Parameter estimation using cyclic statistics," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2217–2221, Sep. 1995.
[13] S. Chen, S. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, 1989.
[14] N. Kalouptsidis, *Signal Processing Systems: Theory & Design*. New York: Wiley, 1997.
[15] A. Carini, G. Sicuranza, and V. Mathews, "On the inversion of certain nonlinear systems," *IEEE Signal Process. Lett.*, vol. 4, no. 12, pp. 334–336, Dec. 1997.
[16] S. Alshebeili, A. Venetsanopoulos, and E. Cetin, "Cumulant based identification approaches for nonminimum phase FIR systems," *IEEE Trans. Signal Process.*, vol. 41, no. 4, pp. 1576–1588, Apr. 1993.
[17] V. Leonov and A. Shiryaev, "On a method of calculation of semi-invariants," *Theor. Probab. Appl.*, vol. 4, pp. 319–329, 1959.

MMSE-Based MDL Method for Robust Estimation of Number of Sources Without Eigendecomposition

Lei Huang, Teng Long, Erke Mao, and H. C. So

***Abstract*—It is well known that the conventional eigenvalue-based minimum description length (MDL) approach for source number estimation suffers from high computational load and performs optimally only in the presence of spatially and temporally white noise. To improve the robustness of the MDL methodology, we propose to utilize the minimum mean square error (MMSE) of the multistage Wiener filter to calculate the required description length for encoding the observed data, instead of relying on the eigenvalues of the data covariance matrix. As there is no need to calculate the covariance matrix and its eigenvalue decomposition, our derived MMSE-based MDL (mMDL) method is also more computationally efficient than the traditional counterparts. Numerical examples are included to demonstrate the robustness of the mMDL detector in nonuniform noise.**

***Index Terms*—Eigenvalue decomposition (EVD), minimum description length (MDL), multistage Wiener filter (MSWF), sensor array processing, source number estimation.**

I. INTRODUCTION

Development of computationally efficient and robust methods for source number estimation is of significant interest in the field of array processing [1]–[6]. One reason is that the involved computational load of the classical source enumeration methods based on Akaike information criterion and minimum description length (MDL) is quite heavy particularly for a large array, making real-time processing infeasible. Another reason is that the required assumption of spatially and temporally white noise in the conventional schemes might not be valid in practical situations due to changing noise environment, receiver nonidealities and sensor nonuniformities [7], [8]. In the presence of unknown nonuniform noise, the classical eigenvalue-based MDL detectors, which exploit the multiplicity of the smallest eigenvalues of the estimated data covariance matrix, will fail to achieve reliable source enumeration.

Although many methods [1]–[3] have been proposed for robust source enumeration, they need to be further improved in terms of computational complexity. As eigenvectors of the data covariance matrix also contains source number information but is less sensitive to the noise models, eigenvector-based methods [1], [2] have been developed to accurately estimate the number of sources in nonuniform

Manuscript received March 03, 2009; accepted April 27, 2009. First published May 27, 2009; current version published September 16, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Biao Chen. This work was supported by the Natural Science Foundation of China (NSFC) under Grant 60702068.

L. Huang, T. Long, and E. Mao are with the Department of Electronic Engineering, Beijing Institute of Technology, Beijing 100081, China (e-mail: lhuang8sasp@hotmail.com).

H. C. So is with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, China (e-mail: hcso@ee.cityu.edu.hk).

Digital Object Identifier 10.1109/TSP.2009.2024043

noise environment. However, they need to calculate the observed covariance matrix and its eigenvalue decomposition (EVD), which corresponds to a computationally intensive task. The robust-MDL (rMDL) method proposed by Fishler and Poor [3] is robust against both spatial and statistical mismodeling but it involves N iterations, where N is the sensor number, and each iteration needs EVD computation. As the complexity is of $O(N^4)$ flops, the rMDL detector is also computationally burdensome especially for a large array. Recently, we propose a Gerschgorin disk source enumerator [4] that does not require EVD. Although it is more robust and computationally efficient than the classical MDL methods, as in [1], its detection performance relies on a nonincreasing function that needs to be carefully designed for practical applications.

In this correspondence, source enumeration in a computationally efficient as well as robust manner is addressed. Unlike the eigenvalue-based MDL methods, we propose an alternative MDL approach which uses the minimum mean square error (MMSE) of the multistage Wiener filter (MSWF) to calculate the minimum description length for encoding the observed data. As a result, the latter requires a lower computational cost than the classical schemes because calculation of the observed covariance matrix and its EVD is not required. Moreover, the proposed MMSE-based MDL (mMDL) detector is superior in terms of robustness to nonuniform noise as it does not rely on the eigenvalues of the data covariance matrix. Note that the mMDL methodology is different from our previous work of [6] since it is based on the variances of the desired signals of the MSWF and is designed for spatially and temporally white noise environments.

II. Data Model

Consider an array with N sensors receiving q narrowband far-field sources from distinct directions $\theta_1, \ldots, \theta_q$. For simplicity, we assume that the array and sources are in the same plane. In the sequel, the ℓth snapshot vector consisting of the sensor array outputs, excluding the last sensor output, is written as

$$\boldsymbol{x}(t_\ell) = [x_1(t_\ell), \ldots, x_M(t_\ell)]^T = \boldsymbol{A}(\boldsymbol{\theta})\,\boldsymbol{s}(t_\ell) + \boldsymbol{n}(t_\ell) \tag{1}$$

where $M = N - 1$, $(\cdot)^T$ denotes the transpose operation and $\boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_q)]$, $\boldsymbol{s}(t_\ell) = [s_1(t_\ell), \ldots, s_q(t_\ell)]^T$, and $\boldsymbol{n}(t_\ell) = [n_1(t_\ell), \ldots, n_M(t_\ell)]^T$ are the $M \times q$ steering matrix, $q \times 1$ source waveform vector and $M \times 1$ sensor noise vector, respectively. Note that the last sensor output of the array is not included in (1) because in the proposed method developed in Section III, the first M sensor outputs and last sensor output of the array are used as the observed data and reference signal of the MSWF, respectively.

For a uniform linear array (ULA) with inter-sensor spacing d, the steering vector can be of the form $\boldsymbol{a}(\theta_i) = \left[1, e^{j2\pi d \sin(\theta_i)/\lambda}, \ldots, e^{j2\pi d(M-1)\sin(\theta_i)/\lambda}\right]^T$ $(i = 1, \ldots, q)$, where λ denotes the wavelength and $q(q < M)$ denotes the *unknown* number of sources. The source waveform $s_i(t_\ell)$ $(i = 1, \ldots, q)$ is assumed to be a jointly stationary, statistically uncorrelated, zero-mean complex Gaussian random process. The additive noise $\boldsymbol{n}(t_\ell)$ is assumed to be a spatially and temporally white complex Gaussian process, uncorrelated with the sources, with mean zero and covariance matrix $\sigma_n^2 \boldsymbol{I}_M$ where $\boldsymbol{I}_M$ denotes the $M \times M$ identity matrix.

Under these assumptions, the observed data $\boldsymbol{x}(t_\ell)$ is a complex Gaussian random process with mean zero and covariance matrix given by

$$\boldsymbol{R}_x = E\left[\boldsymbol{x}(t_\ell)\boldsymbol{x}^H(t_\ell)\right] = \boldsymbol{A}(\boldsymbol{\theta})\,\boldsymbol{R}_s \boldsymbol{A}^H(\boldsymbol{\theta}) + \sigma_n^2 \boldsymbol{I}_M \tag{2}$$

where $E[\cdot]$ is the expectation operator, $(\cdot)^H$ is the Hermitian transpose and $\boldsymbol{R}_s = E\left[\boldsymbol{s}(t_\ell)\boldsymbol{s}^H(t_\ell)\right]$ is the signal covariance matrix. In the practical applications, however, only a finite number of snapshots, say, L, is available. In the sequel, the sample covariance matrix based on L snapshots is calculated as $\hat{\boldsymbol{R}}_x = (1/L)\sum_{\ell=1}^{L} \boldsymbol{x}(t_\ell)\boldsymbol{x}^H(t_\ell)$.

III. Proposed MMSE-Based MDL Estimator

A. Derivation

It is shown in [9] and [10] that, for a given data set and a family of probabilistic models, the MDL principle is to select the model that offers the shortest description length of the data. Specifically, given an observed data set $\boldsymbol{X} = \{\boldsymbol{x}(t_\ell)\}_{\ell=1}^{L}$ and a probabilistic model $p(\boldsymbol{X}|\boldsymbol{\Theta})$, where $\boldsymbol{\Theta}$ denotes an unknown parameter vector of the model, the shortest description length required to encode the data using the model can be asymptotically written as

$$\mathcal{L}\{\boldsymbol{x}(t_\ell)\} = -\log p(\boldsymbol{X}|\hat{\boldsymbol{\Theta}}) + \frac{1}{2} K \log L \tag{3}$$

where $\hat{\boldsymbol{\Theta}}$ is the maximum likelihood (ML) estimate of $\boldsymbol{\Theta}$ and K denotes the number of free parameters in $\hat{\boldsymbol{\Theta}}$. Since the observed data $\{\boldsymbol{x}(t_\ell)\}$ are assumed to be statistically independent complex Gaussian random vectors with zero means, their joint probability density is

$$p(\boldsymbol{X}|\boldsymbol{\Theta}) = \prod_{\ell=1}^{L} \frac{1}{\pi^q \det(\boldsymbol{R}_x)} \exp\left\{-\boldsymbol{x}(t_\ell)^H \boldsymbol{R}_x^{-1} \boldsymbol{x}(t_\ell)\right\} \tag{4}$$

where $\det(\cdot)$ is the determinant operation. Taking natural logarithm on (4) and omitting the terms independent of $\boldsymbol{\Theta}$, we define the negative log-likelihood function as

$$\mathcal{F}(\boldsymbol{\Theta}) = L \log \det(\boldsymbol{R}_x) + \operatorname{tr}\left(\boldsymbol{R}_x^{-1}\hat{\boldsymbol{R}}_x\right) \tag{5}$$

where tr represents the trace operator.

To proceed to the derivation of the mMDL method, we need to use the following results of the MSWF. The MSWF with the reference signal $d_0(t_\ell) = x_{M+1}(t_\ell)$ and observed data $\boldsymbol{x}_0(t_\ell) = \boldsymbol{x}(t_\ell)$ is given in Appendix A.

Lemma 1: The determinant of $\boldsymbol{R}_{x_0}$ is equal to that of $\boldsymbol{R}_e$, i.e.,

$$\det(\boldsymbol{R}_{x_0}) = \det(\boldsymbol{R}_e) = \prod_{i=1}^{M} \rho_i \tag{6}$$

where $\boldsymbol{R}_{x_0} \triangleq E\left[\boldsymbol{x}_0(t_\ell)\boldsymbol{x}_0^H(t_\ell)\right]$, $\boldsymbol{R}_e \triangleq \operatorname{diag}\left([\rho_1, \ldots, \rho_M]^T\right)$, and $\rho_i \triangleq E\left[|e_i(t_\ell)|^2\right]$ $(i = 1, \ldots, M)$ which is the MMSE at the ith stage of the MSWF. Note that explicit calculation of $\rho_i (i = 1, \ldots, M)$ is provided in Appendix A.

Proof: The proof of Lemma 1 is provided in Appendix B. ■

Lemma 2: The MMSEs of the MSWF satisfy

$$\rho_1 \geq \cdots \geq \rho_q > \rho_{q+1} = \cdots = \rho_M = \sigma_n^2. \tag{7}$$

Proof: The proof of Lemma 2 is provided in Appendix C. ■

In the sequel, updating $\boldsymbol{R}_x$ by $\boldsymbol{R}_{x_0}$, substituting (6) and (7) into (5), and assuming that k is the *supposed* number of sources, we obtain

$$\mathcal{F}(\boldsymbol{\Theta}) = L \log\left(\prod_{i=1}^{k} \rho_i \times \prod_{i=k+1}^{M} \sigma_n^2\right) + \operatorname{tr}\left(\boldsymbol{R}_{x_0}^{-1}\hat{\boldsymbol{R}}_{x_0}\right). \tag{8}$$

It follows from (B5) that $\boldsymbol{R}_{x_0}$ can be reexpressed as

$$\boldsymbol{R}_{x_0} = \boldsymbol{H}\left(\boldsymbol{W}^H\right)^{-1} \boldsymbol{R}_e \boldsymbol{W}^{-1} \boldsymbol{H}^H \tag{9}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M]$, $\boldsymbol{h}_i$ is the ith matched filter of the WSWF, and $\boldsymbol{W}$ is defined in (B4). Note that $\boldsymbol{H}$ is an unitary matrix due to the orthogonality of $\boldsymbol{h}_i$. As a consequence, the parameter vector is

$$\boldsymbol{\Theta}^T = \left[\rho_1, \ldots, \rho_k, \sigma_n^2, w_1, \ldots, w_k, \boldsymbol{h}_1^T, \ldots, \boldsymbol{h}_k^T\right] \quad (10)$$

where w_i $(i = 1, \ldots, k)$ are the scalar Wiener filter coefficients in the backward recursion of the MSWF. Actually, not all the parameters are independent of each other. It follows from the MSWF in Appendix A that both ρ_i and w_i rely on the desired signals $d_i(t_\ell)$ which are obtained by filtering the observed data with the matched filter $\boldsymbol{h}_i$, i.e., $d_i(t_\ell) = \boldsymbol{h}_i^H \boldsymbol{x}_0(t_\ell)$. In the sequel, the parameter vector is finally reduced to be

$$\boldsymbol{\Theta}^T = \left[\sigma_n^2, \boldsymbol{h}_1^T, \ldots, \boldsymbol{h}_k^T\right]. \quad (11)$$

Meanwhile, notice that the matched filters are the orthogonal and normalized vectors, which lead to a reduction of $(2k + 2(1/2)k(k-1))$ in the free parameter number. As a result, the number of free parameters in $\boldsymbol{\Theta}$ can be counted as

$$\begin{aligned} K &= 2Mk + 1 - 2k - 2\left(\frac{1}{2}\right)k(k-1) \\ &= k(2M - k - 1) + 1. \end{aligned} \quad (12)$$

In Appendix D, it is shown that the estimated MMSE yielded by the backward recursion of the MSWF, i.e., $\hat{\rho}_i = \hat{\sigma}_{d_i}^2 - |\hat{\delta}_{i+1}|^2/\hat{\rho}_{i+1}$, is the ML estimate of ρ_i. Consequently, it follows from (7) that the ML estimate of σ_n^2 is

$$\hat{\sigma}_n^2 = \frac{1}{M-k} \sum_{i=k+1}^{M} \hat{\rho}_i. \quad (13)$$

Therefore, substituting the ML estimates of ρ_i $(i = 1, \ldots, k)$, σ_n^2 and $\boldsymbol{R}_{x_0}$ into (8) and (3), omitting the constant terms and applying the same argument used in [9] yields

$$\begin{aligned} \mathcal{L}\{\boldsymbol{x}(t_\ell)\} &= L \log\left(\prod_{i=1}^{k} \hat{\rho}_i \times \prod_{i=k+1}^{M} \hat{\sigma}_n^2\right) \\ &\quad + \frac{1}{2}k(2M - k - 1)\log L \\ &\triangleq \mathcal{F}(k) + \mathcal{P}(k) \end{aligned} \quad (14)$$

where

$$\mathcal{F}(k) = L(M-k)\log\left(\frac{\frac{1}{(M-k)}\sum_{i=k+1}^{M}\hat{\rho}_i}{\left(\prod_{i=k+1}^{M}\hat{\rho}_i\right)^{1/(M-k)}}\right) \quad (15)$$

$$\mathcal{P}(k) = \frac{1}{2}k(2M - k - 1)\log L \quad (16)$$

are the negative log-likelihood function and the penalty function, respectively. Thus, the number of sources can be obtained by minimizing our proposed mMDL criterion

$$\hat{q} = \arg\min_{k=0,1,\ldots,M-1} \mathrm{mMDL}(k) \quad (17)$$

where $\mathrm{mMDL}(k) = \mathcal{F}(k) + \mathcal{P}(k)$.

B. Reduced-Rank MMDL Method

It is implied in (15)–(17) that the mMDL estimator can be constructed provided that the number of the smallest MMSEs is not less than two. This thereby indicates that a reduced number of the MMSEs, say $D(q + 2 \le D < M)$, is sufficient to form a reduced-rank mMDL method for source enumeration with an even lower computational cost. The reduced-rank mMDL method is formulated as

$$\hat{q} = \arg\min_{k=0,1,\ldots,D-1} \mathrm{mMDL}^{(D)}(k) \quad (18)$$

where $\mathrm{mMDL}^{(D)}(k) = \mathcal{F}^{(D)}(k) + \mathcal{P}^{(D)}(k)$ with

$$\mathcal{F}^{(D)}(k) = L(D-k)\log\left(\frac{\frac{1}{(D-k)}\sum_{i=k+1}^{D}\hat{\rho}_i}{\left(\prod_{i=k+1}^{D}\hat{\rho}_i\right)^{1/(D-k)}}\right) \quad (19)$$

$$\mathcal{P}^{(D)}(k) = \frac{1}{2}k(2D - k - 1)\log L. \quad (20)$$

It is important to correctly determine the reduced rank D so that the reduced-rank mMDL estimator can correctly enumerate the sources. To this end, an adaptive detector of D can be defined, similar to [4] and [6], by

$$D = \max\left\{i : |\hat{\delta}_i| > \epsilon\right\} \quad (21)$$

where ϵ is a small positive constant.

In Appendix E, we have proven that the reduced-rank mMDL method offers the property of strong consistency. That is, as the number of snapshots tends to infinity, it correctly estimates the source number with probability one.

C. Computational Requirement

It is well known that the EVD-based MDL methods [9] necessarily involve the estimated covariance matrix and its EVD, requiring around $O(N^2L) + O(N^3)$ flops. To correctly detect the sources, the rMDL method [3] does not terminate the iterative procedure until a stationary point is reached, generally requiring N iterations, and each iteration includes the EVD of an updated covariance matrix. As a result, the rMDL needs around $O(N^4)$ flops in EVD computation and $O(N^2L)$ flops in calculating the observed covariance matrix. However, in the mMDL method, the MMSEs $\rho_i (i = 1, \ldots, D)$ can be directly obtained from the recursions of the MSWF, avoiding the calculation of the observed covariance matrix and its EVD. Meanwhile, note that the forward recursion only involves complex vector-vector products, and does not include any complex matrix-vector products, thereby requiring around $O(M)$ flops for each snapshot and each stage. In the sequel, after performing D $(q + 2 \le D < M)$ forward recursions, the required computational cost is only about $O(DML)$ flops. On the other hand, notice that the backward recursion only involves complex scalar-scalar products that are fiddling in computational complexity. Therefore, the mMDL method eventually requires only around $O(DML)$ flops that is much less than that of the EVD-based MDL methods [9] and the rMDL method [3], particularly when $D \ll M$.

IV. Numerical Results

We consider a ULA of ten sensors with $d = \lambda/2$ and assume two narrowband and uncorrelated sources with equal power impinging upon the array from the directions $[\theta_1, \theta_2] = [2.5°, 7.8°]$. To evaluate the insensitiveness to the Gaussian and non-Gaussian assumptions, two cases are considered as follows. The first case corresponds to complex Gaussian sources with mean zero and variance σ^2 and we examine complex Laplace sources [13] whose real and imaginary components are independent random variables with mean zero and variance

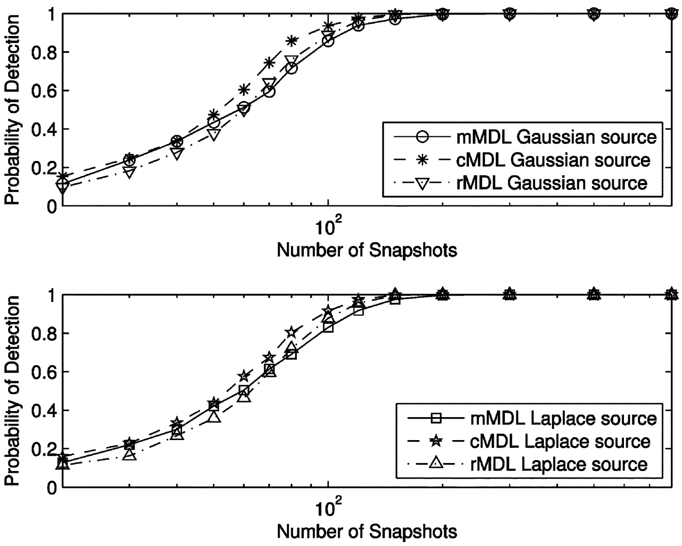

Fig. 1. Probability of correct detection versus number of snapshots in spatially and temporally white noise. $[\theta_1, \theta_2] = [2.5°, 7.8°]$, SNR $= -3$ dB, $N = 10$ and $\epsilon = 0.05$.

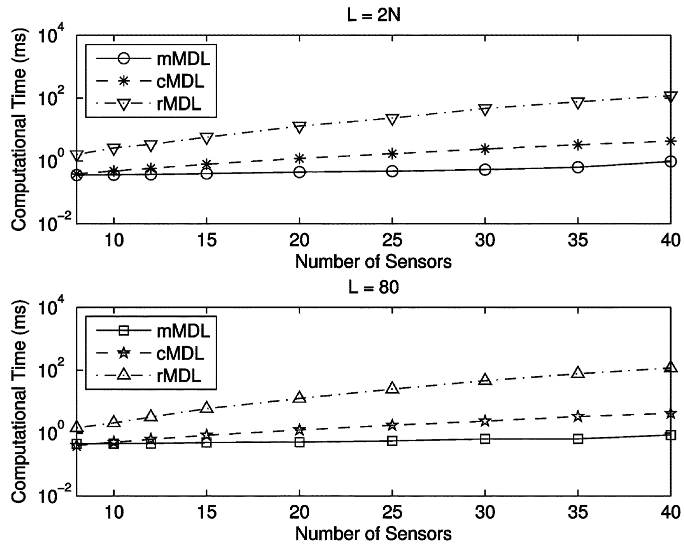

Fig. 2. Computational time versus number of sensors. $[\theta_1, \theta_2] = [2.5°, 7.8°]$, SNR $= -3$ dB, $D = 4$ and all sources are Gaussian.

σ^2 in the second case. Accordingly, the univariate Laplace distribution is given as $p_Z(z) = \left(\sqrt{2}/2\sigma\right) e^{\left(-\sqrt{2}|z|/\sigma\right)}$.

We first consider the spatially and temporally white noise. In this case, the signal-to-noise ratio (SNR) is defined as the ratio of the power of signals to the power of noise at each sensor. Five hundred independent trials have been run to obtain the empirical results of the proposed reduced-rank mMDL, classical MDL (cMDL) [9] and rMDL [3] methods. Fig. 1 shows the empirical probabilities of correct detection versus the number of snapshots. When the number of snapshots tends to infinity, it is observed that the mMDL, cMDL and rMDL methods converge to one in probability of correct detection for both Gaussian and Laplace sources, thereby indicating that all the MDL methods are consistent and work well for the spatially and temporally white noise model. Nevertheless, when the number of snapshots is $L < 200$, the cMDL method has a higher detection probability than the mMDL and rMDL schemes. As pointed out in [3], the rMDL method is less accurate than the cMDL method because it ignores the *a priori* knowledge that the sensor noise is spatially and temporally white. Note that the mMDL method only employs 9 sensor outputs as the observed data to calculate the MMSEs of the MSWF, reducing the aperture of the array from 10 to 9, and is thereby less accurate than the cMDL and rMDL methods as the number of snapshots varies from 60 to 200. When $L \leq 60$, however, the proposed mMDL algorithm outperforms the rMDL method and its detection performance is very close to that of the cMDL scheme. The computational times versus number of sensors for $L = 2N$ and $L = 80$ are depicted in Fig. 2. Without the need of calculating the observed covariance matrix and its EVD, the mMDL method requires much less computational time than those of the cMDL and rMDL approaches. As a result, we see that the mMDL method is computationally simple at the expense of small degradation in detection performance for spatially and temporally white noise.

To examine the robustness against the deviations from the assumption of spatially and temporally white noise, five hundred independent trials have been run to calculate the probability of correct detection in the presence of nonuniform noise. Similar to [3], we use the following noise covariance matrix:

$$\sigma_n^2 \boldsymbol{I}_N + \left(\frac{\sigma_n^2}{2}\right) \mathrm{diag}([-0.9, -0.7, -0.5, -0.3, -0.1, 0.1, 0.3, 0.5, 0.7, 0.9]^T) \quad (22)$$

to simulate the deviations of up to 3 dB from the normal noise power level σ_n^2. In this case, SNR is defined as the ratio of the power of signals to the *averaged* power of noises. Fig. 3 depicts the empirical probabilities of correct detection versus the number of snapshots. It is observed that the mMDL and rMDL methods can correctly detect the sources as the number of snapshots increases and their outstanding performance is independent of the source distribution. The eigenvalue-based cMDL method, however, fails to correctly enumerate the sources as the number of snapshots is greater than 100, and converges to zero in probability of correct detection when L is sufficiently large. This phenomenon can be easily understood by noticing that the cMDL method relies on the equality of the smallest eigenvalues. The unequal noise powers make the smallest eigenvalues to be different, as depicted in Fig. 4 where the MMSEs versus the rank of the MSWF and the eigenvalues versus the number of sensors are shown, resulting in more sources or the so-called "virtual" sources [3] to be detected in the classical scheme. When the number of snapshots is not large enough, the cMDL method cannot detect the differences in the smallest eigenvalues, and thereby will not detect the "virtual" sources. As the number of snapshots is large enough, however, the "virtual" sources are detected by the cMDL method, thereby leading to an error event. The enhanced robustness of the mMDL method can be readily interpreted by examining Fig. 4. The deviations make the smallest eigenvalues to be significantly unequal, as noted above, but scarcely affect the smallest MMSEs. As a result, the mMDL method, which is based on the multiplicity of the smallest MMSEs, is more robust to the nonuniform noise than the eigenvalue-based cMDL method. Fig. 5 depicts the empirical probability of correct detection versus the angle separation. As an eigenvalue-based method, the cMDL algorithm fails to correctly detect the sources no matter how large the angle separation is. On the other hand, as the mMDL detector only uses the MMSEs of the MSWF instead of the eigenvalues, it is more robust against the deviations than the cMDL method for both Gaussian and Laplace sources.

V. Discussion and Conclusion

We have devised the mMDL method for source enumeration in this correspondence. Since the mMDL method only needs to calculate the MMSEs by the recursions of the MSWF, and avoids to compute the observed covariance matrix and its EVD, making it to be computationally attractive. Meanwhile, the unequal noise power levels in nonuniform

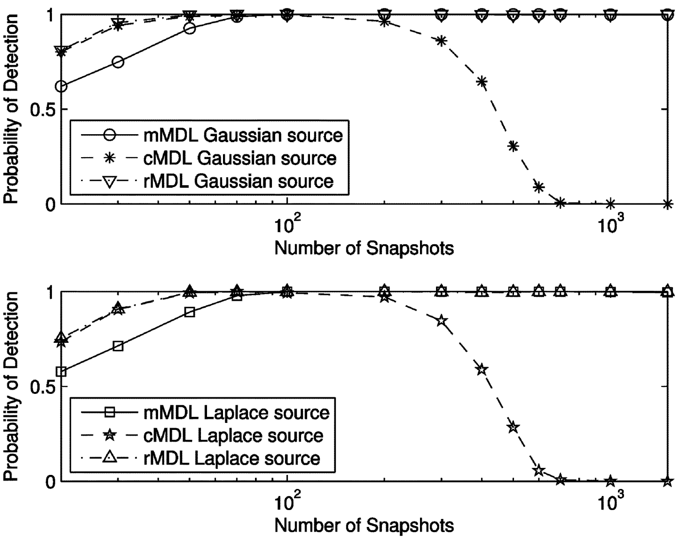

Fig. 3. Probability of correct detection versus number of snapshots in nonuniform noise. $[\theta_1, \theta_2] = [2.5^\circ, 7.8^\circ]$, SNR $= 0$ dB, $N = 10$ and $\epsilon = 0.1$.

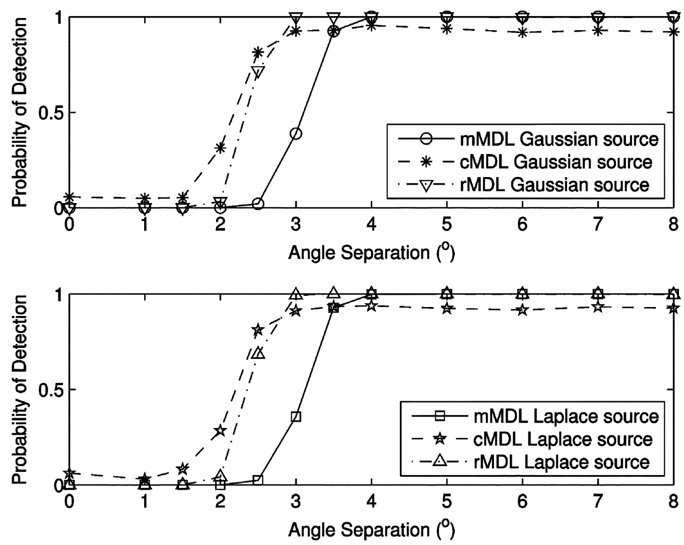

Fig. 5. Probability of correct detection versus angular separation in nonuniform noise. $[\theta_1, \theta_2] = [2.5^\circ, 2.5^\circ + \Delta\theta]$, SNR $= 0$ dB, $N = 10$, $L = 250$ and $\epsilon = 0.1$.

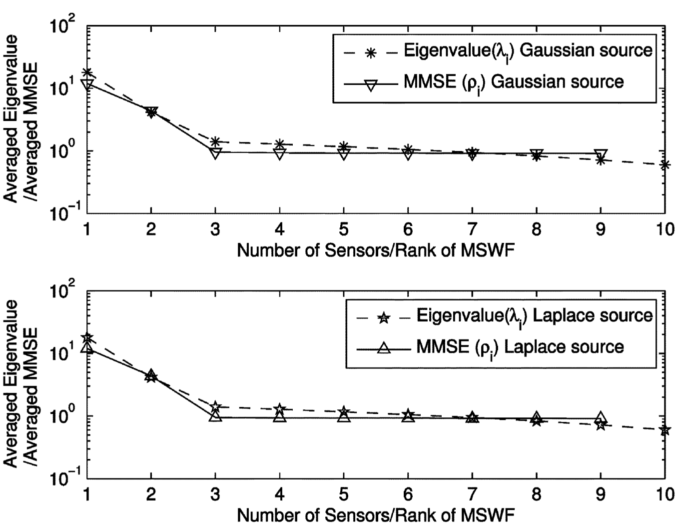

Fig. 4. Averaged MMSE of the MSWF, $\overline{\rho}_i$, versus rank of MSWF and averaged eigenvalue, $\overline{\lambda}_i$, versus number of sensors. $[\theta_1, \theta_2] = [2.5^\circ, 7.8^\circ]$, SNR $= 0$ dB, $N = 10$ and $L = 1500$.

noise only make the smallest eigenvalues to be significantly different but scarcely affect the smallest MMSEs, the mMDL method is therefore more robust against the deviations from the assumption of spatially and temporally white noise than the eigenvalue-based MDL methods. However, when the deviations become more severe, say, larger than 3 dB, the mMDL method might not provide reliable estimate of the number of sources because such large deviations may lead to severe fluctuations of the MMSEs. Developing an efficient strategy to further enhance the robustness of the mMDL method will be our future work.

Appendix A
MSWF

The MSWF was first proposed by Goldstein *et al.* [11] to solve the classical Wiener filtering problem with reduced computational complexity. A full-rank MSWF algorithm is given as follows:

Initialization:

$d_0(t_\ell) = x_{M+1}(t_\ell),$

$\boldsymbol{x}_0(t_\ell) = [x_1(t_\ell), \ldots, x_M(t_\ell)]^T.$

Forward Recursion: For $i = 1, \ldots, M$:

$\boldsymbol{r}_{x_{i-1}d_{i-1}} = E\left[\boldsymbol{x}_{i-1}(t_\ell) d_{i-1}^*(t_\ell)\right];$

$\delta_i = \|\boldsymbol{r}_{x_{i-1}d_{i-1}}\|_2,$

$\boldsymbol{h}_i = \boldsymbol{r}_{x_{i-1}d_{i-1}} / \delta_i;$

$d_i(t_\ell) = \boldsymbol{h}_i^H \boldsymbol{x}_{i-1}(t_\ell), \quad \sigma_{d_i}^2 = E\left[|d_i(t_\ell)|^2\right];$

$\boldsymbol{x}_i(t_\ell) = \boldsymbol{x}_{i-1}(t_\ell) - \boldsymbol{h}_i d_i(t_\ell).$

Backward Recursion: For $i = M, \ldots, 1$ with $\rho_M = E\left[|d_M(t_\ell)|^2\right]$ and $e_M(t_\ell) = d_M(t_\ell)$:

$w_i = \delta_i / \rho_i;$

$e_{i-1}(t_\ell) = d_{i-1}(t_\ell) - w_i^* e_i(t_\ell);$

$\rho_{i-1} = \sigma_{d_{i-1}}^2 - |\delta_i|^2 / \rho_i.$

Here, we use $\|\cdot\|_2$ and $|\cdot|$ to denote the Euclidean norm of a vector and the absolute value of a number, respectively. The desired signal $d_i(t_\ell)$ is obtained by filtering the observed data $\boldsymbol{x}_{i-1}(t_\ell)$ with the matched filters $\{\boldsymbol{h}_i\}$, but annihilated in the calculation of the next observed data $\boldsymbol{x}_i(t_\ell)$. To obtain the output of the MSWF, we only need to linearly combine the outputs of the matched filters $\{\boldsymbol{h}_i\}$ with the scalar Wiener filters $\{w_i\}$.

Appendix B
Proof of Lemma 1

After performing M forward recursions, we obtain M desired signals of the MSWF, which are collected as

$$\boldsymbol{d}(t_\ell) \triangleq [d_1(t_\ell), \ldots, d_M(t_\ell)]^T = \boldsymbol{H}^H \boldsymbol{x}_0(t_\ell) \tag{B1}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M]$. It is shown in [4], [11], that the covariance matrix of the desired signals, given by

$$\boldsymbol{R}_d \triangleq E[\boldsymbol{d}(t_\ell)\boldsymbol{d}^H(t_\ell)] = \boldsymbol{H}^H \boldsymbol{R}_{x_0} \boldsymbol{H} \tag{B2}$$

is a tridiagonal matrix. Meanwhile, it is indicated in [11] that the backward recursion is equivalent to a procedure of a Gram-Schmidt transform that diagonalizes the tridiagonal matrix $\boldsymbol{R}_d$, yielding the uncorrelated errors of the MSWF

$$\boldsymbol{e}(t_\ell) = \boldsymbol{W}^H \boldsymbol{d}(t_\ell) = \boldsymbol{W}^H \boldsymbol{H}^H \boldsymbol{x}_0(t_\ell), \tag{B3}$$

where $\boldsymbol{W}$ is the Gram–Schmidt operator which has the form of

$$\boldsymbol{W} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -w_2 & 1 & \cdots & 0 & 0 \\ w_2 w_3 & -w_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (-1)^M \prod_{i=2}^{M-1} w_i & (-1)^{M-1} \prod_{i=3}^{M-1} w_i & \cdots & 1 & 0 \\ (-1)^{M+1} \prod_{i=2}^{M} w_i & (-1)^{M} \prod_{i=3}^{M} w_i & \cdots & -w_M & 1 \end{pmatrix}. \tag{B4}$$

In the sequel, the covariance matrix of the uncorrelated errors is a diagonal matrix, given as

$$\begin{aligned} \boldsymbol{R}_e &\triangleq E\left[\boldsymbol{e}(t_\ell)\boldsymbol{e}^H(t_\ell)\right] \\ &= \operatorname{diag}\left([\rho_1, \ldots, \rho_M]^T\right) = \boldsymbol{W}^H \boldsymbol{H}^H \boldsymbol{R}_{x_0} \boldsymbol{H} \boldsymbol{W} \end{aligned} \tag{B5}$$

where $\rho_i = E[e_i(t_\ell)e_i^*(t_\ell)]$ can be recursively calculated by the backward recursion of the MSWF. Therefore, using the multiplicative property of determinants, i.e., $\det(\boldsymbol{U} \times \boldsymbol{V}) = \det(\boldsymbol{U}) \times \det(\boldsymbol{V})$ for $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{C}^{M \times M}$ and noticing that $\boldsymbol{H}$ is a unitary matrix, we obtain

$$\begin{aligned} \det(\boldsymbol{R}_e) &= \det\left(\boldsymbol{W}^H\right) \times \det\left(\boldsymbol{H}^H\right) \\ &\quad \times \det(\boldsymbol{R}_{x_0}) \times \det(\boldsymbol{H}) \times \det(\boldsymbol{W}) \\ &= \det(\boldsymbol{R}_{x_0}) = \prod_{i=1}^{M} \rho_i \end{aligned} \tag{B6}$$

which completes the proof of Lemma 1.

Appendix C
Proof of Lemma 2

It follows from the MSWF given in Appendix A that

$$\begin{aligned} \delta_i^2 &= \|\boldsymbol{r}_{x_{i-1}d_{i-1}}\|_2^2 = \boldsymbol{r}_{x_{i-1}d_{i-1}}^H \boldsymbol{r}_{x_{i-1}d_{i-1}} \\ &= E\left\{\boldsymbol{x}_{i-1}^H(t_\ell) d_{i-1}(t_\ell)\right\} \boldsymbol{h}_i \delta_i = E\left[d_i^*(t_\ell) d_{i-1}(t_\ell)\right] \delta_i. \end{aligned} \tag{C1}$$

In the sequel, we obtain

$$\delta_i = E\left[d_i^*(t_\ell) d_{i-1}(t_\ell)\right]. \tag{C2}$$

Using the results in [6] that the last $(M-q)$ desired signals of the MSWF are uncorrelated whereas the first q desired signals are correlated with the adjacent desired signals, we obtain $\delta_i \neq 0$ $(i = 1, \ldots, q)$ and $\delta_i = 0$ $(i = q+1, \ldots, M)$. Meanwhile, it follows from [5] and [6] that the first q matched filters span the signal subspace while the last $(M-q)$ matched filters span the noise subspace, yielding $\boldsymbol{h}_i^H \boldsymbol{A}(\boldsymbol{\theta}) = \boldsymbol{0}$ $(i = q+1, \ldots, M)$. In the sequel, for $i = q+1, \ldots, M$, we obtain the MMSE associated with the ith stage as

$$\begin{aligned} \rho_i &= \sigma_{d_i}^2 - \frac{|\delta_{i+1}|^2}{\rho_{i+1}} = E\left[d_i(t_\ell) d_i^*(t_\ell)\right] \\ &= \boldsymbol{h}_i^H \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{R}_s \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{h}_i + \sigma_n^2 \boldsymbol{h}_i^H \boldsymbol{h}_i = \sigma_n^2, \ (i = q+1, \ldots, M). \end{aligned} \tag{C3}$$

For $i = 1, \ldots, q$, however, $\boldsymbol{h}_i^H \boldsymbol{A}(\boldsymbol{\theta}) \neq \boldsymbol{0}$ and $\delta_i \neq 0$. In the sequel, the variance associated with the ith error can be accordingly represented as

$$\begin{aligned} \rho_i &= \sigma_{d_i}^2 - \frac{|\delta_{i+1}|^2}{\rho_{i+1}} \\ &= \left(\boldsymbol{h}_i^H \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{R}_s \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{h}_i - \frac{|\delta_{i+1}|^2}{\rho_{i+1}}\right) + \sigma_n^2 \quad (i = 1, \ldots, q) \end{aligned} \tag{C4}$$

which is greater than σ_n^2 due to $\boldsymbol{h}_i^H \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{R}_s \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{h}_i \gg |\delta_{i+1}|^2/\rho_{i+1}$. As a result, we obtain $\rho_i > \rho_{q+1} = \cdots = \rho_M = \sigma_n^2$ $(i = 1, \ldots, q)$. Without loss of generality, we assume that the MMSEs of the MSWF are arranged in a decreasing order:

$$\rho_1 \geq \cdots \geq \rho_q > \rho_{q+1} = \cdots = \rho_M = \sigma_n^2. \tag{C5}$$

Additionally, we assume that $\boldsymbol{h}_i$ is the ith matched filter corresponding to ρ_i. This proves Lemma 2.

Appendix D
ML Estimate of ρ_i

Following the results in [10] and [12], we can readily obtain

$$\hat{\sigma}_{d_i}^2 = \sigma_{d_i}^2 + O\left(\sqrt{\frac{\log\log L}{L}}\right) \tag{D1}$$

$$\hat{\delta}_i = \delta_i + O\left(\sqrt{\frac{\log\log L}{L}}\right) \tag{D2}$$

where $\hat{\sigma}_{d_i}^2 \triangleq (1/L)\sum_{\ell=1}^{L} d_i^*(t_\ell) d_i(t_\ell)$ and $\hat{\delta}_i \triangleq (1/L)\sum_{\ell=1}^{L} d_i^*(t_\ell) d_{i-1}(t_\ell)$. In the sequel, considering that $1/(1+x) = 1 - x + x^2 - x^3 + \cdots$ for $|x| < 1$ and $\hat{\rho}_M = \hat{\sigma}_{d_M}^2$ yields

$$\begin{aligned} \hat{\rho}_{M-1} &= \hat{\sigma}_{d_{M-1}}^2 - \frac{|\hat{\delta}_M|^2}{\hat{\rho}_M} \\ &= \sigma_{d_{M-1}}^2 + O\left(\sqrt{\frac{\log\log L}{L}}\right) - \frac{|\delta_M + O\left(\sqrt{\frac{\log\log L}{L}}\right)|^2}{\sigma_{d_M}^2 + O\left(\sqrt{\frac{\log\log L}{L}}\right)} \\ &= \sigma_{d_{M-1}}^2 - \frac{|\delta_M|^2}{\sigma_{d_M}^2}\left(\frac{1}{1+O\left(\sqrt{\frac{\log\log L}{L}}\right)}\right) + O\left(\sqrt{\frac{\log\log L}{L}}\right) \\ &= \sigma_{d_{M-1}}^2 - \frac{|\delta_M|^2}{\sigma_{d_M}^2} + O\left(\sqrt{\frac{\log\log L}{L}}\right) \\ &= \rho_{M-1} + O\left(\sqrt{\frac{\log\log L}{L}}\right). \end{aligned} \tag{D3}$$

Consequently, employing the iteration $\hat{\rho}_{i-1} = \hat{\sigma}^2_{d_{i-1}} - |\hat{\delta}_i|^2/\hat{\rho}_i$ and (D1)–(D2), we eventually have

$$\hat{\rho}_i = \rho_i + O\left(\sqrt{\frac{\log\log L}{L}}\right) \quad (i = 1, \ldots, M). \tag{D4}$$

Exploiting the same arguments used in [10], we obtain that $\hat{\rho}_i$ is the ML estimate of ρ_i.

Appendix E
Proof of Consistency of Reduced-Rank MMDL Method

It follows from (18) that

$$\frac{1}{L}\left(\mathrm{mMDL}^{(D)}(k) - \mathrm{mMDL}^{(D)}(p)\right) = \Upsilon(k) + \frac{1}{2}(k-q)(2D-k-q-1)\frac{\log L}{L} \tag{E1}$$

where

$$\Upsilon(k) \triangleq \frac{\left(\mathcal{F}^{(D)}(k) - \mathcal{F}^{(D)}(q)\right)}{L}. \tag{E2}$$

We first consider the case $k < q$. Substituting (19) into (E2), with some straightforward manipulations, yields (E3), shown at the bottom of the page. Since $\hat{\rho}_i$ $(i = k+1, \ldots, D)$ are not all equal with probability one (w.p.1) as L tends to infinity, we obtain by the inequality between the arithmetic and geometric means

$$\log\left(\frac{\left(\frac{1}{q-k}\sum_{i=k+1}^{q}\hat{\rho}_i\right)^{q-k}}{\prod_{i=k+1}^{q}\hat{\rho}_i}\right) > 0 \text{ a.s. as } L \to \infty. \tag{E4}$$

Here we use the standard abbreviation "a.s." for "almost sure" to describe an event occurring w.p.1. Meanwhile, using the generalized arithmetic-mean geometric-mean inequality, i.e., $\beta_1^{\alpha_1}\beta_2^{\alpha_2} \leq \alpha_1\beta_1 + \alpha_2\beta_2$, where $\alpha_1 + \alpha_2 = 1$, we have (E5), shown at the bottom of the page. As a result, it follows from (E3)–(E5) that $\Upsilon(k) > 0$ a.s. as $L \to \infty$. Meanwhile, noticing that $\log L/L$ approaches to zero as L increases, we attain from (E1) that for $k < q$:

$$\mathrm{mMDL}^{(D)}(k) > \mathrm{mMDL}^{(D)}(q) \text{ a.s. as } L \to \infty. \tag{E6}$$

Consider now the case $k \geq q$. It follows from (C5) and (D4) that $\hat{\rho}_i - \sigma_n^2 = O\left(\sqrt{\log\log L/L}\right)$ $(i = q+1, \ldots, D)$. In the sequel, applying the Taylor series around a small number x, i.e., $\log(1+x) = x - x^2/2 + \cdots$, we obtain (E7), shown at the top of the next page, and

$$\frac{\mathcal{F}^{(D)}(q)}{L} = \frac{1}{2}\left(\sum_{i=q+1}^{D}\left(\frac{\hat{\rho}_i - \sigma_n^2}{\sigma_n^2}\right)^2 - \frac{1}{D-q}\left(\sum_{i=q+1}^{D}\frac{\hat{\rho}_i - \sigma_n^2}{\sigma_n^2}\right)^2\right) + O\left(\left(\frac{\log\log L}{L}\right)^{(3/2)}\right). \tag{E8}$$

Substituting (E7) and (E8) into (E2) yields

$$\begin{aligned}\Upsilon(k) &= \frac{1}{2}\left(\frac{1}{D-q}\left(\sum_{i=q+1}^{D}\frac{\hat{\rho}_i - \sigma_n^2}{\sigma_n^2}\right)^2 - \frac{1}{D-k}\left(\sum_{i=k+1}^{D}\frac{\hat{\rho}_i - \sigma_n^2}{\sigma_n^2}\right)^2 - \sum_{i=q+1}^{k}\left(\frac{\hat{\rho}_i - \sigma_n^2}{\sigma_n^2}\right)^2\right) + O\left(\left(\frac{\log\log L}{L}\right)^{(3/2)}\right) \\ &= O\left(\frac{\log\log L}{L}\right) \quad (k = q, \ldots, D-1).\end{aligned} \tag{E9}$$

Consequently, substituting (E9) into (E1), we obtain for $k = q$:

$$\mathrm{mMDL}^{(D)}(k) = \mathrm{mMDL}^{(D)}(q) \text{ a.s. as } L \to \infty. \tag{E10}$$

Furthermore, considering that $\log L/\log\log L \to \infty$ as $L \to \infty$, for $k > q$

$$\begin{aligned}&\frac{1}{L}\left(\mathrm{mMDL}^{(D)}(k) - \mathrm{mMDL}^{(D)}(q)\right) \\ &= O\left(\frac{\log\log L}{L}\right) + \frac{1}{2}(k-q)(2D-k-q-1)\frac{\log L}{L} \\ &> 0 \text{ a.s. as } L \to \infty.\end{aligned} \tag{E11}$$

Therefore, we eventually have for $k > q$:

$$\mathrm{mMDL}^{(D)}(k) > \mathrm{mMDL}^{(D)}(q) \text{ a.s. as } L \to \infty. \tag{E12}$$

This proves the consistency of the proposed mMDL estimator.

$$\Upsilon(k) = \log\left(\frac{(\frac{1}{q-k}\sum_{i=k+1}^{q}\hat{\rho}_i)^{q-k}}{\prod_{i=k+1}^{q}\hat{\rho}_i}\right) + (D-k)\log\left(\frac{\frac{1}{D-k}\sum_{i=k+1}^{D}\hat{\rho}_i}{(\frac{1}{q-k}\sum_{i=k+1}^{q}\hat{\rho}_i)^{(q-k/D-k)} \times (\frac{1}{D-q}\sum_{i=q+1}^{D}\hat{\rho}_i)^{(D-q/D-k)}}\right). \tag{E3}$$

$$(D-k)\log\left(\frac{\frac{1}{D-k}\sum_{i=k+1}^{D}\hat{\rho}_i}{\left(\frac{1}{q-k}\sum_{i=k+1}^{q}\hat{\rho}_i\right)^{(q-k/D-k)} \times \left(\frac{1}{D-q}\sum_{i=q+1}^{D}\hat{\rho}_i\right)^{(D-q/D-k)}}\right) \geq 0 \text{ a.s. as } L \to \infty. \tag{E5}$$

$$
\begin{aligned}
\frac{\mathcal{F}^{(D)}(k)}{L} &= (D-k)\log\left(\frac{1}{D-k}\sum_{i=k+1}^{D}\hat{\rho}_i\right) - \sum_{i=k+1}^{D}\log\hat{\rho}_i \\
&= (D-k)\log\left(\sigma_n^2\left(1+\frac{\sum\limits_{i=k+1}^{D}\left(\hat{\rho}_i-\sigma_n^2\right)}{(D-k)\,\sigma_n^2}\right)\right) - \sum_{i=k+1}^{D}\log\left(\sigma_n^2\left(1+\frac{\hat{\rho}_i-\sigma_n^2}{\sigma_n^2}\right)\right) \\
&= (D-k)\left(\frac{\sum\limits_{i=k+1}^{D}\left(\hat{\rho}_i-\sigma_n^2\right)}{(D-k)\,\sigma_n^2} - \frac{1}{2}\left(\frac{\sum\limits_{i=k+1}^{D}\left(\hat{\rho}_i-\sigma_n^2\right)}{(D-k)\,\sigma_n^2}\right)^2 + O\left(\left(\frac{\log\log L}{L}\right)^{(3/2)}\right)\right) \\
&\quad - \sum_{i=k+1}^{D}\left(\frac{\hat{\rho}_i-\sigma_n^2}{\sigma_n^2} - \frac{1}{2}\left(\frac{\hat{\rho}_i-\sigma_n^2}{\sigma_n^2}\right)^2 + O\left(\left(\frac{\log\log L}{L}\right)^{(3/2)}\right)\right) \\
&= \frac{1}{2}\left(\sum_{i=k+1}^{D}\left(\frac{\hat{\rho}_i-\sigma_n^2}{\sigma_n^2}\right)^2 - \frac{1}{D-k}\left(\sum_{i=k+1}^{D}\frac{\hat{\rho}_i-\sigma_n^2}{\sigma_n^2}\right)^2\right) + O\left(\left(\frac{\log\log L}{L}\right)^{(3/2)}\right),
\end{aligned}
\tag{E7}
$$

REFERENCES

[1] H.-T. Wu, J.-F. Yang, and F.-K. Chen, "Source number estimation using transformed Gerschgorin radii," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1325–1333, Jun. 1995.

[2] M. Nezafat, M. Kaveh, and W. Xu, "Estimation of the number of sources based on the eigenvectors of the covariance matrix," in *Proc. Sensor Array Multichannel Signal Processing Workshop*, Sitges, Spain, Jul. 2004, pp. 465–469.

[3] E. Fishler and H. V. Poor, "Estimation of the number of sources in unbalanced arrays via information theoretic criteria," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3543–3553, Sep. 2005.

[4] L. Huang, T. Long, and S. Wu, "Source enumeration for high-resolution array processing using improved Gerschgorin radii without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5916–5925, Dec. 2008.

[5] L. Huang and S. Wu, "Low-complexity MDL method for accurate source enumeration," *IEEE Signal Process. Lett.*, vol. 14, pp. 581–584, Sep. 2007.

[6] L. Huang, S. Wu, and X. Li, "Reduced-rank MDL method for source enumeration in high-resolution array processing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5658–5667, Dec. 2007.

[7] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1413–1426, Jun. 1995.

[8] M. Pesavento and A. B Gershman, "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1310–1324, Jul. 2001.

[9] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, pp. 387–392, Apr. 1985.

[10] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 1190–1196, Aug. 1989.

[11] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.

[12] I. A. Ibragimov and R. Z. Hasminski, *Statistical Estimation-Asymptotic Theory*. New York: Springer-Verlag, 1981.

[13] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

Joint Transmitter and Receiver Polarization Optimization for Scattering Estimation in Clutter

Jin-Jun Xiao and Arye Nehorai, *Fellow, IEEE*

***Abstract*—Controlling the polarization information in transmitted waveforms enables improving the performance of radar systems. We consider the design of optimal polarizations at both the radar transmitter and receiver for the estimation of target scattering embedded in clutter. The goal is to minimize the mean squared error of the scattering estimation subject to an average radar pulse power constraint. Under the condition that the target and clutter scattering covariance matrices are known a priori, we show that such a problem is equivalent to the optimal design of a radar sensing matrix that contains the polarization information. We formulate the optimal design as a nonlinear optimization problem and then recast it in a convex form and is thus efficiently solvable by semi-definite programming (SDP). We compare the sensing performance of the optimally selected polarization over conventional approaches. Our numerical results demonstrate that a significant amount of power gain is achieved in the target scattering estimation through such an optimal design.**

***Index Terms*—Adaptive estimation, optimization methods, radar polarimetry, scattering matrices.**

I. INTRODUCTION

Advances in digital signal processing and computing technology have resulted in the emergence of increasingly adaptive radar systems.

Manuscript received June 19, 2008; accepted March 11, 2009. First published May 12, 2009; current version published September 16, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Antonio Napolitano. The work is supported in part by the Department of Defense under Air Force Office of Scientific Research MURI Grant FA9550-05-0443 and ONR Grant N000140810849.

The authors are with the Department of Electrical and Systems Engineering, Washington University, St. Louis, MO 63130 USA (e-mail: xiao@ese.wustl.edu; nehorai@ese.wustl.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2022887

1053-587X/$26.00 © 2009 IEEE