

# On the Cramér–Rao Bound for Estimating the Mixing Matrix in Noisy Sparse Component Analysis

Hadi Zayyani, Massoud Babaie-Zadeh, Farzan Haddadi, Christian Jutten

► **To cite this version:**

Hadi Zayyani, Massoud Babaie-Zadeh, Farzan Haddadi, Christian Jutten. On the Cramér–Rao Bound for Estimating the Mixing Matrix in Noisy Sparse Component Analysis. IEEE Signal Processing Letters, Institute of Electrical and Electronics Engineers, 2008, 15, pp.609-612. 10.1109/LSP.2008.2003989 . hal-00327590

**HAL Id: hal-00327590**

**<https://hal.archives-ouvertes.fr/hal-00327590>**

Submitted on 8 Oct 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the Cramér–Rao Bound for Estimating the Mixing Matrix in Noisy Sparse Component Analysis

Hadi Zayyani, *Student Member, IEEE*, Masoud Babaie-zadeh, *Member, IEEE*,  
Farzan Haddadi, *Student Member, IEEE*, and Christian Jutten, *Fellow, IEEE*

**Abstract**—In this letter, we address the theoretical limitations in estimating the mixing matrix in noisy sparse component analysis (SCA) for the two-sensor case. We obtain the Cramér–Rao lower bound (CRLB) error estimation of the mixing matrix. Using the Bernoulli–Gaussian (BG) sparse distribution, and some simple assumptions, an approximation of the Fisher information matrix (FIM) is calculated. Moreover, this CRLB is compared to some of the main methods of mixing matrix estimation in the literature.

**Index Terms**—Blind source separation, Cramér–Rao bound, mixing matrix estimation, sparse component analysis.

## I. INTRODUCTION

**S**PARSE component analysis (SCA) [1] is a semi-blind source separation approach, in which the prior information about the sources is their sparsity. A sparse signal is a signal whose most samples are nearly zero (say they are “inactive”), and just a few percent take significant values (say they are “active”). This prior information enables us to separate sources with less sensors than sources [2]–[8]. The mathematical model of the instantaneous underdetermined blind source separation (BSS) in the noisy case is

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{v} \quad (1)$$

where  $\mathbf{A} \in \mathbb{R}^{n \times m}$  is the mixing matrix, and  $\mathbf{x}$  and  $\mathbf{s}$  are the observation and source vectors, respectively. In the underdetermined case, the number of observations is less than the number of sources ( $n < m$ ). Therefore, estimating the mixing matrix is not sufficient to recover the sources, since the mixing matrix is not invertible. Therefore, underdetermined SCA consists of two steps: first, estimating the mixing matrix and then estimating the sparse sources. In this letter, we focus on the first step.

Several approaches have been proposed to address the mixing matrix estimation in SCA in the underdetermined case. The potential-function-based method is proposed in [2]. A similar approach is described in [9], using a histogram rather than a potential function. In [10], the Laplacian mixture model (LMM) is assumed for the distribution of  $\varphi = \arctan(x_2/x_1)$  in the case of a two-sensor set up where  $x_1$  and  $x_2$  are the two observations. Then, an EM algorithm finds the ML estimation of the pa-

rameters of this LMM. So, this method is called the EM-LMM method. A geometrical approach was proposed in [11] for estimating the mixing matrix. Recently, [12] proposed a potential-function-based clustering method constructed by a Laplacian-like window function.

In most of the previously mentioned approaches (namely, [9]–[12]), the ratio of the two observations or the polar angle is used for estimating the mixing matrix in the two-sensor case. In this letter, we present a minimum error bound for estimating the mixing matrix based on this ratio of observations. The Cramér–Rao lower bound (CRLB), which is the inverse of the Fisher information matrix (FIM), bounds the performance of any unbiased parametric estimator in terms of the mean square estimation error [18]. The CRLB is calculated in this letter assuming a Bernoulli–Gaussian (BG) sparse distribution for sources and assuming a high signal-to-noise ratio (SNR). Moreover, it is assumed that the columns of the mixing matrix are not too close to each other. In the context of BSS, in [13], the asymptotic Cramér–Rao bound has been calculated in the case of instantaneous mixture of nonstationary source signals when the source distributions are known. In [14], the CRLB for the estimation of the unmixing parameters has been evaluated for the case of Gaussian auto-regressive (AR) sources in the determined case. Reference [15] derives a closed-form expression for the Cramér–Rao bound in estimating the source signals in the linear independent component analysis (ICA). Moreover, [16] derives the Cramér–Rao-induced bound for blind separation of stationary parametric Gaussian sources. All of these papers calculate CRLB in the determined case of BSS, but we calculate the CRLB in estimating the mixing matrix in underdetermined SCA.

## II. SYSTEM MODEL AND PRELIMINARIES

Consider the problem of estimating the mixing matrix in the two-sensor case. We use the polar model

$$\begin{cases} x_1 = (\cos \theta_1)s_1 + (\cos \theta_2)s_2 + \dots + (\cos \theta_m)s_m + n_1 \\ x_2 = (\sin \theta_1)s_1 + (\sin \theta_2)s_2 + \dots + (\sin \theta_m)s_m + n_2 \end{cases} \quad (2)$$

where  $x_1$  and  $x_2$  are the two sensor outputs, and  $n_1$  and  $n_2$  are two independent Gaussian noises with variance  $\sigma_n^2$ . Using this model, the estimation of the mixing matrix reduces to the estimation of the parameter vector  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_m]^T$ .

We assume a BG distribution to model the sparsity. This model, which has also been used in [7] and [8], can model the sparsity in a simple manner and has simple computational properties. Therefore, the probability density function (PDF) of the sources is assumed to be

$$p(s_i) = p\delta(s_i) + (1-p)\mathcal{N}(0, \sigma_r^2) \quad (3)$$

where  $p$  is the probability of inactivity of the sources and is near one (by the sparsity assumption), and  $\sigma_r^2$  is the variance of

Manuscript received March 1, 2008; revised June 7, 2008. This work was supported in part by the Iran NSF (INSF) under contract number 86/994 and in part by the Iran Telecommunication Research Center (ITRC). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jen-Tzung Chien.

H. Zayyani, M. Babaie-zadeh, and F. Haddadi are with the Electrical Engineering Department, Sharif University of Technology, Tehran, Iran (e-mail: zayyani@ee.sharif.edu; mbzadeh@sharif.edu; farzanhaddadi@ee.sharif.edu).

C. Jutten is with the GIPSA-LAB, Institut National Polytechnique de Grenoble (INPG), Grenoble, France (e-mail: Christian.Jutten@inpg.fr).

Digital Object Identifier 10.1109/LSP.2008.2003989

the active samples of the  $i$ th sources. We also define the source activity vector as an  $m$ -tuple vector  $\mathbf{q} = (q_1, q_2, \dots, q_m)^T$ , where the  $i$ th component shows the activity of the  $i$ th source

$$q_i \triangleq \begin{cases} 1, & \text{if } s_i \text{ is active} \\ 0, & \text{if } s_i \text{ is inactive.} \end{cases}$$

As stated in the introduction, in most algorithms of mixing matrix estimation in the two-dimensional space, and for sparse sources, the ratio of the two observations is used for the estimation. The ratio is equal to  $\tan \theta_i$  when only one source (the  $i$ th source) is active and the noise terms are negligible. Therefore, in this work, we deduce the CRLB from the ratio of the observations  $y \triangleq x_2/x_1$ . Using the total probability theorem, the likelihood  $p(y|\boldsymbol{\theta})$  can be written as

$$p(y|\boldsymbol{\theta}) = \sum_{\mathbf{q}} p(\mathbf{q})p(y|\mathbf{q}, \boldsymbol{\theta}) \quad (4)$$

where  $p(\mathbf{q}) = p^{m-n_a}(1-p)^{n_a}$  is the probability of  $\mathbf{q}$ , in which  $n_a$  is the number of active sources, and the summation is taken over the  $2^m$  possible values for the source activity vector. Since  $(1-p) \ll 1$ , the terms with  $n_a > 1$  can be neglected. Therefore, the sparse approximation of the distribution is

$$p(y|\boldsymbol{\theta}) \approx p^m p(y|\boldsymbol{\theta}, H_0) + p^{m-1}(1-p) \sum_{i=1}^m p(y|\boldsymbol{\theta}, H_i) \quad (5)$$

where  $H_0$  is the hypothesis that all sources are inactive (all sources are zero) and  $H_i$  is the hypothesis that only the  $i$ th source is active (all other sources are zero). These hypotheses can be written as

$$H_0 : \begin{cases} x_1 = n_1 \\ x_2 = n_2 \end{cases} \quad H_i : \begin{cases} x_1 = (\cos \theta_i)s_i + n_1 \\ x_2 = (\sin \theta_i)s_i + n_2. \end{cases}$$

In both hypotheses, the variable  $y$  is obtained from division of the two normal random variables. It is well known that the quotient of two jointly normal random variables ( $y = x_2/x_1$ ) is a Cauchy random variable with the following distribution [17]:

$$f_y(y) = \frac{\frac{\alpha}{\pi}}{(y - \mu)^2 + \alpha^2} \quad (6)$$

where the two Cauchy parameters are  $\alpha = (\sigma_2/\sigma_1)\sqrt{1-r^2}$  and  $\mu = r(\sigma_2/\sigma_1)$ , in which  $\sigma_1^2$  and  $\sigma_2^2$  are the variance of  $x_1$  and  $x_2$ , and  $r$  is their correlation coefficient. Therefore, each of the probability distributions  $p(y|\boldsymbol{\theta}, H_i)$ ,  $i = 0, 1, \dots, m$  is a Cauchy distribution. So,  $p(y|\boldsymbol{\theta})$  is a mixture of Cauchy (MoC) distributions. Simple computations show that the parameters of these Cauchy distributions assuming  $H_i$  are

$$\alpha_i \simeq \frac{\sigma_{r_i}\sigma_n}{\sigma_n^2 + \sigma_{r_i}^2 \cos^2 \theta_i} \quad \mu_i = \frac{\sigma_{r_i}^2 \sin \theta_i \cos \theta_i}{\sigma_n^2 + \sigma_{r_i}^2 \cos^2 \theta_i} \quad (7)$$

where  $1 \leq i \leq m$ . In the computation of  $\alpha_i$ , we have neglected the term  $\sigma_n^2$  in comparison with  $\sigma_{r_i}^2$  ( $\sigma_{r_i}^2 \gg \sigma_n^2$ ). In other words, the case of high SNR is assumed. For hypothesis  $H_0$ , we have  $\alpha_0 = 1$  and  $\mu_0 = 0$ . In this case, the ratio of observations  $y$  has a Cauchy distribution around zero. In this case, it is well known [17] that  $\arctan(y)$ , the polar phase, has a uniform distribution in  $[-(\pi/2), (\pi/2)]$ . So, there are no informative points in the scatter plot. This provides a mathematical argument to methods such as [3] and [10], which throw away the observation points

near the origin. Finally, using the above discussions, the probability density (5) is

$$p(y|\boldsymbol{\theta}) = \sum_{i=0}^m p_i \frac{\frac{\alpha_i}{\pi}}{(y - \mu_i)^2 + \alpha_i^2} \quad (8)$$

where  $p_i = (1-p)p^{m-1}$  for  $i \neq 0$  and  $p_i = p^m$  for  $i = 0$ .

### III. CRAMÉR-RAO LOWER BOUND

The CRLB of a vector of parameters  $\boldsymbol{\theta}$  estimated from data  $y$  is the inverse of the FIM and bounds the performance of any unbiased estimator. The FIM is [18]

$$\mathbf{I}_{\boldsymbol{\theta}} = E_y \left[ \left( \frac{\partial \ln(p(y|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial \ln(p(y|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \right)^T \right]. \quad (9)$$

Then, the covariance of the estimated parameters are bounded as<sup>1</sup>

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \geq \text{CRB}_{\boldsymbol{\theta}} = \mathbf{I}_{\boldsymbol{\theta}}^{-1}. \quad (10)$$

In our case, we want to compute the FIM from the observation ratio  $y = x_2/x_1$ . Therefore, the elements of the FIM should be calculated as

$$I_{ij} = E_y \left[ \left( \frac{\partial \ln(p(y|\boldsymbol{\theta}))}{\partial \theta_i} \right) \left( \frac{\partial \ln(p(y|\boldsymbol{\theta}))}{\partial \theta_j} \right) \right]. \quad (11)$$

To compute  $(\partial \ln(p(y|\boldsymbol{\theta}))/\partial \theta_i)$ , we use (7) and (8). After some straightforward manipulations,<sup>2</sup> this term can be written as

$$\frac{A_i}{p(y|\boldsymbol{\theta})} \left\{ \frac{\sin 2\theta_i}{(y - \mu_i)^2 + \alpha_i^2} + \frac{2(y - \mu_i) \frac{\sigma_{r_i}^2 \cos^2 \theta_i}{\sigma_n^2 + \sigma_{r_i}^2 \cos^2 \theta_i}}{[(y - \mu_i)^2 + \alpha_i^2]^2} \right\} \quad (12)$$

where  $A_i \triangleq p_i/\pi(\sigma_n \sigma_{r_i}^3/(\sigma_n^2 + \sigma_{r_i}^2 \cos^2 \theta_i)^2)$ .

To compute  $I_{ij}$  from (11), the expectation is replaced with an integral on the pdf  $p(y|\boldsymbol{\theta})$ . So, we have

$$I_{ij} = \int_{-\infty}^{+\infty} \left( \frac{\partial \ln p(y|\boldsymbol{\theta})}{\partial \theta_i} \right) \left( \frac{\partial \ln p(y|\boldsymbol{\theta})}{\partial \theta_j} \right) p(y|\boldsymbol{\theta}) dy. \quad (13)$$

Therefore, we must multiply the expressions in (12) by a similar expression and by the likelihood  $p(y|\boldsymbol{\theta})$  and then integrate them with respect to  $y$ . This is equivalent to compute the sum of the four integrals due to each of the multiplication terms.

Now, to obtain an approximate closed-form relation for (13), we assume that any pair of Cauchy distributions,  $p(y|\boldsymbol{\theta}, H_i)$  and  $p(y|\boldsymbol{\theta}, H_j)$ ,  $i \neq j$ , whose densities are proportional to  $1/((y - \mu_i)^2 + \alpha_i^2)$  and  $1/((y - \mu_j)^2 + \alpha_j^2)$ , are far from each other. In other words, the angles  $\theta_i$  are assumed to be far from each other, and hence, the centers of these Cauchy distributions, which are approximately to  $\mu_i \approx \tan \theta_i$  are far from each other.<sup>3</sup> Moreover, the parameter  $\alpha_i$ , which determines the width of the Cauchy distribution, is assumed to be small enough compared to the distance between the centers of the

<sup>1</sup>where  $\mathbf{A} \geq \mathbf{B}$  means that  $(\mathbf{A} - \mathbf{B})$  is positive semidefinite.

<sup>2</sup>By neglecting the terms of orders higher than 2 in  $(\sigma_n/\sigma_{r_i})^k$ .

<sup>3</sup>This can be seen by assuming  $\sigma_{r_i}^2 \cos^2 \theta_i \gg \sigma_n^2$  in (7), which is approximately equivalent to  $\tan \theta_i \ll \sigma_{r_i}/\sigma_n$ . It also means that we assume that there is no angle near  $\pi/2$ . However, this restriction is only due to the parametrization. For example, if there is any angle there, you can rotate the angles by a proper angle to solve the problem.

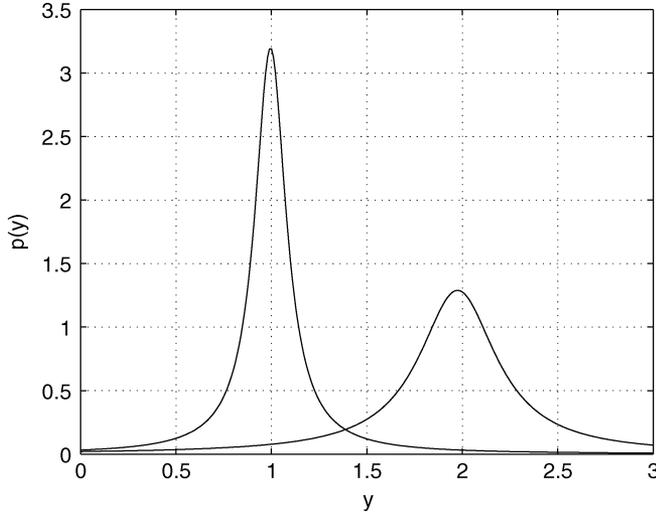


Fig. 1. Typical display of two Cauchy distribution. The parameters are  $\theta_1 = \arctan(1)$ ,  $\theta_2 = \arctan(2)$ ,  $p = .9$ ,  $\sigma_{r_i} = 1$  for  $i = 1, 2$  and  $\sigma_n = .05$

Cauchy distributions.<sup>4</sup> A typical figure of these two Cauchy distributions is depicted in Fig. 1. Therefore, in the case of  $i \neq j$ , the four terms of the integrand of the integral (13) are small enough, and so the nondiagonal elements of the FIM are approximately zero. This theoretical result will also be experimentally verified in Section IV.

To compute the diagonal elements,  $I_{ii}$ , among the four previously mentioned terms, the two cross terms are odd functions around  $\mu_i$ , and so these two integrals have zero values. Therefore, we only should compute the two following integrals:

$$c_1 \triangleq A_i^2 \sin^2 2\theta_i \int_{-\infty}^{+\infty} \frac{1}{[(y - \mu_i)^2 + \alpha_i^2]^2} \frac{1}{p(y|\boldsymbol{\theta})} dy \quad (14)$$

$$c_2 \triangleq A_i^2 \int_{-\infty}^{+\infty} \frac{4(y - \mu_i)^2 \left( \frac{\sigma_{r_i}^2 \cos^2 \theta_i}{\sigma_n^2 + \sigma_{r_i}^2 \cos^2 \theta_i} \right)^2}{[(y - \mu_i)^2 + \alpha_i^2]^4} \frac{1}{p(y|\boldsymbol{\theta})} dy. \quad (15)$$

Now, we use the approximation  $p(y|\boldsymbol{\theta}) \approx 1/\pi(p_i\alpha_i/(y - \mu_i)^2 + \alpha_i^2)$  around  $\mu_i$ ,  $p_i p(y|\boldsymbol{\theta}, H_i)$ . It means that the mixture of Cauchy distributions is approximately equal to one of the Cauchy distributions around  $\mu_i$ . Using this approximation and some manipulations and computations of the above integrals, we finally obtain

$$c_1 \simeq A_i^2 \sin^2 2\theta_i \frac{\pi^2}{p_i \alpha_i}, \quad c_2 \simeq A_i^2 \frac{B\pi^2}{8p_i \alpha_i^4} \quad (16)$$

where  $B \triangleq 4(\sigma_{r_i}^2 \cos^2 \theta_i / \sigma_n^2 + \sigma_{r_i}^2 \cos^2 \theta_i)^2$ . After simplifications, the value of the integral will be computed. In summary, the FIM is a diagonal matrix  $\mathbf{I} = \text{diag}(I_{ii})$ , and hence, the CRLB matrix is easily found by inverting the diagonal elements,  $\text{CRB} = \text{diag}(1/I_{ii})$ . Therefore, (10) gives the lower bound on

<sup>4</sup>It is equivalent to  $\alpha_i \approx \sigma_n / \sigma_{r_i} (1 + \tan^2 \theta_i) \ll |\tan \theta_i - \tan \theta_j|$ . Some manipulations result in  $|\sin(\Delta\theta_{ij})| \gg (|\cos \theta_i / \cos \theta_j|) \sigma_n / \sigma_{r_i}$ , where  $\Delta\theta_{ij} = \theta_i - \theta_j$ .

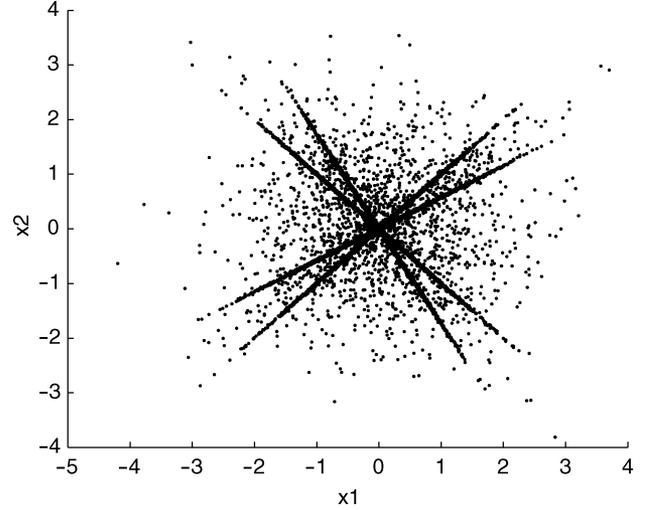


Fig. 2. Typical scatter plot of observation points. The angle parameter vector is  $\boldsymbol{\theta} = [(-\pi/3), (-\pi/4), (\pi/6), (\pi/4)]^T$ , and other parameters are  $p = .9$ ,  $\sigma_{r_i} = 1$  for all  $i$  and  $\sigma_n = .01$ .

the error estimation of any angle of the mixing matrix. This lower bound using  $N$  samples of  $y$  will be [18]

$$E \left\{ (\theta_i - \hat{\theta}_i)^2 \right\} \geq \frac{1}{4Np_1} \frac{1}{\gamma_i} \frac{\left(1 + \frac{1}{\gamma_i} \tan^2 \theta_i\right)^2}{0.125 + \frac{1}{\gamma_i} \tan^2 \theta_i} \quad (17)$$

where  $\gamma_i \triangleq \sigma_{r_i}^2 / \sigma_n^2$  is a measure of the SNR of the  $i$ th source and  $p_1 \triangleq (1 - p)p^{m-1}$ . The above Cramér–Rao bound shows that the estimation error increases when the absolute value of  $|\tan \theta_i|$  increases. In fact, if the SNR is high and the angles are not near  $\pi/2$ , we can neglect small terms in (17) and obtain a simple formula for CRLB equal to  $(2/Np_1)(1/\gamma_i)$ . Note that the assumption of avoiding angles near  $\pi/2$  is only due to the parametrization. Moreover, (17) shows that in the higher SNRs, the CRLB becomes smaller.

#### IV. SIMULATION RESULTS

In this section, the performance of a few methods of mixing matrix estimation is compared to the calculated CRLB. In this simulation, we chose the mixing matrix as in (2) with parameter vector  $\boldsymbol{\theta} = [(-\pi/3), (-\pi/4), (\pi/6), (\pi/4)]^T$ . Therefore, we have  $m = 4$  sources and  $n = 2$  observations. The sparse sources are generated artificially using the model (3), with parameters  $p = 0.9$ ,  $\sigma_{r_i} = 1$  for all  $i$  and  $\sigma_n = .01$ . The scatter plot of the observation points are shown in Fig. 2.

To verify the approximations assumed in obtaining the closed-form CRLB in (17), we compute all the entries (13) of the FIM by numerical integration, which results in

$$\mathbf{I} = \begin{pmatrix} 362.3189 & -0.2830 & -0.0012 & -0.0019 \\ -0.2830 & 362.1377 & -0.0024 & -0.0012 \\ -0.0012 & -0.0024 & 356.7162 & -0.2834 \\ -0.0019 & -0.0012 & -0.2834 & 362.1494 \end{pmatrix}$$

while the diagonal approximation derived from (16) gives

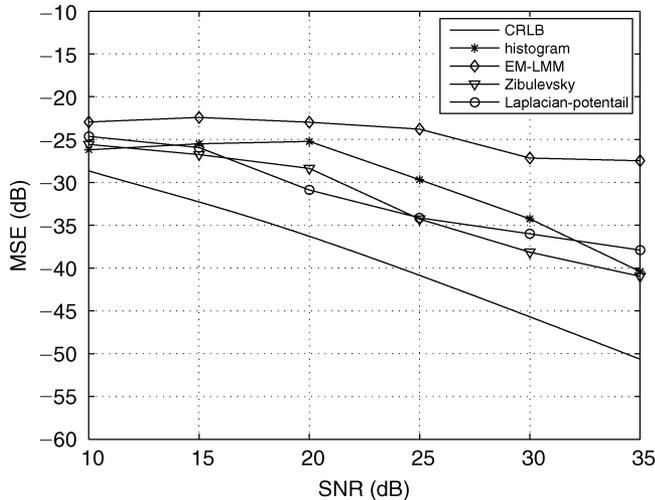


Fig. 3. MSE of the first mixing matrix angle  $\theta_1 = (-\pi)/3$  versus common SNR of all sources for various methods and their comparison with new calculated CRLB. The angle parameter vector is  $\theta = [(-\pi/3), (-\pi/4), (\pi/6), (\pi/4)]^T$ , and other parameters are  $p = .9$ ,  $\sigma_{r_i} = 1$  for all  $i$ .

$$\mathbf{I}_{\text{approx}} = \text{diag}(365.1557, 364.7187, 364.5729, 364.7187)$$

which is close to the actual matrix.

The results of four different methods of mixing matrix estimation are now compared with each other and also with the CRLB(17). These four methods are potential-based clustering [3], histogram method [9], Laplacian-window potential-function-based clustering [12], and EM-LMM method [2]. In potential-based clustering [10], the parameters are chosen as  $\lambda = N/10$  (adjusting the angular width),  $K = N/2$  (number of angle bins), and  $h = 0.3$  (the threshold) [2], where  $N$  is the total number of observations and is equal to 1000 in our simulations.

To evaluate and compare the accuracy of the algorithms in estimating the mixing matrix, we define the mean-square error (MSE) of the  $i$ th mixing matrix angle ( $\theta_i$ ) as

$$\text{MSE}_{\theta_i} = 10 \log_{10} \left( \frac{\sum_{n=1}^{100} (\theta_i - \hat{\theta}_{in})^2}{100} \right). \quad (18)$$

The simulations are then repeated 100 times with new random sparse sources and the MSEs were averaged over all experiments. For the EM-LMM method, because of its iterative nature and its sensitivity to the initialization, it may not converge to the true mixing matrix angle. Therefore, to investigate the success of this algorithm, the resulting errors were averaged only over the successful experiments, where by a successful experiment, we mean an experiment in which  $|\theta_i - \hat{\theta}_i| < 0.1$ . Fig. 3 shows the results of simulation for the first angle of the mixing matrix ( $\theta_1 = (-\pi)/3$ ). The results show a minimum of 4-dB gap in MSE for the CRLB and the various methods. This shows that there could be better methods for estimating the mixing matrix especially in the two-sensor case.

## V. CONCLUSIONS

In this letter, the CRLB on the estimation error of the mixing matrix in the two-sensor case are computed for the ratio of the

observations. To do that, a few assumptions and approximations were made which seem reasonable in a typical application. This lower bound shows the theoretical dependence of the error variance of the estimated mixing matrix on the model parameters which is basically the sparsity parameters, noise power, and the number of samples of the observations. This CRLB is compared to the performance of some of the previous methods, showing a gap between the previous methods and this CRLB. Further investigations include a parametrization which provides angle-invariant bound and computation of CRLB for higher (than 2) dimension source separation problem.

## ACKNOWLEDGMENT

The authors would like to thank Dr. R. Gribonval for useful discussions and suggestions and the three reviewers for their insightful comments.

## REFERENCES

- [1] R. Gribonval and S. Lesage, "A survey of sparse component analysis for blind source separation: Principles, perspectives, and new challenges," in *Proc. ESANN'06*, Apr. 2006, pp. 323–330.
- [2] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representation," *Signal Process.*, vol. 81, pp. 2353–2362, 2001.
- [3] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Comput.*, vol. 13, no. 4, pp. 863–882, 2001.
- [4] M. Davies and N. Mitianoudis, "Simple mixture model for sparse overcomplete ICA," *Proc. Inst. Elect. Eng., Vis., Image, Signal Process.*, vol. 151, no. 1, pp. 35–43, Feb. 2004.
- [5] M. Babaie-Zadeh, C. Jutten, and A. Mansour, "Sparse ICA via cluster-wise PCA," *Neurocomputing*, vol. 69, pp. 1458–1466, Aug. 2006.
- [6] Y. Q. Li, S. Amari, A. Cichocki, D. W. C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 423–437, Feb. 2006.
- [7] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "Source estimation in noisy sparse component analysis," in *Proc. DSP'07*, Jul. 2007, pp. 219–222.
- [8] L. Vielva, D. Erdogmus, and C. Principe, "Underdetermined blind source separation using a probabilistic source sparsity model," in *Proc. ICA'01*, 2001, pp. 675–679.
- [9] A. Prieto, B. Prieto, C. G. Puntonet, A. Canas, and P. Martin-Smith, "Geometric separation of linear mixtures of sources: Application to speech signals," in *Proc. ICA'99*, Jan. 1999, pp. 295–300.
- [10] N. Mitianoudis and T. Stathaki, "Overcomplete Source separation using Laplacian Mixture models," *IEEE Signal Process. Lett.*, vol. 12, no. 4, pp. 277–280, Apr. 2005.
- [11] F. J. Theis, W. E. Lang, and C. G. Puntonet, "A geometric algorithm for overcomplete linear ICA," *Neurocomputing*, vol. 56, pp. 381–398, 2004.
- [12] W. Zhang, J. Liu, J. Sun, and S. Bai, "A new two-stage approach to underdetermined blind source separation using sparse representation," in *Proc. ICASSP'07*, 2007, pp. 953–956.
- [13] J. F. Cardoso, "Separation of non stationary sources: Achievable performance," in *Proc. 10th IEEE Workshop Statistical Signal and Array Processing*, 2000, pp. 359–363.
- [14] S. Degerine and A. Zaidi, "Separation of an instantaneous mixture of Gaussian autoregressive sources by the exact maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1492–1512, Jun. 2004.
- [15] P. Tichavsky, Z. Koldovsky, and E. Oja, "Performance analysis of the Fastlca algorithm and Cramer-Rao bounds for linear independent component analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1189–1203, Apr. 2006.
- [16] E. Doron, A. Yeredor, and P. Tichavsky, "Cramer-Rao-induced bound for blind separation of stationary parametric Gaussian sources," *IEEE Signal Process. Lett.*, vol. 14, no. 6, pp. 417–420, Jun. 2007.
- [17] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York: McGraw-Hill, 2002.
- [18] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.