Published in IEEE Signal Processing Letters, Vol. 15, 2008, pp. 609-612

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

H. Zayyani[1] *Student Member*, M. Babaie-zadeh[1] *Member*, F. Haddadi[1] *Student Member*, and C. Jutten[2] *Fellow*

***Abstract*— In this paper, we address the theoretical limitations in estimating the mixing matrix in noisy Sparse Component Analysis (SCA) for the two-sensor case. We obtain the Cramer-Rao Lower Bound (CRLB) error estimation of the mixing matrix. Using the Bernouli-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*-Sparse component analysis, Blind source separation, Mixing matrix estimation, Cramer-Rao bound.**

I. Introduction

Sparse 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 percents takes signi£cant values (say they are "active"). This prior information enables us to separate sources with less sensors than sources [2], [3], [4], [5], [6], [7], [8]. The mathematical model of the instantaneous underdetermined Blind Source Separation (BSS) in the noisy case is:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{v} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{n\times m}$ is the mixing matrix, $\mathbf{x}$ and $\mathbf{s}$ are the observation and source vectors respectively. In underdetermined case, the number of observations is less than the number of sources ($n < m$). Therefore, estimating the mixing matrix is not suf£cient to recover the sources, since the mixing matrix is not invertible. Therefore, underdetermined SCA consists of two steps: £rst, estimating the mixing matrix and then estimating the sparse sources. In this paper, we focus on the £rst 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 $\varphi = \arctan(\frac{x_2}{x_1})$ in the case of two-sensor set up where x_1 and x_2 are the two observations. Then, an EM algorithm £nds the ML estimation of the parameters of this LMM. So, this method is called 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 Laplacian-like window function.

In most of the pre-mentioned approaches (namely [9], [10], [11], [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 paper, we present a minimum error bound for estimating the mixing matrix based on this ratio of observations. The Cramer-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 paper 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 Blind Source Separation (BSS), in [13], the asymptotic Cramer-Rao bound has been calculated in the case of instantaneous mixture of non stationary 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. [15] derives a closed-form expression for the Cramer-Rao bound in estimating the source signals in the linear Independent Component Analysis (ICA). Moreover, [16] derives the Cramer-Rao-induced bound for blind separation of stationary parametric Gaussian sources. All 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 priliminaries

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

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

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

We assume a Bernoulli-Gaussian (BG) distribution to model the sparsity. This model, which has also been used in [7], [8], can model the sparsity in a simple manner, and has simple

[1]Electrical engineering department, Sharif university of technology, Tehran, Iran. This work has been partially funded by Iran NSF (INSF) under contract number 86/994, and by Iran Telecommunication Research Center (ITRC).

[2]GIPSA-LAB, Institut National Polytechnique de Grenoble (INPG), France.

computational properties. Therefore, the Probability Density Function (PDF) of the sources is assumed to be:

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

where p is the probability of inactivity of the sources and is near one (by the sparsity assumption), and $\sigma_{r_i}^2$ is the variance of the active samples of the i'th sources. We also de£ne the source activity vector as an m-tuple vector $\mathbf{q} = (q_1, q_2, ..., q_m)^T$, 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 $y \triangleq \frac{x_2}{x_1}$. 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:

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

In both hypotheses, the variable y is obtained from division of the two normal random variable. It is well known that the quotient of two jointly normal random variables ($y = \frac{x_2}{x_1}$) 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 Caushy parameters are $\alpha = \frac{\sigma_2}{\sigma_1}\sqrt{1-r^2}$ and $\mu = r\frac{\sigma_2}{\sigma_1}$, in which σ_1^2, σ_2^2 are the variance of x_1 and x_2 and r is their correlation coef£cient. Therefore, each of the probability distributions $p(y|\boldsymbol{\theta}, H_i), i = 0, 1, ..., m$ 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 $1 \le i \le m$. In the computation of α_i, we have neglected the term σ_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 $[-\frac{\pi}{2}, \frac{\pi}{2}]$. 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. Cramer-Rao Lower Bound

The Cramer-Rao Lower Bound of a vector of parameters $\boldsymbol{\theta}$ estimated from data y is the inverse of the Fisher Information Matrix (FIM), and bounds the performance of any unbiased estimator. The Fisher information matrix is [18]:

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

Then, the covariance of the estimated parameters are bounded as[1]:

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

In our case, we want to compute the Fisher information matrix from the observation ratio $y = \frac{x_2}{x_1}$. Therefore, the elements of the Fisher information matrix should be calculated as:

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

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

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

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

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

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

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 $\frac{1}{(y-\mu_i)^2+\alpha_i^2}$ and $\frac{1}{(y-\mu_j)^2+\alpha_j^2}$, are far from each other. In other words, the angles θ_i are assumed to be far from each other, and hence the centers of these Caushy distributions,

[1]where $\mathbf{A} \ge \mathbf{B}$ means that $(\mathbf{A} - \mathbf{B})$ is positive semide£nite.

[2]By neglecting the terms of orders higher than 2 in $(\frac{\sigma_n}{\sigma_{r_i}})^k$.

which are approximately to $\mu_i \approx \tan\theta_i$ are far from each other[3]. Moreover, the parameter α_i which determines the width of the Cauchy distribution, is assumed to be small enough compared to the distance between the centers of the Cauchy distributions[4]. A typical £gure 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 non diagonal elements of the Fisher Information matrix are approximately zero. This theoretical result will also be experimentally veri£ed in Section IV.

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

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

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

Now, we use the approximation $p(y|\boldsymbol{\theta}) \approx \frac{1}{\pi}\frac{p_i\alpha_i}{(y-\mu_i)^2+\alpha_i^2}$ around μ_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 μ_i. Using this approximation and some manipulations and computations of the above integrals, we £nally 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 $B \triangleq 4(\frac{\sigma_{r_i}^2\cos^2\theta_i}{\sigma_n^2+\sigma_{r_i}^2\cos^2\theta_i})^2$. After simpli£cations, the value of the integral will be computed. In summary, the Fisher information matrix is a diagonal matrix $\mathbf{I} = \mathrm{diag}(I_{ii})$, and hence the Cramer-Rao Lower Bound matrix is easily found by inverting the diagonal elements, $\mathrm{CRB} = \mathrm{diag}(\frac{1}{I_{ii}})$. Therefore, (10) gives the lower bound on the error estimation of any angle of the mixing matrix. This lower bound using N samples of y will be [18]:

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

where $\gamma_i \triangleq \frac{\sigma_{r_i}^2}{\sigma_n^2}$ is a measure of the Signal to Noise Ratio (SNR) of the i'th source and $p_1 \triangleq (1-p)p^{m-1}$. The above Cramer-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 $\frac{\pi}{2}$, we can neglect small terms in (17), and obtain a simple formula for CRLB equal to $\frac{2}{Np_1}\frac{1}{\gamma_i}$. Note that the assumption of avoiding angles near $\frac{\pi}{2}$ is only due to the parametrization. Moreover, (17) shows that in the higher SNR's, the CRLB becomes smaller.

[3]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 $\tan\theta_i \ll \frac{\sigma_{r_i}}{\sigma_n}$. It also means that we assume that there is no angle near $\frac{\pi}{2}$. 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.

[4]It is equivalent to $\alpha_i \approx \frac{\sigma_n}{\sigma_{r_i}}(1+\tan^2\theta_i) \ll |\tan\theta_i - \tan\theta_j|$. Some manipulations results in $|\sin(\Delta\theta_{ij})| \gg |\frac{\cos\theta_i}{\cos\theta_j}|\frac{\sigma_n}{\sigma_{r_i}}$ where $\Delta\theta_{ij} = \theta_i - \theta_j$.

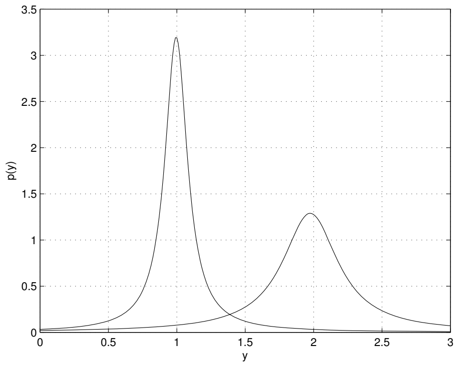


Fig. 1. A typical display of two Cauchy distribution. The parameters are $\theta_1 = \arctan(1)$, $\theta_2 = \arctan(2)$, $p = .9$, $\sigma_{r_i} = 1$ for $i = 1, 2$ and $\sigma_n = .05$.

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 $\boldsymbol{\theta} = [\frac{-\pi}{3}, \frac{-\pi}{4}, \frac{\pi}{6}, \frac{\pi}{4}]^T$. Therefore, we have $m = 4$ sources and $n = 2$ observations. The sparse sources are generated arti£cially 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 Fisher information matrix 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:

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

which is close to the actual matrix.

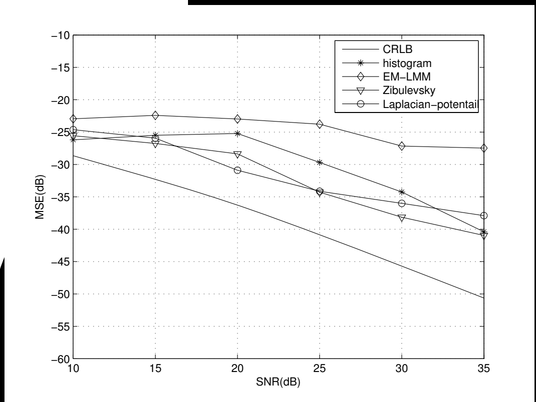
The results of four different methods of mixing matrix estimation are now compared with each other and also with the Cramer-Rao Lower bound (17). These four methods are potential-based clustering [3], histogram method [9], Laplacian-window potential-function-based clustering [12], and EM-LMM method [10]. In potential-based clustering [2], 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 de£ne the Mean Square Error (MSE) of the i'th mixing matrix angle (θ_i) as:

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

the simulations are then repeated 100 times with new random sparse sources and the MSE's were averaged over all experi-

[redacted]g. 2. A typical scatter p[redacted] [redacted]ector is $\boldsymbol{\theta} = [\frac{-\pi}{3}, \frac{-\pi}{4}, \frac{\pi}{6},$ [redacted] [redacted]or all i and $\sigma_n = .01$.



[redacted]

hen[redacted]
ure[redacted]
o th[redacted]
succ[redacted]
only[redacted]
expe[redacted]
Figu[redacted]
the [redacted]
of 4[redacted]
This[redacted]
the r[redacted]