Sparse Recovery Using an Iterative Variational Bayes Algorithm and Application to AoA Estimation

Ahmad Bazzi*†, Dirk T.M. Slock*, and Lisa Meilhac†
*EURECOM Mobile Communications Department, 450 route des Chappes, 06410 Biot Sophia Antipolis, France
Email: {bazzi,slock}@eurecom.fr
†CEVA-RivieraWaves, 400, avenue Roumanille Les Bureaux, Bt 6, 06410 Biot Sophia Antipolis, France
Email:{ahmad.bazzi, lisa.meilhac}@ceva-dsp.com

***Abstract*—This paper presents an iterative Variational Bayes (VB) algorithm that allows sparse recovery of the desired transmitted vector. The VB algorithm is derived based on the latent variables introduced in the Bayesian model in hand. The proposed algorithm is applied to the Angle-of-Arrival (AoA) estimation problem and simulations demonstrate the potential of the proposed VB algorithm when compared to existing sparse recovery and compressed sensing algorithms, especially in the case of closely spaced sources. Furthermore, the proposed algorithm does not require prior knowledge of the number of sources and operates with only one snapshot.**

***Index Terms*—Sparse Recovery, Variational Bayes, Iterative, Latent Variables, Angle-of-Arrival**

I. Introduction

The estimation of the angles of arrival, or AoAs, of multiple sources is a well known problem in the context of array signal processing. In fact, this problem is important in many applications such as positioning, radar, sonar, public safety, etc. [1]. The Maximum Likelihood [2] was one of the first techniques to be investigated, however, it involves a $q-$dimensional search, where q is the number of present sources. To cope with this issue, a tradeoff has been done between complexity and performance, hence suboptimal techniques with reduced complexity have dominated the field. The most famous ones are: Multiple Signal Classification (MUSIC) developed in [3] and [4], independently. Also, less complex algorithms were implemented to replace the 1-D search of MUSIC by a polynomial root finding process [5], or a least squares fit [6]. The performance of these algorithms are inferior to the ML technique.

Recently, sparse recovery optimisation and compressed sensing algorithms have become popular and found many applications in diverse areas in signal processing, speech, imaging, coding, and so forth. The area of compressed sensing was initiated in 2006 by two ground breaking papers, namely [7] by Donoho and [8] by Candes, Romberg, and Tao.

Consider the following linear model, which will be oriented towards the AoA estimation problem in the next section:

$$\boldsymbol{x} = \boldsymbol{\mathcal{A}}\boldsymbol{s} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{\mathrm{N}\times\mathrm{K}}$ is a known overcomplete dictionary. Each column of $\boldsymbol{\mathcal{A}}$ is referred to as an atom. The vector $\boldsymbol{s} \in \mathbb{C}^{\mathrm{K}\times 1}$ is composed of unknown coefficients that we would like to retrieve using the observed vector $\boldsymbol{x} \in \mathbb{C}^{\mathrm{N}\times 1}$. In general, this problem is underdetermined and therefore ill-posed. However, a typical remedy for this indeterminacy is to pose a sparse constraint on $\boldsymbol{s}$, which leads to the following sparse optimisation problem:

$$\hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}} \|\boldsymbol{x} - \boldsymbol{\mathcal{A}}\boldsymbol{s}\|^2 + \lambda\|\boldsymbol{s}\|_p \tag{2}$$

where $\|\boldsymbol{s}\|_p$ is the l_p norm of $\boldsymbol{s}$ and $0 < p \leq 2$. Note that we have excluded $p = 0$ since l_0 is a pseudo-norm (the distance property of norms is not satisfied), which counts the number of non-zero elements. Also note that $\|\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|$.

Sparsity is most favored when $p = 0$. However, the above optimisation problem will become NP-hard [9]. As a response to this issue, greedy algorithms have been implemented to solve the above optimisation problem under the l_0 constraint, such as Matching

Pursuit (MP) [10] and Orthogonal MP (OMP) [11]. An alternative to the NP-hard problem is to relax the constraint so as the problem is convex, i.e. this happens when $p \geq 1$ [12]. Popular algorithms that are used for the l_1 optimisation problem are the Iterative Shrinkage Thresholding Algorithm (ISTA) [13] and the Basis Pursuit Denoising (BPDN) [14]. Note that as opposed to classical approaches, i.e. subspace methods, the sparse approach could resolve coherent sources and does not require the knowledge of the number of sources.

Inspired by the papers [16]–[18], we take a Bayesian approach. The papers [16]–[18] focus on introducing latent, or hidden, variables and imposing prior distributions on these variables that favor sparsity. In this paper, we also introduce the latent variables discussed in [16]–[18], which leads to an iterative Variational Bayes [20] algorithm that allows recovering $\boldsymbol{s}$ from a single observation $\boldsymbol{x}$ with the help of the latent variables that are introduced. The algorithm iterates between parameters related to the latent variables and the variables of interest, i.e. $\boldsymbol{s}$. In addition, the proposed algorithm could discriminate closely spaced sources due to the excellent sparse solution it provides. This is shown in simulations when compared to other existing and popular algorithms.

This paper is organised as follows: Section II presents the System Model. In Section III, we take a Bayesian approach and introduce the latent variables as done in [17]. A Recap of Variational Bayes (VB) and the proposed algorithm based on VB is presented in Section VI. Section V demonstrates our simulation results. We conclude the paper in Section VI.

Notations: Upper-case and lower-case boldface letters denote matrices and vectors, respectively. $(.)^{\mathrm{T}}$ and $(.)^{\mathrm{H}}$ represent the transpose and the transpose-conjugate operators. $E_x\{.\}$ is the statistical expectation over the distribution of the random variable x. The operator tr denotes "trace". Finally, $|z|$ denotes the magnitude of $z \in \mathbb{C}$.

II. System Model

Consider a planar array composed of N antennas and assume $q < N$ narrowband sources impinge the array from different directions, i.e. $\theta_1 \dots \theta_q$. Sampling at an arbitrary time instance, we can express a single snapshot as [15]

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{t} + \boldsymbol{n} \tag{3}$$

where $\boldsymbol{x} \in \mathbb{C}^{\mathrm{N}\times 1}$ is the observed vector at the output of the antenna array. The vector $\boldsymbol{t} \in \mathbb{C}^{\mathrm{q}\times 1}$ is the $q \times 1$ source vector. The steering manifold, $\boldsymbol{A} \in \mathbb{C}^{\mathrm{N}\times \mathrm{q}}$ is composed of q steering vectors, i.e. $\boldsymbol{A} = [\boldsymbol{a}(\theta_1)\dots\boldsymbol{a}(\theta_q)]$. Each vector $\boldsymbol{a}(\theta_i)$ is the response of the array to a source attacking the antenna array from angle θ_i. For arbitrary arrays, $\boldsymbol{a}(\theta)$ is given as

$$\boldsymbol{a}(\theta) = \frac{1}{\sqrt{N}}\begin{bmatrix} e^{-j\frac{w_c}{c}(\bar{x}_1\sin(\theta)+\bar{y}_1\cos(\theta))} \\ \vdots \\ e^{-j\frac{w_c}{c}(\bar{x}_N\sin(\theta)+\bar{y}_N\cos(\theta))} \end{bmatrix} \tag{4}$$

where $(\bar{x}_i, \bar{y}_i)$ is the position of the i^{th} antenna. The term $w_c = 2\pi f_c$ is the angular frequency, and c is the speed of light in vacuum. The vector $\boldsymbol{n} \in \mathbb{C}^{\mathrm{N}\times 1}$ is background noise. The noise is modelled as a white circular complex Gaussian process of zero mean and covariance $\sigma^2 \boldsymbol{I}_{\mathrm{N}}$ and independent from the source signals.

Now, we recast the problem statement in (3) to the following

$$\boldsymbol{x} = \boldsymbol{\mathcal{A}}\boldsymbol{s} + \boldsymbol{n} \tag{5}$$

where $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{\mathrm{N}\times \mathrm{K}}$ is an overcomplete dictionary given as

$$\boldsymbol{\mathcal{A}} = [\boldsymbol{a}(\theta^1)\dots\boldsymbol{a}(\theta^{\mathrm{K}})] \tag{6}$$

and $\boldsymbol{s} \in \mathbb{C}^{\mathrm{K}\times 1}$ is a $q-$sparse (only q elements of $\boldsymbol{s}$ are not set to zero) vector. Note that the non-zero elements of $\boldsymbol{s}$ are equal to the corresponding elements of $\boldsymbol{t}$.

III. A Bayesian Approach

In this section, we shall take a *Bayesian* approach, i.e. the vector $\boldsymbol{s}$ is random and not an unknown deterministic vector. Adopting the *Bayesian* criterion is equivalent to optimising the *maximum aposteriori* (MAP) [15], which is given as

$$\hat{\boldsymbol{s}} = \arg\max_{\boldsymbol{s}} p(\boldsymbol{s}|\boldsymbol{x}) = \arg\max_{\boldsymbol{s}} \frac{p(\boldsymbol{x}|\boldsymbol{s})p(\boldsymbol{s})}{p(\boldsymbol{x})} \tag{7}$$

where $p(\boldsymbol{x}|\boldsymbol{s})$ is known as the likelihood function and $p(\boldsymbol{s})$ is referred to as the prior. It was noted in [16] and [17] that the following type of prior favors sparsity

$$p(\boldsymbol{s}) = \prod_{k=1}^{K} p(s_k), \quad p(s_k) = p(s_k|\beta_k)\phi(\beta_k) \tag{8}$$

where $\beta_1 \dots \beta_K$ are referred to as latent variables and

$$p(s_k|\beta_k) = \mathcal{N}(s_k; 0, \beta_k^{-1}) \tag{9}$$

and $\phi(\beta_k)$ is a nonnegative function. Now, the latent variables $\beta_1 \dots \beta_K$, which are treated as random variables, should have appropriate corresponding pdfs, i.e.

$\phi(\beta_1)\dots\phi(\beta_K)$, respectively. As explained in [19], the pdf $\phi(\beta_K)$ should be chosen as the conjugate to the Gaussian distribution. One possibility is the Gamma function, i.e.

$$\phi(\beta_k) = \Gamma(\beta_k;\gamma,\delta) \tag{10}$$

Moreover, let $\nu = \frac{1}{\sigma^2}$ be the inverse of the noise variance. Also, we allow ν to follow a Gamma prior, viz.

$$p(\nu) = \Gamma(\nu;\zeta,\eta) \tag{11}$$

The MAP criterion, with the formulation from equations (8) till (11) is now

$$p(\boldsymbol{s},\boldsymbol{\beta},\nu|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{s},\boldsymbol{\beta},\nu)p(\boldsymbol{s},\boldsymbol{\beta},\nu)}{p(\boldsymbol{x})} \tag{12}$$

with $\boldsymbol{\beta} = [\beta_1\dots\beta_K]$. Assuming independency between the signal vector $\boldsymbol{s}$ and the noise, we can say that

$$p(\boldsymbol{s},\boldsymbol{\beta},\nu) = p(\boldsymbol{s}|\boldsymbol{\beta})p(\boldsymbol{\beta})p(\nu) \tag{13}$$

Finally, we notice that the normalisation factor in equation (13) given as

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x}|\boldsymbol{s},\nu)p(\boldsymbol{s}|\boldsymbol{\beta})p(\boldsymbol{\beta})p(\nu)d\boldsymbol{s}d\nu d\boldsymbol{\beta} \tag{14}$$

does not have a closed-form expression; hence we propose to use the Variational Bayes methodology.

IV. Variational Bayes

A. Methodology

Let $\boldsymbol{y} = [\boldsymbol{\beta},\nu]$. The log-likelihood function that does not take into account the latent variables $\boldsymbol{y}$, is given as follows [20]

$$\log\, p(\boldsymbol{x}|\boldsymbol{s}) = \int q(\boldsymbol{y})\,\log\,\Big(\frac{p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{s})}{q(\boldsymbol{y})}\Big)d\boldsymbol{y} + \mathrm{KL}\big(q||p\big) \tag{15}$$

where $\mathrm{KL}\big(q||p\big)$ is the *Kullback-Leilbler* divergence between $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{s})$ and $q(\boldsymbol{y})$. Since $\mathrm{KL}\big(q||p\big) \geq 0$, then

$$\log\, p(\boldsymbol{x}|\boldsymbol{s}) \geq \int q(\boldsymbol{y})\,\log\,\Big(\frac{p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{s})}{q(\boldsymbol{y})}\Big)d\boldsymbol{y} \tag{16}$$

The methodology of Variational Bayes lies in maximising the lower bound in equation (16) by imposing a factorised structure on $\boldsymbol{y}$ as follows [20]

$$q(\boldsymbol{y}) = \prod_{k=1}^{K+1} q_k(y_k) \tag{17}$$

Table 1: Proposed Variational Bayes algorithm for AoA Estimation

INPUT:
Given the observed vector $\boldsymbol{x} = \boldsymbol{\mathcal{A}}\boldsymbol{s} + \boldsymbol{n}$.

INITIALISATION:

- Fix
$$\gamma = \delta = \zeta = \eta = 10^{-6}$$
$$n = 0$$
- Initialise
$$m_{\beta_k}^{(0)} = 10^5 \text{ for all } k$$
$$m_{\nu}^{(0)} = \frac{1}{\sigma_n^2}$$

MAIN LOOP:
while $\|\hat{\boldsymbol{\Theta}}^{(n+1)} - \hat{\boldsymbol{\Theta}}^{(n)}\| > \xi$ *(Pre-defined Threshold)* **do**

- Form
$$\boldsymbol{\Omega}^{(n)} = \text{diag }[m_{\beta_1}^{(n)}\dots m_{\beta_K}^{(n)}]$$
- Compute $\boldsymbol{\Sigma}$ as in equation (25)
$$\boldsymbol{\Sigma}^{(n)} = \Big(\boldsymbol{\Omega}^{(n)} + m_{\nu}^{(n)}\boldsymbol{\mathcal{A}}^{\mathrm{H}}\boldsymbol{\mathcal{A}}\Big)^{-1}$$
- Compute $\boldsymbol{m_s}$ using (26)
$$\boldsymbol{m_s}^{(n)} = m_{\nu}^{(n)}\boldsymbol{\Sigma}^{(n)}\boldsymbol{\mathcal{A}}^{\mathrm{H}}\boldsymbol{x}$$
- For all $k = 1\dots K$, compute m_{β_k} using (30)
$$m_{\beta_k}^{(n+1)} = \frac{2\gamma+1}{2\delta + |(\boldsymbol{m_s}^{(n)})_k|^2 + \boldsymbol{\Sigma}_{k,k}^{(n)}}$$
- Compute m_ν using (31)
$$m_{\nu}^{(n+1)} = \frac{2\zeta+1}{2\eta + \|\boldsymbol{x} - \boldsymbol{\mathcal{A}}\boldsymbol{m_s^{(n)}}\|^2 + \mathrm{tr}\ \{\boldsymbol{\mathcal{A}}\boldsymbol{\Sigma^{(n)}}\boldsymbol{\mathcal{A}}^{\mathrm{H}}\}}$$
- Increment n
$$n \leftarrow n+1$$

OUTPUT:
The estimate of $\boldsymbol{s}$ is
$$\hat{\boldsymbol{s}} = \boldsymbol{m_s}^{(n)}$$

Substituting the form of $q(\boldsymbol{y})$ in (16) and following [20], this lower bound could be expressed as follows:

$$\int q(\boldsymbol{y})\,\log\,\Big(\frac{p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{s})}{q(\boldsymbol{y})}\Big)d\boldsymbol{y} = -\sum_{\substack{k=1\\k\neq i}}^{K+1}\int q_k(y_k)\,\log\,q_k(y_k)dy_k - \mathrm{KL}(q_i||\bar{p}_i) \tag{18}$$

with

$$\hat{p}_i \triangleq \mathbb{E}_{y_k \neq y_i}\Big\{\log\, p(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{y})\Big\} = \int \log\big(p(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{y})\big) \prod_{\substack{k=1 \\ k\neq i}}^{K+1} q_k(y_k) dy_k \tag{19}$$

It is straightforward to see that the lower bound is maximising when $\mathrm{KL}(q_i || \bar{p}_i) = 0$. In other words, each $q_i(y_i)$ should be chosen as

$$\log\, q_i(y_i) = \mathbb{E}_{y_k \neq y_i}\Big\{\log\, p(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{y})\Big\} + C \tag{20}$$

where C is a normalisation constant. Now, following [21], one could solve for $\boldsymbol{s}$, in a Variational Expectation-Maximisation (EM) iterative manner as follows:

- Variational E-step: Given $\boldsymbol{s}^{(n)}$ (i.e. the value of $\boldsymbol{s}$ at iteration n), compute $q_i^{(n)}(y_i)$ for all i using equation (20).
- Variational M-step: Given $q_i^{(n)}(y_i)$ for all i, compute $\boldsymbol{s}^{(n+1)}$ that maximises equation (18).

Now, we are ready to apply the Variational Bayes methodology to the problem in hand.

B. Variational Bayes AoA Estimation

We first start off by deriving the expressions of $q_i(y_i)$ and $q(\boldsymbol{s})$. Following the factorised structure of $\boldsymbol{y}$ in equation (17) and the independency between $\boldsymbol{s}$ and $\boldsymbol{y}$, we can say that the posterier factorises as follows

$$p(\boldsymbol{s}, \boldsymbol{y} | \boldsymbol{x}, \gamma, \delta, \zeta, \eta) = p(\boldsymbol{s}) p(\boldsymbol{y}) = p(\boldsymbol{s}) p(\boldsymbol{\beta}) p(\nu) \tag{21}$$

With the help of equation (20), we now analytically evaluate $q(\boldsymbol{s})$ as follows

$$\begin{aligned} \log\, q(\boldsymbol{s}) &= \mathbb{E}_{\boldsymbol{\beta},\nu}\Big\{\log\, p(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{y})\Big\} \\ &= \mathbb{E}_{\boldsymbol{\beta},\nu}\Big\{\log\, p(\boldsymbol{x}|\boldsymbol{s}, \nu) p(\boldsymbol{s}|\boldsymbol{\beta})\Big\} \\ &= -\frac{1}{2}\mathbb{E}_{\boldsymbol{\beta},\nu}\Big\{\nu \|\boldsymbol{x} - \boldsymbol{\mathcal{A}}\boldsymbol{s}\|^2 + \sum_{k=1}^{K} \beta_k |s_k|^2\Big\} \end{aligned} \tag{22}$$

With some abuse of notation, $\mathbb{E}_{\boldsymbol{\beta},\nu}$ is the average over the joint distributions $q(\boldsymbol{\beta})$ and $q(\nu)$. In addition, we have omitted the constant in equation (22) for the sake of compact presentation. Now, assuming that $\boldsymbol{\beta}$ and ν are independent, we can say

$$\log\, q(\boldsymbol{s}) = -\frac{m_\nu}{2}\|\boldsymbol{x} - \boldsymbol{\mathcal{A}}\boldsymbol{s}\|^2 - \frac{1}{2}\sum_{k=1}^{K} m_{\beta_k} |s_k|^2 \tag{23}$$

where $m_\nu = \mathbb{E}\{\nu\}$ and $m_{\beta_k} = \mathbb{E}\{\beta_k\}$. With some mathematical steps, one could show that $q(\boldsymbol{s})$ is given as follows

$$\log\, q(\boldsymbol{s}) = -\frac{1}{2}\big(\boldsymbol{s} - \boldsymbol{m_s}\big)^{\mathrm{H}} \boldsymbol{\Sigma}^{-1} \big(\boldsymbol{s} - \boldsymbol{m_s}\big) \tag{24}$$

where

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Omega} + m_\nu \boldsymbol{\mathcal{A}}^{\mathrm{H}} \boldsymbol{\mathcal{A}} \tag{25}$$

and

$$\boldsymbol{m_s} = m_\nu \boldsymbol{\Sigma} \boldsymbol{\mathcal{A}}^{\mathrm{H}} \boldsymbol{x} \tag{26}$$

where $\boldsymbol{\Omega} = \mathrm{diag}[m_{\beta_1} \dots m_{\beta_K}]$. Now, we compute $q(\boldsymbol{\beta})$

$$\begin{aligned} &\log\, q(\boldsymbol{\beta}) \\ &= \mathbb{E}_{\boldsymbol{s},\nu}\Big\{\log\, p(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{y})\Big\} \\ &= \mathbb{E}_{\boldsymbol{s},\nu}\Big\{\sum_{k=1}^{K}\Big(\log\, p(\beta_k)\Big) + \log\, p(\boldsymbol{s}|\boldsymbol{\beta})\Big\} \\ &= \sum_{k=1}^{K}\Big((\gamma - 1)\log\, \beta_k - \delta\beta_k + \frac{1}{2}\log\, \beta_k - \beta_k \mathbb{E}|s_k|^2\Big) \end{aligned} \tag{27}$$

where the terms $(\gamma - 1)\log\, \beta_k$ and $\delta\beta_k$ appear due to K independent Gamma distributions, i.e. $p(\beta_k)$ for $k = 1 \dots K$. Again, with some abuse of notation, we have ommited constant terms for the sake of compact presentation. With some straightforward algebra, we could say that

$$q(\beta_k) = \Gamma\Big(\beta_k; \frac{2\gamma + 1}{2}, \frac{2\delta + |(\boldsymbol{m_s})_k|^2 + \boldsymbol{\Sigma}_{k,k}}{2}\Big) \tag{28}$$

where $(\boldsymbol{m_s})_k$ is the k^{th} entry of vector $\boldsymbol{m_s}$ and $\boldsymbol{\Sigma}_{k,k}$ is the element found in the k^{th} diagonal of $\boldsymbol{\Sigma}$. In a similar manner, we could show that

$$q(\nu) = \Gamma\Big(\nu; \frac{2\zeta + 1}{2}, \frac{2\eta + \|\boldsymbol{x} - \boldsymbol{\mathcal{A}}\boldsymbol{m_s}\|^2 + \mathrm{tr}\ \{\boldsymbol{\mathcal{A}}\boldsymbol{\Sigma}\boldsymbol{\mathcal{A}}^{\mathrm{H}}\}}{2}\Big) \tag{29}$$

Knowing that for any random variable following a Gamma distribution with parameters λ and μ, i.e. $X \sim \Gamma(x; \lambda, \mu)$, the mean of X, say m_X, is given as $m_X = \frac{\lambda}{\mu}$. Therefore, it is easy to see from equation (28) that

$$m_{\beta_k} = \frac{2\gamma + 1}{2\delta + |(\boldsymbol{m_s})_k|^2 + \boldsymbol{\Sigma}_{k,k}} \tag{30}$$

Similarly, equation (29) implies that

$$m_\nu = \frac{2\zeta + 1}{2\eta + \|\boldsymbol{x} - \boldsymbol{\mathcal{A}}\boldsymbol{m_s}\|^2 + \mathrm{tr}\ \{\boldsymbol{\mathcal{A}}\boldsymbol{\Sigma}\boldsymbol{\mathcal{A}}^{\mathrm{H}}\}} \tag{31}$$

Before presenting the algorithm in **Table 1**, we find the following notation useful

$$\boldsymbol{\Theta} = [m_{\beta_1} \dots m_{\beta_K}, m_\nu, \boldsymbol{s}] \tag{32}$$

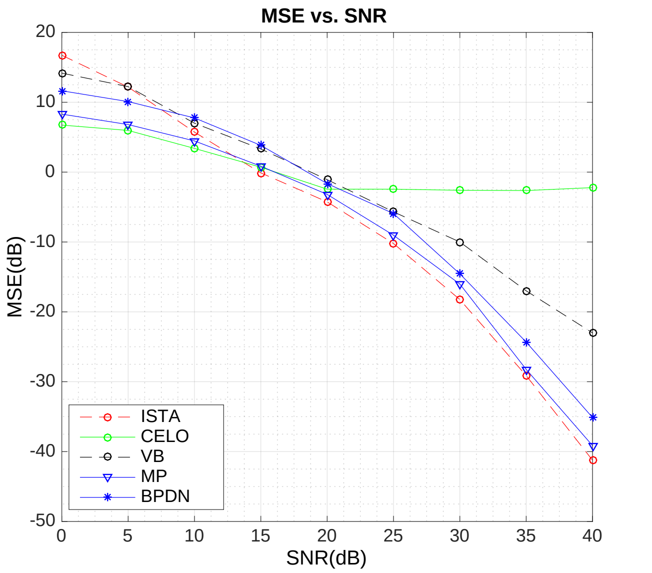

Fig. 1: Two sources impinging the array from directions $\theta_1 = 0°$ and $\theta_2 = 30°$. The number of antennas is 10.

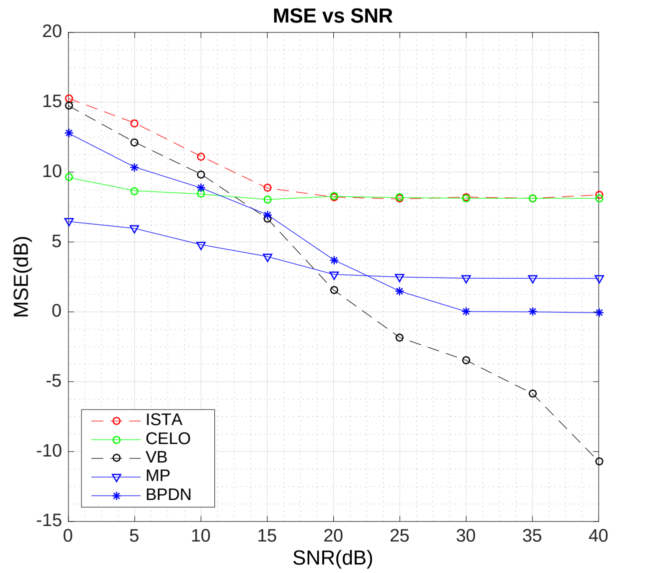

Fig. 2: Two sources impinging the array from directions $\theta_1 = 0°$ and $\theta_2 = 10°$. The number of antennas is 10.

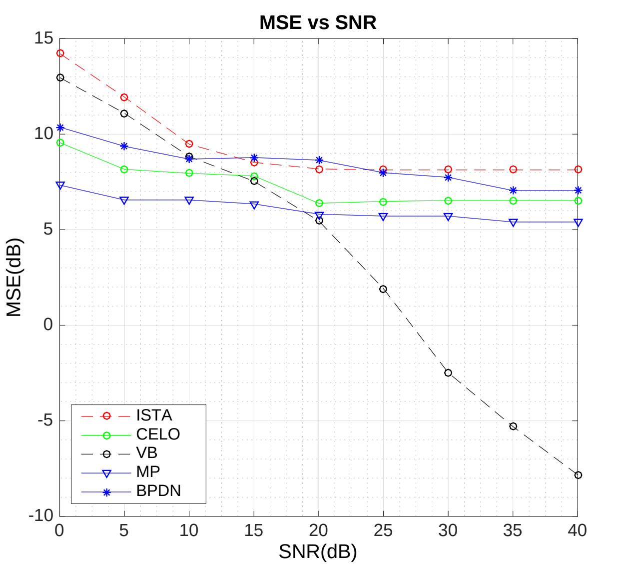

Fig. 3: Two sources impinging the array from directions $\theta_1 = 0°$ and $\theta_2 = 5°$. The number of antennas is 10.

Furthermore, let $x^{(n)}$ denote the value of the quantity x at iteration n. For convenience, $x^{(0)}$ is the initial value of x. Now, we are ready to state the iterative algorithm that is based on Variational EM as explained in Section IV.A. The algorithm is given in **Table 1**.

V. Simulation Results

In this section, we present our simulation results with emphasis on closely spaced sources. In all the experiments done, we fix the following: Consider a Uniform Linear Antenna array composed of $N = 10$ antennas spaced at half a wavelength. Furthermore, assume $q = 2$ sources attacking the array from directions $\theta_1 = 0°$ and θ_2. The dictionary $\boldsymbol{\mathcal{A}}$ is composed of $K = 91$ atoms discretized from $-45°$ till $+45°$ as follows

$$\boldsymbol{\mathcal{A}} = [\boldsymbol{a}(-45), \boldsymbol{a}(-44) \dots \boldsymbol{a}(0) \dots \boldsymbol{a}(44), \boldsymbol{a}(45)] \quad (33)$$

Define the Mean Squared Error (MSE) as follows

$$MSE = \frac{1}{qM} \sum_{i=1}^{q} \sum_{k=1}^{M} (\theta_i - \hat{\theta}_{i,k})^2 \quad (34)$$

where $\hat{\theta}_{i,k}$ is the estimate of the i^{th} AoA (i.e. θ_i) at the k^{th} Monte-Carlo trial and M is the number of Monte-Carlo trials. All our experiments are done using $M = 200$ trials.

We would strongly like to note the following: Generally, in compressive sensing and sparse recovery algorithms, the estimates $\hat{\theta}_1 \dots \hat{\theta}_q$ are obtained by choosing the corresponding q largest magnitudes of the sparse recovered vector $\hat{\boldsymbol{s}}$ that activate the atoms in the dictionary. Indeed, these estimates vary according to the size and elements of the dictionary.

In the 3 experiments, we plot the MSE vs. SNR for different values of θ_2. Furthermore, we compare with MP, BPDN, ISTA, and Continuous Exact l_0 Penalty (CELO), which was recently introduced in [22] and applied to the AoA Estimation problem in [23].

In experiment 1, i.e. Figure 1, θ_2 is chosen to be $30°$. In experiment 2, i.e. Figure 2, θ_2 is brought closer to $\theta_1 = 0°$ and chosen to be $\theta_2 = 10°$. Finally, in experiment 3 (Figure 3), θ_2 is even more close to θ_1 and is tuned to be $\theta_2 = 5°$. We notice that the proposed VB algorithm is able to resolve closely spaced sources, whereas all other algorithms fail in doing so. This is due to the high MSE error (MSE $>$ 5 dB) present in all algorithms even at high SNR.

VI. Conclusion

In this paper, and with the help of latent variables and Variational Bayes, we have derived an iterative algorithm that could estimate the Angles of Arrival (AoA) of the incoming sources with a single snapshot, without the knowledge of the number of sources, and with closely spaced sources at high SNR.

Future work may be oriented towards performance analysis of the proposed Variational Bayes algorithm and towards taking into account prior knowledge of the number of source signals, which may improve the performance of this algorithm.

Acknowledgment

This work was supported in part by RivieraWaves, a CEVA company, and a Cifre scholarship. EURECOM's research is partially supported by its industrial members and by the EU projects ADEL (FP7) and HIGHTS (H2020).

References

[1] Tuncer, E. and B. Friedlander, "Classical and Modern Direction of Arrival Estimation," Elsevier, Burlington, MA, 2009.

[2] W. S. Ligget, "Passive sonar: Fitting models to multiple time series," in NATO ASI on Signal Processing, J. W. R. Griffiths et al., Eds. New York: Academic, pp. 327-345, 1973.

[3] R. 0 . Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, pp. 276- 280, 1986.

[4] G . Bienvenu and L. Kopp, "Adaptivity to background noise spatial coherence for high resolution passive methods," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 80, pp. 307-310, 1980.

[5] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction finding algorithms," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 336-339, 1983.

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

[7] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, 52:12891306, 2006.

[8] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information," *IEEE Transactions on Information Theory*, 52:489-509, 2006.

[9] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, 24, pp. 227-234, 1995.

[10] S.G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, 41, pp. 3397-3415, 1993.

[11] J. Tropp and A.C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, 53, pp. 4655-4666, 2007.

[12] Y.C. Eldar and G. Kutyniok, "Compressed Sensing: Theory and applications," Cambridge, U.K.: Cambridge Univ. Press, 2012.

[13] I. Daubechies, M. Defrise, and C. D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.*, 57, pp. 1413-1457, 2004.

[14] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, 20, pp. 33-61, 1998.

[15] H. L. Van Trees, "Detection, Estimation, and Modulation Theory," Part IV, Optimum Array Processing. New York: Wiley, 2002.

[16] D. Wipf, "Bayesian methods for finding sparse representations," Ph. D., Univ. Calif., San Diego, 2006.

[17] D. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *IEEE Transactions on Information Theory*, 2010.

[18] D. Wipf, J. Palmer, B. Rao, and K. Kreutz-Delgado, "Performance analysis of latent variable models with sparse priors," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr. 2007.

[19] C. M. Bishop, "Pattern Recognition and Machine Learning," *Springer*, 2006.

[20] M. J. Beal, "Variational algorithms for approximate Bayesian inference," PhD Thesis, University College London, 2003.

[21] M. J. Beal and Z. Ghahrarmani, "The variational Bayesian EM algorithm for incomplete data: With application to scoring graphical model structures," in Bayesian Statistics. Oxford, U.K.: Oxford Univ. Press, vol. 7, 2002.

[22] E. Soubies, L. Blanc-Feraud, and G. Aubert, A Continous Exact l_0 penalty (CEL0) for least squares regularized problem, *SIAM J. Imaging Sci.*, pp. 1574-1606, 2015.

[23] A. Chinatto, E. Soubies, C. Junqueira, J. M. T. Romano, P. Larzabal,et al "l_0-Optimization for Channel and DOA Sparse Estimation, "*International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 4, Cancun, Mexico, Dec 2015.