Single Snapshot Direction of Arrival Estimation Using the EP-SURE-SBL Algorithm

Fangqing Xiao, Dirk Slock
Communication Systems Department
Eurecom, France
Email: {fangqing.xiao, dirk.slock}@eurecom.fr

***Abstract*—Grid-based methods in sparse signal reconstruction (SSR) are well-regarded for their efficacy in direction-of-arrival (DoA) estimation. This paper presents the EP (Expectation Propagation)-SURE (Stein's Unbiased Risk Estimate)-SBL (Sparse Bayesian Learning) algorithm, designed for single snapshot DoA estimation. The algorithm divides DoA estimation into two parts: grid-on estimation and off-grid error estimation, employing first-order and second-order Taylor expansions. In grid-on estimation, sparse Bayesian learning is employed for sparse modeling. To tackle hyperparameter estimation challenges within sparse Bayesian learning, the algorithm adopts SURE estimator instead of the commonly-used expectation-maximization (EM) approach. For off-grid error estimation, the algorithm utilizes the EP technique to handle high-dimensional, non-tractable integration in posterior mean calculations. The feasibility and effectiveness of the proposed algorithm are validated through extensive simulations.**

I. Introduction

Grid-based methods, treated as sparse signal reconstruction (SSR), have gained interest in direction-of-arrival (DoA) estimation in recent years. Various algorithms, such as Orthogonal Matching Pursuit (OMP) [1], basis pursuit [2], least absolute shrinkage and selection operator (LASSO) [3], and iterative re-weighted l_1 and l_2 algorithms [4], address the SSR problem. Compared to these algorithms, Bayesian techniques for SSR generally achieve superior performance while [5] provides a detailed overview of SSR algorithms based on l_1 or l_2 norm minimization approaches, such as Basis Pursuit and LASSO, and Sparse Bayesian Learning (SBL) methods, highlighting SBL's superior recovery performance. The SBL algorithm, first introduced by [6] and proposed for SSR by [7], aims to calculate the posterior distribution of the parameter $\mathbf{x}$ given some observations (data) and prior knowledge. Moreover, Wipf and Rao [8] have theoretically shown that the SBL framework can obtain the true (sparsest) solution, whereas assigning informative priors may lead to unstable algorithms or incorrect. In SBL, the unknown hyperparameters (signal variance and noise variance) need to be estimated jointly. One approach uses the expectation maximization (EM) algorithm [9] for this estimation. Although EM can converge to the maximum likelihood estimator, it cannot recover exact zero variances of signals with a finite number of iterations, meaning it cannot recover sparse signals precisely unless manual tuning methods are leveraged. To address this, we use Stein's unbiased risk estimator (SURE) [10] to recover hyperparameters, termed SURE-SBL [11]. For the SURE estimator, as data size approaches infinity, optimizing SURE is equivalent to minimizing mean squared error (MSE). In our previous work [12], we demonstrated that SBL achieves better performance than LASSO from an MSE perspective.

The advantage of grid-based methods can operate effectively, even with a single snapshot, provided that all source spatial frequencies align exactly with the preset grid. This is in contrast to traditional methods such as MUltiple SIgnal Classification (MUSIC) [13] estimator and ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) estimator [14]. However, in practice, this condition is almost impossible to meet, as the region of interest (ROI) contains infinite candidates, leading to grid mismatch when splitting the ROI into a finite number of grids. This off-grid issue has attracted significant research interest in array signal processing over the past decade. Existing solutions to address the off-grid problem can be categorized into three groups. The first group employs denser grids or a coarse-to-fine strategy, as discussed by [15]. However, these methods have two main drawbacks: denser grids significantly increase computational complexity, and overly dense grids may result in weak incoherence among the steering vectors. The second group includes the so-called gridless approaches [16], which are limited to regularly sampled measurements from a uniform linear array (ULA). The third group, which we utilize in this paper, estimates the off-grid bias along with the grids closest to the true spatial frequencies. We extend the measurement matrix at the grid-on place using Taylor expansion. Based on a Bayesian framework, we develop the posterior mean of off-grid error, assuming a uniform distribution. Due to the intractable nature of high-dimensional integration, we employ an expectation propagation (EP)-based [17] method to address it. In this paper, we propose a novel DoA estimation algorithm based on a single snapshot, leveraging Sparse Bayesian Learning (SBL), Expectation Propagation (EP), and Stein's Unbiased Risk Estimate (SURE), coined as the EP-SURE-SBL algorithm. This innovative combination of techniques enhances the precision and robustness of DoA estimation. Numerical simulations validate the robustness and effectiveness of our method.

A. Notations

The notation $\mathcal{CN}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the complex Gaussian distribution evaluated at $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Here, $\boldsymbol{H}_i$ represents the ith column of matrix $\boldsymbol{H}$, while a_i denotes the ith element of vector $\boldsymbol{a}$. The function $\mathrm{diag}(\boldsymbol{a})$ signifies a square diagonal matrix with $\boldsymbol{a}$ as its main diagonal. Additionally, $(\cdot)^H$, $(\cdot)^T$, and $(\cdot)^*$ respectively indicate the Hermitian transpose, transpose, and adjoint of a matrix or vector. $\boldsymbol{\beta}_{\bar{n}}$ denotes the $\boldsymbol{\beta}$ vector with the nth entry set to 0. $\boldsymbol{I}_M$ denote the $M \times M$ identity matrix, and $\Re\{a\}$ represent the real part of the complex number a. $\mathrm{rank}(\boldsymbol{A})$ denotes the rank of matrix $\boldsymbol{A}$.

II. Data Model

Consider a single snapshot measurement and K narrowband far-field sources s_k, $k = 1, \cdots, K$, impinging on an array of M omnidirectional sensors from direction θ_k, $k = 1, \cdots, K$. Time delays at different sensors can be represented by simple phase shifts, leading to the observation model:

$$\boldsymbol{y} = \boldsymbol{\Phi}(\boldsymbol{\theta})\boldsymbol{s} + \boldsymbol{v}, \tag{1}$$

where $\boldsymbol{y} = [y_1, \cdots, y_M]^T$ represent the vector of outputs, and $\boldsymbol{v}(t) = [v_1, \cdots, v_M]^T$ represents the vector of measurement noise for each sensor. Additionally, $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_K]^T$ denotes the vector

of unknown directions of arrival (DoAs), and $\boldsymbol{s} = [s_1, \cdots, s_K]^T$ represents the source signals. The matrix $\boldsymbol{\Phi}(\boldsymbol{\theta}) = [\boldsymbol{\phi}(\theta_1), \cdots, \boldsymbol{\phi}(\theta_K)]$ serves as the array manifold matrix, with $\boldsymbol{\phi}(\theta_k)$ referred to as the steering vector corresponding to the k-th source.

Each entry $\phi_m(\theta_k) = \exp(-2\pi(m-1)\sin(\theta_k)d/\lambda)$ encapsulates the delay information from the k-th source to the m-th sensor. Throughout this paper, we assume the number of sources K is known. Hence, the primary objective is to estimate the unknown DoAs $\boldsymbol{\theta}$ based on the given source number K, $\boldsymbol{y}$, and the mapping $\boldsymbol{\theta} \to \boldsymbol{\Phi}(\boldsymbol{\theta})$.

In the ensuing sections, we derive the off-grid model using linear approximation and establish its connection with the on-grid approach.

Let $\tilde{\boldsymbol{\theta}} = \{\tilde{\theta}_1, \cdots, \tilde{\theta}_N\}$ be a fixed sampling grid in the DoA range $[-\frac{\pi}{2}, \frac{\pi}{2})$, where N represents the number of grid points. Typically, $K < M \ll N$. Without loss of generality, $\tilde{\boldsymbol{\theta}}$ is a uniform grid with a grid interval $l = \tilde{\theta}_N - \tilde{\theta}_{N-1} = \pi/N$. Suppose $\theta_k \notin \{\tilde{\theta}_1, \cdots, \tilde{\theta}_N\}$ for some $k \in \{1, \cdots, K\}$ and that $\tilde{\theta}_{n_k}$, $n_k \in \{1, \cdots, N\}$ is the nearest grid point to θ_k. We approximate the steering vector $\boldsymbol{\phi}(\theta_k)$ using second-order Taylor linearization:

$$\boldsymbol{\phi}(\theta_k) \approx \boldsymbol{a}(\tilde{\theta}_{n_k}) + \boldsymbol{b}(\tilde{\theta}_{n_k})(\theta_k - \tilde{\theta}_{n_k}) + \tfrac{1}{2}\boldsymbol{c}(\tilde{\theta}_{n_k})(\theta_k - \tilde{\theta}_{n_k})^2, \quad (2)$$

with $\boldsymbol{a}(\tilde{\theta}_{n_k}) = \boldsymbol{\phi}(\tilde{\theta}_{n_k})$, $\boldsymbol{b}(\tilde{\theta}_{n_k}) = \frac{\partial \boldsymbol{\phi}(\theta_k)}{\partial \theta_k}\Big|_{\theta_k = \tilde{\theta}_{n_k}}$ and $\boldsymbol{c}(\tilde{\theta}_{n_k}) = \frac{\partial^2 \boldsymbol{\phi}(\theta_k)}{\partial \theta_k^2}\Big|_{\theta_k = \tilde{\theta}_{n_k}}$. Denote $\boldsymbol{A} = [\boldsymbol{a}(\tilde{\theta}_1), \cdots, \boldsymbol{a}(\tilde{\theta}_N)]$, $\boldsymbol{B} = [\boldsymbol{b}(\tilde{\theta}_1), \cdots, \boldsymbol{b}(\tilde{\theta}_N)]$, $\boldsymbol{C} = [\boldsymbol{c}(\tilde{\theta}_1), \cdots, \boldsymbol{c}(\tilde{\theta}_N)]$ and $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_N]^T$, where for $n = 1, \cdots, N$,

$$\begin{aligned} &\beta_n = \theta_k - \tilde{\theta}_{n_k}, \quad \beta_n \in (-\tfrac{l}{2}, \tfrac{l}{2}); \\ &x_n = s_{n_k}, \text{if } n = n_k, \text{ for any } k \in \{1, \cdots, K\}; \\ &x_n = 0, \text{otherwise}, \end{aligned} \quad (3)$$

with $n_k \in \{1, \cdots, N\}$ and $\tilde{\theta}_{n_k}$ being the nearest grid to a source θ_k, $k \in \{1, \cdots, K\}$. By absorbing the approximation error into measurement noise, the observation model in (1) can be re-expressed as follows:

$$\boldsymbol{y} = \left[\boldsymbol{A} + \boldsymbol{B}\ \mathrm{diag}(\boldsymbol{\beta}) + \tfrac{1}{2}\boldsymbol{C}\ \mathrm{diag}(\boldsymbol{\beta})^2\right]\boldsymbol{x} + \boldsymbol{v}. \quad (4)$$

It is worth noting that setting $\boldsymbol{C}$ to be a zero matrix in equation (4) causes the second-order Taylor expansion to reduce to a first-order approximation. To estimate the DoA $\boldsymbol{\theta}$, it is essential to determine both the support of the sparse signal $\boldsymbol{x}$ and the off-grid difference $\boldsymbol{\beta}$. This paper adopts a Bayesian framework to formulate the problem and introduces an iterative algorithm in the subsequent section for the joint estimation of $\boldsymbol{x}$ and $\boldsymbol{\beta}$.

III. Off-Grid Sparse Bayesian Inference

A. Sparse Bayesian Formulation

1) Noise Model: Under the assumption of white (circular symmetric) complex Gaussian noise, we have

$$p(\boldsymbol{v}|\lambda) = \mathcal{CN}(\boldsymbol{v}|\boldsymbol{0}, \lambda \boldsymbol{I}_M), \quad (5)$$

where λ denotes the noise variance. Then we have

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}) = \mathcal{CN}(\boldsymbol{y}|[\boldsymbol{A} + \boldsymbol{B}\mathrm{diag}(\boldsymbol{\beta}) + \tfrac{1}{2}\boldsymbol{C}\mathrm{diag}(\boldsymbol{\beta})^2]\boldsymbol{x}, \lambda \boldsymbol{I}_M). \quad (6)$$

In this paper, we assume that the noise variance λ is unknown and needs to be estimated later on.

2) Sparse Signal Model: In Sparse Bayesian Learning (SBL), the unknown parameters $\boldsymbol{x}$ are represented as decorrelated zero-mean (circular symmetric) complex Gaussian distributions:

$$p(\boldsymbol{x}|\boldsymbol{\gamma}) = \textstyle\prod_{n=1}^N \mathcal{CN}(x_n|0, \gamma_n), \quad (7)$$

where $\boldsymbol{\gamma}$ represents the unknown parameters. γ_n is a nonnegative parameter that controls the sparsity of $\boldsymbol{x}$. Specifically, when $\gamma_n = 0$, x_n is constrained to be 0. Throughout the learning process, many γ_n tend to approach 0 due to the mechanism of automatic relevance determination. In this framework, we treat all $\{\gamma_n\}_{n=1}^N$ as deterministic parameters of unknown values, without assuming a specific prior distribution for them, effectively treating them as random variables. This approach assumes a non-informative prior distribution for $\boldsymbol{\gamma}$.

3) Off-Grid Bias Model: We assume a uniform prior for $\boldsymbol{\beta}$:

$$p(\boldsymbol{\beta}) = \textstyle\prod_{n=1}^N p(\beta_n) = U\left(\left(-\frac{l}{2}, \frac{l}{2}\right)^N\right), \quad (8)$$

where U denotes the uniform distribution, and $\left(-\frac{l}{2}, \frac{l}{2}\right)^N$ specifies the bounds within which each component β_n of $\boldsymbol{\beta}$ is confined. This prior is considered non-informative because it only reflects the bounded nature of $\boldsymbol{\beta}$.

By combining the stages of our hierarchical Bayesian model, the joint probability density function is expressed as:

$$p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\beta}|\boldsymbol{\gamma}) = p(\boldsymbol{x}|\boldsymbol{\gamma})\, p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta})\, p(\boldsymbol{\beta}), \quad (9)$$

where the distributions on the right-hand side are defined by (6), (7), and (8) respectively. This formulation encapsulates how the data $\boldsymbol{y}$, the unknown parameters $\boldsymbol{x}$ governed by $\boldsymbol{\gamma}$, and the regression parameters $\boldsymbol{\beta}$ are interconnected under the assumptions of the Bayesian framework.

B. Stein's Unbiased Risk Estimator for Estimating $\boldsymbol{\gamma}$

We estimate $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ separately. Initially, assuming $\boldsymbol{\beta}$ has been estimated as $\hat{\boldsymbol{\beta}}$. For the sake of convenience, let us define

$$\boldsymbol{H}(\hat{\boldsymbol{\beta}}) = \boldsymbol{A} + \boldsymbol{B}\ \mathrm{diag}(\hat{\boldsymbol{\beta}}) + \tfrac{1}{2}\boldsymbol{C}\ \mathrm{diag}(\hat{\boldsymbol{\beta}})^2, \quad (10)$$

Referring to (10) in (4), we obtain:

$$\boldsymbol{y} = \boldsymbol{H}(\hat{\boldsymbol{\beta}})\boldsymbol{x} + \boldsymbol{v}. \quad (11)$$

According to the Gaussian-Markov theorem, the posterior distribution of $\boldsymbol{x}$ is a complex Gaussian with the probability density function given by:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \mathcal{CN}(\boldsymbol{x}|\boldsymbol{\Gamma}\boldsymbol{H}(\hat{\boldsymbol{\beta}})^H \boldsymbol{R}^{-1}\boldsymbol{y}, \boldsymbol{\Gamma} - \boldsymbol{\Gamma}^H \boldsymbol{H}(\hat{\boldsymbol{\beta}})^H \boldsymbol{R}^{-1}\boldsymbol{H}(\hat{\boldsymbol{\beta}})\boldsymbol{\Gamma}), \quad (12)$$

where $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$ is the diagonal covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{R} = \boldsymbol{H}(\hat{\boldsymbol{\beta}})\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{H}(\hat{\boldsymbol{\beta}})^H + \lambda \boldsymbol{I}_M$ is the covariance matrix of $\boldsymbol{y}$.

In the context of estimating the n-th entry of the signal vector $\boldsymbol{x}$, we adopt the Component-Wise Conditionally Unbiased (CWCU-)LMMSE approach [18]. This approach assumes that the n-th entry of $\boldsymbol{x}$ is deterministic, while treating the other entries as random variables.

Given that we consider the n-th entry of $\boldsymbol{x}$ to be deterministic (with a prior variance assumed to be $+\infty$), and treat the remaining entries as random variables, we can estimate the n-th entry of $\boldsymbol{x}$ and the associated error using the following equations:

$$\widehat{x}_n(0) = x_n + \widetilde{x}_n(0) \quad (13)$$

where $\widetilde{x}_n(0)$ has variance $\sigma^2_{\widetilde{x}_n(0)}$ and $\widehat{x}_n(0)$ represents the CWCU-LMMSE estimated value as below:

$$\hat{x}_n(0) = \frac{\boldsymbol{H}_n^H(\hat{\boldsymbol{\beta}}) \left(\sum_{j\neq n}^N \hat{\gamma}_j \boldsymbol{H}_j(\hat{\boldsymbol{\beta}}) \boldsymbol{H}_j^H(\hat{\boldsymbol{\beta}}) + \lambda \boldsymbol{I}_M\right)^{-1} \boldsymbol{y}}{\boldsymbol{H}_n^H(\hat{\boldsymbol{\beta}}) \left(\sum_{j\neq n}^N \hat{\gamma}_j \boldsymbol{H}_j(\hat{\boldsymbol{\beta}}) \boldsymbol{H}_j^H(\hat{\boldsymbol{\beta}}) + \lambda \boldsymbol{I}_M\right)^{-1} \boldsymbol{H}_n(\hat{\boldsymbol{\beta}})}; \tag{14a}$$

$$\sigma^2_{\widetilde{x}_n(0)} = \left[\boldsymbol{H}_n^H(\hat{\boldsymbol{\beta}}) \left(\sum_{j\neq n}^N \hat{\gamma}_j \boldsymbol{H}_j(\hat{\boldsymbol{\beta}}) \boldsymbol{H}_j^H(\hat{\boldsymbol{\beta}}) + \lambda \boldsymbol{I}_M\right)^{-1} \boldsymbol{H}_n(\hat{\boldsymbol{\beta}})\right]^{-1}. \tag{14b}$$

Incorporating the Gaussian prior information assumed in Sparse Bayesian Learning (SBL) into (13), the posterior mean $\widehat{x}_n$ can be expressed as

$$\widehat{x}_n = \widehat{x}_n(\gamma_n) = \frac{\gamma_n}{\gamma_n + \sigma^2_{\widetilde{x}_n(0)}} \widehat{x}_n(0). \tag{15}$$

Then we obtain the MSE:

$$\begin{aligned} &\mathrm{MSE}_{x_n} = \mathrm{E}\|\widehat{x}_n - x_n\|_2^2 = \mathrm{E}\left\{\|x_n\|_2^2 + \|\widehat{x}_n\|_2^2 - 2\Re\{\widehat{x}_n x_n^*\}\right\} \\ &= \mathrm{E}\Big\{\|x_n\|_2^2 - \|\widehat{x}_n(0)\|_2^2 + \|\widehat{x}_n(0) - \widehat{x}_n\| + 2\sigma^2_{\widetilde{x}_n(0)} \Re\{\tfrac{\partial \widehat{x}_n}{\partial \widehat{x}_n(0)}\}\Big\}. \end{aligned} \tag{16}$$

where E denotes expectation with respect to $\widetilde{x}_n(0)$ (assuming x_n is deterministic). By omitting the expectation, we obtain an instantaneous unbiased estimate of the MSE and the corresponding Stein's Unbiased Risk Estimate (SURE) function, which represents the part of $\widehat{\mathrm{MSE}}$ dependent on $\widehat{x}_n$:

$$\begin{aligned} &\widehat{\mathrm{MSE}}_{x_n} = \|x_n\|_2^2 - \|\widehat{x}_n(0)\|_2^2 + \mathrm{SURE}_{x_n}, \\ &\mathrm{SURE}_{x_n} = \|\widehat{x}_n(0) - \widehat{x}_n\| + 2\sigma^2_{\widetilde{x}_n(0)} \Re\{\tfrac{\partial \widehat{x}_n}{\partial \widehat{x}_n(0)}\}. \end{aligned} \tag{17}$$

In (17), after some algebraic manipulations, SURE_{x_n} can be expressed as

$$\mathrm{SURE}_{x_n}(\gamma_n) = \left(\frac{\sigma^2_{\widetilde{x}_n(0)}}{\gamma_n + \sigma^2_{\widetilde{x}_n(0)}}\right)^2 \|\widehat{x}_n(0)\|_2^2 + 2\frac{\sigma^2_{\widetilde{x}_n(0)}\,\gamma_n}{\gamma_n + \sigma^2_{\widetilde{x}_n(0)}}, \tag{18}$$

where, as a function of γ_n, the first term is decreasing and the second term is increasing. Thus, we obtain

$$\frac{\partial \mathrm{SURE}_{x_n}}{\partial \gamma_n} = 2\sigma^4_{\widetilde{x}_n(0)}(\gamma_n + \sigma^2_{\widetilde{x}_n(0)} - \|\widehat{x}_n(0)\|_2^2)/(\gamma_n + \sigma^2_{\widetilde{x}_n(0)})^3. \tag{19}$$

$\mathrm{SURE}_{x_n}(\gamma_n)$ has a single extremum, a local minimum, at $\gamma_n = \widehat{x}_n^2(0) - \sigma^2_{\widetilde{x}_n(0)}$. So, the minimum of $\mathrm{SURE}_{x_n}(\gamma_n)$ occurs at positive γ_n when $\widehat{x}_n^2(0) > \sigma^2_{\widetilde{x}_n(0)}$, but at negative γ_n in the opposite case. Therefore, to ensure $\gamma_n \geq 0$, we find the optimum as

$$\hat{\gamma}_n = \max\left(\|\widehat{x}_n(0)\|_2^2 - \sigma^2_{\widetilde{x}_n(0)}, 0\right). \tag{20}$$

Therefore, to estimate the entire vector $\hat{\boldsymbol{\gamma}}$, we initialize $\boldsymbol{\gamma}$ first, then update $\hat{\gamma}_n$ step by step until convergence.

C. Expectation Propagation for Estimating $\boldsymbol{\beta}$

For estimating $\boldsymbol{\beta}$, by applying Bayes' formula, we have:

$$p(\boldsymbol{\beta}|\boldsymbol{y}, \hat{\boldsymbol{\gamma}}) = \frac{p(\boldsymbol{y}|\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) p(\boldsymbol{\beta})}{\int p(\boldsymbol{y}|\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) p(\boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta}}, \tag{21}$$

where

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) = \mathcal{CN}(\boldsymbol{y}|\boldsymbol{0}, \boldsymbol{R}(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}})) \tag{22}$$

and covariance matrix $\boldsymbol{R}(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}})$ can be expressed as

$$\boldsymbol{R}(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) = \boldsymbol{H}(\boldsymbol{\beta})\ \mathrm{diag}(\hat{\boldsymbol{\gamma}})\ \boldsymbol{H}(\boldsymbol{\beta})^H + \lambda \boldsymbol{I}_M. \tag{23}$$

Since $p(\boldsymbol{\beta}|\boldsymbol{y}, \hat{\boldsymbol{\gamma}})$ in (21) involves a high-dimensional integral, it is intractable to compute directly. Therefore, an approximation method is necessary to address this challenge. Here, we propose an expectation-propagation (EP)-like algorithm for solving this problem. In this approach,

$$p(\boldsymbol{\beta}|\boldsymbol{y}, \hat{\boldsymbol{\gamma}}) \approx q(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \boldsymbol{y}, \hat{\boldsymbol{\gamma}}) = \frac{p(\boldsymbol{y}|\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) \prod_{n=1}^N f(\beta_n|\hat{\beta}_n)}{\int p(\boldsymbol{y}|\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) \prod_{n=1}^N f(\beta_n|\hat{\beta}_n) \mathrm{d}\boldsymbol{\beta}}, \tag{24}$$

where $f(\beta_n|\hat{\beta}_n) = \delta(\beta_n - \hat{\beta}_n)$ and $\delta(\cdot)$ is delta function which can also be treated as a Gaussian distribution with zero variance. It is obvious that updating $\hat{\boldsymbol{\beta}}$ is quite important. Employing the principles of EP, we outline the algorithm as follows:

- Initialize $\hat{\boldsymbol{\beta}}$
- Until all $\hat{\boldsymbol{\beta}}$ converged
  - Choose a $f(\beta_n|\hat{\beta}_n)$ to refine
  - Remove $f(\beta_n|\hat{\beta}_n)$ from the posterior and integral out $\boldsymbol{\beta}$ except β_n to get a pdf $\psi(\boldsymbol{y}|\beta_n, \hat{\boldsymbol{\beta}}_{\bar{n}}, \hat{\boldsymbol{\gamma}})$ of extrinsic β_n:

$$\psi(\boldsymbol{y}|\beta_n, \hat{\boldsymbol{\beta}}_{\bar{n}}, \hat{\boldsymbol{\gamma}}) = \int p(\boldsymbol{y}|\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) \prod_{i\neq n}^N f(\beta_i|\hat{\beta}_i) \mathrm{d}\boldsymbol{\beta}_{\bar{n}}. \tag{25}$$

  - Combine $\psi(\beta_n|\hat{\boldsymbol{\beta}}_{\bar{n}}, \boldsymbol{y}, \hat{\boldsymbol{\gamma}})$ with real prior $p(\beta_n)$ to update $\hat{\beta}_n$:

$$\hat{\beta}_n = \frac{\int \beta_n \psi(\boldsymbol{y}|\beta_n, \hat{\boldsymbol{\beta}}_{\bar{n}}, \hat{\boldsymbol{\gamma}}) p(\beta_n) \mathrm{d}\beta_n}{\int \psi(\boldsymbol{y}|\beta_n, \hat{\boldsymbol{\beta}}_{\bar{n}}, \hat{\boldsymbol{\gamma}}) p(\beta_n) \mathrm{d}\beta_n}. \tag{26}$$

In (25), $\psi(\boldsymbol{y}|\beta_n, \hat{\boldsymbol{\beta}}_{\bar{n}}, \hat{\boldsymbol{\gamma}})$ can be represented as:

$$\psi(\boldsymbol{y}|\beta_n, \hat{\boldsymbol{\beta}}_{\bar{n}}, \hat{\boldsymbol{\gamma}}) = \frac{1}{\pi^M \det(\boldsymbol{\Xi}_{\bar{n}})} \mathrm{e}^{-\boldsymbol{y}^H \boldsymbol{\Xi}_{\bar{n}}^{-1} \boldsymbol{y}}, \tag{27}$$

where

$$\begin{aligned} \boldsymbol{\Xi}_{\bar{n}} = &(\boldsymbol{H}(\hat{\boldsymbol{\beta}}_{\bar{n}}) + \beta_n \boldsymbol{B}_n + \tfrac{\beta_n^2}{2} \boldsymbol{C}_n) \mathrm{diag}(\hat{\boldsymbol{\gamma}}) \\ &(\boldsymbol{H}(\hat{\boldsymbol{\beta}}_{\bar{n}}) + \beta_n \boldsymbol{B}_n + \tfrac{\beta_n^2}{2} \boldsymbol{C}_n)^H + \lambda \boldsymbol{I}_M. \end{aligned} \tag{28}$$

After some straightforward algebraic manipulations, the calculation for (26) yields:

$$\hat{\beta}_n = \frac{\int_{-\frac{l}{2}}^{\frac{l}{2}} \frac{\beta_n}{\det(\boldsymbol{\Xi}_{\bar{n}})} \mathrm{e}^{-\boldsymbol{y}^H \boldsymbol{\Xi}_{\bar{n}}^{-1} \boldsymbol{y}} \mathrm{d}\beta_n}{\int_{-\frac{l}{2}}^{\frac{l}{2}} \frac{1}{\det(\boldsymbol{\Xi}_{\bar{n}})} \mathrm{e}^{-\boldsymbol{y}^H \boldsymbol{\Xi}_{\bar{n}}^{-1} \boldsymbol{y}} \mathrm{d}\beta_n}. \tag{29}$$

Even though (29) lacks an analytic form, it can be computed numerically as a one dimension integration using appropriate tools.

D. Noise variance λ Estimation

Once $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ have been estimated, we construct the matrix $\boldsymbol{H}(\hat{\boldsymbol{\beta}})$ by selecting its non-zero columns to form a new sensing matrix $\hat{\boldsymbol{\Phi}}$. We then estimate $\boldsymbol{s}$ using the least squares estimator (LSE) as follows:

$$\hat{\boldsymbol{x}} = (\hat{\boldsymbol{\Phi}}^T \hat{\boldsymbol{\Phi}})^{-1} \hat{\boldsymbol{\Phi}}^T \boldsymbol{y}. \tag{30}$$

Then, the hyperparameter noise variance λ can be estimated as:

$$\hat{\lambda} = \frac{\|\boldsymbol{y} - \hat{\boldsymbol{\Phi}} \hat{\boldsymbol{x}}\|_2^2}{M - \mathrm{rank}(\hat{\boldsymbol{\Phi}})}. \tag{31}$$

As previously mentioned, we lack prior information about $\boldsymbol{x}$ (deterministic) except for its sparsity. Therefore estimating λ with respect to $\boldsymbol{\gamma}$ may introduce unknown biases, necessitating further analytical investigation. This consideration underscores the need for additional rigorous analysis to understand the implications and potential biases introduced by estimating λ in relation to $\boldsymbol{\gamma}$. Such exploration is crucial for ensuring the robustness and reliability of the inference process in the context of sparse Bayesian learning.

E. SURE-SBL-EP DoA Estimation Algorithm

The whole algorithm are given in Alg. 1. The initialization of hyperparameters in the numerical simulations of this paper are denoted in IV-A.

IV. Numerical Simulation

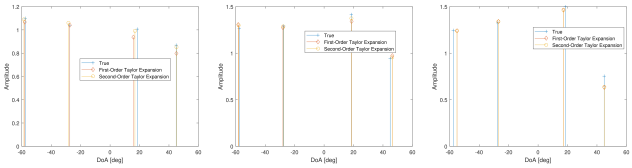
In this section, the performance of EP-SURE-SBL is evaluated by performing numerical simulations.

Algorithm 1 DoA Estimation based on SURE-SBL-EP

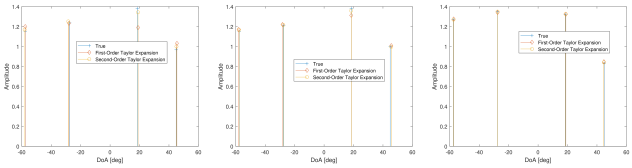
Input: $\boldsymbol{y}$, N, K, ϵ_1, ϵ_2 and ϵ_3

Initialization:$\boldsymbol{\gamma}^1 \leftarrow \mathbf{1}_N, \boldsymbol{\beta}^1 \leftarrow \mathbf{0}_N, \tau \leftarrow 1$, $\tilde{\boldsymbol{\theta}}$, $\hat{\lambda}^1 = 0.1$
Calculate $\boldsymbol{A}$ $\boldsymbol{B}$ and $\boldsymbol{C}$ based on $\tilde{\boldsymbol{\theta}}$
repeat
  repeat
    $\tau \leftarrow \tau + 1$
    repeat
      Update $\hat{x}_n(0)$ and $\sigma^2_{\tilde{x}_n(0)}$ based on (14)
      Update γ_n^τ based on (20)
    until All $\{\gamma_n^\tau\}_{n=1}^N$ updated
  until $\|\boldsymbol{\gamma}^\tau - \boldsymbol{\gamma}^{\tau-1}\|^2 / \|\boldsymbol{\gamma}^\tau\|^2 < \epsilon_1$
  Keep K large γ^τ and set others to be 0
  repeat
    repeat
      if $\gamma_n^\tau \neq 0$, **then** Update $\hat{\beta}_n^\tau$ based on (29)
      else $\hat{\beta}_n^\tau = 0$ **end if**
    until All $\{\gamma_n^\tau\}_{n=1}^N$ updated
  until $\|\boldsymbol{\beta}^\tau - \boldsymbol{\beta}^{\tau-1}\|^2 / \|\boldsymbol{\beta}^\tau\|^2 < \epsilon_2$
  Update $\hat{\lambda}^\tau$ based on (31)
until $\|\hat{\lambda}^\tau - \hat{\lambda}^{\tau-1}\| / \|\hat{\lambda}^\tau\|^2 < \epsilon_3$

Output: Estimate DoAs $\{\hat{\theta}_k\}_{k=1}^K$ based on $\boldsymbol{\beta}^\tau$ and $\tilde{\boldsymbol{\theta}}$



(a) 30 grids and SNR = 10dB (b) 60 grids and SNR = 10dB (c) 90 grids and SNR = 10dB



(d) 30 grids and SNR = 20dB (e) 60 grids and SNR = 20dB (f) 90 grids and SNR = 20dB

Fig. 1: Signal reconstruction performance

A. Simulation Setup

The performance of EP-SURE-SBL is evaluated through numerical simulations. A simple simulation demonstrates the recovery capability of EP-SURE-SBL with parameters set as follows: $M = 10$, $K = 4$, wavelength $\lambda = 10^{-2}$ meter, distance $d = 5 \times 10^{-3}$ meter and true Directions-of-Arrival (DoAs) given by $\boldsymbol{\theta} = [-57.6, -27.4, 18.5, 44.9]^T$ (units: degrees). The complex coefficients $\{s_k\}_{k=1}^K$ are generated with magnitudes drawn from $\mathcal{U}(1,2)$ and phases from $\mathcal{U}(-\pi, \pi)$, not adhering to a complex Gaussian distribution. Parameters ϵ_1, ϵ_2, and ϵ_3 are set to 0.001.

B. Performance versus SNR and grid numbers

The performance of EP-SURE-SBL is analyzed across various Signal-to-Noise Ratios (SNRs). Results are presented in Figs. 1a-1f. The reconstruction of signals and the estimation error of DoAs exhibit higher variability when using second-order Taylor expansion with 30 and 60 grid points. Conversely, with 90 grid points, the performance of first and second-order Taylor expansions becomes comparable. The Mean Squared Error (MSE) of DoA is defined as:

$$\text{MSE} = \frac{1}{K} \|\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\|_2^2. \tag{32}$$

Fig. 2 illustrates 100 trial runs with the conditional Cramér-Rao Bound (CRB) serving as a lower bound. It is evident that with 90 grid points, performance consistently outperforms that of 30 and 60 grid points. It is clear that higher-order Taylor expansions yield improved performance with identical grid points theoretically. Notably, with the same grid points, Fig. 2 also shows second-order Taylor expansion achieves superior performance compared to first-order expansion. In addition, we compared Determinisic Maximum Likelihood (DML) based DoA estimation [19], and our algorithm achieves better performance when the grid is set to 60 or more.

The computational complexity of EP-SURE-SBL is dominated by the matrix inversion, which scales as $\mathcal{O}(MN^2)$, where M is the length of the received data and N is the grid size. The total complexity is $\mathcal{O}(TMN^2)$, with T being the sum of iteration number of EP and SBL. Another factor that affects the computational speed is the use of numerical integration, which could potentially be improved. Empirical results show the following averaged single-operation times:

- First-order Taylor Expansion: 30 grids: 0.1830s, 60 grids: 0.6021s, 90 grids: 1.2532s.
- Second-order Taylor Expansion: 30 grids: 0.2321s, 60 grids: 0.6951s, 90 grids: 1.3834s.

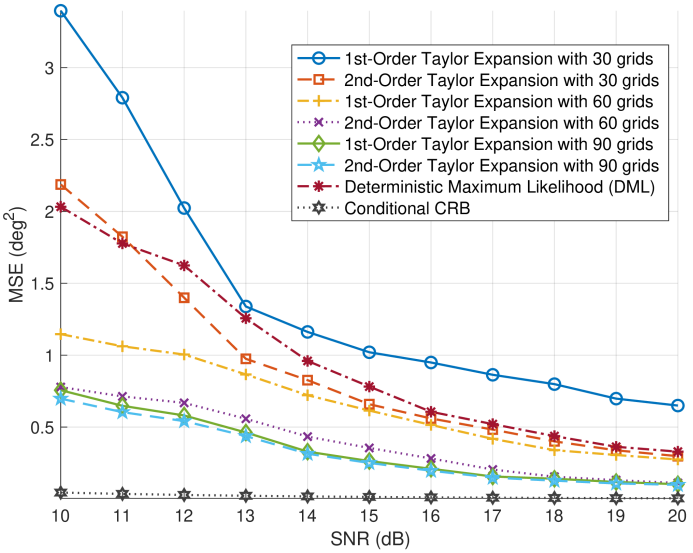


Fig. 2: MSE varying w.r.t. SNR

V. Conclusion

In this paper, we propose the EP-SURE-SBL algorithm for single snapshot Direction-of-Arrival (DoA) estimation. We categorize DoA estimation using both first order and second order Taylor expansion into grid-on estimation and off-grid error estimation. Grid-on estimation employs sparse Bayesian learning for sparse modeling, while the variance hyperparameter in sparse Bayesian learning is estimated using Stein's Unbiased Risk Estimate (SURE) instead of Expectation-Maximization (EM). For off-grid error, we approximate the Minimum Mean Square Error (MMSE) bias using Expectation Propagation (EP) for estimation. We validate the algorithm's feasibility through simulations. Future research will focus on improving recovery in scenarios with an unknown number of signal sources and conducting comparative experiments with other state-of-the-art methods.

Acknowledgements: EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, Principaute de Monaco and by the projects CellFree6G (French ANR), and 5G-OPERA (Franco-German 5G ecosystem project).

REFERENCES

[1] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit ," *IEEE Trans. Inf. Theory*, December 2007.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit ," *SIAM J. Sci. Comput.*, vol. 20, no. 1, 1998.

[3] H. Zou, "The Adaptive Lasso and Its Oracle Properties," *J. Am. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, 2006.

[4] D. Wipf and S. Nagarajan, "Iterative reweighted l_1 and l_2 methods for finding sparse solutions ," *IEEE J. Sel. Topics Sig. Process.*, vol. 4, no. 2, April 2010.

[5] R. Giri and Bhaskar D. Rao, "Type I and type II bayesian methods for sparse signal recovery using scale mixtures," *IEEE Trans. on Sig Process.*, vol. 64, no. 13, 2018.

[6] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *J. Mach. Learn. Res.*, vol. 1, 2001.

[7] D. P. Wipf and B. D. Rao, "Sparse Bayesian Learning for Basis Selection ," *IEEE Trans. on Sig. Proc.*, vol. 52, no. 8, Aug. 2004.

[8] D.P. Wipf, *Bayesian Methods for Finding Sparse Representations*, Ph.d. thesis, University of California, San Diego, 2006.

[9] J. A. Fessler and A. Hero, "Space-alternating generalized expectation-maximization algorithm.," *IEEE Trans. on Sig. Process.*, Oct. 1994.

[10] W. James and C. M. Stein, "Estimation with quadratic loss," *Proc. of Four. Berk. Sympo. on Mathe. Stat. Prob., Berk.: Univ. of Calif. Press,*, 1961.

[11] D. Slock, "Sparse Bayesian learning with Stein's unbiased risk estimator based hyperparameter optimization," in *IEEE Asilomar*, 2022.

[12] F. Xiao and D. Slock, "Towards Hyperparameter Optimization of Sparse Bayesian Learning Based on Stein's Unbiased Risk Estimator," in *First 'Learn to Compress' Workshop ISIT 2024*.

[13] R.O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. Antennas Propagation*, vol. 34, no. 3, pp. 276–280, March 1986.

[14] R. Roy and T. Kailath, "ESPRIT-estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul 1989.

[15] X. Xu and Q. Huang, "MD-DOA: A Model-based Deep Learning DOA Estimation Architecture," *IEEE Sensors Journal*, 2024.

[16] Y. Wu, M. B. Wakin, and P. Gerstoft, "Gridless DOA estimation with multiple frequencies," *IEEE Trans. Signal Process.*, pp. 417–432, 2023.

[17] B. Çakmak and M. Opper, "Expectation Propagation for Approximate Inference: Free Probability Framework," *IEEE Inter. Sympo. On Info. Theo. (ISIT)*, 2018.

[18] M. Triki and D.T.M. Slock, "Component-Wise Conditionally Unbiased Bayesian Parameter Estimation: General Concept and Applications to Kalman Filtering and LMMSE Channel Estimation," in *Conference Record of the Thirty-Ninth Asilomar Conference onSignals, Systems and Computers, 2005.*, 2005, pp. 670–674.

[19] J. Selva, "Efficient computation of ML DOA estimates under unknown nonuniform sensor noise powers," *Signal Processing*, vol. 205, pp. 108879, 2023.