SPARSE BAYESIAN LEARNING WITH UNCERTAIN SENSING MATRIX

Santosh Nannuru, Peter Gerstoft, Kay L. Gemba

Scripps Institution of Oceanography, University of California, San Diego

ABSTRACT

Sparse Bayesian learning is a sparse processing method used for solving high-dimensional, underdetermined linear equations. Often the sensing matrix in the system of equations is assumed known and in presence of perturbations in this matrix performance of sparse processing degrades. We develop a sparse Bayesian learning method that accounts for perturbations in the sensing matrix. We derive an iterative weight update by performing evidence maximization. Beamforming simulations are used to demonstrate the advantages of the proposed method.

Index Terms— Sparse Bayesian learning, compressed sensing, sensing matrix uncertainty, beamforming

1. INTRODUCTION

Many interesting research problems can be formulated as a system of high-dimensional, underdetermined, linear equations whose solutions are sparse. Recent advances in statistics have produced various methods to solve for sparse solutions in a feasible way. Some of these methods are basis pursuit [1, 2], matching pursuit, and sparse Bayesian learning (SBL) [3]. Among these, matching pursuit is a fast but greedy method. Basis pursuit involves numerically solving a convex optimization problem and can be computationally expensive.

Recent literature [3, 4] has focussed on SBL and its applications [5–8]. SBL is computationally much faster than basis pursuit and has good performance among various sparse processing methods [9]. SBL takes a probabilistic approach to solve the system of linear equations and obtains a sparse solution which maximizes the evidence.

Most of the existing literature on sparse processing assumes that the sensing matrix is deterministic and completely known. This is not feasible in many practical applications, some of which include beamforming [10, 11] and matched-field processing [12, 13]. Uncertain sensing matrices have been studied in the context of basis pursuit [14, 15], matching pursuit [16] and message passing [17]. A recent work discussed robust SBL [18] to account for outliers in the signal.

We derive iterative update equation for a SBL method which accounts for uncertainty in the sensing matrix. The statistics of the sensing matrix perturbations are assumed known a-priori. The paper is organized as follows: Section 2 discusses the signal model and the approximate likelihood model. The proposed SBL method is derived in Section 3. Simulations using beamforming example are discussed in Section 4 and conclusions are provided in Section 5.

2. SIGNAL MODEL

In this section we discuss the signal model used in SBL and the assumptions made in this paper. Let $\mathbf{y} \in \mathbb{C}^N$ be the complex signal which is expressed as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where noise $\mathbf{n} \in \mathbb{C}^N$ is zero mean circularly symmetric complex Gaussian with covariance matrix $\sigma^2\mathbf{I}_N$; $\mathbf{I}_N$ is an $N \times N$ identity matrix; $\mathbf{A} \in \mathbb{C}^{N\times M}$ is the sensing matrix; $\mathbf{x} \in \mathbb{C}^M$ is the weight vector. In sparse problem formulation, $\mathbf{x}$ is sparse with at most K non-zero entries where $K \ll M$. In most of the literature it is assumed that the sensing matrix $\mathbf{A}$ is known and has maximal column rank N. In this work we allow $\mathbf{A}$ to be random with known statistics.

2.1. Prior

SBL models $\mathbf{x}$ as a zero mean circularly symmetric Gaussian with prior density $p(\mathbf{x}) = \mathcal{CN}(\mathbf{x}; \mathbf{0}, \mathbf{\Gamma})$, where the unknown covariance matrix $\mathbf{\Gamma}$ is assumed to be diagonal $\mathbf{\Gamma} = \text{diag}(\boldsymbol{\gamma})$ where $\boldsymbol{\gamma} = [\gamma_1 \dots \gamma_M]$. We use $\mathbf{0}$ to denote a vector (or matrix) of all zeroes with appropriate dimensions. Noise $\mathbf{n}$ and weights $\mathbf{x}$ are assumed to be independent. The variance γ_m is allowed to be zero and in fact at convergence many $\gamma_m \to 0$. The limiting Gaussian distribution (density does not exist since the covariance Gamma is not invertible) thus obtained is sparse. The notation $\mathcal{CN}(\mathbf{z}; \boldsymbol{\mu}, \mathbf{\Sigma})$ denotes a circularly symmetric complex Gaussian density function with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$,

$$\mathcal{CN}(\mathbf{z}; \boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{\pi^k \det(\mathbf{\Sigma})} \exp\big(- (\mathbf{z} - \boldsymbol{\mu})^H \mathbf{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})\big),$$

where $\mathbf{\Sigma}$ is a Hermitian and non-negative definite matrix.

2.2. Likelihood

Very often the sensing matrix $\mathbf{A}$ is assumed deterministic, denote this by $\mathbf{A} = \mathbf{A}^o$. Since the noise is assumed Gaussian, given the weight vector $\mathbf{x}$ and the sensing matrix $\mathbf{A}^o$, the likelihood function is written as

$$p(\mathbf{y}|\mathbf{x}) = p(\mathbf{y}|\mathbf{x};\mathbf{A}^o) = \mathcal{CN}(\mathbf{y};\mathbf{A}^o\mathbf{x},\sigma^2\mathbf{I}_N). \tag{2}$$

Sensing matrix perturbation: The assumption $\mathbf{A} = \mathbf{A}^o$ does not always hold, especially when there is uncertainty in the model or parameters used to construct the matrix $\mathbf{A}$. For example, in plane wave beamforming entries of the matrix $\mathbf{A}$ depend on array positions and the sound speed. These parameters may not be accurately known or can change over time. To account for perturbations we model the matrix $\mathbf{A}$ as a random matrix. Express the matrix $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{A}^o + \mathbf{A}^e, \tag{3}$$

where $\mathbf{A}^o$ is a known matrix and $\mathbf{A}^e$ is a random matrix which accounts for the perturbation in $\mathbf{A}$. The perturbation model in (3) has been studied in [12, 14–16, 19]. In the context of beamforming, sensing matrix perturbations have been studied in [20, 21]. We now derive a likelihood function by approximately averaging out the perturbations in model (3).

Approximate likelihood: For computational tractability we impose certain restrictions on $\mathbf{A}^e$. Denote the mth column of matrix $\mathbf{A}^e$ by $\mathbf{a}^e_m$. We assume that the random vectors $\mathbf{a}^e_m$ have zero mean, zero relation matrix, and known covariance matrix $\mathbf{\Sigma}^e_m$ with columns of $\mathbf{A}^e$ such that

$$\mathsf{E}(\mathbf{a}^e_m) = \mathbf{0}\,; \quad \mathsf{E}(\mathbf{a}^e_m\mathbf{a}^{eT}_n) = \mathbf{0} \tag{4}$$

$$\mathsf{E}(\mathbf{a}^e_m\mathbf{a}^{eH}_n) = \delta(m-n)\mathbf{\Sigma}^e_m \tag{5}$$

where $\delta(m)$ is the Kronecker delta function. Since the perturbation $\mathbf{A}^e \in \mathbb{C}^{N\times M}$, full specification of its second order moments require $\mathcal{O}(NM\times NM)$ terms. The assumptions (4) and (5) reduce the number of free parameters and keeps the problem tractable. The perturbation $\mathbf{A}^e$ and the weights $\mathbf{x}$ are assumed independent.

The signal model (1) when combined with the matrix perturbation model (3) can be written as

$$\mathbf{y} = (\mathbf{A}^o + \mathbf{A}^e)\mathbf{x} + \mathbf{n} = \mathbf{A}^o\mathbf{x} + \boldsymbol{\zeta}, \tag{6}$$

where we have defined $\boldsymbol{\zeta} = \mathbf{A}^e\mathbf{x}+\mathbf{n}$. Here $\boldsymbol{\zeta}$ can be considered as the modified noise term (which also depends on the weights $\mathbf{x}$). To simplify the likelihood model, we compute the mean and covariance of $\boldsymbol{\zeta}$:

$$\mathsf{E}(\boldsymbol{\zeta}) = \mathsf{E}(\mathbf{A}^e\mathbf{x} + \mathbf{n}) = \mathbf{0} \tag{7}$$

$$\mathsf{E}(\boldsymbol{\zeta}\boldsymbol{\zeta}^H) = \mathsf{E}(\mathbf{A}^e\mathbf{x}\mathbf{x}^H\mathbf{A}^{eH}) + \mathsf{E}(\mathbf{n}\mathbf{n}^H) \tag{8}$$

$$= \mathsf{E}\Big(\sum_{m,n} x_m x^H_n \mathbf{a}^e_m\mathbf{a}^{eH}_n\Big) + \sigma^2\mathbf{I}_N \tag{9}$$

$$= \sum_{m,n} \mathsf{E}(x_m x^H_n)\mathsf{E}(\mathbf{a}^e_m\mathbf{a}^{eH}_n) + \sigma^2\mathbf{I}_N \tag{10}$$

$$= \sum_m \gamma_m\mathbf{\Sigma}^e_m + \sigma^2\mathbf{I}_N = \mathbf{\Sigma}_{\boldsymbol{\zeta}}. \tag{11}$$

In above simplification we have used independence of $\mathbf{x}$ and $\mathbf{A}^e$. The covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\zeta}}$ depends on 1) the variance of the signal i.e. γ_m; 2) the covariance of the error in columns of sensing matrix i.e. $\mathbf{\Sigma}^e_m$; and 3) the covariance of the additive white Gaussian noise i.e. $\sigma^2\mathbf{I}_N$. For analytical simplification we approximate the density of $\boldsymbol{\zeta}$ to be Gaussian

$$p(\boldsymbol{\zeta}) \approx \mathcal{CN}(\boldsymbol{\zeta};\mathbf{0},\mathbf{\Sigma}_{\boldsymbol{\zeta}}). \tag{12}$$

To justify this we write $\boldsymbol{\zeta} = \sum_m x_m\mathbf{a}^e_m + \mathbf{n}$. Now since $\mathbf{x}$ is a high dimensional vector, $\boldsymbol{\zeta}$ is a sum of large number of random vectors. Hence applying the central limit theorem, probability distribution of $\boldsymbol{\zeta}$ converges to a Gaussian distribution as $M\to\infty$. When $\mathbf{x}$ is K-sparse, the accuracy of the approximation increases with K. The likelihood for the signal model (6) is now approximately expressed as

$$p(\mathbf{y}|\mathbf{x}) \approx \mathcal{CN}(\mathbf{y};\mathbf{A}^o\mathbf{x},\mathbf{\Sigma}_{\boldsymbol{\zeta}}), \tag{13}$$

$$\mathbf{\Sigma}_{\boldsymbol{\zeta}} = \sigma^2\mathbf{I}_N + \sum_m \gamma_m\mathbf{\Sigma}^e_m \tag{14}$$

where $\boldsymbol{\zeta}$ and $\mathbf{x}$ are assumed independent for analytical tractability of the evidence term in Section 3.

Multiple snapshots: To increase the signal-to-noise ratio (SNR) we usually process multiple observations (snapshots) simultaneously. Let $\mathbf{Y} \in \mathbb{C}^{N\times L}$ denote the collection of L consecutive snapshots arranged column wise in a matrix. For the multi snapshot case (1) becomes

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N}, \tag{15}$$

where $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_L]$ and $\mathbf{N} = [\mathbf{n}_1 \dots \mathbf{n}_L]$. The weights $\mathbf{x}_l$ are assumed to be independent identically distributed with Gaussian density. Similarly, noise is assumed to be independent across snapshots. Thus

$$p(\mathbf{X}) = \prod_{l=1}^{L} p(\mathbf{x}_l) = \prod_{l=1}^{L} \mathcal{CN}(\mathbf{x}_l;\mathbf{0},\mathbf{\Gamma}), \tag{16}$$

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{l=1}^{L} p(\mathbf{y}_l|\mathbf{x}_l), \tag{17}$$

where the single snapshot likelihood $p(\mathbf{y}_l|\mathbf{x}_l)$ is given by (13).

3. SPARSE BAYESIAN LEARNING

In the SBL framework [3,4], the prior parameter $\mathbf{\Gamma}$ is assumed to be unknown and estimated using the observed signal $\mathbf{Y}$. It is estimated by maximizing the evidence (also called type-II maximum likelihood). The evidence $p(\mathbf{Y})$ is obtained by averaging over all realizations of $\mathbf{X}$

$$p(\mathbf{Y};\boldsymbol{\gamma}) = \int p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})d\mathbf{X} \tag{18}$$

$$= \int \prod_{l=1}^{L} \mathcal{CN}(\mathbf{y}_l;\mathbf{A}^o\mathbf{x}_l,\mathbf{\Sigma}_{\boldsymbol{\zeta}})\,\mathcal{CN}(\mathbf{x}_l;\mathbf{0},\mathbf{\Gamma})d\mathbf{X} \tag{19}$$

$$= \prod_{l=1}^{L} \mathcal{CN}(\mathbf{y}_l;\mathbf{0},\mathbf{\Sigma}_{\boldsymbol{\zeta}} + \mathbf{A}^o\mathbf{\Gamma}\mathbf{A}^{oH}) = \prod_{l=1}^{L} \mathcal{CN}(\mathbf{y}_l;\mathbf{0},\mathbf{\Sigma}_{\mathbf{y}}),$$

where we have defined $\boldsymbol{\Sigma}_{\mathbf{y}} = \boldsymbol{\Sigma}_{\boldsymbol{\zeta}} + \mathbf{A}^o \boldsymbol{\Gamma} \mathbf{A}^{oH}$ for brevity. Note that $\boldsymbol{\Sigma}_{\mathbf{y}}$ contains both the parameters σ^2 and $\boldsymbol{\Gamma}$. Ignoring the constant terms independent of σ^2 and $\boldsymbol{\Gamma}$, the logarithm of the evidence can be expressed as

$$\log p(\mathbf{Y};\boldsymbol{\gamma}) = \sum_{l=1}^{L} -\log\left((\pi)^N |\boldsymbol{\Sigma}_{\mathbf{y}}|\right) - \sum_{l=1}^{L} \mathbf{y}_l^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}_l \quad (20)$$
$$\propto -L \log|\boldsymbol{\Sigma}_{\mathbf{y}}| - \mathrm{Tr}(\mathbf{Y}^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{Y}), \quad (21)$$

where Tr() denotes the trace of a matrix.

3.1. Fixed point update

The estimate $\hat{\boldsymbol{\Gamma}}$ is the argument which maximizes the evidence

$$\hat{\boldsymbol{\Gamma}} = \arg\max_{\boldsymbol{\Gamma}} \log p(\mathbf{Y};\boldsymbol{\gamma}) \quad (22)$$
$$= \arg\min_{\boldsymbol{\Gamma}} \left\{ L \log|\boldsymbol{\Sigma}_{\mathbf{y}}| + \mathrm{Tr}(\mathbf{Y}^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{Y}) \right\}. \quad (23)$$

One approach to solve this problem is using the EM algorithm [22] but the resulting update equations have slow convergence properties [3, 4]. We perform differentiation of the objective function in (23) to obtain a local minima. We have the following derivative relations for the matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$

$$\frac{\partial \log|\boldsymbol{\Sigma}_{\mathbf{y}}|}{\partial \gamma_m} = \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathbf{y}}}{\partial \gamma_m}\right), \quad (24)$$

$$\frac{\partial \boldsymbol{\Sigma}_{\mathbf{y}}^{-1}}{\partial \gamma_m} = -\boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\mathbf{y}}}{\partial \gamma_m} \boldsymbol{\Sigma}_{\mathbf{y}}^{-1}, \quad \frac{\partial \boldsymbol{\Sigma}_{\mathbf{y}}}{\partial \gamma_m} = \boldsymbol{\Sigma}_m^e + \mathbf{a}_m^o \mathbf{a}_m^{oH}. \quad (25)$$

Differentiating (23) with respect to the mth diagonal element γ_m of the matrix $\boldsymbol{\Gamma}$ we get

$$\frac{\partial}{\partial \gamma_m} \left\{ L \log|\boldsymbol{\Sigma}_{\mathbf{y}}| + \mathrm{Tr}(\mathbf{Y}^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{Y}) \right\}$$
$$= L\, \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}}^{-1} [\boldsymbol{\Sigma}_m^e + \mathbf{a}_m^o \mathbf{a}_m^{oH}]\right) -$$
$$\mathrm{Tr}\left(\mathbf{Y}^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} [\boldsymbol{\Sigma}_m^e + \mathbf{a}_m^o \mathbf{a}_m^{oH}] \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{Y}\right). \quad (26)$$

Equating the derivative of the objective function to zero

$$1 = \frac{1}{L} \frac{\mathrm{Tr}\left(\mathbf{Y}^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} [\boldsymbol{\Sigma}_m^e + \mathbf{a}_m^o \mathbf{a}_m^{oH}] \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{Y}\right)}{\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}}^{-1} [\boldsymbol{\Sigma}_m^e + \mathbf{a}_m^o \mathbf{a}_m^{oH}]\right)} \quad (27)$$

$$\frac{\gamma_m}{\gamma_m} = \left(\frac{1}{L} \frac{\mathrm{Tr}\left(\mathbf{Y}^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} [\boldsymbol{\Sigma}_m^e + \mathbf{a}_m^o \mathbf{a}_m^{oH}] \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{Y}\right)}{\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}}^{-1} [\boldsymbol{\Sigma}_m^e + \mathbf{a}_m^o \mathbf{a}_m^{oH}]\right)} \right)^b \quad (28)$$

where we introduced γ_m terms to obtain an iterative update equation. The update equation can then be formulated as

$$\gamma_m^{\text{new}} \leftarrow \gamma_m^{\text{old}} \left(\frac{\mathrm{Tr}\left(\mathbf{S}_{\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} [\boldsymbol{\Sigma}_m^e + \mathbf{a}_m^o \mathbf{a}_m^{oH}] \boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\right)}{\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}}^{-1} [\boldsymbol{\Sigma}_m^e + \mathbf{a}_m^o \mathbf{a}_m^{oH}]\right)} \right)^b. \quad (29)$$

where $\mathbf{S}_{\mathbf{y}}$ is the sample covariance matrix $\mathbf{S}_{\mathbf{y}} = \frac{1}{L}\mathbf{Y}\mathbf{Y}^H$. In the above update equation, γ_m^{old} appears explicitly as well as implicitly in the expression for $\boldsymbol{\Sigma}_{\mathbf{y}}$.

Remark: There are multiple ways to formulate a fixed point update equation. Our formulation is inspired by some of the equations used in the literature [3, 4, 8] and convergence properties of the simulation results. It is not clear for what values of b, if any, convergence of the update (29) can be guaranteed. When we set $\boldsymbol{\Sigma}_m^e$ to zero, value of $b = 1$ gives the update equation used in [3,4] and $b = 0.5$ gives the update equation used in [8].

Posterior mean: Applying Bayes rule, the posterior distribution $p(\mathbf{X}|\mathbf{Y})$ is Gaussian with mean given by [8]

$$\mathsf{E}_{p(\mathbf{X}|\mathbf{Y})}(\mathbf{X}) = \boldsymbol{\Gamma} \mathbf{A}^{oH} \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{Y}. \quad (30)$$

Relation to other methods: In [12] the perturbation vectors $\mathbf{a}_m^e$ are assumed stochastic and an elastic net regression problem is formulated by averaging out the perturbations. Parametric modeling of the perturbations $\mathbf{a}_m^e$ is considered in [21] for plane wave beamforming. The parameters are estimated within the iterative framework of SBL but only specific perturbations are considered and cannot be generalized to include broader class of errors.

3.2. Noise estimate

Similar to the derivation of the update rule for γ_m, we can develop an update equation for σ^2 by computing derivative of the evidence with respect to σ^2. But the resulting update equation is not very useful in practice [4, 7, 8] and this is possibly because of the identifiability issue [4]. Hence we use traditional method to estimate σ^2 in this paper. Let $\mathbf{A}_{\mathcal{M}}$ denote the matrix formed by K columns of $\mathbf{A}$ indexed by $\mathcal{M}$, where the set $\mathcal{M}$ indicates the location of non-zero entries of $\mathbf{x}$, $|\mathcal{M}| = K$. We can estimate $\mathcal{M}$ using $\boldsymbol{\gamma}$. The noise variance estimate is then given by

$$\hat{\sigma}^2 = \frac{1}{N-K} \mathrm{Tr}\left((\mathbf{I}_N - \mathbf{A}_{\mathcal{M}} \mathbf{A}_{\mathcal{M}}^+) \mathbf{S}_{\mathbf{y}}\right), \quad (31)$$

where $\mathbf{A}_{\mathcal{M}}^+$ denotes the Moore-Penrose pseudo-inverse of the matrix $\mathbf{A}_{\mathcal{M}}$. This noise estimate has been used in [7, 8, 23].

4. SIMULATIONS

We use beamforming simulations to demonstrate the proposed SBL method. In plane wave beamforming the observed signal can be expressed as linear combination of plane waves arriving from different angles. The arrival angles are in the range $[-90, 90]^\circ$. Since the number of sources (arrival angles) is usually small, finely dividing the angle space results in a vector $\mathbf{x}$ of amplitudes which is sparse. SBL is used to recover the angle of arrivals. We assume narrow band processing and array sensor spacing is half the wavelength. The sensing matrix columns for beamforming are

$$\mathbf{a}_m^o = \frac{1}{\sqrt{N}}[1, e^{-j\pi\sin(\theta_m)}, \ldots, e^{-j(N-1)\pi\sin(\theta_m)}]^T \quad (32)$$

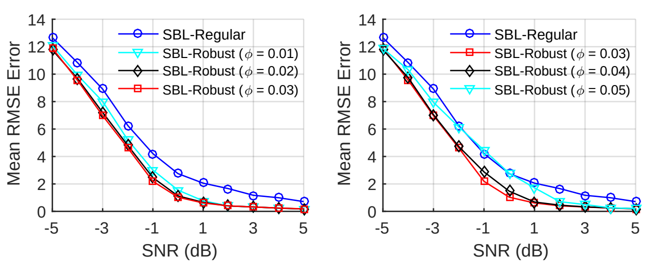
for $m = 1 \ldots M$ where θ_m is the mth discretized angle. The angle space is discretized with $1°$ separation giving $M = 181$. We consider $N = 20$ sensors. In SBL implementation we set $b = 1$. To simplify the simulations we make the assumption

$$\boldsymbol{\Sigma}_m^e = \phi \mathbf{I}_N, \quad \forall m = 1, 2, \ldots, M. \quad (33)$$

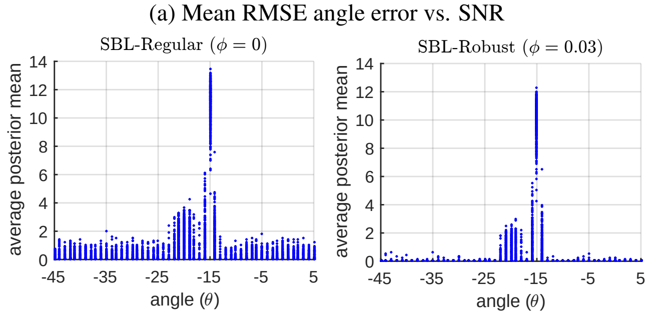
The parameter ϕ is a tuning parameter of the algorithm. When $\phi = 0$ we get SBL in [4, 8] which we refer to as SBL-Regular. The proposed SBL (29) is referred to as SBL-Robust in this section. For noise variance computation we estimate $\mathcal{M}$ by picking K highest peaks from $\boldsymbol{\gamma}^{\text{new}}$ at each iteration. On convergence, angle of arrivals are estimated corresponding to K highest peaks of $\boldsymbol{\gamma}^{\text{new}}$. Algorithm pseudo code can be found in [8] with modified update using (29). A convergence error threshold of 10^{-6} is used and we initialize $\boldsymbol{\gamma} = \mathbf{1}$.

4.1. Beamforming without perturbations

We consider three sources located at angles $[-20, -15, 75]°$ with amplitudes $[4, 13, 10]$ respectively. In this simulation the sensing matrix columns used for generating signals are given by (32) and there is no perturbation i.e. $\mathbf{a}_m^e = \mathbf{0}$. We process $L = 30$ snapshots.



(a) Mean RMSE angle error vs. SNR



(b) Posterior mean ($\phi = 0$) (c) Posterior mean ($\phi = 0.03$)

Fig. 1: Beamforming without perturbations: (a) Mean RMSE angle error vs. SNR for different ϕ; Scatter plot of average absolute posterior mean at (b) $\phi = 0$ and (c) $\phi = 0.03$.

We compute the angle RMSE at different SNR by varying values of ϕ. The RMSE is $\sqrt{\frac{1}{K}\sum_{k=1}^{K}(\widehat{\theta}_k - \theta_k^o)}$ where θ_k^o and $\widehat{\theta}_k$ are the true and estimated angles. The estimated angles correspond to the first K peaks in $\boldsymbol{\gamma}$. The mean error over 500 random trials is shown in Figure 1a. SBL-Regular (blue) is shown in both plots for comparison. For the proposed SBL as ϕ is increased from 0 to 0.03 the mean RMSE error reduces (left) and on further increasing ϕ from 0.03 to 0.05 the error starts increasing (right). The proposed SBL thus offers advantage even when there is no sensing matrix perturbation.

Scatter plots of posterior mean at 0 dB SNR are shown in Figures 1b and 1c. The plots show absolute posterior mean computed using (30) and averaged across snapshots. From Figure 1b, at low SNR, SBL-Regular often identifies spurious peaks which compete with the weaker source peak located at $-20°$. In Figure 1c, SBL-Robust ($\phi = 0.03$) is able to significantly suppress the amplitudes of the spurious peaks. The proposed SBL improves support estimate as reflected by the reduced average RMSE error in Figure 1a, though at the same time the source amplitudes are slightly underestimated as seen by comparing Figures 1b and 1c.

4.2. Beamforming with Gaussian perturbations

In this simulation we add zero mean complex Gaussian perturbations to sensing matrix. For generating signals, we sample the perturbations $\mathbf{a}_m^e$ from $\mathcal{CN}(\mathbf{a}_m^e; \mathbf{0}, \psi\mathbf{I}_N)$ independently for each snapshot. Note that the parameter ψ is different from ϕ that is used in the SBL implementation. We set $\psi = 0.02$. The average RMSE error as a function of SNR is shown in Figure 2. The error is higher in general because of the perturbations and SBL-Robust reduces the error for $\phi \in [0.01, 0.04]$.

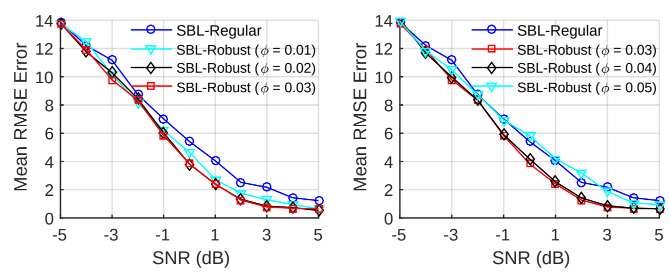


Fig. 2: Beamforming with Gaussian perturbations: Mean RMSE angle error vs. SNR for different values of ϕ.

5. CONCLUSIONS

We proposed a modified SBL to address the issue of perturbations present in the sensing matrix. This is achieved by computing approximate likelihood which integrates out the perturbations using the known statistics. This likelihood is used to compute the approximate evidence which on maximization gives an update rule for SBL. Simulations demonstrate that the proposed method is able to improve support recovery.

6. REFERENCES

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[2] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[3] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Machine Learning Research*, vol. 1, pp. 211–244, Jun. 2001.

[4] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Sig. Proc.*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.

[5] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Sig. Proc.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[6] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Sig. Proc.*, vol. 5, no. 5, pp. 912–926, 2011.

[7] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, "An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning," *IEEE Trans. Wireless Comm.*, vol. 11, no. 10, pp. 1–11, 2012.

[8] P. Gerstoft, C. F. Mecklenbräuker, A. Xenaki, and S. Nannuru, "Multi snapshot sparse Bayesian learning for DOA," *IEEE Sig. Proc. Letters*, vol. 23, no. 10, pp. 1469–1473, Oct. 2016.

[9] R. Giri and B. Rao, "Type I and type II Bayesian methods for sparse signal recovery using scale mixtures," *IEEE Trans. Sig. Proc.*, vol. 64, no. 13, pp. 3418–3428, Jul. 2016.

[10] A. Xenaki, P. Gerstoft, and K. Mosegaard, "Compressive beamforming," *J. Acoust. Soc. Am.*, vol. 136, no. 1, pp. 260–271, 2014.

[11] P. Gerstoft, A. Xenaki, and C. F. Mecklenbräuker, "Multiple and single snapshot compressive beamforming," *J. Acoust. Soc. Am.*, vol. 138, no. 4, pp. 2003–2014, 2015.

[12] P. A. Forero and P. A. Baxley, "Shallow-water sparsity-cognizant source-location mapping," *J. Acoust. Soc. Am.*, vol. 135, no. 6, pp. 3483–3501, 2014.

[13] K. L. Gemba, W. S. Hodgkiss, and P. Gerstoft, "Adaptive and compressive matched field processing," *J. Acoust. Soc. Am.*, vol. 141, no. 1, pp. 92–103, Jan. 2017.

[14] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE J. Sel. Topics Sig. Proc.*, vol. 4, no. 2, pp. 342–349, Apr. 2010.

[15] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Sig. Proc.*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[16] O. Teke, A. C. Gurbuz, and O. Arikan, "Perturbed orthogonal matching pursuit," *IEEE Trans. Sig. Proc.*, vol. 61, no. 24, pp. 6220–6231, Dec. 2013.

[17] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing - part I: Derivation," *IEEE Trans. Sig. Proc.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.

[18] I. Fedorov, R. Giri, B. D. Rao, and T. Q. Nguyen, "Robust Bayesian method for simultaneous block sparse signal recovery with applications to face recognition," in *IEEE Intl. Conf. Image Proc.*, Phoenix, Arizona, USA, Sep. 2016.

[19] W. Zhang and S. A. Vorobyov, "Joint robust transmit/receive adaptive beamforming for MIMO radar using probability-constrained optimization," *IEEE Sig. Proc. Letters*, vol. 23, no. 1, pp. 112–116, Jan. 2016.

[20] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Sig. Proc.*, vol. 51, no. 2, pp. 313–324, 2003.

[21] Z. M. Liu and Y. Y. Zhou, "A unified framework and sparse Bayesian perspective for direction-of-arrival estimation in the presence of array imperfections," *IEEE Trans. Sig. Proc.*, vol. 61, no. 15, pp. 3786–3798, Aug. 2013.

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. Series B (Methodological)*, pp. 1–38, 1977.

[23] P. Stoica and A. Nehorai, "On the concentrated stochastic likelihood function in array signal processing," *Circuits, Sys., Sig. Proc.*, vol. 14, no. 5, pp. 669–674, 1995.