SPARSE SIGNAL RECOVERY IN THE PRESENCE OF COLORED NOISE AND RANK-DEFICIENT NOISE COVARIANCE MATRIX: AN SBL APPROACH

Vinuthna Vinjamuri *Ranjitha Prasad* *Chandra R. Murthy*

Dept. of ECE, Indian Institute of Science, Bangalore, India
Email: {vinuthna, ranjitha.p, cmurthy}@ece.iisc.ernet.in

ABSTRACT

In this work, we address the recovery of sparse and compressible vectors in the presence of colored noise possibly with a rank-deficient noise covariance matrix, from overcomplete noisy linear measurements. We exploit the structure of the noise covariance matrix in a Bayesian framework. In particular, we propose the *CoNo-SBL* algorithm based on the popular and efficient Sparse Bayesian Learning (SBL) technique. We also derive Bayesian and Marginalized Cramér Rao lower Bounds (CRB) for the problem of estimating compressible vectors. We consider an unknown compressible vector drawn from a Student-t prior distribution, and derive CRBs that encompass the random nature of the unknown compressible vector and the parameters of the prior distribution, in the presence of colored noise and rank-deficient noise covariance matrix. Using Monte Carlo simulations, we demonstrate the efficacy of the proposed CoNo-SBL algorithm as compared to compressed sensing and greedy techniques. Further, we demonstrate the mean squared error performance of the proposed estimator compared to the CRBs, for different ranks of the noise covariance matrix.

Index Terms— Sparse Bayesian learning, colored noise, rank-deficient noise covariance matrix, expectation maximization, Cramér Rao lower bounds

1. INTRODUCTION

Recently, the problem of sparse signal recovery has received immense interest as it enjoys numerous applications in signal processing and machine learning. Compressed Sensing (CS) [1] and Bayesian techniques [2–4] have been proposed for obtaining robust solutions to the problem of sparse recovery, which involves estimating a sparse vector $\mathbf{x} \in \mathbb{R}^{N\times 1}$ from an overcomplete system of linear equations given by

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

where $\mathbf{\Phi} \in \mathbb{R}^{m\times N}$ ($m << N$) represents the overcomplete basis, and $\mathbf{y} \in \mathbb{R}^{m\times 1}$ represents the observations. In the conventional sparse recovery framework [1, 2], the ambient noise $\mathbf{n} \in \mathbb{R}^{m\times 1}$ is distributed as $\mathbf{n} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m)$, i.e., $\mathbf{n}$ is modeled as Additive White Gaussian Noise (AWGN). However, in real-world scenarios, we often encounter situations where the noise $\mathbf{n}$ is not white, i.e., $\mathbb{E}[nn^T] = \mathbf{Q}$, and $\mathbf{Q}$ is a non-negative definite matrix.

Among the existing recovery methods, the ones based on CS such as Basis Pursuit Denoising (BPDN), LASSO [5] etc., assume that the noise is bounded in magnitude. Another popular class of algorithms used for sparse recovery constitute the greedy approaches such as Matching Pursuit (MP) [6], Orthogonal MP (OMP) [7], CoSAMP [8], etc. These algorithms are oblivious to the structure of the noise, and hence, are not designed to exploit the known properties of the noise covariance matrix. On the other hand, Bayesian approaches are capable of elegantly incorporating the structure in the noise covariance matrix into the problem of sparse recovery. However, the algorithms proposed in the Bayesian framework thus far [4, 9, 10] are based on the AWGN model.

In order to cater to specific applications, extensions of the basic CS and greedy approaches have been proposed for structured noise scenarios. For e.g., in [11], the authors consider the problem of estimating a UWB channel impulse response in the presence of colored noise and propose a Matching Pursuit (MP) approach. In [12], the authors consider the colored noise due to noise folding in spread spectrum based receivers. Such techniques pre-whiten the observations prior to sparse signal recovery instead of explicitly incorporating the noise structure into the sparse recovery formulation.

In the Bayesian framework, a family of techniques known as Sparse Bayesian Learning (SBL) has been developed to find robust solutions to the sparse signal recovery problems. A feature of these algorithms is that it is simple to incorporate the structure and correlation constraints inherent to the sparse vector [13, 14]. In this work, we demonstrate that it is possible to incorporate the underlying noise structure in the SBL framework. Specifically, in the SBL framework, we model the prior distribution on the sparse vector $\mathbf{x}$ as $\mathbf{x} \sim \mathcal{N}(0, \Gamma)$, where $\Gamma = \text{diag}(\gamma(1), \ldots, \gamma(N))$ represents the unknown hyperparameters. Further, we model the noise as $\mathbf{n} \sim \mathcal{N}(0, \mathbf{Q})$, where $\mathbf{Q}$ is the noise covariance matrix. The results presented in this paper address the following:

1. Colored noise: Typically, in distributed sensor network applications, every sensor has a set of measurements of a sparse signal, and the goal is to recover the signal from their collective measurements at a minimal communication cost and low computational complexity. Since the observations at various sensors experience an ambient noise of different noise variance [15], one essentially deals with a sparse recovery problem where $\mathbf{Q}$ is diagonal, but consists of unequal values along the diagonal.

2. Low-rank noise covariance matrix: In [16, 17], the authors consider a problem of recovering sparse signals from undersampled measurements corrupted by very large but correlated noise. Such scenarios are often encountered in real-time video surveillance and layering [18]. Note that our framework can handle such scenarios by considering $\mathbf{Q}$ to be low-rank and with nonzero off diagonal entries.

This work was financially supported in part by research grants from the Defense Research Development Organization, the Department of Electronics and Information Technology, Govt. of India, and the TCS research scholar program.

1.1. Problem Statement and Contributions

In this section, we describe the problem addressed and the contributions of this work. The sparse recovery problem given by (1) can be generalized to the several scenarios listed in the previous section by considering $\mathbf{Q}$ to be a colored covariance matrix. We express $\mathbf{Q}$ using the eigenvalue decomposition, as follows:

$$\mathbf{Q} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T, \tag{2}$$

where $\mathbf{V} \in \mathbb{R}^{m\times m}$ consists of m orthonormal columns, and $\boldsymbol{\Lambda}$ is a diagonal matrix consisting of eigenvalues of $\mathbf{Q}$. In the case when $\mathbf{Q}$ is full-rank, $\boldsymbol{\Lambda}$ has nonzero diagonal entries. When $\mathbf{Q}$ is rank-deficient, it can be written as

$$\mathbf{Q} = [\mathbf{V}_1\mathbf{V}_2]\begin{bmatrix} \mathbf{D} & \mathbf{0}_{p\times m-p} \\ \mathbf{0}_{m-p\times p} & \mathbf{0}_{m-p\times m-p}\end{bmatrix}\begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T\end{bmatrix}, \tag{3}$$

where p is the rank of $\mathbf{Q}$, $\mathbf{0}_{p\times m}$ represents a $p\times m$ matrix of zeros, and $\mathbf{D}$ is a diagonal matrix consisting of the p nonzero eigenvalues of $\mathbf{Q}$. This leads to a system of linear equations given by

$$\begin{aligned}\mathbf{y}_1 &= \boldsymbol{\Phi}_1\mathbf{x} + \mathbf{n}_1 \\ \mathbf{y}_2 &= \boldsymbol{\Phi}_2\mathbf{x} + \mathbf{n}_2.\end{aligned} \tag{4}$$

Here, the observations $\mathbf{y}$ are projected onto two orthogonal subspaces, such that in one of the subspaces the measurements are noisy and drawn from a Gaussian distribution governed by a diagonal covariance matrix, while in the other subspace, the measurements are noiseless. Accordingly, $\mathbf{y}_1 = \mathbf{V}_1^T\mathbf{y}$ and $\boldsymbol{\Phi}_1 = \mathbf{V}_1^T\boldsymbol{\Phi}$, and on similar lines, $\mathbf{y}_2 = \mathbf{V}_2^T\mathbf{y}$ and $\boldsymbol{\Phi}_2 = \mathbf{V}_2^T\boldsymbol{\Phi}$. Note that since $\mathbf{n}_1 = \mathbf{V}_1^T\mathbf{n}$, $\mathbb{E}[\mathbf{n}_1\mathbf{n}_1^T] = \mathbf{D}$ and similarly since $\mathbf{n}_2 = \mathbf{V}_2^T\mathbf{n}$, $\mathbb{E}[\mathbf{n}_2\mathbf{n}_2^T] = \mathbf{0}_{m-p}$.

In this work, we propose an Expectation Maximization (EM) based CoNo-SBL algorithm for recovery of *sparse* and *compressible* vectors in the presence of correlated noise, which may be full-rank or rank-deficient. We demonstrate that Mean Squared Error (MSE) performance of the proposed CoNo-SBL estimator is superior to CS based LASSO and greedy methods such as OMP. Further, we derive Cramér Rao type bounds assuming that the vector $\mathbf{x}$ is drawn from a compressible Student-t prior distribution. In particular, for the estimation problem stated in this paper, Bayesian (B) and Marginalized (M) Cramér Rao Bounds (CRB) [19] are derived to obtain lower bounds on the MSE performance of the proposed estimator, by incorporating the prior distribution on $\mathbf{x}$ and correlation structure in $\mathbf{Q}$. We demonstrate while MCRB is tighter than the BCRB, the performance of the proposed estimator is just 3 dB away from MCRB. When $\mathbf{Q}$ is rank deficient, with rank $p < m$, it is equivalent to having p noisy and $m-p$ noiseless measurements. To the best of our knowledge, the problem of sparse signal recovery with a combination of noisy and noiseless measurements has not been considered in the literature.

In the following section, we present our proposed algorithm for recovering a sparse vector from noisy linear measurements, for the case when the noise covariance matrix $\mathbf{Q}$ is colored, and in particular for the case when $\mathbf{Q}$ is colored and rank-deficient.

2. PROPOSED ALGORITHMS

In this section, we propose the CoNo-SBL algorithm for the recovery of sparse/compressible vectors in the presence of a noise distributed as $\mathbf{n} \sim \mathcal{N}(0, \mathbf{Q})$ for the observation model given by (1). The conventional SBL framework [9] uses a parameterized prior to induce sparsity in the solution, given by

$$p(\mathbf{x};\boldsymbol{\gamma}) = \prod_{i=1}^{N}(2\pi\gamma(i))^{-1}\exp\left(-\frac{|x(i)|^2}{\gamma(i)}\right). \tag{5}$$

In the prior density given in (5), the hyperparameters $\boldsymbol{\gamma}$ are unknown, and can be estimated using the type-II Maximum Likelihood (ML) procedure [20], i.e., by maximizing the marginalized pdf $p(\mathbf{y};\boldsymbol{\gamma})$ as

$$\hat{\boldsymbol{\gamma}}_{ML} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{N\times 1}}{\arg\max}\, p(\mathbf{y};\boldsymbol{\gamma}). \tag{6}$$

Since the above problem cannot be solved in closed form, iterative estimators such as the EM algorithm is employed. The sparse/compressible vector $\mathbf{x}$ is considered as the hidden variable and the ML estimate of $\boldsymbol{\gamma}$ is obtained in the M-step. The steps of the algorithm can be given as

$$\text{E-step}: Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right) = \mathbb{E}_{\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y},\mathbf{x};\boldsymbol{\gamma})] \tag{7}$$

$$\text{M-step}: \boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{N\times 1}}{\arg\max}\, Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right). \tag{8}$$

The E-step above involves computation of the posterior density of $\mathbf{x}$, where the hyperparameters $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(r)}$, i.e., in order to obtain the posterior distribution in the $(r+1)^{\text{th}}$ iteration, we utilize the hyperparameter update obtained in the M-step of the r^{th} iteration. Accordingly, the posterior density of $\mathbf{x}$ can be expressed as

$$p\left(\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}^{(r)}\right) = \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}), \tag{9}$$

where $\boldsymbol{\Sigma} = \Gamma^{(r)} - \Gamma^{(r)}\boldsymbol{\Phi}^T\left(\mathbf{Q} + \boldsymbol{\Phi}\Gamma^{(r)}\boldsymbol{\Phi}^T\right)^{-1}\boldsymbol{\Phi}\Gamma^{(r)}$, and $\boldsymbol{\mu} = \boldsymbol{\Sigma}\boldsymbol{\Phi}^T\mathbf{Q}^{-1}\mathbf{y}$. The M-step in (8) can be simplified, to obtain

$$\gamma^{(r+1)}(i) = \underset{\gamma(i)\in\mathbb{R}_+}{\arg\max}\, \mathbb{E}_{\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}^{(r)}}\left[\log p(\mathbf{x};\boldsymbol{\gamma})\right] \tag{10}$$

$$= \mathbb{E}_{\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}^{(r)}}\left[|x(i)|^2\right] = \Sigma(i,i) + |\mu(i)|^2. \tag{11}$$

In (10), the term $\mathbb{E}_{\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y}|\mathbf{x};\boldsymbol{\gamma})]$ has been dropped, as it is not a function of $\gamma(i)$. Note that, since all the algorithms proposed in this paper use EM updates, they have monotonicity property, i.e., the likelihood is guaranteed to increase at each iteration [21, 22].[1] In the case of rank-deficient matrix $\mathbf{Q}$, it is necessary that we derive EM based update equations considering the zero eigenvalues as given in (3). In order to derive the update equations accomodating for a rank-deficient noise covariance matrix, we start with the observation model (4) and consider $\tilde{\mathbf{y}} = \tilde{\boldsymbol{\Phi}}\mathbf{x} + \tilde{\mathbf{n}}$, where $\tilde{\mathbf{y}} = [\mathbf{y}_1^T\mathbf{y}_2^T]^T$, $\tilde{\boldsymbol{\Phi}} = [\boldsymbol{\Phi}_1^T\boldsymbol{\Phi}_2^T]^T$ and $\tilde{\mathbf{n}} \sim \mathcal{N}(0,\boldsymbol{\Lambda})$ with $\boldsymbol{\Lambda}$ as defined in (2) and (3).

In the case when $\mathbf{Q}$ is rank-deficient, we first let $\mathbb{E}[\mathbf{n}_2\mathbf{n}_2^T] = \sigma_2^2\mathbf{I}_{m-p}$. The EM updates then take the form

$$\boldsymbol{\Sigma} = \Gamma^{(r)} - \Gamma^{(r)}\left(\sum_{m=1}^{2}\sum_{n=1}^{2}\boldsymbol{\Phi}_n^T\mathbf{B}_{nm}\boldsymbol{\Phi}_m\right)\Gamma^{(r)}, \tag{12}$$

where

$$\begin{bmatrix}\mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22}\end{bmatrix} = \begin{bmatrix}\mathbf{D} + \boldsymbol{\Phi}_1^T\Gamma^{(r)}\boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_1\Gamma^{(r)}\boldsymbol{\Phi}_2^T \\ \boldsymbol{\Phi}_2\Gamma^{(r)}\boldsymbol{\Phi}_1^T & \sigma_2^2\mathbf{I}_{m-p} + \boldsymbol{\Phi}_2\Gamma^{(r)}\boldsymbol{\Phi}_2^T\end{bmatrix}^{-1}. \tag{13}$$

Also, we have $\mu = \boldsymbol{\Sigma}\boldsymbol{\Phi}_1^T\mathbf{D}^{-1}\mathbf{y}_1 + \sigma_2^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}_2^T\mathbf{y}_2$.

Applying $\sigma_2^2 \to 0$, using straightforward block matrix inversion rules [23] and the identity $\lim_{\delta\to 0}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \delta\mathbf{I}_P)^{-1} = \mathbf{A}^{\dagger}$,

[1] We have found, empirically, that a straightforward initialization such as $\Gamma^{(0)} = \mathbf{I}_N$ leads to accurate solutions.

where $\mathbf{A}^\dagger$ represents the Moore-Penrose pseudo-inverse of $\mathbf{A} \in \mathbb{R}^{P\times L}$ [24], we obtain the following expression for $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \Gamma^{(r)} - \Gamma^{(r)^{\frac{1}{2}}}\Theta_1^T\mathbf{B}_{11}\Theta_1\Gamma^{(r)^{\frac{1}{2}}} - \Gamma^{(r)^{\frac{1}{2}}}\Theta_2^\dagger\mathbf{U}_1\Theta_2\Gamma^{(r)\frac{1}{2}} + \Gamma^{(r)\frac{1}{2}}\Theta_1^T\mathbf{\Sigma}_{t1}^{-1}\Theta_1\Theta_2^\dagger\mathbf{U}_1\Theta_2\Gamma^{(r)\frac{1}{2}} + \Gamma^{(r)\frac{1}{2}}\Theta_2^\dagger\Theta_2\Theta_1^T\mathbf{B}_{11}\Theta_1\Gamma^{(r)\frac{1}{2}}. \tag{14}$$

where $\mathbf{B}_{11} = (\mathbf{\Sigma}_v - \Theta_1\Theta_2^\dagger\Theta_2\Theta_1^T)^{-1}$, $\mathbf{\Sigma}_v = (\mathbf{D} + \mathbf{\Phi}_1\Gamma^{(r)}\mathbf{\Phi}_1^T)$, $\Theta_1 = \mathbf{\Phi}_1\Gamma^{(r)\frac{1}{2}}$, $\Theta_2 = \mathbf{\Phi}_2\Gamma^{(r)\frac{1}{2}}$, and $\mathbf{U}_1 = \mathbf{I}_{m-p} + \Theta_2\Theta_1^T\mathbf{B}_{11}\Theta_1\Theta_2^\dagger$. Further, using the identity $\lim_{\delta\to 0}(\mathbf{A}^T\mathbf{A} + \delta\mathbf{I}_L)^{-1}\mathbf{A}^T = \mathbf{A}^\dagger$, the posterior mean $\boldsymbol{\mu}$ is given as

$$\boldsymbol{\mu} = \mathbf{\Sigma}\mathbf{\Phi}_1^T\mathbf{D}^{-1}\mathbf{y}_1 + \Gamma^{(r)\frac{1}{2}}\mathbf{U}_2^{\frac{1}{2}}(\Theta_2\mathbf{U}_2^{\frac{1}{2}})^\dagger\mathbf{y}_2, \tag{15}$$

where $\mathbf{U}_2 = (\mathbf{I}_N + \Theta_1^T\mathbf{D}^{-1}\Theta_1)^{-1}$. Since $\Theta_1^T\mathbf{D}^{-1}\Theta_1$ is rank-deficient, $\mathbf{U}_2$ can be found using the *Sherman-Morrison-Woodbury update* as

$$\mathbf{U}_2 = \mathbf{I}_N - \Theta_1^T\mathbf{\Sigma}_v^{-1}\Theta_1. \tag{16}$$

Thus, the final EM updates evaluate $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$ iteratively using above expressions until convergence. The steps are summarized in Algorithm 1.

Algorithm 1 CoNo-SBL Algorithm

1: Initialize $\Gamma \leftarrow \mathbb{I}_N$,
2: **while** $(\Gamma^{(r+1)} - \Gamma^{(r)}) < 10^{-6}$ or $r < 300$ **do**
3: $\mathbf{\Sigma} = \Gamma^{(r)} - \Gamma^{(r)\frac{1}{2}}\Theta_1^T\mathbf{B}_{11}\Theta_1\Gamma^{(r)\frac{1}{2}} - \Gamma^{(r)\frac{1}{2}}\Theta_2^\dagger\mathbf{U}_1\Theta_2\Gamma^{(r)\frac{1}{2}} + \Gamma^{(r)\frac{1}{2}}\Theta_1^T\mathbf{\Sigma}_{t1}^{-1}\Theta_1\Theta_2^\dagger\mathbf{U}_1\Theta_2\Gamma^{(r)\frac{1}{2}} + \Gamma^{(r)\frac{1}{2}}\Theta_2^\dagger\Theta_2\Theta_1^T\mathbf{B}_{11}\Theta_1\Gamma^{(r)\frac{1}{2}}$
4: $\boldsymbol{\mu} = \mathbf{\Sigma}\mathbf{\Phi}_1^T\mathbf{D}^{-1}\mathbf{y}_1 + \Gamma^{(r)\frac{1}{2}}\mathbf{U}_2^{\frac{1}{2}}(\Theta_2\mathbf{U}_2^{\frac{1}{2}})^\dagger\mathbf{y}_2$,
5: $\gamma_i^{(r+1)} \leftarrow |\boldsymbol{\mu}_i|^2 + \mathbf{\Sigma}_{ii}$
6: $(r) \leftarrow (r+1)$
7: **end while**
8: Output $\boldsymbol{\mu}$

In the following section, we derive lower bounds on the MSE performance of CoNo-SBL estimator.

3. CRAMÉR RAO TYPE BOUNDS: BCRB AND MCRB

In this section, we derive Bayesian and marginalized Cramér Rao type lower bounds (BCRB and MCRB) for the system in (4), where the unknown vector is given by $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$ and the signal $\mathbf{x}$ is drawn from a compressible prior distribution [25]. We model the sparse vector as being random and $\boldsymbol{\gamma}$ as being random or marginalized [26]. However, in contrast to [26], we derive the lower bounds in the presence of noise with a general covariance matrix $\mathbf{Q}$.

3.1. BCRB for $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$

In this subsection, we consider the unknown vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, where the compressible vector $\mathbf{x}$ is distributed according to a Gaussian distribution parameterized by $\boldsymbol{\gamma}$. For deriving the BCRB, a hyperprior distribution is considered on $\boldsymbol{\gamma}$, and, as a result, the vector $\mathbf{x}$ is drawn from a compressible prior distribution. We consider the Inverse Gamma (IG) hyperprior distribution [20], where, $\gamma_i, i = 1, 2, \ldots, N$ are distributed as $\mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right)$. The IG distribution is given by

$$p(\gamma_i) \triangleq \left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1}\left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}}\gamma_i^{\left(-\frac{\nu}{2}-1\right)}\exp\left\{-\frac{\nu}{2\lambda\gamma_i}\right\}, \tag{17}$$

where $\gamma_i \in (0, \infty)$, $\nu, \lambda > 0$. From the definition of the BCRB, we state the following proposition.

Proposition 1 *For the signal model in* (4) *with noise covariance matrix* $\mathbf{Q}$, *the BCRB on the MSE matrix* $\mathbf{E}^{\boldsymbol{\theta}}$ *of the unknown random vector* $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, *where the conditional distribution of the compressible signal* $\mathbf{x}|\boldsymbol{\gamma}$ *is* $\mathcal{N}(0, \Gamma)$, *and the hyperprior distribution on* $\boldsymbol{\gamma}$ *is* $\prod_{i=1}^N \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right)$, *is given by* $\mathbf{E}^{\boldsymbol{\theta}} \succeq (\mathbf{B}^{\boldsymbol{\theta}})^{-1}$, *where*

$$\mathbf{B}^{\boldsymbol{\theta}} \triangleq \begin{bmatrix} \mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}) & \mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\gamma}) \\ (\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\gamma}))^T & \mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) \end{bmatrix} = \begin{bmatrix} \frac{1}{\lambda}(\mathbf{I}_m - \tilde{\mathbf{\Phi}}(\lambda\Lambda + \tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^T)\tilde{\mathbf{\Phi}}^T) & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \frac{\lambda^2(\nu+2)(\nu+7)}{2\nu}\mathbf{I}_N \end{bmatrix}. \tag{18}$$

Proof: Using the definitions in [26], $\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x})$ can be computed as

$$\begin{aligned} \mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}) &\triangleq -\mathbb{E}_{\mathbf{Y},\mathbf{x},\Gamma}\left[\nabla_{\mathbf{x}}^2 \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\gamma})\right] \\ &= -\mathbb{E}_{\mathbf{Y},\mathbf{x},\Gamma}\left[\nabla_{\mathbf{x}}\left(\tilde{\mathbf{\Phi}}^T\mathbf{\Lambda}^{-1}(\mathbf{y} - \tilde{\mathbf{\Phi}}\mathbf{x}) - \Gamma^{-1}\mathbf{x}\right)\right] \\ &= \tilde{\mathbf{\Phi}}^T\mathbf{\Lambda}^{-1}\tilde{\mathbf{\Phi}} + \mathbb{E}_\Gamma\left[\Gamma^{-1}\right] \\ &= \tilde{\mathbf{\Phi}}^T\mathbf{\Lambda}^{-1}\tilde{\mathbf{\Phi}} + \lambda\mathbf{I}_N. \end{aligned} \tag{19}$$

However, note that in the case when $\mathbf{Q}$ is rank-deficient, $\mathbf{\Lambda}$ is not invertible. In this case, $(\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}))^{-1}$ can be expressed as

$$(\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}))^{-1} \triangleq \frac{1}{\lambda}(\mathbf{I}_m - \tilde{\mathbf{\Phi}}(\lambda\mathbf{\Lambda} + \tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^T)\mathbf{\Phi}^T). \tag{20}$$

Further, $\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\gamma})$ and $\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ remains the same as in [26]. ∎

It is known that the MCRB is the lower bound to the BCRB [26]. In the following subsection, we derive the MCRB by marginalizing $\boldsymbol{\gamma}$ from the joint distribution of $\mathbf{x}$ and $\boldsymbol{\gamma}$.

3.2. MCRB for $\boldsymbol{\theta} = [\boldsymbol{\gamma}]$

In this subsection, we consider an IG hyperprior on $\boldsymbol{\gamma}$ as in the conventional SBL framework. Effectively, this leads to a compressible vector $\mathbf{x}$ with a Student-t distribution. Such bounds have been derived in [26] where the authors obtain MCRB for vectors sampled from a Student-t distribution with parameters ν and λ, i.e., a ν-compressible $\mathbf{x}$ [25]. The Student-t prior is given by

$$p(\mathbf{x}) \triangleq \left(\frac{\Gamma((\nu+1)/2)}{\Gamma(\frac{\nu}{2})}\right)^N \left(\frac{\lambda}{\pi\nu}\right)^{\frac{N}{2}} \prod_{i=1}^N \left(1 + \frac{\lambda x_i^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}, \tag{21}$$

where $x_i \in (-\infty, \infty)$, $\nu, \lambda > 0$, and ν represents the number of degrees of freedom and λ represents the inverse variance of the distribution. Accordingly, we state the following theorem for deriving the MCRB.

Proposition 2 *For the signal model in* (4), *the MCRB on the MSE matrix* $\mathbf{E}^{\mathbf{x}}$ *of the unknown compressible random vector* $\boldsymbol{\theta} = [\mathbf{x}]$ *distributed as* (21), *is given by* $\mathbf{E}^{\mathbf{x}} \succeq (\mathbf{M}^{\mathbf{x}})^{-1}$, *where*

$$(\mathbf{M}^{\mathbf{x}})^{-1} = \frac{(\nu+3)}{\lambda(\nu+1)}\left(\mathbf{I} - \tilde{\mathbf{\Phi}}\left(\frac{(\nu+1)\lambda}{(\nu+3)}\mathbf{\Lambda} + \tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^T\right)\tilde{\mathbf{\Phi}}^T\right) \tag{22}$$

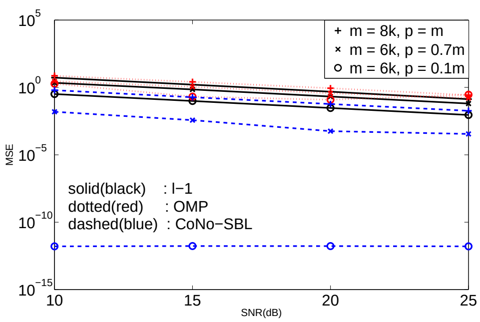
Proof: Omitted due to lack of space; but the proof follows along the lines of Proposition 1 and [26]. ∎

4. SIMULATION RESULTS

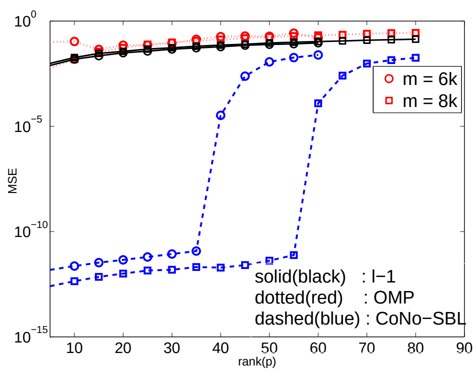
In this section, we illustrate the Mean Squared Error (MSE) performance of the proposed CoNo-SBL algorithm and compare it with existing methods. We also compare against the Cramér Rao type bounds derived in Sec. 3. We generate a k-sparse vector $\mathbf{x} \in \mathbb{R}^N$, whose nonzero entries are i.i.d. according to an equiprobable Bernoulli ($x(i) \in \{+1, -1\}$) distribution. In each trial, the measurement matrix $\mathbf{\Phi}$ is generated as a random overcomplete matrix, whose entries are i.i.d. and standard Gaussian distributed and the columns are normalized to have unit euclidean norm. The experiment is repeated for 1000 trials. The noisy measurements are corrupted by white noise with known noise variance σ^2. We consider $N = 100$ and $k = 10$, m represents the number of measurements and p the rank of the noise covariance matrix. We fix the number of iterations of the proposed algorithm to 300 and the convergence criterion is given by $||\Gamma^{r+1} - \Gamma^r||_2 < 10^{-6}$.

4.1. CoNo-SBL Algorithm

In this section, we simulate the proposed CoNo-SBL algorithm, and compare its performance with OMP [7], and the convex optimization based approach known as LASSO [5].



(a) MSE vs. SNR for different p and m



(b) MSE vs. p for different m, with SNR$=$ 10dB

Fig. 1. Comparison of the MSE performance of the CoNo-SBL with OMP [7] and LASSO [5].

The MSE performance of the CoNo-SBL algorithm across SNR for different values of p, the rank of the noise covariance matrix $\mathbf{Q}$, is depicted in Fig. 1(a). We observe that the MSE performance when as little as 10% of the measurements are noiseless is markedly superior to the case when all the measurements are noisy. Although similar behavior is observed with OMP and LASSO, the proposed CoNo-SBL algorithm has a superior performance compared to the other schemes as it is able to exploit the structure in the noise covariance matrix. Interestingly, in the case of CoNo-SBL algorithm, we see that even at low SNR, having a few noiseless measurements leads to significant MSE improvements, and the sparse signal can be recovered with an MSE in the order of 10^{-2}.

In Fig. 1(b), we demonstrate the MSE performance of the CoNo-SBL algorithm as a function of the rank p. Again, we observe that, compared to OMP and LASSO, the CoNo-SBL algorithm utilizes the rank-deficient structure of $\mathbf{Q}$ effectively, leading to a better MSE performance. We note that there is an interesting tradeoff between the number of measurements and the rank of Q: when the rank of Q is low, fewer measurements suffice to achieve the same MSE. For instance, the MSE of CoNo-SBL algorithm with $m = 60$ and $p = 0.7m$ is better than the MSE with $m = 80$ and $p = m$.

4.2. Cramér Rao type bounds

In Fig. 2, we compare the performance of the proposed CoNo-SBL algorithm for recovering a compressible vector $\mathbf{x}$, with the Cramér Rao type bounds derived in Sec. 3. Specifically, we consider an IG hyperprior distribution on Γ and a conditional Gaussian distribution on $\mathbf{x}$, parameterized by Γ. Hence, the resulting compressible signal is viewed as being drawn from Student-t distribution, which is compressible [25]. We consider the parameters of the hyperprior given by $\nu = 2.05$ and $\lambda = 2000$ [26]. The figure compares the performance of CoNo-SBL algorithm as a function of rank, p, for different values of m, with the MCRB and the BCRB given in (22) and (18), respectively. First, note that MCRB is tighter compared to BCRB [26]. However, a more interesting point, which has been observed empirically but is not shown in the plot to avoid clutter is that a slight rank deficiency in the noise covariance matrix $\mathbf{Q}$ is sufficient to get a considerable improvement in the MSE performance. Note that rank deficiency leads to a system model given by (4), encompassing a few noiseless measurements which results in a significant improvement in the recovery performance when the recovery algorithm explicitly accounts for the known noise statistics.

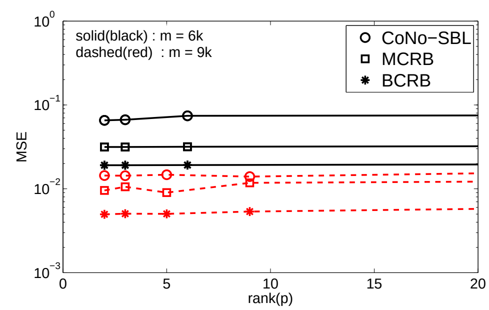


Fig. 2. MSE of the proposed CoNo-SBL technique compared to BCRB and MCRB as a function of rank p of $\mathbf{Q}$ for different m.

5. CONCLUSIONS

In this work, we proposed a novel CoNo-SBL algorithm for recovery of sparse and compressible signals contaminated by colored noise with a rank-deficient noise covariance matrix. We showed that the rank-deficient structure of the noise covariance matrix leads to noiseless measurements, which can be utilized to improve the MSE performance as compared to the existing methods. In the context of compressible signal estimation, we derived the Bayesian and marginalized CRB for the case where the noise covariance matrix is rank-deficient. We saw that MCRB is tighter than BCRB, and that the performance of the CoNo-SBL algorithm is close to the MCRB. Future work could extend the CoNo-SBL algorithm to perform joint sparse vector recovery and noise covariance matrix estimation.

6. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] M. E. Tipping, "The relevance vector machine," in *Advances in NIPS*, vol. 12, 2000.

[3] S. Babacan, R. Molina, and A. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, pp. 53–64, 2010.

[4] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, 2008.

[5] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[6] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[7] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[8] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[9] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, 2004.

[10] P. Schniter, L. C. Potter, and J. Ziniel, "Fast Bayesian matching pursuit," in *Info. Th. and Appli. Workshop*, 2008, pp. 326–333.

[11] M. Al Atassi and I. Abou-Faycal, "A reconstruction algorithm for noisy compressed sensing; the UWB channel estimation test case," in *Intl. Conf. on Telecom.*, 2012, pp. 1–6.

[12] K. Fyhn, T. L. Jensen, T. Larsen, and S. H. Jensen, "Compressive sensing for spread spectrum receivers," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2334–2343, 2013.

[13] R. Prasad, C. Murthy, and B. D. Rao, "Nested sparse Bayesian learning for block-sparse signals with intra-block correlation," in *Proc. ICASSP*, 2010, pp. 1–6.

[14] Z. Zhang and B. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, 2013.

[15] S. Patterson, Y. C. Eldar, and I. Keidar, "Distributed sparse signal recovery for sensor networks," in *Proc. ICASSP*, 2013, pp. 4494–4498.

[16] C. Qiu and N. Vaswani, "Recursive sparse recovery in large but correlated noise," in *Proc. Allerton Conf. on Commun., Control and Comput.*, 2011, pp. 752–759.

[17] C. Qiu, N. Vaswani, and L. Hogben, "Recursive robust PCA or recursive sparse recovery in large but structured noise," in *Proc. ICASSP*, 2013, pp. 5954–5958.

[18] T. Bouwmans and E. H. Zahzah, "Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance," *Computer Vision and Image Understanding*, vol. 122, pp. 22–34, 2014.

[19] H. V. Trees, *Detection, Estimation, and Modulation Theory - Part I*. New York: Wiley, 1968, vol. 2.

[20] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–214, 2001.

[21] C. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, 1983.

[22] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*. Wiley New York, 1997, vol. 274.

[23] K. Petersen and M. Pedersen, "The matrix cookbook," *Online: http://matrixcookbook.com*, 2008.

[24] G. H. Golub and C. F. van Van Loan, *Matrix computations (Johns Hopkins studies in mathematical sciences)*. The Johns Hopkins University Press, 1996.

[25] V. Cevher, "Learning with compressible priors," *Advances in NIPS*, pp. 7–12, 2008.

[26] P. Ranjitha and C. R. Murthy, "Cramér-Rao–type bounds for sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 622–632, 2013.