NESTED SPARSE BAYESIAN LEARNING FOR BLOCK-SPARSE SIGNALS WITH INTRA-BLOCK CORRELATION

Ranjitha Prasad[†] Chandra R. Murthy[†] Bhaskar D. Rao^{*}

[†] Dept. of ECE, Indian Institute of Science, Bangalore, India *Dept. of ECE, UC San Diego, La Jolla, CA, USA email: [†]{ranjitha.p, cmurthy}@ece.iisc.ernet.in *brao@ece.ucsd.edu

ABSTRACT

In this work, we address the recovery of block sparse vectors with intra-block correlation, i.e., the recovery of vectors in which the correlated nonzero entries are constrained to lie in a few clusters, from noisy underdetermined linear measurements. Among Bayesian sparse recovery techniques, the cluster Sparse Bayesian Learning (SBL) is an efficient tool for block-sparse vector recovery, with intrablock correlation. However, this technique uses a heuristic method to estimate the intra-block correlation. In this paper, we propose the Nested SBL (NSBL) algorithm, which we derive using a novel Bayesian formulation that facilitates the use of the monotonically convergent nested Expectation Maximization (EM) and a Kalman filtering based learning framework. Unlike the cluster-SBL algorithm, this formulation leads to closed-form EM updates for estimating the correlation coefficient. We demonstrate the efficacy of the proposed NSBL algorithm using Monte Carlo simulations.

1. INTRODUCTION

In recent literature, techniques such as Compressed Sensing(CS) [1] and Bayesian methods [2-5] have been proposed for efficiently reconstructing sparse signals from an underdetermined system of linear equations. In this paper, we consider the recovery of sparse signals which exhibit additional structure, wherein, the nonzero entries are constrained to occur in a few clusters, i.e., signals are block-sparse, and the entries within a nonzero block are correlated with each other. There are several applications where block-sparsity and intra-block correlation arise naturally (see [6] and references therein). In particular, strong intra-block correlation has been observed in EEG, ECG and several physiological signals [7]. Popular CS based approaches exploit block-sparsity in linear models using mixed penalty, such as the $\ell_1 - \ell_2$ and $\ell_1 - \ell_\infty$ [8–10], block matching pursuit, block orthogonal matching pursuit [9], and block-CoSamp [11]. However, none of the techniques based on CS exploit the intra-block correlation in the block-sparse signal. In the Bayesian framework, a block-sparse vector recovery algorithm known as the cluster-SBL algorithm [12, 13] is proposed, which, in addition to incorporating the block-sparse structure into the prior probability density function (pdf), also exploits the intrablock correlation. However, when the intra-block correlation is not known, the cluster-SBL framework uses an approximate heuristic to compute this intra-block correlation.

In the context of Bayesian estimation, it is known that adding hidden variables to the problem space can lead to enhanced interaction between the observed and hidden variables, and, in turn, simplify the problem [14]. In this work, we reformulate the block-sparse recovery problem in the Bayesian framework. Typically, the SBL family of algorithms consider the unknown sparse vector as the hidden variable. Here, we introduce a set of hidden variables in order decompose the block-sparse vector recovery problem into a set of low-dimensional sparse vector recovery problems and explicitly impose the block-sparse structure of the signal. We propose a Nested SBL (NSBL) algorithm for block-sparse vector recovery, by employing the Nested Expectation Maximization (EM) [15] approach and a Kalman filtering based framework [16] for learning the intra-block correlation. In essence, NSBL is based on a divide and conquer approach: the problem of estimating a high-dimensional block sparse vector is reformulated as a set of smaller problems each involving the estimation of low-dimensional correlated group-sparse vectors.

2. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a *BM*-length block-sparse vector \mathbf{x} consisting of *B* blocks denoted by $\mathbf{b}_1, \ldots, \mathbf{b}_B$, and arranged as follows:

$$\mathbf{x} = [\underbrace{x_{11}, x_{12}, \dots, x_{1M}}_{\mathbf{b}_1^T : \mathbf{b}_1 \in \mathbb{R}^{M \times 1}}; \dots; \underbrace{x_{B1}, x_{B2}, \dots, x_{BM}}_{\mathbf{b}_P^T : \mathbf{b}_B \in \mathbb{R}^{M \times 1}}].$$
(1)

The M entries of each block \mathbf{b}_i are constrained to be either allzero or all-nonzero. In the cluster-SBL framework, the block-sparse structure is exploited by modeling $\mathbf{b}_i \sim \mathcal{N}(0, \gamma_i \mathbf{B}_i)$, where γ_i is an unknown hyperparameter such that when $\gamma_i = 0$, the i^{th} block of \mathbf{x} is zero [12]. Here, $\mathbf{B}_i \in \mathbb{R}^{M \times M}$ is a positive-definite covariance matrix that captures the intra-block correlation of the i^{th} block, which is also unknown. Moreover, different blocks are mutually uncorrelated, and hence, the block-sparse vector $\mathbf{x} \sim \mathcal{N}(0, \mathbf{\Sigma}_0)$, where $\mathbf{\Sigma}_0$ is a block-diagonal matrix with principal blocks given by $\gamma_i \mathbf{B}_i$, $1 \leq i \leq B$.

The noisy observations $\mathbf{y} \in \mathbb{R}^{N \times 1}$ are obtained as a weighted combination of the columns of a measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{N \times MB}$, as follows:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n},\tag{2}$$

where the components of the additive noise $\mathbf{n} \in \mathbb{R}^{N \times 1}$ are independent, zero mean, and Gaussian distributed: $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N)$.

Restructuring the block-sparse vector \mathbf{x} , the problem of recovering \mathbf{x} from \mathbf{y} is equivalent to finding the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$, as depicted in Fig. 1. Since $\mathbf{b}_i \sim \mathcal{N}(0, \gamma_i \mathbf{B}_i)$ for $1 \leq i \leq B$, $\mathbf{x}_i \sim \mathcal{N}(0, \Gamma)$ where $\Gamma = \text{diag}(\gamma(1), \ldots, \gamma(B))$, i.e., $\mathbf{x}_1, \ldots, \mathbf{x}_M$ represent group-sparse vectors.

This work was supported in part by the Indo-US Virtual Institute for Mathematical and Statistical Sciences, the Tata Consultancy Services Research Scholar Program, a research grant from the Indo-UK Advanced Technology Centre, and NSF grant CCF-1144258.



Fig. 1. Restructuring the block-sparse recovery problem such that the *B* length vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are *group-sparse vectors*.

By rearranging the columns of Φ , the system model in (2) can be equivalently written as

$$\mathbf{y} = \sum_{m=1}^{M} \mathbf{t}_{m}, \text{ where } \mathbf{t}_{m} \triangleq \mathbf{\Phi}_{m} \mathbf{x}_{m} + \mathbf{n}_{m}, 1 \le m \le M.$$
 (3)

In (3), $\Phi_m \in \mathbb{R}^{N \times B}$ consists of the columns of Φ such that the coefficients corresponding to its columns are given by \mathbf{x}_m . Although \mathbf{n}_m cannot be explicitly obtained, we note that its covariance can be written as $\mathbf{n}_m \sim \mathcal{N}(0, \beta_m \sigma^2 \mathbf{I}_N)$ where, $0 \leq \beta_m \leq 1$ and $\sum_{m=1}^M \beta_m = 1$. If \mathbf{t}_m is known, recovering \mathbf{x}_m from \mathbf{t}_m is a multiple measurement vector based group-sparse recovery problem [17] in a lower dimensional space (B), as compared to the dimension of the original problem (MB). In this work, we focus on recovering the block-sparse vector by recovering its group-sparse components $\mathbf{x}_1, \ldots, \mathbf{x}_M$, using the restructured problem given by (3).

In the following section, we propose two algorithms which recover \mathbf{x} when $\mathbf{B}_1 = \ldots = \mathbf{B}_M = \mathbf{B}$. In the first case, $\mathbf{B} = \mathbf{I}_B$, while in the second case, \mathbf{B} need not be a scaled identity matrix.

3. PROPOSED ALGORITHMS

In this section, we propose two algorithms for block-sparse vector recovery: (a) Parallel Cluster-SBL (PC-SBL) algorithm when the entries within a block are not correlated ($\mathbf{B} = \mathbf{I}_B$), (b) NSBL algorithm when there is nonzero intra-block correlation (\mathbf{B} need not equal \mathbf{I}_B).

The conventional SBL framework treats (\mathbf{y}, \mathbf{x}) in (2) as the complete data, and \mathbf{x} as the hidden variable. However, for the reformulated system model in (3), it is necessary to augment the set of hidden variables \mathbf{x} with $\mathbf{t} = [\mathbf{t}_1^T, \dots, \mathbf{t}_M^T]^T$ since \mathbf{t} is also hidden [18]. Accordingly, the complete information is given by $(\mathbf{y}, \mathbf{t}, \mathbf{x})$, and (\mathbf{t}, \mathbf{x}) constitute the hidden variables. Since closed-form expressions for the maximum likelihood estimates of the unknown parameter γ cannot be obtained, we adopt the iterative EM algorithm for estimating γ , as follows:

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

M-step : $\boldsymbol{\gamma}^{(r+1)} = \operatorname*{arg\,max}_{\boldsymbol{\gamma} \in \mathbb{R}^{B \times 1}_{+}} Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right).$ (4)

The E-step in (4) requires the computation of $p(\mathbf{t}, \mathbf{x} | \mathbf{y}; \boldsymbol{\gamma}^{(r)})$, which is given by

$$p(\mathbf{t}, \mathbf{x} | \mathbf{y}; \boldsymbol{\gamma}^{(r)}) = p(\mathbf{x} | \mathbf{t}, \mathbf{y}; \boldsymbol{\gamma}^{(r)}) p(\mathbf{t} | \mathbf{y}; \boldsymbol{\gamma}^{(r)})$$
$$= p(\mathbf{x} | \mathbf{t}; \boldsymbol{\gamma}^{(r)}) p(\mathbf{t} | \mathbf{y}; \boldsymbol{\gamma}^{(r)}).$$
(5)

Hence, the E-step can be rewritten as

E-step :
$$Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right) = \underbrace{\mathbb{E}_{\mathbf{t}|\mathbf{y};\boldsymbol{\gamma}^{(r)}}}_{\mathbb{E}_{\mathbf{t}}}\underbrace{\mathbb{E}_{\mathbf{x}|\mathbf{t};\boldsymbol{\gamma}^{(r)}}}_{\mathbb{E}_{\mathbf{x}}}[\log p(\mathbf{y},\mathbf{t},\mathbf{x};\boldsymbol{\gamma})].$$
 (6)

To compute $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)})$, we first compute the posterior distribution $p(\mathbf{t}|\mathbf{y};\boldsymbol{\gamma}^{(r)})$ using the likelihood $p(\mathbf{t}_m|\mathbf{x}_m) = \mathcal{N}(\boldsymbol{\Phi}_m\mathbf{x}_m, \beta_m\sigma^2\mathbf{I}_N)$ for $1 \leq m \leq M$, and the prior $p(\mathbf{x};\boldsymbol{\gamma}) = \mathcal{N}(0,\Gamma_B)$. Given $\mathbf{H} = \mathbf{1}_M \otimes \mathbf{I}_N$, where $\mathbf{1}_M$ is a M length vector of ones, and $\mathbf{y} = \mathbf{H}\mathbf{t}$, we have $p(\mathbf{t}|\mathbf{y};\boldsymbol{\gamma}^{(r)}) = \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$, where

$$\mu_{t} = (\mathbf{R} + \boldsymbol{\Phi}_{B}\boldsymbol{\Gamma}_{B}\boldsymbol{\Phi}_{B}^{T})\mathbf{H}^{T}(\mathbf{H}(\mathbf{R} + \boldsymbol{\Phi}_{B}\boldsymbol{\Gamma}_{B}\boldsymbol{\Phi}_{B}^{T})\mathbf{H}^{T})^{-1}\mathbf{y}$$

$$\boldsymbol{\Sigma}_{t} = (\mathbf{R} + \boldsymbol{\Phi}_{B}\boldsymbol{\Gamma}_{B}\boldsymbol{\Phi}_{B}^{T}) - (\mathbf{R} + \boldsymbol{\Phi}_{B}\boldsymbol{\Gamma}_{B}\boldsymbol{\Phi}_{B}^{T})\mathbf{H}^{T}$$

$$(\mathbf{H}(\mathbf{R} + \boldsymbol{\Phi}_{B}\boldsymbol{\Gamma}_{B}\boldsymbol{\Phi}_{B}^{T})\mathbf{H}^{T})^{-1}\mathbf{H}(\mathbf{R} + \boldsymbol{\Phi}_{B}\boldsymbol{\Gamma}_{B}\boldsymbol{\Phi}_{B}^{T}). \quad (7)$$

Here, $\Phi_B \in \mathbb{R}^{NM \times BM}$ is a block diagonal matrix with Φ_1, \ldots, Φ_M along the diagonal, and $\Gamma_B = \mathbf{B} \otimes \Gamma$, where $\Gamma = \text{diag}(\boldsymbol{\gamma})$. The diagonal matrix \mathbf{R} has m^{th} diagonal entry $\mathbf{R}_m = \beta_m \sigma^2 \mathbf{I}_N$. Note that the posterior mean $\boldsymbol{\mu}_t \in \mathbb{R}^{MN \times 1}$ consists of M vectors, $\boldsymbol{\mu}_{\mathbf{t}_1}, \ldots, \boldsymbol{\mu}_{\mathbf{t}_M}$ such that $\mathbf{H}\boldsymbol{\mu}_t = \mathbf{y}$, i.e., $\mathbf{y} = \sum_{m=1}^M \boldsymbol{\mu}_{\mathbf{t}_m}$. Further, the posterior distribution $p(\mathbf{x}|\mathbf{t};\boldsymbol{\gamma}^{(r)})$ depends on the correlation between the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$. In the following subsection, we provide an algorithm for block-sparse vector recovery when $\mathbf{B} = \mathbf{I}_B$.

3.1. Parallel Cluster-SBL: $\mathbf{B} = \mathbf{I}_B$

In the literature on block-sparse vector recovery, past work has mainly focused on the case where the sparse vectors are uncorrelated, i.e., $\mathbf{B} = \mathbf{I}_{B}$ [8,9]. The PC-SBL algorithm proposed in this subsection is also designed to handle such a scenario.

When $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are uncorrelated, we have $p(\mathbf{x}|\mathbf{t}; \boldsymbol{\gamma}) = \prod_{m=1}^M p(\mathbf{x}_m|\mathbf{t}_m; \boldsymbol{\gamma})$. This decomposes the block-sparse recovery problem in (2) into a multiple measurement vector problem [17], where the goal is recovering group-sparse vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$ from multiple measurements $\boldsymbol{\mu}_{\mathbf{t}_1}, \ldots, \boldsymbol{\mu}_{\mathbf{t}_M}$. The posterior distribution of \mathbf{x}_m is given by $p(\mathbf{x}_m|\mathbf{t}_m; \boldsymbol{\gamma}^{(r)}) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}_m}, \boldsymbol{\Sigma}_{\mathbf{x}_m})$, where $\boldsymbol{\mu}_{\mathbf{x}_m} = \beta_m^{-1} \sigma^{-2} \boldsymbol{\Sigma}_{\mathbf{x}_m} \boldsymbol{\Phi}_m^T \mathbf{t}_m$ and $\boldsymbol{\Sigma}_{\mathbf{x}_m} = \left(\frac{\boldsymbol{\Phi}_m^T \boldsymbol{\Phi}_m}{\beta_m \sigma^2} + \Gamma^{(r)-1}\right)^{-1}$. Using the posterior distribution computed above, the update for $\boldsymbol{\gamma}$ is obtained as follows:

$$\boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma} \in \mathbb{R}^{B \times 1}_{+}}{\arg \max} \mathbb{E}_{\mathbf{t}, \mathbf{x} | \mathbf{y}; \boldsymbol{\gamma}^{(r)}} [\log p(\mathbf{t}, \mathbf{x}; \boldsymbol{\gamma})] = \underset{\boldsymbol{\gamma} \in \mathbb{R}^{B \times 1}_{+}}{\arg \max} (c'$$
$$- \mathbb{E}_{\mathbf{t} | \mathbf{y}; \boldsymbol{\gamma}^{(r)}} \mathbb{E}_{\mathbf{x} | \mathbf{t}; \boldsymbol{\gamma}^{(r)}} [\frac{\mathbf{x}^{T} \Gamma_{B}^{-1} \mathbf{x}}{2} + \frac{1}{2} \log |\Gamma_{B}|])$$
(8)

In the above expression, $\log |\Gamma_B|$ simplifies as $M \log |\Gamma|$ and $\mathbf{x}^T \Gamma_B^{-1} \mathbf{x} = \sum_{m=1}^M \mathbf{x}_m^T \Gamma^{-1} \mathbf{x}_m$. Further, $\mathbb{E}_{\mathbf{x}|\mathbf{t};\boldsymbol{\gamma}^{(r)}}[\mathbf{x}_m^T \Gamma^{-1} \mathbf{x}_m] = \operatorname{Tr}(\Gamma^{-1}(\boldsymbol{\Sigma}_{\mathbf{x}_m} + \boldsymbol{\mu}_{\mathbf{x}_m} \boldsymbol{\mu}_{\mathbf{x}_m}^T))$. Substituting for $\boldsymbol{\mu}_{\mathbf{x}_m}$, we obtain the overall optimization problem in the M-step as

$$\boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma} \in \mathbb{R}^{B \times 1}_{+}}{\arg\min} (c' + \frac{M}{2} \log |\Gamma| + \frac{1}{2} \sum_{m=1}^{M} \operatorname{Tr}(\Gamma^{-1} \boldsymbol{\Sigma}_{\mathbf{x}_{m}})$$
$$+ \operatorname{Tr}\left(\Gamma^{-1} \frac{\boldsymbol{\Sigma}_{\mathbf{x}_{m}} \boldsymbol{\Phi}_{m}^{T} \mathbf{R}_{m} \boldsymbol{\Phi}_{m} \boldsymbol{\Sigma}_{\mathbf{x}_{m}}}{\beta_{m}^{2} \sigma^{4}}\right) \right), \tag{9}$$

where $\mathbf{R}_m = \boldsymbol{\Sigma}_{\mathbf{t}_m} + \boldsymbol{\mu}_{\mathbf{t}_m} \boldsymbol{\mu}_{\mathbf{t}_m}^T, \boldsymbol{\Sigma}_{\mathbf{t}_m} \in \mathbb{R}^{N \times N}$ is the m^{th} entry of blkdiag($\boldsymbol{\Sigma}_{\mathbf{t}}$) and blkdiag(\mathbf{A}) returns the block diagonal matrices of **A**. Maximizing (9) w.r.t. $\boldsymbol{\gamma}$, we get

$$\boldsymbol{\gamma}^{(r+1)} = \frac{1}{M} \sum_{m=1}^{M} \operatorname{diag}\left(\boldsymbol{\Sigma}_{\mathbf{x}_{m}} + \frac{\boldsymbol{\Sigma}_{\mathbf{x}_{m}} \boldsymbol{\Phi}_{m}^{T} \mathbf{R}_{m} \boldsymbol{\Phi}_{m} \boldsymbol{\Sigma}_{\mathbf{x}_{m}}}{\beta_{m}^{2} \sigma^{4}}\right).$$
(10)

The proposed PC-SBL and the cluster-SBL algorithm [12] are mathematically equivalent for the case when $\mathbf{B} = \mathbf{I}_B$. However, the PC-SBL approach allows for parallel implementation of the algorithm, since the block-sparse vector is recovered by solving Mparallel problems. Further, from (10), we see that the overall M-step is simply the average of the hyperparameter updates obtained from the M parallel problems.

A drawback of the PC-SBL algorithm is that it cannot handle the case when $\mathbf{B} \neq \mathbf{I}_B$, since the inner expectation $\mathbb{E}_{\mathbf{x}}$ does not split as M separate problems unless $\mathbf{B} = \mathbf{I}_B$. In the following subsection, we derive a novel NSBL algorithm for block-sparse vector recovery with intra-block correlation.

3.2. Nested SBL: B need not equal I_B

In this subsection, we model the intra-block correlation using a firstorder AR model, and propose an NSBL algorithm to learn the unknown parameters γ and the correlation coefficient ρ . The first order AR model is a widely accepted model, and is used in a variety of applications [19–21]. It also has the advantage that it avoids overfitting [13] and allows for a Kalman filtering based learning framework. The evolution of the m^{th} group-sparse vector is modeled as

$$\mathbf{x}_m = \rho \mathbf{x}_{m-1} + \mathbf{u}_m, \quad m = 1, \dots, M, \tag{11}$$

where the driving noise \mathbf{u}_m is distributed as $\mathbf{u}_m(i) \sim \mathcal{N}(0, (1 - \rho^2)\gamma(i)), \rho \in \mathbb{R}$ is the AR coefficient and $0 \leq \rho \leq 1$. Overall, this leads to a common correlation matrix given by $\mathbf{B}_1 = \ldots = \mathbf{B}_B = \mathbf{B} = \text{Toep}([1, \rho, \ldots, \rho^{M-1}])$, where Toep(a) represents the symmetric Toeplitz matrix defined by its first row a [13]. The state space model for \mathbf{t}_m and \mathbf{x}_m is given as

$$\mathbf{t}_m = \mathbf{\Phi}_m \mathbf{x}_m + \mathbf{n}_m,\tag{12}$$

$$\mathbf{x}_m = \rho \mathbf{x}_{m-1} + \mathbf{u}_m, \quad m = 1, \dots, M.$$
(13)

Since $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are group-sparse, from the above model, we have

$$p(\mathbf{t}, \mathbf{x}_1, \dots, \mathbf{x}_M; \boldsymbol{\gamma}) = \prod_{m=1}^M p(\mathbf{t}_m | \mathbf{x}_m) p(\mathbf{x}_m | \mathbf{x}_{m-1}; \boldsymbol{\gamma}), \quad (14)$$

where $p(\mathbf{x}_1|\mathbf{x}_0; \boldsymbol{\gamma}) \triangleq p(\mathbf{x}_1; \boldsymbol{\gamma})$. Using (14), the posterior distribution of the sparse vectors $p(\mathbf{x}_1, \dots, \mathbf{x}_M | \mathbf{t}; \boldsymbol{\gamma}^{(r)})$ is computed using the recursive Kalman Filter and Smoother (KFS) equations for $1 \leq m \leq M$ as follows [22,23]:

for
$$m = 1, \dots, M$$
 do
Prediction: $\hat{\mathbf{x}}_{m|m-1} = \rho \hat{\mathbf{x}}_{m-1|m-1}$ (15)

$$\mathbf{P}_{m|m-1} = \rho^2 \mathbf{P}_{m-1|m-1} + (1-\rho^2)\Gamma$$
(16)

Filtering:
$$\mathbf{G}_m = \mathbf{P}_{m|m-1} \boldsymbol{\Phi}_m^T \left(\sigma^2 \mathbf{I}_N + \boldsymbol{\Phi}_m \mathbf{P}_{m|m-1} \boldsymbol{\Phi}_m^T \right)^{-1}$$
(17)

$$\hat{\mathbf{x}}_{m|m} = \hat{\mathbf{x}}_{m|m-1} + \mathbf{G}_m(\mathbf{t}_m - \mathbf{\Phi}_m \hat{\mathbf{x}}_{m|m-1})$$
(18)

$$\mathbf{P}_{m|m} = (\mathbf{I}_B - \mathbf{G}_m \mathbf{\Phi}_m) \mathbf{P}_{m|m-1},$$
end
(19)

for
$$j = M, M - 1, ..., 2$$
 do

Smoothing:
$$\hat{\mathbf{x}}_{j-1|m} = \hat{\mathbf{x}}_{j-1|j-1} + \mathbf{J}_{j-1}(\hat{\mathbf{x}}_{j|m} - \hat{\mathbf{x}}_{j|j-1})$$
 (20)

$$\mathbf{P}_{j-1|m} = \mathbf{P}_{j-1|j-1} + \mathbf{J}_{j-1} (\mathbf{P}_{j|m} - \mathbf{P}_{j|j-1}) \mathbf{J}_{j-1}^{T}$$
(21)
end,



Fig. 2. Illustration of the inner and outer EM loops, which consist of $\mathbb{E}_{\mathbf{x}}$ and $\mathbb{E}_{\mathbf{t}}$, respectively.

where $\mathbf{J}_{j-1} = \rho \mathbf{P}_{j-1|j-1} \mathbf{P}_{j|j-1}^{-1}$ and \mathbf{G}_m is the Kalman gain matrix. The above mentioned KFS equations are initialized by setting $\hat{\mathbf{x}}_{0|0} = \mathbf{0}$, i.e., a zero vector, and $\mathbf{P}_{0|0} = \Gamma$. The E-step requires the computation of $\mathbb{E}_{\mathbf{x}_1,...,\mathbf{x}_M|\mathbf{t};\boldsymbol{\gamma}^{(r)}}[\mathbf{x}_j\mathbf{x}_{j-1}^T] \triangleq \mathbf{P}_{j,j-1|m} + \hat{\mathbf{x}}_{j|m}\hat{\mathbf{x}}_{j-1|m}^T$ for $m = M, M-1, \ldots, 2$, which we obtain from [22] as follows:

$$\mathbf{P}_{j-1,j-2|m} = \mathbf{P}_{j-1|j-1} \mathbf{J}_{j-2}^{T} + \mathbf{J}_{j-1}^{T} (\mathbf{P}_{j,j-1|m} - \rho \mathbf{P}_{j-1|j-1}) \mathbf{J}_{j-2}.$$
(22)

The above recursion is initialized using $\mathbf{P}_{m,m-1|M} = \rho(\mathbf{I}_B - \mathbf{G}_m \mathbf{\Phi}_m) \mathbf{P}_{m-1|m-1}$. Note that $\mathbf{x}_{i|M}$ and $\mathbf{P}_{i|M}$, $1 \le i \le M$ represent the posterior mean and covariance of \mathbf{x} given \mathbf{t} , respectively.

The expectation $\mathbb{E}_{\mathbf{t}}$ involves computing $p(\mathbf{t}|\mathbf{y}; \boldsymbol{\gamma}^{(r)})$ using (7). As mentioned earlier, Γ_B is given by $\Gamma = \mathbf{B} \otimes \Gamma$. The KFS equations in (15)-(23) constitute the $\mathbb{E}_{\mathbf{x}}$ step, after which we compute $\mathbb{E}_{\mathbf{t}}$. However, due to the recursive nature of the inner E-step, $\mathbb{E}_{\mathbf{x}}$, the expectation of $\boldsymbol{\mu}_{\mathbf{x}_m}$ w.r.t. the posterior density of \mathbf{t} is a recursive function of $\mathbf{t}_m, \ldots, \mathbf{t}_1$. As M increases, the complexity of computing such a recursive expectation becomes prohibitive. In order to circumvent this problem, we employ an alternate technique, known as the Nested EM approach [15]. This monotonically convergent approach allows us to simplify the overall algorithm into an inner and outer EM loop, while the unknown parameter $\boldsymbol{\gamma}$ is the common factor between the two loops. We call this algorithm as the NSBL algorithm, where the nested E- and M-steps are given as

$$E-step: Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r+\frac{k}{K})}, \boldsymbol{\gamma}^{(r)}\right) \\ = \mathbb{E}_{\mathbf{t}|\mathbf{y};\boldsymbol{\gamma}^{(r)}} \left[\mathbb{E}_{\mathbf{x}_{1},\ldots,\mathbf{x}_{M}|\mathbf{t};\boldsymbol{\gamma}^{(r+\frac{k}{K})}} \left[\log p(\mathbf{y},\mathbf{t},\mathbf{x}_{1},\ldots,\mathbf{x}_{M};\boldsymbol{\gamma})\right] \right] \\ M-step: \boldsymbol{\gamma}^{(r+\frac{k+1}{K})} = \underset{\boldsymbol{\gamma}\in\mathbb{R}^{H\times 1}_{+}}{\arg\max} Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r+\frac{k}{K})}, \boldsymbol{\gamma}^{(r)}\right).$$
(23)

The inner EM loop is initialized by $\gamma^{(r+\frac{R}{K})} = \gamma^{(r)}$. Note that, when γ is updated in every iteration, only the inner E-step ($\mathbb{E}_{\mathbf{x}} = \mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_M|\mathbf{t};\gamma^{(r+\frac{R}{K})}}[\cdot]$) is updated. The overall NSBL algorithm is executed by nesting one EM loop within the other, as depicted in Fig. 2. The inner EM loop consists of $\mathbb{E}_{\mathbf{x}}$ and the corresponding posterior distribution is given by (15)-(19). Further, the M-step for the inner EM loop is given by [16]

$$\gamma^{(r+\frac{(k+1)}{K})}(i) = \frac{1}{M} \left(\sum_{j=2}^{M} \frac{\mathbf{M}_{j|M}(i,i)}{(1-\rho^2)} + \mathbf{M}_{1|M}(i,i) \right)$$
(24)

for $1 \leq i \leq B$, where $\mathbf{M}_{j|M} \triangleq \mathbf{P}_{j|M} + \hat{\mathbf{x}}_{j|M} \hat{\mathbf{x}}_{j|M}^T + \rho^2 (\mathbf{P}_{j-1|M} + \hat{\mathbf{x}}_{j-1|M} \hat{\mathbf{x}}_{j-1|M}^T) - 2\rho (\mathbf{P}_{j,j-1|M} + \hat{\mathbf{x}}_{j|M} \hat{\mathbf{x}}_{j-1|M}^T)$ and $\mathbf{M}_{1|M} \triangleq$



Fig. 3. MSE of the proposed algorithm as compared to cluster-SBL (CSBL), least-squares (Oracle estimator) and SBL, with M = 8, B = 32 and $\rho = 0.8$.

 $\mathbf{P}_{1|M} + \hat{\mathbf{x}}_{1|M} \hat{\mathbf{x}}_{1|M}^T$. After K iterations of the inner EM loop, we obtain $\gamma^{r+\frac{K}{K}} = \gamma^{r+1}$, which affects the posterior distribution of **t**. The outer EM loop consists of updating the posterior distribution of **t** given in (7).

An update step for the unknown correlation coefficient ρ can also be incorporated into the M-step of the NSBL algorithm. The correlation coefficient $\rho^{(r+\frac{(k+1)}{K})}$ in the $(r+\frac{(k+1)}{K})^{\text{th}}$ iteration is obtained as a solution to the cubic equation

$$(2B(M-1))\rho^{3} + \operatorname{Tr} \{\mathbf{T}_{2} + \mathbf{T}_{3}\}\rho^{2} - \operatorname{Tr} \{\mathbf{T}_{2} + \mathbf{T}_{3}\} - [2B(M-1) - 2\operatorname{Tr} \{\mathbf{T}_{1} + \mathbf{T}_{4}\}]\rho = 0, \quad (25)$$

where the matrices T_1 through T_4 are defined as

$$\mathbf{T}_1 = \Gamma^{-1} \sum_{j=2}^{M} [\mathbf{P}_{j|M} + \hat{\mathbf{x}}_{j|M} \hat{\mathbf{x}}_{j|M}^T], \qquad (26)$$

$$\mathbf{T}_{2} = \Gamma^{-1} \sum_{j=2}^{M} [\mathbf{P}_{j,j-1|M} + \hat{\mathbf{x}}_{j|M} \hat{\mathbf{x}}_{j-1|M}^{T}], \qquad (27)$$

$$\mathbf{T}_{3} = \Gamma^{-1} \sum_{j=2}^{M} [\mathbf{P}_{j,j-1|M} + \hat{\mathbf{x}}_{j-1|M} \hat{\mathbf{x}}_{j|M}^{T}], \qquad (28)$$

$$\mathbf{\Gamma}_4 = \Gamma^{-1} \sum_{j=2}^{M} [\mathbf{P}_{j-1|M} + \hat{\mathbf{x}}_{j-1|M} \hat{\mathbf{x}}_{j-1|M}^T].$$
(29)

Among the possible solutions of the above cubic equation, we pick the $\rho \in \mathbb{R}$ that satisfies $0 \le \rho \le 1$.

Using a flop-count analysis [24], we note that the computations in cluster-SBL are dominated by the E-step, which incurs a computational complexity of $\mathcal{O}(N^2MB)$. On the other hand, NSBL consists of two EM loops, where the maximum complexity of the outer and inner EM loop are $\mathcal{O}(N^2MB)$ and $\mathcal{O}(MB^3)$, respectively. Typically, in the nested EM approach, the number of inner EM iterations are fixed, so that the outer EM loop is guaranteed to converge, and the inner EM loop ensures likelihood increase [15]. Consequently, the complexity of the NSBL algorithm is dominated by $\mathcal{O}(N^2MB)$. However, since the number of outer EM iterations are far lower than that of cluster-SBL, the NSBL entails a lower computational complexity than the cluster-SBL approach.

In the following section, we demonstrate the efficacy of the proposed algorithms using Monte Carlo simulations.

4. SIMULATION RESULTS

The experimental set-up used to evaluate the Mean Square Error (MSE) and the support recovery performance of the proposed algorithms is as follows. We consider a block-sparse vector of length 256 consisting of B = 32 blocks of length M = 8 each, with 5 nonzero blocks. We set the value of $\beta_m = 1/M$. In each trial, the



Fig. 4. Success rates of the NSBL and CSBL algorithms, with M = 8, B = 32 and $\rho = 0.99$ (left) and 0.8 (right).

matrix Φ is generated as a random underdetermined (N < MB) measurement matrix, whose entries are *i.i.d.* and standard Bernoulli ({+1, -1}) distributed. For fair comparison, we fix the number of iterations of the cluster-SBL [12] to 125, and the number of inner and outer EM loop iterations of the NSBL algorithm to 5 and 25, respectively. The outcome of the experiment is averaged over 1,000 trials.

In Fig. 3, we compare the MSE performance of the proposed NSBL algorithm with the cluster-SBL [12], SBL and the oracle least-squares estimator, i.e., the least squares estimator which is aware of the support of \mathbf{x} . We see that the MSE performance of NSBL algorithm is 2 dB better than the CSBL algorithm, while being very close to the MSE performance of the oracle estimator. A similar trend is observed at different Signal to Noise Ratios (SNR) and N. The SBL algorithm fails to recover the block-sparse vector for small values of N and SNR, which demonstrates the advantage of exploiting the block-sparse structure.

In Fig. 4, we plot the support recovery performance of the NSBL algorithm and the cluster-SBL algorithm at high SNR of 80dB for $\rho = 0.8$ and 0.99. We see that NSBL algorithm has a better support recovery performance at smaller values of N, even with ρ being learnt by the algorithm. However, as N increases, the two algorithms have similar performance. In the cluster-SBL approach, we use the heuristic algorithm to estimate ρ in the unknown- ρ case [12]. We see that the performance degradation due to the learning of ρ is marginal, which makes the NSBL algorithm particularly attractive for practical implementations.

5. CONCLUSIONS

In this work, we proposed novel algorithms for the recovery of block-sparse vectors with intra-block correlation from underdetermined noisy linear measurements. First, we reformulated the blocksparse vector recovery problem as a group-sparse vector recovery problem by introducing hidden variables. Using the reformulated framework, we proposed the PC-SBL algorithm for the scenario when the nonzero blocks of the block-sparse vector has uncorrelated entries. We showed that, unlike the cluster-SBL algorithm, the proposed PC-SBL approach allows for parallel implementation. Next, we proposed the NSBL algorithm for the case when the entries of a non-zero block are correlated. In contrast to the cluster-SBL approach, we were able to provide closed-form EM updates for estimating the correlation coefficient. Using simulations, we showed that the NSBL algorithm offers nearly the same performance as the oracle estimator, with respect to MSE, and an improved support recovery performance compared to the cluster-SBL approach.

6. REFERENCES

- 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] 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.
- [4] S. Babacan, R. Molina, and A. Katsaggelos, "Bayesian Compressive Sensing using Laplace Priors," *IEEE Trans. Image Process.*, vol. 19, pp. 53–64, 2010.
- [5] S. Ji, Y. Xue, and L. Carin, "Bayesian Compressive Sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [6] A. Juditsky, F. K. Karzan, A. Nemirovski, and B. Polyak, "On the accuracy of 11-filtering of signals with block-sparse structure," *Advances in NIPS*, vol. 24, 2011.
- [7] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ecg via block sparse bayesian learning," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 300–309, 2013.
- [8] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 57, no. 8, pp. 3075–3085, 2009.
- [9] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.
- [10] S. Negahban and M. J. Wainwright, "Joint support recovery under high-dimensional scaling: Benefits and perils of $\ell_1 \ell_{\infty}$ -regularization," *Advances in NIPS*, vol. 21, pp. 1161–1168, 2008.
- [11] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
- [12] Z. Zhang and B. D. Rao, "Recovery of block sparse signals using the framework of block sparse bayesian learning," in *Proc. ICASSP*. IEEE, 2012, pp. 3345–3348.
- [13] 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.
- [14] G. Elidan, N. Lotner, N. Friedman, D. Koller *et al.*, "Discovering hidden variables: A structure-based approach," in *Advances in NIPS*, vol. 13, 2000, pp. 479–485.
- [15] D. A. Van Dyk, "Nesting EM algorithms for computational efficiency," *Statistica Sinica*, vol. 10, no. 1, pp. 203–226, 2000.
- [16] R. Prasad, C. R. Murthy, B. D. Rao, "Joint Approximately Sparse Channel Estimation and Data Detection in OFDM Systems using Sparse Bayesian Learning," *Submitted to IEEE Trans. Signal Process.*, 2013.
- [17] D. Wipf and B. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, 2007.

- [18] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust.*, *Speech, Signal Process.*, vol. 36, no. 4, pp. 477–489, 1988.
- [19] H. Akaike, "Fitting autoregressive models for prediction," Annals of the institute of Statistical Mathematics, vol. 21, no. 1, pp. 243–247, 1969.
- [20] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, 1999.
- [21] W. Wei, *Time series analysis*. Addison-Wesley Redwood City, California, 1994.
- [22] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Tech. Rep., 1996.
- [23] B. Anderson and J. Moore, *Optimal filtering*. Courier Dover Publications, 2005.
- [24] R. Hunger, "Floating point operations in matrix-vector calculus," *Munich University of Technology*, TUM-LNS-TR-05-05, 2005.