Sparse Recovery from Multiple Measurement Vectors Using Exponentiated Gradient Updates

Saurabh Khanna, *Student Member, IEEE* and Chandra R. Murthy, *Senior Member, IEEE*

***Abstract*—In this paper, we address the problem of reconstructing the common nonzero support of multiple joint sparse vectors from their noisy and underdetermined linear measurements. The support recovery problem is formulated as the selection of nonnegative hyperparameters of a correlation-aware, joint sparsity inducing Gaussian prior. The hyperparameters are then recovered as a nonnegative sparse solution of covariance matching constraints formulated in the observation space by solving a sequence of proximal regularized convex optimization problems. For proximal regularization based on Von Neumann Bregman matrix divergence, an exponentiated gradient (EG) update is proposed, which when applied iteratively, converges to hyperparameters with the correct sparse support. Compared to existing MMV support recovery algorithms, the proposed multiplicative EG update has a significantly lower computational and storage complexity and takes fewer iterations to converge. We empirically demonstrate that the support recovery algorithm based on the proposed EG update can solve million variable support recovery problems in tens of seconds. Additionally, by leveraging its correlation-awareness property, the proposed algorithm can recover supports of size as high as $O(m^2)$ from only m linear measurements per joint sparse vector.**

***Index Terms*—Compressive Sensing, Sparse Recovery, Joint Sparsity, Multiple Measurement Vectors, Covariance Matching, Von Neumann Divergence, Exponentiated Gradient Updates.**

I. Introduction

A canonical problem in multi-sensor signal processing is the reconstruction of the common nonzero support of a set of joint sparse[1] vectors from their noisy and underdetermined linear measurements. Traditionally referred to as the *MMV support recovery problem*, it concerns locating the nonzero rows of a row sparse matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L] \in \mathbb{R}^{n \times L}$, where the k-sparse vector $\mathbf{x}_j$ denotes the j^{th} column of $\mathbf{X}$. Since $\mathbf{X}$ is a row-sparse matrix, it implies that the columns of $\mathbf{X}$ are *jointly sparse*, i.e., they share a common nonzero support. Let us denote the common support of the columns in $\mathbf{X}$ by the index set $\mathcal{S} \in [n]$ where $|\mathcal{S}| = k$. Then, the goal of the MMV support recovery problem is to recover the unknown support set $\mathcal{S}$ from the noisy and compressive linear measurements $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_L] \in \mathbb{R}^{m \times L}$ generated as per the linear measurement model

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}. \quad (1)$$

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India (E-mail: saurabh@iisc.ac.in; cmurthy@iisc.ac.in).

[1] A set of vectors are said to be *jointly sparse* if their nonzero coefficients belong to the same set of rows, and the number of nonzero rows is small compared to the ambient signal dimension.

The measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ in (1) is assumed to be known, and $\mathbf{W} \in \mathbb{R}^{m \times L}$ models the additive noise in the measurements. The noise matrix $\mathbf{W}$ is assumed to be entrywise i.i.d Gaussian with zero mean and known variance σ^2. The columns of $\mathbf{Y}$ are collectively called multiple measurement vectors or MMVs.

Joint sparsity offers a compelling way to model the inherent mutual structure within a given set of multiple signal vectors [1], [2]. In various multi-sensor configurations, the high-resolution data generated by the multiple sensors tends to exhibit joint sparsity under a suitable basis transformation. This can be attributed to overlapping of the signal subspaces perceived by the different sensors sensing the same physical process. In such scenarios, identifying the true low-dimensional and common signal subspace of the compressively acquired high-resolution multi-sensor data can be formulated as the canonical MMV support recovery problem. For instance, in a cognitive radio network, the estimation of sparse spectral occupancy of the licensed users as perceived by the secondary users in the cellular network using wideband compressive measurements can be cast as the MMV support recovery problem [3]. The support recovery problem also arises in distributed source coding [1], event detection/localization in wireless sensor networks [4], and array signal processing [5].

The conventional MMV algorithms for sparse support reconstruction rely on techniques that include (i) solution space regularization using joint-sparsity-inducing convex and nonconvex penalties (Row-LASSO [6] and CRL-1 [7]), (ii) greedy support reconstruction (SOMP [8] and SCo-SAMP [9]), (iii) iterative hard thresholding (SIHT) [9] and (iv) MuSiC criterion (CS-MuSiC [10] and SA-MuSiC [11]). These conventional techniques implicitly assume that the support size, k, is smaller than m, the number of linear measurements per MMV. In MMV literature, this upper limit on maximum size of recoverable support is known as the ℓ_0-bound. This work focuses on a recently proposed *correlation-aware covariance matching* framework which can recover supports of size significantly larger the ℓ_0-bound, by exploiting a latent correlation structure in the joint sparse columns of $\mathbf{X}$. In particular, this framework has been empirically [12], [13] as well as theoretically [14], [15] validated to be capable of recovering supports of size as high as $\mathcal{O}(m^2)$ from m linear measurements per MMV.

We propose a novel MMV support recovery algorithm which is inspired by Co-LASSO [12], an existing covariance matching algorithm. The proposed algorithm, Co-LASSO-EXPGRD, iteratively applies exponentiated gradient (EG) updates to recover the shared nonzero support of the joint sparse columns of $\mathbf{X}$ as the nonnegative hyperparameters of a correlation-aware, joint sparsity inducing Gaussian prior. The proposed EG update is empirically shown to converge significantly faster than the parent Co-LASSO algorithm and also requires significantly lower storage space for execution. The proposed Co-LASSO-EXPGRD is an ideal candidate for solving very large scale support recovery problems which involve millions of variables.

The rest of the paper is organized as follows. In Section II, we introduce the concept of correlation-aware covariance matching principle and discuss the Co-LASSO algorithm, which serves as the starting point for developing the proposed algorithm. In Sections III and IV, we give an overview of the EG updates for the nonnegative parameter estimation, and in Section V, we adapt these EG updates to solve the Co-LASSO's optimization problem. In Section VI, we present the results of the numerical experiments showcasing the support recovery performance of the proposed algorithm. Final concluding remarks are made in Section VII.

II. Correlation-Aware Support Recovery via Covariance Matching

The key idea behind the correlation-aware covariance matching framework [12] for sparse support recovery is to assume that *the nonzero elements in different rows of the unknown sparse ensemble are uncorrelated*. For the observation model described in (1), this assumption translates into latent uncorrelatedness of the nonzero elements belonging to the individual columns of $\mathbf{X}$, and it can be enforced by using a suitable correlation aware prior. In [12], [16], the authors impose a common Gaussian prior on the unknown joint sparse columns of $\mathbf{X}$ as shown below.

$$\mathbf{x}_j \overset{i.i.d}{\sim} \mathcal{N}(0, \mathbf{\Gamma}), \;\; j \in [L], \tag{2}$$

where $\mathbf{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$ and $\boldsymbol{\gamma} \in \mathbb{R}^n_+$. Since $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$ are zero mean and share a common covariance matrix, their posterior means also share a common support, which in turn equals the support of $\boldsymbol{\gamma}$. Owing to the diagonal nature of the covariance matrix $\mathbf{\Gamma}$, the above Gaussian prior is deemed *correlation-aware* in the sense that it naturally captures the lack of intra-vector correlation in the individual columns of $\mathbf{X}$. Additionally, from the linearity of the measurement model in (1), it follows that the MMVs are Gaussian distributed as

$$\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m + \mathbf{A}\mathbf{\Gamma}\mathbf{A}^T), \;\; j \in [L], \tag{3}$$

where $\sigma^2\mathbf{I}_m$ denotes the known covariance matrix of the additive white measurement noise. Under these assumptions, the row-support of $\mathbf{X}$ can now be recovered as $\texttt{support}(\hat{\boldsymbol{\gamma}})$, where $\hat{\boldsymbol{\gamma}}$ is a nonnegative sparse solution of the following covariance matching problem.

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma} \succeq 0}{\mathrm{argmin}} \; ||\boldsymbol{\gamma}||_1 \text{ subject to } \left|\left|\left|\mathbf{R}_{\mathbf{yy}} - \mathbf{A}\mathbf{\Gamma}\mathbf{A}^T\right|\right|\right|_F^2 \leq \epsilon. \tag{4}$$

The ℓ_1-norm objective in (4) promotes sparsity in $\hat{\boldsymbol{\gamma}}$, while simultaneously the constraints on $\boldsymbol{\gamma}$ demand a good fit between the parameterized covariance matrix $\mathbf{A}\hat{\mathbf{\Gamma}}\mathbf{A}^T$ and the sample covariance matrix $\mathbf{R}_{\mathbf{yy}} \triangleq \frac{1}{L}\mathbf{Y}\mathbf{Y}^T$. The positive constant ϵ introduces slackness in the covariance matching constraints to account for the modeling errors arising from the presence of measurement noise, the Gaussian source assumption, and the use of a finite sample estimate of $\mathbf{R}_{\mathbf{yy}}$ in place of true MMV covariance $\mathbb{E}\left[\mathbf{y}_j\mathbf{y}_j^T\right]$. An equivalent but unconstrained version of the constrained optimization in (4) has been investigated in [12] as the following nonnegative LASSO problem.

$$\text{Co-LASSO: } \hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma} \in \mathbb{R}^n_+}{\mathrm{argmin}} \frac{1}{2}\left|\left|\mathrm{vec}(\mathbf{R}_{\mathbf{yy}}) - (\mathbf{A} \odot \mathbf{A})\boldsymbol{\gamma}\right|\right|_2^2 + \lambda||\boldsymbol{\gamma}||_1, \tag{5}$$

where $\lambda > 0$ is the regularization parameter and $\mathbf{A} \odot \mathbf{A}$ denotes the columnwise self Khatri-Rao product of the measurement matrix $\mathbf{A}$ with itself. The $m^2 \times n$ sized Khatri-Rao product $\mathbf{A} \odot \mathbf{A}$ is evaluated as the columnwise Kronecker product and it arises as a consequence of vectorization of the covariance matching constraint: $\mathbf{R}_{\mathbf{yy}} \approx \mathbf{A}\mathbf{\Gamma}\mathbf{A}^T$. Once $\hat{\boldsymbol{\gamma}}$ is found, $\texttt{support}(\hat{\boldsymbol{\gamma}})$ is declared as the estimated row-support of $\mathbf{X}$.

Similar to Co-LASSO, MSBL [16] is another popular MMV algorithm which also employs the Gaussian prior in (2). In MSBL, the hyperparameter vector $\boldsymbol{\gamma}$ is estimated via covariance matching by seeking to minimize the gap between the empirical $\mathbf{R}_{\mathbf{yy}}$ and the parameterized $\mathbf{A}\mathbf{\Gamma}\mathbf{A}^T$ in the LogDet Bregman matrix divergence sense [15]. In [12], [14], [15], it is shown that both MSBL and Co-LASSO can perfectly recover supports of size as high as $\mathcal{O}(m^2)$ from m linear measurements per MMV. The ℓ_0-bound breaching support recovery performance of the covariance matching based algorithms instigates our interest in developing their fast variants for practical applications.

In the following sections, we develop a new type of support recovery algorithm which employs a multiplicative update in $\boldsymbol{\gamma}$ to find the nonnegative sparse solution $\hat{\boldsymbol{\gamma}}$ of Co-LASSO's optimization problem in (5). The new algorithm retains the attractive ℓ_0-bound breaching performance of Co-LASSO, while simultaneously requires significantly lesser storage resources and takes fewer iterations to converge.

III. Exponentiated Gradient Updates for Nonnegative Parameter Estimation

We now discuss the matrix exponentiated gradient (MEG) updates, first introduced in [17] for learning symmetric positive definite matrices. In the next section, we will adapt these MEG updates to find the sparse nonnegative hyperparameter vector $\hat{\boldsymbol{\gamma}}$ which solves the Co-LASSO optimization in (5).

Consider a batch learning setting wherein the goal is to learn a positive definite matrix $\mathbf{W}^*$, which minimizes a real valued loss function: $L : S^n_{++} \rightarrow \mathbb{R}^+$, i.e.,

$$\mathbf{W}^* = \underset{\mathbf{W} \in S^n_{++}}{\arg\min} \; L(\mathbf{W}). \tag{6}$$

We assume that the loss $L(\mathbf{W})$ is convex in $\mathbf{W}$, and that the gradient $\nabla_{\mathbf{W}} L(\mathbf{W})$ is well defined. As motivated in [17], the optimal $\mathbf{W}^*$ can be found by solving the following sequence of convex optimization problems,

$$\mathbf{W}_{t+1} = \underset{\mathbf{W}}{\mathrm{argmin}} \; \mathcal{D}_F(\mathbf{W}, \mathbf{W}_t) + \eta L(\mathbf{W}), \tag{7}$$

where $\mathcal{D}_F(\mathbf{W}, \mathbf{W}_t)$ denotes the Bregman matrix divergence between positive definite matrices $\mathbf{W}$ and $\mathbf{W}_t$. For a given convex seed function $F : S^n_{++} \rightarrow \mathbb{R}$, the Bregman matrix divergence $\mathcal{D}_F(\mathbf{W}, \tilde{\mathbf{W}})$ between any two positive definite matrices $\mathbf{W}$ and $\tilde{\mathbf{W}}$ is defined as

$$\mathcal{D}_F(\mathbf{W}, \tilde{\mathbf{W}}) \triangleq F(\mathbf{W}) - F(\tilde{\mathbf{W}}) - \mathrm{tr}\left(f(\tilde{\mathbf{W}})^T(\mathbf{W} - \tilde{\mathbf{W}})\right), \quad (8)$$

where $f(\tilde{\mathbf{W}}) = \nabla F(\tilde{\mathbf{W}})$ is the first order derivative of the matrix function F evaluated at $\tilde{\mathbf{W}}$.

The matrix update in (7) serves the dual purpose of achieving a small value for the loss function while simultaneously ensuring that the new parameter $\mathbf{W}_{t+1}$ stays close to the old parameter $\mathbf{W}_t$. The *learning rate* parameter $\eta > 0$ controls the trade-off between these two conflicting goals.

A. Matrix Exponentiated Gradient (MEG) Updates

Since both $\mathcal{D}_F(\mathbf{W}, \tilde{\mathbf{W}})$ and $L(\mathbf{W})$ are convex functions of $\mathbf{W}$, the $\arg\min$ in (7) can be eliminated by setting $\mathbf{W}_{t+1}$ such that the gradient of the objective, $\mathcal{D}_F(\mathbf{W}, \tilde{\mathbf{W}}) + \eta L(\mathbf{W})$, with respect to $\mathbf{W}$ vanishes, i.e.,

$$\mathbf{W}_{t+1} = f^{-1}\left(f(\mathbf{W}_t) - \eta \nabla L(\mathbf{W}_{t+1})\right). \quad (9)$$

Here we assume that the matrix functions f, f^{-1} and ∇L, all three preserve the symmetry and positive semidefiniteness of their respective outputs. It is usually difficult to obtain a closed form expression for $\mathbf{W}_{t+1}$ from the zero gradient condition in (9). In [18], Kivinen and Warmuth proposed a way to circumvent this issue by approximating $\nabla L(\mathbf{W}_{t+1})$ by $\nabla L(\mathbf{W}_t)$, leading to the following explicit update rule.

$$\mathbf{W}_{t+1} = f^{-1}\left(f(\mathbf{W}_t) - \eta \nabla L(\mathbf{W}_t)\right). \quad (10)$$

A choice of the seed function F leads to multiplicative updates of the exponentiated gradient (EG) form. As motivated in [17], by choosing F as the *Von Neumann entropy*, i.e., $F(\mathbf{W}) = \mathrm{tr}\left(\mathbf{W}\log\mathbf{W} - \mathbf{W}\right)$, the proximal function $\mathcal{D}_F$ becomes the Von Neumann Bregman matrix divergence and the update in (10) takes the following multiplicative form.

$$\mathbf{W}_{t+1} = \exp\left(\log\mathbf{W}_t - \eta \nabla L(\mathbf{W}_t)\right). \quad (11)$$

The above matrix update[2] preserves the symmetry and positive definite nature of the iterates $\mathbf{W}_t$ for all $t \geq 1$, provided that $\mathbf{W}_0$ is initialized as a symmetric positive definite matrix. For an in-depth analysis of how the loss function $L(\mathbf{W})$ decays with the number of iterations of the Von Neumann update in (11), the reader is referred to the excellent exposition in [17].

IV. The Proposed Co-LASSO-EXPGRD Algorithm

Using the matrix EG updates described in Section III-A, we now develop Co-LASSO-EXPGRD, an iterative algorithm to find a sparse nonnegative solution of the Co-LASSO optimization problem in (5). According to the correlation-aware covariance matching framework for support recovery, we seek a diagonal positive semidefinite matrix $\mathbf{\Gamma}$ that minimizes the below convex loss function $L(\mathbf{\Gamma})$,

$$L(\mathbf{\Gamma}) \triangleq \frac{1}{2}\left|\left|\left|\mathbf{R}_{\mathbf{yy}} - \mathbf{A}\mathbf{\Gamma}\mathbf{A}^T\right|\right|\right|_F^2 + \lambda\left|\left|\boldsymbol{\gamma}\right|\right|_1. \quad (12)$$

[2]In (11), the matrix functions $\exp\mathbf{A}$ and $\log\mathbf{A}$ are evaluated as $\exp\mathbf{A} \triangleq \mathbf{U}\mathrm{diag}\left(e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n}\right)\mathbf{U}^T$ and $\log\mathbf{A} \triangleq \mathbf{U}\mathrm{diag}\left(\log\lambda_1, \log\lambda_2, \ldots, \log\lambda_n\right)\mathbf{U}^T$, respectively for any $n \times n$ sized positive definite matrix $\mathbf{A}$ admitting the eigenvalue decomposition $\mathbf{A} = \mathbf{U}\mathrm{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_n\right)\mathbf{U}^T$.

Under the assumption that the iterates $\boldsymbol{\gamma}_t \succeq 0$, a consequence of the nonnegativity preserving updates, the gradient of $L(\mathbf{\Gamma})$ with respect to $\boldsymbol{\gamma}_i$ can be evaluated as

$$\nabla_{\boldsymbol{\gamma}_i} L(\mathbf{\Gamma}) = \mathbf{a}_i^T\left(\mathbf{A}\mathbf{\Gamma}\mathbf{A}^T - \mathbf{R}_{\mathbf{yy}}\right)\mathbf{a}_i + \lambda, \;\; \forall i \in [n], \quad (13)$$

where $\mathbf{a}_i$ denotes the i^{th} column of the measurement matrix $\mathbf{A}$. For $F(\mathbf{\Gamma}) = \mathrm{tr}\left(\mathbf{\Gamma}\log\mathbf{\Gamma} - \mathbf{\Gamma}\right)$, by using the above expression for $\nabla_{\boldsymbol{\gamma}_i} L(\mathbf{\Gamma})$ in (11), we obtain the following multiplicative elementwise update for $\boldsymbol{\gamma}$,

$$\boldsymbol{\gamma}_{t+1}(i) = \boldsymbol{\gamma}_t(i) e^{-\eta\left[\mathbf{a}_i^T\left(\mathbf{A}\mathbf{\Gamma}_t\mathbf{A}^T - \mathbf{R}_{\mathbf{yy}}\right)\mathbf{a}_i + \lambda\right]}, \;\; i \in [n]. \quad (14)$$

Since the objective $L(\Gamma)$ contains an ℓ_1 regularization term, the multiplicative updates in (14) converge to a nonnegative sparse vector $\hat{\boldsymbol{\gamma}}$, and $\texttt{support}(\hat{\boldsymbol{\gamma}})$ is declared as an estimate of the row-support of $\mathbf{X}$. In practice, only those coefficients of $\hat{\boldsymbol{\gamma}}$ whose magnitude exceeds a carefully chosen threshold value are deemed as active and determine the nonzero support of $\hat{\boldsymbol{\gamma}}$. Finally, we summarize Co-LASSO-EXPGRD in Algorithm 1.

Algorithm 1: Co-LASSO-EXPGRD

Input: $\{\mathbf{y}_j\}_{j=1}^L$ and $\mathbf{A}$.
Initializations: $\boldsymbol{\gamma}_0(i) \leftarrow 1 \;\; \forall i \in [n]$, $t \leftarrow 0$, $\Delta = 1$
while $(t < t_{max})$ *and* $\left(\Delta > 10^{-6}\right)$ **do**
1. $\boldsymbol{\gamma}_{t+1}(i) = \boldsymbol{\gamma}_t(i) e^{-\eta\left[\mathbf{a}_i^T\left(\mathbf{A}\mathbf{\Gamma}_t\mathbf{A}^T - \mathbf{R}_{\mathbf{yy}}\right)\mathbf{a}_i + \lambda\right]}$, $i \in [n]$.
2. $\Delta \leftarrow ||\boldsymbol{\gamma}_{t+1} - \boldsymbol{\gamma}_t||_2 / ||\boldsymbol{\gamma}_t||_2$ and $t \leftarrow t + 1$.

end
Output: $\hat{\boldsymbol{\gamma}} \leftarrow \boldsymbol{\gamma}_{t+1}$, and $\hat{\mathcal{S}} \leftarrow \{i \in [n] : \hat{\boldsymbol{\gamma}}(i) \geq \tau\}$.

In Table I, we compare the computational and storage complexity of Co-LASSO-EXPGRD with the existing algorithms, MSBL [16] and Co-LASSO [12]. MSBL has the highest per-iteration computational complexity and combined with the slow convergence rate of its underlying Expectation-Maximization (EM) iterations, it quickly becomes impractical for large signal dimensions. The Co-LASSO algorithm, on the other hand, has the highest storage complexity, as it needs to store an $m^2 \times n$ sized Khatri-Rao product matrix $\mathbf{A} \odot \mathbf{A}$, which quickly becomes prohibitive for larger values of m and n. In contrast, the proposed Co-LASSO-EXPGRD algorithm based on the EG update in (14) has the lowest storage complexity which scales conveniently with the problem dimensions. Co-LASSO-EXPGRD and its parent algorithm, Co-LASSO, have identical per-iteration computational complexities which interestingly do not depends on L. However, in practice, Co-LASSO-EXPGRD converges in significantly fewer iterations.

V. Numerical Experiments

In this section, we benchmark the support recovery performance of Co-LASSO-EXPGRD against the parent Co-LASSO

TABLE I: Computational & Storage Complexity Comparison

Algorithm	Per iteration computational complexity	Storage complexity
MSBL	$\mathcal{O}(m^3 + nm^2 + n^2L)$	$\mathcal{O}(nm + mL + nL)$
Co-LASSO	$\mathcal{O}(nm^2)$	$\mathcal{O}(nm^2 + mL + m^2)$
Co-LASSO-EXPGRD	$\mathcal{O}(nm^2)$	$\mathcal{O}(nm + mL + m^2)$

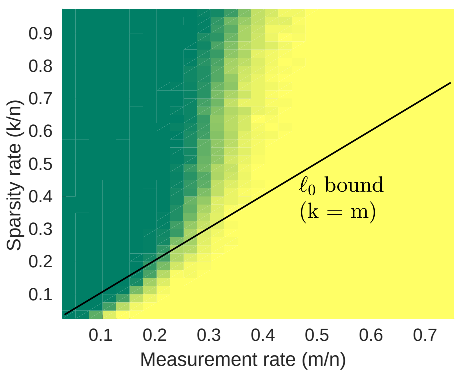

Fig. 1: Support recovery phase transition for the proposed Von-Neumann updates. Other simulation parameters: $n = 200$, $L = 400$, SNR = 20 dB, number of trials = 100.

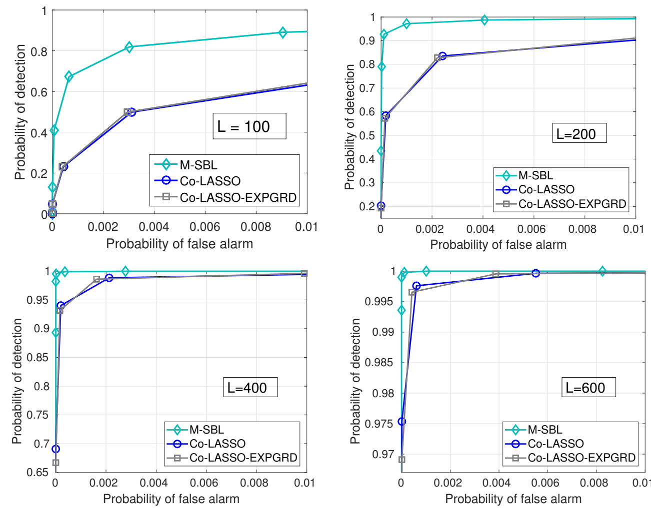

Fig. 2: ROC plots indicating the achievable combinations of support detection and false alarm probabilities for growing number of MMVs. Other simulation parameters: $n = 500$, $k = 200$, $m = 100$, SNR = 10 dB and number of trials = 100.

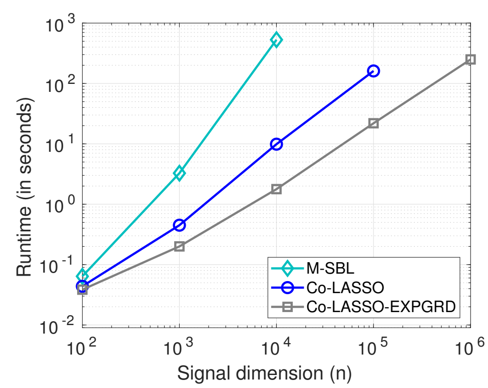

Fig. 3: Average runtimes of algorithms for signal dimension (n) varying from 100 to 10^6, and for $k = 20 \log n$, $m = \frac{k}{2}$, $L = 200 \log n$ and SNR = 10dB. Simulations were run on an Intel Xeon machine with 16 CPU cores and 64 gigabytes memory.

algorithm (implemented as an ℓ_1-regularized least squares program with nonnegativity constraints [19] using MATLAB code from *http://stanford.edu/~boyd/l1_ls*) and MSBL, the current state-of-the-art covariance matching method. Conventional MMV algorithms like SOMP, SCo-SAMP, SIHT, Row-LASSO and CS/SA-MUSIC are not included in our comparison as they fail to recover the true support when the support size, k, exceeds m. The entries of the measurement matrix $\mathbf{A}$ are drawn independently according to $\mathcal{N}(0, 1/\sqrt{m})$. The common k-sparse support of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$ is uniformly selected from $\binom{n}{k}$ possible combinations and the nonzero signal coefficients are i.i.d $\mathcal{N}(0, 1)$. For Co-LASSO and Co-LASSO-EXPGRD, the regularization parameter λ is set to $0.1\sigma\sqrt{2 \log n}$ and the learning rate parameter η is set to 0.5. In all three algorithms considered here, the active support is identified by hard-thresholding the converged $\hat{\gamma}$ using the threshold $0.65\sigma^2 + 0.35\sigma^2_{\text{sig}}$, where σ^2_{sig} denotes the signal variance which can be roughly estimated as $\hat{\sigma}^2_{\text{sig}} = \left(\text{tr}\left(\mathbf{R}_{\mathbf{yy}}\right) - m\sigma^2\right) / ||\mathbf{A}||^2_F$.

Figure 1 shows the support recovery phase transition of the proposed Co-LASSO-EXPGRD algorithm. The lighter and darker regions of the phase transition plots correspond to the success and failure in terms of at least 99% support reconstruction, respectively. The quadratic behavior of the support recovery phase transition empirically validates that the proposed algorithm is capable of recovering supports of size even beyond the ℓ_0 bound and up to $\mathcal{O}(m^2)$.

Next, we compare the Receiver Operation Characteristic (ROC) of the support recovery algorithms. Figure 2 plots the ROCs for $L = \{k/2, k, 2k, 3k\}$, which are generated by varying the threshold used to obtain hard support estimates. The asymmetric nature of the sparse support recovery problem demands that we achieve a low false alarm probability. For a fixed false alarm rate, the support detection improves as the number of MMVs grow as expected. MSBL consistently offers the best support detection rate for a fixed false alarm probability, however, it is computationally feasible only when n is at most a few hundred. Co-LASSO and Co-LASSO-EXPGRD have almost identical ROCs, which is expected, as both optimize the same cost function.

Finally, in Figure 3, we compare the average run-times of the algorithms for problem dimension n ranging from 100 to 1 million. For sparsity growing logarithmically with n, the Co-LASSO-EXPGRD is at least 10 times faster than MSBL and 2 to 5 times faster than Co-LASSO for n in thousands, and this gap increases as n grows. While MSBL and Co-LASSO are either computationally or storage wise bottlenecked, the proposed Co-LASSO-EXPGRD algorithm is able to scale easily to solve a million variable problem in tens of seconds.

VI. Final Remarks

Despite the ℓ_0-bound surpassing performance of the existing correlation-aware covariance matching MMV algorithms, their high computational and storage complexities has been a major bottleneck to their use in very large scale settings involving millions of variables. We have addressed this issue to a significant extent by proposing a new type of multiplicative support recovery updates based on the Co-LASSO algorithm. The new updates consume lesser memory resources for execution and converge faster leading to significantly shorter runtimes compared to the parent Co-LASSO and the current state-of-the-art algorithm, MSBL. The future extensions of this work can focus on analyzing the convergence of the proposed EG updates and developing their online/stochastic extensions.

REFERENCES

[1] M. Duarte, S. Sarvotham, D. Baron, M. Wakin, and R. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. Asilomar Conf. on Signals, Syst., and Comput.*, Oct. 2005, pp. 1537–1541.

[2] J. Tropp, A. Gilbert, and M. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proc. ICASSP*, vol. 5, 2005, pp. 721–724.

[3] J. Bazerque and G. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010.

[4] E. Lagunas, S. K. Sharma, S. Chatzinotas, and B. Ottersten, "Compressive sensing based target counting and localization exploiting joint sparsity," in *Proc. ICASSP*, March 2016, pp. 3231–3235.

[5] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4997–5008, Oct 2014.

[6] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec 2006.

[7] Q. Ling, Z. Wen, and W. Yin, "Decentralized jointly sparse optimization by reweighted l-q minimization," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1165–1170, 2013.

[8] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct 2004.

[9] J. D. Blanchard, M. Cermak, D. Hanle, and Y. Jing, "Greedy algorithms for joint sparse recovery," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1694–1704, April 2014.

[10] J. M. Kim, O. K. Lee, and J. C. Ye, "Compressive music: Revisiting the link between compressive sensing and array signal processing," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 278–301, Jan 2012.

[11] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3613–3641, June 2012.

[12] P. Pal and P. P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 711–726, Feb. 2015.

[13] S. Khanna and C. R. Murthy, "Rényi divergence based covariance matching pursuit of joint sparse support," in *Sig. Process. Advances in Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1 – 5.

[14] O. Balkan, K. Kreutz-Delgado, and S. Makeig, "Localization of more sources than sensors via jointly-sparse Bayesian learning," *IEEE Signal Processing Letters*, vol. 21, no. 2, pp. 131–134, Feb 2014.

[15] S. Khanna and C. R. Murthy, "On the support recovery of jointly sparse Gaussian sources using sparse Bayesian learning," *CoRR*, vol. abs/1703.04930. [Online]. Available: http://arxiv.org/abs/1703.04930

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

[17] K. Tsuda, G. Rätsch, and M. K. Warmuth, "Matrix exponentiated gradient updates for on-line learning and Bregman projection," *J. Mach. Learn. Res.*, vol. 6, pp. 995–1018, 2005.

[18] J. Kivinen and M. K. Warmuth, "Exponentiated gradient versus gradient descent for linear predictors," *Inf. Comput.*, vol. 132, no. 1, pp. 1–63, Jan 1997.

[19] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ_1-regularized least squares," *IEEE J. Sel. Topics Signal Process*, vol. 1, no. 4, pp. 606–617, Dec 2007.