

Improving M-SBL for Joint Sparse Recovery using a Subspace Penalty

Jong Chul Ye, Jong Min Kim and Yoram Bresler

Abstract

The multiple measurement vector problem (MMV) is a generalization of the compressed sensing problem that addresses the recovery of a set of jointly sparse signal vectors. One of the important contributions of this paper is to reveal that the seemingly least related state-of-art MMV joint sparse recovery algorithms - M-SBL (multiple sparse Bayesian learning) and subspace-based hybrid greedy algorithms - have a very important link. More specifically, we show that replacing the $\log \det(\cdot)$ term in M-SBL by a rank proxy that exploits the spark reduction property discovered in subspace-based joint sparse recovery algorithms, provides significant improvements. In particular, if we use the Schatten- p quasi-norm as the corresponding rank proxy, the global minimiser of the proposed algorithm becomes identical to the true solution as $p \rightarrow 0$. Furthermore, under the same regularity conditions, we show that the convergence to a local minimiser is guaranteed using an alternating minimization algorithm that has closed form expressions for each of the minimization steps, which are convex. Numerical simulations under a variety of scenarios in terms of SNR, and condition number of the signal amplitude matrix demonstrate that the proposed algorithm consistently outperforms M-SBL and other state-of-the art algorithms.

Index Terms

Compressed sensing, joint sparse recovery, multiple measurement vector problem, subspace method, M-SBL, generalized MUSIC criterion, rank proxy, Schatten- p norm

Correspondence to:

Jong Chul Ye, Ph.D. Professor

Dept. of Bio and Brain Engineering, KAIST

291 Daehak-ro Yuseong-gu, Daejeon 305-701, Republic of Korea

Email: jong.ye@kaist.ac.kr

Tel: 82-42-350-4320

Fax: 82-42-350-4310

I. INTRODUCTION

The multiple measurement vector problem (MMV) is a generalization of the compressed sensing problem, which addresses the recovery of a set of sparse signal vectors that share a common support [1], [2], [3], [4], [5], [6]. In the MMV model, let m and N denote the number of sensor elements and snapshots, respectively; and $n > m$ denote the length of the signal vectors. Then, for a given noisy observation matrix $Y = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbb{C}^{m \times N}$ and a sensing matrix $A \in \mathbb{C}^{m \times n}$, the multiple measurement vector (MMV) problem can be formulated as:

$$\begin{aligned} & \text{minimize} \quad \|X\|_0 \\ & \text{subject to} \quad \|Y - AX\|_F < \delta, \end{aligned} \tag{1}$$

where $\mathbf{x}_j \in \mathbb{R}^n$ is the j -th signal, $X = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{n \times N}$, \mathbf{x}^i is the i -th row of X , and $\|X\|_0 = |\text{supp}X|$, where $\text{supp}X = \{1 \leq i \leq n : \mathbf{x}^i \neq 0\}$ is the set of indices of nonzero rows in X . The Frobenius norm is used to measure the discrepancy between the data and the model. Classically, pursuit algorithms such as alternating minimization algorithm (AM) and MUSIC (multiple signal classification) algorithm [7], S-OMP (simultaneous orthogonal matching pursuit) [8], [2], M-FOCUSS [3], randomized algorithms such as REduce MMV and BOost (ReMBo)[4], and model-based compressive sensing using block-sparsity [9], [10] have been applied to the MMV problem.

An algebraic bound for the recoverable sparsity level has been theoretically studied by Feng and Bresler [7], and by Chen and Huo [2] for noiseless measurement Y . More specifically, if $X \in \mathbb{R}^{n \times N}$ satisfies $AX = Y$ and

$$\|X\|_0 < \frac{\text{spark}(A) + \text{rank}(Y) - 1}{2}, \tag{2}$$

where $\text{spark}(A)$ denotes the smallest number of linearly dependent columns of A , then X is the unique solution of (1). This indicates that the recoverable sparsity level may increase with an increasing number of measurement vectors. Indeed, for noiseless measurement, a MUSIC algorithm by Feng and Bresler [7] is shown to achieve the performance limit when the measurement matrix is full rank. However, except for MUSIC in full rank cases, the performance of the aforementioned classical MMV algorithms is not generally satisfactory, falling far short of (2) even for the noiseless case, when only a finite number of snapshots are available.

In a noisy environment, Obozinski *et al* showed that a near optimal sampling rate reduction up to $\text{rank}(Y)$ can be achieved using l_1/l_2 mixed norm penalty [11]. A similar gain was observed in

computationally inexpensive greedy approaches such as compressive MUSIC (CS-MUSIC) [1] and subspace augmented MUSIC (SA-MUSIC) [6]. More specifically, Kim *et al* [1] and Lee *et al* [6] independently showed that a class of hybrid greedy algorithms that combine greedy steps with a so called *generalized MUSIC* subspace criterion [1], or equivalently, with subspace augmentation [6], can reduce the required number of measurements by up to $\text{rank}(Y)$ in noisy environment. Furthermore, using a large system MMV model, Kim *et al* further showed that for an i.i.d. Gaussian sensing matrix, their algorithm can asymptotically achieve the algebraic performance limit when $\text{rank}(Y)$ increases with a particular scaling law [1]. Lee *et al* [6] also showed that MUSIC can do this in the noisy case and full rank, non-asymptotically with finite data, and for realistic Fourier sensing matrices.

While the aforementioned mixed norm approach and subspace based greedy approaches provide theoretical performance guarantees, there also exist a very different class of powerful MMV algorithms that are based on empirical Bayesian and Automatic Relevance Determination (ARD) principles from machine learning. Among these, the so-called multiple sparse Bayesian learning (M-SBL) algorithm is best known [12]. Even though M-SBL is more computationally expensive than greedy algorithms such as CS-MUSIC or SA-MUSIC, empirical results show that M-SBL is quite robust to noise and to unfavorable restricted isometry property constant (RIC) of the sensing matrix [13]. Moreover, M-SBL is more competitive than mixed norm approaches. Since Bayesian approaches are very different from classical compressed sensing, such high performance appears mysterious at first glance. However, a recent breakthrough by Wipf *et al* unveiled that M-SBL can be converted to a standard compressed sensing framework with an additional $\log \det(\cdot)$ (log determinant) penalty - a non-separable sparsity inducing prior [14]. The presence of the non-separable penalty term is so powerful that M-SBL performs almost as well as MUSIC. However, the guarantee only applies to the full row rank case of X with noise-free measurement vectors [15], [12]. However, despite its excellent performance, compared to the mixed norm approaches or subspace greedy algorithms, other than the work by Wipf *et al* [14], the fundamental theoretical analysis of M-SBL has been limited.

Therefore, one of the main goals of this paper is to continue the effort by Wipf *et al* [14] and analyze the origin of the high performance of M-SBL, as well as to investigate its limitations. One of the important contributions of this paper is to show that the seemingly least related algorithms - M-SBL and subspace-based hybrid greedy algorithms - have a very important link. More specifically, we show that the $\log \det(\cdot)$ term in M-SBL is a proxy for the rank of a partial sensing matrix corresponding to the true support. We then show that minimising the rank that was discovered in subspace-based hybrid greedy algorithm to exploit the spark reduction property of MMV can indeed provide a true solution for

the MMV problem. Accordingly, replacing $\log \det(\cdot)$ term in M-SBL by a Schatten- p quasi-norm rank proxy provides significant performance improvements.

The resulting new algorithm is no longer Bayesian due to the use of a *deterministic* penalty based on a geometric argument, so we call the new algorithm *subspace-penalized sparse learning (SPL)* by excluding term “Bayesian”. We show that as $p \rightarrow 0$ in the Schatten p -norm rank proxy, the global minimizers of the SPL cost function are identical to those of the original l_0 minimization problem. Furthermore, we show that SPL can be easily implemented as an alternating minimization approach.

Using numerical simulations, we demonstrate that compared to the current state-of-art MMV algorithms such as mixed norm approaches, M-SBL, CS-MUSIC/SA-MUSIC, and sequential CS-MUSIC [13], SPL provides superior recovery performance. Moreover, the results show that SPL is very robust to noise, and to the condition number of the unknown signal matrix.

A. Notation

Throughout the paper, \mathbf{x}^i and \mathbf{x}_j correspond to the i -th row and the j -th column of matrix X , respectively. The (i, j) element of X is represented by x_{ij} . When S is an index set, X^S , and A_S correspond to a submatrix collecting corresponding rows of X and columns of A , respectively. For a matrix A , $\text{Tr}(A)$ is the trace of a matrix A , A^* is its adjoint, A^\dagger denotes the Penrose-Moor psuedo-inverse, $|A|$ refers the determinant, $R(A)$ denotes the range space of A , and P_A (or $P_{R(A)}$) and P_A^\perp (or $P_{R(A)}^\perp$) are the projection on the range space and its orthogonal complement, respectively. The vector \mathbf{e}_i denotes an elementary unit vector whose i -th element is 1, and I denotes an identity matrix.

A sensing matrix $A \in \mathbb{R}^{m \times n}$ is said to have a k -restricted isometry property (RIP) if there exist left and right RIP constants $0 \leq \delta_k^L, \delta_k^R < 1$ such that

$$(1 - \delta_k^L) \|\mathbf{x}\|^2 \leq \|A\mathbf{x}\|^2 \leq (1 + \delta_k^R) \|\mathbf{x}\|^2$$

for all $\mathbf{x} \in \mathbb{R}^n$ such that $\|\mathbf{x}\|_0 \leq k$. A single RIP constant $\delta_k = \max\{\delta_k^L, \delta_k^R\}$ is often referred to as the RIP constant.

II. M-SBL: A REVIEW

A. Algorithm Description

Under appropriate assumptions of noise and signal Gaussian statistics, one can show that M-SBL minimizes the following cost function in a so-called γ space [12]:

$$\mathcal{L}^\gamma(\gamma) = \text{Tr}(\Sigma_y^{-1} Y Y^*) + N \log |\Sigma_y| \quad (3)$$

where

$$\Sigma_y = \lambda I + A\Gamma A^* \quad , \quad \Gamma = \text{diag}(\gamma) \quad . \quad (4)$$

With an estimate of Γ , which typically has a nearly sparse diagonal and may be thresholded to be exactly sparse, the solution of M-SBL is given by

$$X = \Gamma A^* (\lambda I + A\Gamma A^*)^{-1} Y \quad . \quad (5)$$

One of the most important contributions by Wipf is that the minimization problem of the cost function (3) can be equivalently represented as the following standard sparse recovery framework [14]:

$$\min_X \mathcal{L}^x(X), \quad \mathcal{L}^x(X) = \|Y - AX\|_F^2 + \lambda g_{msbl}(X) \quad (6)$$

where $g_{msbl}(X)$ is a penalty given by

$$g_{msbl}(X) = \min_{\gamma \geq 0} G(X; \gamma) \quad (7)$$

where

$$G(X; \gamma) \equiv \text{Tr}(X^* \Gamma^{-1} X) + N \log |\lambda I + A\Gamma A^*| \quad . \quad (8)$$

Wipf *et al* [14] gave a heuristic argument showing that g_{msbl} corresponds to a non-separable sparsity promoting penalty, and proposed the following alternating minimization approach to solve the minimization problem (6).

1) *Step 1: Minimization with respect to X* : For a given estimate $\gamma^{(t)}$ at the t -th iteration, we can find a closed form solution for X in (6):

$$X^{(t)} = \Gamma^{(t)} A^* (\lambda I + A\Gamma^{(t)} A^*)^{-1} Y, \quad \Gamma^{(t)} = \text{diag}(\gamma^{(t)}).$$

For large scale problems, this can be computed using a standard conjugate gradient algorithm with an appropriate preconditioner.

2) *Step 2: Minimization with respect to γ* : In this step, for a given $X^{(t)}$ we need to solve the following minimization problem:

$$\gamma^{(t+1)} = \arg \min_{\gamma \geq 0} G(X^{(t)}, \gamma)$$

where

$$G(X^{(t)}, \gamma) = \text{Tr} \left(X^{(t)*} \Gamma^{-1} X^{(t)} \right) + N \log |\Sigma_y| . \quad (9)$$

Wipf *et al* [14] find the solution to $\nabla G(X^{(t)}, \gamma) = \mathbf{0}$. More specifically, the derivative with respect to each component is given by

$$\frac{\partial G(X^{(t)}, \gamma)}{\partial \gamma_i} = -\frac{\sum_j |x_{ij}^{(t)}|^2}{\gamma_i^2} + N \mathbf{a}_i^H (\lambda I + A \Gamma A^*)^{-1} \mathbf{a}_i \quad (10)$$

since

$$\frac{\partial |\Sigma_y|}{\partial \gamma_i} = |\Sigma_y| \text{Tr} \left(\Sigma_y^{-1} \frac{\partial \Sigma_y}{\partial \gamma_i} \right) .$$

Setting the derivative to zero after fixing $\Gamma := \Gamma^{(t)}$, this observation leads to the following fixed point update of γ :

$$\gamma_i^{(t+1)} = \left(\frac{\frac{1}{N} \sum_j |x_{ij}^{(t)}|^2}{\mathbf{a}_i^H (\lambda I + A \Gamma^{(t)} A^*)^{-1} \mathbf{a}_i} \right)^{\frac{1}{2}} . \quad (11)$$

B. Role of the Non-separable Penalty in M-SBL

In order to develop a new joint sparse recovery algorithm that improves on M-SBL, we provide here a new interpretation of the role of the regularization term in M-SBL. Note that due to the non-negativity constraint for γ , a critical solution to the minimization problem in (7) should satisfy the following first order Karush-Kuhn-Tucker (KKT) necessary conditions [16]:

$$\gamma_i \frac{\partial G(X, \gamma)}{\partial \gamma_i} = \gamma_i \left(-\frac{\sum_j |x_{ij}|^2}{\gamma_i^2} + N \mathbf{a}_i^H (\lambda I + A \Gamma A^*)^{-1} \mathbf{a}_i \right) = 0, \quad \forall i$$

Hence, as $\lambda \rightarrow 0$, this leads to the following fixed point equation:

$$\lim_{\lambda \rightarrow 0} \gamma_i = \lim_{\lambda \rightarrow 0} \frac{\frac{1}{N} \|\mathbf{x}^i\|^2}{\gamma_i \mathbf{a}_i^* (\lambda I + A \Gamma A^*)^{-1} \mathbf{a}_i} .$$

If $\|\gamma\|_0 < m$, using the matrix inversion lemma, we have

$$\begin{aligned} \lambda (\lambda I + A \Gamma A^*)^{-1} &= I - A \Gamma^{\frac{1}{2}} \left(\lambda I + \Gamma^{\frac{1}{2}} A^* A \Gamma^{\frac{1}{2}} \right)^{-1} \Gamma^{\frac{1}{2}} A^* \\ &= P_{A_S}^\perp + \lambda (A_S^\dagger)^* (\lambda (A_S^* A_S)^{-1} + \Gamma_S)^{-1} A_S^\dagger , \end{aligned} \quad (12)$$

where $S = \text{supp}(\gamma)$ denotes a nonzero support of γ and P_{A_S} denotes the orthogonal projection on the span of the columns of A indexed by S . Accordingly,

$$\lim_{\lambda \rightarrow 0} \mathbf{a}_i^* (\lambda I + A\Gamma A^*)^{-1} \mathbf{a}_i = \mathbf{a}_i^* (A_S^\dagger)^* \Gamma_S^{-1} A_S^\dagger \mathbf{a}_i = \frac{1}{\gamma_i}, \quad i \in S.$$

Therefore, we have

$$\lim_{\lambda \rightarrow 0} \gamma_i = \frac{1}{N} \|\mathbf{x}^i\|^2, \quad i \in S. \quad (13)$$

Substituting (13) into (7) yields

$$\begin{aligned} g_{msbl}(X) &= \min_{\gamma \geq \mathbf{0}} \text{Tr}(X^* \Gamma^{-1} X) + N \log |\lambda I + A\Gamma A^*| \\ &\approx N |S| + N \log |\lambda I + A\Gamma A^*| \\ &= N \|\gamma\|_0 + N \log |\lambda I + A\Gamma A^*| \end{aligned} \quad (14)$$

Note that the first term in (14) imposes row sparsity on X since $\gamma_i = 0$ for $\|\mathbf{x}^i\| = 0$ due to (13). Hence, the first term of M-SBL penalty is in fact $\|X\|_0$. Then, what is the meaning of the $\log |\cdot|$ term? Wipf *et al* [14] showed that the superior performance of the M-SBL is owing to the non-separability of the term $\log |\lambda I + A\Gamma A^*|$ with respect to γ , which can avoid many local minimizers. In addition to this interpretation, the following section shows another important geometric implications of the $\log \det(\cdot)$ term.

III. SUBSPACE-PENALIZED SPARSE LEARNING

A. Key Observation

In this section, we provide another interpretation of the M-SBL penalty, which suggests a new algorithm called subspace-penalized sparse learning (SPL) that overcomes the limitation of M-SBL. Note that for any matrix $Z \in \mathbb{R}^{m \times n}$ with $m \leq n$, we have

$$\log |ZZ^* + \lambda I| = \sum_{i=1}^m \log(\sigma_i^2(Z) + \lambda), \quad (15)$$

where $\sigma_i(Z)$ denote the singular values of Z . Therefore, the $\log \det(\cdot)$ function is a concave proxy for nonzero singular values, hence in the limit of $\lambda \rightarrow 0$, (15) acts as a proxy for $\text{rank}(Z)$ [17], [18]. This leads us to another interpretation: the penalty term in M-SBL is equivalent to

$$\lim_{\lambda \rightarrow 0} g_{msbl}(X) = N \|\gamma\|_0 + N \text{Rprox}(A\Gamma^{\frac{1}{2}}) \quad (16)$$

where $\text{Rprox}(\cdot)$ denotes a rank proxy. Thus, the penalty simultaneously imposes the row sparsity of X as well as the low rank of the matrix $A\Gamma^{\frac{1}{2}}$. By inspection, the first sparsity penalty term in (16) is quite intuitive, but it is not clear why $\text{Rank}(A\Gamma^{\frac{1}{2}})$ needs to be minimized.

In fact, the main contribution of this paper is that we need to replace the second term, $\text{Rprox}(A\Gamma^{\frac{1}{2}})$, by geometrically more intuitive rank proxy as follows:

$$g_{SPL}(X) = N\|\gamma\|_0 + N \text{Rprox}(Q^* A\Gamma^{\frac{1}{2}}) \quad (17)$$

where Q denotes a basis for the noise subspace denoted as $R(Q) = R^\perp(Y)$. In the following, we will describe in detail how we arrive at the new rank penalty.

B. Subspace Criteria

A solution X for $Y = AX$ that satisfies $\|X\|_0 \leq m$ is called a *basic feasible solution* (BFS) [14]. Among BFSs, a solution of the following l_0 MMV problem is called *maximally sparse* solution:

$$(P0) : \min_X \|X\|_0, \quad \text{subject to } Y = AX. \quad (18)$$

To address (18), subspace-based greedy algorithms such as CS-MUSIC [1] and SA-MUSIC [6] exploit the *spark reduction principle* or an equivalent subspace criterion using an *augmented signal subspace*. More specifically, if $r = \text{rank}(Y)$ and k denotes the number of the non-zero rows, the algorithms first estimate $k - r$ partial support index I_{k-r} , then the remaining r components of the support are found using the subspace criterion.

One of the main contributions of this paper is that this two step approach is not necessary. Instead, for noiseless measurements, a direct minimization of the rank of $Q^* A_I$ with respect to index I , $|I| \geq k$, still guarantees to obtain the true support as shown in Theorem 3.1. We believe that this is an extremely powerful result that provides an important clue to overcome the limitation of existing greedy subspace methods [1], [6].

Theorem 3.1: Assume that $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times r}$, $Y \in \mathbb{R}^{m \times r}$ satisfy $AX = Y$, where $\|X\|_0 = k$, and the columns of Y are linearly independent. If A satisfies a RIP condition $0 \leq \delta_{2k-r+1}^L(A) < 1$, then we have

$$k - r = \min_{|I| \geq k} \text{rank}(Q^* A_I),$$

and

$$\text{supp}X = \arg \min_{|I| \geq k} \text{rank}(Q^* A_I).$$

Proof: Since $\min_{|I|=l} \text{rank}(Q^* A_I)$ is a nondecreasing function of l , we may consider $\min_{|I|=k} \text{rank}(Q^* A_I)$. By the rank-nullity Theorem, $\dim(N(Q^* A_I)) + \dim(R(Q^* A_I)) = k$ for $I \subset \{1, \dots, n\}$ with $|I| = k$. Furthermore, because $N(Q^*) = Y$,

$$N(Q^* A_I) = \{\mathbf{v} \in \mathbb{R}^k : A_I \mathbf{v} \in R(Y)\} \quad (19)$$

and because $r \leq k$, $N(A_I) = \{0\}$ by the RIP condition $0 \leq \delta_{2k-r+1}(A) < 1$. Since $N(A_I) = \{0\}$, we have $\dim(N(Q^* A_I)) = \dim(R(Y) \cap R(A_I))$ so that

$$\dim(N(Q^* A_I)) = \dim(R(Y) \cap R(A_I)) \leq \dim(R(Y)) = r, \quad (20)$$

which also implies that $\text{rank}(Q^* A_I) \geq k - r$ for any $|I| = k$. Hence, denoting $\text{supp}(X) = S$, it is enough to show that

$$\text{rank}(Q^* A_S) = k - r \quad (21)$$

and

$$\text{rank}(Q^* A_I) > k - r, \text{ for } |I| = k \text{ and } I \neq S. \quad (22)$$

First we will show that (21) holds. Because $N(A_S) = \{0\}$ and $Y = A_S X^S$, we have that $\text{rank}(Y) = \text{rank}(X^S)$ or $\dim(R(X^S)) = r$. Also, since $Y = A_S X^S$, by (19), we have $R(X^S) \subset N(Q^* A_S)$. Hence $\dim(N(Q^* A_S)) \geq r$. On the other hand, by (20), the dimension of $N(Q^* A_S)$ is at most r , which implies $\dim(N(Q^* A_S)) = r$, so that we have $\text{rank}(Q^* A_S) = k - r$.

Then, (22) is the only a remaining part to prove. Suppose that we have an index set $I \subset \{1, \dots, n\}$ such that $|I| = k$ and $\text{rank}(Q^* A_I) = k - r$. Then, it must hold that

$$\dim(N(Q^* A_I)) = \dim(R(Y) \cap R(A_I)) = r = \dim R(Y).$$

It follows that $R(Y) = R(Y) \cap R(A_I) \subset R(A_I)$. Then, for each column \mathbf{y}_i of Y , $\mathbf{y}_i \in R(Y) \subset R(A_I)$ so that there exists $\mathbf{z}_i \in \mathbb{R}^{|I|}$ such that $A_I \mathbf{z}_i = \mathbf{y}_i$. Then, $A_I [\mathbf{z}_1 \ \dots \ \mathbf{z}_r] = [\mathbf{y}_1 \ \dots \ \mathbf{y}_r] = Y$ so that there is a $\tilde{X} \in \mathbb{R}^{n \times r}$ such that $Y = A \tilde{X}$ with $\text{supp} \tilde{X} \subset I$.

Since $\text{rank}(X) = r$ and $(\mathbf{x}^i)^T = 0$ for any $i \notin S = \text{supp} X$, it follows that X^S has rank r . Hence, because the row rank of a matrix equals its column rank, X^S must have r linearly independent rows. Therefore, there is a subset Z of $\text{supp} X$ such that $|Z| = r$, and the rows of X^Z are linearly independent.

Since $X^Z \in \mathbb{R}^{r \times r}$, for every $i \in Z$ there is a nonzero vector $\mathbf{w}_i \in N(X^{Z \setminus \{i\}})$, so that we have $\|X \mathbf{w}_i\|_0 \leq k - r + 1$ and $i \in \text{supp}(X \mathbf{w}_i)$, since $\mathbf{x}^i \mathbf{w}_i \neq 0$ by the linearly independence of the rows of

X^Z . Since $Y = AX = A\tilde{X}$, we have $A(X\mathbf{w}_i - \tilde{X}\mathbf{w}_i) = 0$.

Now, because $\text{supp}(\tilde{X}) \subset I$, we have $\text{supp}(\tilde{X}\mathbf{w}_i) \subset \text{supp}(\tilde{X}) \subset I$. Hence $\|\tilde{X}\mathbf{w}_i\| \leq |I| = k$. It follows that

$$\|X\mathbf{w}_i - \tilde{X}\mathbf{w}_i\|_0 \leq k - r + 1 + |I| = 2k - r + 1.$$

Hence, by the RIP of A , we must have $X\mathbf{w}_i = \tilde{X}\mathbf{w}_i$. Since $\text{supp}(\tilde{X}\mathbf{w}_i) \subset I$, we also have $\text{supp}(\tilde{X}\mathbf{w}_i) = \text{supp}(X\mathbf{w}_i) \subset I$, which implies that $i \in \text{supp}(X\mathbf{w}_i) \subset I$. Since i can be any element in $\text{supp}X$, we have $i \in I$ for any $i \in \text{supp}X$. It implies that $\text{supp}X \subset I$ so that $I = \text{supp}X$ since $|I| = |\text{supp}X| = k$. Hence, in order to satisfy $\text{rank}(Q^*A_I) = k - r$ and $|I| = k$, we must have $I = \text{supp}X$. ■

C. The SPL Penalty

Note that minimizing $\text{rank}(Q^*A\Gamma^{\frac{1}{2}})$ with respect to γ is equivalent to finding the index set I that minimizes $\text{rank}(Q^*A_I)$. Hence, Theorem 3.1 implies that minimizing $\text{rank}(Q^*A\Gamma^{\frac{1}{2}})$ under the constraint $\|\gamma\|_0 \geq k$ will find γ_* that has non-zero values for indices corresponding to $\text{supp}X_*$, where $\|X_*\|_0 = k$ and $Y = AX_*$. This observation leads to the second term in the SPL penalty of (17) as a rank proxy to exploit this geometric finding.

Moreover, rather than just using $\log \det(\cdot)$ as in M-SBL, in this paper, we use more general family of rank proxies that still satisfy our goals. Specifically, our rank proxy is based on *Schatten- p quasi norm* with $0 < p \leq 1$ that includes the popular nuclear norm as a special case. For a matrix $W \in \mathbb{R}^{m \times n}$, the Schatten p -norm proxy for the rank is defined as

$$\text{Tr}|W|^p = \text{Tr}\left((WW^*)^{\frac{p}{2}}\right) = \sum_{i=1}^m \sigma_i^p(W), \quad (23)$$

which corresponds to the nuclear norm when $p = 1$. Following the derivation that leads to (7), we propose the following SPL penalty:

$$g_{SPL}(X) \equiv \min_{\gamma \geq 0} \mathcal{G}_{SPL}(\gamma, X) \quad (24)$$

where

$$\mathcal{G}_{SPL}(\gamma, X) = \text{Tr}(X^*\Gamma^{-1}X) + N\text{Tr}\left((Q^*A\Gamma A^*Q)^{\frac{p}{2}}\right). \quad (25)$$

Using the proposed SPL penalty, we formulate the following noiseless SPL minimization problem:

$$\min_X g_{SPL}(X), \quad \text{subject to } Y = AX. \quad (26)$$

Note that $\text{Tr} \left((Q^* A \Gamma A^* Q)^{\frac{p}{2}} \right)$ is a concave function with respect to its singular values, so we can find its convex conjugate:

$$\mathcal{G}_{SPL}(\gamma, X) \equiv \min_{\Psi \in \mathbf{S}_{0+}} \mathfrak{G}_{SPL}(\gamma, X, \Psi) \quad (27)$$

where $\mathfrak{G}_{SPL}(\gamma, X, \Psi)$ is given as

$$\mathfrak{G}_{SPL}(\gamma, X, \Psi) \equiv \text{Tr} (X^* \Gamma^{-1} X) + Np \left(\text{Tr} ((Q^* A \Gamma A^* Q) \Psi) - \frac{2}{q} \text{Tr} (\Psi^{\frac{q}{2}}) \right). \quad (28)$$

for q such that $1/(p/2) + 1/(q/2) = 1$; and \mathbf{S}_{0+} denotes the set of symmetric positive semi-definite matrices:

$$\mathbf{S}_{0+} = \{X \in \mathbf{S} : X \succeq 0\}.$$

The relationship between (28) and (25) can be clearly understood by minimizing (28) with respect to Ψ . Indeed, using $\partial \text{Tr}(A\Psi)/\partial \Psi = A^*$ and $\partial \text{Tr}(\Psi)^{q/2} \partial \Psi = \Psi^{q/2-1}$ [19], we have

$$\Psi = (Q^* A \Gamma A^* Q)^{\frac{1}{q/2-1}} \quad (29)$$

and

$$\min_{\Psi \in \mathbf{S}_{0+}} \text{Tr} ((Q^* A \Gamma A^* Q) \Psi) - \frac{2}{q} \text{Tr} (\Psi^{\frac{q}{2}}) = \frac{1}{p} \text{Tr} \left((Q^* A \Gamma A^* Q)^{\frac{p}{2}} \right)$$

Here, $(Q^* A \Gamma A^* Q)^{\frac{1}{q/2-1}}$ should be understood as applying the power operation to the non-zero singular values of $Q^* A \Gamma A^* Q$ while retaining zero singular values at zero.

Notice that $\mathfrak{G}_{SPL}(\gamma, X, \Psi)$ is a surrogate function that majorizes $\mathcal{G}_{SPL}(X, \gamma)$. Although like (25), (28) is not jointly convex with respect to the different variables, the reason to prefer (28) over (25) is that (28) is convex with respect to each of the variables γ , X , and Ψ with the other held constant, and we can obtain a closed-form expression in each step of alternating minimization. Specifically, recall that the SPL penalty is given by

$$g_{SPL}(X) = \min_{\gamma \geq 0, \Psi \in \mathbf{S}_{0+}} \mathfrak{G}_{SPL}(\gamma, X, \Psi) \quad (30)$$

where

$$\mathfrak{G}_{SPL}(\gamma, X, \Psi) = \text{Tr} (X^* \Gamma^{-1} X) + Np \left(\text{Tr} ((Q^* A \Gamma A^* Q) \Psi) - \frac{2}{q} \text{Tr} (\Psi^{\frac{q}{2}}) \right).$$

Let S denotes the non-zero support set of X . Using the KKT condition with respect to γ , we have

$$\frac{\partial \mathfrak{G}_{SPL}(\gamma, X, \Psi)}{\partial \gamma_i} - \mu_i = -\frac{\|\mathbf{x}^i\|^2}{\gamma_i^2} + N\mathbf{a}_i^* Q \Psi Q^* \mathbf{a}_i - \mu_i = 0, \quad \forall i \quad (31)$$

$$\mu_i \gamma_i = 0, \quad \mu_i \geq 0, \quad \gamma_i \geq 0, \quad \forall i \quad (32)$$

which leads to $\gamma_i = 0$ for $i \notin S$, whereas for $i \in S$

$$\gamma_i \frac{\partial \mathfrak{G}_{SPL}}{\partial \gamma_i} = 0 = -\frac{\|\mathbf{x}^i\|^2}{\gamma_i} + N\gamma_i \mathbf{a}_i^* Q (Q^* A \Gamma A^* Q)^{\frac{1}{q/2-1}} Q^* \mathbf{a}_i.$$

Hence, $\text{Tr}(X^* \Gamma^{-1} X) = N \text{Tr} \left((Q^* A \Gamma A^* Q)^{1+\frac{1}{q/2-1}} \right) = N \text{Tr} \left((Q^* A \Gamma A^* Q)^{\frac{p}{2}} \right)$ and we have

$$g_{SPL}(X) = 2N \text{Tr} \left((Q^* A \Gamma A^* Q)^{\frac{p}{2}} \right) = 2N \text{Tr} |Q^* A \Gamma^{\frac{1}{2}}|^p.$$

This implies that at the KKT point, the SPL penalty has cost function values equivalent to the Schatten- p quasi-norm rank penalty for $Q^* A \Gamma^{\frac{1}{2}}$.

IV. THE SPL ALGORITHM

A. Alternating Minimization Algorithm

So far, we have analyzed the global minimizer for the noiseless SPL algorithm. For noisy measurement, we propose the following cost function:

$$\min_X \|Y - AX\|_F^2 + \lambda g_{SPL}(X). \quad (33)$$

By letting $\lambda \rightarrow 0$, the solution of (33) becomes a solution of (26) when $\text{rank}(A \Gamma^{1/2}) = m$ since then the constraint is automatically satisfied as follows:

$$\lim_{\lambda \rightarrow 0} AX(\lambda) = \lim_{\lambda \rightarrow 0} A \Gamma A^* (\lambda I + A \Gamma A^*)^{-1} Y = Y$$

Similar equivalence can be hold for $\text{rank}(A \Gamma^{1/2}) < m$ if $Y \in R(A \Gamma^{1/2})$. Therefore, rather than dealing with Eqs. (26) and (33) separately, we use (33) and the limiting argument to discuss a noiseless SPL optimization problem.

Then, using (30), a noisy SPL formulation can be written as

$$\min_{X, \gamma \geq \mathbf{0}, \Psi \in \mathbf{S}_{0+}} C(X, \gamma, \Psi) \quad (34)$$

where the augmented cost function is given by

$$C(X, \gamma, \Psi) = \|Y - AX\|_F^2 + \lambda \left\{ \text{Tr}(X^* \Gamma^{-1} X) + Np \left[\text{Tr}((Q^* A \Gamma A^* Q) \Psi) - \frac{2}{q} \text{Tr}(\Psi^{\frac{q}{2}}) \right] \right\} \quad (35)$$

While $C(X, \gamma, \Psi)$ is not convex for all these variables simultaneously due to the presence of the bi-convex terms $\text{Tr}(X^* \Gamma^{-1} X)$ and $\text{Tr}((Q^* A \Gamma A^* Q) \Psi)$, it is convex with respect to each variable X, γ and Ψ separately. Indeed, this is a typical example of the d.c. algorithm (DCA) for the difference of convex functions programming [20], [21], and the alternating minimization algorithm converges to a *local* minimizer or a critical point.

Specifically, a critical solution should satisfy the following first order Karush-Kuhn-Tucker (KKT) necessary conditions [16]:

$$\frac{\partial C(X, \gamma, \Psi)}{\partial X} = -2A^*(Y - AX) + \lambda \Gamma^{-1} X = \mathbf{0} \quad (36)$$

$$\frac{\partial C(X, \gamma, \Psi)}{\partial \Psi} = Np(Q^* A \Gamma A^* Q) - Np\Psi^{q/2-1} = 0 \quad (37)$$

$$\frac{\partial C(X, \gamma, \Psi)}{\partial \gamma_i} - \mu_i = -\frac{\sum_j |x_{ij}|^2}{\gamma_i^2} + N\mathbf{a}_i^* Q \Psi Q^* \mathbf{a}_i - \mu_i = 0, \quad \forall i \quad (38)$$

$$\mu_i \gamma_i = 0, \quad \mu_i \geq 0, \quad \gamma_i \geq 0, \quad \forall i \quad (39)$$

This leads us to the following fixed point iterations:

1) *Minimization with respect to X* : For a given estimate $\gamma^{(t)}$, (36) yields a closed form solution for $X^{(t+1)}$:

$$X^{(t+1)} = \Gamma^{(t)} A^* (\lambda I + A \Gamma^{(t)} A^*)^{-1} Y, \quad \Gamma^{(t)} = \text{diag}(\gamma^{(t)}).$$

2) *Determination of Ψ* : For a given estimate $\gamma^{(t)}$, using (37), we can find a closed-form solution for $\Psi^{(t)}$: i.e. $\Psi^{(t)} = (Q^* A \Gamma^{(t)} A^* Q)^{\frac{1}{q/2-1}}$.

3) *Estimation of γ* : For a given $\Psi^{(t)}$ and $X^{(t)}$, using Eqs. (38) and (39), we have

$$\gamma_i \frac{\partial C(X^{(t)}, \gamma, \Psi^{(t)})}{\partial \gamma_i} = \gamma_i \left(-\frac{\|\mathbf{x}^{(t)i}\|^2}{\gamma_i^2} + N\mathbf{a}_i^* Q \Psi^{(t)} Q^* \mathbf{a}_i \right) = 0. \quad (40)$$

Here, if $\mathbf{a}_i^* Q \Psi^{(t)} Q^* \mathbf{a}_i \neq 0$, we have the following update equation:

$$\gamma_i^{(t)} = \left(\frac{\frac{1}{N} \|\mathbf{x}^{(t)i}\|^2}{\mathbf{a}_i^* Q \Psi^{(t)} Q^* \mathbf{a}_i} \right)^{\frac{1}{2}} = \left(\frac{\frac{1}{N} \|\mathbf{x}^{(t)i}\|^2}{\mathbf{a}_i^* Q (Q^* A \Gamma^{(t)} A^* Q)^{\frac{1}{q/2-1}} Q^* \mathbf{a}_i} \right)^{\frac{1}{2}}. \quad (41)$$

Note that the SPL updates appears similar to those of M-SBL except the γ update by (41), which is now modified based on subspace geometry. This is the main ingredient for the performance improvement

of SPL over M-SBL. In the following, we further discuss several important properties of the SPL penalty.

B. Properties of the SPL Penalty

An interesting case occurs when $\mathbf{a}_i^* Q \Psi^{(t)} Q^* \mathbf{a}_i = 0$. In this case, based on (40), we have the following two observations: 1) $\gamma_i^{(t)} \rightarrow \infty$ when $\|\mathbf{x}^{(t)i}\|^2 \neq 0$; and 2) $\gamma_i^{(t)}$ can be an arbitrary positive number C when $\|\mathbf{x}^{(t)i}\|^2 = 0$ since the equality in (40) is satisfied regardless of the choice of C . Therefore, we define the following γ_i update¹:

$$\gamma_i^{(t)} = \begin{cases} \infty, & \text{if } \mathbf{a}_i^* Q \Psi^{(t)} Q^* \mathbf{a}_i = 0 \text{ and } \|\mathbf{x}^{(t)i}\|^2 \neq 0 \\ C \gg 0, & \text{if } \mathbf{a}_i^* Q \Psi^{(t)} Q^* \mathbf{a}_i = 0 \text{ and } \|\mathbf{x}^{(t)i}\|^2 = 0 \\ \left(\frac{\frac{1}{N} \|\mathbf{x}^{(t)i}\|^2}{\mathbf{a}_i^* Q \Psi^{(t)} Q^* \mathbf{a}_i} \right)^{\frac{1}{2}} & \text{if } \mathbf{a}_i^* Q \Psi^{(t)} Q^* \mathbf{a}_i \neq 0 \end{cases} \quad (42)$$

Thanks to (42), even if $\|\mathbf{x}^{(t)i}\|^2$ becomes erroneously zero during the iterations, there is a possibility, when $\Psi^{(t)} Q^* \mathbf{a}_i = 0$, for $\gamma_i^{(t)}$ to become nonzero; hence, the corresponding row of $X^{(t)}$ can become nonzero once γ_i turns into nonzero. Note that this is very different from M-SBL, since in (11) the denominator term cannot be zero even under the most relaxed RIP constraint $\delta_{k+1}^L < 1$, so the condition $\|\mathbf{x}^{(t)i}\|^2 = 0$ will set the corresponding $\gamma_i^{(k)}$ to zero. Therefore, in M-SBL, once a row of $X^{(t)}$ is set to zero in error, it will stay zero for all subsequent iterations and the algorithm is unable to recover from this error.

Second, it is important note that since $1/(p/2) + 1/(q/2) = 1$, we have $\lim_{q \rightarrow 0} p = \lim_{q \rightarrow 0} \frac{2}{1-2/q} = 0$; so

$$\lim_{q \rightarrow 0} g_{SPL}(X) = \lim_{q \rightarrow 0} 2N \text{Tr} |Q^* A \Gamma^{\frac{1}{2}}|^p = 2N \text{Rank}(Q^* A_I), \quad (43)$$

where I denotes the index set of non-zero diagonal elements of Γ . Hence, in this case, the SPL algorithm with $p \rightarrow 0$ is the algorithm that directly minimizes the rank of $Q^* A_I$. In this case, the corresponding update rule is given by

$$X^{(t+1)} = \Gamma^{(t)} A^* (\lambda I + A \Gamma^{(t)} A^*)^{-1} Y, \quad \gamma_i^{(t)} = \left(\frac{\frac{1}{N} \|\mathbf{x}^{(t)i}\|^2}{\mathbf{a}_i^* Q (Q^* A \Gamma^{(t)} A^* Q)^{-1} Q^* \mathbf{a}_i} \right)^{\frac{1}{2}}. \quad (44)$$

The main technical challenge is, however, that for $p = q = 0$ the cost function (35) is not well-defined. Therefore, the aforementioned interpretation of the SPL should be understood as an asymptotic result such that p and q approach zero, but are not exactly zero.

¹In a practical implementation, a tolerance around 0 and finites values for $\gamma_i^{(t)}$ have to be used.

Next, as a by product of Theorem 3.1, the SPL algorithm is computationally more efficient than M-SBL. Note that the computational bottleneck of M-SBL (or SPL) is due to the the inversion of $A\Gamma^{(t)}A^*$ (or $Q^*A\Gamma^{(t)}A^*Q$, respectively). Specifically, unlike the $X^{(t)}$ update step that can be done using the conjugate gradient (CG) algorithm, the matrix inversion cannot be performed using CG and usually is performed using the singular value decomposition (SVD). Now, note that the size of matrix $Q^*A\Gamma^{(t)}A^*Q$ in SPL is $(m-r) \times (m-r)$ compared to $m \times m$ for $A\Gamma^{(t)}A^*$, which reduces the cost of matrix inversion for SPL compared to M-SBL. In particular, for the case of MUSIC where $m = k + 1$ and $r = k$, matrix inversion is not necessary for SPL whereas M-SBL still requires the $m \times m$ matrix inversion.

Finally, note that the hyper-parameter γ is closely related to spectral estimation. For example, for the case of MUSIC where $m = k + 1$ and $r = k$, the term $(Q^*A\Gamma^{(t)}A^*Q)$ in (44) reduces a scalar and we have

$$\gamma_i^{(t)} = \left(\frac{\frac{1}{N} \|\mathbf{x}^{(t)i}\|^2}{\mathbf{a}_i^* Q (Q^* A \Gamma^{(t)} A^* Q)^{-1} Q^* \mathbf{a}_i} \right)^{\frac{1}{2}} \quad (45)$$

$$= \frac{1}{\sqrt{\mathbf{a}_i^* Q Q^* \mathbf{a}_i}} \times \frac{\|\mathbf{x}^{(t)i}\|}{\sqrt{N(Q^* A \Gamma^{(t)} A^* Q)}} \quad (46)$$

where the first term is the MUSIC spectrum and the second term is related to the magnitude of the i -th row of $X^{(t)}$. Hence, in the case of full-row rank X (i.e., the MUSIC case), SPL can be regarded as an algorithm that initialises the non-zero support estimation using a spectral estimation technique, followed by alternating modification using the data fidelity matching criterion.

V. NUMERICAL RESULTS

In this section, we perform extensive numerical experiments to validate the proposed algorithm under various experimental conditions, and compare it with respect to existing joint sparse recovery algorithms. In particular, we are interested in the SPL algorithm in the asymptotic region of $p \rightarrow 0$ since it directly minimises the rank of Q^*A_I .

The elements of a sensing matrix A were generated from a Gaussian distribution with zero mean and variance of $1/m$, and then each column of A was normalized to have a unit norm. An unknown signal X_* with $\text{rank}(X_*) = r \leq k$ was generated using the same procedure as in [6]. Specifically, we randomly generated a support I , and then the corresponding nonzero signal components were obtained by

$$X_*^I = U \Sigma V, \quad (47)$$

where $U \in \mathbb{R}^{k \times r}$ was set to random orthonormal columns, and $\Sigma = \text{diag}([\sigma_i]_{i=1}^r)$ is a diagonal matrix

whose i -th element is given by

$$\sigma_i = \tau^i, \quad 0 < \tau < 1, \quad (48)$$

and $V \in \mathbb{R}^{r \times N}$ was generated using Gaussian random distribution with zero mean and variance of $1/N$. After generating noiseless data, we added zero mean white Gaussian noise. We declared success if an estimated support from a certain algorithm was the same as a true $\text{supp}X$.

As the proposed algorithm does not require a prior knowledge of the sparsity level, we need to define a stopping criterion. Here, the stopping criterion is defined by monitoring the normalized change in the variable γ :

$$\frac{\|\gamma^{(t)} - \gamma^{(t-1)}\|_2}{\|\gamma^{(t)}\|_2} < 10^{-3}.$$

From our experiments, usually 20-30 iterations are required for SPL to converge.

A. Local Minima Property

We first perform experiments to confirm that SPL produces the true solution under milder conditions than M-SBL. To show this, using k -sparse signal X_* generated by (47) with $\tau = 1$, we produced measurements $Y = AX_*$ such that $r := \text{rank}(Y) \leq k$. Then, we initialized both algorithms with $X^{(0)}$ that satisfies the following:

$$Y = AX^{(0)}, \quad \|X^{(0)}\|_0 = m, \quad |\text{supp}X_* \cap \text{supp}X^{(0)}| = s, \quad (49)$$

where $s < k - r$, $s = k - r$ and $s > k - r$, respectively. Note that the initialization corresponds to a local minimiser and we are interested in confirming that SPL can escape from the local minimizers thanks to the update in (42). Recall that it is difficult for M-SBL to avoid this type of local minimizers since $X^{(0)}$ has zeros rows at the i -th row where $i \in S \setminus \text{supp}X^{(0)}$ and the M-SBL update rule in (11) cannot make the corresponding γ_i nonzero in the subsequent iterations.

Figs. 1(a)-(c) illustrate the perfect recovery ratio from the initialization using SPL and M-SBL at various SNR conditions for (a) $s = k - r = 3$, (b) $s = 2 < k - r$, and (c) $s = 5 > k - r$, respectively. The results clearly demonstrate that SPL finds the global minimizer nearly perfectly, whereas M-SBL fails most of the time. This clearly confirms the theoretical advantages for SPL.

B. Comparison with Other State-of-Art Algorithms

To compare the proposed algorithm with various state-of-art joint sparse recovery methods, the recovery rates of various state-of-art joint sparse recovery algorithms such as MUSIC, S-OMP, SA-MUSIC,

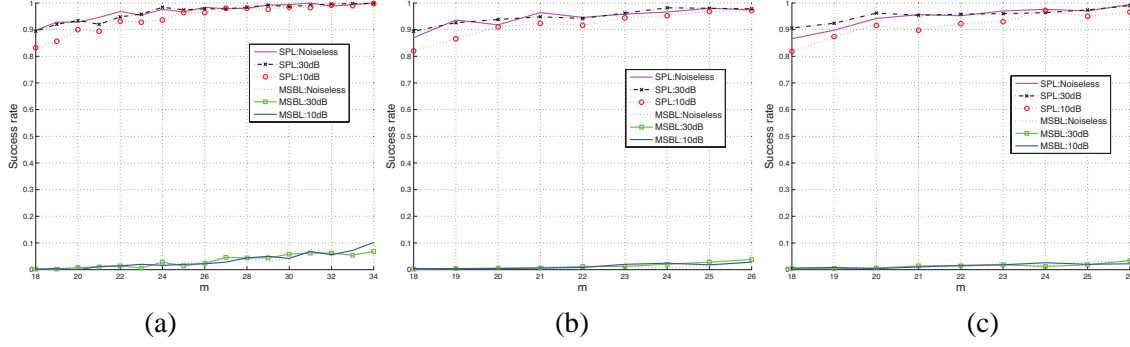


Fig. 1. Perfect recovery ratio from initialization using local minimizer that satisfy (49). The results are averaged after 500 runs and the simulation parameters are: $k = 10, r = 7, n = 128$ and $N = 64$. (a) $s = k - r = 3$, (b) $s = 2 < k - r$, and (c) $s = 5 > k - r$, respectively.

sequential CS-MUSIC, M-SBL, and the l_1/l_2 mixed norm approach are plotted in Fig. 2 along with those of SPL. Among the various implementation of mixed norm approaches, we used high performance SGPL1 software [22], which can be downloaded from <http://www.cs.ubc.ca/labs/scl/spgl1/>. Since M-SBL, the mixed norm approach, as well as SPL do not provide an exact k -sparse solution, we used the support for the largest k coefficients as a support estimate in calculating the perfect recovery ratio. For MUSIC, S-OMP, SA-MUSIC, sequential CS-MUSIC, we assume that k is known. For subspace based algorithms such as MUSIC, SA-MUSIC, sequential CS-MUSIC as well as SPL, we determine the signal subspace using the following criterion

$$\max_{i \in \{1, \dots, m\}} \frac{\sigma_i - \sigma_{i+1}}{\sigma_i - \sigma_m} > 0.1,$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m$ denotes the singular values of YY^* . A theoretical motivation for such subspace determination is given in [6]. Here, the success rates were averaged over 1000 experiments. The simulation parameters were as follows: $m \in \{1, 2, \dots, 50\}$, $n = 128$, $k = 8$, $r = 5$, $SNR = 30dB, 10dB$, and $N \in \{32, 128\}$, respectively. Figs. 2(a)-(d) illustrates the comparison results under various snapshot number and SNR conditions. Note that SPL consistently outperforms all other algorithms at various snapshots numbers. In particular, the gain increases with increasing number of snapshots, since it provides better subspace estimation. Also, note that SPL consistently outperforms M-SBL at all SNR ranges. Figs. 2(e)(f) illustrates that SPL significantly outperforms M-SBL when X is badly conditioned. Moreover, as the subspace estimation becomes accurate with increasing N , the performance gain becomes more significant.

Figs. 3(a)(b)(c) compares the performance of various MMV algorithm by varying the sparsity

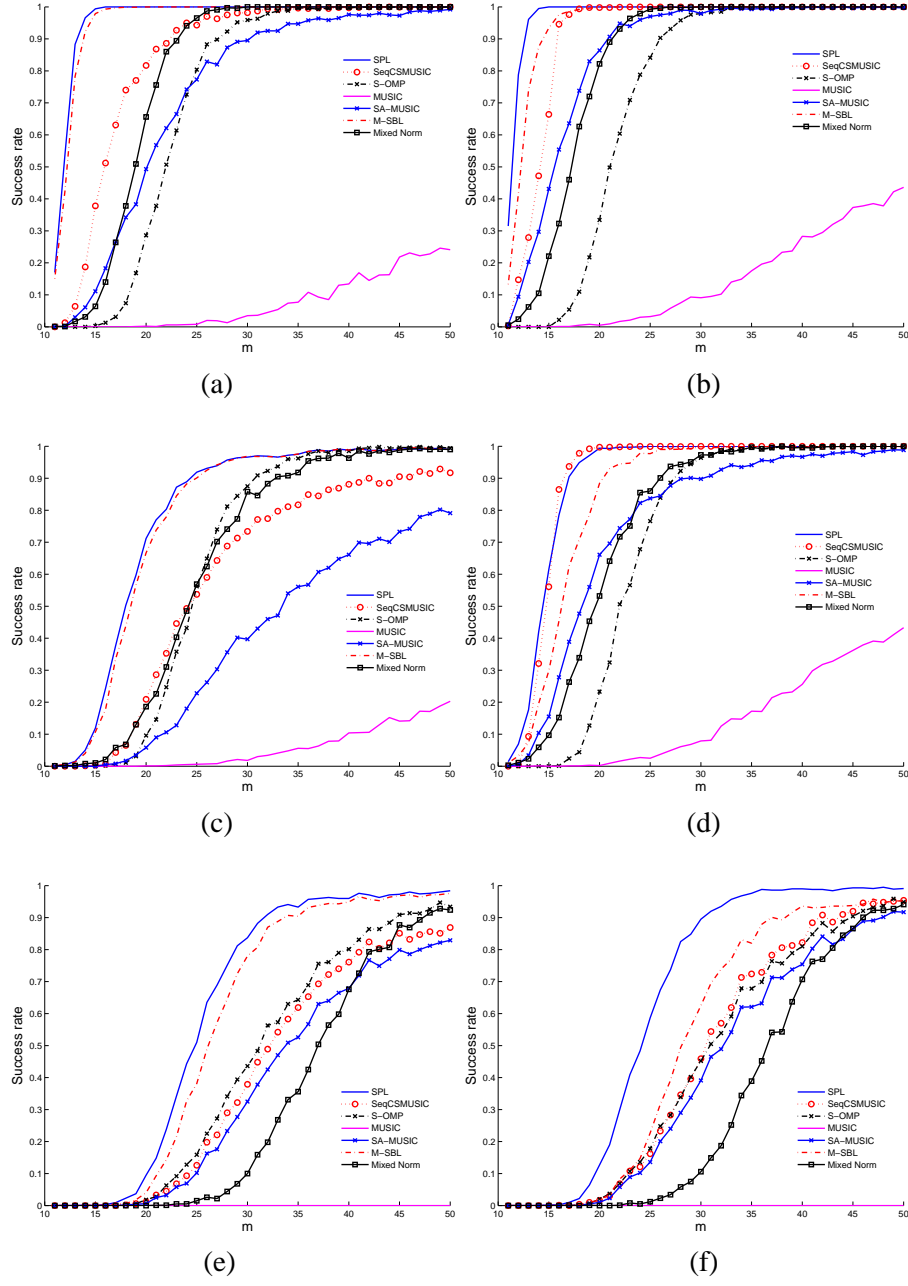


Fig. 2. Performance of various joint sparse recovery algorithms at $n = 128, k = 10, r = 6$ when (a) $SNR = 30dB, N = 16, \tau = 1$, (b) $SNR = 30dB, N = 256, \tau = 1$, (c) $SNR = 10dB, N = 16, \tau = 1$, (d) $SNR = 10dB, N = 256, \tau = 1$, (e) $SNR = 30dB, N = 16, \tau = 0.1$, (f) $SNR = 30dB, N = 256, \tau = 0.1$, respectively.

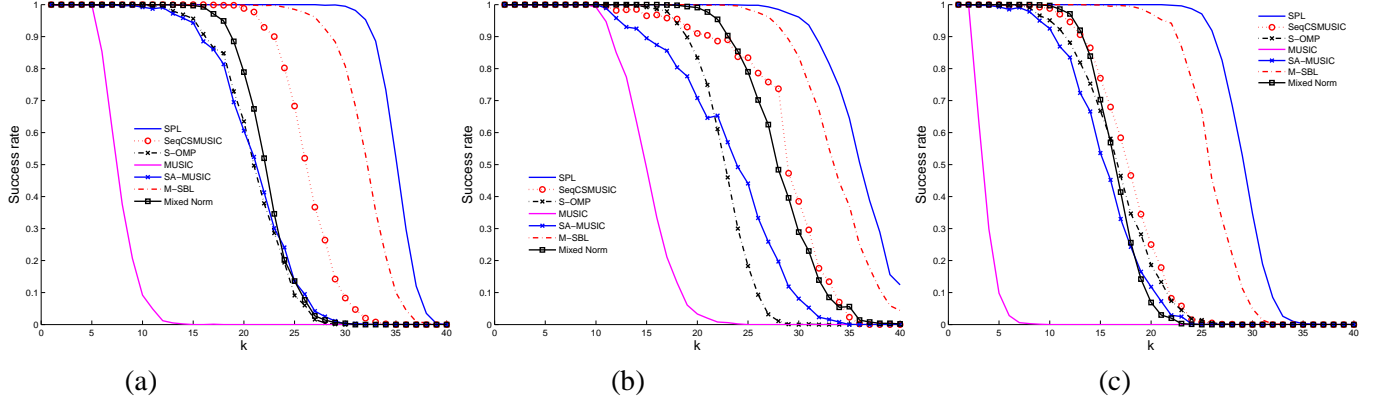


Fig. 3. Performance of various joint sparse recovery algorithms for varying sparsity level at $N = 256$. The simulation parameters are (a) $m = 40, r = 5, \tau = 1$ and $\text{SNR}=30\text{dB}$, and (b) $m = 40, r = 12, \tau = 1$ and $\text{SNR}=10\text{dB}$, and (c) $m = 40, r = 15, \tau = 0.5$ and $\text{SNR}=30\text{dB}$, respectively.

level. Here, m and $\text{rank}(Y)$ are fixed and the sparsity levels changes, and we calculated the perfect reconstruction ratio. Again, SPL outperforms all existing methods for various SNR and conditions numbers.

C. Fourier Measurements Cases

Fig. 4 illustrates the results of the comparison when the measurement are from Fourier sensing matrix. Similar to Gaussian sensing matrix, consistent improvement of SPL over M-SBL and other algorithms under various conditions have been observed.

VI. CONCLUSION

In this paper, we derived a new MMV algorithm called subspace penalized sparse learning (SPL) to address a joint sparse recovery problem, in which the unknown signals share a common non-zero support. The SPL algorithm was inspired by the observation that the $\log \det(\cdot)$ term in M-SBL is a rank proxy for a partial sensing matrix, and similar rank criteria exist in subspace-based greedy MMV algorithms like CS-MUSIC and SA-MUSIC. Furthermore, we proved that instead of $\text{rank}(A\Gamma^{1/2})$, minimizing $\text{rank}(Q^*A\Gamma^{1/2})$ is a more direct way of imposing joint sparsity since its global minimizer can provide the true joint support. To impose such a subspace constraint as a penalty, the SPL algorithm employs the Schatten- p quasi norm rank penalty and was implemented as an alternating minimisation method. Theoretical analysis showed that as $p \rightarrow 0$, the global minimizer of the SPL is equivalent to the global minimiser of the l_0 MMV solution. We further demonstrated that compared to M-SBL, our SPL is

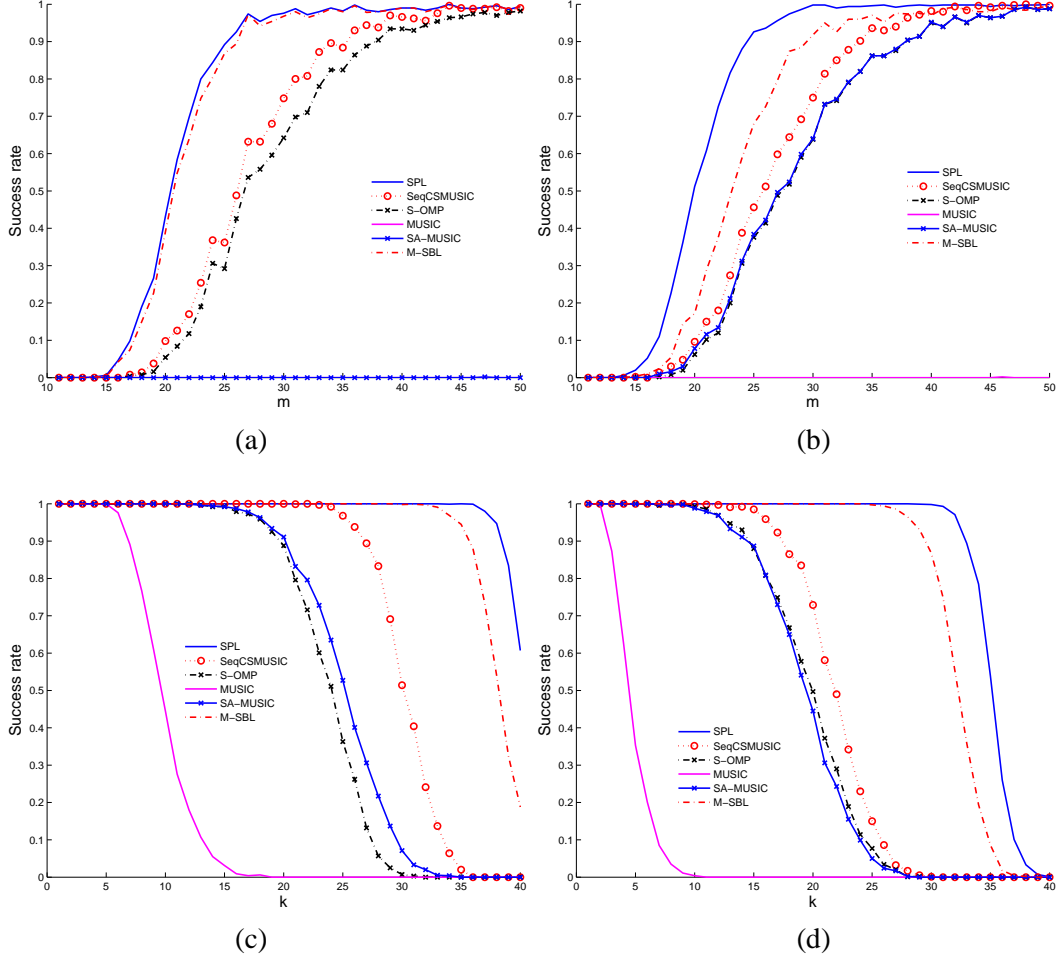


Fig. 4. Performance of various joint sparse recovery algorithms at $n = 128$ when the sensing matrix is from Fourier matrix and $\text{SNR}=30\text{dB}$. The simulation parameters are (a) $r = 8, N = 16, \tau = 1, k = 10$, (b) $r = 8, N = 256, \tau = 1, k = 10$, (c) $r = 5, m = 40, N = 256, \tau = 1$, (d) $r = 15, m = 40, N = 256, \tau = 0.5$, respectively.

more robust to recovering badly conditioned X_* . With numerical simulations, we demonstrated that SPL consistently outperforms all existing state-of-the-art algorithms including M-SBL.

ACKNOWLEDGEMENTS

This work was supported by Korea Science and Engineering Foundation under Grant NRF-2014R1A2A1A11052491.

REFERENCES

- [1] J. Kim, O. Lee, and J. Ye, "Compressive MUSIC: revisiting the link between compressive sensing and array signal processing," *IEEE Trans. on Information Theory*, vol. 58, no. 1, pp. 278–301, 2012.

- [2] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple measurement vectors," *IEEE Trans. on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.
- [3] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. on Signal Processing*, vol. 53, no. 7, p. 2477, 2005.
- [4] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. on Signal Processing*, vol. 56, pp. 4692–4702, 2009.
- [5] E. Berg and M. P. Friedlander, "Theoretical and empirical results for recovery from multiple measurements," *IEEE Trans. on Information Theory*, vol. 56, no. 5, pp. 2516–2527, 2010.
- [6] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. on Information Theory*, vol. 58, no. 6, pp. 3613–3641, 2012.
- [7] P. Feng, "Universal minimum-rate sampling and spectrum-blind reconstruction for multiband signals," Ph.D. Dissertation, University of Illinois, Urbana-Champaign, 1997.
- [8] J. Tropp, A. Gilbert, and M. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.
- [9] Y. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. on Signal Processing*, vol. 58, pp. 3042–3054, 2010.
- [10] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
- [11] G. Obozinski, M. Wainwright, and M. Jordan, "Support union recovery in high-dimensional multivariate regression," *The Annals of Statistics*, vol. 39, no. 1, pp. 1–47, 2011.
- [12] D. Wipf and B. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. on Signal Processing*, vol. 55, no. 7 Part 2, pp. 3704–3716, 2007.
- [13] J. M. Kim, O. K. Lee, and J. C. Ye, "Improving noise robustness in subspace-based joint sparse recovery," *IEEE Trans. on Signal Processing (in press)*, 2012.
- [14] D. Wipf, B. Rao, and S. Nagarajan, "Latent variable bayesian models for promoting sparsity," *IEEE Trans. on Information Theory*, vol. 57, no. 9, p. 6236, 2011.
- [15] D. Wipf, "Bayesian methods for finding sparse representations," Ph.D. dissertation, University of California, San Diego, 2006.
- [16] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*. New York: Wiley-Interscience, 1996.
- [17] K. Mohan and M. Fazel, "Iterative reweighted least squares for matrix rank minimization," in *48th IEEE Annual Allerton Conference on Communication, Control, and Computing*, 2010, pp. 653–661.
- [18] M. Fazel, H. Hindi, and S. Boyd, "Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices," in *American Control Conference, 2003. Proceedings of the 2003*, vol. 3. IEEE, 2003, pp. 2156–2162.
- [19] K. Petersen and M. Pedersen, "The matrix cookbook," *Technical University of Denmark*, pp. 7–15, 2008.
- [20] P. Tao and L. An, "Convex analysis approach to dc programming: Theory, algorithms and applications," *Acta Mathematica Vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.
- [21] P. Tao and L. T. H. An, "A DC optimization algorithm for solving the trust-region subproblem," *SIAM Journal on Optimization*, vol. 8, no. 2, pp. 476–505, 1998.
- [22] E. Van Den Berg and M. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, p. 890, 2008.