Support Recovery for Sparse Recovery and Non-stationary Blind Demodulation

Youye Xie, Michael B. Wakin, and Gongguo Tang
Department of Electrical Engineering
Colorado School of Mines
Golden, USA
{youyexie,mwakin,gtang}@mines.edu

***Abstract*—The problem of estimating a sparse signal from its modulated low-dimensional observations appears naturally in many applications like signal deconvolution and self-calibration. In the presence of noise, however, the sparse signal and modulation parameters cannot be exactly recovered. Therefore, in this paper, we study the support recovery problem for the sparse recovery and non-stationary blind demodulation model, in which each dictionary atom composing the low-dimension observations undergoes a distinct modulation process. Using the lifting technique and with the assumption that the modulating signals live in a common subspace, we reformulate this problem as one of recovering a column-wise sparse matrix from structured linear observations. We propose to solve the reformulated problem via block ℓ_1 norm regularized quadratic minimization. We derive sufficient conditions on the sample complexity and regularization parameter such that the support of the ground truth sparse signal can be exactly recovered with overwhelming probability, and we bound the recovery error of the sparse signal and modulation process on the support. Simulations illustrate and support our theoretical results.**

***Index Terms*—Support recovery, blind demodulation, sparse matrix recovery, group lasso, compressive sensing**

I. Introduction

A. Motivation and Problem Formulation

The problem of estimating a sparse signal from its modulated low-dimensional observations arises naturally in many applications including non-stationary super-resolution [1] and self-calibration [2]. Mathematically, suppose a system observes $\boldsymbol{y} = \mathbf{D}\mathbf{A}\boldsymbol{c} + \boldsymbol{n} \in \mathbf{C}^N$, where $\mathbf{D} \in \mathbf{C}^N$ is a diagonal matrix, $\mathbf{A} \in \mathbf{C}^{N\times M}(N < M)$ is a dictionary matrix, $\boldsymbol{c} \in \mathbf{C}^M$ is a sparse signal vector, and $\boldsymbol{n}$ is the additive noise. Recovering a sparse signal $\boldsymbol{c}$ from a known but under-determined system of equations has been studied by the compressive sensing community [3], [4], [5]. In this scenario, $\mathbf{D}$ performs the element-wise multiplication (also known as modulation in signal processing) and is assumed to be unknown. Therefore, assuming $\mathbf{A}$ is known, recovering $\boldsymbol{c}$ and $\mathbf{D}$ from $\boldsymbol{y}$ is referred to as simultaneous sparse recovery and blind demodulation [2], [6].

In this paper, we actually consider a more general model such that each signal atom undergoes a distinct modulation process. In this case, the observations can be represented as

$$\boldsymbol{y} = \sum_{j=1}^{M} c_j \mathbf{D}_j \boldsymbol{a}_j + \boldsymbol{n} \in \mathbf{C}^N, \tag{I.1}$$

where c_j is the j-th entry of $\boldsymbol{c}$ and $\boldsymbol{a}_j$ is the j-th column of the dictionary matrix $\mathbf{A}$. To make this problem well-posed, we assume that at most J of the coefficients c_j are non-zero and also that the modulation signals live in a common subspace:

$$\mathbf{D}_j = \mathrm{diag}(\mathbf{B}\boldsymbol{h}_j), \tag{I.2}$$

where $\mathbf{B} \in \mathbf{C}^{N\times K}(N > K)$ is a known, orthonormal subspace matrix and $\boldsymbol{h}_j$ is the corresponding coefficient vector. Thus, recovering $\boldsymbol{h}_j$ is equivalent to recovering $\mathbf{D}_j$. A similar subspace assumption can be found in the deconvolution and demixing literature [7], [8].

According to Proposition 1 in [9], by using the lifting technique, we can construct a column-wise sparse matrix $\mathbf{X} = [c_1\boldsymbol{h}_1 \quad c_2\boldsymbol{h}_2 \quad \cdots \quad c_M\boldsymbol{h}_M] \in \mathbf{C}^{K\times M}$ which contains all the unknown parameters. When $\boldsymbol{n} = 0$, the observation (I.1) can be represented as

$$\boldsymbol{y}(n) = \boldsymbol{b}_n'^H \mathbf{X} \boldsymbol{a}_n'$$

where $\boldsymbol{b}_n'$ and $\boldsymbol{a}_n'$ are the n-th column of $\mathbf{B}^H$ and $\mathbf{A}^T$ respectively. We use $\boldsymbol{y} = \mathcal{L}(\mathbf{X})$ to denote this linear process and thus (I.1) takes the equivalent form $\boldsymbol{y} = \mathcal{L}(\mathbf{X}) + \boldsymbol{n}$. We denote the ground truth matrix as $\mathbf{X}_0$, and to estimate $\mathbf{X}_0$ from $\boldsymbol{y}$, we propose to adopt the block ℓ_1 ($\ell_{2,1}$) norm regularized quadratic minimization (also known as the group lasso in the statistics literature [10], [11]) as follows:

$$\underset{\mathbf{X}\in\mathbf{C}^{K\times M}}{\text{minimize}} \frac{1}{2}||\boldsymbol{y} - \mathcal{L}(\mathbf{X})||_2^2 + \lambda||\mathbf{X}||_{2,1}, \tag{I.3}$$

where the $\ell_{2,1}$ norm is defined as $||\mathbf{A}||_{2,1} = \sum_{j=1}^{M} ||\boldsymbol{a}_j||_2$. Note that due to the random noise, we can not recover the ground truth $\mathbf{X}_0$ perfectly. Instead, we aim to determine the indices of the non-zero columns in $\mathbf{X}_0$ which indicate the support of the sparse signal $\boldsymbol{c}$ when there is no trivial null modulation (namely, all $\mathbf{D}_j \neq 0$). We also aim to bound the recovery error of $\mathbf{X}_0$ in terms of the $\ell_{2,\infty}$ norm, which is defined as $||\mathbf{A}||_{2,\infty} = \max_j ||\boldsymbol{a}_j||_2$. In addition, note that (I.3) has an equivalent form

$$\underset{\mathbf{X}\in\mathbf{C}^{K\times M}}{\text{minimize}} \frac{1}{2}||\boldsymbol{y} - \boldsymbol{\Phi}\cdot\mathrm{vec}(\mathbf{X})||_2^2 + \lambda\sum_{i=1}^{M}||\boldsymbol{x}_i||_2 \tag{I.4}$$

where $\boldsymbol{x}_i$ is the i-th column of $\mathbf{X}$ and $\boldsymbol{\Phi}\cdot\mathrm{vec}(\mathbf{X}) = \mathcal{L}(\mathbf{X})$ is the matrix multiplication form of the linear operator. Specifically,

$$\boldsymbol{\Phi} = [\phi_{1,1} \quad \cdots \quad \phi_{K,1} \quad \cdots \quad \phi_{1,M} \quad \cdots \quad \phi_{K,M}] \in \mathbf{C}^{N\times KM}$$

where $\phi_{i,j} = \text{diag}(\boldsymbol{b}_i)\boldsymbol{a}_j \in \mathbf{C}^{N\times 1}$ and $\boldsymbol{b}_i$ is the i-th column of $\mathbf{B}$.

B. Contributions

The contributions are twofold. First, we study the support recovery problem of the general sparse recovery and non-stationary blind demodulation model and adopt the $\ell_{2,1}$ norm regularized quadratic minimization to solve it. Second, we show that, for exact support recovery, the sufficient number of observations, N, is proportional to the number of degrees of freedom, $O(JK)$, and that the regularization parameter, λ, should be proportional to the standard deviation of the random noise. We further bound the recovery error in terms of the $\ell_{2,\infty}$ norm.

C. Related Work

Due to the signal model that we consider, our linear operator has a special block structure with randomness which distinguishes our work from the existing works in group lasso literature [10], [11], [12]. For instance, [11] requires each block of $\boldsymbol{\Phi}$ to be orthonormal. [13], [14] consider a deterministic $\boldsymbol{\Phi}$, and [15] allows $\boldsymbol{\Phi}$ to have independent sub-exponential rows (which is different from our operator) and bounds the recovery error in terms of the ℓ_2 norm instead of the $\ell_{2,\infty}$ norm.

Most of the sparse recovery and blind demodulation literature considers a common modulation process [2], [8], [16], [17], [18], in which $\mathbf{D}_j$ are the same for all dictionary atoms. [8] studies a spikes deconvolution problem and its dictionary consists of complex sinusoids over a continuous frequency range. [1] extends the work of [8] to incorporate non-stationary modulation. However, they still study the complex sinusoids dictionary and have a random 'sign' assumption on $\boldsymbol{h}_j$ which makes it challenging to consider the noise. Moreover, [2] considers the random Gaussian and Fourier dictionaries but still requires all $\mathbf{D}_j$ to be the same. [7] generalizes the model in [2] but needs the knowledge of the number of source signals. Furthermore, we have studied the same model but with bounded additive noise in [6], [9]. In those works, we considered a constrained $\ell_{2,1}$ norm minimization problem instead of group lasso and did not study the support recovery problem.

The rest of the paper is organized as follows. In Section II, we present our main theorem regarding the sample complexity and regularization parameter for support recovery, and we bound the recovery error. Numerical simulations are conducted in Section III to illustrate the important scaling relationships. Section IV is devoted to the conclusion.

II. Main Result

In this section, we present the main theorem of this paper, which provides the sufficient conditions on sample complexity and regularization parameter for support recovery and bounds the recovery error, when solving (I.3) (or equivalently (I.4)). We refer to a matrix whose entries follow the i.i.d standard normal distribution as a random Gaussian matrix.

Theorem II.1. *Consider the observation model in equation (I.1), assume that $\mathbf{A} \in \mathbf{R}^{N\times M}$ $(N < M)$ is a random Gaussian matrix, at most J $(< M)$ coefficients c_j are nonzero, and the real and imaginary parts of each entry of the noise vector $\boldsymbol{n} \in \mathbf{C}^{N\times 1}$ follow the i.i.d Gaussian distribution with 0 mean and σ^2 variance. Suppose also that each modulation matrix $\mathbf{D}_j$ satisfies the subspace constraint* (I.2)*, where $\mathbf{B}^H\mathbf{B} = \mathbf{I}_K$. If the number of observations*

$$N \geq C_{\alpha,1}\mu_{max}^2 KJ\left[\log(M-J) + \log^2(N)\right] \tag{II.1}$$

and the regularization parameter

$$\lambda \geq \sqrt{C_{\alpha,2}\sigma^2\mu_{max}^2 K\left[\log(M-J) + \log(N)\right]} \tag{II.2}$$

where $C_{\alpha,1}$ and $C_{\alpha,2}$ are constants that grow linearly with $\alpha > 1$ and the coherence parameter

$$\mu_{max} = \max_{i,j} \sqrt{N}|\mathbf{B}_{ij}|,$$

then the following properties hold with probability at least $1 - O(N^{-\alpha+1})$:

1) *Problem* (I.4) *has a unique solution $\hat{\mathbf{X}} \in \mathbf{C}^{K\times M}$ with its support, the set of indices of the non-zero columns in $\hat{\mathbf{X}}$, contained within the support T of the ground truth solution, $\mathbf{X}_0$.*
2) *The recovery error between the solution, $\hat{\mathbf{X}}$, and the ground truth, $\mathbf{X}_0$, satisfies*

$$||\hat{\mathbf{X}} - \mathbf{X}_0||_{2,\infty} \leq \sqrt{C_\alpha\sigma^2\mu_{max}^2 JK\left[\log(J) + \log(N)\right]} + 4\sqrt{J}\lambda \tag{II.3}$$

where C_α is a constant that grows linearly with α. If in addition the non-zero columns of $\mathbf{X}_0$ are bounded below

$$\min_{j\in T} ||\boldsymbol{x}_{0,j}||_2 > \sqrt{C_\alpha\sigma^2\mu_{max}^2 JK\left[\log(J) + \log(N)\right]} + 4\sqrt{J}\lambda, \tag{II.4}$$

then $\hat{\mathbf{X}}$ and $\mathbf{X}_0$ have exactly the same support which implies exact support recovery.

From (II.3), we can obtain that for any $\hat{\boldsymbol{x}}_j = \hat{c}_j\hat{\boldsymbol{h}}_j$, which is the j-th columns of the solution $\hat{\mathbf{X}}$, and $\boldsymbol{x}_{0,j} = c_{0,j}\boldsymbol{h}_{0,j}$, which is the j-th columns of the ground truth $\mathbf{X}_0$, $||\hat{c}_j\hat{\mathbf{D}}_j - c_{0,j}\mathbf{D}_{0,j}||_F = ||\hat{c}_j\hat{\boldsymbol{h}}_j - c_{0,j}\boldsymbol{h}_{0,j}||_2 \leq \sqrt{C_\alpha\sigma^2\mu_{max}^2 JK\left[\log(J) + \log(N)\right]} + 4\sqrt{J}\lambda$. In addition, because $\mathbf{B}$ consists of orthonormal columns, $\mu_{max} \in [1, \sqrt{N}]$. And when all system parameters are fixed, (II.1) requires $1 \leq \mu_{max} \leq \sqrt{\frac{N}{C_{\alpha,1}KJ[\log(M-J)+\log^2(N)]}}$. Moreover, results in Theorem II.1 may be of interest outside the support recovery problem and to the group lasso problem with random block structured linear operators.

In the proof of Theorem II.1, we first derive the optimality and uniqueness conditions for the solution to (I.4) and then construct such a solution satisfying the optimality and uniqueness conditions via the primal-dual witness method [19]. We use $T := T(\mathbf{X}_0)$ with $|T| = J$ to denote the set containing the indices of the non-zero columns of the ground-truth matrix $\mathbf{X}_0$. The complement set of T is T^C. We define $\mathbf{\Phi}_T \in \mathbf{C}^{N\times KM}$ to be the same as $\mathbf{\Phi}$ in the $K(j-1)+1$ to $K(j-1)+K$-th columns for all $j \in T$ and zero otherwise. Removing those zero columns results in $\tilde{\mathbf{\Phi}}_T \in \mathbf{C}^{N\times KJ}$ which is a submatrix of $\mathbf{\Phi}$. Similarly, we define $\tilde{\mathbf{X}}_T \in \mathbf{C}^{K\times J}$ to be a submatrix of $\mathbf{X}$ but only contains the $j \in T$ columns.

A. Optimality and Uniqueness Conditions

Lemma II.1.

1) *A matrix $\hat{\mathbf{X}} \in \mathbf{C}^{K\times M}$ is an optimal solution to* (I.4) *if and only if there exists a subgradient vector* $\boldsymbol{s} = \begin{bmatrix} \boldsymbol{s}_1 \\ \vdots \\ \boldsymbol{s}_M \end{bmatrix} \in vec\left(\partial\|\hat{\mathbf{X}}\|_{2,1}\right)$, *such that*

$$\mathbf{\Phi}^H\mathbf{\Phi}\cdot vec(\hat{\mathbf{X}}) - \mathbf{\Phi}^H\boldsymbol{y} + \lambda\cdot\begin{bmatrix} \boldsymbol{s}_1 \\ \vdots \\ \boldsymbol{s}_M \end{bmatrix} = \mathbf{0}$$

which is equivalent to

$$\mathbf{\Phi}^H\mathbf{\Phi}\cdot\left(vec(\hat{\mathbf{X}}) - vec(\mathbf{X}_0)\right) - \mathbf{\Phi}^H\boldsymbol{n} + \lambda\boldsymbol{s} = \mathbf{0} \tag{II.5}$$

where $\boldsymbol{s}_i \in \mathbf{C}^K$ is the subgradient of $\|\cdot\|_2$ at $\hat{\boldsymbol{x}}_i$ defined as

$$\boldsymbol{s}_i = \begin{cases} \frac{\hat{\boldsymbol{x}}_i}{\|\hat{\boldsymbol{x}}_i\|_2} & \text{if } \|\hat{\boldsymbol{x}}_i\|_2 \neq 0; \\ \{\boldsymbol{z} : \|\boldsymbol{z}\|_2 \leq 1\} & \text{if } \|\hat{\boldsymbol{x}}_i\|_2 = 0. \end{cases}$$

2) *If the subgradient vectors of the optimal solution $\hat{\mathbf{X}}$ satisfy $\|\boldsymbol{s}_i\|_2 < 1$ for all $i \notin T(\hat{\mathbf{X}})$, then any optimal solution, $\check{\mathbf{X}}$, to* (I.4) *satisfies $\check{\boldsymbol{x}}_i = 0$ for all $i \notin T(\hat{\mathbf{X}})$.*
3) *When conditions in (2) are satisfied, if in addition $\tilde{\mathbf{\Phi}}^H_{T(\hat{\mathbf{X}})}\tilde{\mathbf{\Phi}}_{T(\hat{\mathbf{X}})} \in \mathbf{C}^{KJ\times KJ}$ is invertible, then $\hat{\mathbf{X}}$ is the unique solution to* (I.4).

B. Primal-Dual Witness Construction

1) Conditioned on $\tilde{\mathbf{\Phi}}^H_T\tilde{\mathbf{\Phi}}_T \in \mathbf{C}^{KJ\times KJ}$ is invertible, we first obtain $\tilde{\tilde{\mathbf{X}}}_T \in \mathbf{C}^{K\times J}$ by solving the support restricted problem

$$\tilde{\tilde{\mathbf{X}}}_T = \arg\min_{\mathbf{X}\in\mathbf{C}^{K\times J}}\left\{\frac{1}{2}\|\boldsymbol{y} - \tilde{\mathbf{\Phi}}_T\cdot\mathrm{vec}(\mathbf{X})\|_2^2 + \lambda\|\mathbf{X}\|_{2,1}\right\}.$$

The solution $\tilde{\tilde{\mathbf{X}}}_T$ is unique under the invertibility condition on $\tilde{\mathbf{\Phi}}^H_T\tilde{\mathbf{\Phi}}_T$. And we set $\tilde{\tilde{\mathbf{X}}}_{T^C} \in \mathbf{C}^{K\times(M-J)} = \mathbf{0}$. Thus, $\hat{\mathbf{X}}$ has support contained within the support T of the ground truth solution $\mathbf{X}_0$.

2) We calculate the subgradient vector $\tilde{\boldsymbol{s}}_T \in \mathbf{C}^{JK}$ based on $\tilde{\tilde{\mathbf{X}}}_T$, where $\tilde{\boldsymbol{s}}_T$ is a sub-vector of $\boldsymbol{s}$ consisting of $\boldsymbol{s}_j$ for all $j \in T$.
3) We solve for a vector $\tilde{\boldsymbol{s}}_{T^C} \in \mathbf{C}^{(M-J)K}$ satisfying (II.5) and check whether $\|\boldsymbol{s}_i\|_2 < 1$ for all $i \notin T$.

The primal-dual witness construction succeeds only if the problem (I.4) has a unique solution $\hat{\mathbf{X}}$ whose support is contained within the support of the ground truth $\mathbf{X}_0$. And without loss of generality, we assume the support of $\mathbf{X}_0$ is the first J columns so that $T = \{1, 2, ..., J\}$. Then (II.5) can be written into a matrix multiplication form

$$\begin{bmatrix} \tilde{\mathbf{\Phi}}^H_T\tilde{\mathbf{\Phi}}_T & \tilde{\mathbf{\Phi}}^H_T\tilde{\mathbf{\Phi}}_{T^C} \\ \tilde{\mathbf{\Phi}}^H_{T^C}\tilde{\mathbf{\Phi}}_T & \tilde{\mathbf{\Phi}}^H_{T^C}\tilde{\mathbf{\Phi}}_{T^C} \end{bmatrix}\begin{bmatrix} \mathrm{vec}(\tilde{\tilde{\mathbf{X}}}_T) - \mathrm{vec}(\tilde{\mathbf{X}}_{0,T}) \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \tilde{\mathbf{\Phi}}^H_T \\ \tilde{\mathbf{\Phi}}^H_{T^C} \end{bmatrix}\boldsymbol{n} + \lambda\begin{bmatrix} \tilde{\boldsymbol{s}}_T \\ \tilde{\boldsymbol{s}}_{T^C} \end{bmatrix} = 0. \tag{II.6}$$

When $\tilde{\mathbf{\Phi}}^H_T\tilde{\mathbf{\Phi}}_T$ is invertible, (II.6) implies that

$$\Delta(\mathbf{X}) = \mathrm{vec}(\tilde{\tilde{\mathbf{X}}}_T) - \mathrm{vec}(\tilde{\mathbf{X}}_{0,T}) = (\tilde{\mathbf{\Phi}}^H_T\tilde{\mathbf{\Phi}}_T)^{-1}\left(\tilde{\mathbf{\Phi}}^H_T\boldsymbol{n} - \lambda\tilde{\boldsymbol{s}}_T\right)$$

and

$$\tilde{\boldsymbol{s}}_{T^C} = \frac{1}{\lambda}\left(\tilde{\mathbf{\Phi}}^H_{T^C}\boldsymbol{n} - \tilde{\mathbf{\Phi}}^H_{T^C}\tilde{\mathbf{\Phi}}_T\Delta(\mathbf{X})\right). \tag{II.7}$$

Substituting $\Delta(\mathbf{X})$ into (II.7) yields

$$\tilde{\boldsymbol{s}}_{T^C} = \tilde{\mathbf{\Phi}}^H_{T^C}\left(\mathbf{I}_N - \tilde{\mathbf{\Phi}}_T(\tilde{\mathbf{\Phi}}^H_T\tilde{\mathbf{\Phi}}_T)^{-1}\tilde{\mathbf{\Phi}}^H_T\right)\frac{\boldsymbol{n}}{\lambda} + \tilde{\mathbf{\Phi}}^H_{T^C}\tilde{\mathbf{\Phi}}_T(\tilde{\mathbf{\Phi}}^H_T\tilde{\mathbf{\Phi}}_T)^{-1}\tilde{\boldsymbol{s}}_T.$$

The rest of the proof involves deriving the conditions on the sample complexity N and regularization parameter λ such that $\|\tilde{\boldsymbol{s}}_{T^C}\|_{2,\infty} = \max_{i\in T^C}\|\boldsymbol{s}_i\|_2 < 1$ utilizing the tail bound property of the quadratic forms of random Gaussian vectors [20] and bounding the recovery error using the strongly convex property of the support restricted problem conditioned on $\tilde{\mathbf{\Phi}}^H_T\tilde{\mathbf{\Phi}}_T \in \mathbf{C}^{KJ\times KJ}$ being invertible. Full details of the proof are available in [21].

III. Numerical Simulations

We conduct several numerical simulations to demonstrate the important scaling relationships in Theorem II.1. Specifically, we set $\mathbf{B} \in \mathbf{C}^{N\times K}$ to be the first K columns of the normalized $N\times N$ DFT matrix. The real and imaginary parts of the entry of $c_j \in \mathbf{C}$ and $\boldsymbol{h}_j \in \mathbf{C}^{K\times 1}$ follow the i.i.d standard normal distribution. The support, T with $|T| = J$, of the ground truth solution $\mathbf{X}_0$ is sampled uniformly. If not explicitly stated otherwise, we use $\sigma = 0.1$, $J = K = 3$, $N = 100$, and $M = 150$. Problem (I.4) is solved via CVX [22]. To achieve exact support recovery, (II.2) gives a lower bound for λ and (II.4) provides an upper bound. To verify those bounds, we define $\gamma_0 = \sqrt{\sigma^2\mu^2_{max}K\left[\log(M-J) + \log(N)\right]}$ and $\gamma = \frac{\gamma_0}{\min_{j\in T}\|\boldsymbol{x}_{0,j}\|_2}$. When all system parameters except λ are fixed, based on (II.2) and (II.4), $\lambda = k\gamma_0$ should satisfy

$$C_1\gamma_0 \leq \lambda < \frac{\min_{j\in T}\|\boldsymbol{x}_{0,j}\|_2 - C_2}{C_3} \Rightarrow C_1 \leq k < \frac{C_4}{\gamma} - C_5$$

where C_is are constants. We change k and γ and the exact support recovery rate for different (k, γ) pairs is recorded in

Fig. 1. Fifty trials are run for each (k, γ) pair and we can observe that C_1 is about 1.2 in this setting and k and γ have an inverse relationship.

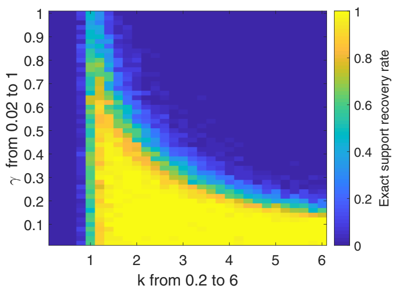
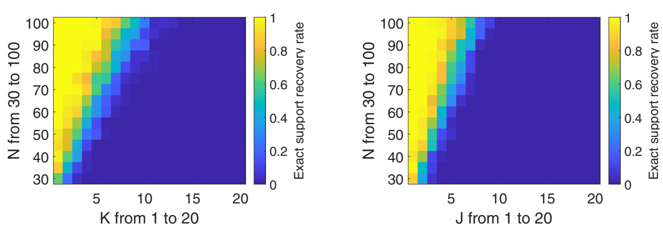


Fig. 1. The relation between k and γ in terms of the exact support recovery rate where $\lambda = k\gamma_0$ and $\gamma = \frac{\gamma_0}{\min_{j \in T} ||\boldsymbol{x}_{0,j}||_2}$.

To achieve exact support recovery, (II.1) shows that the sufficient number of observations, N, scales nearly linearly with respect to the subspace dimension K and the sparsity J. Thus, we record the exact support recovery rate with fixed $\gamma = 0.02$ and $k = 3$ and varying N and K ($J = 3$) in Fig. 2 (a). The result of a similar simulation changing the role of K and J is recorded in Fig. 2 (b). We run 50 simulations for each setting. The nearly linear scaling relations between N and K and N and J are consistent with the theoretical results.



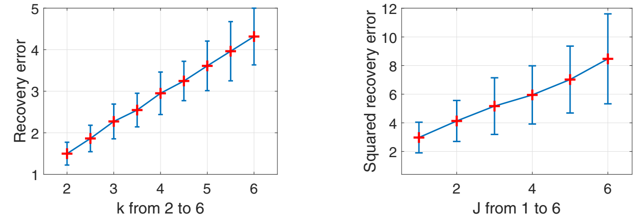
(a) (b)

Fig. 2. The nearly linear relation between the number of observations N and (a) the subspace dimension K and (b) the sparsity J for exact support recovery.

According to (II.3), the recovery error scales linearly with respect to λ and nearly linearly with respect to $\sqrt{J}$. To verify that, we set $\gamma = 0.02$ and vary k for $\lambda = k\gamma_0$. The recovery error, $||\hat{\mathbf{X}} - \mathbf{X}_0||_{2,\infty}$, when exact support recovery is achieved, is recorded in Fig. 3 (a). The squared recovery error, $||\hat{\mathbf{X}} - \mathbf{X}_0||_{2,\infty}^2$, with different values of J is recorded in Fig. 3 (b). 100 trials are run for each setting. In the results, we do observe the linear scaling of the recovery error with λ and nearly linear scaling of the squared recovery error with J.

IV. Conclusion

In this paper, we study the support recovery problem of the sparse recovery and non-stationary blind demodulation model. With a common subspace assumption on the modulation matrix, we recast this problem as the recovery problem of a column-wise sparse matrix and solve it via $\ell_{2,1}$ norm regularized quadratic minimization. We derive sufficient conditions



(a) (b)

Fig. 3. (a) The linear relation between the recovery error, $||\hat{\mathbf{X}} - \mathbf{X}_0||_{2,\infty}$, and the regularization parameter $\lambda = k\gamma_0$. (b) The nearly linear relation between the squared recovery error, $||\hat{\mathbf{X}} - \mathbf{X}_0||_{2,\infty}^2$, and the sparsity J. The red plus signs and the blue horizontal sticks indicate the mean and standard deviation of the corresponding error.

on the sample complexity and regularization parameter for exact support recovery and bound the recovery error in terms of the $\ell_{2,\infty}$ norm. Simulation results are consistent with the theoretical results.

Acknowledgment

This work was supported by NSF grant CCF-1704204.

References

[1] D. Yang, G. Tang, and M. B. Wakin, "Super-resolution of complex exponentials from modulations with unknown waveforms," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5809–5830, 2016.

[2] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, p. 115002, 2015.

[3] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[4] E. J. Candes and Y. Plan, "A probabilistic and ripless theory of compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.

[5] Y. Xie, S. Li, G. Tang, and M. B. Wakin, "Radar signal demixing via convex optimization," in *2017 22nd International Conference on Digital Signal Processing (DSP)*, pp. 1–5, IEEE, 2017.

[6] Y. Xie, M. B. Wakin, and G. Tang, "Simultaneous sparse recovery and blind demodulation," *arXiv preprint arXiv:1902.05023*, 2019.

[7] S. Ling and T. Strohmer, "Blind deconvolution meets blind demixing: Algorithms and performance bounds," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4497–4520, 2017.

[8] Y. Chi, "Guaranteed blind sparse spikes deconvolution via lifting and convex optimization," *J. Sel. Topics Signal Processing*, vol. 10, no. 4, pp. 782–794, 2016.

[9] Y. Xie, M. B. Wakin, and G. Tang, "Sparse recovery and non-stationary blind demodulation," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5566–5570, IEEE, 2019.

[10] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American statistical association*, vol. 101, no. 476, pp. 1418–1429, 2006.

[11] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[12] Y. Nardi, A. Rinaldo, *et al.*, "On the asymptotic properties of the group lasso estimator for linear models," *Electronic Journal of Statistics*, vol. 2, pp. 605–633, 2008.

[13] X. Lv, G. Bi, and C. Wan, "The group lasso for stable recovery of block-sparse signal representations," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1371–1382, 2011.

[14] J. Huang, T. Zhang, *et al.*, "The benefit of group sparsity," *The Annals of Statistics*, vol. 38, no. 4, pp. 1978–2004, 2010.

[15] V. Sivakumar, A. Banerjee, and P. K. Ravikumar, "Beyond sub-gaussian measurements: High-dimensional structured estimation with sub-exponential designs," in *Advances in neural information processing systems*, pp. 2206–2214, 2015.

[16] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1711–1732, 2014.
[17] Y. C. Eldar, W. Liao, and S. Tang, "Sensor calibration for off-the-grid spectral estimation," *Applied and Computational Harmonic Analysis*, 2018.
[18] A. Flinth, "Sparse blind deconvolution and demixing through $\ell_{1,2}$-minimization," *Advances in Computational Mathematics*, vol. 44, no. 1, pp. 1–21, 2018.
[19] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso)," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2183–2202, 2009.
[20] D. Hsu, S. Kakade, T. Zhang, *et al.*, "A tail inequality for quadratic forms of subgaussian random vectors," *Electronic Communications in Probability*, vol. 17, 2012.
[21] Y. Xie, M. B. Wakin, and G. Tang, "Support recovery for sparse signals with non-stationary modulation," *arXiv preprint arXiv:1910.13104*, 2019.
[22] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," 2008.