SPARSE RECOVERY AND NON-STATIONARY BLIND DEMODULATION

Youye Xie, Michael B. Wakin, and Gongguo Tang

Department of Electrical Engineering, Colorado School of Mines, Golden, CO, USA

ABSTRACT

In this paper, we consider a general sparse recovery and blind demodulation model. Different from the ones in the literature, in our general model, each dictionary atom undergoes a distinct modulation process; we refer to this as non-stationary modulation. We also assume that the modulation matrices live in a known subspace. Through the lifting technique, the sparse recovery and blind demodulation problem can be reformulated as a column-wise sparse matrix recovery problem, and we are able to recover both the sparse source signal and a cluster of modulation matrices via atomic norm and the induced $\ell_{2,1}$ norm minimizations. Moreover, we show that the sampling complexity for exact recovery is proportional to the number of degrees of freedom up to log factors in the noiseless case. We also bound the recovery error in terms of the norm of the noise when the observation is noisy. Numerical simulations are conducted to illustrate our results.

Index Terms— Sparse recovery, blind demodulation, atomic norm minimization, sparse matrix recovery

1. INTRODUCTION

1.1. Motivation

The sparse recovery and blind demodulation problem arises naturally in a wide range of applications such as self calibration [1] and blind super-resolution [2]. Mathematically, the system receives $\boldsymbol{y} = \mathbf{D}\mathbf{A}\boldsymbol{c} \in \mathbf{C}^N$ where $\mathbf{D} \in \mathbf{C}^{N\times N}$, $\mathbf{A} \in \mathbf{R}^{N\times M}(N < M)$ and $\boldsymbol{c} \in \mathbf{R}^M$ are a diagonal matrix, the dictionary matrix and the sparse signal coefficient vector, respectively [3, 4]. The diagonal matrix $\mathbf{D}$ performs pointwise multiplication, also known as modulation in signal processing. The goal is to recover both the sparse vector $\boldsymbol{c}$ and the diagonal modulation matrix $\mathbf{D}$ from the observation $\boldsymbol{y}$. $\mathbf{A}$ is known but the whole system is under-determined.

In this paper, we consider a more general model and allow each dictionary atom to undergo a distinct modulation process. Thus,

$$\boldsymbol{y} = \sum_{j=1}^{M} c_j \mathbf{D}_j \boldsymbol{a}_j \in \mathbf{C}^N \tag{1.1}$$

This work was supported by NSF grant CCF−1704204. Email: {youyexie, mwakin, gtang}@mines.edu

where $\boldsymbol{a}_j$ is the j-th column of $\mathbf{A}$ and most c_j are zero due to the sparsity. We assume only J of c_j are non-zero. In addition, inspired by [5, 6], we assume each diagonal matrix $\mathbf{D}_j$ lives in a known low dimension subspace, that is,

$$\mathbf{D}_j = \mathrm{diag}(\mathbf{B}\boldsymbol{h}_j) \tag{1.2}$$

where $\mathbf{B} \in \mathbf{C}^{N\times K}(N > K)$ contains the bases of the K dimension subspace and $\boldsymbol{h}_j$ is the unknown coefficient vector. When there is measurement noise, the observation becomes

$$\boldsymbol{y} = \sum_{j=1}^{M} c_j \mathbf{D}_j \boldsymbol{a}_j + \boldsymbol{n} \tag{1.3}$$

where $\boldsymbol{n}$ is the noise vector. We will recover both $\mathbf{D}_j$ and c_j through atomic norm minimization [7], or equivalently the $\ell_{2,1}$ norm minimization. The atomic norm, which promotes the sparsity with respect to a given dictionary, has been widely applied in many inverse problems [8, 9].

1.2. Contributions and Related Work

In this paper, we study the general sparse recovery and non-stationary blind demodulation problem in equation (1.1) and recover all c_j and $\mathbf{D}_j$ simultaneously via atomic norm minimization. Specifically, under the assumption that $\mathbf{D}_j$'s diagonal entries obey a subspace constraint, we show that the unknown parameters can be exactly recovered with overwhelming probability when the observation is noiseless and the required number of measurement is proportional to the number of degrees of freedom, $O(JK)$. When the observation is contaminated with noise, we establish that the recovery error can be bounded in terms of the norm of the noise. Particularly, we consider the random Gaussian dictionary where each entry of $\mathbf{A}$ is sampled independently from a Gaussian distribution with 0 mean and unit variance.

The majority of the sparse recovery and blind demodulation literatures assume a common modulation process $\mathbf{D}_j = \mathbf{D}$ for all dictionary atoms [1, 3, 6]. Specifically, [6] considers a spikes deconvolution problem whose dictionary consists of complex sinusoids and [1] studies the self calibration problem with random Gaussian and Fourier dictionaries when all $\mathbf{D}_j$ are the same. [2] extends the work of [6] to non-stationary modulation. However, they consider the complex sinusoids dictionary and their technical development requires a random

assumption on $\boldsymbol{h}_j$. The random 'sign' assumption makes it challenging to develop noisy performance guarantees. Inspired by [1], [5] studies the deconvolution and demixing problem but the method requires knowledge of the number of source signals. More variations of the sparse recovery and blind demodulation problem with a common $\mathbf{D}$ can be found in [3, 10, 11].

The rest of this paper is organized as follows. In Section 2, we recast sparse recovery and blind demodulation to a column-wise sparse matrix recovery problem. The main theorems regarding the sampling complexity and noisy error bound are presented in Section 3. Section 4 and 5 are devoted to numerical simulations and the conclusion.

2. PROBLEM FORMULATION

According to the signal model in equation (1.1), if $\mathbf{B}^H = [\boldsymbol{b}_1' \quad \boldsymbol{b}_2' \quad \cdots \quad \boldsymbol{b}_N'] \in \mathbf{C}^{K\times N}$ where $\boldsymbol{b}' \in \mathbf{C}^K$, each entry of $\boldsymbol{y}$ takes the form

$$\begin{aligned}\boldsymbol{y}(n) &= \sum_{j=1}^{M} c_j \bar{\boldsymbol{a}}_j^H \boldsymbol{e}_n \boldsymbol{b}_n'^H \boldsymbol{h}_j \\ &= \mathrm{Tr}\left(\boldsymbol{e}_n \boldsymbol{b}_n'^H \sum_{j=1}^{M} c_j \boldsymbol{h}_j \bar{\boldsymbol{a}}_j^H\right) \\ &= \langle \sum_{j=1}^{M} c_j \boldsymbol{h}_j \bar{\boldsymbol{a}}_j^H, \boldsymbol{b}_n' \boldsymbol{e}_n^H \rangle \\ &:= \langle \mathbf{G}, \boldsymbol{b}_n' \boldsymbol{e}_n^H \rangle\end{aligned} \tag{2.1}$$

where $\mathbf{G} = \sum_{j=1}^{M} c_j \boldsymbol{h}_j \bar{\boldsymbol{a}}_j^H$, $\bar{\boldsymbol{a}}$ is the complex conjugate of $\boldsymbol{a}$ and $\boldsymbol{e}_n$ is the n-th column of the $N \times N$ identity matrix. Thus we can view $\boldsymbol{y}$ as the result of applying a linear measurement operator to the structured matrix $\mathbf{G}$, which we denote as $\boldsymbol{y} = \mathcal{L}'(\mathbf{G})$. To partially resolve the scaling ambiguity in c_j and $\boldsymbol{h}_j$, we assume $\boldsymbol{h}_j$ has unit norm. In addition, by defining an atomic set $\mathcal{A} = \{\boldsymbol{h}\bar{\boldsymbol{a}}^H : \bar{\boldsymbol{a}} \in \{\bar{\boldsymbol{a}}_1, ..., \bar{\boldsymbol{a}}_M\}, ||\boldsymbol{h}||_2 = 1\}$, we recover $\mathbf{G}$, and ultimately c_j and $\boldsymbol{h}_j$ via solving the following atomic norm minimization problem

$$\underset{\mathbf{G}\in\mathbf{C}^{K\times N}}{\text{minimize}} ||\mathbf{G}||_{\mathcal{A}} \quad \text{subject to } \boldsymbol{y} = \mathcal{L}'(\mathbf{G}) \tag{2.2}$$

where the atomic norm is $||\mathbf{G}||_{\mathcal{A}} := \inf\{\sum_k |\tilde{c}_k| : \mathbf{G} = \sum_k \tilde{c}_k \boldsymbol{g}_k, \boldsymbol{g}_k \in \mathcal{A}\}$. We can write $\mathbf{G} = \sum_k \tilde{c}_k \boldsymbol{g}_k = \sum_{j=1}^{M} (\sum_{k\in\mathcal{N}_j} \tilde{c}_k \tilde{\boldsymbol{h}}_k) \bar{\boldsymbol{a}}_j$ where $\mathcal{N}_j = \{k : \boldsymbol{g}_k = \tilde{\boldsymbol{h}}_k \bar{\boldsymbol{a}}_j\}$. Define $\boldsymbol{h}_j = \frac{\sum_{k\in\mathcal{N}_j} \tilde{c}_k \tilde{\boldsymbol{h}}_k}{||\sum_{k\in\mathcal{N}_j} \tilde{c}_k \tilde{\boldsymbol{h}}_k||_2}$ and $c_j = ||\sum_{k\in\mathcal{N}_j} \tilde{c}_k \tilde{\boldsymbol{h}}_k||_2$, since $|c_j| \le \sum_{k\in\mathcal{N}_j} |\tilde{c}_k|$, taking the infimum yields $||\mathbf{G}||_{\mathcal{A}} = \inf\{\sum_j |c_j| : \mathbf{G} = \sum_{j=1}^{M} c_j \boldsymbol{h}_j \bar{\boldsymbol{a}}_j^H, ||\boldsymbol{h}_j|| = 1\}$.

Proposition 1. *The atomic norm optimization problem (2.2) can be equivalently expressed as the following $\ell_{2,1}$ norm optimization problem*

$$\underset{\mathbf{X}\in\mathbf{C}^{K\times M}}{\text{minimize}} ||\mathbf{X}||_{2,1} \quad \text{subject to } \boldsymbol{y} = \mathcal{L}(\mathbf{X}) \tag{2.3}$$

where $\mathbf{X} = [c_1\boldsymbol{h}_1 \quad c_2\boldsymbol{h}_2 \quad \cdots \quad c_M\boldsymbol{h}_M] \in \mathbf{C}^{K\times M}$. $\mathcal{L}$ represents the following linear sensing process

$$\boldsymbol{y}(n) = \langle \mathbf{X}, \boldsymbol{b}_n' \boldsymbol{e}_n^H \bar{\mathbf{A}} \rangle = \boldsymbol{b}_n'^H \mathbf{X} \boldsymbol{a}_n'. \tag{2.4}$$

in which $\boldsymbol{b}_n'$ and $\boldsymbol{a}_n'$ are the n-th column of $\mathbf{B}^H$ and $\mathbf{A}^T$.

Proof. Define $\mathbf{x}_j = c_j \mathbf{h}_j$ and $\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_M]$.

$$\begin{aligned}&||\mathbf{G}||_{\mathcal{A}} \\ &= \inf\left\{\sum_{j=1}^{M} |c_j| : \mathbf{G} = \sum_{j=1}^{M} c_j \boldsymbol{h}_j \bar{\boldsymbol{a}}_j^H, ||\boldsymbol{h}_j||_2 = 1\right\} \\ &= \inf\left\{\sum_{j=1}^{M} ||\boldsymbol{x}_j||_2 : \mathbf{G} = \sum_{j=1}^{M} \boldsymbol{x}_j \bar{\boldsymbol{a}}_j^H\right\} \\ &= \inf\left\{\sum_{j=1}^{M} ||\boldsymbol{x}_j||_2 : \mathbf{G} = [\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_M] \begin{bmatrix} \bar{\boldsymbol{a}}_1^H \\ \bar{\boldsymbol{a}}_2^H \\ \vdots \\ \bar{\boldsymbol{a}}_M^H \end{bmatrix}\right\} \\ &= \inf\left\{||\mathbf{X}||_{2,1} : \mathbf{G} = \mathbf{X}\bar{\mathbf{A}}^H\right\}.\end{aligned} \tag{2.5}$$

In addition, equation (2.1) indicates

$$\boldsymbol{y}(n) = \langle \mathbf{G}, \boldsymbol{b}_n' \boldsymbol{e}_n^H \rangle = \langle \mathbf{X}, \boldsymbol{b}_n' \boldsymbol{e}_n^H \bar{\mathbf{A}} \rangle = \boldsymbol{b}_n'^H \mathbf{X} \boldsymbol{a}_n'. \tag{2.6}$$

□

Note that the $\ell_{2,1}$ norm has been applied in the sparse representation with multiple measurement vectors (MMV) problem to exploit the joint sparsity structure [12, 13]. But if we vectorize their linear measurement matrix, their sensing matrix becomes a block diagonal matrix and all the diagonal block matrices are the same which is different from our measurement model. Specifically, if we reformulate the equality condition in equation (2.3) to the matrix-vector multiplication form, we obtain $\mathcal{L}(\mathbf{X}) = \mathbf{\Phi} \cdot \mathrm{vec}(\mathbf{X})$ and $\mathbf{\Phi} \in \mathbf{C}^{N\times KM}$ is

$$\mathbf{\Phi} = [\phi_{1,1} \quad \cdots \quad \phi_{K,1} \quad \cdots \quad \phi_{1,M} \quad \cdots \quad \phi_{K,M}] \tag{2.7}$$

in which $\phi_{i,j} = \mathrm{diag}(\boldsymbol{b}_i)\boldsymbol{a}_j \in \mathbf{C}^{N\times 1}$ and $\boldsymbol{b}_i$ is the i-th column of the matrix $\mathbf{B}$. After solving problem (2.3), $\mathbf{D}_j$ and c_j can be easily recovered from the solution $\hat{\mathbf{X}}$ via

$$c_j = ||\hat{\boldsymbol{x}}_j||_2, \quad \boldsymbol{h}_j = \frac{\hat{\boldsymbol{x}}_j}{||\hat{\boldsymbol{x}}_j||_2} \text{ and } \mathbf{D}_j = \mathrm{diag}(\mathbf{B}\boldsymbol{h}_j) \tag{2.8}$$

for all j and $\hat{\boldsymbol{x}}_j \neq 0$.

Furthermore, when the observed signal contains additive noise, $\boldsymbol{n}$, and assume $||\boldsymbol{n}||_2 \le \eta$, in order to make the ground truth solution feasible, we relax the constraint in equation (2.3) and the optimization problem becomes

$$\underset{\mathbf{X}\in\mathbf{C}^{K\times M}}{\text{minimize}} ||\mathbf{X}||_{2,1} \quad \text{subject to } ||\boldsymbol{y} - \mathcal{L}(\mathbf{X})||_2 \le \eta. \tag{2.9}$$

3. MAIN RESULTS

In this section, we present the main results of this paper. Theorem 3.1 considers the noiseless observation and the result for the noisy case is in Theorem 3.2.

Theorem 3.1. *(Noiseless case) Consider the model in equation (1.1). Assume that $\mathbf{B}^H\mathbf{B} = \mathbf{I}_K$ and only J atoms of $\mathbf{A}$ are committed to the observed signal. The solution $\hat{\mathbf{X}}$ to problem (2.3) is the ground truth solution, $\mathbf{X}_0$, which means c_j and $\mathbf{D}_j$ for all j can be exactly recovered, with probability at least $1-O(N^{-\alpha+1})$ if $\mathbf{A} \in \mathbf{R}^{N\times M}(N < M)$ is a random Gaussian matrix and*

$$\frac{N}{\log^2 N} \geq C_\alpha \mu_{max}^2 KJ(\log(M-J)+\log(N)). \quad (3.1)$$

Here C_α is a constant defined for $\alpha > 1$ and

$$\mu_{max} = \max_{i,j} \sqrt{N}|\mathbf{B}_{ij}|. \quad (3.2)$$

Theorem 3.2. *(Noisy case) Consider the model in equation (1.3). Assume that $\mathbf{B}^H\mathbf{B} = \mathbf{I}_K$, only J atoms of $\mathbf{A}$ are committed to the observed signal and the norm of the noise is bounded, $||\boldsymbol{n}||_2 \leq \eta$. Then with probability at least $1 - O(N^{-\alpha+1})$, if $\mathbf{A} \in \mathbf{R}^{N\times M}(N < M)$ is a random Gaussian matrix, the solution $\hat{\mathbf{X}}$ to problem (2.9) satisfies*

$$||\hat{\mathbf{X}} - \mathbf{X}_0||_F \leq \left(C_1 + C_2\sqrt{J}\right)\eta \quad (3.3)$$

when

$$\frac{N}{\log^2 N} \geq C_\alpha \mu_{max}^2 KJ\left(\log(C\mu_{max}\sqrt{KJ})C+1\right)\cdot (\log(M-J)+\log(MK)+\log(N)) \quad (3.4)$$

where C, C_1 and C_2 are constant. C_α is defined for $\alpha > 1$.

The assumption $\mathbf{B}^H\mathbf{B} = \mathbf{I}_K$ requires the columns of $\mathbf{B}$ to be orthonormal because $\mathbf{B}$ contains the bases of the K dimension subspace. In addition, we require that the energy of each column of $\mathbf{B}$ spreads across the whole column. In this case, the coherence parameter $\mu_{max} = \max_{i,j}\sqrt{N}|\mathbf{B}_{ij}| \approx 1$ and the linear sensing process $\mathcal{L}$ possesses the isometry property which is critical in the proof. The main ingredient of the proof of Theorem 3.1 involves the construction of a dual certificate matrix. The existence of such a matrix satisfying the properties in Proposition 2 guarantees that the ground truth $\mathbf{X}_0$ is the unique solution to the minimization problem (2.3).

Proposition 2. *The matrix $\mathbf{X}_0 \in \mathbf{C}^{K\times M}$ with support T, which is a set containing the indices of non-zero columns, is the unique solution to the inverse problem (2.3), if there exists a matrix $\mathbf{Y}$ in the range space of $\mathcal{L}^*$, such that*

$$||\mathbf{Y}_T - \text{sign}(\mathbf{X}_{0,T})||_F \leq \frac{1}{4\sqrt{2}\gamma} \quad \textit{and} \quad ||\mathbf{Y}_{T^C}||_{2,\infty} \leq \frac{1}{2} \quad (3.5)$$

and the operator $\mathcal{L}$ satisfies

$$||\mathcal{L}_T^*\mathcal{L}_T - \mathbf{I}_T|| \leq \frac{1}{2} \quad \textit{and} \quad ||\mathcal{L}|| \leq \gamma. \quad (3.6)$$

When $\mathbf{A}$ is a random Gaussian matrix, $||\mathcal{L}|| \leq \gamma := \sqrt{M\log(MN/2)+\alpha\log(N)}$ with probability at least $1 - N^{-\alpha}$ [1, 3]. Moreover, the desired dual certificate matrix can be constructed through

$$\text{vec}(\mathbf{Y}) = \mathbf{\Phi}^H\mathbf{\Phi}_T(\mathbf{\Phi}_T^H\mathbf{\Phi}_T)^{-1}\text{vec}(\text{sign}(\mathbf{X}_{0,T})) \in \mathbf{C}^{KM\times 1} \quad (3.7)$$

where $\mathbf{\Phi}_T \in \mathbf{C}^{N\times KJ}$ consists of columns $\phi_{i,j}$ with $j \in T$ and $\mathbf{X}_{0,T} \in \mathbf{C}^{K\times J}$ is composed of the columns of $\mathbf{X}_0$ on the support T. Moreover, $\text{sign}(\boldsymbol{x}) = \boldsymbol{x}/||\boldsymbol{x}||_2$ and $\text{sign}(\mathbf{X}) = [\text{sign}(\boldsymbol{x}_1), \cdots, \text{sign}(\boldsymbol{x}_M)]$. Therefore, we already have $\mathbf{Y}_T = \text{sign}(\mathbf{X}_{0,T})$ and it remains to derive the conditions under which $||\mathcal{L}_T^*\mathcal{L}_T - \mathbf{I}_T|| \leq \frac{1}{2}$ and $||\mathbf{Y}_{T^C}||_{2,\infty} \leq \frac{1}{2}$, which can be shown using the Orlicz norm version of the Bernstein inequality [3] and a tail inequality [14] respectively. When the observation is contaminated with noise, we can no longer recover the unknown parameters exactly but we can bound the recovery error through the theorem below. It is a variation of Theorem 4.33 in [15] adapted to our problem (2.9).

Theorem 3.3. *Define $\mathbf{\Phi} \in \mathbf{C}^{N\times KM}$ and $\mathbf{\Phi}\cdot vec(\mathbf{X}) = \mathcal{L}(\mathbf{X})$. The ground truth $\mathbf{X}_0$ has J non-zero columns with support T and observation $\boldsymbol{y} = \mathcal{L}(\mathbf{X}_0) + \boldsymbol{n}$ with $||\boldsymbol{n}||_2 \leq \eta$. For $\delta, \beta, \theta, \gamma, \tau > 0$ and $\delta < 1$, assume that*

$$\max_{i\in T^C} ||\mathbf{\Phi}_T^H[\mathbf{\Phi}_{K(i-1)+1}\cdots \quad \mathbf{\Phi}_{K(i-1)+K}]|| \leq \beta, \quad (3.8)$$

$$||\mathbf{\Phi}_T^H\mathbf{\Phi}_T - \mathbf{I}_T|| \leq \delta \quad (3.9)$$

and that there exists a matrix $\mathbf{Y} = \mathcal{L}^*(\boldsymbol{p}) \in \mathbf{C}^{K\times M}$ such that

$$||\mathbf{Y}_T - \text{sign}(\mathbf{X}_{0,T})||_F \leq \frac{1}{4\sqrt{2}\gamma}, \quad ||\mathbf{Y}_{T^C}||_{2,\infty} \leq \theta, \quad \textit{and} \quad ||\boldsymbol{p}||_2 \leq \tau\sqrt{J}. \quad (3.10)$$

If $\rho := \theta + \frac{\beta}{4\sqrt{2}\gamma(1-\delta)} < 1$, then the minimizer, $\hat{\mathbf{X}}$, to problem (2.9) satisfies

$$||\hat{\mathbf{X}} - \mathbf{X}_0||_F \leq \left(C_1 + C_2\sqrt{J}\right)\eta \quad (3.11)$$

where C_1 and C_2 are two constants depending on $\delta, \beta, \theta, \gamma, \tau$.

Most of the constants, like δ, θ and γ, have already been specified during the construction of the dual certificate matrix. Full details of the proof and additional results when $\mathbf{A}$ is the random Fourier dictionary are available in [16].

4. NUMERICAL SIMULATIONS

Three simulations are conducted to support our theoretical results. We set $\mathbf{B}$ to be the first K columns of the normalized DFT matrix, and sample c_j and $\boldsymbol{h}_j$ from i.i.d standard normal distribution. The support of the ground truth solution $\mathbf{X}_0 = [c_1\boldsymbol{h}_1 \cdots c_M\boldsymbol{h}_M]$ are selected uniformly. In the first simulation, we fix $M = 200$, $N = 100$ and vary J and K. We run 40 trials for each setting and record the success recovery rate in Fig. 1. A successful recovery is declared when the relative error, $||\hat{\mathbf{X}} - \mathbf{X}_0||_F / ||\mathbf{X}_0||_F$, is smaller than 10^{-5}. In the second simulation, we examine J and K individually. From Fig. 2 and 3, we observe that the sufficient number of measurements for exact recovery scales nearly linearly with respect to J and K. In Fig. 4, we fix $J = K = 5$ and show that the recovery error scales linearly with respect to the norm of the additive Gaussian noise. The dashed line shows the theoretical error bound from Theorem 3.2. When the noise dominates the signal, the minimization (2.9) returns $\hat{\mathbf{X}} = 0$ and the relative error (dB) becomes 0.

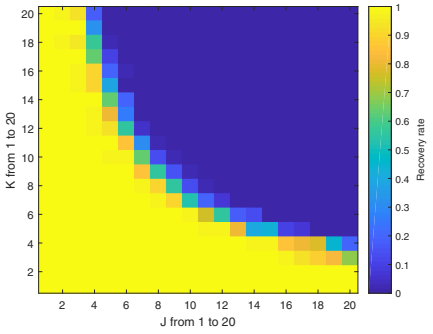

Fig. 1. The relation between the subspace dimension of the modulation matrix, K, and the number of committed atoms, J, in terms of the success recovery rate.

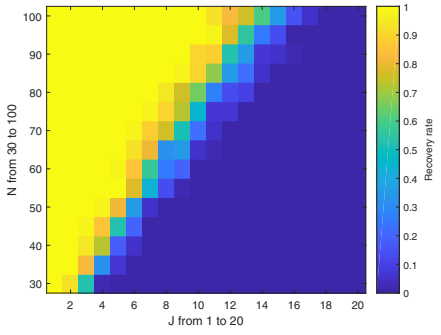

Fig. 2. The nearly linear relation between the dimension of the observed signal, N, and the number of committed atoms, J, in terms of the success recovery rate.

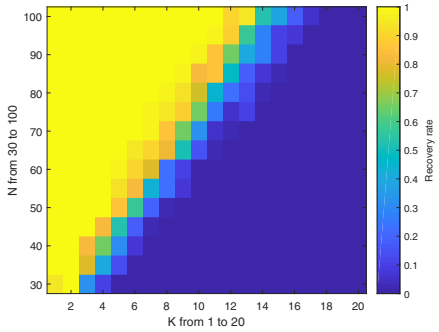

Fig. 3. The nearly linear relation between the dimension of the observed signal, N, and subspace dimension, K, in terms of the success recovery rate.

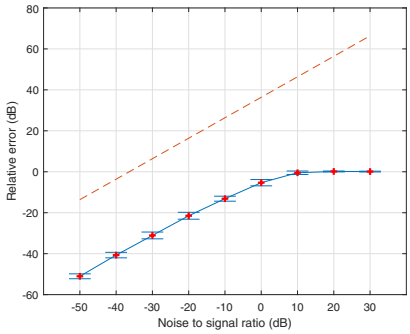

Fig. 4. The relation between the relative error (dB) and noise to signal ratio (dB). The blue horizontal sticks and red plus sign indicate the range of the standard deviation and the mean of the relative error (dB) given a specific noise to signal ratio (dB). The dashed line shows the theoretical error bound from Theorem 3.2.

5. CONCLUSION

In this paper, we consider a general sparse recovery and non-stationary blind demodulation problem with a random Gaussian dictionary. We recast the problem to a column-wise sparse matrix recovery problem which can be solved via atomic norm minimization and its induced $\ell_{2,1}$ norm minimization. With noiseless observation, we derive the sampling complexity for exact recovery which is proportional to the number of degrees of freedom up to log factors. Moreover, we bound the recovery error in terms of the strength of the noise when the observation is contaminated with noise. Numerical simulations verify our results.

6. REFERENCES

[1] Shuyang Ling and Thomas Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, pp. 115002, 2015.

[2] Dehui Yang, Gongguo Tang, and Michael 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.

[3] Ali Ahmed, Benjamin Recht, and Justin Romberg, "Blind deconvolution using convex programming," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1711–1732, 2014.

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

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

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

[7] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[8] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht, "Compressive sensing off the grid," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2012, pp. 778–785.

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

[10] Yonina C Eldar, Wenjing Liao, and Sui Tang, "Sensor calibration for off-the-grid spectral estimation," *arXiv preprint arXiv:1707.03378*, 2017.

[11] Axel Flinth, "Sparse blind deconvolution and demixing through 1, 2-minimization," *Advances in Computational Mathematics*, vol. 44, no. 1, pp. 1–21, 2018.

[12] Jie Chen and Xiaoming Huo, "Sparse representations for multiple measurement vectors (mmv) in an overcomplete dictionary," in *2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2005, vol. 4, pp. iv–257.

[13] Shane F Cotter, Bhaskar D Rao, Kjersti Engan, and Kenneth Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[14] Daniel Hsu, Sham Kakade, Tong Zhang, et al., "A tail inequality for quadratic forms of subgaussian random vectors," *Electronic Communications in Probability*, vol. 17, 2012.

[15] Simon Foucart and Holger Rauhut, *A mathematical introduction to compressive sensing*, Birkhäuser Basel, 2013.

[16] Youye Xie, Michael B Wakin, and Gongguo Tang, "Simultaneous sparse recovery and blind demodulation," *arXiv preprint arXiv:1902.05023*, 2019.