Blind Deconvolution from Multiple Sparse Inputs

Liming Wang, *Member, IEEE,* and Yuejie Chi, *Member, IEEE*

***Abstract*—Blind deconvolution is an inverse problem when both the input signal and the convolution kernel are unknown. We propose a convex algorithm based on ℓ_1-minimization to solve the blind deconvolution problem, given multiple observations from sparse input signals. The proposed method is related to other problems such as blind calibration and finding sparse vectors in a subspace. Sufficient conditions for exact and stable recovery using the proposed method are developed which shed light on the sample complexity. Finally, numerical examples are provided to showcase the performance of the proposed method.**

***Index Terms*—blind deconvolution, blind calibration, convex programming, dictionary learning, sparsity**

I. Introduction

Blind deconvolution is a classical inverse problem that ubiquitously appears in various areas of signal processing [1], communications [2] and array processing [3]. For such a problem, the observation is often posed as a convolution of the signal with some kernel, or filter. In addition, the explicit knowledge of the convolution kernel nor the signal is unknown *a priori*. The goal of blind deconvolution is to recover both the signal and the kernel from the observation.

Motivated by multi-channel blind deconvolution [4] and the joint recovery problem of signals and sensor parameters in array signal processing [3], the observation can be formulated as $\mathbf{y} = \mathbf{g} \circledast \mathbf{x} \in \mathbb{R}^n$, where $\circledast$ denotes the circular convolution, $\mathbf{g} \in \mathbb{R}^n$ denotes the kernel and $\mathbf{x} \in \mathbb{R}^n$ denotes the input signal. Recently, this problem has drawn lots of research attention, where algorithms with provable performance guarantees are developed by assuming both $\mathbf{g}$ and $\mathbf{x}$ satisfy certain sparsity or subspace constraints [5]–[8]. However, these assumptions may be difficult to verify in practice, particularly for the kernel.

In the case when multiple input signals are present, the observations are given as $\mathbf{y}_i = \mathbf{g} \circledast \mathbf{x}_i$, for $1 \leq i \leq p$, where p is the number of observations. It is known that [9], as long as p is sufficiently large, the problem is identifiable and can be solved via a least-squares approach by exploiting the cross-correlations of the observations under some mild conditions of the kernel. In this letter, we reconsider this problem by assuming the input signals $\mathbf{x}_i$'s are sparse, which is motivated by applications of compressed sensing [10]. Identifiability under this setup is recently studied in [11]. It is natural to seek a kernel $\mathbf{g}$ such that the inputs are made as sparse as possible, however, such a direct consideration is not computationally feasible. Alternatively, with a mild assumption that the kernel is invertible, we propose a convex optimization algorithm based on ℓ_1-minimization, which can be solved efficiently. Sufficient conditions for exact and stable recovery using the proposed method are developed under a Bernoulli-Subgaussian model of the sparse inputs, which shed light on the sample complexity. In contrast, the alternating minimization algorithm in [12] is heuristic and lack performance analysis.

L. Wang is with Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA. Email: wang.8482@osu.edu.

Y. Chi is with Department of Electrical and Computer Engineering and Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210, USA. Email: chi.97@osu.edu.

This work is supported in part by AFOSR under the grant FA9550-15-1-0205 and by ONR under the grant N00014-15-1-2387.

Our approach is mostly inspired by [13], [14] for exact dictionary learning with sparse input signals, where the problem can be regarded as a special case of learning an invertible circulant dictionary. Furthermore, the problem is also related to finding sparse vectors in a subspace [15], [16] and blind calibration [17], [18], which will be detailed later.

II. Problem Formulation

Let $\mathbf{x}_i \in \mathbb{R}^n$ denote the i-th sparse input signal and the i-th observation $\mathbf{y}_i \in \mathbb{R}^n$ can be expressed as

$$\mathbf{y}_i = \mathbf{g} \circledast \mathbf{x}_i = \mathcal{C}(\mathbf{g})\mathbf{x}_i, \quad i = 1, \ldots, p. \tag{1}$$

The common kernel or filter $\mathcal{C}(\mathbf{g}) \in \mathbb{R}^{n\times n}$ is the circulant matrix spanned by $\mathbf{g} = [g_1, \ldots, g_n]^T$, given as

$$\mathcal{C}(\mathbf{g}) = \begin{bmatrix} g_1 & g_n & \cdots & g_2 \\ g_2 & g_1 & \cdots & g_3 \\ \vdots & \vdots & \ddots & \vdots \\ g_n & g_{n-1} & \cdots & g_1 \end{bmatrix}. \tag{2}$$

The filter $\mathbf{g}$ is called invertible if $\mathcal{C}(\mathbf{g})$ is invertible. Denote $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_p] \in \mathbb{R}^{n\times p}$ and $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_p] \in \mathbb{R}^{n\times p}$, we can rewrite (1) as

$$\mathbf{Y} = \mathcal{C}(\mathbf{g})\mathbf{X}. \tag{3}$$

Furthermore, the sparse signal matrix $\mathbf{X}$ is assumed of the following Bernoulli-Subgaussian model, which is standard for modeling sparse signals.

Definition 1 (Bernoulli-Subgaussian model, [13])**.** *$\mathbf{X}$ is said to satisfy the Bernoulli-Subgaussian model with parameter $\theta \in (0,1)$, if $\mathbf{X} = \mathbf{\Omega} \odot \mathbf{R}$, where $\mathbf{\Omega}$ is an i.i.d. Bernoulli matrix with parameter θ, and $\mathbf{R}$ is an independent random matrix with i.i.d. symmetric random variables that satisfy*

$$\mu = \mathbb{E}[|R_{i,j}|] \in [1/10, 1], \quad \mathbb{E}[R_{i,j}^2] \leq 1,$$
$$\text{and} \qquad \mathbb{P}(|R_{i,j}| > t) \leq 2\exp(-t^2/2), \quad \forall t > 0.$$

We note that the sparsity of $\mathbf{X}$ is manipulated by the Bernoulli distribution, and the non-zero entries of $\mathbf{X}$ obey the Subgaussian distribution, thereby facilitating a very general model of the sparse signal $\mathbf{X}$.

Our goal is to recover both $\mathbf{g}$ and $\mathbf{X}$ from $\mathbf{Y}$. Clearly, the problem is not uniquely identifiable, due to the fact that we can always rewrite (1) as $\mathbf{y}_i = (\beta \mathbf{R}_k \mathbf{g}) \circledast (\beta^{-1}\mathbf{R}_k \mathbf{x}_i)$, where $\mathbf{R}_k$ is a circulant shift matrix by k, $k = 1, \ldots, n-1$, and $\beta \neq 0$ is an arbitrary scalar. Hence, recovery should be interpreted in the sense that $\mathbf{g}$ and $\mathbf{X}$ are accurately recovered up to a circulant shift and a scaling factor.

Since $\mathbf{X}$ is sparse, we aim to seek the sparsest $\mathbf{X}$ that reproduces the observations:

$$\{\hat{\mathbf{g}}, \hat{\mathbf{X}}\} = \operatorname*{argmin}_{\mathbf{g}\in\mathbb{R}^n, \mathbf{X}\in\mathbb{R}^{n\times p}} \|\mathbf{X}\|_0, \quad \text{subject to} \quad \mathcal{C}(\mathbf{g})\mathbf{X} = \mathbf{Y}, \mathbf{X} \neq 0, \tag{4}$$

where $\|\cdot\|_0$ is the entry-wise ℓ_0 norm, which is unfortunately computationally infeasible. A natural question is under what conditions it is guaranteed that there exists a unique pair $\{\mathbf{g}, \mathbf{X}\}$ such that $\mathbf{Y} = \mathcal{C}(\mathbf{g})\mathbf{X}$, *i.e.,* the identifiability of blind deconvolution problem. One sufficient condition is established in [11], [13], where under the Bernoulli-Subgaussian model on $\mathbf{X}$, it follows that the problem (4) is identifiable with high probability, provided that $\mathbf{g}$ is invertible, $\theta \in (1/n, 1/4)$ and $p > Cn \log n$ for some constant C. Throughout the paper, we focus on the case where $\mathbf{g}$ is invertible.

III. A Convex Optimization Approach

A. The Convex Formulation

Although it is desirable to develop a more tractable formulation than (4) to efficiently solve the problem, the problem is still non-convex even if we relax the ℓ_0 norm to ℓ_1 norm due to the bilinear constraint. Hence, we seek an alternative approach. Recall that any circulant matrix $\mathcal{C}(\mathbf{g})$ admits the eigenvalue decomposition [19]:

$$\mathcal{C}(\mathbf{g}) = \mathbf{F}^H \cdot \operatorname{diag}(\hat{\mathbf{g}}) \cdot \mathbf{F}, \tag{5}$$

where $\mathbf{F} \in \mathbb{C}^{n\times n}$ is the discrete Fourier transform (DFT) matrix, and $\hat{\mathbf{g}} = \mathbf{F}\mathbf{g}$. Since $\mathbf{g}$ is invertible, $\hat{g}_i \neq 0$ for all $1 \leq i \leq n$, then the inverse of $\mathcal{C}(\mathbf{g})$ is also circulant and can be written as

$$\mathcal{C}(\mathbf{g})^{-1} = \mathbf{F}^H \cdot \operatorname{diag}(\hat{\mathbf{g}})^{-1} \cdot \mathbf{F} := \mathcal{C}(\mathbf{h}), \tag{6}$$

where $\mathbf{h}$ denotes the inverse filter of $\mathbf{g}$. In particular, $\mathcal{C}(\mathbf{g})\mathcal{C}(\mathbf{h}) = \mathcal{C}(\mathbf{h})\mathcal{C}(\mathbf{g}) = \mathbf{I}_n$, where $\mathbf{I}_n$ denotes the identity matrix of size n.

Motivated by the observation that $\mathcal{C}(\mathbf{h})\mathbf{Y} = \mathcal{C}(\mathbf{g})^{-1}\mathbf{Y} = \mathbf{X}$ is sparse, rather than aiming at reconstructing $\mathcal{C}(\mathbf{g})$, we alternatively seek to recover $\mathcal{C}(\mathbf{h})$, by considering the following convex optimization algorithm:

$$\hat{\mathbf{v}} = \operatorname*{argmin}_{\mathbf{v}\in\mathbb{R}^n} \|\mathcal{C}(\mathbf{v})\mathbf{Y}\|_1, \tag{7}$$

where $\|\cdot\|_1$ is the entry-wise ℓ_1 norm to motivate sparse solutions. However, the above algorithm (7) admits a trivial solution $\hat{\mathbf{v}} = \mathbf{0}$. In order to avoid this case, an additional linear constraint can be added to (7), yielding the algorithm

$$\hat{\mathbf{v}} = \operatorname*{argmin}_{\mathbf{v}\in\mathbb{R}^n} \|\mathcal{C}(\mathbf{v})\mathbf{Y}\|_1, \quad \text{subject to} \quad \mathbf{e}_1^T \mathbf{v} = 1. \tag{8}$$

where $\mathbf{e}_1 = [1, 0, \cdots, 0]^T$. This constraint not only avoids the trivial solution, but also eliminates the scaling ambiguity. The reason for the choice of $\mathbf{e}_1$ is made clear when we analyze the performance. Once $\hat{\mathbf{v}}$ is obtained, one can recover $\mathbf{X}$ as $\hat{\mathbf{X}} = \mathcal{C}(\mathbf{v})\mathbf{Y}$.

B. Connection to Other Problems

First, the problem under study is related to learning an invertible dictionary [13], [14], where one aims to recover an invertible $\mathbf{A} \in \mathbb{R}^{n\times n}$ and a sparse coefficient matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ from their product $\mathbf{Y} = \mathbf{A}\mathbf{X}$. The trick used in developing (8) is also reminiscent of the ER-SpUD algorithm in [13]. Although it is possible to use ER-SpUD for our problem by ignoring the circulant structure of the dictionary, it requires solving p different subproblems, with each one similar to the complexity of (8), which is much more demanding.

By writing $\mathcal{C}(\mathbf{v}) = \mathbf{F}^H \operatorname{diag}(\mathbf{F}\mathbf{v})\mathbf{F}$, the objective function of (8) can be rewritten as

$$\begin{aligned}\|\mathbf{F}^H \operatorname{diag}(\mathbf{F}\mathbf{v})\mathbf{F}\mathbf{Y}\|_1 &= \left\|\operatorname{vec}[\mathbf{F}^H \operatorname{diag}(\mathbf{F}\mathbf{v})\mathbf{F}\mathbf{Y}]\right\|_1 \\ &= \left\|(\mathbf{Y}^T\mathbf{F} \odot \mathbf{F}^H)\mathbf{F}\mathbf{v}\right\|_1,\end{aligned}$$

where $\operatorname{vec}[\cdot]$ vectorizes the argument matrix, and $\odot$ denotes the Khatri-Rao product [20]. Hence, the problem (8) can then be interpreted as finding a sparse vector in the structured subspace $\mathbf{S} = (\mathbf{Y}^T\mathbf{F} \odot \mathbf{F}^H)\mathbf{F} \in \mathbb{C}^{pn\times n}$ [15], [16].

Finally, note that we can rewrite (3) as $\mathbf{F}\mathbf{Y} = \operatorname{diag}(\hat{\mathbf{g}})\mathbf{F}\mathbf{X}$ by multiplying $\mathbf{F}$ on both sides. The problem of simultaneously recovering $\hat{\mathbf{g}}$ and $\mathbf{X}$ is then equivalent to blind calibration of a compressed sensing system [17], [18], where the sparsifying basis is the DFT matrix and $\hat{\mathbf{g}}$ is the unknown calibration vector. In this case, the proposed algorithm (8) becomes equivalent to the algorithm in [17], [18] for gain calibration, wherein it is only studied numerically without performance analysis.

IV. Theoretical Analysis

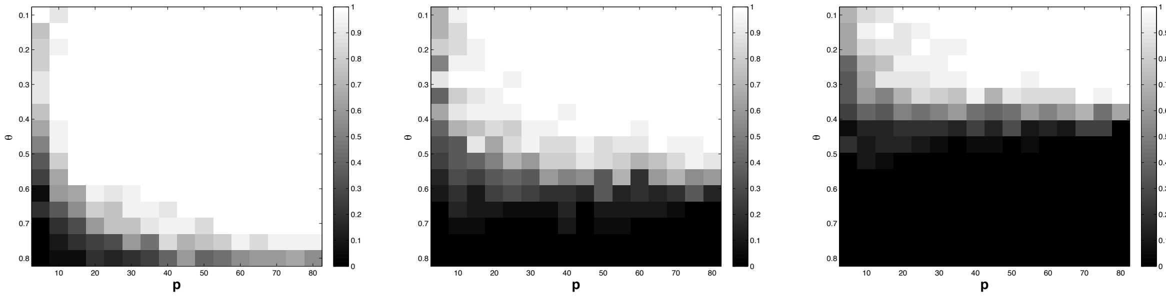
We now provide theoretical analysis of the proposed method in (8). We first establish the sufficient conditions under which the proposed method allows exact and stable recovery, then we comment on when these conditions hold with a high probability under the Bernoulli-Subgaussian model of $\mathbf{X}$. Let S be the support of $\mathbf{X}$ and $[\cdot]_S$ be the argument matrix restricted on the support S. Denote $|h|_{(i)}$ as the ith largest entry of $\mathbf{h}$ in the absolute value, and without loss of generality let $|h|_{(1)} = 1$ to eliminate the scaling ambiguity. We first establish following results that are useful later. For any $\mathbf{v} \in \mathbb{R}^n$ we have

$$\begin{aligned}\mathbb{E}\|\mathcal{C}(\mathbf{v})\mathbf{X}\|_1 &= \mathbb{E}\sum_{i=1}^p \|\mathcal{C}(\mathbf{v})\mathbf{x}_i\|_1 = np\mathbb{E}\left|\sum_{i=1}^n x_{n+1-i}v_i\right|, \\ \mathbb{E}\|[\mathcal{C}(\mathbf{v})\mathbf{X}]_S\|_1 &= |S|\mathbb{E}\left|\sum_{i=1}^n x_{n+1-i}v_i\right|,\end{aligned} \tag{9}$$

and if $n\theta \geq 2$,

$$\mathbb{E}\left|\sum_{i=1}^n x_{n+1-i}v_i\right| \geq \frac{\mu}{4}\sqrt{\frac{\theta}{n}}\|\mathbf{v}\|_1, \tag{10}$$

where the last inequality follows from [13, Lemma 16]. We further make the following assumptions:



(a) $\beta = 0.05$ (b) $\beta = 0.1$ (c) $\beta = 0.2$

Figure 1: Phase transition of the proposed algorithm with respect to θ and p for various values of β. The gray level for each cell denotes the recovery success rate.

A1) There exists $0 < \delta < 1$ such that for all $\mathbf{w} \in \mathbb{R}^n$:

$$\left| \|\mathcal{C}(\mathbf{w})\mathbf{X}\|_1 - \mathbb{E}\|\mathcal{C}(\mathbf{w})\mathbf{X}\|_1 \right| \leq \delta \mathbb{E}\|\mathcal{C}(\mathbf{w})\mathbf{X}\|_1, \quad (11)$$

and

$$\left| \|[\mathcal{C}(\mathbf{w})\mathbf{X}]_S\|_1 - \mathbb{E}\|[\mathcal{C}(\mathbf{w})\mathbf{X}]_S\|_1 \right| \leq \delta \mathbb{E}\|[\mathcal{C}(\mathbf{w})\mathbf{X}]_S\|_1. \quad (12)$$

A2) There exists $0 < \delta_1, \delta_2 < 1$ such that

$$(1-\delta_1)\mu\theta np \leq \|\mathbf{X}\|_1 \leq (1+\delta_1)\mu\theta np \quad (13)$$

and

$$(1-\delta_2)\theta np \leq |S| \leq (1+\delta_2)\theta np. \quad (14)$$

With the above assumptions, we present the main theorem which quantifies the sufficient conditions for exact recovery.

Theorem 1. *Assume $2/n \leq \theta \leq (1-\delta)/(2(1+\delta)(1+\delta_2))$. Under the assumptions A1) and A2), the proposed algorithm (8) achieves exact recovery,* i.e., *uniquely identifies* $\mathbf{h}$ *up to the shift and scaling ambiguity, provided that* $|h|_{(2)} \leq \frac{(1-\delta)-2(1+\delta)(1+\delta_2)\theta}{4(1+\delta_1)\sqrt{\theta n}}$.

Proof. Rather than working with the original problem (8), we first transform it into an equivalent problem. Applying a change of variable $\mathbf{v} = \mathcal{C}(\mathbf{h})\mathbf{w} = \mathbf{h} \circledast \mathbf{w}$, we have

$$\mathcal{C}(\mathbf{v})\mathbf{Y} = \mathcal{C}(\mathbf{h})\mathcal{C}(\mathbf{w})\mathcal{C}(\mathbf{g})\mathbf{X} = \mathcal{C}(\mathbf{w})\mathcal{C}(\mathbf{h})\mathcal{C}(\mathbf{g})\mathbf{X} = \mathcal{C}(\mathbf{w})\mathbf{X},$$

and

$$\mathbf{e}_1^T\mathbf{v} = \mathbf{e}_1^T\mathcal{C}(\mathbf{h})\mathbf{w} = \mathbf{r}^T\mathbf{w},$$

where $\mathbf{r} = [h_1, h_n, \ldots, h_2]^T$ is the first row of $\mathcal{C}(\mathbf{h})$, and a permuted version of $\mathbf{h}$. Then (8) is equivalent to

$$\hat{\mathbf{w}} = \underset{\mathbf{w}\in\mathbb{R}^n}{\operatorname{argmin}} \|\mathcal{C}(\mathbf{w})\mathbf{X}\|_1, \quad \text{subject to} \quad \mathbf{r}^T\mathbf{w} = 1. \quad (15)$$

Although we cannot directly implement (15) since $\mathbf{X}$ and $\mathbf{r}$ are unknown, but it is more convenient to analyze. We order the entries of $\mathbf{r}$ in magnitude as $|r|_{(1)} \geq |r|_{(2)} \geq \cdots \geq |r|_{(n)}$. Without loss of generality, we assume $j^* = 1$ is the index of the largest entry of $\mathbf{r}$ in magnitude, and $r_1 = 1$. Clearly $\mathbf{e}_1$ is a feasible solution of (15). Our goal is to demonstrate that under the assumptions of Theorem 1, it is also the unique solution of (15), which will in turn imply that $\mathbf{h}$ is the unique solution of (8).

Denote the solution of (15) as $\hat{\mathbf{w}} = \mathbf{w}_1 + \mathbf{w}_2 = w_1\mathbf{e}_1 + \mathbf{w}_2$, where $\mathbf{w}_1$ is supported on the first entry, and $\mathbf{w}_2$ is supported on $\{2, \ldots, n\}$. Due to feasibility of $\hat{\mathbf{w}}$, we have

$$\mathbf{r}^T(\mathbf{w}_1 + \mathbf{w}_2) = w_1 + \mathbf{r}^T\mathbf{w}_2 = 1,$$

and it results in $w_1 = 1 - \mathbf{r}^T\mathbf{w}_2$. Recall S as the support of $\mathbf{X}$, and its complement set is denoted as S^c. Let $\alpha = \mathbb{E}\left|\sum_{i=1}^n x_{n+1-i}(w_2)_i\right|$ and we have

$$\begin{aligned}
&\|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X}\|_1 = \|\mathcal{C}(\mathbf{w}_1)\mathbf{X} + [\mathcal{C}(\mathbf{w}_2)\mathbf{X}]_S + [\mathcal{C}(\mathbf{w}_2)\mathbf{X}]_{S^c}\|_1 \\
&= \|w_1\mathbf{X} + [\mathcal{C}(\mathbf{w}_2)\mathbf{X}]_S\|_1 + \|[\mathcal{C}(\mathbf{w}_2)\mathbf{X}]_{S^c}\|_1 \quad (16)\\
&= \|w_1\mathbf{X} + [\mathcal{C}(\mathbf{w}_2)\mathbf{X}]_S\|_1 + \|\mathcal{C}(\mathbf{w}_2)\mathbf{X} - [\mathcal{C}(\mathbf{w}_2)\mathbf{X}]_S\|_1 \\
&\geq \left(1 - |\mathbf{r}^T\mathbf{w}_2|\right)\|\mathbf{X}\|_1 - 2\left\|[\mathcal{C}(\mathbf{w}_2)\mathbf{X}]_S\right\|_1 + \|\mathcal{C}(\mathbf{w}_2)\mathbf{X}\|_1 \\
&\geq \left(1 - |\mathbf{r}^T\mathbf{w}_2|\right)\|\mathbf{X}\|_1 - 2(1+\delta)\mathbb{E}\left\|[\mathcal{C}(\mathbf{w}_2)\mathbf{X}]_S\right\|_1 \\
&+ (1-\delta)\mathbb{E}\left\|\mathcal{C}(\mathbf{w}_2)\mathbf{X}\right\|_1 \quad (17)\\
&\geq \|\mathbf{X}\|_1 - |r|_{(2)}\|\mathbf{w}_2\|_1\|\mathbf{X}\|_1 - 2(1+\delta)|S|\alpha + (1-\delta)np\alpha
\end{aligned}$$

where (16) follows from the decomposability of the ℓ_1 norm on disjoint support sets and (17) follows from A1).

In order to show $-|r|_{(2)}\|\mathbf{w}_2\|_1\|\mathbf{X}\|_1 - 2(1+\delta)|S|\alpha + (1-\delta)np\alpha \geq 0$, we need

$$|r|_{(2)} \leq \frac{(1-\delta)np - 2(1+\delta)(1+\delta_2)\theta np}{(1+\delta_1)\mu\theta np\|\mathbf{w}_2\|_1}\alpha, \quad (18)$$

and above inequality holds if

$$|r|_{(2)} \leq \frac{(1-\delta) - 2(1+\delta)(1+\delta_2)\theta}{(1+\delta_1)\mu\theta}\frac{\mu}{4}\sqrt{\frac{\theta}{n}}, \quad (19)$$

which is assumed in the statement of the theorem. Note that (18) and (19) follow from A2) and (10), respectively. Therefore, we show that

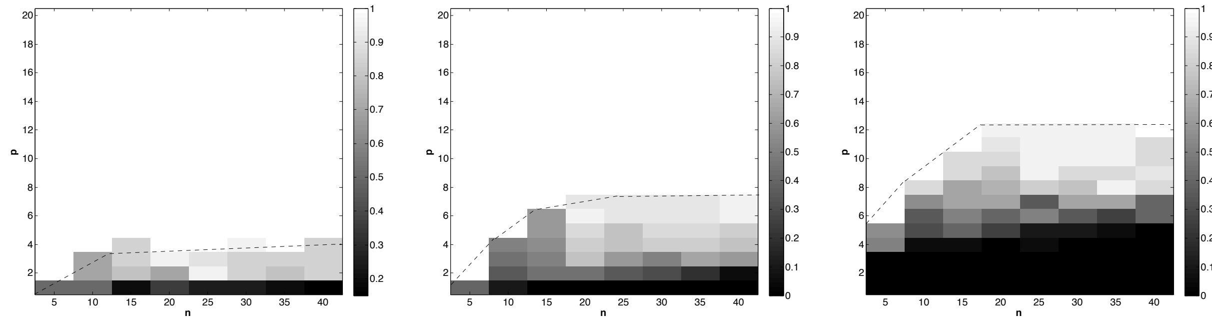
$$-|r|_{(2)}\|\mathbf{w}_2\|_1\|\mathbf{X}\|_1 - 2(1+\delta)|S|\alpha + (1-\delta)np\alpha \geq 0$$

Meanwhile, we have $\|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X}\|_1 \leq \|\mathbf{X}\|_1$. As a result, $\|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X}\|_1 = \|\mathbf{X}\|_1$ and $\mathbf{e}_1$ is the unique solution of (15). ☐

Note that the assumption A2) essentially postulates $\mathbf{h}$ to be a vector with most of its energy concentrated on one entry. Consequently, its inverse filter $\mathbf{g}$ also possesses this property. Therefore, Theorem 1 implies that (8) achieves exact recovery, provided that the filter $\mathbf{g}$ does not deviate too far from $\mathbf{e}_1$.

A. Discussions of the Assumptions

With additional conditions, A1) and A2) can be established with high probabilities. For example, it can be shown that A1) holds with probability $1 - o(1)$, provided that $p \geq C_0 n \log^4 n$



(a) $\theta = 0.3$ (b) $\theta = 0.4$ (c) $\theta = 0.5$

Figure 2: Phase transition of the proposed algorithm with respect to n and p for various values of θ when $\beta = 0.3$. The gray level for each cell denotes the recovery success rate. The dashed line depicts the envelope of the plot.

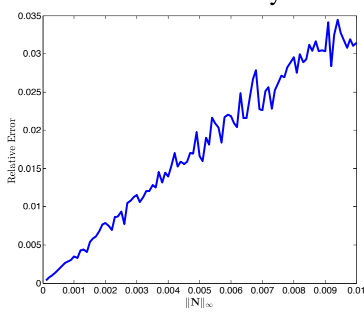


Figure 3: Relative recovery error of the proposed algorithm with respect to the observation noise.

for some positive constant C_0 [14]. For the assumption A2), it is known from [13, Lemma 16 and 18] that (13) holds with probability at least $1 - 4\exp(-C_1\delta_1^2\theta np)$, where C_1 is a positive constant, and (14) holds with probability at least $1 - 2\exp(-2\delta_2^2\theta np/3)$. Combining these results with Theorem 1, it is straightforward to argue that the proposed algorithm (8) admits exact recovery as long as $p \geq C_0 n \log^4 n$ and $\mathbf{h}$ is spiky enough. This sample complexity matches the identifiability bound in [11], [13] up to logarithmic factors.

B. Stability of the Proposed Method

We now establish a stability result for the proposed method under observation noise, by showing that the recovery only suffers a small deviation provided the noise level is small enough, under the same conditions as Theorem 1.

Theorem 2. *Consider the model $\mathbf{Y} = \mathcal{C}(\mathbf{g})\mathbf{X} + \mathbf{N}$, where $\mathbf{N}$ is the observation noise. Under the same condition as Theorem 1, the solution $\hat{\mathbf{v}}$ of the proposed algorithm* (8) *satisfies*

$$\|\hat{\mathbf{v}} - \mathbf{h}\|_1 \leq \frac{4np\|\mathbf{h}\|_1^2\|\mathbf{N}\|_\infty}{L - np\|\mathbf{h}\|_1\|\mathbf{N}\|_\infty}, \tag{20}$$

where $L = ((1-\delta)np - 2(1+\delta)(1+\delta_2)np\theta)\frac{\mu}{4}\sqrt{\frac{\theta}{n}}$, provided that $|h|_{(2)} \leq \frac{(1-\delta)-2(1+\delta)(1+\delta_2)\theta}{8(1+\delta_1)\sqrt{\theta n}}$ and $\|\mathbf{N}\|_\infty \leq \frac{L}{np\|\mathbf{h}\|_1}$. Here, $\|\cdot\|_\infty$ denotes the entry-wise ℓ_∞ norm.

Due to lack of space, the proof of Theorem 2 is presented in the supplemental material.

V. Numerical Example

We present several numerical examples to showcase the performance of the proposed algorithm. In these examples, the entries of $\mathbf{X}$ are drawn i.i.d. from a Bernoulli-Gaussian distribution. The parameter for the Bernoulli distribution is θ and the Gaussian distribution is chosen as the standard normal distribution $\mathcal{N}(0,1)$. The filter $\mathbf{g}$ is generated as $\mathbf{g} = \mathbf{e}_1 + \beta\mathbf{r}$ to avoid shift ambiguity, where $\mathbf{r}$ is composed of i.i.d. standard Gaussian entries and β is a small constant. We rescale the estimate $\hat{\mathbf{X}}$ to have the same ℓ_1 norm as $\mathbf{X}$, and an exact recovery is declared whenever the relative error $\|\hat{\mathbf{X}} - \mathbf{X}\|_1/\|\mathbf{X}\|_1 < 10^{-3}$. The proposed algorithm (8) is a standard convex programming which is solved numerically via the CVX package [21].

Fig. 1 shows the phase transition of the proposed algorithm with respect to θ and p when $n = 10$ for various values of β, where the recovery success rate is calculated over 20 Monte Carlo simulations per cell. It can be seen that the performance improves when $\mathbf{g}$ becomes more peaky, which also serves as a validation of Theorem 1. In order to explore the sample complexity of the proposed algorithm, Fig. 2 shows the phase transition with respect to n and p under $\beta = 0.3$ and various values of θ, where the recovery success rate is calculated over 20 Monte Carlo simulations per cell. From the numerical results, we conjecture that the algorithm succeeds with a high probability as soon as p scales as a polynomial of $\log n$, indicating very few samples are sufficient for blind deconvolution with sparse inputs. We refer the readers to additional numerical experiments in the supplementary material.

In Fig. 3, we demonstrate how the relative recovery error of the proposed algorithm changes with respect to the amplitude of the observation noise. We set $n = 10$, $\beta = 0.3$, $\theta = 0.3$ and $p = 5$. A uniform random noise on $[-\|\mathbf{N}\|_\infty, -\|\mathbf{N}\|_\infty]$ is applied, and the recovery error is averaged over 10 trails. We can see that the proposed algorithm yields a stable recovery, provided that the amplitude of the noise is small enough.

VI. Conclusion

Blind deconvolution with multiple sparse inputs has been studied in this letter. Because of the sparsity assumption of the signal, an ℓ_1-minimization algorithm has been proposed to solve the problem efficiently, where sufficient conditions for exact and stable recovery have been developed. We have demonstrated that our algorithm yields promising results. Through numerical simulations we conjecture that both the identifiably and performance guarantee of the proposed algorithm have rooms for improvement in terms of the sample complexity, which we leave for future work.

REFERENCES

[1] H. Q. Nguyen, S. Liu, and M. N. Do, "Subspace methods for computational relighting," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013, pp. 865 703–865 703.

[2] S.-i. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*. IEEE, 1997, pp. 101–104.

[3] A. Paulraj and T. Kailath, "Direction of arrival estimation by eigenstructure methods with unknown sensor gain and phase," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 10. IEEE, 1985, pp. 640–643.

[4] F. Sroubek and J. Flusser, "Multichannel blind iterative image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 9, pp. 1094–1106, 2003.

[5] K. Lee, Y. Li, M. Junge, and Y. Bresler, "Stability in blind deconvolution of sparse signals and reconstruction by alternating minimization," in *Int. Conf. Sampling Theory and Applications (SampTA)*, 2015.

[6] Y. Chi, "Guaranteed blind sparse spikes deconvolution via lifting and convex optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. PP, no. 99, pp. 1–1, 2016.

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

[8] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *arXiv preprint arXiv:1501.06864*, 2015.

[9] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *Signal Processing, IEEE Transactions on*, vol. 43, no. 12, pp. 2982–2993, 1995.

[10] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, April 2006.

[11] Y. Li, K. Lee, and Y. Bresler, "A unified framework for identifiability analysis in bilinear inverse problems with applications to subspace and sparsity models," *arXiv preprint arXiv:1501.06120*, 2015.

[12] J. Gao, E. Ertin, S. Kumar, and M. al'Absi, "Contactless sensing of physiological signals using wideband RF probes," in *2013 Asilomar Conference on Signals, Systems and Computers*. IEEE, 2013, pp. 86–90.

[13] D. A. Spielman, H. Wang, and J. Wright, "Exact recovery of sparsely-used dictionaries," *arXiv preprint arXiv:1206.5882*, 2012.

[14] K. Luh and V. Vu, "Dictionary learning with few samples and matrix concentration," *arXiv preprint arXiv:1503.08854*, 2015.

[15] Q. Qu, J. Sun, and J. Wright, "Finding a sparse vector in a subspace: Linear sparsity using alternating directions," in *Advances in Neural Information Processing Systems*, 2014, pp. 3401–3409.

[16] L. Demanet and P. Hand, "Scaling law for recovering the sparsest element in a subspace," *Information and Inference*, p. iau007, 2014.

[17] R. Gribonval, G. Chardon, and L. Daudet, "Blind calibration for compressed sensing by convex optimization," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2713–2716.

[18] C. Bilen, G. Puy, R. Gribonval, and L. Daudet, "Convex optimization approaches for blind sensor calibration using sparsity," *Signal Processing, IEEE Transactions on*, vol. 62, no. 18, pp. 4847–4856, 2014.

[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 2012.

[20] C. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 167–180, 1968.

[21] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," *Online accessible: http://stanford. edu/˜ boyd/cvx*, 2008.

Blind Deconvolution from Multiple Sparse Inputs: Supplementary Material

Liming Wang, *Member, IEEE,* and Yuejie Chi, *Member, IEEE*

I. Proof of Theorem 2

Proof. We follow the same steps and symbols in the proof of Theorem 1, and (8) is now equivalent to

$$\hat{\mathbf{w}} = \operatorname*{argmin}_{\mathbf{w}\in\mathbb{R}^n} \|\mathcal{C}(\mathbf{w})\mathbf{X} + \mathcal{C}(\mathbf{h})\mathcal{C}(\mathbf{w})\mathbf{N}\|_1, \text{subject to} \quad \mathbf{r}^T\mathbf{w} = 1.$$

By the proof of Theorem 1, we have

$$\begin{aligned}
\|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X}\|_1 &\geq \|\mathbf{X}\|_1 - |r|_{(2)}\|\mathbf{w}_2\|_1\|\mathbf{X}\|_1 - 2(1+\delta)|S|\alpha + (1-\delta)np\alpha \\
&\geq \|\mathbf{X}\|_1 + ((1-\delta)np - 2(1+\delta)|S|)\frac{\mu}{4}\sqrt{\frac{\theta}{n}} - |r|_{(2)}\|\mathbf{X}\|_1)\|\mathbf{w}_2\|_1 \qquad (1)\\
&\geq \|\mathbf{X}\|_1 + ((1-\delta)np - 2(1+\delta)(1+\delta_2)np\theta)\frac{\mu}{4}\sqrt{\frac{\theta}{n}} - |r|_{(2)}\|\mathbf{X}\|_1)\|\mathbf{w}_2\|_1. \qquad (2)
\end{aligned}$$

Via the assumption on $|r|_{(2)}$ and A2), we have

$$\begin{aligned}
|r|_{(2)} &\leq \frac{(1-\delta) - 2(1+\delta)(1+\delta_2)\theta}{(1+\delta_1)\mu\theta}\frac{\mu}{8}\sqrt{\frac{\theta}{n}} \qquad (3)\\
&\leq \frac{(1-\delta)np - 2(1+\delta)(1+\delta_2)np\theta}{\|\mathbf{X}\|_1}\frac{\mu}{8}\sqrt{\frac{\theta}{n}}. \qquad (4)
\end{aligned}$$

Plug (4) in (2), we have

$$\|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X}\|_1 \geq \|\mathbf{X}\|_1 + ((1-\delta)np - 2(1+\delta)(1+\delta_2)np\theta)\frac{\mu}{4}\sqrt{\frac{\theta}{n}}\|\mathbf{w}_2\|_1 \qquad (5)$$

Denote $L = ((1-\delta)np - 2(1+\delta)(1+\delta_2)np\theta)\frac{\mu}{4}\sqrt{\frac{\theta}{n}}$, and we simply have $\|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X}\|_1 \geq \|\mathbf{X}\|_1 + L\|\mathbf{w}_2\|_1$.

Hence,

$$\begin{aligned}
&\|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X} + \mathcal{C}(\mathbf{h})\mathcal{C}(\hat{\mathbf{w}})\mathbf{N}\|_1 \geq \|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X}\|_1 - \|\mathcal{C}(\mathbf{h})\mathcal{C}(\hat{\mathbf{w}})\mathbf{N}\|_1 \\
&\geq \|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X}\|_1 - \|w_1\mathcal{C}(\mathbf{h})\mathbf{N}\|_1 - \|\mathcal{C}(\mathbf{h})\mathcal{C}(\mathbf{w}_2)\mathbf{N}\|_1 \qquad (6)\\
&\geq \|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X}\|_1 - \|\mathcal{C}(\mathbf{h})\mathbf{N}\|_1 - np\|\mathbf{w}_2\|_1\|\mathcal{C}(\mathbf{h})\mathbf{N}\|_\infty \qquad (7)\\
&\geq \|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X}\|_1 - \|\mathcal{C}(\mathbf{h})\mathbf{N}\|_1 - np\|\mathbf{w}_2\|_1\|\mathbf{h}\|_1\|\mathbf{N}\|_\infty \\
&\geq \|\mathbf{X}\|_1 + L\|\mathbf{w}_2\|_1 - \|\mathcal{C}(\mathbf{h})\mathbf{N}\|_1 - np\|\mathbf{w}_2\|_1\|\mathbf{h}\|_1\|\mathbf{N}\|_\infty, \qquad (8)
\end{aligned}$$

where (6) and (7) follow from the inequalities $\|\mathcal{C}(\mathbf{h})\mathcal{C}(\mathbf{w}_2)\mathbf{N}\|_1 \leq np\|\mathcal{C}(\mathbf{h})\mathbf{N}\|_\infty\|\mathbf{w}_2\|_1$ and $\|\mathcal{C}(\mathbf{h})\mathbf{N}\|_\infty \leq \|\mathbf{N}\|_\infty\|\mathbf{h}\|_1$.

On the other hand, we have $\|\mathcal{C}(\hat{\mathbf{w}})\mathbf{X} + \mathcal{C}(\mathbf{h})\mathcal{C}(\hat{\mathbf{w}})\mathbf{N}\|_1 \leq \|\mathbf{X}\|_1 + \|\mathcal{C}(\mathbf{h})\mathbf{N}\|_1$ due to optimality of $\hat{\mathbf{w}}$. Combining the lower and upper bounds, we have

$$\|\mathbf{X}\|_1 + L\|\mathbf{w}_2\|_1 - \|\mathcal{C}(\mathbf{h})\mathbf{N}\|_1 - np\|\mathbf{w}_2\|_1\|\mathbf{h}\|_1\|\mathbf{N}\|_\infty \leq \|\mathbf{X}\|_1 + \|\mathcal{C}(\mathbf{h})\mathbf{N}\|_1. \qquad (9)$$

Therefore, we obtain

$$\|\mathbf{w}_2\|_1 \leq \frac{2\|\mathcal{C}(\mathbf{h})\mathbf{N}\|_1}{L - np\|\mathbf{h}\|_1\|\mathbf{N}\|_\infty} \tag{10}$$

$$\leq \frac{2np\|\mathbf{h}\|_1\|\mathbf{N}\|_\infty}{L - np\|\mathbf{h}\|_1\|\mathbf{N}\|_\infty}, \tag{11}$$

provided that

$$\|\mathbf{N}\|_\infty \leq \frac{L}{np\|\mathbf{h}\|_1}. \tag{12}$$

We know that $\mathbf{e}_1$ is the solution of (8) when $\mathbf{N} = \mathbf{0}$, and the estimation perturbation caused by the noise can be expressed as

$$\|\hat{\mathbf{w}} - \mathbf{e}_1\|_1 = \|(w_1 - 1)\mathbf{e}_1 + \mathbf{w}_2\|_1 \tag{13}$$

$$\leq |\mathbf{r}^T\mathbf{w}_2| + \|\mathbf{w}_2\|_1 \tag{14}$$

$$\leq 2\|\mathbf{w}_2\|_1, \tag{15}$$

where the last inequality follows from $\|\mathbf{r}\|_\infty = 1$. Let $\hat{\mathbf{v}}$ denote the solution to (8). Due to the transformation $\mathbf{v} = \mathcal{C}(\mathbf{h})\mathbf{w} = \mathbf{h} \circledast \mathbf{w}$, we have

$$\|\hat{\mathbf{w}} - \mathbf{e}_1\|_1 = \|\mathcal{C}(\mathbf{g})\hat{\mathbf{v}} - \mathcal{C}(\mathbf{g})\mathbf{h}\|_1 \leq 2\|\mathbf{w}_2\|_1. \tag{16}$$

Multiply $\|\mathcal{C}(\mathbf{h})\|_{\ell_1}$ on both sides of the inequality where $\|\cdot\|_{\ell_1}$ denotes the ℓ_1 operator norm of the argument matrix, we have

$$\|\mathcal{C}(\mathbf{h})\|_{\ell_1} 2\|\mathbf{w}_2\|_1 \geq \|\mathcal{C}(\mathbf{h})\|_{\ell_1}\|\mathcal{C}(\mathbf{g})\hat{\mathbf{v}} - \mathcal{C}(\mathbf{g})\mathbf{h}\|_1 \tag{17}$$

$$\geq \|\hat{\mathbf{v}} - \mathbf{h}\|_1, \tag{18}$$

where the last inequality follows from the fact $\|\mathcal{C}(\mathbf{h})\|_{\ell_1} \geq \frac{\|\mathcal{C}(\mathbf{h})\mathbf{v}\|_1}{\|\mathbf{v}\|_1}$, for any non-vanishing $\mathbf{v} \in \mathbb{R}^n$. Moreover, since $\|\mathcal{C}(\mathbf{h})\|_{\ell_1} = \|\mathbf{h}\|_1$, we obtain

$$\|\hat{\mathbf{v}} - \mathbf{h}\|_1 \leq \frac{4np\|\mathbf{h}\|_1^2\|\mathbf{N}\|_\infty}{L - np\|\mathbf{h}\|_1\|\mathbf{N}\|_\infty}, \tag{19}$$

provided that

$$\|\mathbf{N}\|_\infty \leq \frac{L}{np\|\mathbf{h}\|_1}. \tag{20}$$

□

II. Additional Numerical Example

We note that the condition in Theorem 1 is only a sufficient condition. Practically, we found that the algorithm works well for many other situations. For example, in the following experiment, we consider the kernel to be generated by i.i.d. standard Gaussian $\mathcal{N}(0,1)$. We set $n = 10$, $\beta = 0.3$, $\theta = 0.3$ and $p = 5$. The generated $\mathbf{g}$ is

$$\mathbf{g} = \begin{pmatrix} 0.8404 & -0.888 & 0.1001 & -0.5445 & 0.3035 & -0.6003 & 0.49 & 0.7394 & 1.712 & -0.1941 \end{pmatrix}^T,$$

and the generated data $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} 0 & -1.089 & 0 & 0 & 0 \\ 0 & 0 & 0.7481 & 0.2916 & 0 \\ -1.214 & 0 & 0 & 0.1978 & 0 \\ 0 & 0 & 0 & 1.588 & 0 \\ 0 & 0 & 0 & -0.8045 & -0.6669 \\ 1.533 & 0.08593 & 0 & 0 & 0 \\ -0.7697 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.1961 & -1.166 & 0 \end{pmatrix}.$$

As we discussed in the letter, recovery should be interpreted in the sense that $\mathbf{g}$ and $\mathbf{X}$ are accurately recovered up to a scaling factor. The estimated kernel $\hat{\mathbf{g}}$ after rescaling is

$$\hat{\mathbf{g}} = \begin{pmatrix} 0.8404 & -0.888 & 0.1001 & -0.5445 & 0.3035 & -0.6003 & 0.49 & 0.7394 & 1.712 & -0.1941 \end{pmatrix}^T,$$

and the estimated data $\hat{\mathbf{X}}$ after rescaling is

$$\hat{\mathbf{X}} = \begin{pmatrix} 2.725 \cdot 10^{-7} & -1.089 & 2.386 \cdot 10^{-7} & -4.214 \cdot 10^{-7} & 6.012 \cdot 10^{-9} \\ -7.748 \cdot 10^{-7} & -4.716 \cdot 10^{-7} & 0.7481 & 0.2916 & 3.531 \cdot 10^{-8} \\ -1.214 & -1.881 \cdot 10^{-7} & 3.338 \cdot 10^{-7} & 0.1978 & -1.124 \cdot 10^{-7} \\ -2.258 \cdot 10^{-7} & 7.215 \cdot 10^{-8} & 1.279 \cdot 10^{-7} & 1.588 & -2.883 \cdot 10^{-7} \\ 3.282 \cdot 10^{-7} & 4.697 \cdot 10^{-8} & -9.976 \cdot 10^{-8} & -0.8045 & -0.6669 \\ 1.533 & 0.08593 & -4.976 \cdot 10^{-9} & -8.278 \cdot 10^{-8} & -2.883 \cdot 10^{-7} \\ -0.7697 & 4.697 \cdot 10^{-8} & 2.399 \cdot 10^{-7} & -7.026 \cdot 10^{-8} & -1.124 \cdot 10^{-7} \\ -4.469 \cdot 10^{-7} & 7.215 \cdot 10^{-8} & -3.979 \cdot 10^{-8} & -1.102 \cdot 10^{-7} & 3.531 \cdot 10^{-8} \\ -1.999 \cdot 10^{-7} & -1.881 \cdot 10^{-7} & -1.244 \cdot 10^{-7} & -2.69 \cdot 10^{-8} & 6.012 \cdot 10^{-9} \\ 9.122 \cdot 10^{-8} & -4.716 \cdot 10^{-7} & -0.1961 & -1.166 & -2.046 \cdot 10^{-7} \end{pmatrix}.$$

Even though the settings do not necessarily satisfy the condition in Theorem 1, an exact recovery is still achieved.