Sparse Recovery from Convolved Output in Underwater Acoustic Relay Networks

Sunav Choudhary and Urbashi Mitra
Ming Hsieh Dept. of Electrical Engineering
University of Southern California, Los Angeles
{*sunavcho,ubli*}*@usc.edu*

***Abstract*—This paper explores criteria for unique recovery from blind deconvolution under sparsity priors. Additionally regularizing functions stemming from this problem framework are developed. For key cases, it is possible to ensure unique recoverability given the regularized problem statement. The uniqueness results are informed by a matrix completion-based viewpoint of blind deconvolution. Furthermore, this perspective enables characterization of why blind deconvolution with two sparse inputs is an inherently hard problem. Two blind deconvolution algorithms are proposed which do not rely on alternating between the estimation of one input signal, while holding the other constant. Evaluation of the algorithms is done via simulation and shown to significantly outperform a previously proposed method. Furthermore, numerical illustration of recovery failure considering sparsity of input signals that do not satisfy the recovery constraints is also provided.**

***Index Terms*—sparse recovery, blind deconvolution, underwater acoustic communications**

I. INTRODUCTION

Underwater acoustic communication (UAC) systems have the potential to enable or enhance key ocean–related applications such as: scientific data collection, pollution monitoring, tactical surveillance and disaster prevention [1]. Cooperative communication with multi–hopping for terrestrial sensor networks has been extensively studied enabling power savings and improved fidelity [2]. Given the significant attenuating effects in underwater acoustics [3], cooperative communications are of major interest. Most schemes (such as distributed space-time coding and equalization in [4]) require channel state information. In a previous paper [5], we developed a structured channel model based on the *multichannel approximation* for the second hop in a relay assisted communication topology shown in Fig. 1. Herein, we develop two channel estimation approaches explicitly exploiting this channel structure along with an investigation of uniqueness of the estimate. We observe that the network topology in Fig. 1 naturally leads to a *blind deconvolution* problem where both signals are known to be *sparse*. It is well known [6] that, without additional regularization, blind deconvolution falls into the category of nonlinear ill-posed inverse problems, showing poor stability to the recovery process and admitting multiple solutions (the number of solutions could grow exponentially in length of the convolved output). We derive regularizers specific to our scenario for the blind deconvolution problem (under mild assumptions). In particular, we investigate whether physical deployment of relays in a network can make the blind deconvolution problem well-posed.

This research has been funded in part by the following grants and organizations: ONR N00014-09-1-0700, NSF CNS-0832186, NSF CNS-0821750 (MRI), NSF CCF-0917343, NSF CCF-1117896 and DOT CA-26-7084-00.

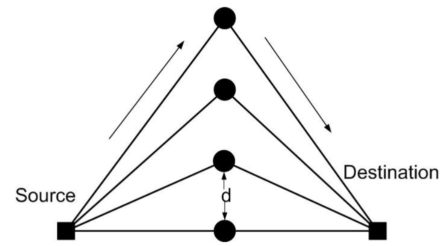

Fig. 1. Two hop relay assisted communication link topology with four cooperating nodes.

Blind deconvolution under smoothness priors and statistical priors has been well studied in the signal processing as well as wireless communications literature (e.g. see [7], [8]). Only recently has signal sparsity been exploited for inverse problems in image processing [9]. Our interest is motivated by the fact that point to point links in underwater acoustic communications exhibit sparse channels [10]. Much recent work on blind deconvolution with sparsity priors [5], [11]–[13] takes an algorithmic approach based on the very popular *alternating minimization* heuristic, i.e. estimating one sparse signal while holding the other fixed and iterating. Although the alternating minimization approach to blind deconvolution is guaranteed to converge to a local minimum of the objective function [6], the point of convergence is highly dependent on initialization.

Blind deconvolution was also examined in [14]. We shall adopt several observations from [14] relating to the operation of convolution herein. However, there is a key difference in that [14] allows for a random linear precoding operation which enables a mapping to more classical compressed sensing [15] problems. The relevant linear map for our recovery problem thus does not possess the same desirable properties as that of [14]. As such, the question of establishing uniqueness for recovery remains open.

The contributions of this work are as follows:

1) We map the two hop cooperative communication problem

to that of blind deconvolution.

2) We identify good regularizers for this problem.
3) We show that a matrix rank minimization approach towards blind deconvolution allows us to provide a clean answer to the unique recoverability question.
4) We argue using matrix completion ideas that a sparsity prior on both input signals leads to poor recovery performance.
5) We provide results on unique recoverability for both the "well separated" as well as the "not well separated" cases of blind deconvolution.
6) We propose two different channel estimation algorithms for the second hop estimation, neither of them based on alternating minimization or convex relaxation techniques, and provide a numerical comparison to an alternating minimization based technique with significant improvement.

This paper is organized as follows. In Section III, the system model for the UWA relay channel is developed along with a discussion of possible regularizers. We investigate conditions for unique recoverability for the regularized versions of the blind deconvolution problem in Section IV. We also show that blind deconvolution of two sparse signals is inherently hard. We provide uniqueness results for both "well separated" and "not well separated" scenarios of the problem. Section V describes two channel estimation algorithms for the second hop and numerical comparisons are provided in Section VI. Section VII concludes the paper. Appendix A provides a uniqueness proof for the "not well separated" scenario.

II. Notation

We shall use lowercase boldface alphabets to denote column vectors (e.g. $\boldsymbol{z}$). The k^{th} element of a vector $\boldsymbol{z}$ will be denoted by z_k or $(\boldsymbol{z})_k$. In contrast, $\boldsymbol{z}_k$ denotes the k^{th} vector in a sequence of vectors (which is not the same as $(\boldsymbol{z})_k$). For a vector $\boldsymbol{z}$, $\mathrm{N}_{\boldsymbol{z}}$ shall denote the length of $\boldsymbol{z}$.

Matrices are denoted by uppercase boldface letters (e.g. $\boldsymbol{A}$). The $(i,j)^{\text{th}}$ element of the matrix $\boldsymbol{A}$ will be denoted by A_{ij}, $A_{i,j}$, $(\boldsymbol{A})_{ij}$ or $(\boldsymbol{A})_{i,j}$. The j^{th} column of the matrix $\boldsymbol{A}$ will be denoted by $(\boldsymbol{A})_j$. The uppercase and lowercase boldface versions of the same alphabet, respectively representing a matrix and a vector (e.g. $\boldsymbol{H}$ and $\boldsymbol{h}$), are *unrelated* unless explicitly mentioned otherwise.

The Schur product will be denoted by $\odot$ and the standard Euclidean inner product will be denoted by $\langle\cdot,\cdot\rangle$ (for matrices it denotes the trace inner product). For a set M, $|M|$ shall denote its cardinality. The sign of a real number x will be denoted by $\mathrm{sgn}(x)$ and is defined as,

$$\mathrm{sgn}(x) \triangleq \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases} \tag{1}$$

Unless otherwise specified, functions defined on real numbers are assumed to be extended to operate elementwise on vectors as well as matrices. For example, given a vector $\boldsymbol{z}$ and a $m \times n$ matrix $\boldsymbol{A}$, the vectors $|\boldsymbol{z}|$ and $\mathrm{sgn}(\boldsymbol{z})$ and the matrices $|\boldsymbol{A}|$ and $\mathrm{sgn}(\boldsymbol{A})$ are defined by

$$\begin{aligned} (|\boldsymbol{z}|)_k &= |z_k|\,, & (\mathrm{sgn}(\boldsymbol{z}))_k &= \mathrm{sgn}(z_k)\,, \\ (|\boldsymbol{A}|)_{ij} &= |A_{ij}| \quad \text{and,} & (\mathrm{sgn}(\boldsymbol{A}))_{ij} &= \mathrm{sgn}(A_{ij})\,, \end{aligned} \tag{2}$$

for all $1 \le k \le \mathrm{N}_{\boldsymbol{z}}$, $1 \le i \le m$ and $1 \le j \le n$. For a vector $\boldsymbol{z}$ and a matrix $\boldsymbol{A}$, the support will be denoted by $\mathrm{supp}(\boldsymbol{z})$ and $\mathrm{supp}(\boldsymbol{A})$, and defined as,

$$\mathrm{supp}(\boldsymbol{z}) \triangleq |\mathrm{sgn}(\boldsymbol{z})| \quad \text{and,} \quad \mathrm{supp}(\boldsymbol{A}) \triangleq |\mathrm{sgn}(\boldsymbol{A})|\,, \tag{3}$$

respectively. The ℓ_0-pseudonorm, which counts the number of nonzero elements in its argument, will be denoted by $\|\cdot\|_0$ for both vectors and matrices. Specifically, for any vector $\boldsymbol{z}$ and $m \times n$ matrix $\boldsymbol{A}$, we have

$$\|\boldsymbol{z}\|_0 \triangleq \sum_k \mathrm{supp}(\boldsymbol{z}) \quad \text{and,} \quad \|\boldsymbol{A}\|_0 \triangleq \sum_{i,j} \mathrm{supp}(\boldsymbol{A})\,, \tag{4}$$

for all $1 \le k \le \mathrm{N}_{\boldsymbol{z}}$, $1 \le i \le m$ and $1 \le j \le n$.

III. System Model

We shall consider the two-hop relay assisted communication topology shown in Fig. 1. We show that the signal model in [5] maps to blind deconvolution. We derive regularizers for the inverse problem. We make some sparse recovery related terminology precise via the following definitions:

Definition 1. The support of a vector $\boldsymbol{z}$, denoted by $\mathrm{supp}(\boldsymbol{z})$, is defined as the binary vector,

$$\mathrm{supp}(\boldsymbol{z}) \triangleq |\mathrm{sgn}(\boldsymbol{z})|\,. \tag{5}$$

We say that $\boldsymbol{z}$ is supported on $\boldsymbol{x}$, or equivalently, $\boldsymbol{x}$ is the support vector for $\boldsymbol{z}$ if $\boldsymbol{x} = \mathrm{supp}(\boldsymbol{z})$.

Definition 2. A vector $\boldsymbol{z}$ is said to be exactly N-sparse, or equivalently, $\boldsymbol{z}$ is said to have sparsity N if

$$\|\boldsymbol{z}\|_0 = N. \tag{6}$$

We say that $\boldsymbol{z}$ is "sparse" in the standard basis if $\|\boldsymbol{z}\|_0 \ll \mathrm{N}_{\boldsymbol{z}}$.

We make the following assumptions about our system:

Assumption 1. Each point-to-point link is sparse in the time domain.

Assumption 2. All relay-destination channels on the second hop have a common sparse support.

Assumption 3. All relays use a common training sequence for the purpose of channel estimation at the destination.

Assumption 4. The relays transmit at equal power and can compensate for relative phase differences.

Assumptions 1 and 2 are justified in [5]. Assumption 3 is justified by practical considerations (e.g. see [16]). Assumption 4 can be justified by recent developments in full duplex wireless (e.g. see [17]).

For our two-hop link in Fig. 1, there are, in fact, two different channel estimation problems to be solved. The first hop estimation problem is that of a single-input-single-output

(SISO) source-relay channel impulse response (CIR) which can be solved using standard sparse recovery techniques (e.g. see [15]) and shall not be examined here. The second hop estimation problem involves estimating each of the relay-destination point-to-point links at the destination and, unlike the first hop estimation problem, the estimation tasks for the relay-destination channels are, in general, *not* independent. In particular, the channel seen by the destination is a multiple-input-single-output (MISO) channel with structural relations given by Assumption 2.

Assuming K participating relays, we denote the K SISO relay-destination CIRs by $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_K$ and their common sparse support vector by $\boldsymbol{b}$. We also define the square diagonal matrix $\boldsymbol{H}_k = \operatorname{diag}(\boldsymbol{h}_k)$, for each $1 \leq k \leq K$. We have by definition,

$$\boldsymbol{h}_k \odot \boldsymbol{b} = \boldsymbol{h}_k = \boldsymbol{H}_k \boldsymbol{b}, \quad \forall\, 1 \leq k \leq K. \tag{7}$$

Let the propagation delays for the K relay-destination links in the tapped-delay line model be denoted respectively by $\tau_1, \tau_2, \ldots, \tau_K$. Let the τ-delay (τ-downshifting) matrix operator be denoted by $\boldsymbol{D}_\tau$ which is defined by,

$$(\boldsymbol{D}_\tau)_{ij} = \begin{cases} 1, & i - j = \tau, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

for non-negative integers τ. The second hop MISO CIR, as seen by the destination, will be a superposition of delayed versions of the K SISO relay-destination CIRs $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_K$ with the delays being $\tau_1, \tau_2, \ldots, \tau_K$ respectively. Denoting the second hop MISO CIR by $\boldsymbol{\beta}$ we have,

$$\boldsymbol{\beta} = \sum_{k=1}^{K} \boldsymbol{D}_{\tau_k} \boldsymbol{h}_k = \sum_{k=1}^{K} \boldsymbol{D}_{\tau_k} \boldsymbol{H}_k \boldsymbol{b} = \boldsymbol{L}\boldsymbol{b}, \tag{9}$$

where $\boldsymbol{L}$ is a tall banded matrix and is defined as,

$$\boldsymbol{L} = \sum_{k=1}^{K} \boldsymbol{D}_{\tau_k} \boldsymbol{H}_k. \tag{10}$$

By virtue of the above definition, $\boldsymbol{L}$ will be zero above its first (main) diagonal and below its last diagonal. Further, atmost K diagonals in $\boldsymbol{L}$ can have nonzero elements. Finally, if such a diagonal (with nonzero elements) is written out as a column vector, then this vector is supported on $\boldsymbol{b}$.

As the channel is finite-impulse-response (FIR), the output to the common training sequence is a convolution. Let the common training sequence be denoted by the lower triangular Toeplitz matrix $\boldsymbol{T}$. Then the noiseless received signal at the destination is given by,

$$\sum_{k=1}^{K} \boldsymbol{D}_{\tau_k} \boldsymbol{T} \boldsymbol{h}_k = \sum_{k=1}^{K} \boldsymbol{T} \boldsymbol{D}_{\tau_k} \boldsymbol{h}_k = \boldsymbol{T} \left(\sum_{k=1}^{K} \boldsymbol{D}_{\tau_k} \boldsymbol{h}_k \right) = \boldsymbol{T}\boldsymbol{\beta}. \tag{11}$$

Note that if $\boldsymbol{T}$ is an invertible matrix, then our second hop channel estimation problem is the same as the inverse problem of estimating $\boldsymbol{L}$ and $\boldsymbol{b}$ from the observation of $\boldsymbol{\beta} = \boldsymbol{L}\boldsymbol{b}$ taking into account the structural properties of $\boldsymbol{L}$ and $\boldsymbol{b}$ described previously.

By some elementary manipulations, we can convert the expression $\boldsymbol{\beta} = \boldsymbol{L}\boldsymbol{b}$ into a form which is linear in the unknown with a rank constraint on the domain. An analogous observation was made in [14]. Let the k^{th} diagonal of the matrix $\boldsymbol{L}$ be written out as a row vector and form the k^{th} row of a matrix $\boldsymbol{W}$. In particular we have,

$$W_{ij} = L_{j,i+j-1}, \tag{12}$$

for the appropriate range of indices i and j. It is easy to see that the elements of the vector $\boldsymbol{\beta}$ are given by the matrix inner product expressions,

$$\beta_k = \langle \boldsymbol{W}, \boldsymbol{S}_k \rangle, \tag{13}$$

where the matrices $\boldsymbol{S}_k$ are given by,

$$(\boldsymbol{S}_k)_{ij} = \begin{cases} 1, & i + j = k + 1, \\ 0, & \text{otherwise}, \end{cases} \tag{14}$$

for $1 \leq k \leq \mathrm{N}_{\boldsymbol{\beta}}$.

We note that each row of $\boldsymbol{W}$ is supported on the sparse support vector $\boldsymbol{b}^{\mathrm{T}}$ so that the matrix $\boldsymbol{W}$ is column-sparse, i.e. only a few columns of $\boldsymbol{W}$ have nonzero elements. A similar argument also applies if the roles of the rows and columns of $\boldsymbol{W}$ are interchanged, since $\boldsymbol{W}$ has at most K nonzero rows by construction, which makes $\boldsymbol{W}$ row-sparse as well. These row and column sparsity features of $\boldsymbol{W}$ form our first set of regularizers for the inverse problem. Under the assumption of similar multipath propagation environments for each relay-destination link (see Assumption 2) the CIRs $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_K$ are roughly expected to be scalar multiples of each other. This means that the rows of $\boldsymbol{W}$ are approximately scalar multiples of a single row which gives us a low rank (rank-1) regularizer for the inverse problem.

By Assumption 4, the CIRs $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_K$ are observed to have the same phase at the destination. Combined with the rank-1 approximation of $\boldsymbol{W}$ already described this would mean that all nonzero rows of $\boldsymbol{W}$ are equal (or approximately equal) resulting in yet another regularizing constraint for our inverse problem.

IV. Design Considerations for Unique Recovery

Based on the rank-1 domain constraints discussed in Section III, we could pose our inverse problem as,

$$\begin{aligned} &\text{find} && \boldsymbol{W} \\ &\text{subject to} && \operatorname{rank}(\boldsymbol{W}) = 1, \\ & && \langle \boldsymbol{W}, \boldsymbol{S}_k \rangle = \beta_k, \quad \forall 1 \leq k \leq \mathrm{N}_{\boldsymbol{\beta}}, \end{aligned} \tag{P$_0$}$$

Given the specific structure of the matrices $\boldsymbol{S}_k$, for all $1 \leq k \leq \mathrm{N}_{\boldsymbol{\beta}}$, and the rank-1 constraint on the domain of $\boldsymbol{W}$, it is not difficult to see that Problem (P$_0$) translates into deconvolving $\boldsymbol{\beta}$ into the two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ which respectively form appropriately scaled versions of $(\boldsymbol{W})_1$ and $\left(\boldsymbol{W}^{\mathrm{T}}\right)_1$ (a blind deconvolution problem). Thus, Problem (P$_0$) restated in terms of the optimization variables $\boldsymbol{x}$ and $\boldsymbol{y}$, becomes,

$$\begin{aligned} &\text{find} && \boldsymbol{x}, \boldsymbol{y} \\ &\text{subject to} && \boldsymbol{x} \star \boldsymbol{y} = \boldsymbol{\beta}. \end{aligned} \tag{P$_1$}$$

It was noted in Section III that $\boldsymbol{W}$, as an optimization variable in Problem (P_0), is row-sparse as well as column-sparse. To make this notion concrete, we introduce the pseudonorms $\|\boldsymbol{W}\|_{\mathrm{r},0}$ and $\|\boldsymbol{W}\|_{\mathrm{c},0}$ respectively denoting the number of nonzero rows and the number of nonzero columns in $\boldsymbol{W}$. If matrix $\boldsymbol{W}$ is of dimension $m \times n$, $\|\boldsymbol{W}\|_{\mathrm{r},0}$ and $\|\boldsymbol{W}\|_{\mathrm{c},0}$ are defined as,

$$\|\boldsymbol{W}\|_{\mathrm{r},0} \triangleq \left\| \sum_{i=1}^{m} \left(|\boldsymbol{W}|^{\mathrm{T}} \right)_i \right\|_0 \tag{15}$$

and,

$$\|\boldsymbol{W}\|_{\mathrm{c},0} \triangleq \left\| \sum_{j=1}^{n} \left(|\boldsymbol{W}| \right)_j \right\|_0 . \tag{16}$$

Both $\|\cdot\|_{\mathrm{r},0}$ and $\|\cdot\|_{\mathrm{c},0}$ exhibit the same properties as the ℓ_0-pseudonorm, i.e. they satisfy all properties of a norm except positive homogeneity. We thus expect $\|\cdot\|_{\mathrm{r},0}$ (respectively $\|\cdot\|_{\mathrm{c},0}$) to encourage row sparsity (respectively column sparsity) in the regularized version of Problem (P_0), similar to the way $\|\cdot\|_0$ encourages sparsity in standard compressed sensing.

A. Column Sparsity Regularization

In the column-sparse regularized version of Problem (P_0) we have the following recovery problem,

$$\begin{aligned} \underset{\boldsymbol{W}}{\text{minimize}} \quad & \|\boldsymbol{W}\|_{\mathrm{c},0} \\ \text{subject to} \quad & \mathrm{rank}(\boldsymbol{W}) = 1, \\ & \langle \boldsymbol{W}, \boldsymbol{S}_k \rangle = \beta_k, \quad \forall\, 1 \le k \le \mathrm{N}_{\boldsymbol{\beta}}. \end{aligned} \tag{P$_2$}$$

For ease of exposition, we also define a condition which we call the *non-overlapping constraint*. A similar (albeit more restrictive) condition was also proposed in [11] to avoid circular shift ambiguities for the blind *circular* deconvolution problem. Practically, this condition can be interpreted as "no inter-tap interference".

Definition 3. For the vectors $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ satisfying the relationship: $\boldsymbol{z} = \boldsymbol{x} \star \boldsymbol{y}$, we say that vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ satisfy the non-overlapping constraint w.r.t. $\boldsymbol{z}$ if the following holds,

$$\mathrm{supp}(\boldsymbol{z}) = \mathrm{supp}(\boldsymbol{x}) \star \mathrm{supp}(\boldsymbol{y}) . \tag{17}$$

Here, $\star$ denotes the linear convolution operator. Further, if $\mathrm{N}_{\boldsymbol{x}} < \mathrm{N}_{\boldsymbol{y}}$ and each pair of nonzero taps in $\boldsymbol{y}$ is atleast $\mathrm{N}_{\boldsymbol{x}}$ taps apart then $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be "well separated" w.r.t. $\boldsymbol{z}$.

We state, without proof, the following sufficient condition for the solution of Problem (P_2) to be unique:

Theorem 1. *If there exists an optimal solution $\boldsymbol{W}_*$ to Problem* (P_2) *for which $(\boldsymbol{W}_*)_1$ and $\left(\boldsymbol{W}_*^T\right)_1$ are well separated w.r.t. $\boldsymbol{\beta}$, then $\boldsymbol{W}_*$ is the unique solution to Problem* (P_2).

It is worthy of notice that the optimal solution to Problem (P_2) tends to be a row-dense matrix. This is explained by the fact that $\|\boldsymbol{\beta}\|_0 = \|\boldsymbol{W}\|_{\mathrm{r},0} \cdot \|\boldsymbol{W}\|_{\mathrm{c},0}$ when $(\boldsymbol{W})_1$ and $\left(\boldsymbol{W}^{\mathrm{T}}\right)_1$ satisfy the non-overlapping constraint w.r.t. $\boldsymbol{\beta}$. Thus if one of $\|\boldsymbol{W}\|_{\mathrm{r},0}$ or $\|\boldsymbol{W}\|_{\mathrm{c},0}$ decreases then the other must increase as $\|\boldsymbol{\beta}\|_0$ is a constant.

B. Joint Effect of Row and Column Sparsity

In [13], the authors developed the BCD algorithm for blind deconvolution; an observation made therein, based on numerical results, was that assuming a sparisty prior on both input signals yielded degraded performance relative to the assumption on only one input. We endeavor to formalize this result for the general case, independent of recovery algorithm, using ideas from matrix completion [18].

The matrix completion problem studied in [18] has several similarities with our recovery set up in Problem (P_0). In particular, the matrices $\boldsymbol{S}_k$ describing the linear constraints in Problem (P_0) are pairwise orthogonal, sparse in the standard Euclidean basis and their elements assume values from the set $\{0, 1\}$. These characteristics are also true for the matrices describing the linear constraints in the matrix completion problem in [18].

We modify Problem (P_2) to include row-sparse constraint and state it as the following recovery problem,

$$\begin{aligned} \underset{\boldsymbol{W}}{\text{minimize}} \quad & \|\boldsymbol{W}\|_{\mathrm{c},0} \\ \text{subject to} \quad & \|\boldsymbol{W}\|_{\mathrm{r},0} = K, \\ & \mathrm{rank}(\boldsymbol{W}) = 1, \\ & \langle \boldsymbol{W}, \boldsymbol{S}_k \rangle = \beta_k, \quad \forall\, 1 \le k \le \mathrm{N}_{\boldsymbol{\beta}}. \end{aligned} \tag{P$_3$}$$

Consider a rank-1 matrix $\boldsymbol{W}$ which is very sparse in both $\|\cdot\|_{\mathrm{r},0}$ and $\|\cdot\|_{\mathrm{c},0}$ pseudonorms, and $(\boldsymbol{W})_1$ and $\left(\boldsymbol{W}^{\mathrm{T}}\right)_1$ satisfy the non-overlapping constraint w.r.t. $\boldsymbol{\beta}$. As $\boldsymbol{W}$ is rank-1, it is quite sparse in the standard Euclidean basis which leads to high coherence between the sparsity bases for the projection operators $\boldsymbol{S}_k$ and the matrix $\boldsymbol{W}$ in the Problem (P_3). From the arguments in [18], it is now easy to see that the recovery of such a matrix is expected to fail with high probability (due to non-uniqueness of recovery) in the absence of further constraints, regardless of which recovery strategy is used. As $\|\boldsymbol{\beta}\|_0 = \|\boldsymbol{W}\|_{\mathrm{r},0} \cdot \|\boldsymbol{W}\|_{\mathrm{c},0}$, the above mentioned situation arises when the MISO CIR at the destination is quite sparse.

C. Constraints from Assumption 4

We shall adopt the notion of *feasibility* of an optimization problem and the definition of a *feasible point* as described in [19]. We denote by $S(\boldsymbol{z})$, the set of all distinct nonzero values assumed by the elements of a vector $\boldsymbol{z}$. We let $\boldsymbol{x}(z)$ denote the polynomial equivalent of vector $\boldsymbol{x}$, i.e.

$$\boldsymbol{x}(z) = \sum_{j=1}^{\mathrm{N}_{\boldsymbol{x}}} x_j z^{j-1} \tag{18}$$

As noted in Subsection IV-B, we need additional constraints to determine the correct support for $\boldsymbol{W}$. To this end, we invoke Assumption 4 and restrict $(\boldsymbol{W})_1$ to be a binary vector. Then, in terms of the optimization variables $\boldsymbol{x}$ and $\boldsymbol{y}$, Problem (P$_3$) becomes,

$$\begin{aligned} \underset{\boldsymbol{x},\boldsymbol{y}}{\text{minimize}} \quad & \|\boldsymbol{y}\|_0 \\ \text{subject to} \quad & \|\boldsymbol{x}\|_0 = K, \\ & \boldsymbol{x} \star \boldsymbol{y} = \boldsymbol{\beta}, \\ & \boldsymbol{x} \in \{0,1\}^{\mathrm{N}_{\boldsymbol{x}}}. \end{aligned} \tag{P$_4$}$$

Remark 1. If Problem (P$_4$) is feasible then K divides $\|\boldsymbol{\beta}\|_0$.

Remark 2. If $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ is a feasible point for Problem (P$_4$), with $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ satisfying the non-overlapping constraint w.r.t. $\boldsymbol{\beta}$, then $S(\boldsymbol{\beta}) = S(\tilde{\boldsymbol{y}})$.

Theorem 2. *Let K^* be the maximum value of K such that Problem* (P$_4$) *has a feasible point $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ with $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ satisfying the non-overlapping constraint w.r.t. $\boldsymbol{\beta}$. For $K = K^*$, Problem* (P$_4$) *has a unique solution.*

Proof: See Appendix A. ■

V. Algorithmic Approaches to Signal Recovery

We shall consider two recovery strategies. The first one is specific to the set up described in Subsection IV-C (where one of the vectors being convolved is a binary vector) and is called "Decoupled MAP Estimation" (DMAP). The second strategy is based on jointly estimating both input vectors in an iterative fashion different from the alternating minimization approach and is dubbed "Stable Projection" (SP).

A. Decoupled MAP Estimation

Suppose that the convolved vectors are $\boldsymbol{y}$ and $\boldsymbol{x}$ with $\boldsymbol{x}$ binary and $\|\boldsymbol{x}\|_0 = K$ known at the destination. DMAP is a step by step estimation algorithm based on sequentially estimating/detecting $S(\boldsymbol{y})$, supp$(\boldsymbol{y})$ and $\boldsymbol{x}$ in that order, where $S(\boldsymbol{y})$ denotes the set of distinct nonzero elements in $\boldsymbol{y}$. We shall assume that $|S(\boldsymbol{y})| = \|\boldsymbol{y}\|_0$, i.e. all nonzero values in $\boldsymbol{y}$ are distinct. Under the assumption of no noise,

$$M = \|\boldsymbol{y}\|_0 = \frac{\|\boldsymbol{\beta}\|_0}{\|\boldsymbol{x}\|_0} = \frac{\|\boldsymbol{\beta}\|_0}{K}, \tag{19}$$

is known at the receiver. For noisy observation of $\boldsymbol{\beta}$, we assume that we have a good estimate of M.

1) Estimation of $S(\boldsymbol{y})$:
   a) Take the $M \cdot K$ largest magnitude tap values from $\boldsymbol{\beta}$ and sort in descending order in $\boldsymbol{\alpha}$.
   b) Let $S(\boldsymbol{y}) = \{y_1, y_2, \dots, y_M\}$ with $y_i > y_j$ for $i > j$. We estimate $y_j, \forall 1 \le j \le M$ as,
   $$\widehat{y}_j = \frac{1}{K} \sum_{i=1}^{K} \alpha_{(j-1)\cdot K + i}. \tag{20}$$
   Assuming correct ordering of the noiseless contributions, this is the maximum likelihood estimate for additive Gaussian noise.
2) Estimation of supp$(\boldsymbol{y})$:
   c) Let $\Pr(\mathcal{H}_{ij})$ denote the *a priori* probability that the hypothesis $\mathcal{H}_{ij}\colon \beta_j = \widehat{y}_i + \eta_{ij}$ is true, where η_{ij} denotes additive noise and $1 \le j \le \mathrm{N}_{\boldsymbol{\beta}}$. Then *a posteriori* probability p_{ij} of hypothesis $\mathcal{H}_{ij}$ being true is calculated as,
   $$p_{ij} = \frac{\Pr(\beta_j \,|\, \mathcal{H}_{ij}) \Pr(\mathcal{H}_{ij})}{\sum_{i=1}^{M+1} \Pr(\beta_j | \mathcal{H}_{ij}) \Pr(\mathcal{H}_{ij})}, \tag{21}$$
   where $\widehat{y}_{M+1} = 0$.
   d) We calculate the average expected delay of the channel tap value $\widehat{y}_i$ as,
   $$\widehat{G}_i = \frac{1}{K} \sum_{j=1}^{\mathrm{N}_{\boldsymbol{\beta}}} j \cdot p_{ij}. \tag{22}$$
   e) If τ_j represents the delay between tap values $\widehat{y}_j$ and $\widehat{y}_{j-1}$ in the vector $\boldsymbol{y}$, then $\widehat{\tau}_j = \widehat{G}_j - \widehat{G}_{j-1}$. Assuming first nonzero tap in $\boldsymbol{y}$ has zero delay, an estimate of supp$(\boldsymbol{y})$ is completely determined by the estimates $\widehat{\tau}_2, \widehat{\tau}_3, \dots, \widehat{\tau}_M$.
3) Estimation of $\boldsymbol{x}$:
   f) We denote the sub-vector formed by $\beta_i, \beta_{i+1}, \dots, \beta_j$ as $\boldsymbol{\beta}(i:j)$ and the vector supp$(\boldsymbol{y}) \odot \boldsymbol{\beta}(j : j + \mathrm{N}_{\boldsymbol{y}} - 1)$ by $\boldsymbol{\gamma}_j$. Let $\Pr(\boldsymbol{\gamma}_j \mid \widehat{\boldsymbol{y}})$ denote the probability of observing $\boldsymbol{\gamma}_j$ from additive noise corruption of $\widehat{\boldsymbol{y}}$. Initialize set $C = \emptyset$.
   g) We repeatedly update set C as,
   $$C = C \cup \underset{j \in \{1,\dots,\mathrm{N}_{\boldsymbol{x}}\} \setminus C}{\arg\max} \Pr(\boldsymbol{\gamma}_j \mid \widehat{\boldsymbol{y}}), \tag{23}$$
   until the stopping criterion $|C| = K$ is reached. Then, the values in set C are the nonzero tap locations of $\boldsymbol{x}$.

Under the assumption of no noise, the algorithm would yield the true values of y_j which results in perfect reconstruction of supp$(\boldsymbol{y})$ and hence a perfect estimate of $\boldsymbol{x}$ as well. If we relax the assumption of $|S(\boldsymbol{y})| = \|\boldsymbol{y}\|_0$, then estimation of supp$(\boldsymbol{y})$ is not perfect. Rather, if one tap value is associated with two different locations i and j in $\boldsymbol{y}$, then the estimated index for that tap value would be some weighted average of i and j and would be associated with a nontrivial discrete probability distribution. It is not possible to address this problem in a one step estimation approach for supp$(\boldsymbol{y})$ and we have to resort to iterations based on belief propagation technique. One step estimation performance for noisy measurements is provided in Section VI-D.

B. Stable Projection

This is a joint estimation algorithm based on projecting the current estimate onto a convex domain at each iteration to obtain the next estimate. Let $\boldsymbol{Y}$ and $\boldsymbol{X}$ denote the Toeplitz matrices which act as convolution operators w.r.t. $\boldsymbol{y}$ and $\boldsymbol{x}$ defined as,

$$\boldsymbol{Y}\boldsymbol{x} = \boldsymbol{X}\boldsymbol{y} = \boldsymbol{y} \star \boldsymbol{x}. \tag{24}$$

This algorithm exploits the fact that elements of $\boldsymbol{y}$ and $\boldsymbol{x}$ are bounded and without loss of generality, we assume that $\|\boldsymbol{y}\|_\infty \le 1$ and $\|\boldsymbol{x}\|_\infty \le 1$.

The gradient of the convolved output with respect to the inputs is given by,

$$\nabla\left(\boldsymbol{y}\star\boldsymbol{x}\right)=\begin{bmatrix}\nabla_{\boldsymbol{y}}\left(\boldsymbol{y}\star\boldsymbol{x}\right)\\\nabla_{\boldsymbol{x}}\left(\boldsymbol{y}\star\boldsymbol{x}\right)\end{bmatrix}=\begin{bmatrix}\nabla_{\boldsymbol{y}}\left(\boldsymbol{X}\boldsymbol{y}\right)\\\nabla_{\boldsymbol{x}}\left(\boldsymbol{Y}\boldsymbol{x}\right)\end{bmatrix}=\begin{bmatrix}\boldsymbol{X}^{\mathrm{T}}\\\boldsymbol{Y}^{\mathrm{T}}\end{bmatrix}. \tag{25}$$

Let $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{x}}$ be intermediate estimates to the true values of $\boldsymbol{y}$ and $\boldsymbol{x}$ at any given step and $\tilde{\boldsymbol{Y}}$ and $\tilde{\boldsymbol{X}}$ be the corresponding operators for the convolution operation with $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{x}}$ respectively. Then, the error in the convolved output is $\boldsymbol{\beta}-\tilde{\boldsymbol{y}}\star\tilde{\boldsymbol{x}}$. To design an iterative update strategy we note that for small enough step size μ, this error is approximated by the first term of the Taylor series of $\boldsymbol{y}\star\boldsymbol{x}$ as,

$$\mu\left(\boldsymbol{\beta}-\tilde{\boldsymbol{y}}\star\tilde{\boldsymbol{x}}\right)\approx\nabla\left(\tilde{\boldsymbol{y}}\star\tilde{\boldsymbol{x}}\right)^{\mathrm{T}}\begin{bmatrix}\boldsymbol{\delta y}\\\boldsymbol{\delta x}\end{bmatrix}=\begin{bmatrix}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Y}}\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta y}\\\boldsymbol{\delta x}\end{bmatrix}. \tag{26}$$

Observing that,

$$2\left(\tilde{\boldsymbol{y}}\star\tilde{\boldsymbol{x}}\right)=\begin{bmatrix}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Y}}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{y}}\\\tilde{\boldsymbol{x}}\end{bmatrix}, \tag{27}$$

a simple update rule for the next iterate is obtained by adding together (26) and (27),

$$\mu\boldsymbol{\beta}+\left(2-\mu\right)\tilde{\boldsymbol{y}}\star\tilde{\boldsymbol{x}}\approx\begin{bmatrix}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Y}}\end{bmatrix}\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{x}\end{bmatrix}, \tag{28}$$

where $\boldsymbol{y}=\tilde{\boldsymbol{y}}+\boldsymbol{\delta y}$ and $\boldsymbol{x}=\tilde{\boldsymbol{x}}+\boldsymbol{\delta x}$.

The Stable Projection strategy is outlined below.

1) Initialize $\boldsymbol{y}$ and $\boldsymbol{x}$ randomly in their respective domains.
2) Calculate the matrices $\boldsymbol{Y}$ and $\boldsymbol{X}$.
3) Calculate the maximum singular value σ_{max} of the matrix $\begin{bmatrix}\boldsymbol{X}\boldsymbol{Y}\end{bmatrix}$. Set step size $\mu=\max\left\{1,\dfrac{1}{\sigma_{\text{max}}}\right\}$.
4) Solve the quadratic program,

$$\begin{aligned}&\underset{\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}}{\text{minimize}} && \left\|\begin{bmatrix}\boldsymbol{X}\boldsymbol{Y}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{y}}\\\tilde{\boldsymbol{x}}\end{bmatrix}-\left(\mu\boldsymbol{\beta}+\left(2-\mu\right)\boldsymbol{y}\star\boldsymbol{x}\right)\right\|_2\\&\text{subject to} && \left\|\begin{bmatrix}\tilde{\boldsymbol{y}}\\\tilde{\boldsymbol{x}}\end{bmatrix}\right\|_\infty\le 1.\end{aligned} \tag{P$_5$}$$

   to obtain the next estimate for $\boldsymbol{y}$ and $\boldsymbol{x}$ in $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{x}}$.
5) Go back to Step 2 and iterate till convergence is achieved.

Observe that we do not add an ℓ_1-norm penalty term to our objective function due to the discussion in Subsection IV-B.

VI. Numerical Results

A. Data Generation

We generated signals with dimensions given by $\mathrm{N}_{\boldsymbol{x}}=500$ and $\mathrm{N}_{\boldsymbol{y}}=100$. We define the normalized sparsity of a vector $\boldsymbol{x}$ as the quantity $\left\|\boldsymbol{x}\right\|_0/\mathrm{N}_{\boldsymbol{x}}$ and denote it by $\rho_{\boldsymbol{x}}$. We fixed $\rho_{\boldsymbol{y}}=0.05$ and varied $\rho_{\boldsymbol{x}}$ from 0.01 to 0.05 in steps of 0.01. For each value of $\rho_{\boldsymbol{x}}$ considered, we generated 100 random realizations of vector pairs $(\boldsymbol{x},\boldsymbol{y})$ which satisfy (17). We found that for $\rho_{\boldsymbol{x}}>0.05$ it was computationally prohibitive to randomly generate enough vector pairs $(\boldsymbol{x},\boldsymbol{y})$ which satisfy (17). For simulating estimation performance in the presence of noise, we generated 100 additive white Gaussian noise (AWGN) corrupted realizations for each random $(\boldsymbol{x},\boldsymbol{y})$ vector pair realization.

B. Comparison Metrics

We considered two comparison metrics which are described below. The first one tends to capture error energy and represents the conventional way to measure estimation performance. The second one is based on the ℓ_1-norm of the estimation error and tends to penalize support detection errors more that the former metric.

1) Normalized Root Mean Square Error: For a fixed realization of the vector pair $(\boldsymbol{x},\boldsymbol{y})$, the squared error between the true value of the concatenated vector $\left(\boldsymbol{x}^{\mathrm{T}},\boldsymbol{y}^{\mathrm{T}}\right)$ and the estimated value of the concatenated vector $\left(\hat{\boldsymbol{x}}^{\mathrm{T}},\hat{\boldsymbol{y}}^{\mathrm{T}}\right)$ is calculated. In the noiseless setting, this squared error value is averaged over all 100 realizations of the vector pair $(\boldsymbol{x},\boldsymbol{y})$ generated under the same value of $\rho_{\boldsymbol{x}}$ to obtain the mean squared error. In the noisy setting the averaging is done over both, the AWGN realizations as well as the $(\boldsymbol{x},\boldsymbol{y})$ vector pair realizations. The square root of this mean squared error is divided by $\mathrm{N}_{\boldsymbol{y}}+\mathrm{N}_{\boldsymbol{x}}=600$ to obtain the normalized root mean squared error.

2) Normalized Mean Deviation Error: For a fixed realization of the vector pair $(\boldsymbol{x},\boldsymbol{y})$, the ℓ_1-norm of the error between the true value of the concatenated vector $\left(\boldsymbol{x}^{\mathrm{T}},\boldsymbol{y}^{\mathrm{T}}\right)$ and the estimated value of the concatenated vector $\left(\hat{\boldsymbol{x}}^{\mathrm{T}},\hat{\boldsymbol{y}}^{\mathrm{T}}\right)$ is calculated. In the noiseless setting, this ℓ_1-norm value is averaged over all 100 realizations of the vector pair $(\boldsymbol{x},\boldsymbol{y})$ generated under the same value of $\rho_{\boldsymbol{x}}$ to obtain the mean deviation. In the noisy setting the averaging is done over both, the AWGN realizations as well as the $(\boldsymbol{x},\boldsymbol{y})$ vector pair realizations. This mean deviation error is divided by $\mathrm{N}_{\boldsymbol{x}}+\mathrm{N}_{\boldsymbol{y}}=600$ to obtain the normalized mean deviation error.

It was observed in our experiments that both metrics tend to show similar results with the differences being usually more pronounced for the *Normalized Mean Deviation Error*. Thus, we have presented results for this metric alone.

C. Noiseless Estimation Performance

1) Performance Comparison: The Dense-$\boldsymbol{h}$ Block Coordinate Descent (DhBCD) algorithm from [13], the DMAP estimation algorithm from Subsection V-A and the SP algorithm from Subsection V-B were compared when measurement noise was absent. For the DhBCD algorithm we set the number of source signals to one. The results are presented in Fig. 2 with a logarithmic scale for the error axis.

It was observed that the DMAP algorithm perfectly recovers the vector pair $(\boldsymbol{y},\boldsymbol{x})$ in this setting. This is to be expected as the algorithm was designed to arrive at the unique solution if one exists via the sufficiency conditions in Theorem 2. It is also observed that the other two algorithms perform much worse than DMAP and among them SP tends to perform better than DhBCD. This can be explained in part by the fact that both simulation based and real life signals tend to be bounded which is not exploited correctly by DhBCD as it relies on energy renormalization alone. This was verified during simulations with the observation that estimated channel taps from the DhBCD algorithm would often assume very large values leading to instability of the algorithm. The SP ensures algorithmic stability by limiting the step size at each iteration

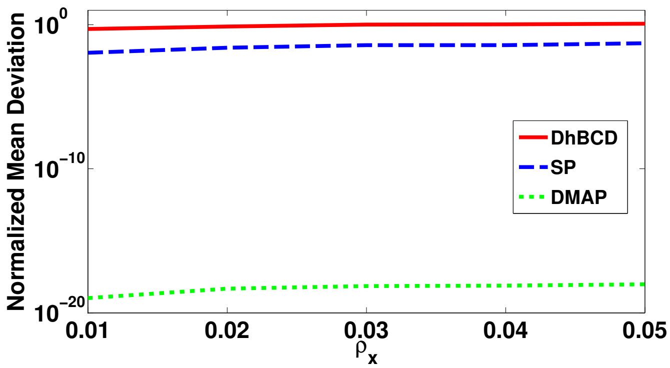

Fig. 2. Comparison of performance of the DhBCD, SP and DMAP estimation algorithms under noiseless setting.

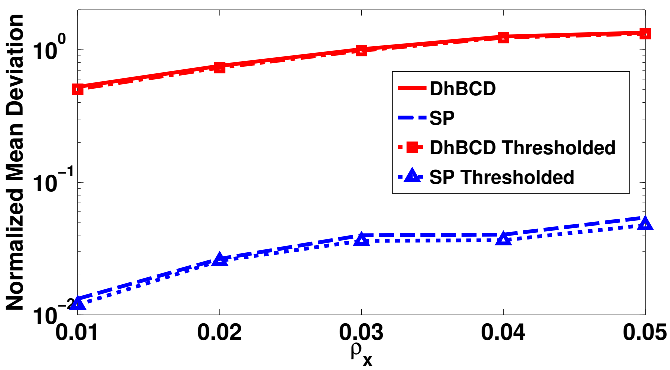

Fig. 3. Comparison of performance gain for DhBCD and SP algorithms on thresholding at final step.

as well as bounding the ℓ_∞-norm of the estimated vectors at the end of each iteration.

2) Thresholding at Final Step: To improve the performance of the DhBCD and SP algorithms, we introduce a thresholding operation as an end step for both algorithms. This step tries to exploit the binary nature of $\boldsymbol{x}$ ($\boldsymbol{x} \in \{0,1\}^{\mathrm{N}_{\boldsymbol{x}}}$) so an intuitive choice for the threshold is 0.5. The performance improvement turns out to be negligible at best, as can be seen in Fig. 3.

D. Noisy Estimation Performance

Because the DhBCD and the SP algorithms perform much much worse than the DMAP algorithm in the noiseless setting, we shall only consider the DMAP algorithm when measurement noise is present. We study the following two cases.

1) $\|\boldsymbol{y}\|_0$ known accurately at receiver: Fig. 4 shows the noisy estimation performance for the DMAP algorithm for different levels of normalized sparsity $\rho_{\boldsymbol{x}}$ under the assumption of $\|\boldsymbol{y}\|_0$ being known accurately at the receiver. The estimation performance degrades with increasing noise power as well as with increasing $\rho_{\boldsymbol{x}}$. Indeed, increasing $\rho_{\boldsymbol{x}}$ means more taps need to be estimated and hence there is a degradation in overall estimation performance.

2) $\|\boldsymbol{y}\|_0$ estimated at the receiver: Fig. 5 shows the noisy estimation performance of the DMAP algorithm under both metrics for different levels of normalized sparsity $\rho_{\boldsymbol{x}}$ when

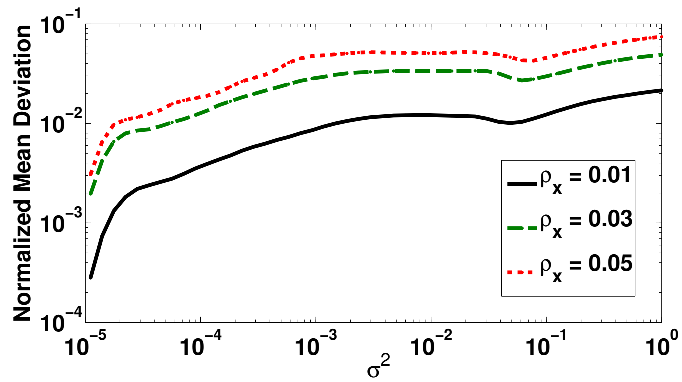

Fig. 4. Performance of DMAP estimation algorithm under noisy setting with prior knowledge of $\|\boldsymbol{y}\|_0$.

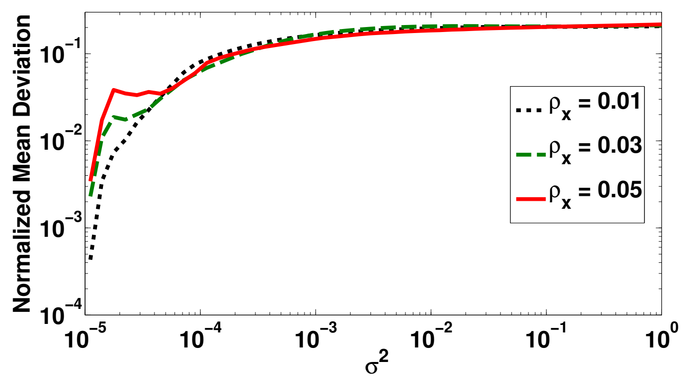

Fig. 5. Performance of DMAP estimation algorithm under noisy setting without prior knowledge of $\|\boldsymbol{y}\|_0$.

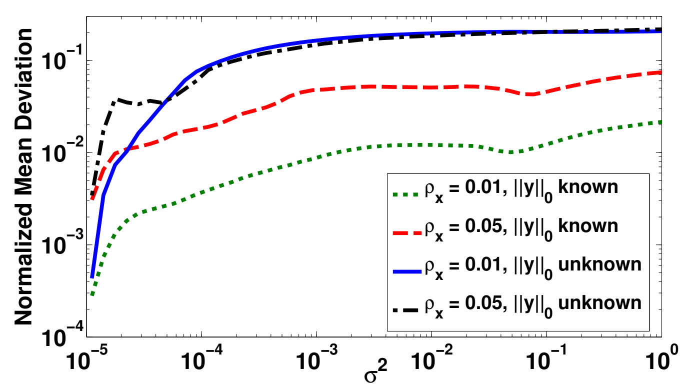

Fig. 6. Effect of prior knowledge of $\|\boldsymbol{y}\|_0$.

prior knowledge of $\|\boldsymbol{y}\|_0$ is not available at the receiver. Beyond small levels of noise power, estimation performance is independent of $\rho_{\boldsymbol{x}}$. This suggests that support detection of $\boldsymbol{y}$ and sparsity estimation of $\boldsymbol{x}$ is the major performance bottleneck. Fig. 6 directly compares estimation performance when $\|\boldsymbol{x}\|_0$ is known versus when it is unknown for two values of $\rho_{\boldsymbol{x}}$. This clearly shows that detecting support of $\boldsymbol{y}$ forms the major performance bottleneck when $\|\boldsymbol{y}\|_0$ is unknown at the receiver.

VII. Conclusions

In this paper we investigated some regularizers for the blind deconvolution problem which arise from physical considerations in underwater acoustic communication channels and proposed some conditions on unique recoverability under these settings. Using an elegant comparison with the low rank matrix completion problem, we argued that imposing sparsity priors on both the input signals is expected to lead to poorer recovery performance, a fact that was discovered in an earlier work but for which no explanation was provided. In this process we also establish the importance of a matrix completion viewpoint for the blind deconvolution problem, an approach not explicitly considered before. We provide results for unique recoverability for the blind deconvolution problem under both the "well separated" and "not well separated" scenarios. To the best of our knowledge, results for the "not well separated" setting have not been proposed before in this context. Two channel estimation strategies have been proposed for the second hop estimation problem apart from the popular alternating minimization and the convex relaxation techniques. We also provide numerical results for comparing estimation performance of the proposed algorithms and demonstrate the failure of assuming sparsity priors on both inputs simultaneously.

An important direction of our future endeavours would be to weaken the strict non-overlapping constraint and extend the ideas developed herein to other ill-posed inverse problems which conform to a bilinear structure (e.g. blind circular deconvolution, matrix factorization, etc.).

References

[1] E. M. Sozer, M. Stojanovic, and J. G. Proakis, "Underwater Acoustic Networks," *IEEE J. Ocean. Eng.*, vol. 25, no. 1, pp. 72–83, 2000.

[2] Y.-W. Hong, W.-J. Huang, F.-H. Chiu, and C.-C. J. Kuo, "Cooperative Communications in Resource-Constrained Wireless Networks," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 47–57, May 2007.

[3] M. Stojanovic, "On the Relationship Between Capacity and Distance in an Underwater Acoustic Communication Channel," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 11, no. 4, pp. 34–43, Oct. 2007.

[4] M. Vajapeyam, S. Vedantam, U. Mitra, J. C. Preisig, and M. Stojanovic, "Distributed Space–Time Cooperative Schemes for Underwater Acoustic Communications," *IEEE J. Ocean. Eng.*, vol. 33, no. 4, pp. 489–501, 2008.

[5] N. Richard and U. Mitra, "Sparse Channel Estimation for Cooperative Underwater Communications: A Structured Multichannel Approach," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008.*, 2008, pp. 5300–5303.

[6] M. Burger and O. Scherzer, "Regularization Methods for Blind Deconvolution and Blind Source Separation Problems," *Math. Control Signals Systems*, vol. 14, no. 4, pp. 358–383, 2001.

[7] D. Kundur and D. Hatzinakos, "Blind Image Deconvolution," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 43–64, May 1996.

[8] C. R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind Equalization Using the Constant Modulus Criterion: A Review," *Proc. IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct. 1998.

[9] M. Elad, M. A. T. Figueiredo, and Y. Ma, "On the Role of Sparse and Redundant Representations in Image Processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, Jun. 2010.

[10] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse Channel Estimation for Multicarrier Underwater Acoustic Communication: From Subspace Methods to Compressed Sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708–1721, Mar. 2010.

[11] C. Hegde and R. G. Baraniuk, "Sampling and Recovery of Pulse Streams," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1505–1517, 2011.

[12] K. Herrity, R. Raich, and A. O. Hero, III, "Blind Reconstruction of Sparse Images with Unknown Point Spread Function," C. A. Bouman, E. L. Miller, and I. Pollak, Eds., vol. 6814, no. 1. SPIE, 2008, p. 68140K.

[13] D. Barchiesi and M. D. Plumbley, "Dictionary Learning of Convolved Signals," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 5812–5815.

[14] M. S. Asif, W. Mantzel, and J. K. Romberg, "Random Channel Coding and Blind Deconvolution," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, Oct. 2009, pp. 1021–1025.

[15] D. L. Donoho, "Compressed Sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[16] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey on Sensor Networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[17] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving Single Channel, Full Duplex Wireless Communication," in *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '10. ACM, 2010, pp. 1–12.

[18] E. J. Candès and T. C. Tao, "The Power of Convex Relaxation: Near-Optimal Matrix Completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

Appendix A
Proof of Theorem 2

We shall employ contradiction. Let $K = K^*$ and $(\tilde{\boldsymbol{x}}(z)\,, \tilde{\boldsymbol{y}}(z))$ be a feasible point for Problem (P_4). Suppose $S(\tilde{\boldsymbol{y}}) = \{\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_r\}$ for some $1 \leq r \leq K^*$ with $\|\tilde{\boldsymbol{x}}\|_0 = K^*$. We have,

$$\tilde{\boldsymbol{y}}(z) = \sum_{j=1}^{\mathrm{N}_{\boldsymbol{y}}} (\tilde{\boldsymbol{y}})_j \, z^{j-1} = \sum_{j=1}^{r} \tilde{y}_j \boldsymbol{p}_j(z) \tag{29}$$

for some polynomials $\boldsymbol{p}_j(z) \in \mathbb{F}_2[z]$ where $\mathbb{F}_2$ is the binary field. We put $\tilde{\boldsymbol{x}}(z)\,\boldsymbol{p}_j(z) = \boldsymbol{q}_j(z)$. Then independent of the feasible point $(\tilde{\boldsymbol{x}}(z)\,, \tilde{\boldsymbol{y}}(z))$, the unique representation of $\boldsymbol{\beta}(z)$ in terms of polynomials in $\mathbb{F}_2[z]$ is,

$$\boldsymbol{\beta}(z) = \tilde{\boldsymbol{x}}(z)\,\tilde{\boldsymbol{y}}(z) = \sum_{j=1}^{r} \tilde{y}_j \tilde{\boldsymbol{x}}(z)\,\boldsymbol{p}_j(z) = \sum_{j=1}^{r} \tilde{y}_j \boldsymbol{q}_j(z) \tag{30}$$

The uniqueness of representation in (30) is due to $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ satisfying the non-overlapping constraint w.r.t. $\boldsymbol{\beta}$. Let $\boldsymbol{q}(z) = \gcd(\boldsymbol{q}_1(z)\,, \boldsymbol{q}_2(z)\,, \ldots, \boldsymbol{q}_r(z))$. It is clear that $\tilde{\boldsymbol{x}}(z)\,\boldsymbol{r}(z) = \boldsymbol{q}(z)$ for some polynomial $\boldsymbol{r}(z)$. Because $\boldsymbol{q}(z)$ divides $\boldsymbol{\beta}(z)$, let $\boldsymbol{s}(z)\,\boldsymbol{q}(z) = \boldsymbol{\beta}(z)$. Then $(\boldsymbol{q}(z)\,, \boldsymbol{s}(z))$ is a feasible point of Problem (P_4). If degree of $\boldsymbol{r}(z)$ equals zero then we have a unique solution to Problem (P_4). If degree of $\boldsymbol{r}(z)$ is greater than zero then $\|\boldsymbol{q}\|_0 > \|\tilde{\boldsymbol{x}}\|_0 = K^*$. This would imply that Problem (P_4) is feasible for some $K > K^*$ which is a contradiction. Hence degree of $\boldsymbol{r}(z)$ equals zero and we have a unique solution to Problem (P_4).