Compressive Topology Identification of Interconnected Dynamic Systems via Clustered Orthogonal Matching Pursuit

Borhan M. Sanandaji, Tyrone L. Vincent, and Michael B. Wakin

***Abstract*— In this paper, we consider topology identification of large-scale interconnected dynamical systems. The system topology under study has the structure of a directed graph. Each edge of the directed network graph represents a Finite Impulse Response (FIR) filter with a possible transport delay. Each node is a summer, whose inputs are the signals from the incoming edges, while the output of the summer is sent to outgoing edges. Edges of the graph can be of different unknown orders and delays. Both the graph topology and the FIR filters and delays that make up the edges are unknown. We aim to do the topology identification from the *smallest possible* number of node observations when there is limited data available and for this reason, we call this problem Compressive Topology Identification (CTI).**

Inspired by Compressive Sensing (CS) which is a recent paradigm in signal processing for sparse signal recovery, we show that in cases where network interconnections are suitably *sparse* (i.e., the network contains sufficiently few links), it is possible to perfectly identify the network topology along with the filter orders and delays from small numbers of node observations, even though this leaves an apparently ill-conditioned identification problem.

The main technical novelty of our approach is in casting the identification problem as the recovery of a *clustered-sparse* signal $\boldsymbol{z} \in \mathbb{R}^N$ from the measurements $\boldsymbol{b} = A\boldsymbol{z} \in \mathbb{R}^M$ with $M < N$, where the measurement matrix A is a block-concatenation of Toeplitz matrices. To this end, we introduce a greedy algorithm called Clustered Orthogonal Matching Pursuit (COMP) that tackles the problem of recovering clustered-sparse signals from few measurements. In a clustered-sparse model, in contrast to block-sparse models, there is no prior knowledge of the locations or the sizes of the clusters. We discuss the COMP algorithm and support our discussions with simulations.

I. Introduction

Systems that have a large number of inputs and outputs present a particular challenge for system identification. An important choice for the system identification practitioner is that of proper model structure — if the correct model structure can be chosen, then the parameters can be identified from data consistently. However, if the incorrect model structure is chosen, the estimates can have excessive bias or variance. In this paper, we focus on systems with a large number of observable variables, where the relationships between these variables can be described by a signal flow graph with nodes of low maximum degree. Examples of such systems come when modeling building thermal dynamics [1], [2], biological systems [3], and economics [4]. The model structure is specified by the connections that occur in the graph, but with a large number of signals, there are a large number of potential connections, only a small number of which will be present. Thus, in order to proceed with system identification, there is a need for effective "topological identification" procedures [5]–[8] which, given measurements of the nodes of an interconnected dynamical system over a finite time interval, determine the correct interconnection topology.

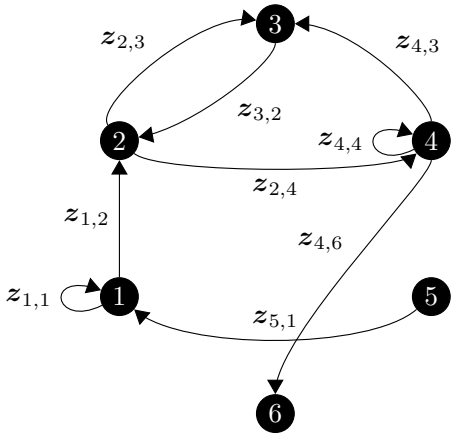


Fig. 1. Network model of 6 interconnected nodes. Each edge of the directed network graph ($\boldsymbol{z}_{j,i}$) represents an FIR filter with possible transport delay. Each node is a summer, whose inputs are the signals from the incoming edges, while the output of the summer is sent to outgoing edges. Edges of the graph can be of different unknown orders and delays.

One solution for the topology identification problem comes from Materassi and Innocenti [7] in the case that the interconnection graph has a tree structure and enough data is available to form reliable estimates of cross-power spectral densities. In this paper, we consider a more general setting, allowing arbitrary interconnections (including trees, loops, and self-loops) between nodes in the network, but we assume that the interconnection graph is sparse in the sense that each node has a relatively low in-degree. Fig. 1 shows one such example network. However, we also assume that each node contains a measurable exogenous input signal, while in [7] the input is an unmeasured random process. We also focus on the case of a small number of measurements, indeed, so small that the parameter estimation problem would not be invertible if connections to all nodes were postulated.

In this paper, the identification problem is formulated as the recovery of a *clustered-sparse* signal $\boldsymbol{z} \in \mathbb{R}^N$ from the measurements $\boldsymbol{b} = A\boldsymbol{z} \in \mathbb{R}^M$, where the measurement matrix A is a block-concatenation of Toeplitz matrices. A signal $\boldsymbol{z} \in \mathbb{R}^N$ is clustered-sparse when it only has K non-

All authors are with the Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO 80401, USA. {bmolazem, tvincent, mwakin}@mines.edu. This work was partially supported by AFOSR Grant FA9550-09-1-0465, NSF Grant CCF-0830320 and NSF Grant CNS-0931748.

zero entries ($K \ll N$) and these non-zero elements are contained in a number of contiguous clusters. As compared to a block-sparse structure [9]–[11] in which the non-zero entries appear in blocks of same size, in a clustered-sparse structure, the clusters can be of different sizes and can appear in arbitrary locations. We present an algorithm called Clustered Orthogonal Matching Pursuit (COMP) that can estimate the proper support and values for $\boldsymbol{z}$, even when the number of measurements is less than N. This algorithm is inspired by the field of Compressive Sensing (CS) [12]–[15], which is concerned with the recovery of sparse signals from limited numbers of measurements.

The connection between CS and network topology identification was first noted in [6], which provided a greedy algorithm in the case that the nodal inputs are unmeasured random sequences, and the data set is not necessarily assumed to be limited. We have previously discussed the limited-data network identification problem in [8], which also contained an analysis of the recovery algorithm Block Orthogonal Matching Pursuit (BOMP) [9]–[11]. The present work documents a new algorithm with improved recovery performance. A recent work [16] also considers a network model similar to [8] but derives conditions under which the Group Lasso (gLasso) procedure consistently estimates sparse network structure.

II. Notation

In this section, we establish our notation. Denote the set of real numbers by $\mathbb{R}$. All signals are discrete time, defined over a finite non-negative time interval, and represented equivalently as either the function of integer time $a(t)$ or grouped into vector form using the boldface $\boldsymbol{a}$. Given a finite sequence $\boldsymbol{a}$, define the mapping to a Toeplitz matrix for $i \geq j$ as

$$\mathcal{T}(\boldsymbol{a})_i^j := \begin{bmatrix} a(0) & 0 & \dots & 0 \\ a(1) & \ddots & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a(i-j) & & \ddots & a(0) \\ \vdots & \ddots & & a(1) \\ & & \ddots & \vdots \\ a(i-1) & & \dots & a(i-j) \end{bmatrix},$$

where zeros are applied if the index goes outside the defined range of $\boldsymbol{a}$. A matrix with the same entries along all its diagonals is called Toeplitz.

III. Problem Setup

A. Network Model

Fig. 1 shows an example of an interconnected network. Any types of interconnections between nodes such as trees, loops, and self-loops are allowed in the network topology. As compared to the previous paper [8], the edges of the network graph include unknown delays and unknown FIR filters of different order.

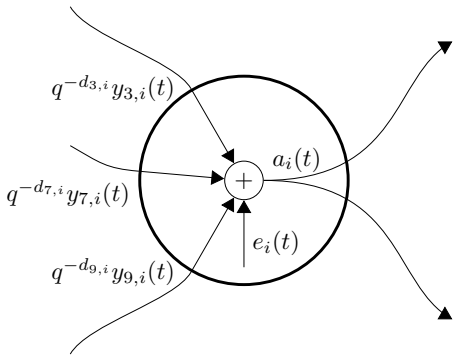


Fig. 2. Single node model. Each node is a summer, whose inputs are the signals from the incoming edges, while the output of the summer is sent to outgoing edges. In this illustration, node i sums the signals from nodes 3, 7, and 9 plus a node specific input term e_i.

Given an interconnected network of P nodes, let the time series $a_i(t)$, $t = 1, 2, \dots, M$, denote the output of node i. An edge in the graph, labeled $\boldsymbol{z}_{j,i}$, represents a dynamic system that filters the output of node j (that is, $a_j(t)$) and passes the result (which we call $y_{j,i}(t)$) through an unknown transport delay $q^{-d_{j,i}}$. The overall result, which is $q^{-d_{j,i}}y_{j,i}(t)$, will be an input to node i.

Let $\mathcal{N}_i$ denote the set of nodes whose outputs are processed and fed to node i. As shown in Fig. 2, we assume that each node i simply sums the signals that terminate upon it $\left\{q^{-d_{j,i}}y_{j,i}\right\}_{j\in\mathcal{N}_i}$ and adds a node-specific input term $e_i(t)$ that may or may not be known. In other words, the output of node i for $t = 1, 2, \cdots, M$, is given by

$$a_i(t) = \sum_{j\in\mathcal{N}_i} q^{-d_{j,i}}y_{j,i}(t) + e_i(t). \tag{1}$$

The filter in each edge $\boldsymbol{z}_{j,i}$ is modeled as a causal FIR filter with impulse response $\boldsymbol{x}_{j,i} \in \mathbb{R}^{n_{j,i}}$, so that $\boldsymbol{y}_{j,i} = \boldsymbol{a}_j * \boldsymbol{x}_{j,i}$. Assuming that $a_j(t) = 0$ for $t \leq 0$, the convolution can be written as

$$y_{j,i}(t) = \sum_{s=0}^{n_{j,i}-1} x_{j,i}(s)a_j(t-s), \tag{2}$$

for $t = 1, 2, \cdots, M$. Note that we have assumed no feedthrough term, so $x_{j,i}(0) = 0$ and therefore, $y_{j,i}(1) = 0$.

We can further incorporate the delay term as part of the filter coefficients. Assuming a maximum length of m (including the delay term) for all of the links in the graph, we define $\boldsymbol{z}_{j,i} \in \mathbb{R}^m$ as

$$\boldsymbol{z}_{j,i} = [\underbrace{0 \quad \cdots \quad 0}_{d_{j,i}} \quad \boldsymbol{x}_{j,i}^T \quad 0 \quad \cdots \quad 0]^T \tag{3}$$

where $(n_{j,i} + d_{j,i}) \leq m, \forall j, i$. In (3), the first $d_{j,i}$ zeros represent the delay term $q^{-d_{j,i}}$. Combining (1) with (3), the output of each node $\boldsymbol{a}_i \in \mathbb{R}^M$ can be written as

$$\boldsymbol{a}_i = \sum_{j\in\mathcal{N}_i} A_j\boldsymbol{z}_{j,i} + \boldsymbol{e}_i, \qquad i = 1, 2, \cdots, P, \tag{4}$$

where $A_j = \mathcal{T}(\boldsymbol{a}_j)_M^m$ is an $M \times m$ Toeplitz matrix, $\boldsymbol{z}_{j,i} \in \mathbb{R}^m$ and $\boldsymbol{e}_i \in \mathbb{R}^M$. Note that in (4) the only knowledge is

on the maximum length m while the filter order $n_{j,i}$ and transport delay $d_{j,i}$ are unknown.

Setting $\boldsymbol{z}_{j,i} = \mathbf{0}$ for $j \notin \mathcal{N}_i$, (4) can be rewritten as

$$\boldsymbol{a}_i = \sum_{j=1}^{P} A_j \boldsymbol{z}_{j,i} + \boldsymbol{e}_i, \qquad i = 1, 2, \cdots, P. \tag{5}$$

Equation (5) can be expanded as

$$\boldsymbol{a}_i = \underbrace{\begin{bmatrix} A_1 & \cdots & A_j & \cdots & A_P \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \boldsymbol{z}_{1,i} \\ \vdots \\ \boldsymbol{z}_{j,i} \\ \vdots \\ \boldsymbol{z}_{P,i} \end{bmatrix}}_{\boldsymbol{z}^i} + \quad \boldsymbol{e}_i, \tag{6}$$

or equivalently as

$$\boldsymbol{a}_i = A\boldsymbol{z}^i + \boldsymbol{e}_i, \qquad i = 1, 2, \cdots, P, \tag{7}$$

where $\boldsymbol{a}_i \in \mathbb{R}^M$, $\boldsymbol{z}^i \in \mathbb{R}^{Pm}$, and $A \in \mathbb{R}^{M \times Pm}$ is a matrix formed by the concatenation of P Toeplitz matrices.

B. Topology Identification

Given an interconnected graph of P nodes, the topology identification problem can be viewed as recovering the set of interconnected links ($\mathcal{N}_i$) for each node i in the graph. The links include unknown FIR filters of different orders followed by unknown different transport delays. The only knowledge is on the total number of nodes in the network (P) and on the maximum length (m) of each link. By the formulation given in the previous section, the topology identification problem is equivalent to recovering $\{\boldsymbol{z}^i\}_{i=1}^P$ given observations. We assume the inputs are decomposed into $\boldsymbol{e}_i = \widehat{\boldsymbol{e}}_i + \widetilde{\boldsymbol{e}}_i$, where $\widehat{\boldsymbol{e}}_i$ is known and $\widetilde{\boldsymbol{e}}_i$ is unknown. We also assume that $\widetilde{\boldsymbol{e}}_i$ is an independent and identically-distributed (i.i.d.) Gaussian sequence. Therefore, the measurements available to us consist of all outputs $\{\boldsymbol{a}_i\}_{i=1}^P$ and the known components $\{\widehat{\boldsymbol{e}}_i\}_{i=1}^P$ of the inputs. Our goal is to estimate the $\widehat{\boldsymbol{z}}_{j,i}$ that best match these observations data in an appropriate sense; we then determine that a link exists whenever $\|\widehat{\boldsymbol{z}}_{j,i}\|$ exceeds some threshold (we set the threshold to zero when striving for perfect recovery of all $\widehat{\boldsymbol{z}}_{j,i}$).

In order to solve this problem, we will utilize the following minimization problem:

$$\min_{\{\boldsymbol{z}^i\}_{i=1}^P} \quad \sum_{i=1}^{P} \|A\boldsymbol{z}^i - (\boldsymbol{a}_i - \widehat{\boldsymbol{e}}_i)\|_2^2. \tag{8}$$

Equivalently, the objective function in (8) can be minimized by solving

$$\min_{\boldsymbol{z}^i} \quad \|A\boldsymbol{z}^i - (\boldsymbol{a}_i - \widehat{\boldsymbol{e}}_i)\|_2^2 \tag{9}$$

separately for each node i in the network. We note that the same matrix A is used for recovery of all $\boldsymbol{z}^i$. For simplicity and without loss of generality, we will suppose henceforth that $\boldsymbol{a}_i - \widehat{\boldsymbol{e}}_i = \boldsymbol{b}$ and $\boldsymbol{z}^i = \boldsymbol{z}$ for each specific node and solve the optimization problem

$$\min_{\boldsymbol{z}} \quad \|A\boldsymbol{z} - \boldsymbol{b}\|_2^2, \tag{10}$$

where by letting $N = Pm$, we have $\boldsymbol{b} \in \mathbb{R}^M$, $\boldsymbol{z} \in \mathbb{R}^N$, and $A \in \mathbb{R}^{M \times N}$ is a matrix consisting of a concatenation of P Toeplitz matrices.

IV. Compressive Topology Identification

The optimization problem (10) has a unique solution if we collect $M \geq N$ measurements and if the matrix A is full rank. Then from standard linear algebra, we know that exact recovery of $\boldsymbol{z}$ when $\widetilde{\boldsymbol{e}}_i = \mathbf{0}$ is possible from

$$\boldsymbol{z}^\star = A^\dagger \boldsymbol{b}, \tag{11}$$

where $A^\dagger = (A^T A)^{-1} A^T$ is the Moore-Penrose pseudoinverse of A. However, as N linearly depends on the number of nodes in the network, for large-scale interconnected networks, the number of measurements required for exact topology identification scales linearly with the number of nodes in the network. This requires a large data set for topology identification.

In this paper, we aim to do exact topology identification from the smallest possible number of node observations (M). This is inspired by the field of CS and for this reason, we call it Compressive Topology Identification (CTI). We show that under the assumption of *sparsity* of the node interconnections (that is, assuming only a few nodes contribute to the output of each node), there will be a distinct structure to the solutions $\boldsymbol{z}$ that we are searching for. In particular, a typical vector $\boldsymbol{z}$ under our model assumptions will have very few non-zero entries, and these non-zero entries will be clustered in few locations. The number of clusters corresponds to the number of links that contribute to the output of the current node of interest (i.e., the cardinality of the set $\mathcal{N}_i$ for node i), while the size of each cluster depends on the order of the corresponding FIR filter connected to node i. From tools in CS and for signals obeying such structures, it is known that we can recover $\boldsymbol{z} \in \mathbb{R}^N$ exactly from measurements $\boldsymbol{b} = A\boldsymbol{z} \in \mathbb{R}^M$ even when $M \ll N$.

In the previous work [8], all links were assumed to be of the same order without considering any unknown link delays in the interconnections. Therefore, the non-zero entries of $\boldsymbol{z}$ appeared in locations with same length. In the field of CS, such a structure is known as *block-sparsity* [9]–[11]. However, in this paper the links are allowed to have different orders and different unknown link delays. Thus, the vector $\boldsymbol{z}$ will no longer have a block-sparse structure. In fact, the vector $\boldsymbol{z}$ has a *clustered-sparse* structure [17]. In the following sections, we will discuss block-sparse and clustered-sparse structures more formally. For simplicity, we assume that $\widetilde{\boldsymbol{e}}_i = \mathbf{0}$, but it is possible to extend our arguments from exact recovery in noise-free settings to robust recovery in noisy settings.

V. CS Background

First introduced by Candès, Romberg and Tao [12]–[14], and Donoho [15], CS is a paradigm which enables the recovery of an unknown signal from its underdetermined set of measurements under the assumption of sparsity of the signal and under certain conditions on the measurement

Algorithm 1 The BOMP – block-sparse recovery

Require: matrix A, measurements $\boldsymbol{b}$, block size n, stopping criteria

Ensure: $\boldsymbol{r}^0 = \boldsymbol{b}$, $\boldsymbol{z}^0 = \boldsymbol{0}$, $\Lambda^0 = \emptyset$, $l = 0$

repeat

1. match: $\boldsymbol{h}_i = A_i^T \boldsymbol{r}^l, \quad i = 1, 2, \cdots, P$

2. identify support: $\lambda = \arg\max_i \|\boldsymbol{h}_i\|_2$

3. update the support: $\Lambda^{l+1} = \Lambda^l \cup \lambda$

4. update signal estimate: $\boldsymbol{z}^{l+1} = \arg\min_{\boldsymbol{s}:\text{supp}(\boldsymbol{s})\subseteq\Lambda^{l+1}} \|\boldsymbol{b} - A\boldsymbol{s}\|_2$, where $\text{supp}(\boldsymbol{s})$ indicates the blocks on which $\boldsymbol{s}$ may be non-zero

5. update residual estimate: $\boldsymbol{r}^{l+1} = \boldsymbol{b} - A\boldsymbol{z}^{l+1}$

6. increase index l **by** 1

until stopping criteria true

output: $\widehat{\boldsymbol{z}} = \boldsymbol{z}^l = \arg\min_{\boldsymbol{s}:\text{supp}(\boldsymbol{s})\subseteq\Lambda^l} \|\boldsymbol{b} - A\boldsymbol{s}\|_2$

matrix A. The CS recovery problem can be viewed as recovery of a K-sparse signal $\boldsymbol{z} \in \mathbb{R}^N$ from its observations $\boldsymbol{b} = A\boldsymbol{z} \in \mathbb{R}^M$ where $A \in \mathbb{R}^{M\times N}$ is the measurement matrix with $M < N$ (in many cases $M \ll N$). A K-sparse signal $\boldsymbol{z} \in \mathbb{R}^N$ is a signal of length N with K non-zero entries where $K < N$. Since the null space of A is non-trivial, there are infinitely many candidate solutions to the equation $\boldsymbol{b} = A\boldsymbol{z}$; however, CS recovery algorithms exploit the fact that, under certain conditions on A, only one candidate solution is suitably sparse. The Restricted Isometry Property (RIP) [12], [18], the Exact Recovery Condition (ERC) [19], and mutual coherence [20], [21] are among the most important conditions on A that have been proposed in the CS literature.

There exist several CS recovery algorithms that exploit the sparsity of the signal to be recovered. In general, these recovery algorithms can be classified into two main types: 1) greedy algorithms such as Orthogonal Matching Pursuit (OMP) [21] and Compressive Sampling Matching Pursuit (CoSaMP) [22], and 2) convex optimization algorithms such as Basis Pursuit (BP) [23].

There also exist several extensions of the standard CS recovery algorithms to account for additional structure in the sparse signal to be recovered [11], [24]. Among these, the BOMP algorithm [9]–[11] is designed to exploit block sparsity. In previous work [8], the BOMP algorithm was considered for the topology identification problem due to its ease of implementation and its flexibility in recovering block-sparse signals of different sparsity levels.

Definition 1 (Block-Sparse Signal [11]): Consider $\boldsymbol{z} \in \mathbb{R}^N$ as a concatenation of P vector-blocks $\boldsymbol{z}_i \in \mathbb{R}^n$ of the same length where $N = Pn$, i.e.,

$$\boldsymbol{z} = [\boldsymbol{z}_1^T \cdots \boldsymbol{z}_i^T \cdots \boldsymbol{z}_P^T]^T. \tag{12}$$

A signal $\boldsymbol{z} \in \mathbb{R}^N$ is called block K-sparse if it has $K < P$ non-zero blocks.

Also consider a matrix $A \in \mathbb{R}^{M\times N}$ as a concatenation of P matrix-blocks $A_i \in \mathbb{R}^{M\times n}$ as

$$A = [A_1 \cdots A_i \cdots A_P]. \tag{13}$$

To find a block sparse solution to the equation $\boldsymbol{b} = A\boldsymbol{z}$, the formal steps of BOMP are listed in Algorithm 1. The basic intuition behind BOMP is as follows. Due to the block sparsity of $\boldsymbol{z}$, the vector of observations $\boldsymbol{b}$ can be written as a succinct linear combination of the columns of A, with the selections of columns occurring in clusters due to the block structure of the sparsity pattern in $\boldsymbol{z}$. BOMP attempts to identify the participating indices by correlating the measurements $\boldsymbol{b}$ against the columns of A and comparing the correlation statistics among different blocks. Once a significant block has been identified, its influence is removed from the measurements $\boldsymbol{b}$ via an orthogonal projection, and the correlation statistics are recomputed for the remaining blocks. This process repeats until the residual equals zero.

Eldar et al. [11] proposed a sufficient condition for BOMP to recover any sufficiently concise block-sparse signal $\boldsymbol{z}$ from compressive measurements. This condition depends on the properties of A, specifically on two coherence metrics, block and sub-block coherence of matrix A. For a detailed description of these metrics see [11].

VI. Clustered Orthogonal Matching Pursuit

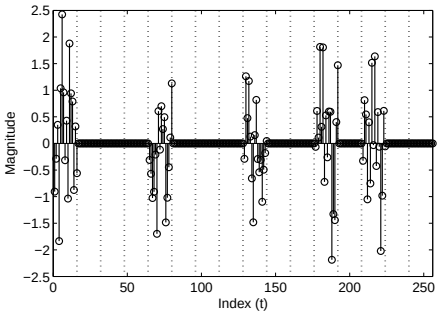
In a block-sparse structure as mentioned in Definition 1, the non-zero coefficients appear in blocks of the same length n. The BOMP algorithm is designed for recovering such block-sparse signals. As mentioned in Algorithm 1, the block size n is assumed to be known as one of the inputs to the algorithm. However, in this paper we are interested in recovering signals whose non-zero entries appear in clusters of different sizes. The only assumption is on the maximum cluster length. In the context of CS, such signals are called *clustered-sparse*.

Definition 2 (Clustered-Sparse Signal [17]): A signal $\boldsymbol{z} \in \mathbb{R}^N$ is called (K, C)-clustered sparse if it contains a total of K nonzero coefficients, spread among C disjoint clusters of arbitrary sizes and locations.

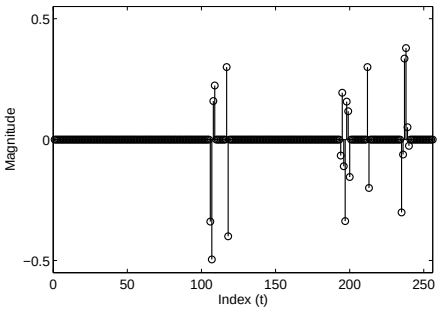
Fig. 3 shows a comparison between a block-sparse signal (Fig. 3(a)) and a clustered-sparse signal (Fig. 3(b)). As can be seen, in a block-sparse signal, the non-zero entries appear in blocks of the same size while in a clustered-sparse signal, they can appear in any locations and have any sizes.

In this section, we provide an algorithm that can be used for recovering clustered-sparse signals. The proposed method is an iterative greedy algorithm that is based on the well-known OMP algorithm. Its idea is intuitive and simple and also easy to implement.

The idea behind COMP is to exploit the knowledge that the non-zero entries of the signal appear in clusters, although of an arbitrary size and location. We modify the iterations of OMP in a way that exploits the clustered-sparsity pattern of the signal. The steps of the COMP algorithm are listed in Algorithm 2. The first two steps of COMP are the same as the first two steps of OMP. The outcome of step 2 at each iteration is a candidate for the true support. Let λ^l



(a) A block-sparse signal. Each block has the same size.



(b) A clustered-sparse signal. Clusters have different sizes.

Fig. 3. Clustered sparsity compared to block sparsity. Both signals have same length $N = 256$ with same cluster/block sparsity level of 5.

denote the support candidate at iteration l of the algorithm. If λ^0 is a valid candidate, i.e., $\lambda^0 \in T$ where T is the true support, then we can use our extra knowledge about the clustered-sparsity of the signal. In fact, we can use λ^0 as an indicator for the location of one of the clusters in the signal. Therefore, if we consider a window with proper length centered around λ^0, the extended support candidate is the window $\Lambda^1 = \widehat{\lambda}^0 = \{\lambda^0 - w + 1, \cdots, \lambda^0, \cdots, \lambda^0 + w - 1\}$ with window size $2w - 1$. Because the algorithm does not know where exactly λ^0 is located in the true cluster, the window length $2w - 1$ should be large enough such that the true cluster which by assumption is at most of size m, will be contained in the extended support candidate Λ^1. Apparently, the most conservative value for w is m. In the next step, the algorithm updates the signal estimate on the extended support candidate Λ^1. Having this estimate, the algorithm continues by updating the residual estimate. In the next iteration of COMP, the algorithm finds the column that is most correlated with the current residual (steps 1 and 2). The new support candidate λ^1 will not be one of the already chosen indices due to orthogonal projection properties, i.e., $\lambda^1 \notin \Lambda^1$. Again the algorithm considers a window of length $2w - 1$ centered around λ^1 and combines it with the previous support, i.e., $\Lambda^2 = \Lambda^1 \cup \{\lambda^1 - w + 1, \cdots, \lambda^1, \cdots, \lambda^1 + w - 1\}$. COMP continues until stopping criteria are met.

Note that Λ^f (the final support candidate found by COMP) should contain the true support, i.e., $T \subset \Lambda^f$ while the reverse $\Lambda^f \subset T$ is not a necessity. In addition, the cardinality of Λ^f should be smaller than the number of measurements M in order to have a unique least-squares solution while updating the signal estimate, i.e., $|\Lambda^f| < M$. The latter condition depends on the window length w that COMP considers at each iteration. As mentioned earlier, the most conservative choice for w is m, where m is the maximum cluster size. However, adding $2m-1$ elements to the support candidate at each iteration of COMP may make the uniqueness condition ($|\Lambda^l| < M$ for $l = 1, \cdots, f$) fail before the stopping criteria are met. Therefore, a simple iteration on the considered window length is added to the algorithm to improve the recovery performance of COMP. If the algorithm does not converge with $w = m$, the next set of iterations will be carried out starting with $w = m - 1$. This continues until the stopping criteria are met.

Algorithm 2 The COMP – clustered-sparse recovery

Require: matrix A, measurements $\boldsymbol{b}$, maximum cluster size m, stopping criteria

Ensure: $\boldsymbol{r}^0 = \boldsymbol{b}$, $\boldsymbol{z}^0 = \boldsymbol{0}$, $\Lambda^0 = \emptyset$, $l = 0$, $w = m$

repeat

1. **match:** $\boldsymbol{h}^l = A^T \boldsymbol{r}^l$
2. **identify support indicator:** $\lambda^l = \arg\max_j |h^l(j)|$
3. **extend support:** $\widehat{\lambda}^l = \{\lambda^l - w + 1, \cdots, \lambda^l, \cdots, \lambda^l + w - 1\}$
4. **update the support:** $\Lambda^{l+1} = \Lambda^l \cup \widehat{\lambda}^l$
5. **update signal estimate:** $\boldsymbol{z}^{l+1} = \arg\min_{\boldsymbol{s}:\text{supp}(\boldsymbol{s}) \subseteq \Lambda^{l+1}} \|\boldsymbol{b} - A\boldsymbol{s}\|_2$, where $\text{supp}(\boldsymbol{s})$ indicates the indices on which $\boldsymbol{s}$ may be non-zero
6. **update residual estimate:** $\boldsymbol{r}^{l+1} = \boldsymbol{b} - A\boldsymbol{z}^{l+1}$
7. **increase index** l **by** 1

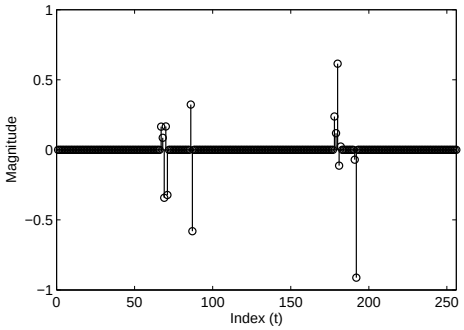
until stopping criteria true

output: $\widehat{\boldsymbol{z}} = \boldsymbol{z}^l = \arg\min_{\boldsymbol{s}:\text{supp}(\boldsymbol{s}) \subseteq \Lambda^l} \|\boldsymbol{b} - A\boldsymbol{s}\|_2$

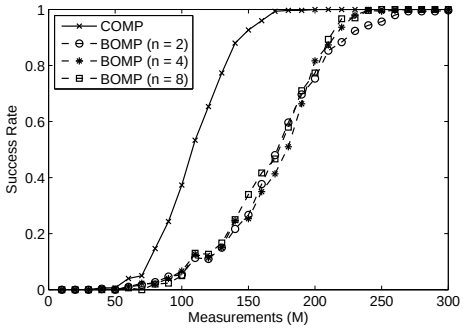
VII. Numerical Simulations

In this section, we test the proposed algorithm for identifying the topology of a dynamical network based on compressive observations, and with random but known inputs. As explained earlier, we cast the topology identification problem as recovery of a clustered-sparse signal whose few non-zero coefficients appear in clustered locations. The clusters are of arbitrary sizes. The only knowledge is on the maximum cluster size m. In order to compare the performance of the COMP algorithm, we also consider recovery using the BOMP algorithm. Moreover, in order to make a fair comparison between the two algorithms, we consider recovery using the BOMP algorithm with several block sizes n.

Fig. 4 shows a network of $P = 32$ nodes. Each edge of the directed network graph represents an FIR filter with possible transport delay. Each node is a summer, whose inputs are the signals from the incoming edges, while the output of the summer is sent to outgoing edges. Edges of the graph can be of different unknown orders and delays. Both the graph topology and the FIR filters and delays that make up the edges are unknown. The only knowledge is on the maximum



(a)



(b)

Fig. 5. Recovery performance corresponding to node 10. (a) Signal $\boldsymbol{z}$ corresponding to node 10 in the network graph of Fig. 4. The cluster-sparsity level corresponds to the in-degree of node 10. (b) Recovery performance comparison between COMP and BOMP with different block sizes n. An initial value of $w = m = 8$ is chosen in COMP. The algorithm iterates by reducing w until stopping criteria are met. For comparison, BOMP is tested with three different block sizes ($n = \{2, 4, 8\}$). Success rate is calculated over 300 realizations of the network for a given number of measurements.

cluster size $m = 8$. Therefore, for each node, the signal $\boldsymbol{z}$ has length $N = Pm = 256$.

Fig. 5 shows the recovery performance corresponding to node 10 of the network graph of Fig. 4. The corresponding signal $\boldsymbol{z}$ to be recovered is shown in Fig. 5(a). As can be seen the signal has a clustered-sparse structure with 4 clusters of different sizes. The number of clusters corresponds to the in-degree of node 10 while the size of each cluster depends on the order of the FIR filter of incoming edges. Fig. 5(b) shows the recovery performance comparison between COMP and BOMP with different block sizes n. An initial value of $w = m = 8$ is chosen in COMP. The algorithm iterates by reducing w until stopping criteria are met. For comparison, BOMP is tested with three different block sizes ($n = \{2, 4, 8\}$). The success rate is calculated over 300 realizations of the network for a given number of measurements. As can be seen, the COMP algorithm outperforms the BOMP algorithm. For this signal, the recovery performance of BOMP does not significantly improve by changing the block size n.

Fig. 6 shows the recovery rate comparison of nodes 10, 23, and 32 in the network of Fig. 4. The success rate is calculated over 300 realizations of the network for a given number of measurements. Node 10 has in-degree 4 and

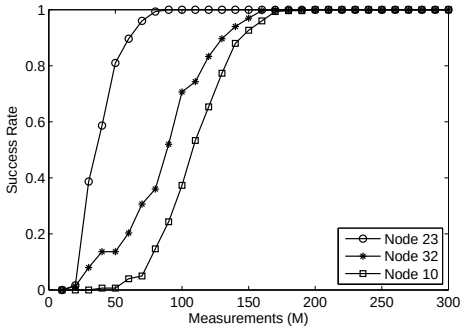


Fig. 6. Recovery rate comparison of nodes 10, 23, and 32 in the network of Fig. 4. An initial value of $w = m = 8$ is chosen in COMP. Nodes 23 and 32 have in-degree 2 and node 10 has in-degree 4. Success rate is calculated over 300 realizations of the network for a given number of measurements.

nodes 23 and 32 have in-degree 2. We observe how the probability of successful recovery changes for different nodes in the network based on the local sparsity and the type of interconnection. For example, node 10 which has in-degree 4 requires more measurements compared to nodes 23 and 32 which have in-degree 2. In addition to the local sparsity of each node, we observe that nodes of same in-degree have different recovery performance. For example, nodes 23 and 32 both have in-degree 2. However, node 32 is much easier to recover with the COMP algorithm, i.e., it requires a smaller number of measurements for perfect recovery as compared to node 23. This difference may be related to the type of incoming interconnections to each node. The incoming edges to node 32 have a tree structure while the incoming edges to node 23 include a loop.

VIII. Conclusions

We considered the exact topology identification of an interconnected dynamical system from measurements of the individual nodes. We have aimed at doing the topology identification from the *smallest possible* number of node observations when there is limited data available.

The system topology under study has the structure of a directed graph. Each edge of the directed network graph represents a FIR filter with possible transport delay. Each node is a summer, whose inputs are the signals from the incoming edges, while the output of the summer is sent to outgoing edges. Edges of the graph can be of different unknown orders and delays. Both the graph topology and the FIR filters and delays that make up the edges are unknown. We showed that exact topology identification is indeed possible form compressive node measurements under the assumption that the network contains nodes of low maximum degree. We cast the topology identification problem in the context of CS and the recovery of a clustered-sparse signal that encode the network topology. To this end, we proposed a greedy algorithm called COMP for recovery of such signals. We supported our proposed algorithm with illustrative simulations on a complicated network graph including trees, loops, and self-loops. We observed that the COMP algorithm outperforms the BOMP algorithm with several block sizes.

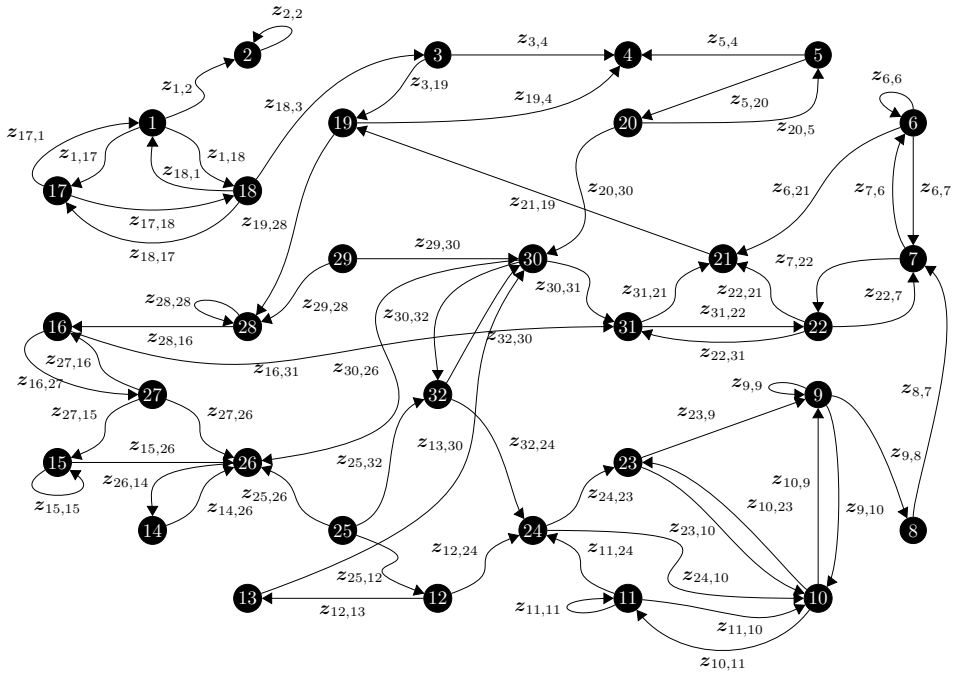


Fig. 4. A network of 32 interconnected nodes including trees, loops and self-loops. Each edge of the directed graph ($z_{j,i}$) represents an FIR filter.

We also observed that how the probability of successful recovery changes for different nodes in the network based on the local sparsity and the type of interconnection. Future work includes having a more formal understanding of COMP performance, dependence of recovery performance on *network coherence* [8], etc.

REFERENCES

[1] J. Pakanen and S. Karjalainen, "Estimating static heat flows in buildings for energy allocation systems," *Energy and Buildings*, vol. 38, no. 9, pp. 1044–1052, 2006.

[2] N. Chaturvedi and J. Braun, "An inverse gray-box model for transient building load prediction," *International Journal of Heating, Ventilating, Air-Conditioning and Refrigerating Research*, vol. 8, no. 1, pp. 73–100, 2002.

[3] E. Ravasz, A. Somera, D. Mongru, Z. Oltvai, and A. Barabási, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551–1555, 2002.

[4] R. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B*, vol. 11, no. 1, pp. 193–197, 1999.

[5] G. Innocenti and D. Materassi, "A modeling approach to multivariate analysis and clusterization theory," *Journal of Physics A: Mathematical and Theoretical*, vol. 41, p. 205101, 2008.

[6] D. Materassi, G. Innocenti, and L. Giarré, "Reduced complexity models in the identification of dynamical networks: links with sparsification problems," *Proceedings of the* 48*th IEEE Conference on Decision and Control*, pp. 4796–4801, 2009.

[7] D. Materassi and G. Innocenti, "Topological identification in networks of dynamical systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1860–1871, 2010.

[8] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, "Exact topology identification of large-scale interconnected dynamical systems from compressive observations," *Proceedings of the* 2011 *American Control Conference*, pp. 649 – 656, 2011.

[9] Y. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[10] ——, "Block-sparsity and sampling over a union of subspaces," *Proceedings of the* 16*th international conference on Digital Signal Processing*, pp. 1–8, 2009.

[11] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[12] E. Candès, "Compressive sampling," *Proceedings of the International Congress of Mathematicians*, vol. 3, pp. 1433–1452, 2006.

[13] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[14] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

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

[16] A. Bolstad, B. Van Veen, and R. Nowak, "Causal network inference via group sparse regularization," *IEEE Trans. Signal Processing*, vol. 59, no. 6, pp. 2628–2641, 2001.

[17] V. Cevher, P. Indyk, C. Hegde, and R. Baraniuk, "Recovery of clustered sparse signals from compressive measurements," *Proceedings of the International Conference on Sampling Theory and Applications*, 2009.

[18] E. Candès and T. Tao, "Decoding via linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[19] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 3, pp. 1030–1051, 2006.

[20] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[21] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[22] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[23] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.

[24] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.