A Review of Sufficient Conditions for Structure Identification in Interconnected Systems ⋆

Borhan M. Sanandaji * **Tyrone L. Vincent** *
Michael B. Wakin *

* *Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO 80401 USA (e-mail: {bmolazem, tvincent, mwakin}@mines.edu).*

Abstract: Structure identification of large-scale but sparse-flow interconnected dynamical systems from limited data has recently gained much attention in the control and signal processing communities. This paper reviews some of the recent results on Compressive Topology Identification (CTI) of such systems with a particular focus on sufficient recovery conditions. We list and discuss the key elements that influence the recovery performance of CTI, namely, the network topology, the number of measurements, and the input sequence. In regards to the last element, we analyze the recovery conditions with respect to an experiment design.

Keywords: Topology Identification; Interconnected Dynamical Systems; Compressive Sensing

1. INTRODUCTION

System identification is usually concerned with developing a model of a dynamical system from data for use in control design or for prediction. Large scale systems, which we define as a system with a large number of observable signals, present particular challenges for identification, particularly in the choice of model structure and the potentially large number of parameters that characterize this structure. One could attempt to meet this challenge by performing the identification with different model structures, and evaluating the prediction error using cross-validation or a prediction-error criterion that includes a penalty for model complexity, such as the AIC (Ljung [1987]). However, when the model structure is truly unknown, this could require an unacceptably large number of identification problems to be solved.

For the types of interconnected systems of interest here, we assume that each observable signal is linearly and causally dependent on a small subset of the remaining observable signals, plus an independent signal (i.e., input) that may or may not be measured. Examples of such systems include gene networks, where the expression of one gene may affect the expression of other genes in a dynamic way, coupled oscillators, large-scale dynamical systems such as power distribution networks or financial systems, and many others. These systems could have hundreds or thousands of observable variables, and identification of a model without simultaneously determining the model structure would be very difficult.

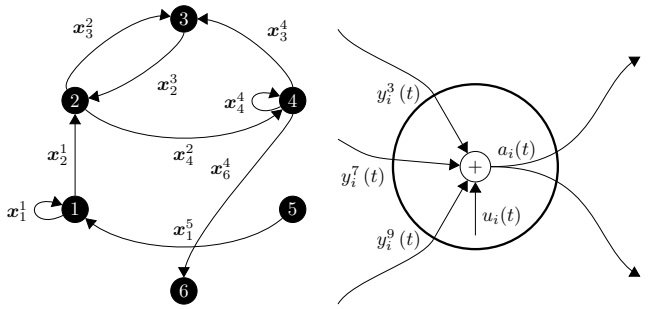
Recently, there has been some work suggesting that in some cases, the model structure and system parameters can be identified at the same time, using greedy algorithms or by solving convex optimization problems (e.g., Timme [2007], Napoletani and Sauer [2008], Materassi et al. [2009], Chiuso and Pillonetto [2010], Bolstad et al. [2011]). These methods generally use a fully parameterized model, but with modifications that ensure that most of the parameters will be zero when identified. Along with this empirical evidence, theoretical results are available concerning the guarantees of success for such algorithms to recover the true parameters in inverse problems with a large numbers of parameters but few non-zero parameters. However, the fact that the system is dynamic induces a specific structure in the problem that the performance guarantees do not specifically take into account. This paper reviews the existing results for guaranteed recovery and discusses the applicability of these results to the network identification problem.

2. INTERCONNECTED DYNAMICAL SYSTEMS

In order to provide an analysis, we assume that the interconnected dynamical system is a realization of a specific model structure. In what follows, 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 $\boldsymbol{a}$. The model structure can be defined using a *directed graph*, such as shown in Fig. 1(a). Each measured signal is represented by a node in the graph. Edges represent filters that operate on the node signal, and for this paper are assumed to be Finite Impulse Response (FIR) filters of different orders. Any types of interconnections between nodes—such as trees, loops, and self-loops—are allowed.

Given an interconnected network of P nodes, let the time series $a_i(t)$, $t = 1, 2, \ldots, M$, denote the output of node i. Each edge in the graph, labeled $\boldsymbol{x}_i^j \in \mathbb{R}^{n_i^j}$, filters the signal $a_j(t)$ at the tail of the edge, and passes the result,

⋆ This work was partially supported by AFOSR Grant FA9550-09-1-0465, NSF Grant CCF-0830320 and NSF Grant CNS-0931748.



(a) (b)

Fig. 1. (a) Network model of 6 interconnected nodes. Each edge of the directed graph ($\boldsymbol{x}_i^j$) represents an FIR filter, including a possible transport delay. Edges of the graph can be of different unknown orders and delays. (b) 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 the outgoing edges. In this illustration, node i sums the signals from nodes 3, 7, and 9 (i.e., $\mathcal{N}_i = \{3, 7, 9\}$) plus a node-specific input term $u_i(t)$.

$y_i^j(t) = \sum_{s=1}^{n_i^j} x_i^j(s)\, a_j(t-s)$, to the node at the head of the edge. The signal $x_i^j(t)$ is the impulse response of the filter for the edge from node j to node i. Note that we have assumed no feedthrough term.

Let $\mathcal{N}_i$ denote the set of nodes whose outputs are processed and fed to node i. As shown in Fig. 1(b), we assume that each node i simply sums the signals that terminate upon it and adds a node-specific input term $u_i(t)$ plus measurement noise $e_i(t)$. In other words, the output of node i for $t = 1, 2, \ldots, M$, is given by

$$a_i(t) = \sum_{j \in \mathcal{N}_i} y_i^j(t) + u_i(t) + e_i(t) \tag{1}$$

for $i = 1, 2, \ldots, P$. Using a more convenient matrix-vector notation, the output of each node $\boldsymbol{a}_i \in \mathbb{R}^M$ for $i = 1, 2, \cdots, P$, can be written as

$$\boldsymbol{a}_i = \sum_{j \in \mathcal{N}_i} A_j \boldsymbol{x}_i^j + \boldsymbol{u}_i + \boldsymbol{e}_i. \tag{2}$$

For uniformity, we take $m = \max_{i,j}(n_i^j)$ and represent $\boldsymbol{x}_i^j \in \mathbb{R}^m$ with trailing zeros as necessary. Observe that with this assumption, $\{\boldsymbol{x}_i^j\}_{i,j=1}^P$ may have different support patterns due to different possible unknown transport delays and different numbers of trailing zeros. In (2), A_j is an $M \times m$ Toeplitz matrix, $\boldsymbol{u}_i \in \mathbb{R}^M$, and $\boldsymbol{e}_i \in \mathbb{R}^M$. We use the notation $A_j = \mathcal{T}(\boldsymbol{a}_j)_M^m$ where $\mathcal{T}(\boldsymbol{a}_j)_M^m$ defines a mapping from a finite sequence $\boldsymbol{a}$ to a Toeplitz matrix as

$$\mathcal{T}(\boldsymbol{a})_M^m := \begin{bmatrix} a(0) & 0 & \ldots & 0 \\ a(1) & \ddots & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a(M-m) & & \ddots & a(0) \\ \vdots & \ddots & & a(1) \\ & & \ddots & \vdots \\ a(M-1) & & \ldots & a(M-m) \end{bmatrix},$$

for $M \geq m$ 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.

Setting $\boldsymbol{x}_i^j = \boldsymbol{0}$ for $\forall j \notin \mathcal{N}_i$, (2) can be rewritten as

$$\boldsymbol{a}_i = \sum_{j=1}^{P} A_j \boldsymbol{x}_i^j + \boldsymbol{u}_i + \boldsymbol{e}_i, \tag{3}$$

which can be expanded as

$$\boldsymbol{a}_i = \underbrace{[\, A_1 \ \cdots \ A_j \ \cdots \ A_P \,]}_{A} \underbrace{\begin{bmatrix} \boldsymbol{x}_i^1 \\ \vdots \\ \boldsymbol{x}_i^j \\ \vdots \\ \boldsymbol{x}_i^P \end{bmatrix}}_{\boldsymbol{x}_i} + \boldsymbol{u}_i + \boldsymbol{e}_i, \tag{4}$$

or equivalently as

$$\boldsymbol{a}_i = A\boldsymbol{x}_i + \boldsymbol{u}_i + \boldsymbol{e}_i, \tag{5}$$

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

3. NETWORK TOMOGRAPHY

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, including unknown transport delays. The assumed *a priori* knowledge is the total number of nodes P in the network and the maximum possible degree m of each link. By the formulation given in the previous section, the topology identification problem is equivalent to recovering $\{\boldsymbol{x}_i\}_{i=1}^P$ given node observations. In particular, the measurements available to us consist of all node inputs $\{\boldsymbol{u}_i\}_{i=1}^P$ and all node outputs $\{\boldsymbol{a}_i\}_{i=1}^P$. One possible approach to this problem would be to solve

$$\min_{\{\boldsymbol{x}_i\}_{i=1}^P} \quad \sum_{i=1}^{P} \| (\boldsymbol{a}_i - \boldsymbol{u}_i) - A\boldsymbol{x}_i \|_2^2. \tag{6}$$

The objective function in (6) can be minimized by solving

$$\min_{\boldsymbol{x}_i} \quad \| (\boldsymbol{a}_i - \boldsymbol{u}_i) - A\boldsymbol{x}_i \|_2^2 \tag{7}$$

separately for each node i in the network. Observe that the same matrix A is used for recovery of all $\boldsymbol{x}_i$. For simplicity and without loss of generality, we will suppose henceforth that $\boldsymbol{a}_i - \boldsymbol{u}_i = \boldsymbol{b}$ and $\boldsymbol{x}_i = \boldsymbol{x}$ for each given node and focus on the task of solving

$$\min_{\boldsymbol{x}} \quad \|\boldsymbol{b} - A\boldsymbol{x}\|_2^2, \tag{8}$$

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

The optimization problem (8) has a unique solution if we collect $M \geq N$ measurements and if the matrix A is full rank. From standard linear algebra, we would know that exact recovery of $\boldsymbol{x}$ when $\boldsymbol{e}_i = \boldsymbol{0}$ is possible from $\boldsymbol{x}^\star = A^\dagger \boldsymbol{b}$, where $A^\dagger = (A^T A)^{-1} A^T$ is the Moore-Penrose pseudoinverse of A. However, since N depends linearly on P, the condition $M \geq N$ requires large M for topology identification for large-scale interconnected networks (i.e.,

large P). On the other hand, if the system topology was known *a priori*, then only the parameters for the links that actually exist would need to be identified. The number of such parameters is independent of P.

In order to move towards a data requirement closer to the case of *a priori* knowledge of the topology, one can attempt to apply additional information or assumptions that can add additional constraints and reduce the effective degrees of freedom. In the case of interconnected dynamical systems, one assumption that is valid for a large number of applications of interest is low node in-degree (i.e., each node has only a small number of incoming edges). With this assumption, there will be a distinct structure to the solutions $\boldsymbol{x}$ that we are searching for. In particular, a typical vector $\boldsymbol{x}$ under our model assumptions will have very few non-zero entries, and these non-zero entries will be grouped in a few locations. The number of groups 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 group depends on the order and structure of the corresponding FIR filter connected to node i.

If all links in the network have the same order with no transport delay (i.e., $\forall i, j,\ n_i^j = m$), the non-zero entries of $\boldsymbol{x}$ appear in locations of the same size. In the literature, such structure is known as *block sparsity*.

Definition 1. (Eldar et al. [2010]). Let $\boldsymbol{x} \in \mathbb{R}^N$ be a concatenation of P vector-blocks $\boldsymbol{x}^j \in \mathbb{R}^m$ of the same length where $N = Pm$, i.e., $\boldsymbol{x} = [\boldsymbol{x}^{1^T} \cdots \boldsymbol{x}^{j^T} \cdots \boldsymbol{x}^{P^T}]^T$. The vector $\boldsymbol{x} \in \mathbb{R}^N$ is called *K-block sparse* if it has $K < P$ non-zero blocks.

On the other hand, if links are allowed to have different orders and different unknown transport delays, the vector $\boldsymbol{x}$ will no longer have a block-sparse structure. Instead, $\boldsymbol{x}$ has a *clustered-sparse* (Cevher et al. [2009]) structure. While network tomography via clustered-sparse recovery has also been addressed (Sanandaji et al. [2011a]), our focus in this paper is on block-sparse recovery.

Armed with this additional assumption, our estimation algorithm can be modified to a search for $\boldsymbol{x}$ such that $\|\boldsymbol{b} - A\boldsymbol{x}\|_2^2$ is small and $\boldsymbol{x}$ is sufficiently block/clustered sparse. This problem has been the subject of great interest in statistics and signal processing especially in the case where A has many fewer rows than columns, i.e., there are more parameters than measurements. This is the case of interest in this paper: when the number of measurements available for recovering $\boldsymbol{x}$ is smaller than the number of nodes, which indicates that both the parameters and the system structure have to be identified simultaneously. Inspired by Compressive Sensing (CS), these problems have been termed Compressive Topology Identification (CTI) (Sanandaji et al. [2011a]).

4. COMPRESSIVE TOPOLOGY IDENTIFICATION

In this section, we discuss two block-sparse recovery algorithms and their recovery guarantees. Then, we provide an analysis of the recovery conditions of these algorithms when they are applied to CTI. We start with an overview of CS theory and algorithms.

4.1 CS Background

First introduced by Candès [2006], Candès and Tao [2006], Candès et al. [2006], and Donoho [2006], CS is a paradigm that enables the recovery of an unknown signal from an underdetermined set of measurements under the assumption of sparsity of the signal and under certain conditions on the measurement matrix. The CS recovery problem can be viewed as recovery of a K-sparse signal $\boldsymbol{x} \in \mathbb{R}^N$ from observations $\boldsymbol{b} = A\boldsymbol{x} \in \mathbb{R}^M$ 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 $\boldsymbol{x} \in \mathbb{R}^N$ 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 $\boldsymbol{b} = A\boldsymbol{x}$; however, CS recovery algorithms, under certain conditions on A, are able to efficiently search for the candidate solution that is suitably sparse. The Restricted Isometry Property (RIP) (Candès and Tao [2005], Candès [2006]), the Exact Recovery Condition (ERC) (Tropp [2006]), and mutual coherence (Donoho and Huo [2001], Tropp [2004]) are among the most important conditions on A that have been proposed in the CS literature. See Candès and Wakin [2008] for an introduction to CS.

There exist several CS recovery algorithms that can be applied. In general, these recovery algorithms can be classified into two main types: 1) greedy algorithms such as Orthogonal Matching Pursuit (OMP) (Tropp [2004]) and Compressive Sampling Matching Pursuit (CoSaMP) (Needell and Tropp [2009]), and 2) convex optimization algorithms such as Basis Pursuit (BP) (Chen et al. [1998]). These algorithms simply seek a sparse solution to an underdetermined set of linear equations.

4.2 Algorithms for Block-Sparse Recovery

As mentioned earlier, if we assume that all the links in the network share the same order with no transport delay, the CTI problem boils down to the recovery of a block-sparse signal from few measurements.

Several extensions of the standard CS recovery algorithms have been proposed to account for the block sparsity of $\boldsymbol{x}$ (see, e.g., Eldar et al. [2010], Yuan and Lin [2006]). Similar to standard CS algorithms, the block-sparse recovery algorithms can be classified into two main groups: greedy algorithms and convex optimization algorithms.

Block Orthogonal Matching Pursuit (BOMP) Eldar and Mishali [2009b,a] and Eldar et al. [2010] proposed a greedy algorithm called BOMP for block-sparse recovery which is given in Algorithm 1. The basic intuition behind BOMP is as follows. Due to the block sparsity of $\boldsymbol{x}$, 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 groups due to the block structure of the sparsity pattern in $\boldsymbol{x}$. 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

Algorithm 1 The BOMP – block-sparse recovery

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

$\boldsymbol{r}^0 = \boldsymbol{b}$, $\boldsymbol{z}^0 = \boldsymbol{0}$, $\Lambda^0 = \emptyset$, $\ell = 0$

repeat

  1. match: $\boldsymbol{h}_j = A_j^T \boldsymbol{r}^\ell, \quad j = 1, 2, \cdots, P$

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

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

  4. update signal estimate:

    $\boldsymbol{z}^{\ell+1} = \arg\min_{\boldsymbol{s}:\text{supp}(\boldsymbol{s})\subseteq\Lambda^{\ell+1}} \|\boldsymbol{b} - A\boldsymbol{s}\|_2$,

    where $\text{supp}(\boldsymbol{s})$ indicates the blocks

    on which $\boldsymbol{s}$ may be non-zero

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

  6. increase index ℓ **by** 1

until stopping criteria true

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

blocks. This process repeats until the stopping criteria are met.

Group LASSO (GLASSO) As an alternative to greedy algorithms such as BOMP for block-sparse recovery, there exist algorithms that can be categorized as convex optimization algorithms. In this section, we review the GLASSO algorithm that can be used for CTI. This algorithm is an extension of the famous Least Absolute Shrinkage and Selection Operator (LASSO) algorithm (Tibshirani [1996]) first introduced for linear regression in statistics. The LASSO is defined as

$$\widehat{\boldsymbol{x}}(\lambda) = \arg\min_{\boldsymbol{x}} \|\boldsymbol{b} - A\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1. \tag{9}$$

The first term in the objective function in (9) corresponds to the matching error while the second term promotes sparsity in the solution. In CS literature, it is known that minimizing the ℓ_1-norm of a vector ($\|\boldsymbol{x}\|_1 = \sum_{i=1}^N |x(i)|$) promotes sparsity and can be used as a convex relaxation for minimizing the ℓ_0-norm (a pseudo-norm that counts the number of non-zero entries of a vector). The parameter λ is a tuning knob that controls the trade-off between these two objectives in (9). As can be understood from the minimization (9), the LASSO algorithm simply seeks a *sparse* solution with no further constraints about the non-zero entries.

As an extension of the LASSO algorithm, Yuan and Lin [2006] proposed the GLASSO algorithm to consider another piece of information about the true solution, namely, the group appearance of the non-zero entries. The GLASSO is an ℓ_2/ℓ_1 minimization problem defined as

$$\widehat{\boldsymbol{x}}(\lambda) = \arg\min_{\boldsymbol{x}} \|\boldsymbol{b} - A\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{\mathcal{E}}_{\boldsymbol{x}}\|_1. \tag{10}$$

In (10), $\boldsymbol{\mathcal{E}}_{\boldsymbol{x}} \in \mathbb{R}^P$ is a vector containing the energy (in terms of the ℓ_2-norm) of the P partitions $\left\{\boldsymbol{x}^j\right\}_{j=1}^P$ associated with the vector $\boldsymbol{x}$ and is defined as

$$\boldsymbol{\mathcal{E}}_{\boldsymbol{x}} = \left[\|\boldsymbol{x}^1\|_2 \ \|\boldsymbol{x}^2\|_2 \ \cdots \|\boldsymbol{x}^P\|_2\right]^T. \tag{11}$$

In general each partition can have an arbitrary size, i.e., $\boldsymbol{x}^j \in \mathbb{R}^{m^j}$ with $\sum_{j=1}^P m^j = N$. In a block-sparse structure, it is assumed that all the partitions have the same length m (i.e., all $m^j = m$ and $N = Pm$). Such information about the partitioning of the signal must be known *a priori* for this kind of recovery algorithm.

4.3 Recovery Conditions

In this section, we provide recovery conditions for the algorithms mentioned in the previous section.

Eldar et al. [2010] proposed a sufficient condition for BOMP to recover any sufficiently concise block-sparse signal $\boldsymbol{x}$ from compressive measurements. This condition depends on the properties of A, specifically on two coherence metrics, block and sub-block coherence.

Assume for the moment that the matrix A (defined in (4)) has columns of unit norm. The block-coherence (Eldar et al. [2010]) of A is defined as

$$\mu_{\text{block}}(A) := \max_{i,j\neq i} \frac{1}{m}\|(A_i^T A_j)\|_2 \tag{12}$$

where $\|A\|_2$ is the spectral norm of a matrix A. In the case where $m = 1$, this matches the conventional definition of coherence (Donoho and Huo [2001], Tropp [2004]),

$$\mu(A) := \max_{i,j\neq i} |\boldsymbol{a}_i^T \boldsymbol{a}_j|, \tag{13}$$

where $\{\boldsymbol{a}_i\}_{i=1}^P$ are columns of A. While μ_{block} characterizes the intra-block relationships within A, the inter-block properties can be quantified by the sub-coherence (Eldar et al. [2010]) of A defined as

$$\mu_{\text{sub-block}}(A) := \max_k \max_{i,j\neq i} |\boldsymbol{a}_{k_i}^T \boldsymbol{a}_{k_j}|, \tag{14}$$

where $\boldsymbol{a}_{k_i}, \boldsymbol{a}_{k_j}$ are columns of the matrix-block A_k.

The recovery condition is stated in terms of the block-adapted coherence metrics, μ_{block} and $\mu_{\text{sub-block}}$ of the matrix A.

Theorem 1. [Eldar et al., 2010, Theorem 3] If $\boldsymbol{x}$ is K-block sparse with blocks of length m, then BOMP will recover $\boldsymbol{x}$ from the measurements $\boldsymbol{b} = A\boldsymbol{x}$ if

$$Km < \mu_T, \tag{15}$$

where

$$\mu_T = \frac{1}{2}\left(\mu_{\text{block}}^{-1} + m - (m-1)\frac{\mu_{\text{sub-block}}}{\mu_{\text{block}}}\right). \tag{16}$$

When $m = 1$, (15) is equivalent to the recovery guarantee proposed for the OMP algorithm (Tropp [2004]), namely, $K < \frac{1}{2}(\mu^{-1} + 1)$. Theorem 1 indicates that for a given matrix A with certain block-adapted coherence metrics (i.e., μ_{block} and $\mu_{\text{sub-block}}$), the BOMP algorithm is guaranteed exact recovery of block-sparse signals of a certain sparsity level specified by (15). The larger the μ_T, the higher the permitted value of K, and the broader the class of signals that can be recovered via BOMP.

As a result of the Karush-Kuhn-Tucker (KKT) conditions for the GLASSO problem, one can derive necessary and sufficient conditions for a solution to (10). Before proceeding, observe that $A\boldsymbol{x} = \sum_{j=1}^P A_j \boldsymbol{x}^j$ where $A_j \in \mathbb{R}^{M\times m}$ and $\boldsymbol{x}^j \in \mathbb{R}^m$. The corresponding subgradient equations for (10) for $j = 1, 2, \ldots, P$ are

$$-2A_j^T(\boldsymbol{b} - \sum_{j=1}^P A_j \boldsymbol{x}^j) + \lambda\frac{\partial\|\boldsymbol{x}^j\|_2}{\partial \boldsymbol{x}^j} = \boldsymbol{0}. \tag{17}$$

When $\boldsymbol{x}^j \neq \boldsymbol{0}$, $\frac{\partial\|\boldsymbol{x}^j\|_2}{\partial \boldsymbol{x}^j} = \frac{\boldsymbol{x}^j}{\|\boldsymbol{x}^j\|_2}$ while if $\boldsymbol{x}^j = \boldsymbol{0}$, the subgradient $\frac{\partial\|\boldsymbol{x}^j\|_2}{\partial \boldsymbol{x}^j}$ is any vector with ℓ_2-norm smaller than

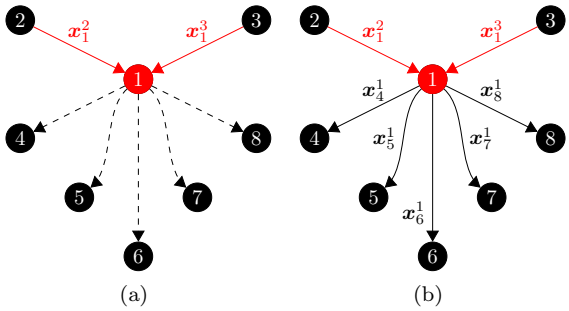


Fig. 2. (a) Disconnected network. Node 1 has in-degree 2 but is disconnected from the rest of the network (i.e., out-degree zero). (b) Connected network. Node 1 has in-degree 2 and is connected to the rest of the network (i.e., in this example out-degree 5).

one. Combining these with (17), we have the following proposition.

Proposition 2. [Yuan and Lin, 2006, Proposition 2.1] A necessary and sufficient condition for $\boldsymbol{x} = [\boldsymbol{x}^{1^T} \cdots \boldsymbol{x}^{P^T}]^T$ to be a solution to (10) is

$$A_j^T(\boldsymbol{b} - \sum_{j=1}^{P} A_j \boldsymbol{x}^j) = \frac{\lambda \boldsymbol{x}^j}{2\|\boldsymbol{x}^j\|_2}, \qquad \forall \boldsymbol{x}^j \neq \boldsymbol{0} \tag{18}$$

$$\|A_j^T(\boldsymbol{b} - \sum_{j=1}^{P} A_j \boldsymbol{x}^j)\|_2 \leq \frac{\lambda}{2}, \qquad \forall \boldsymbol{x}^j = \boldsymbol{0}. \tag{19}$$

4.4 Analysis of the BOMP Recovery Conditions for CTI

In this section, we analyze the recovery conditions of the algorithms when they are applied for CTI. We start with the BOMP algorithm.

As mentioned earlier, the block-adapted coherence metrics give a sufficient condition for recovery via BOMP. Thus, it is of interest to investigate these metrics in the context of CTI. To highlight the important role of these metrics, we refer to them collectively as the *network coherence*. In what follows, we list the key elements that influence the recovery performance of CTI.

1) Network Topology: One of the factors that affects the recovery performance is the network structure itself. Fig. 2 illustrates two similar network structures where recovery of incoming links to node 1 is of interest. In both structures, node 1 has in-degree 2. However, in Fig. 2(a) node 1 has out-degree zero while in Fig. 2(b) node 1 has out-degree 5. Simulations show that the disconnected network (Fig. 2(a)) has better recovery performance compared to the connected network (Fig. 2(b)) even though in both structures, node 1 has the same in-degree. This can be explained with network coherence, as in the disconnected network has a smaller typical network coherence, resulting in better recovery performance. We do not provide the simulation results for this section for the sake of saving space. For a detailed discussion on network coherence, see Sanandaji et al. [2011b].

2) Number of Measurements: Typically, taking more measurements improves the recovery performance of any algorithm. This is true for CTI using BOMP. However, [Sanandaji et al., 2011b, Theorem 2] showed that this can not be anticipated using existing recovery guarantees. Considering a simple network structure, they showed that the typical network coherence is bounded below as number of measurements increases, when the network is perturbed with independent and identically distributed (i.i.d.) Gaussian random sequences. While finding network coherence bounds for more complicated interconnected networks (e.g., networks with loops and nodes of high out-degree) is a harder task, these important characteristics are observed: (i) In the limit, the network coherence metrics are independent of the number of measurements. (ii) The network coherence metrics are bounded below by a non-zero value that depends on the link impulse responses.

The asymptotic behavior of the network coherence metrics is contrary to the conventional behavior in CS, in which increasing the number of rows of a dense matrix (number of measurements M) populated with i.i.d. Gaussian random variables will make the coherence approach a zero value, guaranteeing the recovery of signals with more and more non-zero coefficients. Therefore, the latter phenomenon may suggest an ultimate limitation of the coherence metrics in the analysis of interconnected dynamical networks. Nevertheless, the simulations do indicate that as the number of measurements M increases, recovery remains possible for a range of interesting problem sizes. Investigating the gap that exists between coherence-based recovery guarantees and actual recovery performance is a current and challenging topic.

3) Input Sequence: A more critical factor that has an influence on the recovery performance is the type of input perturbations $u_i(t)$. For example, if instead of random inputs, we perturb a given network with constant inputs, most probably the recovery performance will be decreased. On the other hand, in most of the existing analysis, the input to each node is assumed to be an i.i.d. Gaussian random sequence due to its amenability to analysis. However, it is not clear if an i.i.d. Gaussian random sequence is an optimal choice for CTI of a given network structure. The effect of the input sequence can be explained via network coherence. Sanandaji et al. [2011c] showed how these coherence-based metrics can be improved via a pre-filtering scheme.

An Input Design Scenario: Coherence metrics only reflect the worst-case inner products between the columns of a given matrix and therefore are blunt tools for measuring the actual recovery performance. To illustrate this, consider a matrix that contains many low-correlated columns (i.e., they are almost orthogonal), except one pair of columns is highly correlated. The coherence of such a matrix is high due to that one pair of columns. In such situations, the coherence measure might be misleading in regards to the actual recovery performance.

We propose a pre-filtering scheme to improve the coherence by translating the pressure from the highly correlated columns to the less correlated columns in a way that decreases the overall coherence (Sanandaji et al. [2011c]). This idea is particularly effective in the analysis of dynamical systems where typically we perturb the system with white persistently exciting input sequences and measure the colored outputs. The corresponding matrix in these cases usually contain some columns that are mostly af-

fected by the input signals ($A_{\boldsymbol{u}}$) and therefore less correlated and contain some columns that are built upon output signals ($A_{\boldsymbol{b}}$) and therefore are more correlated. An input design problem can be considered in which the overall coherence of the matrix is reduced. We design an input sequence that is colored, leading to higher coherence of $A_{\boldsymbol{u}}$ but lower coherence of $A_{\boldsymbol{b}}$ and also the overall matrix A (Sanandaji et al. [2011c]).

4.5 Analysis of the GLASSO Recovery Conditions for CTI

The GLASSO has also been applied for CTI. Haufe et al. [2008] applied GLASSO to multivariate time series. A more recent work by Bolstad et al. [2011] also applied GLASSO to a causal network inference problem and derived conditions under which GLASSO consistently estimates the sparse network structure. Defining the *false connection score*, they show that when the network size and the number of node observations grow to infinity (when the sample size can grow much slower compared to the network size), then the network topology can be inferred with high probability in the limit. The result is based on applying the KKT conditions to (10) and deriving recovery conditions when the number of samples goes to infinity. Among the conditions suggested for structure identification, the false connection score is the most critical condition. However, analyzing this condition reveals that: i) the network should have a sparse flow, and ii) the nodes in the graph are sufficiently independent from each other. While the first condition seems to be true for most problems of our interest, the second condition is more challenging, imposing a restriction on the type of structure of the considered graph. Moreover, the assumption that input perturbations are i.i.d. Gaussian sequences seems restricting while in the network coherence the effect of the input perturbations is coded in the coherence metrics.

REFERENCES

A. Bolstad, B.D. Van Veen, and R. Nowak. Causal network inference via group sparse regularization. *IEEE Trans. Signal Processing*, 59(6):2628–2641, 2011.

E.J. Candès. Compressive sampling. *Proc. of the Int. Congress of Mathematicians*, 3:1433–1452, 2006.

E.J. Candès and T. Tao. Decoding via linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

E.J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

E.J. Candès and M.B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

V. Cevher, P. Indyk, C. Hegde, and R.G. Baraniuk. Recovery of clustered sparse signals from compressive measurements. *Proc. of the Int. Conf. on Sampling Theory and Applications*, 2009.

S. Chen, D.L. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.

A. Chiuso and G. Pillonetto. Nonparametric sparse estimators for identification of large scale linear systems. In *Proc. of the* 49*th IEEE Conf. on Decision and Control*, pages 2942–2947, 2010.

D.L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47 (7):2845–2862, 2001.

Y.C. Eldar and M. Mishali. Block-sparsity and sampling over a union of subspaces. *Proc. of the* 16*th Int. Conf. on Digital Signal Processing*, pages 1–8, 2009a.

Y.C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. on Inform. Theory*, 55(11):5302–5316, 2009b.

Y.C. Eldar, P. Kuppinger, and H. Bölcskei. Block-sparse signals: uncertainty relations and efficient recovery. *IEEE Trans. on Signal Processing*, 58(6):3042–3054, 2010.

S. Haufe, K.R. Müller, G. Nolte, and N. Krämer. Sparse causal discovery in multivariate time series. *Proc. of the NIPS Workshop on Causality*, 6:97–106, 2008.

L. Ljung. System Identification: Theory for the User. *Englewood Cliffs, NJ: Prientice Hall*, 1987.

D. Materassi, G. Innocenti, and L. Giarré. Reduced complexity models in the identification of dynamical networks: links with sparsification problems. *Proc. of the* 48*th IEEE Conf. on Decision and Control*, pages 4796–4801, 2009.

D. Napoletani and T.D. Sauer. Reconstructing the topology of sparsely connected dynamical networks. *Physical Review E*, 77(2):026103, 2008.

D. Needell and J.A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26(3):301–321, 2009.

B.M. Sanandaji, T.L. Vincent, and M.B. Wakin. Compressive topology identification of interconnected dynamic systems via clustered orthogonal matching pursuit. *Proc. of the* 50*th IEEE Conf. on Decision and Control and European Control Conf.*, 2011a.

B.M. Sanandaji, T.L. Vincent, and M.B. Wakin. Exact topology identification of large-scale interconnected dynamical systems from compressive observations. *Proc. of the* 2011 *American Control Conf.*, pages 649–656, 2011b.

B.M. Sanandaji, T.L. Vincent, M.B. Wakin, Roland Tóth, and Kameshwar Poolla. Compressive system identification of LTI and LTV ARX models. *Proc. of the* 50*th IEEE Conf. on Decision and Control and European Control Conf.*, 2011c.

R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996. ISSN 0035-9246.

M. Timme. Revealing network connectivity from response dynamics. *Physical review letters*, 98(22):224101, 2007.

J.A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10): 2231–2242, 2004.

J.A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inform. Theory*, 52(3):1030–1051, 2006.

M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, 68(1):49–67, 2006.