Automatica 68 (2016) 9–17

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Sparse network identifiability via Compressed Sensing☆

David Hayden [a], Young Hwan Chang [b], Jorge Goncalves [a,d], Claire J. Tomlin [c]

[a] Department of Engineering, University of Cambridge, CB2 1PZ, UK
[b] Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR 97239, USA
[c] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA
[d] University of Luxembourg, Faculté des Sciences, de la Technologie et de la Communication, 7 Avenue des Hauts Fourneaux, L-4362 Belval, Luxembourg

ARTICLE INFO

Article history:
Received 14 November 2014
Received in revised form
27 October 2015
Accepted 20 December 2015
Available online 22 February 2016

Keywords:
Closed-loop identification
Identifiability
Directed graphs
Interconnection matrices
Linear equations

ABSTRACT

The problem of identifying sparse solutions for the link structure and dynamics of an unknown linear, time-invariant network is posed as finding sparse solutions x to $Ax = b$. If the matrix A satisfies a rank condition, this problem has a unique, sparse solution. Here each row of A comprises one experiment consisting of input/output measurements and cannot be freely chosen. We show that if experiments are poorly designed, the rank condition may never be satisfied, resulting in multiple solutions. We discuss strategies for designing experiments such that A has the desired properties and the problem is therefore well posed. This formulation allows prior knowledge to be taken into account in the form of known nonzero entries of x, requiring fewer experiments to be performed. Simulated examples are given to illustrate the approach, which provides a useful strategy commensurate with the type of experiments and measurements available to biologists. We also confirm suggested limitations on the use of convex relaxations for the efficient solution of this problem.

1. Introduction

Compressed Sensing (CS) refers to the ability to find a sparse solution x to the under-determined set of equations $Ax = b$ (Donoho, 2006a). This problem is relevant in applications in computer vision and signal processing, where a signal often has a representation that is sparse in some domain and can hence be recovered by making relatively few samples in that domain (Candès & Wakin, 2008). Specifically, suppose some signal $\theta \in \mathbb{R}^n$ can be expressed in a basis $\Phi \in \mathbb{R}^{n\times n}$ such that $x = \Phi\theta$ where x is sparse in the sense that $\|x\|_0 = k < n$ and $\|\cdot\|_0$ denotes the number of nonzero entries of a vector. By taking $m < n$ samples of x via an appropriately chosen matrix $A \in \mathbb{R}^{m\times n}$, known as a *sensing matrix*, we may recover x and hence θ.

A related problem is that of identifying the link structure and dynamics of an unknown network from certain observations of it. This is a general inverse problem, currently of particular importance in cell-biological applications, such as identifying Genetic Regulatory Networks (GRNs) (Bansal, Belcastro, Ambesi-Impiombato, & di Bernardo, 2007; De Smet & Marchal, 2010; Marbach et al., 2002, 2010; Michailidis & d'Alché-Buc, 2013). In this context the problem is typically under-determined due to both a paucity of data and limitations on the number of experiments that can be performed. The underlying network is often known to be sparse in the sense that the degree of each node is bounded, and the assumption of sparsity is commonly used as a heuristic to obtain a solution (August & Papachristodoulou, 2009; Bansal, Della Gatta, & di Bernardo, 2006; Bolstad, Van Veen, & Novak, 2011; Chang & Tomlin, 2011; Chiuso & Pillonetto, 2012; Gardner, di Bernardo, Lorenz, & Collins, 2003; Materassi, Innocenti, Giarré, & Salapaka, 2013; Napoletani & Sauer, 2008; Sanandaji, Vincent, & Wakin, 2011; Seneviratne & Solo, 2012; Yeung, Tegnér, & Collins, 2002). This problem is fundamentally different from typical CS applications in that the sensing matrix cannot be chosen freely, but arises based on the experiments applied and the underlying system itself.

The network reconstruction problem was considered for Linear, Time-Invariant (LTI) systems with full state measurement in Chang and Tomlin (2011) using time-series data of a single perturbation

☆ This research was supported by the Engineering and Physical Sciences Research Council under Grant EP/G066477/1 and by the NIH NCI under the ICBP and PS-OC programs (5U54CA112970-08). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Alessandro Chiuso under the direction of Editor Torsten Söderström.

E-mail addresses: davidphayden@gmail.com (D. Hayden), chanyo@ohsu.edu (Y.H. Chang), jmg77@cam.ac.uk (J. Goncalves), tomlin@eecs.berkeley.edu (C.J. Tomlin).

http://dx.doi.org/10.1016/j.automatica.2016.01.008
0005-1098/

to the network. If the perturbation stimulates all manifest states and sufficiently many time points are observed, the solution is shown to be unique and hence only one such experiment is required. By assuming that the solution is sparse, the required number of time points can be reduced. In Sanandaji et al. (2011), sparse networks of FIR filters are treated, again with the aim of reducing the number of data points required. The parameters of the filters are estimated using Block Orthogonal Matching Pursuit (Eldar, Kuppinger, & Bolcskei, 2010) and the notion of *network coherence* is introduced, referring to the coherence (see end of Section 2.1 and Candès & Romberg, 2007) of a network-derived sensing matrix. Network coherence is observed to have a lower bound for some cases, but must be sufficiently small for exact solution via convex relaxations, suggesting a possible limitation for efficient network reconstruction algorithms.

Other examples from the literature include Chiuso and Pillonetto (2012), in which the problem is posed as sparse input selection for MISO LTI systems and Materassi et al. (2013) which concerns the estimation of sparse MISO Wiener filters. For general MIMO LTI systems with deterministic inputs, it was shown in Gonçalves and Warnick (2008) that a certain number of targeted inputs are required in order for the problem to be well posed. This is equivalent to performing experiments to probe the network, for example in a biological context using genetic mutations to identify GRNs (Hayden, Yuan, & Gonçalves, 2013). Here we suppose that sufficient data points are available and consider whether the assumption of sparsity can be used to reduce the number of such experiments required. Our focus is therefore on the *identifiability* of the network, rather than a particular method, although the approach naturally provides an algorithm for steady state or frequency domain identification.

Our contributions are as follows. First we give conditions for sparse solution uniqueness and show that for a poor choice of experiments even the sparsest solution may not be unique. Specifically, for single-input experiments, multiple sparse solutions may exist; in this case we derive constraints on possible network topologies. Strategies for experiment design to ensure a unique solution are then presented and evaluated in simulation. This illustrates the effectiveness of the experiment design and highlights the importance of low network coherence for solving the problem efficiently using, for example, basis pursuit. As in Sanandaji et al. (2011), the coherence is found to depend on the magnitude of network parameters, presenting a fundamental limitation to the approach.

In Section 2 we review some standard results in CS and network reconstruction for LTI systems. Then in Section 3 we discuss how prior knowledge can be incorporated directly into the CS framework and how this reduces the number of experiments needed for a unique solution. Section 4 addresses the problem of sparse network identifiability, first showing that the standard assumptions of CS are not sufficient for exact reconstruction, then proposing experimental procedures to ensure solution uniqueness. Section 5 then provides a discussion of practical identification issues. In Section 6 simulation examples are presented to support the results and conclusions are given in Section 7.

Notation

Denote by $A(i,j)$, $A(i,:)$ and $A(:,j)$ entry (i,j), row i and column j respectively of matrix A and by A^T its transpose. The diagonal matrix comprising the diagonal entries of A is denoted $\mathrm{Diag}(A)$. The function $\|x\|_0$ (the l_0 "norm") returns the number of nonzero entries in the vector x, which is said to be k-sparse if at most k of its entries are nonzero: $\|x\|_0 \leq k$. The matrix A is row (column) sparse if all of its rows (columns) are sparse.

2. Background

2.1. Compressed sensing

Sparse solutions, $x \in \mathbb{R}^n$, are sought to the following problem:

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are known and $m < n$. The sparsest such solution (or set of solutions) are the minimizing argument(s) of:

$$\min_x \|x\|_0 \quad \text{subject to } Ax = b. \tag{2}$$

The following well-known lemma (see for example Lemma 2.1 of Donoho, 2006b) provides a sufficient condition that the solution to (2) is unique:

Lemma 1. *If the sparsest solution to (2) has $\|x\|_0 = k$ and $m \geq 2k$ and all subsets of $2k$ columns of A are full rank, then this solution is unique.*

Proof. Suppose two solutions exist: $Ax^{(1)} = b$ and $Ax^{(2)} = b$, where $\|x^{(1)}\|_0 = \|x^{(2)}\|_0 = k$ and subtract one equation from the other:

$$A(x^{(1)} - x^{(2)}) = 0.$$

Let $l := \|x^{(1)} - x^{(2)}\|_0 \leq 2k$ be the number of nonzero entries in $x^{(1)} - x^{(2)}$; then construct the vector $\hat{x} \in \mathbb{R}^l$ by removing all zero entries and construct the matrix $\hat{A} \in \mathbb{R}^{m\times l}$ by removing the corresponding columns from A. From the above equation, $\hat{A}$ and $\hat{x}$ satisfy: $\hat{A}\hat{x} = 0$, where (since $l \leq m$) $\hat{A}$ is full column rank and $\hat{x} \neq 0$, which is a contradiction.

Convex relaxations of (2) are typically sought, such as l_1 minimization (basis pursuit Candès & Wakin, 2008), which can be solved by linear programming:

$$\min_x \|x\|_1 \quad \text{subject to } Ax = b. \tag{3}$$

It has been shown that l_1 minimization solves exactly (2) if m is sufficiently large and the matrix A is sufficiently *incoherent* (Candès & Romberg, 2007; Donoho, 2006b). Coherence is a measure of similarity between pairs of columns of a matrix, defined as follows:

$$\mu(A) = \max_{i<j} \frac{|A_i^T A_j|}{\|A_i\|_2 \|A_j\|_2}$$

where A_i denotes the ith column of A. If all columns are orthogonal, the coherence is zero; if any two columns are linearly dependent, the coherence is unity. The coherence property allows one to distinguish problems with a unique solution (that can therefore be solved by exhaustive minimization of (2)) from those that, in addition, can be solved efficiently by the convex relaxation (3). Numerical simulations suggest that in practice, most k-sparse signals require $m \geq 4k$ in order to be recovered exactly by l_1 minimization (Candès & Romberg, 2007).

2.2. Dynamical networks

Consider a vector of directly observed variables $y(t) \in \mathbb{R}^p$, whose entries $y_i(t)$ are governed by the following set of LTI equations for $i = 1, \ldots, p$:

$$y_i(t) = \sum_{j\neq i} q_{ij}(t) * y_j(t) + \sum_{k=1}^{r} p_{ik}(t) * u_k(t) + v_i(t) \tag{4}$$