

Compressed Sensing for Network Reconstruction [★]

David Hayden ^a, Young Hwan Chang ^b, Jorge Goncalves ^{a,c}, Claire Tomlin ^b

^a*Department of Engineering, University of Cambridge, CB2 1PZ, UK*

^b*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA*

^c*University of Luxembourg, Faculté des Sciences, de la Technologie et de la Communication, 7 Avenue des Hauts Fourneaux, L-4362 BELVAL, Luxembourg*

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 sensing 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 experimental strategies for designing experiments such that the sensing matrix 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. A number of 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.

Key words: Closed-loop identification; directed graphs; identifiability; interconnections matrices; linear equations.

1 Introduction

Compressed Sensing (CS) refers to the ability to find a sparse solution x to the underdetermined set of equations $Ax = b$ [1]. 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 [2]. Specifically, suppose some signal $\theta \in \mathbb{R}^n$ can be expressed in an orthonormal 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 sensing matrix $A \in \mathbb{R}^{m \times n}$, 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) [3–7]. In this context the problem is typically underdetermined 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 an heuristic to obtain a solution [8–18]. 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 problem was considered for Linear, Time-Invariant (LTI) systems with full state measurement in [14] using time-series data of a single perturbation to the network. If 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 [13], 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 [19] and the notion of *network coherence* is introduced, referring to the

[★] This paper was not presented at any IFAC meeting. Corresponding author D. Hayden.

Email addresses: dph34@cam.ac.uk (David Hayden), yhchang@berkeley.edu (Young Hwan Chang), jmg77@cam.ac.uk (Jorge Goncalves), tomlin@eecs.berkeley.edu (Claire Tomlin).

coherence (see [20]) of a network-derived sensing matrix. The network coherence is observed to have a lower bound for some cases, which suggests a possible limitation to the use of CS for network reconstruction.

Other examples from the literature include [17], in which the problem is posed as sparse input selection for MISO LTI systems and [18] which concerns the estimation of sparse MISO Wiener filters. For general MIMO LTI systems with deterministic inputs, it was shown in [21] 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 [22]. 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 show that for a poor choice of experiments even the sparsest solution may not be unique; then the problem of experiment design to ensure solution uniqueness is addressed; finally, simulated examples demonstrate the effectiveness of the experiment design but reveal similar limitations to those observed in [13] with regards to the network coherence. Hence although the sparse solution may be unique, it may be difficult to find using, for example, basis pursuit.

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 exact reconstruction. Section 4 addresses the problem of underdetermined network reconstruction, first showing that the standard assumptions of CS are not sufficient for exact reconstruction, then proposing experimental procedures to ensure exact reconstruction. In Section 5 a number of simulation examples are presented to support the results and conclusions are given in Section 6.

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 $\text{Diag}(A)$. The function $\|x\|_0$ (the l_0 “norm”) returns the number of nonzero entries in the vector x .

2 Background

2.1 Compressed Sensing

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

$$Ax = b \quad (1)$$

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:

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

The following well-known lemma 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: $A(x^{(1)} - x^{(2)}) = 0$. Since $\|x^{(1)} - x^{(2)}\|_0 \leq 2k$, this equation is equivalent to: $\hat{A}\hat{x} = 0$, where $\hat{A} \in \mathbb{R}^{m \times l}$, $0 \neq \hat{x} \in \mathbb{R}^l$, $l \leq m$ and \hat{A} is full rank, which is a contradiction.

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

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

It has been shown that l_1 minimization solves exactly (2) if m is sufficiently large and the matrix A is sufficiently *incoherent* [20, 23]. The coherence of a matrix is 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 i^{th} column of A . Numerical simulations suggest that in practice, most k -sparse signals require $m \geq 4k$ in order to be recovered exactly [20].

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 $i = 1, \dots, p$:

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

for causal impulse response functions $q_{ij}(t)$ and $p_{ik}(t)$ and vector of inputs $u(t) \in \mathbb{R}^r$. Equation (4) defines a directed graph among observed variables and inputs with no self-loops and in which there is an edge from y_j (u_k) to y_i if and only if $q_{ij}(t) \not\equiv 0$ ($p_{ik}(t) \not\equiv 0$). The functions $q_{ij}(t)$ and $p_{ik}(t)$ therefore define both the topology and edge dynamics of the graph.

By taking the Laplace transform of (4), the system is represented compactly in matrix form as follows:

$$Y = QY + PU \quad (5)$$

where $Y(s)$ and $U(s)$ are the Laplace transforms of $y(t)$ and $u(t)$ respectively; $Q(s)$ is a strictly-proper transfer matrix with entry (i, j) equal to the Laplace transform of $q_{ij}(t)$ and diagonal entries equal to zero; $P(s)$ is a strictly-proper transfer matrix with entry (i, k) equal to the Laplace transform of $p_{ik}(t)$. The couple (Q, P) is termed the Dynamical Structure Function (DSF) and is uniquely defined for any partially-observed LTI system [21].

A state-space realization of (4) can always be made in which the state vector is partitioned into *manifest* states, corresponding to the observed variables y_i and *latent* states, which provide the dynamics of the impulse response functions q_{ij} and p_{ik} . By observing more of the states, a different DSF will be obtained that represents the system in greater detail, hence the DSF may be regarded as a representation of the system at the *resolution* $(\frac{p}{n})$, where n is the total number of states) of the manifest states. Similarly, manifest states may be treated as latent in order to obtain a coarser DSF at a lower resolution.

2.3 Network Reconstruction

We pose the network reconstruction problem as obtaining (Q, P) from input/output data (U, Y) . Suppose m independent experiments have been performed in each of which a different input (or set of inputs) has been applied. The largest number of independent experiments is r , the dimension of the inputs, so $m \leq r$. Denote the Laplace transforms of the inputs and outputs in the i^{th} experiment as $U^{(i)}$ and $Y^{(i)}$ and concatenate these to form the following matrices:

$$\begin{aligned} Y &:= \begin{bmatrix} Y^{(1)} & Y^{(2)} & \dots & Y^{(m)} \end{bmatrix} \\ U &:= \begin{bmatrix} U^{(1)} & U^{(2)} & \dots & U^{(m)} \end{bmatrix} \end{aligned}$$

of dimension $p \times m$ and $r \times m$ respectively. Note that $Y = QY + PU$ and rearrange this to give:

$$\begin{bmatrix} Y^T & U^T \end{bmatrix} \begin{bmatrix} Q^T \\ P^T \end{bmatrix} = Y^T \quad (6)$$

where $\begin{bmatrix} Y^T & U^T \end{bmatrix}$ has dimension $m \times (p + r)$ and we wish to solve for Q and P . By applying the vectorization operator we can write (6) in the form of (1):

$$\begin{aligned} A &\leftarrow I \otimes \begin{bmatrix} Y^T & U^T \end{bmatrix} \\ x &\leftarrow \text{vec} \left(\begin{bmatrix} Q^T \\ P^T \end{bmatrix} \right) \\ b &\leftarrow \text{vec}(Y^T) \end{aligned} \quad (7)$$

where $A(s) \in \mathbb{C}^{M \times N}$, $x(s) \in \mathbb{C}^N$ and $b(s) \in \mathbb{C}^M$ for $M = mp$ and $N = p(p + r)$.

With no other information about the system, in order for (6) to be well posed it is therefore required that:

$$M \geq N \quad \Leftrightarrow \quad m \geq p + r \quad \Leftrightarrow \quad 0 \geq p \quad (8)$$

since $r \geq m$, and hence additional information is always necessary. The assumption that rows of Q and P are sparse may be sufficient to ensure a unique sparse solution. In particular, if each row of $\begin{bmatrix} Q & P \end{bmatrix}$ is k -sparse then the solution to (6) is unique if:

$$M \geq 2kp \quad \Leftrightarrow \quad m \geq 2k$$

and the condition of Lemma 1 is satisfied for A . However, this may not be the case due to the particular way in which the matrix A is constructed in (7).

Alternatively, we may assume some knowledge of how the inputs target the network, for example that P is square ($r = p$) and diagonal. This defines an experimental setup in which each input is associated with a particular manifest state and affects it via the corresponding diagonal entry of P . By removing the $p - 1$ known zero entries in each row of P and the zero diagonal entries of Q , the condition for solution uniqueness (8) becomes:

$$m \geq p \quad \Rightarrow \quad m = p$$

since $m \leq r = p$. This is the main result of [21] – that with P diagonal and no other *a priori* information, $r = p$ is necessary and sufficient for solution uniqueness.

3 Compressed Sensing with prior knowledge

Here we treat the problem of finding a solution $x \in \mathbb{R}^n$ to the following:

$$Ax = b \quad (9)$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are known, $m < n$ and where x is partitioned into sparse and nonzero components. Without loss of generality, we can write (9) as:

$$\begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = b \quad (10)$$

where $x_1 \in \mathbb{R}^{n_1}$ satisfies $\|x_1\|_0 = k$, $x_2 \in \mathbb{R}^{n_2}$ satisfies $\|x_2\|_0 = n_2$ and $n_1 + n_2 = n$. The vector x is therefore $(k + n_2)$ -sparse – it satisfies $\|x\|_0 = k + n_2$.

This will be applied to the sparse reconstruction problem in the following section, where rows of Q in (7) are sparse and part of P is nonzero. Equation (9) has a unique $(k + n_2)$ -sparse solution by Lemma 1 if $m \geq 2(k + n_2)$ and all subsets of $2(k + n_2)$ columns of A are full rank. By making use of the known structure of x , we can solve for x_1 and x_2 separately and hence reduce the number of experiments, m , required.

This problem has been considered in the CS literature (see for example [24], [25]) where it is well known that the number of experiments required for exact reconstruction can be reduced by n_2 , the number of known nonzero entries of x . However, it is generally assumed that the matrix A can be chosen to satisfy Lemma 1, which is not the case in our application. Here we derive a lemma analogous to Lemma 1 that gives conditions for exact reconstruction.

3.1 Conditions for Exact Reconstruction

Take the QR decomposition of $A_2 \in \mathbb{R}^{m \times n_2}$:

$$A_2 = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \quad (11)$$

where $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \in \mathbb{R}^{m \times m}$ is orthogonal and $R_1 \in \mathbb{R}^{n_2 \times n_2}$ is upper triangular. Pre-multiply (10) by $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^T$:

$$\begin{bmatrix} Q_1^T A_1 & R_1 \\ Q_2^T A_1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} Q_1^T b \\ Q_2^T b \end{bmatrix} \quad (12)$$

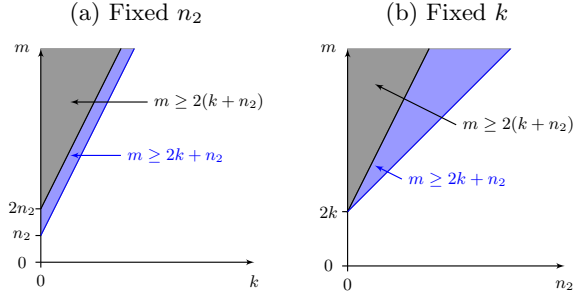


Fig. 1. Values of m that satisfy Lemma 1 (gray) and Lemma 2 (blue) for different values of k (a) and n_2 (b). The blue region represents the increased number of problems that can be solved by incorporating prior information.

We can now solve first for x_1 using the second block row:

$$Q_2^T A_1 x_1 = Q_2^T b \quad (13)$$

where $Q_2^T A_1 \in \mathbb{R}^{(m-n_2) \times n_1}$ and $\|x_1\|_0 = k$. From Lemma 1, (13) has a unique, k -sparse solution if $m - n_2 \geq 2k$ and all subsets of $2k$ columns of $Q_2^T A_1$ are full rank. The number of experiments, m , required for solution uniqueness has therefore been reduced by the number of known components of x from $2(k + n_2)$ to $2k + n_2$.

Given x_1 , we may then solve for x_2 from the first block row of (12):

$$R_1 x_2 = Q_1^T (b - A_1 x_1)$$

This has a unique solution if and only if R_1 is full rank, requiring A_2 to be full column rank. In this case, x_2 is given by:

$$x_2 = R_1^{-1} Q_1^T (b - A_1 x_1) \quad (14)$$

The following lemma summarizes the above conditions for sparse solution uniqueness.

Lemma 2 Suppose the sparsest solution to (10) is known to have $\|x_1\|_0 = k$ and $\|x_2\|_0 = n_2$ and let the QR decomposition of A_2 be given by (11). Then if $m \geq 2k + n_2$, if A_2 is full rank and if all subsets of $2k$ columns of $Q_2^T A_1$ are full rank, then the $(k + n_2)$ -sparse solution to (10) is unique.

By incorporating prior knowledge the number of experiments m required for k -sparse solution uniqueness is reduced by n_2 , the number of known nonzero components. Equivalently, for a given number of experiments, the sparse solution for a higher value of k may be unique. These ideas are illustrated in Fig. 1.

4 Underdetermined Reconstruction

Consider an unknown system defined by its Dynamical Structure Function (DSF): (Q^0, P^0) where for inputs $U(s)$ and outputs $Y(s)$ we have:

$$Y = Q^0 Y + P^0 U$$

Given some set of inputs and outputs we seek to identify Q^0 under the following assumptions.

Assumption 1 *The matrix P^0 is square, diagonal and full rank.*

Assumption 2 *The number of experiments m is fewer than the number of measured states p .*

Assumption 3 *The rows of Q^0 are k -sparse:*

$$\|Q^0(i, :)\|_0 \leq k < p \quad \text{for} \quad i = 1, \dots, p$$

Assumption 1 asserts that the dimension of the input vector is equal to that of the manifest state vector ($r = p$) and that each input directly and uniquely affects one manifest state. Assumption 2 then states that we can only perform $m < p$ experiments, and hence without additional assumptions the reconstruction problem is ill posed.

By Assumption 3, each node in the graph defined by Q^0 has maximum in-degree of k and the sparse solution to the reconstruction problem may be unique. In this case we seek sparse solutions Q and diagonal P to:

$$\begin{bmatrix} Y^T & U^T \end{bmatrix} \begin{bmatrix} Q^T \\ P^T \end{bmatrix} = Y^T \quad (15)$$

This may be converted into the form of (10) either by taking the vectorization operator as in (7) or by solving for each row of $\begin{bmatrix} Q & P \end{bmatrix}$ separately:

$$\begin{bmatrix} Y^T & U^T(:, i) \end{bmatrix} \begin{bmatrix} Q^T(:, i) \\ P^T(i, i) \end{bmatrix} = Y^T(:, i) \quad (16)$$

where $P(i, i) \neq 0$. The $(k + 1)$ -sparse solution to (16) is unique if the conditions of Lemma 2 are satisfied.

This problem is fundamentally different to typical compressed sensing applications in that we do not have free choice of the sensing matrix. First we will demonstrate that for a naïve choice of U , the condition of Lemma 2 may never be met; then we will consider the design of the sensing matrix by choice of U .

4.1 Single Inputs

Suppose that in each experiment only one input is applied, such that U can be written as:

$$U = \begin{bmatrix} U_1 \\ 0 \end{bmatrix} \quad (17)$$

where U_1 is square, diagonal and of dimension $m \times m$. Partitioning Q and P accordingly, the system equations can be written as:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} + \begin{bmatrix} P_{11} \\ 0 \end{bmatrix} U_1 \quad (18)$$

where P_{11} is square and diagonal and the manifest states are therefore partitioned into “perturbed” (Y_1) and “unperturbed” (Y_2)¹. Denote the DSF with respect to the p manifest states (Q, P) as the p-DSF; then by eliminating Y_2 from the right hand side of (18) we can derive two further representations of the system.

From the first block row we can obtain the DSF with respect to only the m perturbed states and denote this the m-DSF. This is a representation of the same system but at a lower resolution ($\frac{m}{n}$) of manifest states. Eliminating Y_2 gives:

$$\begin{aligned} Y_1 &= (Q_{11} + Q_{12}(I - Q_{22})^{-1}Q_{21})Y_1 + P_{11}U_1 \\ &=: \bar{Q}_{11}Y_1 + P_{11}U_1 \end{aligned} \quad (19)$$

where in order to obtain a hollow Q matrix (diagonal entries equal to zero) we must subtract the diagonal part of \bar{Q}_{11} from both sides. Let $D_{11} = \text{Diag}(\bar{Q}_{11})$, then:

$$\begin{aligned} Y_1 &= \bar{Q}_{11}Y_1 + D_{11}Y_1 - D_{11}Y_1 + P_{11}U_1 \\ &= (I - D_{11})^{-1}((\bar{Q}_{11} - D_{11})Y_1 + P_{11}U_1) \\ &=: \hat{Q}_{11}Y_1 + \hat{P}_{11}U_1 \end{aligned} \quad (20)$$

The m-DSF is defined as $(\hat{Q}_{11}, \hat{P}_{11})$, which treats only Y_1 as manifest states and Y_2 as additional latent states. This can be equivalently defined from a state-space realization by changing the partitioning of the state vector to reflect the change in manifest states. The m-DSF can therefore be obtained uniquely from Y_1 and U_1 since U_1 is square, diagonal and full rank [21].

¹ A state is unperturbed if it is not directly driven by an input – it may still be indirectly affected via other perturbed states.

The second block row of (18) gives:

$$\begin{aligned} Y_2 &= (I - Q_{22})^{-1} Q_{21} Y_1 \\ &=: \hat{Q}_{21} Y_1 \end{aligned} \quad (21)$$

where the transfer function \hat{Q}_{21} describes causal relations from states in Y_1 to those in Y_2 that are direct in the sense that they do not involve other states in Y_1 . The matrix \hat{Q}_{21} is also identifiable from input/output data.

Lemma 3 *A particular solution to the sparse network reconstruction problem (15) is:*

$$\hat{Q} = \begin{bmatrix} \hat{Q}_{11} & 0 \\ \hat{Q}_{21} & 0 \end{bmatrix}, \quad \hat{P} = \begin{bmatrix} \hat{P}_{11} & 0 \\ 0 & 0 \end{bmatrix}$$

where $(\hat{Q}_{11}, \hat{P}_{11})$ is defined in (20) and \hat{Q}_{21} in (21).

Hence we can always construct at least one solution for Q , in which the unperturbed states, Y_2 , have no outputs to any other measured states. Clearly if $\|\hat{Q}(i, :)\|_0 \leq k$ for any i then the k -sparse solution to (16) is not unique.

Remark 4 *Given Q for which $\max_i \|Q(i, :)\|_0 = k$ and $\max_i \|\hat{Q}(i, :)\|_0 = \hat{k}$, it is possible that either $\hat{k} \leq k$ or $\hat{k} > k$. This result can be easily seen by example, and in the former case the k -sparse solution will not be unique.*

The assumption of sparsity for any network must therefore be firmly justified *a priori* for the representation in question – when viewed at a different resolution of manifest states, the sparsity of the network can change.

4.2 Constraints on Q

Assume single inputs of the form (17) have been applied and partition Q and \hat{Q} of Lemma 3 as follows:

$$Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}, \quad \hat{Q} = \begin{bmatrix} \hat{Q}_1 & 0 \end{bmatrix}, \quad \hat{Q}_1 := \begin{bmatrix} \hat{Q}_{11} \\ \hat{Q}_{21} \end{bmatrix}$$

such that Q_1 has the same dimension as \hat{Q}_1 . From the identifiable quantity \hat{Q}_1 we may infer something about whether entries of Q are zero or not.

Lemma 5 *For every $i \neq j$, if $\hat{Q}_1(i, j) \neq 0$, then*

$$Q_1(i, j) \neq 0 \quad \text{or} \quad Q_2(i, k) \hat{Q}_{21}(k, j) \neq 0$$

for some k . Else $\hat{Q}_1(i, j) = 0$, then

$$Q_1(i, j) = 0 \quad \text{and} \quad Q_2(i, k) \hat{Q}_{21}(k, j) = 0$$

for every k unless the graph defined by Q contains multiple paths that sum to zero.

PROOF. From the definition of \hat{Q}_{11} in (20), for $i \neq j$:

$$\hat{Q}_{11}(i, j) \neq 0 \quad \Leftrightarrow \quad \bar{Q}_{11}(i, j) \neq 0$$

Then from (20) and (21) we have:

$$\begin{bmatrix} \bar{Q}_{11} \\ \hat{Q}_{21} \end{bmatrix} = \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} + \begin{bmatrix} Q_{12} \\ Q_{22} \end{bmatrix} \hat{Q}_{21} = Q_1 + Q_2 \hat{Q}_{21} \quad (22)$$

If $\hat{Q}_1(i, j) \neq 0$, the result now follows directly from (22). If $\hat{Q}_1(i, j) = 0$ it is possible that $Q_1 + Q_2 \hat{Q}_{21} = 0$ and the condition of the lemma not be satisfied. However, $\hat{Q}_1(i, j)$ comprises the sum of a direct link $Q_1(i, j)$ and a sum of paths:

$$Q_2(i, :)\hat{Q}_{21}(:, j) = Q_2(i, :)(I - Q_{22})^{-1} Q_{21}(:, j)$$

in Q . Hence this case necessitates that this direct link and all the paths sum to zero.

The case of multiple paths summing to zero in Q is considered unlikely to occur in practice. Lemma 5 may be interpreted as low resolution structure \hat{Q}_1 having to be consistent with the structure of Q – a connection that exists in \hat{Q}_1 must be preserved (directly or indirectly) in Q . The following example illustrates how this can be used to place constraints on the unknown Q .

Example 1 *Consider the network of Fig. 2(a) with $p = 6$ and $m = 3$. The matrix Q and the particular solution \hat{Q} from Lemma 3 are:*

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \times \\ \times & 0 & 0 & 0 & 0 & 0 \\ 0 & \times & 0 & 0 & 0 & 0 \\ 0 & 0 & \times & 0 & 0 & 0 \\ 0 & 0 & 0 & \times & 0 & 0 \\ 0 & 0 & 0 & 0 & \times & 0 \end{bmatrix}, \quad \hat{Q} = \begin{bmatrix} 0 & 0 & \times & 0 & 0 & 0 \\ \times & 0 & 0 & 0 & 0 & 0 \\ 0 & \times & 0 & 0 & 0 & 0 \\ 0 & 0 & \times & 0 & 0 & 0 \\ 0 & 0 & 0 & \times & 0 & 0 \\ 0 & 0 & \times & 0 & 0 & 0 \end{bmatrix}$$

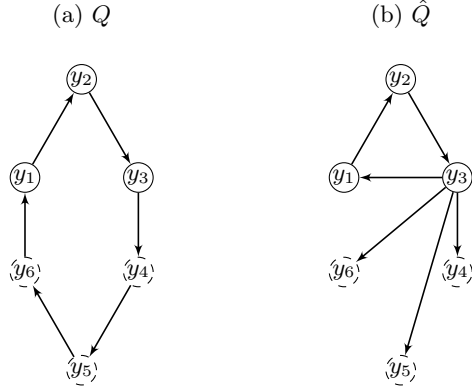


Fig. 2. Network with $p = 6$ and $m = 3$ in which the first three manifest states are perturbed. Solid circles denote perturbed states (Y_1), dashed circles unperturbed states (Y_2) and arrows denote nonzero entries of (a) Q and (b) \hat{Q} .

where \times denotes a nonzero entry. The matrix \hat{Q} is a valid 1-sparse solution to (16) and is shown in Fig. 2(b). Using Lemma 5 we can obtain a matrix Q^c which describes constraints on the solution set of Q :

$$Q^c = \begin{bmatrix} 0 & 0 & ? & ? & ? & ? \\ \times & 0 & 0 & 0 & 0 & 0 \\ 0 & \times & 0 & ? & ? & ? \\ 0 & 0 & ? & 0 & ? & ? \\ 0 & 0 & ? & ? & 0 & ? \\ 0 & 0 & ? & ? & ? & 0 \end{bmatrix}$$

where $?$ denotes an unknown entry. Approximately half of the structure of Q can therefore be found, but the 1-sparse solution is unique for rows two and three only. For any such ring network with $m < p$, only rows $2, \dots, m$ have a unique sparse solution and hence the solution for Q is unique if and only if $m = p$.

With single inputs described by (17), the assumption of sparsity alone is not sufficient to ensure uniqueness of the sparse solution (Q, P) to the network reconstruction problem. This is illustrated by Example 1 in which Q is 1-sparse (unit in-degree) and has multiple 1-sparse solutions for any $m < p$. Applying inputs in this manner does however yield lower resolution structural information that can impose constraints on Q by Lemma 5. In addition, we have not considered the minimality or stability of the solutions – for small problems it may be clear, for example, that the solution of minimal dimension is unique. Next we consider the design of U to satisfy Lemma 2 and hence ensure sparse solution uniqueness.

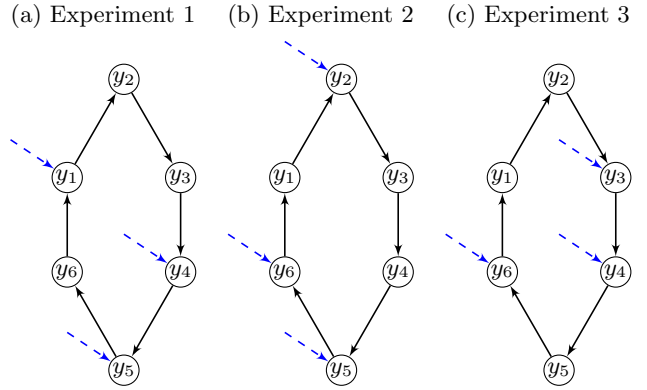


Fig. 3. The network of Fig. 2(a) with $p = 6$, $m = 3$ and three inputs applied in each experiment denoted by the blue dashed arrows.

4.3 Experiment Design

The problem with the diagonal inputs of (17) is that the unperturbed states have no variation independent of their parent states. As a result, subsets of columns of Y^T in (16) may not be full rank and hence may not satisfy Lemma 2. Figure 3 shows the same network as in Fig. 2(a) with three inputs applied in each of the three experiments. It is straightforward to verify that any system with this structure generically satisfies Lemma 2 for $k = 1$ for any particular $s = j\omega$ and therefore has a unique 1-sparse solution. However, without first knowing the structure, one would not be able to design such inputs.

Perturbation design was considered in [26] for fully-observed linear systems using steady-state data. An algorithm is presented in which perturbations are first applied at random until every state has been perturbed in at least one experiment. Then all solutions of a certain sparsity consistent with the data are constructed and a measure of variance for each state is obtained based on how much the outputs of this state vary across all the solutions. The state(s) with the highest variance are then perturbed and the procedure iterated. Simulation results suggested that the more states that are perturbed in each experiment, the fewer experiments are needed.

A similar procedure is presented in [27] for acyclic Boolean networks. At each stage of the algorithm, the next experiment is selected as that which maximizes the decrease in entropy in terms of reducing the set of possible solutions. The set of possible experiments to choose from is taken as given. In [21] and [28] targeted inputs are considered but the number of experiments must be equal to the number of manifest states. In the former, single inputs are considered such that each state is perturbed in turn; in the latter the opposite: every state except one must be perturbed in each experiment.

Three strategies are presented here along similar lines to [26]; in each, experiments are performed iteratively until the solution at a given level of sparsity is unique, according to Lemma 2. At each iteration, a fixed number of inputs, l , are chosen according to:

- (1) Random – inputs are chosen with equal probability
- (2) Biased Random – inputs are chosen at random with a bias towards those that have been applied the least in previous experiments
- (3) Targeted – inputs are chosen to target rank-deficient subsets of columns of the sensing matrix in Lemma 2. Each column of the sensing matrix corresponds to a manifest state and therefore to an input; having identified a deficient subset, the input in this subset that has been applied the least in previous experiments is selected

The third approach will be seen in Section 5 to require the fewest experiments on average to ensure a unique solution; it does however incur the additional computational cost of searching for rank-deficient subsets.

5 Numerical Simulations

5.1 Experimental Design for Solution Uniqueness

Here we compare in simulation the number of experiments required for a unique solution by the three strategies of Section 4.3. Random networks of $p = 20$ measured states were generated with maximum in-degree sparsity of $k = 1, 2, 3, 4$. For each network, we performed experiments with exactly l step inputs applied, for $l = 1, \dots, p$ following each of the three strategies. In each case, the steady-state response of the network was used to assess whether the conditions of Lemma 2 were satisfied. The results are given here for $k = 2$; the same trends were observed for the other values of k .

Figure 4 shows the average number of experiments needed for Lemma 2 to be satisfied over 100 trials for $k = 2$. If only one input is applied in each experiment, the maximum number of experiments is always required and CS hence offers no improvement. Substantial reduction in the number of experiments is observed if more than one input can be applied, and applying these in a biased or targeted manner is more effective than applying them at random. The best results are seen for the targeted approach, although as mentioned this incurs higher computational cost. For a sufficiently large number of inputs per experiment, the number of experiments required for a unique solution is determined by $m = 2k + 1$, independent of the strategy.

5.2 Exact Reconstruction using Basis Pursuit

For 100 random networks of the type treated in Fig. 4 with four inputs per experiment, we attempt to solve

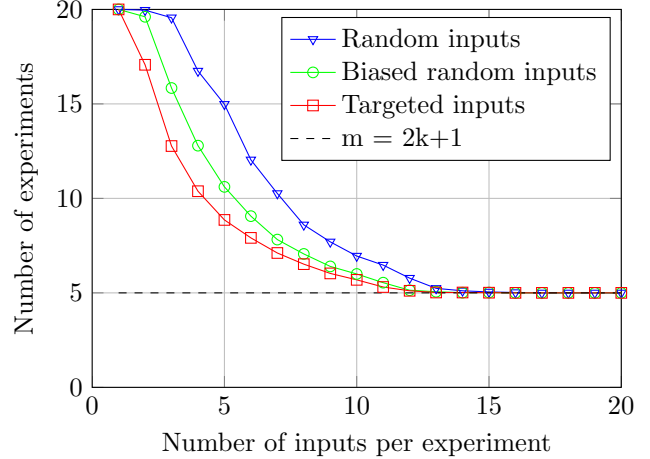


Fig. 4. Mean number of experiments needed for solution uniqueness (Lemma 2) for different numbers of inputs per experiment for three different strategies. The mean was taken over 100 random networks of $p = 20$ manifest states with maximum in-degree sparsity of $k = 2$.

for the 2-sparse solution using basis pursuit (l_1 minimization). A sufficient condition for the success of basis pursuit is that the coherence of the sensing matrix is sufficiently small. In [13], lower bounds on the network coherence of simple networks were derived and also observed in simulation as the number of measurements was increased. In particular the lower bound was increasing in the magnitude of the parameters of the impulse response functions. A similar phenomenon was observed here: increasing the steady-state gain of the entries of Q and P increased the coherence.

For $|Q_{ij}(0)|, |P_{ii}(0)| < 0.5$, Fig. 5 shows the mean coherence of the sensing matrix against the number of experiments. Fig. 6 then shows the success rate of each of the three strategies at recovering the entire network from steady-state data using basis pursuit. The success rate is one if all links are recovered with no false positives and is zero otherwise; for this quite stringent metric, the performance is promising, particularly for the targeted experiments. The relative performance of each of the strategies is consistent with the solution uniqueness results of Fig. 4. The actual number of experiment required for exact reconstruction using basis pursuit can be seen to be higher than that required for a unique solution, as is normally the case [20].

6 Conclusions

We have investigated Compressed Sensing (CS) as a tool for reconstructing dynamical networks from data with a deficient number of experiments. The application differs from the typical use of CS in that the sensing matrix cannot be freely chosen and exact reconstruction hence necessitates appropriate choice of experiments. We provide a formulation of the problem that incorporates prior

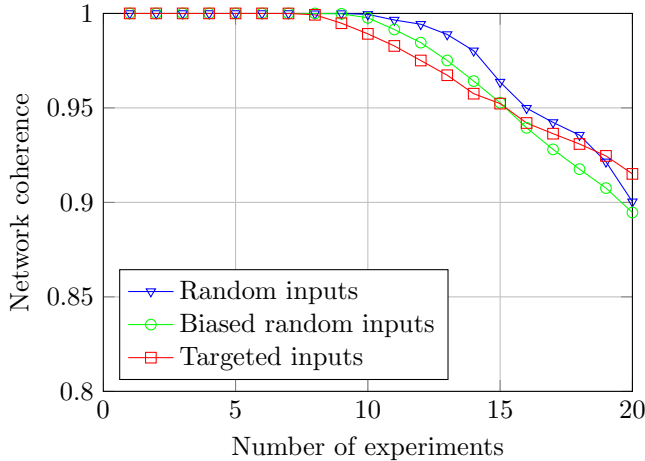


Fig. 5. Mean coherence of the sensing matrix for 100 random networks of $p = 20$ manifest states with maximum in-degree sparsity of $k = 2$ and four inputs applied in each experiment.

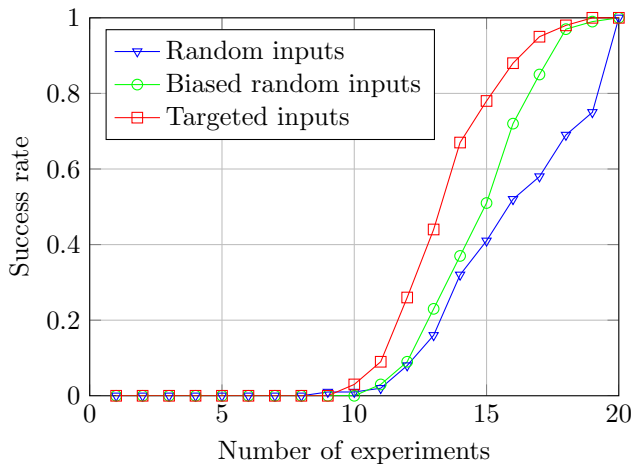


Fig. 6. Mean success rate using basis pursuit for the 100 networks of Fig. 5.

knowledge and present strategies for experimental design to attain sparse solution uniqueness with fewer experiments.

The problem is motivated by biological applications where data are typically scarce; in this context we provide an algorithm for reconstruction at particular frequencies, for example at steady state, where a small number of data points suffice. It is also straightforward to apply this approach when the inputs are unknown, as may be the case in practice. Simulations demonstrate that the problem may be solved efficiently using a convex relaxation, such as basis pursuit, if the network coherence metric is sufficiently small. We also observe previously identified lower bounds on the network coherence, which highlight a potential limitation of the use of CS for this application.

Acknowledgements

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).

References

- [1] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [2] E. J. Candès and M. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, March:21–30, 2008.
- [3] M. Bansal, V. Belcastro, A. Ambesi-Impiombato, and D. di Bernardo. How to infer gene networks from expression profiles. *Molecular Systems Biology*, 3(78):1–10, 2007.
- [4] D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, and G. Stolovitzky. Revealing strengths and weaknesses of methods for gene network inference. *PNAS*, 107:6286–6291, 2010.
- [5] R. De Smet and K. Marchal. Advantages and limitations of current network inference methods. *Nature Reviews Microbiology*, 8:717–729, 2010.
- [6] D. Marbach, J. C. Costello, R. Kuffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison, The DREAM5 Consortium, M. Kellis, J. J. Collins, and G. Stolovitzky. Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8):796–804, 2002.
- [7] G. Michailidis and F. d’Alché-Buc. Autoregressive models for gene regulatory network inference: sparsity, stability and causality issues. *Mathematical Biosciences*, 246:326–334, 2013.
- [8] M. K. S. Yeung, J. Tegnér, and J. J. Collins. Reverse engineering gene networks using singular value decomposition and robust regression. *PNAS*, 99(9):6163–6168, 2002.
- [9] T. S. Gardner, D. di Bernardo, D. Lorenz, and J. J. Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 301:102–105, 2003.
- [10] M. Bansal, G. Della Gatta, and D. di Bernardo. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics*, 22:815–822, 2006.
- [11] D. Napolitano and T. D. Sauer. Reconstructing the topology of sparsely connected dynamical networks. *Phys. Rev. E*, 77:026103, Feb 2008.
- [12] E. August and A. Papachristodoulou. Efficient, sparse biological network determination. *BMC Systems Biology*, 3(25), 2009.
- [13] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin. Exact topology identification of large-scale interconnected dynamical systems from compressive observations. In *Proc. American Control Conference (ACC’11)*, page 649656, San Francisco, California, June 2011.
- [14] Y. H. Chang and C. Tomlin. Inference of temporally evolving network dynamics with applications in biological systems. In *Proc. IEEE Conference on Decision and Control (CDC’11)*, 2011.
- [15] A. Bolstad, B. D. Van Veen, and R. Novak. Causal network inference via group sparse regularization. *Signal Processing, IEEE Transactions on*, 59(6):2628–2641, 2011.

- [16] A. J. Seneviratne and V. Solo. Topology identification of a sparse dynamic network. In *Proc. IEEE Conference on Decision and Control (CDC'12)*, 2012.
- [17] A. Chiuso and G. Pillonetto. A bayesian approach to sparse dynamic network identification. *Automatica*, 48(8):1553 – 1565, 2012.
- [18] D. Materassi, G. Innocenti, L. Giarr, and M. Salapaka. Model identification of a network as compressing sensing. *Systems & Control Letters*, 62(8):664 – 672, 2013.
- [19] Y.C. Eldar, P. Kuppinger, and H. Bolcskei. Block-sparse signals: Uncertainty relations and efficient recovery. *Signal Processing, IEEE Transactions on*, 58(6):3042–3054, June 2010.
- [20] E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23:969–985, 2007.
- [21] J. Gonçalves and S. Warnick. Necessary and sufficient conditions for dynamical structure reconstruction of LTI networks. *IEEE Trans. Automat. Contr.*, 53(7):1670–1674, 2008.
- [22] D. Hayden, Y. Yuan, and J. Gonçalves. Network reconstruction using knock-out and over-expression data. In *Proc. European Control Conference (ECC'13)*, Zürich, July 2013.
- [23] D. L. Donoho. For most large underdetermined systems of linear equations the minimal l_1 -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, 2006.
- [24] R. von Borries, C. J. Miosso, and C. Potes. Compressed sensing using prior information. In *Proc. 2nd IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process*, page 121124, 2007.
- [25] J. Scarlett, J. S. Evans, and S. Dey. Compressed sensing with prior information: information-theoretic limits and practical decoders. *IEEE Transactions on Signal Processing*, 61(2), 2013.
- [26] J. Tegnér, M. K. S. Yeung, J. Hasty, and J. J. Collins. Reverse engineering gene networks: integrating genetic perturbations with dynamic modeling. *PNAS*, 100(10):5944–5949, 2003.
- [27] T. E. Ideker, V. Thorsson, and R. Karp. Discovery of regulatory interactions through perturbation: inference and experimental design. *Pacific Symposium on Biocomputing*, 5:302–313, 2000.
- [28] E. Sontag, A. Kiyatkin, and B. N. Kholodenko. Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics*, 20(12):1877–1886, 2004.