

Exact Reconstruction of Gene Regulatory Networks using Compressive Sensing

Young Hwan Chang¹, Joe W. Gray² and Claire Tomlin^{1,3,*}

¹ Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA

²Department of Biomedical Engineering and the Center for Spatial Systems Biomedicine, Oregon Health and Science University, Portland, OR, USA

³ Faculty Scientist, Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA

* E-mail: tomlin@eecs.berkeley.edu

Abstract

Background: We consider the problem of reconstructing a gene regulatory network structure from limited time series gene expression data, without any a priori knowledge of connectivity. We assume that the network is sparse, meaning the connectivity among genes is much less than full connectivity. We develop a method for network reconstruction based on compressive sensing, which takes advantage of the network's sparseness.

Results: For the case in which all genes are accessible for measurement, and there is no measurement noise, we show that our method can be used to exactly reconstruct the network. For the more general problem, in which hidden genes exist and all measurements are contaminated by noise, we show that our method leads to reliable reconstruction. In both cases, coherence of the model is used to assess the ability to reconstruct the network and to design new experiments. For each problem, a set of numerical examples is presented.

Conclusions: The method provides a guarantee on how well the inferred graph structure represents the underlying system, reveals deficiencies in the data and model, and suggests experimental directions to remedy the deficiencies.

Introduction

Mathematical modeling of biological signaling pathways can provide an intuitive understanding of their behavior [1] [2] [3]. However, since typically only incomplete knowledge of the network structure exists and the system dynamics is known to be sufficiently complex, the challenge has become to show that the identified networks and corresponding mathematical models are enough to adequately represent the underlying system. In the last years, many data-driven mathematical tools have been developed and applied to reconstruct graph representations of gene regulatory networks (GRNs) from data. These include Bayesian networks, regression, correlation, mutual information and system-based approaches [4] [5] [6] [7] [8] [9] [10]. Also, these approaches either focus on static or on time series data. The latter approach has the advantage of being able to identify dynamic relationships between genes.

However, data-driven reconstruction of the network structure itself remains in general a difficult problem; nonlinearities in the system dynamics and measurement noise make this problem even more challenging. For linear time invariant (LTI) systems, there exist necessary and sufficient conditions for network reconstruction [11]. However, for time-varying or nonlinear systems, there has not been as yet any statistical guarantee on how well the inferred model represents the underlying system [12] [13] [14] [15]. The recent work [16] addresses the problem of data-driven network reconstruction, together with measurement noise and unmodelled nonlinear dynamics, yet this work points out that these complications impose a limit on the reconstruction, and with strong nonlinear terms the method fails. Additionally, identifying whether important nodes in the graph

structure are missing, how many are missing, and where these nodes are in the interconnection structure remains a challenging problem.

Moreover, in order to continue to have an impact in systems biology, identification of the graph topology from data should be able to reveal deficiencies in the model and suggest new experimental directions [17]. For example, Steinke *et al.* [18] presented a Bayesian method for identifying a gene regulatory network from micro-array measurements in perturbation experiments and showed how to use optimal design to minimize the number of measurements.

Since biological regulatory networks are known to be sparse, meaning that most proteins interact with only a small number of proteins compared with the total number in the network, many methods [12] [13] [14] [15] [19] [20] [21] [22] take advantage of the sparsity. The methods typically use l_1 -norm optimization, which leads to a sparse representation of the network and improves the ability to find the actual network structure. Even though many methods [14] [15] show that the reconstruction results are fairly good, the methods cannot guarantee *exact recovery*. This stems from the fact that the so-called *incoherence* condition is typically not satisfied for the matrix Ω in a linear measurement model $\mathbf{Y} = \Omega\mathbf{q}$, where \mathbf{q} is the signal to be reconstructed, \mathbf{Y} is the measurement, and Ω is known as the sensing matrix. Roughly, incoherence is a measure of the correlation between columns of the sensing matrix. Since the incoherence condition of the sensing matrix provides a metric of performance, this is one of the motivating factors for the use of compressive sensing (CS) [23]. CS is a signal processing technique for efficiently acquiring and reconstructing a signal by taking advantage of the signal's sparsity and allowing the entire signal to be determined from relatively few measurements under a certain condition, i.e., it requires that the incoherence condition to be satisfied. In the Human Epidermal Growth Factor Receptor2 (HER2) positive breast cancer signaling pathway that we studied in [24] [25], time series data sets consist of only 8 time point measurements of 20 protein signals, and we would like to use this limited data to identify a graph structure which could have 20×20 or 400 edges.

In this paper, we are interested in directed graph representations of signaling pathways. We develop a new algorithm for GRN reconstruction based on CS. First, we focus on sparse graph structures using limited time series data with all nodes accessible and no measurement noise. We test the network reconstruction algorithm on a simulated biological pathway in which the structure is known *a priori*. We explain the limitation of the proposed method's performance when the dataset naturally has high coherence and propose a way to overcome this limitation by designing additional effective experiments.

Next, the proposed algorithm is extended to a more general problem: we consider partially corrupted data with data inconsistencies due to model mismatch, and measurement noise affecting all the data. Typically, data inconsistencies may result from missing nodes in the model; or in some cases arbitrary data corruption may result from human errors such as mislabeling or the improper use of markers or antibodies. The question is whether one can still recover the graph structure reliably under these conditions. Inspired by a robust error correction method [26], the exact recovery of the graph structure can be guaranteed under suitable conditions on the node dynamics, provided that hidden nodes can affect relatively few nodes in the graph structure. Also, a set of numerical examples is provided to demonstrate the method, including some from an RPPA (Reverse Phase Protein Array) dataset [27] collected from HER2 positive breast cancer cell lines.

In this paper, the main contributions are the following.

- The CS framework uses the coherence of the sensing matrix as a performance index, which allows us to assess and optimize mathematically network reconstruction.
- Coherence also provides a guideline for optimizing experiment design for network reconstruction.

tion.

- By utilizing an error correction method in conjunction with the CS framework, network reconstruction may be performed even when there are hidden nodes and measurement noise.

Background

Overview: Compressive Sensing

Consider measurements $\mathbf{Y} \in \mathbb{R}^m$ of a signal $\mathbf{q} \in \mathbb{R}^n$:

$$\mathbf{Y} = \mathbf{\Omega}\mathbf{q} \quad (1)$$

where $\mathbf{\Omega} \in \mathbb{R}^{m \times n}$ is called the sensing matrix.

One key question [28] is how many measurements m are needed to exactly recover the original signal \mathbf{q} from $\mathbf{\Omega}$:

- If $m > n$ and $\mathbf{\Omega}$ is a full rank matrix, then the problem is overdetermined. If $m = n$ and $\mathbf{\Omega}$ is a full rank matrix, the problem is determined and may be solved uniquely for \mathbf{q} .
- If $m < n$, the problem is underdetermined even if $\mathbf{\Omega}$ has full rank. We can restrict $\mathbf{q} \in \mathbb{R}^n$ to the subspace which satisfies $\mathbf{Y} = \mathbf{\Omega}\mathbf{q}$. However, \mathbf{q} cannot be determined uniquely.

For the underdetermined case, the least squares solution $\mathbf{q}^* = \arg \min_{\mathbf{q}} \|\mathbf{q}\|_{l_2} = \mathbf{\Omega}^*(\mathbf{\Omega}\mathbf{\Omega}^*)^{-1}\mathbf{Y}$ is typically used as the “best guess” in many applications. However, if \mathbf{q} is known to be sparse, meaning that many of its components are zero, one might expect that fewer than n measurements are needed to recover it. It is thus of interest to obtain a good estimator for underdetermined problems in which \mathbf{q} is assumed to be *s-sparse*, meaning that s of the elements of \mathbf{q} are nonzero. In principle, the theory of compressive sensing (CS) asserts that the number of measurements needed to recover $\mathbf{q} \in \mathbb{R}^n$ is a number proportional to the compressed size of the signal (s), rather than the uncompressed size (n) [29]. To be able to recover \mathbf{q} , CS relies on two properties: *sparsity*, which pertains to the signals of interest, and *incoherence*, which pertains to the sensing matrix.

Proposition 1 [28]. *Suppose that any $2s$ columns of an $m \times n$ matrix $\mathbf{\Omega}$ are linearly independent (this is a reasonable assumption if $m \geq 2s$). Then, any s -sparse signal $\mathbf{q} \in \mathbb{R}^n$ can be reconstructed uniquely from $\mathbf{\Omega}\mathbf{q}$.*

The proof [28] of the above proposition also shows how to reconstruct an *s-sparse* signal $\mathbf{q} \in \mathbb{R}^n$ from the measurement $\mathbf{Y} = \mathbf{\Omega}\mathbf{q}$ where \mathbf{q} is the unique sparsest solution to $\mathbf{Y} = \mathbf{\Omega}\mathbf{q}$:

$$\mathbf{q}^* = \arg \min_{\mathbf{q}} \|\mathbf{q}\|_{l_0} \quad \text{subject to } \mathbf{Y} = \mathbf{\Omega}\mathbf{q} \quad (2)$$

and $\|\mathbf{q}\|_{l_0} := \sum_{i=1}^n I(\mathbf{q}_i \neq 0)$ is the cardinality of \mathbf{q} . However, the l_0 -minimization is computationally intractable (NP-hard in general). Recent breakthroughs enable approximating l_0 -optimization by using l_1 -minimization which is a convex optimization problem and can be solved in a simple but effective way by linear programming:

$$\mathbf{q}^* = \arg \min_{\mathbf{q}} \|\mathbf{q}\|_{l_1} \quad \text{subject to } \mathbf{Y} = \mathbf{\Omega}\mathbf{q} \quad (3)$$

The l_1 -minimization (3) requires mild oversampling, more specifically, $m \geq c \cdot \mu(\Phi, \Psi)^2 \cdot s \log n$ for some positive constant c where we have m measurement in the Φ domain under bases Ψ and μ represents the coherence defined as follows:

Definition 1 [23]. The coherence of a matrix Ω is given by

$$\mu(\Omega) = \max_{j < k} \frac{|\langle \Omega_j, \Omega_k \rangle|}{\|\Omega_j\|_2 \|\Omega_k\|_2}$$

where Ω_j and Ω_k denote columns of Ω .

Several theoretical results [30] ensure that the l_1 -minimization guarantees exact recovery whenever the sensing matrix Ω is sufficiently *incoherent*. For example, we can say that Ω is *incoherent* if μ is small. Coherence is a key property in the compressed sensing framework because if two columns are closely correlated, it would be very difficult to distinguish whether the influence from the components in the sparse signal comes from one or the other (recall that the measurement $\mathbf{Y} (= \Omega \mathbf{q})$ is a linear combination of each columns of the sensing matrix Ω with the components in \mathbf{q} as coefficients). Also, numerical experiments suggest that in practice, most s -sparse signals are in fact recovered exactly once $m \geq 4s$ [23]. Here, “exact recovery” means that we find the sparsest solution (\mathbf{q}) such that $\mathbf{Y} = \Omega \mathbf{q}$.

Therefore, if the sensing matrix Ω satisfies the incoherence condition ($m \geq c \cdot \mu(\Phi, \Psi)^2 \cdot s \log n$, or in practice, $m \geq 4s$), a sufficiently sparse signal \mathbf{q} can be exactly recovered from the limited dataset without any prior knowledge of the number of nonzero elements, their locations, and their values. On the other hand, if the condition is not satisfied, exact recovery cannot be guaranteed [29] [17]. However, it is possible to use the property of coherence to guide biological experiment design, basically to collect a more informative dataset. As we will discuss in this paper, this can be done by inhibiting or stimulating certain genes to manipulate the gene expression.

CS can help reconstruct GRNs

In graph theory, a digraph can be represented by $G = (V, E)$ where V and E represent nodes and edges respectively. For GRNs, each node represents a gene and each edge represents an influence map which models how genes affect each other. For example, the interactions could be how genes inhibit or stimulate each other. Since the connectivities of GRNs are typically unknown, often the best we can do is to select a set of possible candidate functions encoding possible unknown connectivities between genes.

In this section, we formulate a data-driven network identification problem into CS framework: first, we define a dynamical model of gene regulatory network. Then, we encode system dynamics into the sensing matrix (Ω) and denote unknown connectivities between genes by \mathbf{q} , a signal to be recovered.

A Dynamical Model of Gene Regulatory Networks (GRNs)

We consider a dynamical system described by:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{u} + \mathbf{g}(\mathbf{x}_h) + \mathbf{w} \quad (4)$$

where $\mathbf{x} \in \mathbb{R}^n$ denotes the concentrations of the rate-limiting species which can be measured in experiments; $\dot{\mathbf{x}} = [\dot{x}_1 \ \dot{x}_2 \ \dots \ \dot{x}_n]^\top \in \mathbb{R}^n$ is a vector whose elements are the change in concentrations of the n species over time; $\mathbf{f}(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ represents biochemical reactions, such as those governed by mass action kinetics, Michaelis-Menten, or Hill kinetics. Thus, $\mathbf{f}(\cdot)$ can include functions of known form such as product of monomials, monotonically increasing or decreasing Hill

functions, simple linear terms, and constant terms [19]. $\mathbf{u} \in \mathbb{R}^n$ denotes the control input which could represent inhibitions and stimulations; $\mathbf{g}(\cdot) : \mathbb{R}^{n_h} \rightarrow \mathbb{R}^n$ represents influence from hidden nodes $\mathbf{x}_h \in \mathbb{R}^{n_h}$, which cannot be measured in experiments; n_h is the number of hidden nodes and unknown; and $\mathbf{w} \in \mathbb{R}^n$ represents energy-bounded process noise or measurement noise. Here $\mathbf{x}, \dot{\mathbf{x}}$ which can be measured in experiments, and \mathbf{u} is assumed to be known.

Since we do not know whether important nodes in the gene regulatory network are missing, how many missing nodes there are, and how they affect system dynamics (i.e., \mathbf{x}_h, n_h and $\mathbf{g}(\mathbf{x}_h)$), we denote a vector of (unknown) influence from hidden nodes' dynamics by $\mathbf{v} (\triangleq \mathbf{g}(\mathbf{x}_h))$; Also, without loss of generality, since we know \mathbf{u} , we define \mathbf{y} as follows:

$$\mathbf{y} \triangleq \dot{\mathbf{x}} - \mathbf{u} = \mathbf{f}(\mathbf{x}) + \mathbf{v} + \mathbf{w} \quad (5)$$

Formulating a Dynamical System as a GRN

The nonlinear function $\mathbf{f}(\mathbf{x})$ can be decomposed into a linear sum of scalar basis functions, $f_{b,i}(\mathbf{x}) \in \mathbb{R}$, where we select the set of possible candidate basis functions:

$$\mathbf{f}(\mathbf{x}) = \sum_{i=1}^N f_{b,i}(\mathbf{x}) \begin{bmatrix} q_{i1} \\ q_{i2} \\ \dots \\ q_{in} \end{bmatrix} \quad (6)$$

where N is the number of possible candidate basis functions and q_{ij} are unknown parameters which reflect underlying structure, i.e., influence of $f_{b,i}(\mathbf{x})$ on the j -th state (\dot{x}_j). Typically, we may choose a larger set than necessary, and allow the CS method to indicate the importance of each function, as we shall describe. Thus the equation (5) can be written as follows:

$$\mathbf{y} = \mathbf{S}_q \mathbf{F}_b(\mathbf{x}) + \mathbf{v} + \mathbf{w} \quad (7)$$

where $\mathbf{S}_q = \begin{bmatrix} q_{11} & \dots & q_{N1} \\ \dots & \dots & \dots \\ q_{1n} & \dots & q_{Nn} \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1^\top \\ \dots \\ \mathbf{q}_n^\top \end{bmatrix} \in \mathbb{R}^{n \times N}$ reflects the underlying GRN structure, $\mathbf{q}_i = [q_{1i} \ q_{2i} \ \dots \ q_{Ni}]^\top \in \mathbb{R}^N$ and $\mathbf{F}_b(\mathbf{x}) = [f_{b,1}(\mathbf{x}), \dots, f_{b,N}(\mathbf{x})]^\top \in \mathbb{R}^N$ is the vector field which includes possible candidate basis functions. In this way, any biochemical reactions can be represented by a linear map \mathbf{S}_q and a function $\mathbf{F}_b(\mathbf{x})$ where \mathbf{S}_q encodes the underlying graph structure and $\mathbf{F}_b(\mathbf{x})$ includes all possible candidate functions that could be included in the biochemical reactions.

In practice, we can construct $\mathbf{F}_b(\mathbf{x})$ by selecting the most commonly used candidate basis functions to model GRNs, for example, all monomials, binomials, other combinations and Hill function.

Example 1. Consider the simple nonlinear ordinary differential equations (ODEs):

$$\begin{aligned} y_1 &= \dot{x}_1 = \gamma_1 x_1 + k_1 \frac{1}{1 + x_2^{n_{act}}} \\ y_2 &= \dot{x}_2 = \gamma_2 x_2 + k_2 \frac{x_3^{n_{inh}}}{1 + x_3^{n_{inh}}} \\ y_3 &= \dot{x}_3 = \gamma_3 x_3 + k_3 \frac{1}{1 + x_1^{n_{act}}} \end{aligned} \quad (8)$$

where x_i denotes the concentration of the i -th species, \dot{x}_i is the change in concentration of the i -th species, the γ_i denotes protein decay rate, the k_i denotes the maximum promoter/inhibitor strength. Here, there is no input ($\mathbf{u} = \mathbf{0}$), no hidden node ($\mathbf{v} = \mathbf{0}$) and no process noise ($\mathbf{w} = \mathbf{0}$). Also, n_{act} represents positively cooperative binding (activation) and n_{inh} represents negative cooperative binding (inhibition). The set of ODEs corresponds to a topology where gene 1 is activated by gene 2, gene 2 is inhibited by gene 3, and gene 3 is activated by gene 1 as shown in Figure 1 (a). We can write (8) as follows:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \underbrace{\begin{bmatrix} \gamma_1 & 0 & 0 & 0 & k_2 & 0 \\ 0 & \gamma_2 & 0 & 0 & 0 & k_3 \\ 0 & 0 & \gamma_3 & k_1 & 0 & 0 \end{bmatrix}}_{\mathbf{S}_q} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \frac{1}{1+x_1^{n_{act}}} \\ \frac{1}{1+x_2^{n_{act}}} \\ \frac{x_3^{n_{inh}}}{1+x_3^{n_{inh}}} \end{bmatrix} = \mathbf{S}_q \mathbf{F}_b(\mathbf{x})$$

where $\mathbf{S}_q \in \mathbb{R}^{3 \times 6}$ represents the influence map.

We can also consider a version of (8) in which there exists a hidden node (x_h) affecting (x_1) as shown in Figure 1(b), as well as process noise.

$$\mathbf{y} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \mathbf{S}_q \mathbf{F}_b(\mathbf{x}) + \mathbf{v} + \mathbf{w}$$

where $\mathbf{v} = [v_1, 0, 0]^\top$ and $\mathbf{w} = [w_1, w_2, w_3]^\top$

Formulating GRN into the CS framework

Suppose the time series data are sampled from a real experimental system at discrete time points t_k . By taking the transpose of both sides of equation (7), we obtain

$$\mathbf{y}(t_k)^\top = \mathbf{F}_b^\top(\mathbf{x}(t_k))\mathbf{q} + \mathbf{v}(t_k)^\top + \mathbf{w}(t_k)^\top \quad (9)$$

where $\mathbf{q} = \mathbf{S}_q^\top = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \dots \quad \mathbf{q}_n] \in \mathbb{R}^{N \times n}$. Assuming that M successive data points are sampled, then define:

$$\begin{aligned} \mathbf{y} \triangleq [\mathbf{y}_1 \quad \dots \quad \mathbf{y}_n] &= \begin{bmatrix} y_1(t_1) & \dots & y_n(t_1) \\ y_1(t_2) & \dots & y_n(t_2) \\ \dots & \dots & \dots \\ y_1(t_M) & \dots & y_n(t_M) \end{bmatrix} \in \mathbb{R}^{M \times n} \\ \Phi \triangleq \begin{bmatrix} f_{b,1}(\mathbf{x}(t_1)) & \dots & f_{b,N}(\mathbf{x}(t_1)) \\ f_{b,1}(\mathbf{x}(t_2)) & \dots & f_{b,N}(\mathbf{x}(t_2)) \\ \dots & \dots & \dots \\ f_{b,1}(\mathbf{x}(t_M)) & \dots & f_{b,N}(\mathbf{x}(t_M)) \end{bmatrix} &= \begin{bmatrix} \mathbf{F}_b^\top(\mathbf{x}(t_1)) \\ \mathbf{F}_b^\top(\mathbf{x}(t_2)) \\ \dots \\ \mathbf{F}_b^\top(\mathbf{x}(t_M)) \end{bmatrix} \in \mathbb{R}^{M \times N} \\ [\mathbf{v}_1 \quad \dots \quad \mathbf{v}_n] \triangleq \begin{bmatrix} \mathbf{v}(t_1)^\top \\ \mathbf{v}(t_2)^\top \\ \dots \\ \mathbf{v}(t_M)^\top \end{bmatrix} &\in \mathbb{R}^{M \times n} \\ [\mathbf{w}_1 \quad \dots \quad \mathbf{w}_n] \triangleq \begin{bmatrix} \mathbf{w}(t_1)^\top \\ \mathbf{w}(t_2)^\top \\ \dots \\ \mathbf{w}(t_M)^\top \end{bmatrix} &\in \mathbb{R}^{M \times n} \end{aligned} \quad (10)$$

This leads to n independent equations:

$$\mathbf{y}_i = \Phi \mathbf{q}_i + \mathbf{v}_i + \mathbf{w}_i \quad (i = 1, \dots, n) \quad (11)$$

where $\mathbf{y}_i = [y_i(t_1), y_i(t_2), \dots, y_i(t_M)]^\top \in \mathbb{R}^M$ represents the M successive data points, $\Phi \in \mathbb{R}^{M \times N}$ consists of N possible candidate bases which are functions of given time series data \mathbf{x} and $\mathbf{q}_i \in \mathbb{R}^N$ represents the unknown influence map corresponding to the i -th species. Since a biochemical reaction network is typically sparse, as a consequence, \mathbf{q}_i is sparse and we have $N \gg M$ for Φ because we assume the limited time series data and may choose a larger set of basis functions than necessary.

Although we formulate n independent linear regression problems (11), we consider n independent equations in (11) together by stacking \mathbf{y}_i as follows:

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \dots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \Phi & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \Phi & \dots & \mathbf{O} \\ \dots & \dots & \dots & \dots \\ \mathbf{O} & \mathbf{O} & \dots & \Phi \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \dots \\ \mathbf{v}_n \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \dots \\ \mathbf{w}_n \end{bmatrix} \\ &\triangleq \Omega \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \mathbf{v} + \mathbf{w} \end{aligned} \quad (12)$$

Now, equation (12) is in the form of the CS formulation in (1) where $\mathbf{Y} \in \mathbb{R}^{n \cdot M}$ is the measurement, $\Omega \in \mathbb{R}^{n \cdot M \times n \cdot N}$ is the sensing matrix which consists of basis functions for the given time series data, $\mathbf{v} \triangleq [\mathbf{v}_1; \mathbf{v}_2; \dots; \mathbf{v}_n] \in \mathbb{R}^{n \cdot M}$ represents possibly large corruption from hidden nodes and $\mathbf{w} \triangleq [\mathbf{w}_1; \mathbf{w}_2; \dots; \mathbf{w}_n] \in \mathbb{R}^{n \cdot M}$ represents process or measurement noise.

In this paper, we make the following assumptions:

Assumption 1. We consider two cases:

- (ideal case) We assume that we can measure all states \mathbf{x} (i.e., there is no hidden node (\mathbf{x}_h)) and there is no measurement noise ($\mathbf{w} = \mathbf{0}$). Also, there are enough columns in Φ in (11) to represent the underlying system (6).
- (extension) We consider hidden node (\mathbf{x}_h) where $\mathbf{v}(= \mathbf{g}(\mathbf{x}_h))$ is assumed to be sparse, and process noise ($\mathbf{w} \neq 0$). Here, we can also consider the case where the columns (also known as the dictionaries in the CS literature [23]) of Φ may not be able to represent the underlying system. In this case, we consider the influence from the missing dictionaries as \mathbf{v} .

We first consider the ideal case for simplicity in explaining the main results, and then extend the proposed method to the more general case.

Results and Discussion:

1. Formulating GRN Identification Problem into CS

Many existing algorithms [16] [19] consider the n independent linear regression problems (11) separately. Since the columns of matrix Φ are composed of time series data as in (10), it is difficult to *a priori* guarantee low correlation and sometimes Φ even suffers rank deficiency. Also, Pan *et al.* [19] pointed out that correlation between the columns of Φ is usually high ($\mu(\Phi)$ is close to 1).

Intuitively, if two columns of the sensing matrix are highly correlated, it is hard to distinguish the corresponding components in the sparse signal \mathbf{q} (note that the measurement \mathbf{Y} is a linear combination of each column of the sensing matrix with the components in \mathbf{q} as coefficients). In order to deal with high coherence in (11), many methods combine CS with different techniques such as Bayesian formulation [19], Kalman filter [20], and Granger causality [21]. Also, it is a well known problem in the lasso formulation of network inference: if there is high coherence in the sensing matrix, one can use an elastic net which combines the l_1 and l_2 norms. Although each reconstruction result [19] [20] [21] might be the optimal solution in the sense of its formulation (i.e., maximum likelihood), the identified graph may not represent the underlying GRN. In other words, if the data set is not informative enough to fully explore the underlying system, while the identified graph structure based on the given data set may be an optimal solution of the particular optimization problem, it may not represent the true system.

In this paper, our goal is to get the smallest data informative enough to recover the underlying graph structure exactly. Since the proposed method maintains the CS framework by reducing coherence of the sensing matrix, the method is fundamentally different than any other methods which make use of different techniques in conjunction with the l_1 optimization [19] [20] [21] which leads to a sparse representation of the network. Hence, we use all the properties of CS in order to access the ability to exactly reconstruct the underlying graph structure, reveal deficiencies in the data and model, and design new experiments to remedy the deficiencies if necessary.

While maintaining the CS framework, in order to deal with high coherence, we formulate (12) instead of (11) and we have strongly uncorrelated columns in Ω . In other words, since Ω has many independent columns, we have more degrees of freedom to reduce coherence of Ω . We will show that by using a transformation, the components of the sensing matrix can be made more uniformly distributed so that we could reduce coherence.

Moreover, since each \mathbf{q}_i has different degrees of sparsity in general, if we consider n independent equations in (11), Φ should satisfy the incoherence condition $M \geq 2 \max_i(\|\mathbf{q}_i\|_0)$ stated in **Proposition 1**. On the other hand, Ω only needs to satisfy the condition $M \cdot n \geq 2 \cdot \sum_{i=1}^n \|\mathbf{q}_i\|_{l_0}$. Since

the averaged sparsity ($= \frac{\sum_{i=1}^n \|\mathbf{q}_i\|_{l_0}}{n} \leq \max_i(\|\mathbf{q}_i\|_{l_0})$) is smaller than $\max_i(\|\mathbf{q}_i\|_{l_0})$, we can reduce the required number of samples (M). Also, in case of rank deficiency, we can simply remove the corresponding rows (note that the rank deficiency is more likely caused by the row since $N \gg M$).

2. Optimal Design of Sensing Matrix

2.1. Reducing Coherence by Transformation

In order to reduce coherence, first we rearrange Ω with respect to a spatial information, and then we consider the transformation in order to reduce coherence (see **Method: 1.1. Rearranging the sensing matrix** for mathematical details):

$$\mathbf{Z} \triangleq \begin{bmatrix} \bar{\mathbf{z}}(1) \\ \bar{\mathbf{z}}(2) \\ \dots \\ \bar{\mathbf{z}}(M) \end{bmatrix} = \begin{bmatrix} \Psi_1^s(\mathbf{I}_n \otimes \Phi_{1,:}) \\ \Psi_2^s(\mathbf{I}_n \otimes \Phi_{2,:}) \\ \dots \\ \Psi_M^s(\mathbf{I}_n \otimes \Phi_{M,:}) \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{v}}(1) \\ \bar{\mathbf{v}}(2) \\ \dots \\ \bar{\mathbf{v}}(M) \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{w}}(1) \\ \bar{\mathbf{w}}(2) \\ \dots \\ \bar{\mathbf{w}}(M) \end{bmatrix} \triangleq \bar{\Omega} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \bar{\mathbf{v}} + \bar{\mathbf{w}} \quad (13)$$

where \otimes represents the Kronecker product¹, $\bar{\Omega} \in \mathbb{R}^{M \cdot n \times N \cdot n}$ represents the rearranged sensing matrix multiplied by the transformation Ψ_j^s , and $\bar{\mathbf{z}}(j)$, $\bar{\mathbf{v}}$ and $\bar{\mathbf{w}}$ are defined in (20). We want to find Ψ_j^s ($j = 1, \dots, M$) to minimize

$$\min \mu(\bar{\Omega}) \quad \text{where } \bar{\Omega} = \begin{bmatrix} \psi_{11}^s \Phi_{1,:} & \psi_{12}^s \Phi_{1,:} & \dots & \psi_{1n}^s \Phi_{1,:} \\ \psi_{21}^s \Phi_{2,:} & \psi_{22}^s \Phi_{2,:} & \dots & \psi_{2n}^s \Phi_{2,:} \\ \dots & \dots & \dots & \dots \\ \psi_{M1}^s \Phi_{M,:} & \psi_{M2}^s \Phi_{M,:} & \dots & \psi_{Mn}^s \Phi_{M,:} \end{bmatrix} \quad \text{has full row rank} \quad (14)$$

where $\psi_{ij}^s \in \mathbb{R}^n$ represents the j -th column of Ψ_i^s . In this paper, we propose a heuristic approach and a novel way to find Ψ_i^s by solving the optimization iteratively to reduce coherence (see **Method: Section 1.2 and 1.3** for details).

Example 2. (*reducing coherence for a linear system*) consider a simple linear system ($\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$) where $n = 5, N = 5, M = 4, s = 10$ (note that $n \cdot M \geq 2s$), and the elements of \mathbf{A} are randomly chosen such that there is no isolated node. Figure 2 shows the sensing matrix for both Ω (top left) and the transformed sensing matrix (top right, denoted by Ω_Ψ). By reducing coherence, the components of the transformed sensing matrix (top right) are more uniformly distributed and the coherence is reduced by up to 0.6 although $\mu(\Omega)$ is close to 1 (bottom left) in Figure 2(a). Also, Figure 2(b) shows the result of the inferred graph structure based on given time series data without any a priori information where the x -axis represents indices of the influence map (i.e., the 1st, 2nd, ..., n th rows of influence map \mathbf{S}_q ; note that for a linear system, $\mathbf{S}_q = \mathbf{A}$). Here, there are 5 states ($n = 5$) in a linear system so the influence map \mathbf{A} has 25 elements. Although $L1$ and $L2^2$ norm minimizations fail to recover the exact signal, CS in (17) (see **3. Recovery of**

¹If \mathbf{A} is an $m \times n$ matrix and \mathbf{B} is a $p \times q$ matrix, then the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the $mp \times nq$ block matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{bmatrix}$$

²We compare the performance of CS with the performance of the $L1$ and $L2$ optimization as follows:

$$\min_{\mathbf{q}_i} \|\mathbf{q}_i\|_1 \quad \text{subject to} \quad \mathbf{y}_i = \Phi \mathbf{q}_i \quad (i = 1, \dots, n) \quad (L1)$$

$$\min_{\mathbf{q}} \|\mathbf{q}\|_2 \quad \text{subject to} \quad \mathbf{Y} = \Omega \mathbf{q} \quad (L2)$$

Gene Regulatory Networks for details) recovers the exact signal (bottom right) by first reducing coherence of the sensing matrix. Note that (L1) solves the n independent equations (11) without reducing coherence and (L2) solves (12) with l_2 -regularization.

Example 3. (reducing coherence for a nonlinear system, $n = 3$, $N = 9$, $M = 5$, $s = 6$) consider simple nonlinear ODEs as follows:

$$\begin{aligned}\dot{x}_1 &= \gamma_1 x_1 + k_{+12} \frac{1}{1 + x_2^{n_{act}}} \\ \dot{x}_2 &= \gamma_2 x_2 + k_{-23} \frac{x_3^{n_{ihb}}}{1 + x_3^{n_{ihb}}} \\ \dot{x}_3 &= \gamma_3 x_3 + k_{+13} \frac{1}{1 + x_1^{n_{act}}}\end{aligned}\tag{15}$$

where $(\gamma_1, \gamma_2, \gamma_3) = (-0.3, -0.25, -0.35)$, $k_{+12} = 1.2$, $k_{+13} = 0.9$, $k_{-23} = 2.2$, $n_{act} = 4$ (activation Hill coefficient) and $n_{ihb} = 4$ (inhibition Hill coefficient). The set of ODEs corresponds to a topology $(x_2 \rightarrow x_1 \rightarrow x_3 \dashv x_2)$ as shown in Figure 3 (a). Figure 3(b) shows that we can reduce the coherence by up to 0.8 and (c) shows that only CS recovers the exact graph structure, and L2 regularization does not encourage sparsity but distributes the coefficients to be more similar to each other.

2.2. Designing Effective Experiments

Consider the case where the sensing matrix is not incoherent. If the coherence condition ($M \cdot n > c\mu(\bar{\Omega})^2 \cdot s \log(N \cdot n)$) is not satisfied, exact recovery cannot be guaranteed [29]. We use the transformation (Ψ_i^s or $\mathbf{P} = \mathcal{P}^\top \mathcal{P}$ in (24); see **Method section 1. Reducing Coherence by Transformation**) in order to reduce the coherence but obviously, sometimes we might have inherent limits to how much the coherence can be reduced. There are possible reasons:

- Since we solve the relaxed problem (24) iteratively, \mathbf{P} might be sub-optimal.
- If the time series data of two different gene expressions, \mathbf{x}_i and \mathbf{x}_j are highly correlated, it might be difficult to reduce coherence. In this case, we need to design a new experiment to remedy deficiencies in the data.

As we mentioned, the incoherence of the sensing matrix can be used not only as a good metric to guarantee exact recovery but also as a guideline for designing new experiments. For example, from the coherence distribution, we can identify which columns of the sensing matrix have high coherence, i.e., $f_{b,i}(\mathbf{x})$ and $f_{b,j}(\mathbf{x})$ in (6). Intuitively, in order to reduce ambiguities from the highly correlated columns of the sensing matrix, we should perturb either $f_{b,i}(\mathbf{x})$ or $f_{b,j}(\mathbf{x})$. Thus, it is possible to use this property of coherence to guide biological experiment design, to collect a more informative dataset.

Example 4. (limitation of reducing coherence by Ψ , linear dynamics, $n = 5$, $N = 5$, $M = 4$, $s = 10$) In Figure 4, Exp#1 represents the original experimental data set which has the limitation of reducing coherence by Ψ . Since we consider linear dynamics, i.e., $f_{b,i}(x) = x_i$ and $f_{b,j}(x) = x_j$, we found that x_2 and x_4 cause high coherence as shown in Figure 4 (circle marker). Thus, in order to reduce this high coherence, we should perturb either x_2 or x_4 .

To show the effectiveness of the new experiment, we design two different experiment sets and compare the reconstruction results with each other; for Exp#2, we perturb x_3 and for Exp#3, we perturb x_2 . As we mentioned earlier, intuitively, we expect that Exp#3 to be a more informative experiment to identify the graph structure since we would like to reduce the coherence between x_2 and x_4 . As we expected, in Exp#3, we can reduce the coherence more than that of Exp#2 as shown in Figure 4, and recover the exact graph structure as shown in Figure 5. On the other hand, the coherence of Exp #2 remains almost the same as that of Exp #1 and we fail to recover the exact graph structure. This numerical example illustrates that by using the property of coherence, we can guide biological experiment design more effectively.

Example 5. (limitation of reducing coherence by Ψ , nonlinear dynamics, $n = 3, N = 9, M = 5, s = 9$) consider the following set of ODEs:

$$\begin{aligned}\dot{x}_1 &= \gamma_1 x_1 + k_{+12} \frac{1}{1 + x_2^{n_{act}}} + k_{+13} \frac{1}{1 + x_3^{n_{act}}} \\ \dot{x}_2 &= \gamma_2 x_2 + k_{-21} \frac{x_1^{n_{ihb}}}{1 + x_1^{n_{ihb}}} + k_{-23} \frac{x_3^{n_{ihb}}}{1 + x_3^{n_{ihb}}} \\ \dot{x}_3 &= \gamma_3 x_3 + k_{-31} \frac{x_1^{n_{ihb}}}{1 + x_1^{n_{ihb}}} + k_{-32} \frac{x_2^{n_{ihb}}}{1 + x_2^{n_{ihb}}}\end{aligned}$$

where $(\gamma_1, \gamma_2, \gamma_3) = (-0.25, -0.23, -0.26)$, $k_{+12} = 1.2$, $k_{+13} = 1.25$, $k_{-21} = 2.8$, $k_{-23} = 2.1$, $k_{-31} = 2.7$, $k_{-32} = 1.8$, $n_{act} = 4$ (activation Hill coefficient) and $n_{ihb} = 4$ (inhibition Hill coefficient). The corresponding topology is shown in Figure 6 (b) (note that we intentionally choose the symmetric structure and similar parameters). The reconstruction error using Exp#1 data is shown in Figure 6 (a) (left, bottom) and the reconstruction error illustrates difficulties of resolving ambiguities from x_2 and x_3 . This can be captured by the coherence distribution of the sensing matrix based on the Exp#1 dataset; the correlation between the columns corresponding to x_1 and to x_2 is close to the correlation between the columns corresponding to x_1 and to x_3 . Based on the coherence distribution, we design two trials; for Exp#2, we perturb x_1 and for Exp#3, we perturb x_2 . As we expected, Exp #2 is not an effective experiment in terms of information. On the other hand, by using Exp#3, we can reduce both the maximum coherence and the averaged coherence, and reconstruct the exact graph structure as shown in Figure 6 (a).

Both Examples 4 and 5 illustrate that if the transformed sensing matrix is not incoherent enough to guarantee exact recovery, we can design a new experiment based on the distribution of coherence. Also, we show that an informative experiment can help to reduce coherence and thus reconstruct the exact graph structure. For a fair comparison, we use the same number of time points as M here. However, in practice, we can stack all the experimental data sets together if we assume that the linear map \mathbf{S}_q and the set of basis functions $\mathbf{F}_b(\mathbf{x})$ does not change for different experiment:

$$\begin{bmatrix} \mathbf{Z}_1 \\ \dots \\ \mathbf{Z}_k \end{bmatrix} = \begin{bmatrix} \bar{\Omega}_1 \\ \dots \\ \bar{\Omega}_k \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{v}}_1 \\ \dots \\ \bar{\mathbf{v}}_k \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{w}}_1 \\ \dots \\ \bar{\mathbf{w}}_k \end{bmatrix} \quad (16)$$

where the subscript $\mathbf{Z}_i, \bar{\Omega}_i, \bar{\mathbf{v}}_i, \bar{\mathbf{w}}_i$ represents the i -th experiment. As the number of measurements increase (M), one may be able to reduce the coherence. However, one can reduce the coherence only if the additional measurements provide us more useful information. As a trivial example, one could stack exactly the same data on top of the first, and increase M to $2M$, however the coherence is exactly the same as that of the original dataset.

3. Recovery of Gene Regulatory Networks

In this section, we present reconstruction of the exact graph structure and show how the condition for exact recovery will be used. First, we consider the ideal case where there are no hidden nodes and no measurement noise. Second, we extend the ideal case to the more general case.

3.1. Reconstructing Gene Regulatory Networks (ideal case)

In (13), \mathbf{q} represents the s -sparse network structure which we want to reconstruct from the time series gene expression by solving the l_1 -norm optimization:

$$\min \|\mathbf{q}\|_{l_1} \quad \text{s.t.} \quad \tilde{\mathbf{Z}} = \tilde{\Omega}\mathbf{q} \quad (CS) \quad (17)$$

where $\tilde{\mathbf{Z}} \triangleq \mathcal{P}\mathbf{Z}$, $\tilde{\Omega} \triangleq \mathcal{P}\bar{\Omega}$ and \mathcal{P} is the optimal transformation in (24)

Proposition 2.. *If the sensing matrix $\tilde{\Omega}$ constructed from time series data, multiplied by the optimal transformation, \mathcal{P} , has $2s$ linearly independent columns, then any s -sparse network structure \mathbf{q} can be reconstructed uniquely from $\tilde{\mathbf{Z}} = \tilde{\Omega}\mathbf{q}$.*

Proof. (Suppose not), then there are two s -sparse graph structures $\mathbf{q}_1, \mathbf{q}_2$ with $\tilde{\Omega}\mathbf{q}_1 = \tilde{\Omega}\mathbf{q}_2$ (or $\tilde{\Omega}(\mathbf{q}_1 - \mathbf{q}_2) = 0$). However, $\mathbf{q}_1 - \mathbf{q}_2$ is $2s$ -sparse, so there is a linear dependence between $2s$ columns of $\tilde{\Omega}$ (contradiction). \square

The requirement of $2s$ linearly independent columns in **Proposition 2** may be translated to an incoherence condition on the sensing matrix. That is, if the unknown s -sparse signal \mathbf{q} is reasonably sparse, it is possible to recover \mathbf{q} under the incoherence condition on the sensing matrix. Although the sensing matrix consists of redundant dictionaries, the coherence of the sensing matrix can be reduced. In a heuristic way, we multiply the redundant dictionaries $\Phi_{k,:}$ by a randomly chosen matrix Ψ_k^s at each time step k and iterate this step until the coherence is decreased. Or, we can find the optimal transformation \mathcal{P} (or \mathbf{P}) in (24) to reduce the coherence. In the previous numerical examples, we illustrated that the coherence of the sensing matrix is decreased by transformation and showed the exact reconstruction of graph structure.

Example 6. (Statistics) *Here, we compare the success rate of the proposed method with other methods such as L1 and L2. Figure 7 shows statistics of 50 trials for a simple linear case (for each trial, we randomly generate the influence map). Here, we count the number of successes of each method when any of the methods recover the exact structure. By reducing coherence, we can improve the success rate as shown in Figure 7 (a). Also, L2-regularization does not encourage sparsity but distributes the coefficients to be more similar to each other as shown in Figure 7 (b).*

3.2. Graph Reconstruction with Hidden Nodes

The main contributions of the proposed method in the previous section is the conversion of the problem of inferring graph structure into the CS framework. Then, we demonstrate that one could recover sparse graph structures from only a few measurements. However, for practical use, the proposed method needs to be able to deal with both sparsely corrupted signals ($\mathbf{v} = \mathbf{g}(\mathbf{x}_h)$) and measurement noise (\mathbf{w}) in (5).

In general, the assumption of accessibility or observability of all nodes [17] is not satisfied. Thus, we focus on the case in which the hidden node affects observable nodes directly as shown in Figure 1 (b). Also, without loss of generality, the hidden node dynamics could be any arbitrary

dynamic model. Or, even if there is no hidden node, a small portion of the biological experiment dataset could be in practice contaminated by large error resulting from, for example, mislabeling, or improper use of markers or antibodies. Moreover, all biological datasets are contaminated by at least a small amount of noise from measurement devices. Therefore, the proposed method should be robust. We note this goes beyond the results in [11] due to the consideration of hidden node dynamics with measurement noise.

Here, the question is whether it is still possible to reconstruct the graph structure reliably when measurements are corrupted. Since hidden nodes and measurement noise are considered, the number of time points is assumed to be greater than that of the previous case [17] (i.e., no corruption and no measurement noise). Thus, the number of rows of the sensing matrix is assumed to be greater than the number of columns. If the number of the time points $M (< N)$ is limited, then we can stack $\bar{\mathbf{z}}(j)$ with different Ψ_k^s or including different dataset as described in equation (16).

In CS literature [26], two decoding strategies for recovering the signal from a corrupted measurement are introduced, where the corruption includes both a possible sparse vector of large errors and a vector of small error affecting all the entries. It is shown that two decoding schemes allow the recovery of the signal with nearly the same accuracy as if no sparse large errors occurred. Our contribution is converting the problem of inferring the graph structure with hidden nodes into the highly robust error correction method framework [26] (**see Method: Section 2. Two step refinements**) and showing how this can improve the reliability of reconstruction.

Example 7. (*arbitrary corruption with no measurement noise*) Consider nonlinear ODEs as follows:

$$\begin{aligned}\dot{x}_1 &= -\gamma_1 x_1 + \alpha_{13} h_{act}(x_3) + v_1 \\ \dot{x}_2 &= -\gamma_2 x_2 + \alpha_{21} h_{act}(x_1) + v_2 \\ \dot{x}_3 &= -\gamma_3 x_3 + \alpha_{34} h_{act}(x_4) + \beta_{32} h_{ihb}(x_2) + v_3 \\ \dot{x}_4 &= -\gamma_4 x_4 + \beta_{41} h_{ihb}(x_1) + v_4\end{aligned}\tag{18}$$

where h_{act}, h_{ihb} represents Hill functions for activation and inhibition respectively, and v_i represents arbitrary corruption shown in Figure 8 (a) and assumed to be sparse (at each time step, we choose $\text{card}(v(j)) = 1$). The magnitude of v_i are about 50% of the magnitude of \dot{x} . Since we consider arbitrary corruption, we need more time points ($M > N$). By using two-step refinements, first we estimate sparse large corruption as shown in Figure 8 (b) (top) and then, we reconstruct \mathbf{q} .

In practice, a specific node is corrupted by a hidden node and a small portion of the dataset can be largely corrupted by human error. Also, since we choose the set of possible candidate basis functions of the sensing matrix in (6), the columns of the sensing matrix may not be able to represent the underlying system (i.e., missing dictionaries). Then, we can consider the influence from these missing dictionaries as \mathbf{v} .

Example 8. (*Arbitrary corruption with measurement noise*) Recall a model (18) with different parameters and consider sparse large corruption \mathbf{v} and small magnitude noise \mathbf{w} (1% of the magnitude of \dot{x}). Figure 9 shows the time series data and reconstruction result.

3.3 Geometric view

In equation (17), since we assume all nodes are accessible and perfect measurement (meaning that there is no hidden node, $\mathbf{v} = 0$ and no measurement noise, $\mathbf{w} = 0$), we can solve equation (27)

directly without filtering out the unmodelled dynamics (26) (i.e., $\mathbf{Z} = \bar{\Omega}\mathbf{q}$). If there exist hidden nodes or measurement noise, we can still provide an unambiguous indication of the existence of these corruption ($\mathbf{e} \neq \mathbf{0}$).

The intuition is that $\mathbf{Z}(= \bar{\Omega}\mathbf{q} + \bar{\Psi}\mathbf{e})$ can be decomposed as the superposition of an arbitrary element in $V(= \mathbf{Z} - \mathbf{Q}\mathbf{Q}^*\mathbf{e})$ and of an element in $V^\perp(= \mathbf{Q}\mathbf{Q}^*\mathbf{e})$ as shown in Figure 10. In other words, \mathbf{Z} can be decomposed as the superposition of modelled dynamics and anomalies caused by hidden node or unmodelled dynamics. This geometric view enables us to understand how we could reveal deficiencies in our model:

- $\bar{\mathbf{Z}} = \mathbf{Q}^*\mathbf{Z} = \mathbf{0}$: there is no hidden node $\implies \bar{\mathbf{Z}} \in \mathcal{R}(\bar{\Omega})$
- $\bar{\mathbf{Z}} \neq \mathbf{0}$: \mathbf{Z} cannot be represented by $\bar{\Omega}\mathbf{q}$ so there might be hidden nodes or our dictionaries in the sensing matrix $\bar{\Omega}$ are not sufficient to represent \mathbf{Z} (revealing deficiencies in our model or dictionaries).

4. HER2 Overexpressed Breast Cancer

We apply the proposed algorithm to study a breast cancer signaling pathway by reconstructing the graph structure using an RPPA dataset [27] as shown in Figure 11 (see also Figure S1, S2, and S3 for details: each figure presents the RPPA dataset and the result of graph reconstruction compared with $L1$, $L2$ -optimization). Here, we choose small networks which are composed of 3 nodes and known to be sparsely connected, i.e., $\text{PI3K} \rightarrow \text{PDK} \rightarrow \text{Akt}$ and $\text{PDK} \rightarrow \text{Akt} \rightarrow \text{mTOR}$ in order to satisfy our assumption such that the influence on observable nodes from a hidden node should be sparse (i.e., \mathbf{v} is sparse). The graph structures identified by the proposed method are consistent with the current understanding of the networks, whereas those found using $L1$ - and $L2$ -optimizations fail to reconstruct the known structure as shown in Figure 12.

Also, an abstract model of the breast cancer signal pathway proposed by M. Moasser [31] is considered, as shown in Figure S3(b) where PHLPP isoforms are a pair of protein phosphatases, PHLPP1 and PHLPP2, which are important regulators of Akt serine-threonine kinases (Akt1, Akt2, Akt3) and conventional protein kinase C (PKC) isoforms. PHLPP may act as a tumor suppressor in several types of cancer due to its ability to block growth factor-induced signaling in cancer cells [32]. PHLPP dephosphorylates Ser473 (the hydrophobic motif) in Akt, thus partially inactivating the kinase [33]. Unfortunately, in our RPPA dataset, we do not have PHLPP so we simply consider three nodes ($\text{Akt}_{\text{pT308}}$, $\text{Akt}_{\text{pS473}}$ and mTOR). Figure 11(right) shows the result of the proposed method using the RPPA dataset. The reconstructed graph structure matches up to known structure (Figure S3 (b)). Specifically, our result can capture the partial inactivating characteristics of PHLPP (i.e., $\text{mTOR}(\rightarrow \text{PHLPP}) \dashv \text{Akt}_{\text{pS473}}$).

Conclusion

We proposed a method for reconstructing sparse graph structures based on time series gene expression data without any *a priori* information. We demonstrated that the proposed method can reconstruct graph structure reliably. Also, we illustrated that coherence in the sensing matrix can be used as a guideline for designing effective experiments.

Second, the proposed method is extended to the cases in which dynamics is corrupted by hidden nodes and the measurement is corrupted by human error in addition to the measurement noise. Using a two-step refinement procedure, we demonstrate good performance for the reconstruction of graph structure. A set of numerical examples is implemented to illustrate the method and its

performance. Also, a biological example of HER2 overexpressed breast cancer using an RPPA dataset is studied. We are currently applying our method to recover the HER2 signaling pathway, where a significant part of the network is currently unknown.

Method

1. Reducing Coherence by Transformation

1.1. Rearranging the sensing matrix

Define $\mathbf{z}(j)$ as a vector of each component of \mathbf{y}_i at the j -th time point:

$$\begin{aligned} \mathbf{z}(j) &\triangleq \begin{bmatrix} \mathbf{y}_1^\top \mathbf{e}_j \\ \mathbf{y}_2^\top \mathbf{e}_j \\ \dots \\ \mathbf{y}_n^\top \mathbf{e}_j \end{bmatrix} = \begin{bmatrix} y_1(t_j) \\ y_2(t_j) \\ \dots \\ y_n(t_j) \end{bmatrix} = \begin{bmatrix} \Phi_{j,:} \mathbf{q}_1 \\ \Phi_{j,:} \mathbf{q}_2 \\ \dots \\ \Phi_{j,:} \mathbf{q}_n \end{bmatrix} + \begin{bmatrix} v_1(t_j) \\ v_2(t_j) \\ \dots \\ v_n(t_j) \end{bmatrix} + \begin{bmatrix} w_1(t_j) \\ w_2(t_j) \\ \dots \\ w_n(t_j) \end{bmatrix} \\ &= (\mathbf{I}_n \otimes \Phi_{j,:}) \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \mathbf{v}(j) + \mathbf{w}(j) \end{aligned} \quad (19)$$

where $\Phi_{j,:}$ represents the j -th row of Φ . Consider $\Psi_j^s \in \mathbb{R}^{n \times n}$ and multiply equation (19) by Ψ_j^s :

$$\bar{\mathbf{z}}(j) \triangleq \Psi_j^s \mathbf{z}(j) = \Psi_j^s (\mathbf{I}_n \otimes \Phi_{j,:}) \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \bar{\mathbf{v}}(j) + \bar{\mathbf{w}}(j) \quad (20)$$

where $\bar{\mathbf{v}}(j) \triangleq \Psi_j^s \mathbf{v}(j)$, $\bar{\mathbf{w}}(j) \triangleq \Psi_j^s \mathbf{w}(j)$ and $\text{rank}(\Psi_j^s (\mathbf{I}_n \otimes \Phi_{j,:})) = \text{rank}(\Psi_j^s)$ since $\text{rank}(\mathbf{I}_n \otimes \Phi_{j,:}) = \text{rank}(\mathbf{I}_n) \cdot \text{rank}(\Phi_{j,:}) = n$. By stacking $\bar{\mathbf{z}}(j)$,

$$\mathbf{Z} \triangleq \begin{bmatrix} \bar{\mathbf{z}}(1) \\ \bar{\mathbf{z}}(2) \\ \dots \\ \bar{\mathbf{z}}(M) \end{bmatrix} = \begin{bmatrix} \Psi_1^s (\mathbf{I}_n \otimes \Phi_{1,:}) \\ \Psi_2^s (\mathbf{I}_n \otimes \Phi_{2,:}) \\ \dots \\ \Psi_M^s (\mathbf{I}_n \otimes \Phi_{M,:}) \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{v}}(1) \\ \bar{\mathbf{v}}(2) \\ \dots \\ \bar{\mathbf{v}}(M) \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{w}}(1) \\ \bar{\mathbf{w}}(2) \\ \dots \\ \bar{\mathbf{w}}(M) \end{bmatrix} \triangleq \bar{\Omega} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \bar{\mathbf{v}} + \bar{\mathbf{w}} \quad (21)$$

1.2. Randomly chosen matrix Ψ

The optimization problem (14) is not trivial because of the constraint. One simple and heuristic approach is that we select Ψ_i^s by (normalized) randomly chosen matrix with independent identically distributed (i.i.d.) random variable where $\bar{\Omega}$ has full row rank, calculate $\mu(\bar{\Omega})$ and run this with several times to reduce coherence. Since the randomly chosen matrix, Ψ_i^s , spreads out the component of $\Phi_{i,:}$ uniformly, we can reduce coherence.

1.3. Finding the optimal transformation

This heuristic approach may not be enough to reduce the coherence. Consider a nonsingular matrix $\mathcal{P} \in \mathbb{R}^{M \cdot n \times M \cdot n}$ and $\tilde{\Omega} = \mathcal{P}\bar{\Omega}$ where $\bar{\Omega}$ is constructed by heuristic way, i.e., randomly chosen matrix.

$$\tilde{\mu}_{i,j} \triangleq \frac{|\langle \tilde{\Omega}_i, \tilde{\Omega}_j \rangle|}{\|\tilde{\Omega}_i\|_2 \|\tilde{\Omega}_j\|_2} = \frac{|\langle \mathcal{P}\bar{\Omega}_i, \mathcal{P}\bar{\Omega}_j \rangle|}{\|\mathcal{P}\bar{\Omega}_i\|_2 \|\mathcal{P}\bar{\Omega}_j\|_2} \triangleq \frac{|\bar{\Omega}_i^\top \mathbf{P} \bar{\Omega}_j|}{\sqrt{(\bar{\Omega}_i^\top \mathbf{P} \bar{\Omega}_i)(\bar{\Omega}_j^\top \mathbf{P} \bar{\Omega}_j)}} \quad (22)$$

where $\mathbf{P} \triangleq \mathcal{P}^\top \mathcal{P} = \mathbf{P}^\top \in \mathbb{S}^{M \cdot n}$ is positive definite and $\tilde{\Omega}_i$ denote the i -th column of $\tilde{\Omega}$. Note that if $\mathcal{P}^\top \mathcal{P} = \mathbf{I}$, $\tilde{\mu}_{i,j} = \frac{|\langle \bar{\Omega}_i, \bar{\Omega}_j \rangle|}{\|\bar{\Omega}_i\|_2 \|\bar{\Omega}_j\|_2} = \frac{|\langle \bar{\Omega}_i, \bar{\Omega}_j \rangle|}{\|\bar{\Omega}_i\|_2 \|\bar{\Omega}_j\|_2} \triangleq \bar{\mu}_{i,j}$. Therefore, our goal is finding $\mathbf{P} (= \mathcal{P}^\top \mathcal{P}) \in \mathbb{S}^+$ such that

$$\min_{\mathbf{P}} \mu(\tilde{\Omega}) = \min_{\mathbf{P}} \max_{j < k} \tilde{\mu}_{j,k} = \min_{\mathbf{P}} \|\Gamma\|_\infty \quad (23)$$

where $\Gamma \in \mathbb{R}^{(n \cdot n) \times C_2}$ can be defined as follows:

$$\Gamma \triangleq \begin{bmatrix} \frac{\text{tr}(\mathcal{P}\bar{\Omega}_1\bar{\Omega}_2^\top\mathcal{P}^\top)}{\|\mathcal{P}\bar{\Omega}_1\bar{\Omega}_2^\top\mathcal{P}^\top\|_F} & \cdots & \frac{\text{tr}(\mathcal{P}\bar{\Omega}_1\bar{\Omega}_{N \cdot n}^\top\mathcal{P}^\top)}{\|\mathcal{P}\bar{\Omega}_1\bar{\Omega}_{N \cdot n}^\top\mathcal{P}^\top\|_F} & \frac{\text{tr}(\mathcal{P}\bar{\Omega}_2\bar{\Omega}_3^\top\mathcal{P}^\top)}{\|\mathcal{P}\bar{\Omega}_2\bar{\Omega}_3^\top\mathcal{P}^\top\|_F} & \cdots & \frac{\text{tr}(\mathcal{P}\bar{\Omega}_{N \cdot n-1}\bar{\Omega}_{N \cdot n}^\top\mathcal{P}^\top)}{\|\mathcal{P}\bar{\Omega}_{N \cdot n-1}\bar{\Omega}_{N \cdot n}^\top\mathcal{P}^\top\|_F} \end{bmatrix}^\top$$

In practice, we ignore the denominator and solve the following problem:

$$\min_{\mathbf{P}=\mathbf{P}^\top \succ 0} \left\| \left[\text{vec}(\bar{\Omega}_1\bar{\Omega}_2^\top) \quad \cdots \quad \text{vec}(\bar{\Omega}_1\bar{\Omega}_{N \cdot n}^\top) \quad \cdots \quad \text{vec}(\bar{\Omega}_{N \cdot n-1}\bar{\Omega}_{N \cdot n}^\top) \right]^\top \text{vec}(\mathbf{P}) \right\|_\infty \quad (24)$$

We can also combine $\|\cdot\|_\infty$ and $\|\cdot\|_1$ to reduce the coherence. Note that for $\|\cdot\|_\infty$, we minimize the maxim coherence and for $\|\cdot\|_1$, we minimize the sum of the all possible combinations of the columns of $\tilde{\Omega}$, i.e., $\tilde{\mu}_{i,j}$. Thus, if certain bases are highly correlated, \mathbf{P} or \mathcal{P} makes the components of the sensing matrix spread out enough to differentiate the influences from those bases. Since we ignore the denominator, the optimal solution of (24) may be suboptimal. Thus, we can also combine heuristic way and (24) iteratively to reduce coherence of the sensing matrix in practice.

2. Two-step refinements

2.1. Sparse large corruption with no measurement noise

Recall equation (13) where $\bar{\mathbf{w}} = \mathbf{0}$:

$$\mathbf{Z} = \bar{\Omega} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \text{diag}\{\Psi_1^s, \Psi_2^s, \dots, \Psi_M^s\} \begin{bmatrix} \mathbf{v}(1) \\ \mathbf{v}(2) \\ \dots \\ \mathbf{v}(M) \end{bmatrix} = \bar{\Omega} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \bar{\Psi} \mathbf{v} \quad (25)$$

where $\bar{\Psi} \triangleq \text{diag}\{\Psi_1^s, \Psi_2^s, \dots, \Psi_M^s\}$ and $\mathbf{v}(j)$ represents sparse large corruption at the j -th time point, that could result from the existence of hidden nodes. We assume that the influence from hidden nodes is sparse and unknown (i.e., $\mathbf{v}(j)$ is assumed to be sparse). In other words, hidden nodes can affect only a few nodes' dynamics (intuitively, if hidden nodes affect all nodes, there is no way to reconstruct the graph structure). Then, we consider the reconstruction of graph structure \mathbf{q} from the corrupted signal \mathbf{Z} .

Since we assume the number of rows of the sensing matrix $\bar{\Omega}$ is greater than the number of columns ($M \cdot n > N \cdot n$), we consider \mathbf{Q}^* which annihilates the sensing matrix $\bar{\Omega}$ on the left

($\mathbf{Q}^* \bar{\Omega} = 0$) where \mathbf{Q}^* is any $(M \cdot n - N \cdot n) \times M \cdot n$ matrix whose kernel is the range of $\bar{\Omega}$ in $\mathbb{R}^{M \cdot n}$ ($\text{rank}(\mathbf{Q}^*) + \text{nullity}(\mathbf{Q}^*) = M \cdot n$):

$$\bar{\mathbf{Z}} \triangleq \mathbf{Q}^* \mathbf{Z} = \mathbf{Q}^* \bar{\Omega} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \mathbf{Q}^* \bar{\Psi} \mathbf{v}$$

Then, the following two-step optimization problem enables us to compute \mathbf{q} :

- Filter unmodelled dynamics out from the measurement:

$$\hat{\mathbf{v}} = \arg \min \|\mathbf{v}\|_{l_1} \quad \text{s.t.} \quad \bar{\mathbf{Z}} = \mathbf{Q}^* \bar{\Psi} \mathbf{v} \quad (26)$$

- Reconstruct \mathbf{q} :

$$\begin{aligned} \min \|\mathbf{q}\|_{l_1} \quad \text{s.t.} \quad & \mathbf{Z} - \bar{\Psi} \hat{\mathbf{v}} = \bar{\Omega} \mathbf{q} \\ \text{or} \quad & \tilde{\mathbf{Z}} = \mathcal{P}(\mathbf{Z} - \bar{\Psi} \hat{\mathbf{v}}) = \mathcal{P} \bar{\Omega} \mathbf{q} = \tilde{\Omega} \mathbf{q} \quad (\text{if necessary, } \mathcal{P} \text{ in (23)}) \end{aligned} \quad (27)$$

If we could somehow get an accurate estimate $\hat{\mathbf{v}}$ from equation (26), equation (27) represents the problem of reconstructing the graph structure \mathbf{q} . The intuition is that $\mathbf{Z} (= \bar{\Omega} \mathbf{q} + \bar{\Psi} \mathbf{v})$ can be decomposed as the superposition of modelled dynamics and anomalies caused by hidden node or unmodelled dynamics.

The two step convex optimization problems (26) and (27) are l_1 -norm optimization problems in CS. Thus, if the sensing matrix $\bar{\Psi}$ and $\bar{\Omega}$ (or $\mathcal{P} \bar{\Omega}$) satisfy the incoherence condition, signals \mathbf{v} and \mathbf{q} can be recovered exactly [17] [29]. Here, there are many possible choices of \mathbf{Q}^* but we have to choose \mathbf{Q}^* to satisfy the incoherence condition for the exact recovery of \mathbf{v} [17]. To choose such a \mathbf{Q}^* , we observe that $\bar{\Omega}$ can be denoted as follows using Singular Value Decomposition (SVD):

$$\bar{\Omega} = [\mathbf{U}_1 \quad \mathbf{U}_2] \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^\top \\ \mathbf{V}_2^\top \end{bmatrix} = \mathbf{U}_1 \Sigma \mathbf{V}_1^\top \quad (28)$$

where $\mathbf{U}_1 \in \mathbb{R}^{M \cdot n \times r}$, $\mathbf{U}_2 \in \mathbb{R}^{M \cdot n \times (M \cdot n - r)}$, $\Sigma \in \mathbb{R}^{r \times r}$, $\mathbf{V}_1 \in \mathbb{R}^{N \cdot n \times r}$, $\mathbf{V}_2 \in \mathbb{R}^{N \cdot n \times (N \cdot n - r)}$ and r is the rank of $\bar{\Omega}$. Suppose we choose \mathbf{Q}^* such that $\mathbf{Q}^* = \Xi \mathbf{U}_2^\top$. Then:

$$\begin{aligned} \mathbf{Q}^* \mathbf{Z} &= \mathbf{Q}^* \bar{\Omega} \mathbf{q} + \mathbf{Q}^* \bar{\Psi} \mathbf{v} \\ \bar{\mathbf{Z}} &= (\Xi \mathbf{U}_2^\top) \mathbf{U}_1 \Sigma \mathbf{V}_1^\top \mathbf{q} + (\Xi \mathbf{U}_2^\top) \bar{\Psi} \mathbf{v} \\ \tilde{\mathbf{Z}} &= (\Xi \mathbf{U}_2^\top) \bar{\Psi} \mathbf{v} \end{aligned}$$

where Ξ can be used as a tuning matrix for satisfying the incoherence condition. A geometric view in **Results and Discussion: Section 3.3.** enables us to understand how we could reveal deficiencies in our model.

2.2. Sparse large corruption with measurement noise

While considering influence from hidden nodes is interesting, it still may not be realistic to assume that except for hidden nodes, one is able to measure the node dynamics with infinite precision. A better model would assume that there is measurement noise. Consider the problem of recovering

the graph structure \mathbf{q} from the vector \mathbf{Z} which is corrupted by measurement noise $\bar{\mathbf{w}}$ in equation (13):

$$\mathbf{Z} = \bar{\Omega} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \bar{\Psi} \begin{bmatrix} \mathbf{v}(1) + \mathbf{w}(1) \\ \mathbf{v}(2) + \mathbf{w}(2) \\ \dots \\ \mathbf{v}(M) + \mathbf{w}(M) \end{bmatrix} = \bar{\Omega} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_n \end{bmatrix} + \bar{\Psi} \begin{bmatrix} \mathbf{e}(1) \\ \mathbf{e}(2) \\ \dots \\ \mathbf{e}(M) \end{bmatrix} \quad (29)$$

where $\mathbf{e}(j) = \mathbf{v}(j) + \mathbf{w}(j)$, \mathbf{w} is Gaussian noise $\mathcal{N}(0, \sigma)$ assumed to be bounded $\|\mathbf{w}\|_{l_2} \leq \epsilon$. In general, we can consider any corruption decomposed into sparse large error \mathbf{v} and small magnitude error \mathbf{w} [26]. Then, modified two-step refinements can be applied as follows:

$$\begin{aligned} \bar{\mathbf{Z}} &= \mathbf{Q}^* \bar{\Omega} \mathbf{q} + \mathbf{Q}^* \bar{\Psi} \mathbf{e} = \mathbf{Q}^* \bar{\Psi} (\mathbf{v} + \mathbf{w}) \\ \hat{\mathbf{v}} &= \arg \min \|\mathbf{v}\|_{l_1} \quad \text{s.t.} \quad \|\bar{\mathbf{Z}} - \mathbf{Q}^* \bar{\Psi} \mathbf{v}\|_{l_2} \leq \epsilon_1 \\ \mathbf{q} &= \arg \min \|\mathbf{q}\|_{l_1} \quad \text{s.t.} \quad \|\mathbf{Z} - \bar{\Omega} \mathbf{q} - \hat{\mathbf{e}}\|_{l_2} \leq \epsilon_2, \hat{\mathbf{e}} = \bar{\Psi} \hat{\mathbf{v}} \end{aligned} \quad (30)$$

where the parameters ϵ_1, ϵ_2 above depend on the magnitude of the small errors ϵ , which can be determined as in [26].

Competing interests

The authors declare that they have no competing interests.

Author's contributions

Conceived and developed computational method: YHC CJT,
 Analyzed the data : YHC CJT,
 Supervised the project: JWG CJT,
 Wrote the paper: YHC CJT

Acknowledgements

This research was supported by the NIH NCI under the ICBP and PS-OC programs (5U54CA112970-08), and by the Stand Up To Cancer-American Association for Cancer Research Dream Team Translational Cancer Research Grant SU2C-AACR-DT0409 to JWG.

References

- [1] Ideker, T., Ozier, O., Schwikowski, B., Siegel, A.F.: Discovering regulatory and signaling circuits in molecular interaction networks. *Bioinformatics* **18**, 233–240 (2002)
- [2] Wiley, H.S., Shvartsman, S.Y., Lauffenburger, D.A.: Computational modeling of the egf-receptor system: a paradigm for systems biology. *Trends in Cell Biology* **13**(1), 43–50 (2003)
- [3] Ma, L., Iglesias, P.A.: Quantifying robustness of biochemical network models. *BMC Bioinformatics* **3**(38), 1–13 (2002)

- [4] Sontag, E.D.: Network reconstruction based on steady-state data. *Essays in Biochemistry* **45**, 161–176 (2008)
- [5] Zechner, C., Ruessa, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., Koepl, H.: Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Science of the United States of America* **109**(21), 8340–8345 (2012)
- [6] Zavlanos, M.M., Julius, A.A., Boyd, S.P., Pappas, G.J.: Identification of stable genetic networks using convex programming. *Proceedings of the American Control Conference (ACC)*, 2755–2760 (2008)
- [7] Cooper, N.G., Belta, C.A., Julius, A.A.: Genetic regulatory network identification using multivariate monotone functions. *Proceedings of the IEEE conference on Decision and Control and European Control Conference (CDC-ECC)*, 2208–2213 (2011)
- [8] Porreca, R., Drulhe, S., de Jong, H., Ferrari-Trecate, G.: Structural identification of piecewise-linear models of genetic regulatory networks. *Journal of Computational Biology* **15**(10), 1365–1380 (2008)
- [9] Bernardo, D.D., Gardner, T.S., Collins, J.J.: Robust identification of large genetic networks. *Pacific Symposium on Biocomputing* **9**, 486–497 (2004)
- [10] Richard, G., Julius, A.A., Belta, C.: Optimizing regulation functions in gene network identification. *IEEE Conference on Decision and Control (CDC)*, 745–750 (2013)
- [11] Gonçalves, J.M., Warnick, S.C.: Necessary and sufficient conditions for dynamical structure reconstruction of lti networks. *IEEE Transactions on Automatic Control* **53**(7), 1670–1674 (2008)
- [12] Sontag, E., Kiyatkin, A., Kholodenko, B.N.: Inferring dynamics architecture of cellular network using time series of gene expression, protein and metabolite data. *Bioinformatics* **20**(12), 1877–1886 (2004)
- [13] Han, S., Yoon, Y., Cho, K.-H.: Inferring biomolecular interaction networks based on convex optimization. *Computational Biology and Chemistry* **31**(5-6), 347–354 (2007)
- [14] Napolitani, D., Sauer, T.D.: Reconstructing the topology of sparsely connected dynamical networks. *Physical. Review. E* **77**(2), 026103 (2008)
- [15] Napolitani, D., Sauer, T., Struppa, D.C., Petricoin, E., Liotta, L.: Augmented sparse reconstruction of protein signaling networks. *Journal of Theoretical Biology* **255**(1), 40–52 (2008)
- [16] Yuan, Y., Stan, G.-B., Warnick, S., Gonçalves, J.: Robust dynamical network structure reconstruction. *Automatica* **47**, 1230–1235 (2011)
- [17] Chang, Y.H., Tomlin, C.J.: Data-driven graph reconstruction using compressive sensing. *IEEE Conference on Decision and Control (CDC)*, 1035–1040 (2012)
- [18] Steinke, F., Seeger, M., Tsuda, K.: Experimental design for efficient identification of gene regulatory networks using sparse bayesian models. *BMC Systems Biology*

- [19] Pan, W., Yuan, Y., Gonçalves, J., Stan, G.-B.: Reconstruction of arbitrary biochemical reaction networks: A compressive sensing approach. *IEEE Conference on Decision and Control (CDC)*, 2334–2339 (2012)
- [20] Noor, A., Serpedin, E., Nounou, M., Nounou, H.: Reverse engineering sparse gene regulatory networks using cubature kalman filter and compressed sensing. *Advances in Bioinformatics*, 205763 (2013)
- [21] Deng, M., Emad, A., Milenkovic, O.: Causal compressive sensing for gene network inference. *arXiv:1202.5678* (2012)
- [22] Wang, W.-X., Lai, Y.-C., Grebogi, C., Ye, J.: Network reconstruction based on evolutionary-game data via compressive sensing. *Phys. Rev. X* **1**, 021021 (2011)
- [23] Candès, E., Romberg, J.: Sparsity and incoherence in compressive sampling. *Inverse Problems* **23**, 969–985 (2006)
- [24] Chang, Y.H., Gray, J., Tomlin, C.J.: Optimization-based inference for temporally evolving boolean networks with applications in biology. *Proceedings of the American Control Conference (ACC)*, 4129–4134 (2011)
- [25] Chang, Y.H., Gray, J., Tomlin, C.J.: Optimization-based inference for temporally evolving networks with applications in biology. *Journal of Computational Biology* **19**(12), 1307–1323 (2012)
- [26] Candès, E.J., Randall, P.A.: Highly robust error correction by convex programming. *IEEE Transaction on Information Theory* **54**(7), 2829–2840 (2008)
- [27] Hennessy, B.T., Lu, Y., Gonzalez-Angulo, A.M., Carey, M.S., Myhre, S., Ju, Z., Davies, M.A., Liu, W., Coombes, K., Meric-Bernstam, F., Bedrosian, I., McGahren, M., Agarwal, R., Zhang, F., Overgaard, J., Alsner, J., Neve, R.M., Kuo, W.-L., Gray, J.W., Borresen-Dale, A.-L., Mills, G.B.: A technical assessment of the utility of reverse phase protein arrays for the study of the functional proteome in non-microdissected human breast cancers. *Clinical Proteomics* **6**(4), 129–151
- [28] Tao, T.: Compressed sensing (Or: the equation $Ax = b$, revisited)
- [29] Candès, E., Wakin, M.: An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 21–30 (2008)
- [30] Candès, E.J., Romberg, J.K., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transaction on Information Theory* **52**, 489–509 (2004)
- [31] Moasser, M.: Understanding the network topology underlying addiction to HER2, Integrative Cancer Biology Program (ICBP) Days of Science, the Sequel, Waverly Club 1100 Southeast Waverly Drive Portland, OR 97222, June 11-12, 2012.
- [32] Brognard, J., Newton, A.C.: Phlipping the switch on akt and protein kinase c signaling. *Trends in Endocrinology and Metabolism* **19**(6), 223–30 (2008)
- [33] Gao, T., Furnari, F., Newton, A.C.: Phlpp: a phosphatase that directly dephosphorylates akt, promotes apoptosis, and suppresses tumor growth. *Molecular Cell* **18**(1), 13–24 (2005)

Figures

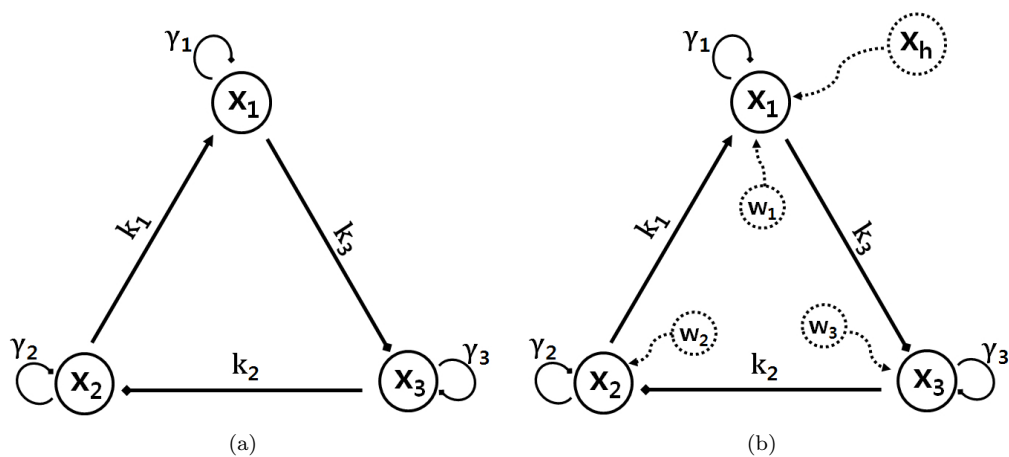


Figure 1. A graph representation of nonlinear ODEs. (a) (Example 1, $n = 3$, $\mathbf{S}_q \in \mathbb{R}^{3 \times 6}$): among 18 ($= 3 \times 6$) components, only 6 components are non-zero (b) (Example 1 with hidden node and measurement noise) there exists hidden node \mathbf{x}_h affecting \mathbf{x}_1 and process noise \mathbf{w} .

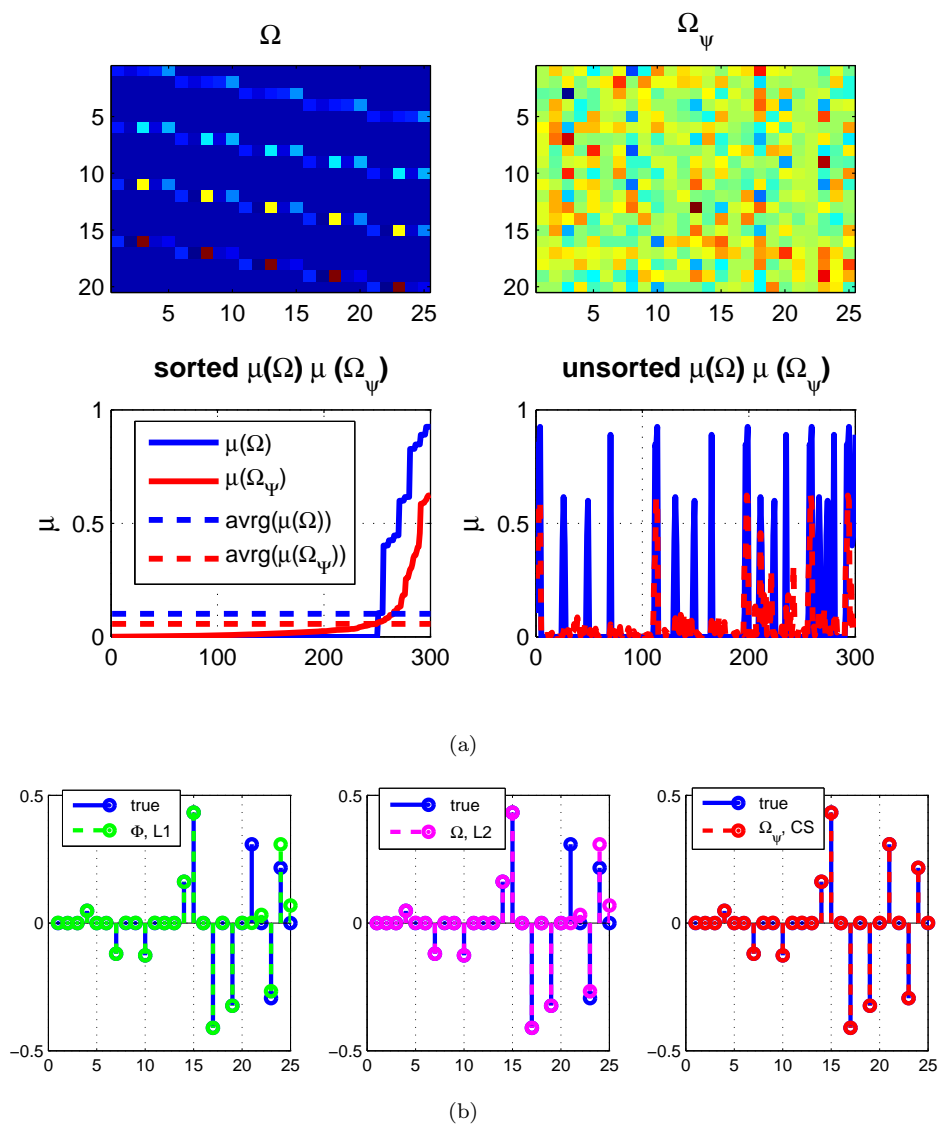


Figure 2. Reducing coherence for a linear system. (Example 2) (a) the sensing matrix Ω , without and with transformation Ψ (top) and coherence (bottom) (b) reconstruction result (left) $L1$ (middle) $L2$ with Ω (right) l_1 optimization with Ω_Ψ (CS).

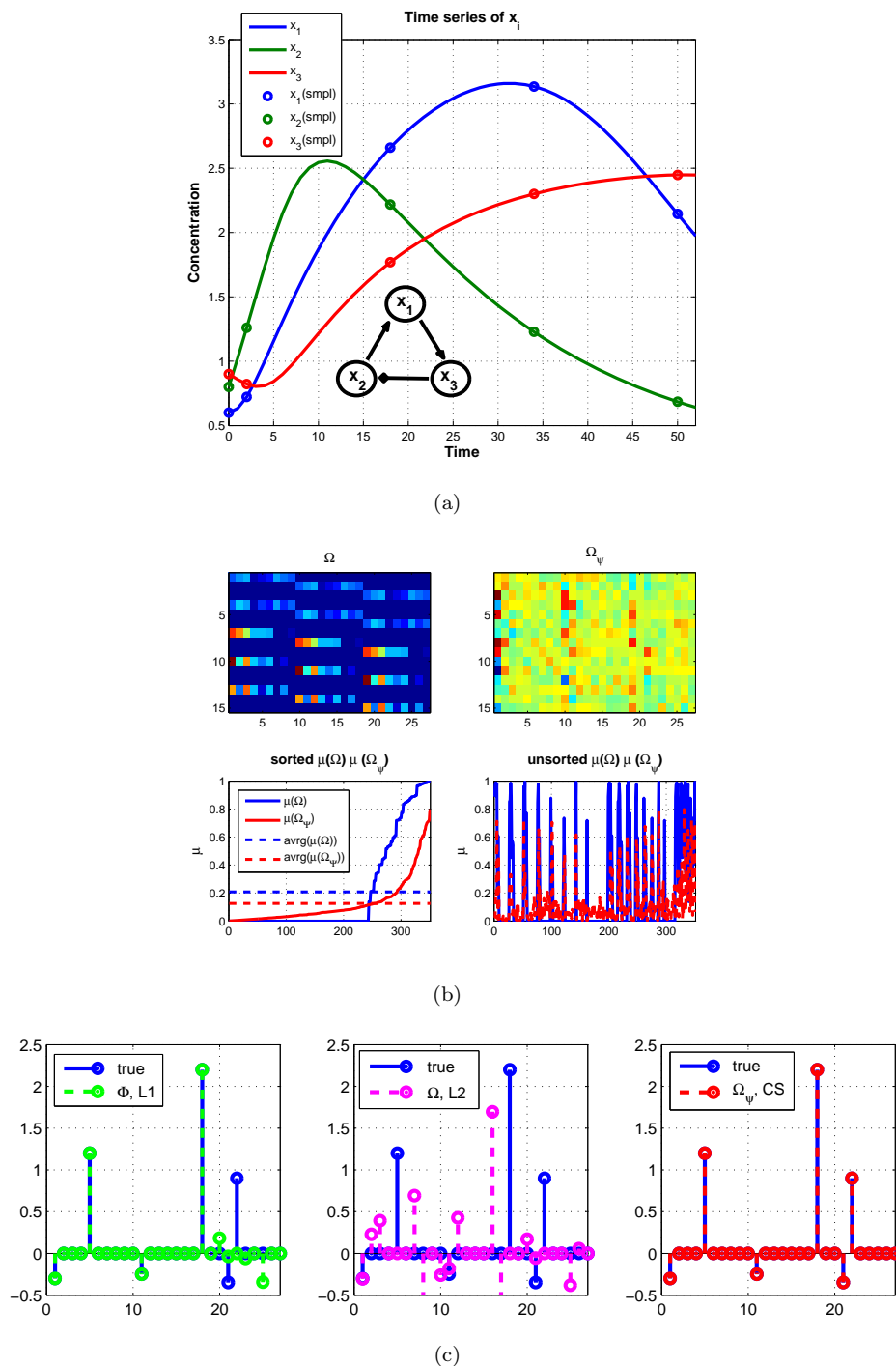


Figure 3. Reducing coherence for a nonlinear system. (Example 3) (a) Time series of x_1 , x_2 , x_3 for model (15) (b) the sensing matrix w/o, and w/ Ψ matrix (top) and coherence (bottom) (c) reconstruction results.

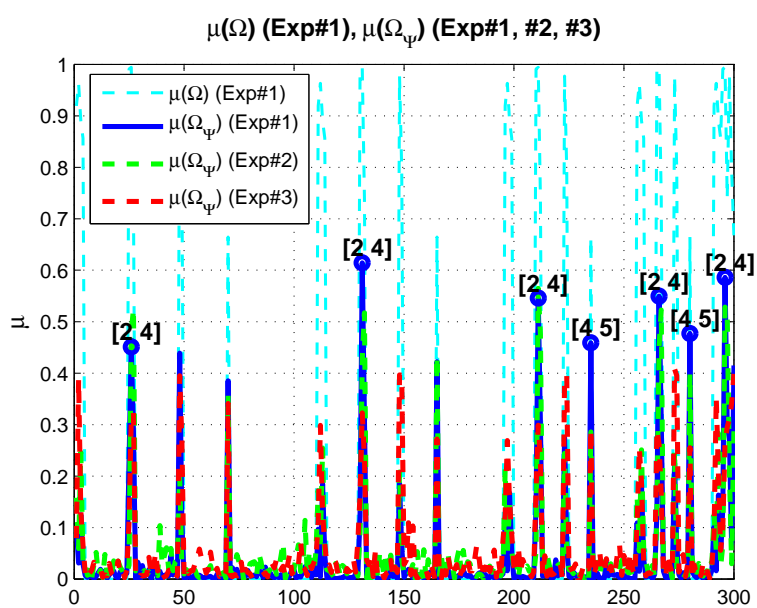


Figure 4. Guideline for designing new experiment based on coherence distribution. (Example 4) coherence distributions of $\tilde{\Omega}$ show that x_2 and x_4 cause high coherence.

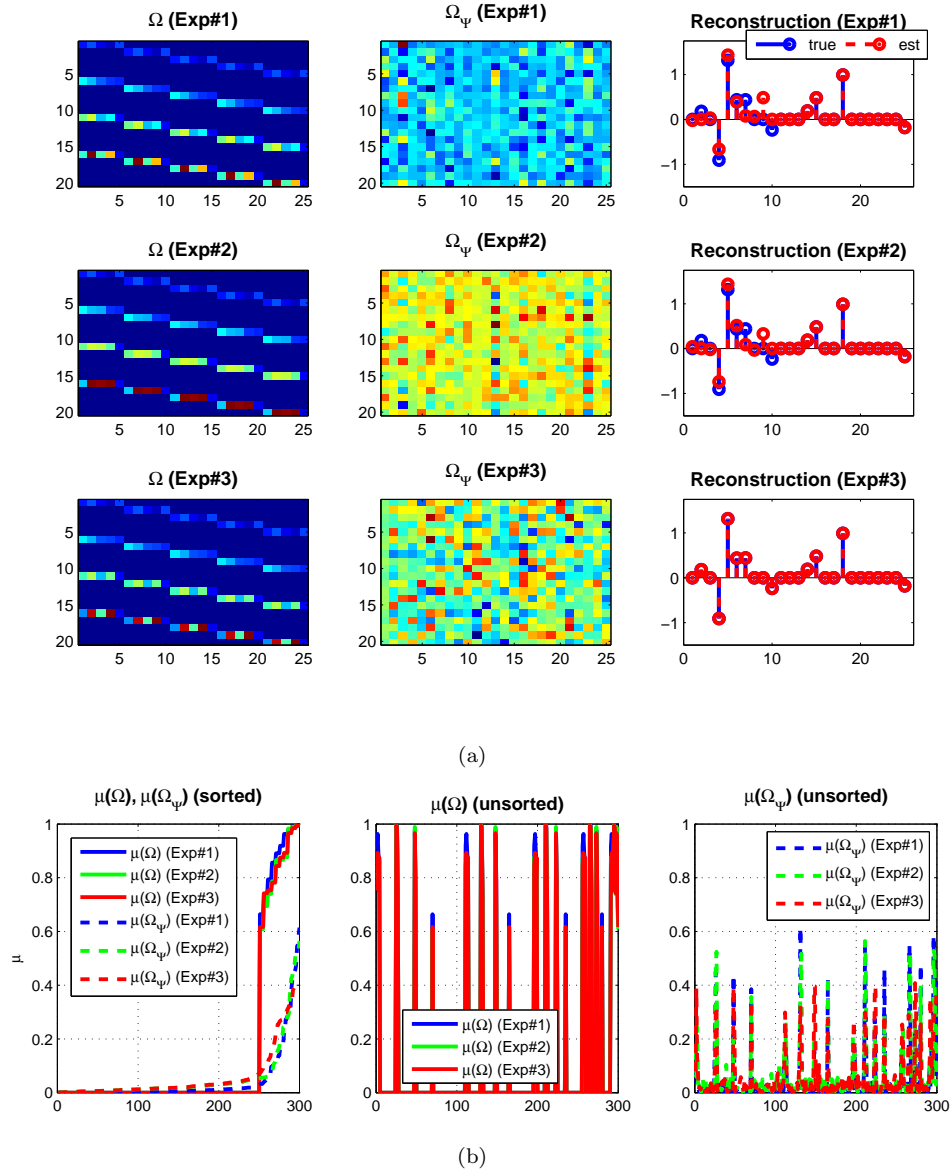
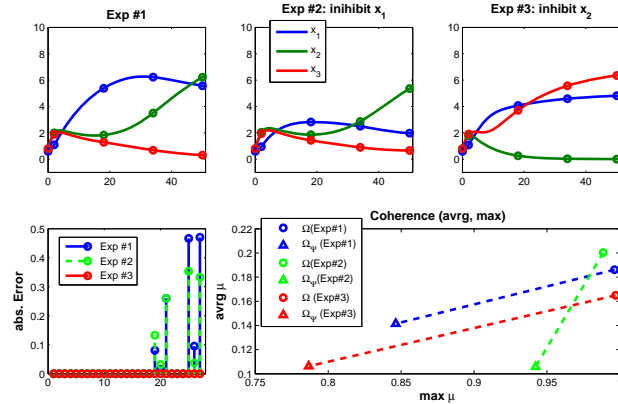
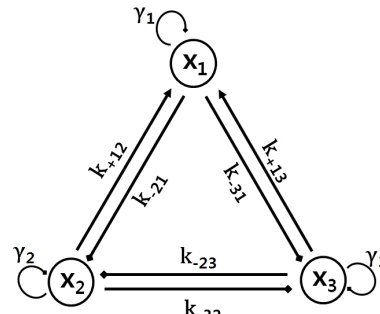


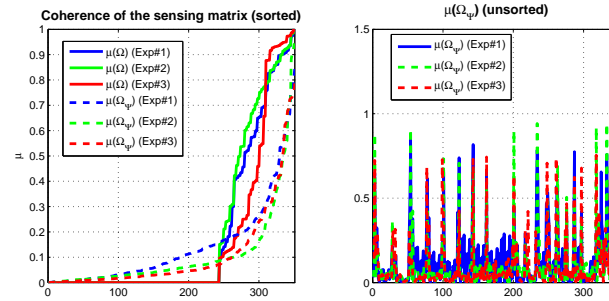
Figure 5. Design effective experiment for a linear system. (Example 4) Reconstruction results based on different experiments: Exp#1 represents the original experimental dataset which has limitation to reduce coherence; Exp#2 represents non-effective experimental dataset; and Exp#3 represents the effective experimental dataset (a) the sensing matrix and reconstruction result (b) coherence comparison



(a)



(b)



(c)

Figure 6. Design effective experiment for a nonlinear system. (Example 5)

Reconstruction result based on different experiment (nonlinear case): Exp#1 represents the original experimental dataset which has limitation to reduce coherence; Exp#2 represents non-effective experimental dataset (inhibit x_1); and Exp#3 represents the effective experimental dataset (inhibit x_2) (a) the time series of x_1 , x_2 , x_3 for each experiment and coherence comparison for each experiment (b) the corresponding topology (c) (detail) coherence of the sensing matrix

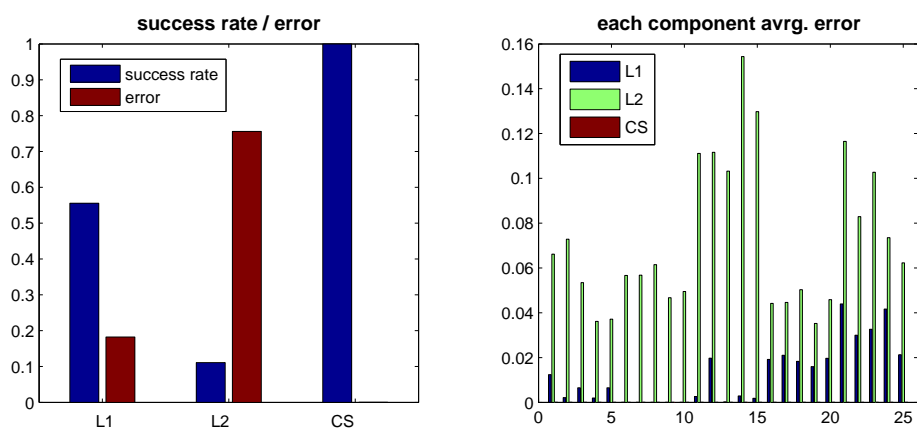
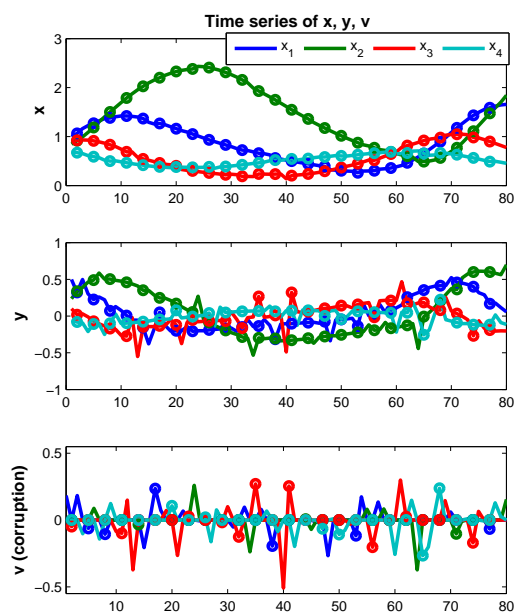
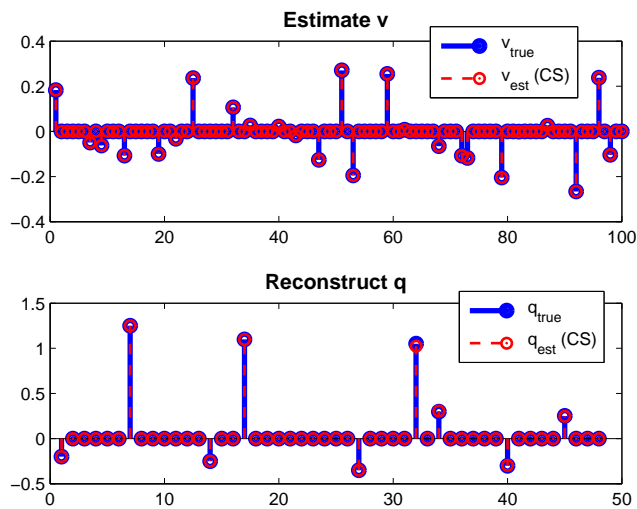


Figure 7. Comparison $L1, L2$, and CS . (Example 6, linear case) statistics of 50 trials: (left) success rate and reconstruction error (right) reconstruction error for each component.

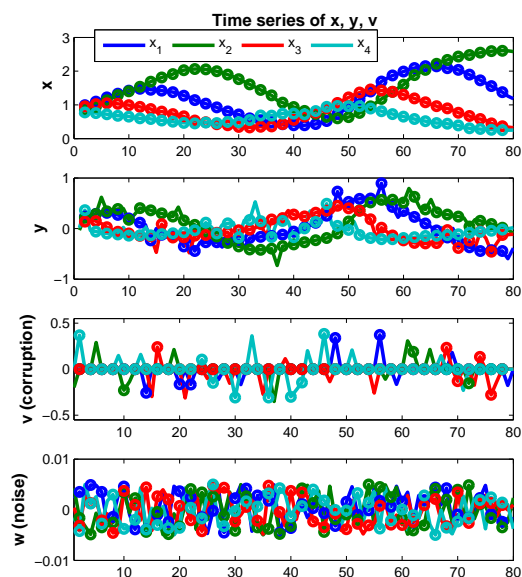


(a)

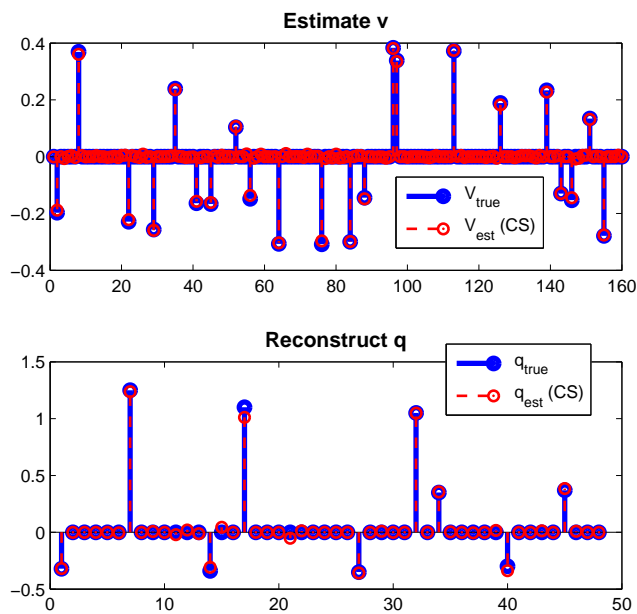


(b)

Figure 8. Arbitrary corruption with no measurement noise. (Example 7) Reconstruction with corrupted signal. (left) time series of x , y , v (right) reconstruction results of v and q where each circle represents sampled time points ($n = 4$, $M = 25$, $N = 12$, $s = 9$).



(a)



(b)

Figure 9. Arbitrary corruption with measurement noise. (Example 8) Reconstruction with corrupted signal. (left) time series of $\mathbf{x}, \mathbf{y}, \mathbf{v}$ (right) reconstruction results of \mathbf{v} and \mathbf{q} where each circle represents sample time points ($n = 4, M = 40, N = 12, s = 9$).

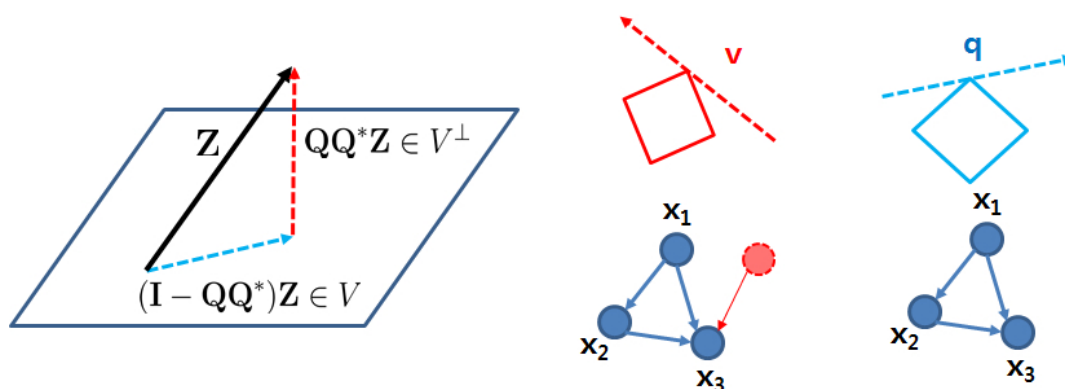


Figure 10. Geometric view of two-step refinement. A geometric view of two consecutive l_1 -norm optimizations.

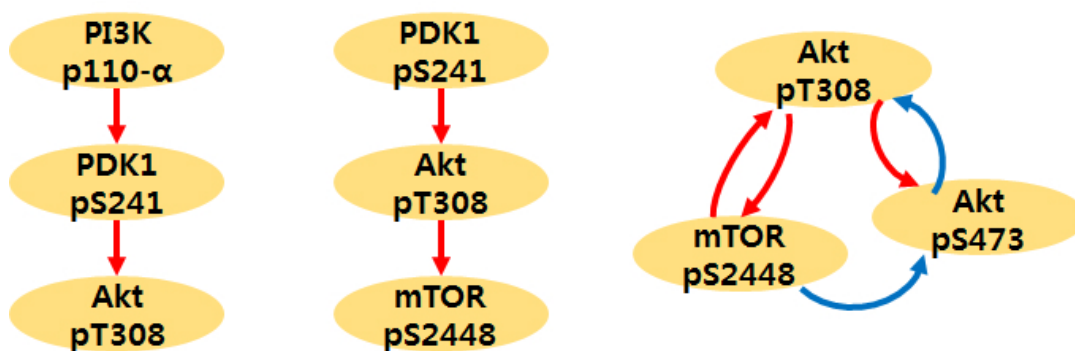


Figure 11. HER2+ overexpressed breast cancer. *CS* reconstruction result using Reverse Phase Protein Array data (SKBR3 cell line, Serum [27]). See Figure S1, S2 and S3 for further details (red: activation, blue: inhibition).

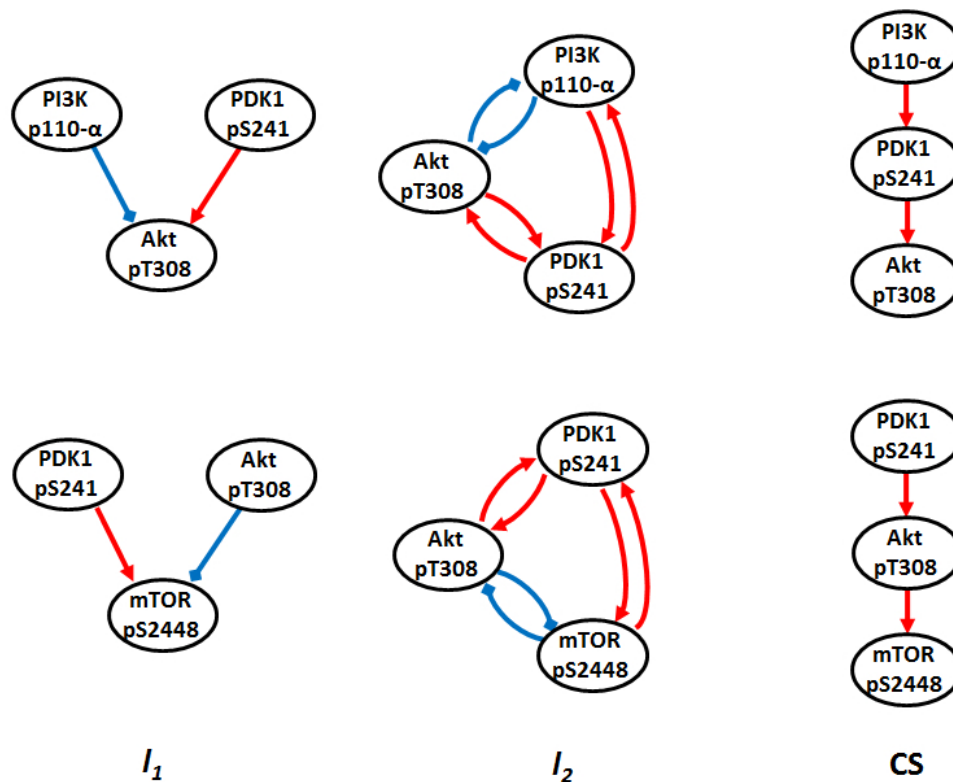
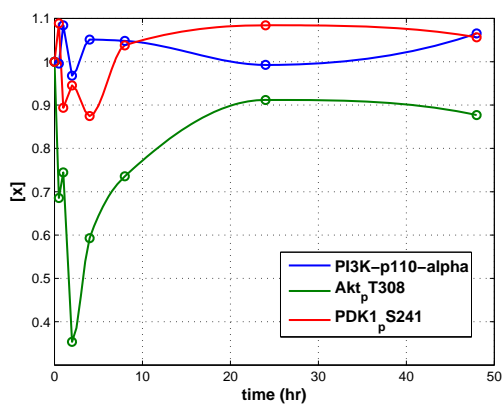
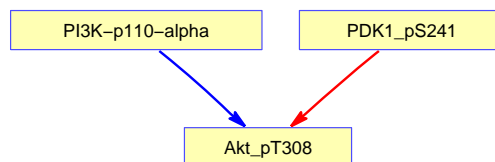


Figure 12. Sub networks inferred for the HER2/3 signaling network from Reverse Phase Protein Array data. The columns show the networks inferred by $L1$ -optimization, $L2$ -optimization, and CS. The network structures identified by CS agree with the current understanding of the network, whereas those found using $L1$ and $L2$ optimization do not. See Figure S1, S2 and S3 for further details (red: activation, blue: inhibition).

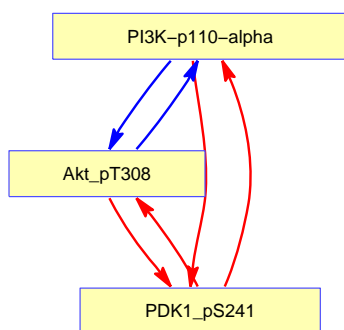
Supporting Information



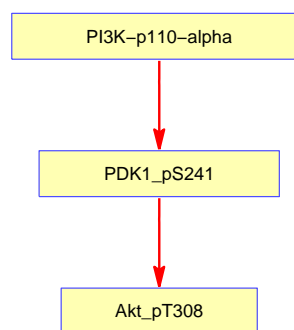
(a)



(b)



(c)



(d)

Figure S1. HER2+ overexpressed breast cancer. RPPA dataset (SKBR3 cell line, Serum [27]) (a) gene expression data [0-48hr] (b) L_1 optimization result (c) L_2 optimization result (d) CS reconstruction result.

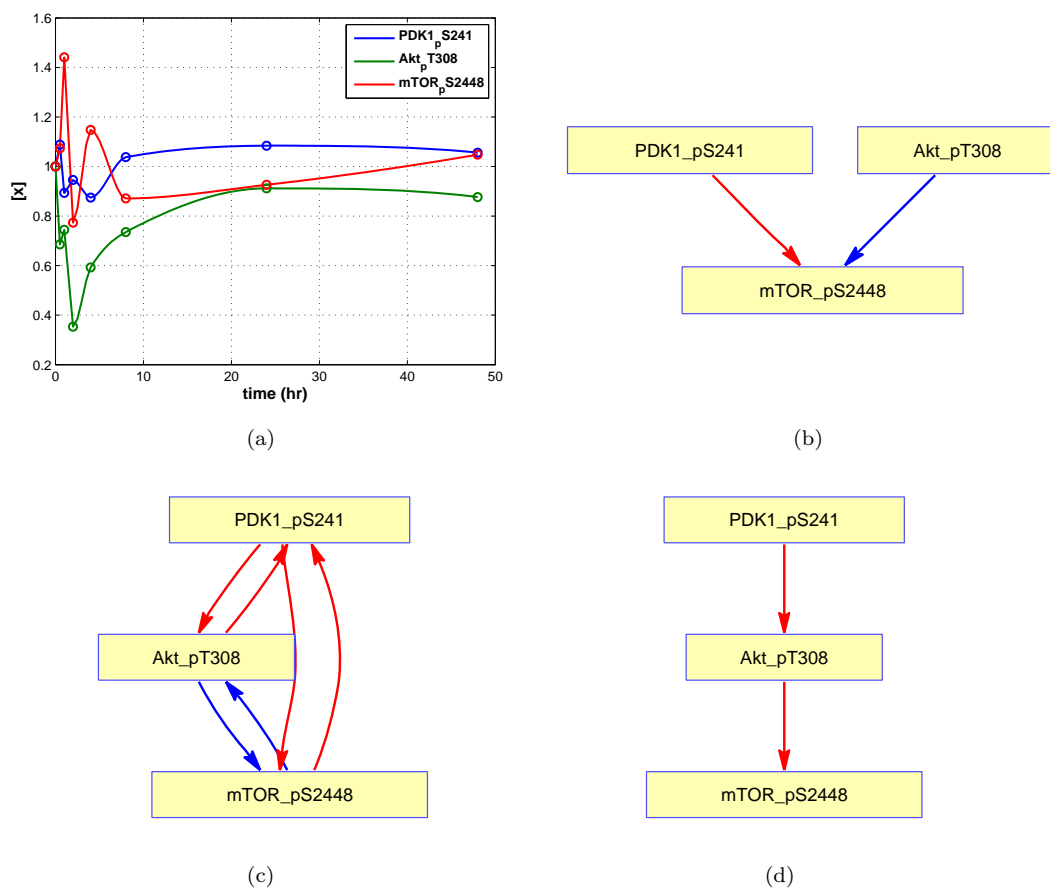
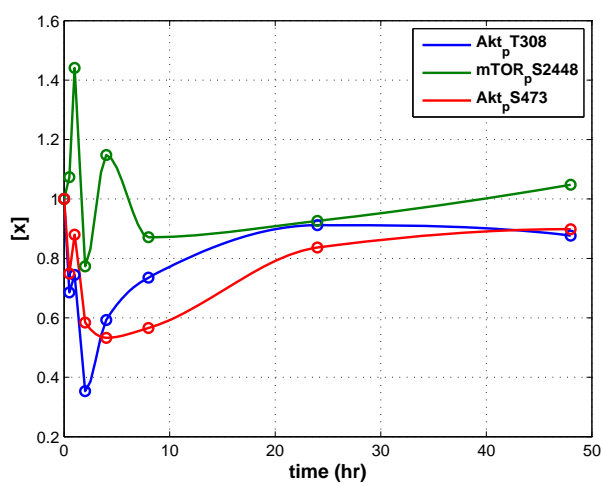
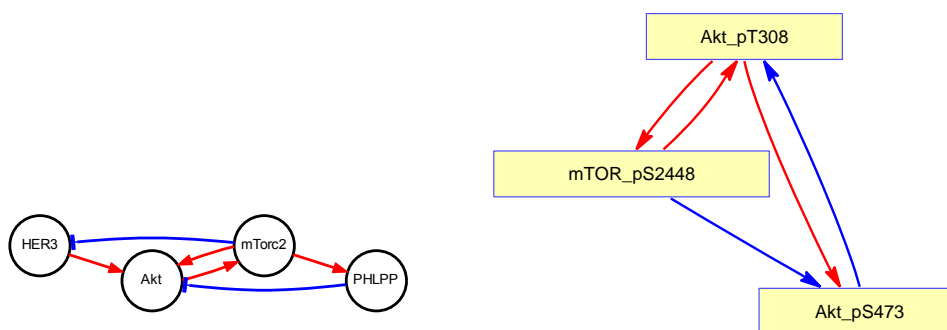


Figure S2. HER2+ overexpressed breast cancer. RPPA dataset (SKBR3 cell line, Serum [27]) (a) gene expression data [0-48hr] (b) $L1$ optimization result (c) $L2$ optimization result (d) CS reconstruction result.



(a)



(b)

(c)

Figure S3. HER2+ overexpressed breast cancer. RPPA dataset (SKBR3 cell line, Serum [27]) (a) gene expression data [0-48hr] (b) abstract model by Dr. Moasser (c) *CS* reconstruction result.