

# Inferring individual network edges - with application to target identification in gene networks

Yu Wang\* Elling W. Jacobsen\*

\* *Division of Decision and Control Systems, EECS, KTH Royal Institute of Technology, Stockholm 10044, Sweden (e-mail: wang3@kth.se, jacobsen@kth.se).*

---

**Abstract:** The paper considers the problem of inferring individual network edges from time-series data. This is the problem faced in target identification, but also important in cases where it is of interest to learn whether two specific network nodes interact directly as well as in cases where there is insufficient information to infer the full network. The proposed inference method is based on taking a geometric perspective on a corresponding regression problem. We show that, by considering the span of individual node response vectors in sample space, it is possible to identify a given edge with a label of confidence even if the available data are not informative to infer other parts of the network. Furthermore, the method points to what further experiments are needed to infer edges for which the available response data are not sufficiently informative. We demonstrate the results on a target identification problem of a nonlinear 20-gene network and show that targets can be identified independently from a single time-series experiment using significantly fewer samples than the number of nodes in the network.

*Keywords:* Network inference, individual edges, time-series data, target identification, gene regulatory networks

---

## 1. INTRODUCTION

Inferring individual network edges is a problem faced in many applications where it is of interest to learn whether two specific network nodes interact directly, a notable example being target identification [Schenone et al. (2013)]. In such applications, rather than seeking to learn the structure of the whole network, the aim is to understand the causal mechanism existing between two or more specific nodes. However, essentially all previously proposed methods for network inference aim at inferring the full network, i.e., the direct causal interaction between all measured nodes. This is true even for methods aimed at target identification; inference of the full network is a prerequisite for identification of the direct targets of external perturbations, e.g., Di Bernardo et al. (2005), Noh et al. (2018). Methods based on inference of the whole network usually require a large number of experiments that can be time-consuming and costly, in particular in biological applications. For instance, in the development of new pharmaceuticals, the experiments needed to infer the gene or protein targets of the active compound is the most time consuming and costly step with current methods [Noh et al. (2018); Di Bernardo et al. (2005)].

It is well known that when applying existing network inference methods to available response data, the resulting network model typically contains a large number of false positives and false negatives, e.g., Marbach et al. (2010). Furthermore, it is difficult to provide a label of confidence on the different inferred edges mainly due to the fact that any computation of significance levels will depend on

having the right model structure in the first place, i.e., the identified model contains no false positives or false negatives, which is self-contradicting. See e.g., Lockhart et al. (2014). Thus, there is a need for a method that can provide some measure of confidence for the existence of individual edges independently of the decision made with respect to the existence of other edges.

In this work, we propose a method for inference of individual edges, based on a geometric perspective on the network inference problem. We consider network models for which the inference problem can be written as a linear regression. Inferring individual edges in a linear model is still an open problem and a prerequisite to solve nonlinear problems. Note that also many nonlinear relations can be inferred using linear regression, e.g., Brunton et al. (2016). By considering the span of individual regressors in the sample space, we show that it is possible to infer individual edges with a label of confidence, even if the available data are not sufficiently informative to infer the full network. Thus, the available data only need to be informative for the edges of interest, while sufficient information for the remaining edges is not required. One consequence is that some individual edges can be inferred with relatively few samples as compared to the size of the full network.

We start the paper by formulating the inference of network edges as a linear regression problem. We then present a geometric perspective on the linear regression problem and the conditions proposed in Nordling and Jacobsen (2011) on the inference of individual edges for the case with no uncertainty in the data. The geometric idea is illustrated

using a simple 3-node network. Then we propose a method for inference of individual edges in the presence of data uncertainty, and illustrate the idea with the same 3-node network in the presence of Gaussian noise. Finally, we demonstrate the results on a target identification problem, using a nonlinear 20-gene network.

## 2. PROBLEM FORMULATION

We consider networks that can be described by a system of linear ordinary differential equations

$$\begin{aligned} \dot{x} &= Ax(t) + Bp(t), \\ y(t) &= x(t), \end{aligned} \quad (1)$$

where  $x(t) = [x_1(t), \dots, x_n(t)]^T \in \mathbb{R}^n$  contains the value of each node activity at time  $t$ ,  $p(t) \in \mathbb{R}^h$  is a vector of external perturbations added at time  $t$ , and  $y(t) = [y_1(t), \dots, y_n(t)]^T$  is a vector of the measurements of  $x(t)$ . A nonzero element  $a_{ij}$  in  $A \in \mathbb{R}^{n \times n}$  corresponds to an edge from node  $j$  to node  $i$ , and a nonzero element  $b_{ij}$  in  $B \in \mathbb{R}^{n \times h}$  corresponds to a direct effect of the external perturbation  $j$  on node  $i$ . For simplicity, we assume that only directly observable nodes are included in the state vector  $x(t)$ . Without loss of generality, we consider the case where only one perturbation is added, i.e.,  $h = 1$ . We first consider the noise-free case, but will later add measurement noise.

Inferring individual edges corresponds to identifying specific elements of interest in  $A$  and  $B$ , from available data. Here, we restrict our attention to the inference problem using time-series data.

With time-series data, the relation between the samples and the perturbations is given by the discretization of (1). Assume fixed sampling time  $T$  and zero-order hold, then the discretization of (1) is

$$\begin{aligned} \Delta x_{k+1} &= (A_d - I)x_k + B_d p_k, \\ y_k &= x_k, \end{aligned} \quad (2)$$

where  $\Delta x_{k+1} = x_{k+1} - x_k$ ,  $A_d = e^{AT}$ ,  $B_d = A^{-1}(A_d - I)B$ , and subscript  $k$  represents time stamp  $k$ . Note that  $A_d - I$  and  $B_d$  in general will be full matrices even if  $A$  and  $B$  are sparse matrices and hence the structure is in some sense lost in sampling. This may seem discouraging since the aim is to identify non-zero elements in  $A$  and/or  $B$ . However, consider the series expansion of  $e^{AT}$  around  $T = 0$

$$e^{AT} = I + AT + \frac{1}{2}A^2T^2 + \dots \quad (3)$$

With reasonably fast sampling, higher order terms can be neglected and we obtain  $A_d - I \approx AT$  and  $B_d \approx BT$  and hence zero elements in  $A$  and  $B$  will be relatively small in  $A_d - I$  and  $B_d$  provided the sampling is reasonably fast. Note that sampling is a problem faced by any inference method based on sampled data. With the uncertainty introduced below, these small elements can not be distinguished from other sources of uncertainty. We will hence consider the problem of inferring non-zero elements in  $A_d - I$  and  $B_d$  below, and will then define these as robustly non-zero.

We collect  $m$  samples of response data for  $n$  nodes as well as the corresponding perturbations. Introduce the time-shifted matrices

$$\begin{aligned} \Phi &= \begin{bmatrix} y_{m-1} & y_{m-2} & \dots & y_1 \end{bmatrix}^T, \\ \Xi &= [\Delta y_m \ \Delta y_{m-1} \ \dots \ \Delta y_2]^T, \end{aligned}$$

where  $\Delta y_{k+1} = y_{k+1} - y_k$ .

According to (2),  $\Phi$  and  $\Xi$  are linearly related

$$\Phi\Theta = \Xi, \quad (4)$$

where  $\Theta = [A_d - I \ B_d]^T \in \mathbb{R}^{(n+1) \times n}$  contains the unknown parameters. An edge from node  $i$  to node  $j$  corresponds to  $\theta_{ij}$ . Similarly,  $\theta_{n+1,j}$  corresponds to the direct effect of the external perturbation on node  $j$ . The relation in (4) corresponds to a linear regression and can be solved independently for each columns of  $\Theta$  and  $\Xi$

$$\Phi\theta_j = \xi_j \quad (5)$$

Inferring an individual edge from node  $k$  to node  $j$  corresponds to determining whether the parameter  $\theta_{kj} \neq 0$ .

Existing methods for fitting regression models to data, e.g., [Efron et al. (2004), Tibshirani (1996)], consider selecting and fitting the vector  $\theta_j$  based on available data, corresponding to inferring all possible edges pointing to node  $j$ . However, the fitted parameters typically contain a large fraction of false positives and false negatives [Marbach et al. (2010)]. Although some significance level of the fitted parameters can be obtained, they are based on the assumption that the inferred model is consistent, i.e., contains no false positives or negatives. If the inferred model is not consistent, the obtained significance levels can be completely misleading. See e.g., Lockhart et al. (2014). Moreover, when the available data are only informative to infer some of the parameters in  $\theta_j$ , existing methods infer all the parameters including the non-informative ones. Thus, with existing methods, it is not possible to infer individual edges in the network with any label of confidence.

## 3. A GEOMETRIC PERSPECTIVE ON LINEAR REGRESSION

To infer individual edges of interest with some label of confidence, we take a geometric perspective on the linear regression problem (5). The linear regression problem (5) can be written as

$$\sum_{k=1}^{m+1} \phi_k \theta_{kj} = \xi_j. \quad (6)$$

In (6), every regressor  $\phi_k$  and regressand  $\xi_j$  can be seen as a vector in the  $(m-1)$ -dimensional sample space. From a geometric point of view, each regressor  $\phi_i$  spans a direction in the  $(m-1)$ -dimensional sample space. With noise-free data, if the direction is unique, i.e.,  $\phi_i$  is not a linear combination of the other regressors, and the same direction is present in the regressand  $\xi_j$ , then we can conclude that the parameter  $\theta_{ij} \neq 0$ . Similarly, if the unique direction in  $\phi_i$  is not present in  $\xi_j$ , then the parameter  $\theta_{ij} = 0$ . On the other hand, if the regressor  $\phi_i$  does not span a unique direction or the regressand  $\xi_j = 0$ , it implies that the data are not sufficiently informative to decide whether  $\theta_{ij}$  is zero or not.

In the noise-free case, necessary and sufficient conditions for the inference of a given edge  $\theta_{ij}$  have been established

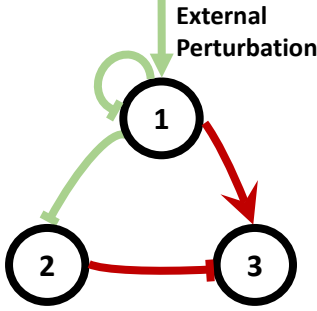


Fig. 1. A 3-node network. The external perturbation perturbs only node 1 directly. The red and green arrows correspond to the edges to be inferred and those that are not of interest, respectively.

in Nordling and Jacobsen (2011), based on linear independence of the regressors.

*Theorem 1.* Consider the linear regression problem (5). Let  $\Phi_{k \neq i}$  be the matrix obtained by removing the  $i^{\text{th}}$  column in  $\Phi$  and  $T_{k \neq i} = (\Phi_{k \neq i}^T)^{\dagger} \Phi_{k \neq i}^T$  be the projection matrix onto the linear subspace spanned by  $\Phi_{k \neq i}$ , where  $(\cdot)^{\dagger}$  denotes the pseudoinverse of a matrix. Then, the coefficient  $\theta_{ij} \forall j \in \{1, \dots, n+1\}$  can be uniquely determined if and only if  $(I - T_{k \neq i})\phi_i \neq 0$ .

Based on Theorem 1, the edges in a network can be classified into three groups

- (1) existing edges, if  $(I - T_{k \neq i})\phi_i \neq 0$  and  $(I - T_{k \neq i})\xi_j \neq 0$ ,
- (2) non-existing edges, if  $(I - T_{k \neq i})\phi_i \neq 0$  and  $(I - T_{k \neq i})\xi_j = 0$ ,
- (3) non-informative edges, if  $(I - T_{k \neq i})\phi_i = 0$  or  $\xi_j = 0$ .

To illustrate the geometric approach, we consider the inference of edges pointing to node 3 in the 3-node network illustrated in Fig. 1. Node 3 is directly regulated by node 1 and 2. We collect 4 samples from one perturbation experiment of the sparse discrete-time model of the network where node 1 is directly perturbed. The applied perturbations correspond to  $\phi_4$ . This gives the noise-free regressor and regressand

$$\Phi = \begin{bmatrix} 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 1 \end{bmatrix}, \quad \xi_3 = \begin{bmatrix} 0 \\ -1 \\ 1.5 \end{bmatrix}. \quad (7)$$

Fig. 2 illustrates the regressors and the regressand  $\xi_3$  in the 3-dimensional sample space. As seen from Fig. 2,  $\phi_1$  spans a unique direction that cannot be expressed as a linear combination of the other regressors, and the unique direction is also present in the regressand  $\xi_3$ . Therefore,  $\theta_{13} \neq 0$ , i.e., the edge from node 1 to node 3 exists, according to Theorem 1. Similarly,  $\phi_4$  spans a unique direction which it is not present in  $\xi_3$ . Therefore,  $\theta_{43} = 0$ , i.e., the external perturbation does not perturb node 3 directly. Since  $\phi_2$  and  $\phi_3$  are linearly dependent, none of them span unique directions. So the available data are not sufficiently informative to infer the existence of the corresponding edges. It implies that further experiments are needed to infer the edges from node 2 and node 3 to node 3.

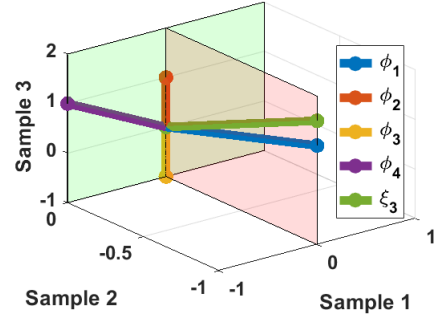


Fig. 2. The noise-free regressors  $\phi_i$  and regressand  $\xi_3$  of the 3-node network in a 3-dimensional sample space. The green and red planes are two orthogonal subspaces spanned by  $\{\phi_2, \phi_3, \phi_4\}$  and  $\{\phi_1, \xi_3\}$ .

#### 4. THE INFERENCE OF INDIVIDUAL EDGES BASED ON DATA WITH UNCERTAINTY

The approach to inferring network edges as outlined above is simple and straightforward in the case of no uncertainty. However, measurement uncertainty is likely to change the span of each regressor and regressand such that they are no longer uniquely spanning. This will in particular be true if the number of samples is less than the number of nodes, i.e.,  $m < n+1$ , and the measurements are contaminated by normally distributed noise; a random matrix in  $\mathbb{R}^{m \times (n+1)}$ , and  $m < n+1$ , will have rank  $m$  with probability 1. That is, the noise will span the full sample space and no regressor or regressand will span a unique direction when  $m < n+1$ . Thus, to utilize Theorem 1 to infer edges from measurement data, we first need to recover the linear subspace spanned by the underlying noise-free regressors, i.e., the span of  $\Phi_{k \neq i}$ . For this purpose, we below employ results from random matrix theory.

Consider adding uncertainty to the regressor matrix  $\Phi$  and regressand matrix  $\Xi$  according to

$$\hat{\Phi} = \Phi + E_{\Phi}, \quad \hat{\Xi} = \Xi + E_{\Xi} \quad (8)$$

where  $E_{\Phi} \in \mathbb{R}^{(m-1) \times (n+1)}$  and  $E_{\Xi} \in \mathbb{R}^{(m-1) \times n}$  are noise matrices with i.i.d  $\mathcal{N}(0, \sigma^2)$  elements, and  $\hat{\Phi}$  and  $\hat{\Xi}$  are the corresponding measurement matrices. Let the measurement matrix of  $\Phi_{k \neq i}$  be

$$\hat{\Phi}_{k \neq i} = \Phi_{k \neq i} + E, \quad (9)$$

where  $\hat{\Phi}_{k \neq i} \in \mathbb{R}^{(m-1) \times n}$  and  $E \in \mathbb{R}^{(m-1) \times n}$  are the matrices obtained by removing the  $i^{\text{th}}$  column in  $\hat{\Phi}$  and  $E_{\Phi}$ , respectively.

A number of methods exist to estimate the noise-free linear subspace from a measurement matrix and are commonly used in e.g., data compression and signal processing. Here we employ a method based on hard thresholding of the singular value decomposition (SVD) of the measurement matrix  $\hat{\Phi}_{k \neq i}$  to obtain an estimate of the linear subspace spanned by  $\Phi_{k \neq i}$  [Gavish and Donoho (2014)]. The basis for this method is that, provided the rank  $r$  of the noise-free matrix  $\Phi_{k \neq i}$  is known, the maximum likelihood estimate of  $\Phi_{k \neq i}$  is obtained by truncating the SVD of  $\hat{\Phi}_{k \neq i}$  to only retain the  $r$  largest singular values. An important result in Gavish and Donoho (2014) is that the singular directions for the singular values corresponding to the

removed noise will asymptotically be orthogonal to the true signal directions. Thus, the span of the maximum likelihood estimate of  $\Phi_{k \neq i}$  should also be the best estimate of the subspace spanned by  $\Phi_{k \neq i}$ .

If the true rank  $r$  of  $\Phi_{k \neq i}$  is unknown, one must first estimate the rank before truncating the SVD of  $\hat{\Phi}_{k \neq i}$  to obtain the estimate  $\check{\Phi}_{k \neq i}$ . This is equivalent to determining a threshold  $\lambda$  for which singular values to retain in the estimate

$$\check{\Phi}_{k \neq i} = \sum_{l=1}^{\min\{m-1, n\}} \hat{s}_l \mathbf{1}_{\{\hat{s}_l \geq \lambda\}} u_l v_l, \quad (10)$$

where  $\hat{s}_1 \geq \dots \geq \hat{s}_{\min\{m-1, n\}}$  are the singular values of  $\hat{\Phi}_{k \neq i}$ ,  $u_l$  and  $v_l$  are the corresponding singular vectors,  $\lambda \in \mathbb{R}$  is a threshold, and  $\mathbf{1}$  is the indicator function. Thus, singular values below the threshold  $\lambda$  are set to 0 and the estimated rank is the number of singular values larger than  $\lambda$ .

The following results on determining the optimal threshold  $\lambda$  are taken from Gavish and Donoho (2014) and are strictly based on asymptotic results for a fixed matrix dimension ratio  $\beta = (m-1)/n \leq 1$  with  $m-1, n \rightarrow \infty$ , but can be applied to given finite-size matrices.

We first consider the case with i.i.d  $\mathcal{N}(0, 1/n)$  measurement noise without loss of generality. Let  $\tilde{s}_1 \geq \dots \geq \tilde{s}_{\min\{m-1, n\}}$  and  $s_1 \geq \dots \geq s_{\min\{m-1, n\}}$  denote the singular values of the noise matrix  $E$  and the noise-free matrix  $\Phi_{k \neq i}$ , respectively. Then, the following properties can be proved [Gavish and Donoho (2014)].

*Theorem 2.* The empirical distribution of the singular values of the noise matrix  $E$  converges almost surely to a non-random distribution and  $\tilde{s}_1 \xrightarrow{a.s.} 1 + \sqrt{\beta}$ , as  $n \rightarrow \infty$ .

*Theorem 3.* For  $1 \leq l \leq r$ ,

$$\lim_{n \rightarrow \infty} \hat{s}_l = \begin{cases} \sqrt{(s_l + \frac{1}{s_l})(s_l + \frac{\beta}{s_l})} & s_l > \beta^{\frac{1}{4}} \\ 1 + \sqrt{\beta} & s_l \leq \beta^{\frac{1}{4}} \end{cases} \quad a.s. \quad (11)$$

Theorem 2 gives an asymptotic upper bound on the singular values of noise matrices  $E$ . Theorem 3 gives the asymptotic singular values of  $\hat{\Phi}_{k \neq i}$ . It implies that singular values of  $\Phi_{k \neq i}$  smaller than  $\beta^{\frac{1}{4}}$  cannot be distinguished from noise singular values, since the corresponding singular values  $\hat{s}_l$  of  $\hat{\Phi}_{k \neq i}$ , lie on the asymptotic upper bound of the singular values of  $E$ , i.e.,  $1 + \sqrt{\beta}$ . Hence, only singular values  $\hat{s}_l > 1 + \sqrt{\beta}$  can be distinguished from noise. From this it seems reasonable to choose the threshold  $\lambda = 1 + \sqrt{\beta}$ . However, as shown in Gavish and Donoho (2014), this threshold will not give the maximum likelihood estimator of  $\Phi_{k \neq i}$ . Rather, based on Theorem 2 and 3, the optimal threshold in the maximum likelihood sense can be shown to be given by [Gavish and Donoho (2014)]

$$\lambda_* = \sqrt{2(\beta + 1) + \frac{8\beta}{(\beta + 1) + \sqrt{\beta^2 + 14\beta + 1}}}. \quad (12)$$

When the entries of the noise matrix  $E$  are i.i.d  $\mathcal{N}(0, \sigma^2)$ , the optimal denoised estimate of  $\Phi_{k \neq i}$  is given by

$$\check{\Phi}_{k \neq i}^* = \sum_{l=1}^{\min\{m-1, n\}} \hat{s}_l \mathbf{1}_{\{\hat{s}_l \geq \sqrt{n}\sigma\lambda_*\}} u_l v_l, \quad (13)$$

Thus, the optimal estimate of the linear subspace spanned by  $\Phi_{k \neq i}$  is given by the linear subspace spanned by  $\check{\Phi}_{k \neq i}^*$ .

In order to apply Theorem 1 to infer individual edges based on projections, we also need to determine whether the residuals after projecting  $\phi_i$  and  $\xi_j$  onto  $\check{\Phi}_{k \neq i}$  are above the noise level or not. For this, we employ the optimal threshold in (13) for a  $(m-1)$ -dimensional vector with i.i.d additive Gaussian noise  $\mathcal{N}(0, \sigma^2)$

$$\tau = \sigma \sqrt{2(\bar{\beta} + 1) + \frac{8\bar{\beta}}{(\bar{\beta} + 1) + \sqrt{\bar{\beta}^2 + 14\bar{\beta} + 1}}}, \quad (14)$$

where  $\bar{\beta} = \frac{1}{m-1}$ . We consider that a  $(m-1)$ -dimensional vector can be distinguished from noise if its 2-norm is larger than the optimal threshold  $\tau$  in the maximum likelihood sense.

#### 4.1 Main results and algorithm

Based on the above results, we propose a method to infer individual edges with a label of confidence in the presence of uncertainty. The proposed method can infer a given edge independently of the inference of other edges. Moreover, the provided data only need to be sufficiently informative for the edges of interest. If the data are not sufficiently informative, the method also provides information on what further experiments are needed to provide the required data.

From a geometric point of view, each regressor  $\hat{\phi}_i$  spans a direction in the  $(m-1)$ -dimensional sample space. Essentially, if the direction is robustly unique, i.e.,  $\hat{\phi}_i$  does not lie in the linear subspace spanned by  $\check{\Phi}_{k \neq i}^*$  and the unique direction obtained by the projection residual of  $\hat{\phi}_i$  onto the linear subspace spanned by  $\check{\Phi}_{k \neq i}^*$  can be distinguished from noise, we can conclude that the direction spanned by  $\hat{\phi}_i$  is robustly unique. If  $\hat{\phi}_i$  spans a robustly unique direction and this direction is robustly present in the regressand  $\hat{\xi}_j$ , i.e., the projection of  $\hat{\xi}_j$  onto the unique direction is above the noise threshold, then we can conclude that the parameter is robustly non-zero  $\theta_{ij} \neq 0$ . Similarly, if the robustly unique direction of  $\hat{\phi}_i$  present in  $\hat{\xi}_j$  cannot be distinguished from noise, then we say the parameter is robustly zero  $\theta_{ij} = 0$ . On the other hand, if the regressor  $\hat{\phi}_i$  does not span a robustly unique direction or the regressand  $\hat{\xi}_j$  cannot be distinguished from a noise vector, it implies that the available data are not sufficiently informative to decide whether  $\theta_{ij}$  is zero or not.

The above can be summarized in the form of the an inference algorithm. Here we assume the noise level  $\sigma$  is known. If the noise level is unknown, the noise level should be estimated before applying the inference algorithm below. See e.g., Gavish and Donoho (2014).

- (1) Determine the thresholds  $\sqrt{n}\sigma\lambda_*$  and  $\tau$ , according to (12) and (14), respectively.
- (2) Determine the optimal estimate of the linear subspace spanned by  $\Phi_{k \neq i}$ , i.e., the span of  $\check{\Phi}_{k \neq i}^*$ , according to (13).
- (3) Determine the unique direction spanned by regressor  $\hat{\phi}_i$  using the projection residual of  $\hat{\phi}_i$  onto the span of  $\check{\Phi}_{k \neq i}^*$ . The projection residual of  $\hat{\phi}_i$  onto the span of

$\hat{\Phi}_{k \neq i}^*$  is given by  $q = (I - \hat{T}_{k \neq i})\hat{\phi}_i$ , where  $\hat{T}_{k \neq i} = ((\hat{\Phi}_{k \neq i}^*)^T)^\dagger \hat{\Phi}_{k \neq i}^*$  is the projection matrix onto the linear subspace spanned by  $\hat{\Phi}_{k \neq i}^*$ . If  $\|q\|_2 > \tau$ , it implies the regressor  $\hat{\phi}_i$  robustly spans a unique direction.

- (4) Classify  $\theta_{ij}$  according to the following conditions.
- If  $\|q\|_2 \leq \tau$  or  $\|\hat{\xi}_j\|_2 \leq \tau$ , it implies that the data are non-informative for inferring  $\theta_{ij}$ .
  - If  $\|q\|_2 > \tau$  and  $\mathbf{proj}_q(\hat{\xi}_j) = \frac{\langle \hat{\xi}_j, q \rangle}{\langle q, q \rangle} > \tau$ , it implies that  $\theta_{ij} \neq 0$ , i.e., robustly non-zero.
  - If  $\|q\|_2 > \tau$  and  $\mathbf{proj}_q(\hat{\xi}_j) = \frac{\langle \hat{\xi}_j, q \rangle}{\langle q, q \rangle} \leq \tau$ , it implies that  $\theta_{ij} = 0$ , i.e., robustly zero.

We define the confidence score of existing edges as

$$\gamma = \min\left\{\frac{\mathbf{proj}_q(\hat{\xi}_j)}{\tau}, \frac{\|q\|_2}{\tau}\right\}.$$

If the confidence score  $\gamma > 1$ , it implies that  $\theta_{ij} \neq 0$ , while  $\|q\|_2/\tau > 1$  and  $\gamma \leq 1$  implies  $\theta_{ij} = 0$ . The confidence in these decisions increases with increasing  $\gamma$  and  $1/\gamma$ , respectively. If  $\|q\|_2/\tau \leq 1$  or  $\|\hat{\xi}_j\|_2/\tau \leq 1$ , it implies that the available data are non-informative for inferring the edge  $\theta_{ij}$ .

Note that the proposed conditions are neither sufficient nor necessary. This is mainly due to the conservative estimate  $\hat{\Phi}_{k \neq i}^*$ . Note, however, that when the number of samples is smaller than the number of nodes, no method can provide sufficient and necessary conditions.

#### 4.2 Illustrating example

We revisit the 3-node network studied above to illustrate the proposed method. As before, the edges pointing to node 3 are assumed to be the edges of interest. To include uncertainty we add measurement noise.

Adding additive Gaussian noise with standard deviation  $\sigma = 0.1$ , corresponding to 15% of the standard deviation of the signal magnitude, to the noise-free matrices in (7), gives the regressor and regressand

$$\hat{\Phi} = \begin{bmatrix} -0.0282 & 0.0138 & 0.0780 & -1.1242 \\ -0.9430 & 0.1184 & 0.0629 & -0.1601 \\ 0.8990 & 0.9743 & -1.1267 & 1.0067 \end{bmatrix}, \quad (15)$$

$$\hat{\xi}_3 = [-0.0152 \quad -1.1896 \quad 1.4534]^T.$$

Applying the proposed algorithm, we obtain the results in Table 1. We can then conclude that the given data are

Table 1. Identification results of edges pointing to node 3 in the network illustrated in Fig. 1 in the presence of measurement noise using the proposed algorithm. N and E.P. are node and external perturbation respectively, while E., N.E. and N.I. are existing, non-existing and non-informative edges respectively.

|                          | N1 $\rightarrow$ N3 | N2 $\rightarrow$ N3 | N3 $\rightarrow$ N3 | E.P. $\rightarrow$ N4 |
|--------------------------|---------------------|---------------------|---------------------|-----------------------|
| $\ q\ _2/\tau$           | 3.48                | 0                   | 0                   | 4.00                  |
| $\ \hat{\xi}_j\ _2/\tau$ | 6.88                | 6.88                | 6.88                | 6.88                  |
| $\gamma$                 | 3.48                | -                   | -                   | 0.15                  |
| Result                   | E.                  | N.I.                | N.I.                | N.E.                  |

sufficiently informative to infer the edges from node 1 and

the external perturbation to node 3, while they are not sufficiently informative to infer the others. The edge from node 1 to node 3 exists with a confidence score  $\gamma = 3.48$ , and node 3 is not directly perturbed by the perturbation.

This simple example serves to illustrate that the proposed method is able to infer specific edges with some label of confidence, even if the available data are not sufficiently informative to infer the full network.

#### 5. APPLICATION: TARGET IDENTIFICATION IN DRUG DISCOVERY

Target identification is an important step in drug discovery based on phenotypic assays, and one of the most time consuming and costly steps in developing new pharmaceuticals [Schenone et al. (2013)]. Target identification determines the direct gene of protein targets of the compound of interest (COI), while the remaining interactions in the gene regulatory network (GRN) are of less concern.

Consider a GRN with  $n$  genes described by

$$\frac{dx_i(t)}{dt} = u_i \prod_{j=1}^n x_j(t)^{A_{ij}} - d_i x_i(t), \quad (16)$$

where  $x_i(t)$  denotes the mRNA concentration of gene  $i$  at time  $t$ ,  $u_i$  and  $d_i$  are the mRNA transcription and degradation rate constants of gene  $i$  respectively.  $A_{ij}$  denotes the regulatory control of gene  $j$  on gene  $i$ . Assume the considered COI only perturbs  $u_i$ . By taking a log2fc transformation, we obtain the model

$$\frac{dz_i}{dt} = d_i \sum_{j=1}^n A_{ij} z_j - d_i z_i + d_i \log_2 \frac{u_i}{u_{i0}}, \quad (17)$$

where  $z_j = \log_2(x_i/x_{i0})$ ,  $x_{i0}$  and  $u_{i0}$  are the steady-state mRNA concentration and the transcription rates of gene  $i$  respectively. Target identification determines which genes are perturbed by the COI, i.e.,  $d_i \log_2 \frac{u_i}{u_{i0}} \neq 0$ .

The effectiveness of the proposed method is here demonstrated on a 20-gene network. The considered COI perturbs the mRNA transcription rate constants of genes 5 and 15 by increasing and decreasing them by 50%, respectively. We conduct an *in-silico* time-series experiment with  $T = 1s$  and obtain 5 log2fc samples containing 30% measurement noise, i.e., additive Gaussian noise with 30% standard deviation relative to the log2fc signal magnitude.

Fig. 3 shows the computed confidence scores for the different genes as target, using the algorithm devised above. If a confidence score exceeds 1, the corresponding gene is deemed a target of the COI. Note that the most informative genes here correspond to the target genes, genes 5 and 15, and these are correctly inferred as targets. There are no false positives. Also note the proposed method shows that the available data do not contain sufficient information to infer whether the non-informative genes are targets or not, and further experiments are hence needed.

As a comparison, we consider Lasso [Tibshirani (1996)], the basis of many widely applied methods, that is based on penalizing the  $L_1$  norm of the obtained coefficients. To determine whether a gene is target or not using Lasso, all possible edges pointing to the gene of interest are needed to be inferred.

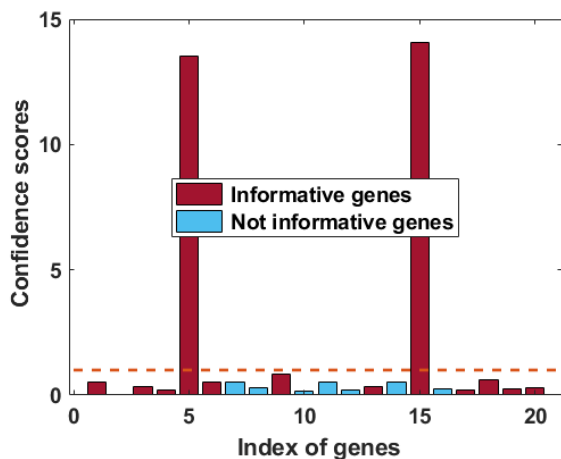


Fig. 3. Confidence scores for targets of 20 gene network using 5 time samples with 30% measurement noise. The red (blue) bars correspond to genes for which the available data are (not) sufficiently informative to determine whether they are targets.

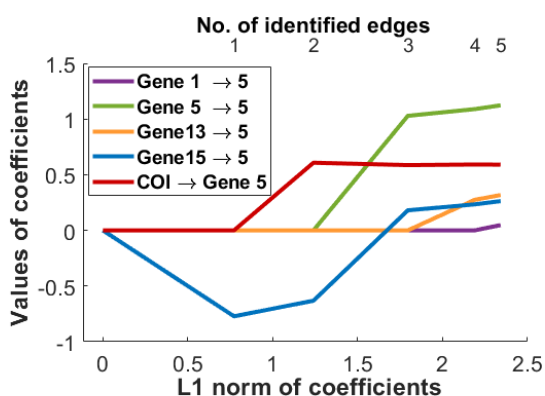


Fig. 4. Lasso trace plots of the values of the coefficients corresponding to the edges pointing to gene 5 against the  $L_1$  norm of the coefficients.

We here consider gene 5 as an example. Gene 5 is directly regulated by gene 5 and 10, and directly perturbed by the COI. Fig. 4 shows the complete Lasso solution path obtained from LARS [Efron et al. (2004)]. Note that only 5 edges can be inferred along the obtained complete Lasso solution path, since 5 regressors can span the full 5-dimensional sample space. Fig. 5 illustrates the inferred edges pointing to gene 5 using Lasso. Note that the non-existing edge from gene 15 to gene 5 is always inferred prior to the edge from the COI to gene 5, and false positives and false negatives occur using Lasso. This implies that the resulting network using Lasso can be misleading.

The target identification example demonstrates that the proposed method offers an efficient and attractive alternative to target identification.

## 6. SUMMARY AND CONCLUSIONS

Methods for inferring individual network edges are needed for applications like target identification and other prob-

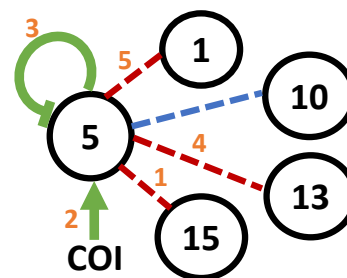


Fig. 5. Inferred edges pointing to gene 5 using Lasso. The green arrows correspond to the edges that can be inferred correctly; the blue and red dash lines correspond to possible false negatives and false positives respectively. The yellow numbers indicate the order of inferred edges along the Lasso solution path obtained from LARS.

lems where one is mainly seeking to determine the direct interactions between a selected set of nodes, but also when available data are not sufficiently informative to infer the complete network structure. In this work, we propose a method based on a geometric approach to network inference based on regression. Instead of fitting available data to a full network model, the proposed method can identify individual edges independently of the rest of the network with a label of confidence, even if the data are not sufficiently informative to infer the rest of the network. Furthermore, the method points to what further experiments are needed to infer non-informative edges.

## REFERENCES

- Brunton, S.L. et al. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *PNAS*, 113(15), 3932–3937.
- Di Bernardo, D. et al. (2005). Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nature Biotechnology*, 23(3), 377.
- Efron, B. et al. (2004). Least angle regression. *Annals of Statistics*, 32(2), 407–499.
- Gavish, M. and Donoho, D.L. (2014). The optimal hard threshold for singular values is  $4/\sqrt{3}$ . *IEEE Transactions on Information Theory*, 60(8), 5040–5053.
- Lockhart, R. et al. (2014). A significance test for the lasso. *Annals of Statistics*, 42(2), 413.
- Marbach, D. et al. (2010). Revealing strengths and weaknesses of methods for gene network inference. *PNAS*, 107(14), 6286–6291.
- Noh, H. et al. (2018). Network perturbation analysis of gene transcriptional profiles reveals protein targets and mechanism of action of drugs and influenza a viral infection. *Nucleic Acids Research*, 46(6), e34.
- Nordling, T.E. and Jacobsen, E.W. (2011). On sparsity as a criterion in reconstructing biochemical networks. In *Proceedings of 18th IFAC World Congress*. Milan.
- Schenone, M. et al. (2013). Target identification and mechanism of action in chemical biology and drug discovery. *Nature Chemical Biology*, 9(4), 232.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.