Robust reconstruction of sparse network dynamics

November 14, 2014

Outline

Outline

1. Problem formulation and contributions
2. General setting for direct influence inference methods
3. Results

Problem formulation and contributions

- Problem: Direct causal inference using noisy transient time series data
- Contributions:
  - A method to reconstruct dynamical system from transient time series that has the following that guarantee sparsity, robust to noise.
  - Analytical bounds on constraints of the $l_1 - \min$ formulation, make use of standard solvers.

Outline

Summary of three direct influence inference methods

- Feizi et. al. (Nature 1):

$$S(I+G)=G \tag{1}$$

S : direct influence
G: total influence, observable/computable

- Barabasi (Nature 2): $dx_i/dt = f(x_1, ..., x_N)$, consider at local of equilibrium, compute $(S = \partial x_i/\partial x_j)$?

$$SG = G - I + \mathscr{D}(SG) \tag{2}$$

- Sontag:

$$d\boldsymbol{x}/dt = \boldsymbol{f}(\boldsymbol{x},\boldsymbol{p}), \tag{3}$$

$$P = SR \tag{4}$$

$$\begin{aligned} R_{ij}(t) &\approx x_i(t, p_j + \Delta p_j) - x_i(t, p_j) \\ P_{ij}(t) &\approx (R_{ij}(t+\Delta t) - R_{ij}(t))/\Delta t \end{aligned}$$

Relaxation

$$S = \arg\min_{S} ||S||_{l_1} \text{ s.t } ||SB - C||_F \leq \varepsilon \quad (5)$$

- Feizi et. al (nature 1): $B = (I + G), C = G$
- Barabasi (nature 2): $B = C = G$ for off diagonal elements
- Sontag: $B = R, C = P$

1. Can solution of (5) solve (2-4)?
2. Estimate ε when known noise bound of the perturbation/quantization error on B, C?

Outline

Relaxation - deconvolution

$$S^0(I+G^0) \quad = \quad G^0 \tag{6}$$

- Noise, quantization error, limit data size: $G = G + \Delta G \rightarrow \Delta S$

$$(S^0+\Delta S)(G^0+\Delta G+I) \quad = \quad G \tag{7}$$

$$\Rightarrow S^0(G+I)-G \quad = \quad -\Delta SG - \Delta S \tag{8}$$

- In general $S^0(G+I)-G \neq 0$

$$\min |S|_{l_1} s.t. ||S(G+I)-G||_F \leq \varepsilon \tag{9}$$

or equivalently

$$\min_{\boldsymbol{s}_i} ||\boldsymbol{s}_i||_1 \text{ s.t } ||(G+I)^T \boldsymbol{s}_i - \boldsymbol{g}_i||_2 \leq \varepsilon \tag{10}$$

Constraint bounds

1. Representation using rows

$$\varepsilon_i = (||\Delta G \boldsymbol{s}_i^0||_2 + ||\boldsymbol{g}_i - \boldsymbol{g}_i^0||_2)^2 \leq 2(||(\Delta G)_i||_2^2 + \\ + (||\Delta G||_F \frac{1}{\sqrt{1-\delta_K}} ||\boldsymbol{g}_i^0||_2)^2) \tag{11}$$

2. Representation using matrices
    1. Let γ be the largest eigenvalue of ΔG

    $$||\Delta SG + \Delta S||_F \leq \varepsilon_1 = (1 + ||G||_F)\gamma \tag{12}$$

    2. If $1 - ||G||_F - ||\Delta G||_F > 0$ and $1 - ||G||_F > 0$

    $$||\Delta SG + \Delta S||_2 \leq \varepsilon_2 = (1 + ||G||_F) \frac{||\Delta G||_F}{(1 - ||G||_F - ||\Delta G||_F)(1 - ||G||_F)} \tag{13}$$

Robustness of the $l_1 - \min$ method

Theorem

The error of the l_1 - min approach is bounded by

$$||S_{l_1} - S^0||_F^2 \leq C_1 \mathscr{E} \tag{14}$$

where $\mathscr{E}$ is an upper bound of the total perturbation estimated by any of (11),(12),(13)

Compare l_1 – min using different bounds and network deconvolution

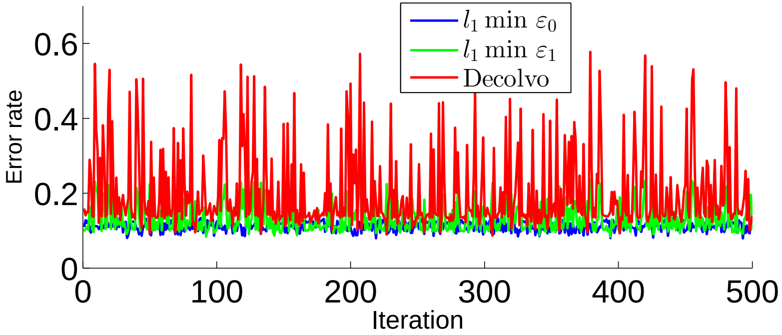

Figure: Reconstruction error

Compare $l_1 - \min$ using different bounds and network deconvolution

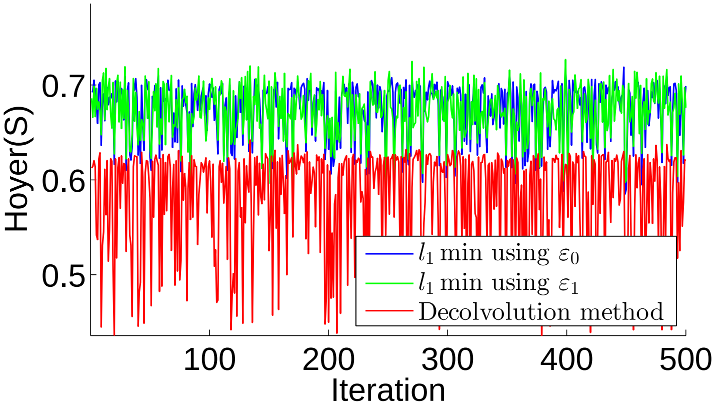

Figure: Sparsity comparision

Compare $l_1 - \min$ using different bounds and network deconvolution

- Error rate: $\frac{||\hat{S}-S^0||_F}{||S^0||_F}$ where $\hat{S}$ is computed using different methods.

Method	ε	Mean	Var
$l_1 - \min$	$\varepsilon_0 = ||(G+I)^T \mathbf{s}_i^0 - \mathbf{g}_i||_2$	0.111	1.62×10^{-4}
$l_1 - \min$	$\varepsilon = \varepsilon_1$(Eq. 12)	0.135	0.002
$l_1 - \min$	$\varepsilon = \varepsilon_2$ (Eq. 13)	0.73	0.97
Network	deconvolution	0.203	0.016

Table: Performance comparision

Compare $l_1 -$ min using different bounds and Network deconvolution

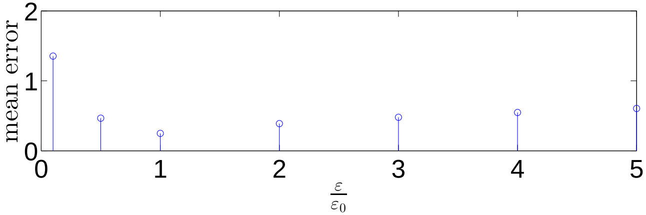

Figure: Reconstruction error trends linearly with ε

- $\frac{\varepsilon}{\varepsilon_0} > 1$: the reconstruction error trends linearly with ε.
- $\frac{\varepsilon}{\varepsilon_0} < 1$, as it decreases toward 0, the reconstruction error increases. When it approaches 0, the model is overfitted.

Effect of training data size to $l_1 - \min$ problem

	ε_0	$2\varepsilon_0$	$0.5\varepsilon_0$	$0.1\varepsilon_0$
50	2.10	2.11	2.24	2.48
60	1.71	1.78	1.89	2.15
70	1.46	1.59	1.65	2.00
80	1.21	1.42	1.43	1.86
90	0.89	1.21	1.11	1.54
100	0.42	0.96	0.64	1.36

Table: The dependence of accuracy $||S - S^0||_F$ on training data length

- The average of the error $||S - S^0||_F$ of the NDM is 1.55
- Do not need full total influence matrix G to get a good approximation of the network.
- Example: $\varepsilon = \varepsilon_0$, we only need 70% of the total influence matrix G to get an equivalent accuracy with the NDM.

Dynamical network: Kuramoto model on scale free network

$$\dot{x}_i(t) = w_i + \sum_j a_{ij} \sin(x_j - x_i), i = 1..n \quad (15)$$

$$\dot{x}_i(t) = f_i(\boldsymbol{x}, \boldsymbol{p}_i), i = 1..n \quad (16)$$

Corollary

$$(\Delta G)_{ij}(t) = (e_{ij}^{(1)}(t) - e_{ij}^{(2)}(t))_{n\times n} \quad (17)$$

$$(\Delta P)_{ij}(t) = (e_{ij}^{(1)}(t) - e_{ij}^{(2)}(t))_{n\times n}/\Delta t \quad (18)$$

where $e_{ij}^{(1)}(t), e_{ij}^{(2)}(t)$ *are error incurred when measuring* $x_i^0(t, p_j), x_i^0(t, p_j + \Delta p_j)$ *respectively.*

$$P = SR \quad (19)$$

Problems to solve

- Robust analysis for general setting
- Robustness with respect to x_i in general case when G is computed from x_i.
- Kuramoto model:
  - Computing $\frac{\partial f_i}{\partial x_i}$ is imprecise, most of the case equal to 0.
  - Evaluate the performance: compare $S_{l_1}(t), S_{decolvo}(t)$ with $S^0(t)$?