Sample Complexity for Topology Estimation in Networks of LTI Systems

Vincent Tan*† and Alan Willsky†

Department of Electrical and Computer Engineering,
University of Wisconsin-Madison*

Department of Electrical Engineering and Computer Science,
Massachusetts Institute of Technology†

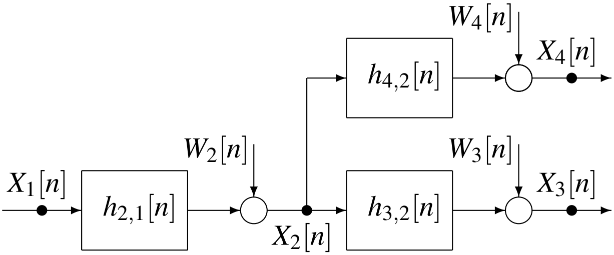
CDC (Dec 2011)

Motivation I

- Tree network $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with p nodes and $p - 1$ directed arcs

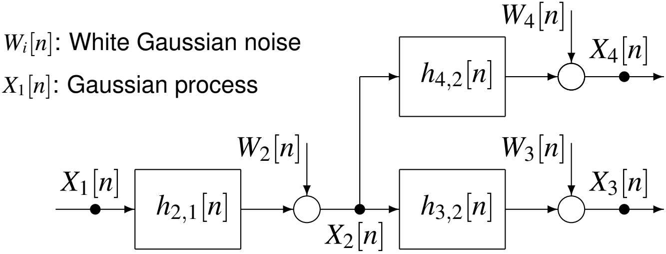
Motivation I

- Tree network $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with p nodes and $p-1$ directed arcs



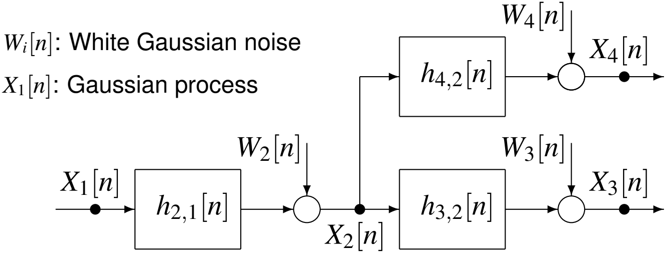
Motivation I

- Tree network $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with p nodes and $p-1$ directed arcs



Motivation I

- Tree network $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with p nodes and $p-1$ directed arcs



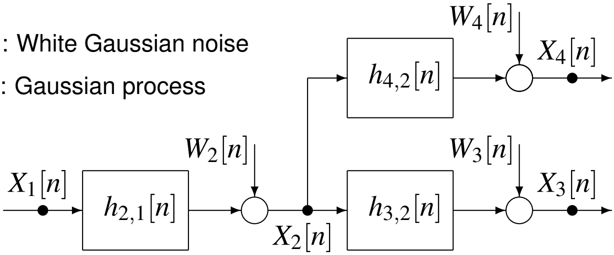
- $p = 4$, $\mathcal{V} = \{1, 2, 3, 4\}$ and $\mathcal{E} = \{(1, 2), (2, 3), (2, 4)\}$

Motivation I

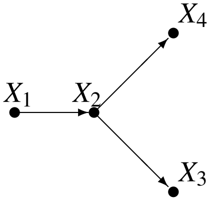
- Tree network $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with p nodes and $p-1$ directed arcs

$W_i[n]$: White Gaussian noise

$X_1[n]$: Gaussian process

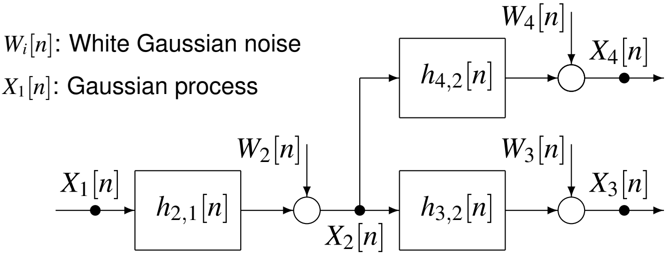


- $p = 4$, $\mathcal{V} = \{1, 2, 3, 4\}$ and $\mathcal{E} = \{(1, 2), (2, 3), (2, 4)\}$

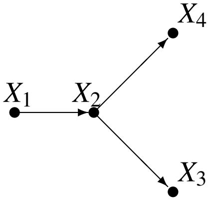


Motivation I

- Tree network $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with p nodes and $p-1$ directed arcs

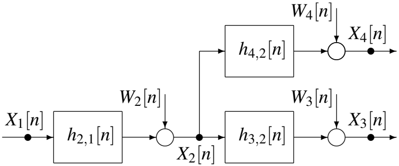


- $p = 4$, $\mathcal{V} = \{1,2,3,4\}$ and $\mathcal{E} = \{(1,2),(2,3),(2,4)\}$



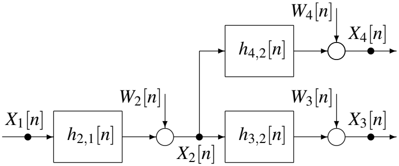
$$P_{X_1,X_2,X_3,X_4} = P_{X_1} P_{X_2|X_1} P_{X_3|X_2} P_{X_4|X_2}$$

Motivation II



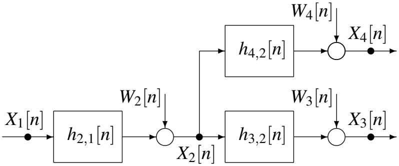
- Observe discrete-time process $X_i := \{X_i[n]\}_{n=0}^{\infty}$ at each node $i \in \mathcal{V}$

Motivation II



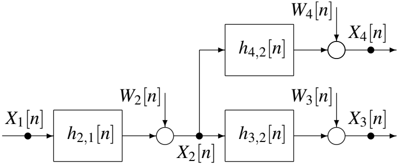
- Observe discrete-time process $X_i := \{X_i[n]\}_{n=0}^{\infty}$ at each node $i \in \mathcal{V}$
- How long do we need to observe the processes $(X_1, \ldots, X_p)$ in order to get a good estimate of the undirected tree network $\mathcal{T}$?

Motivation II



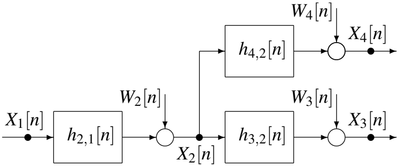
- Observe discrete-time process $X_i := \{X_i[n]\}_{n=0}^{\infty}$ at each node $i \in \mathcal{V}$
- How long do we need to observe the processes $(X_1, \ldots, X_p)$ in order to get a good estimate of the undirected tree network $\mathcal{T}$?
- Materassi and Innocenti (TAC-2011) provided an algorithm

Motivation II



- Observe discrete-time process $X_i := \{X_i[n]\}_{n=0}^{\infty}$ at each node $i \in \mathcal{V}$
- How long do we need to observe the processes $(X_1, \ldots, X_p)$ in order to get a good estimate of the undirected tree network $\mathcal{T}$?
- Materassi and Innocenti (TAC-2011) provided an algorithm
- We analyze a related algorithm here

Motivation II



- Observe discrete-time process $X_i := \{X_i[n]\}_{n=0}^{\infty}$ at each node $i \in \mathcal{V}$
- How long do we need to observe the processes $(X_1, \ldots, X_p)$ in order to get a good estimate of the undirected tree network $\mathcal{T}$?
- Materassi and Innocenti (TAC-2011) provided an algorithm
- We analyze a related algorithm here
- Many applications in system identification and model selection

Relation to Graphical Models

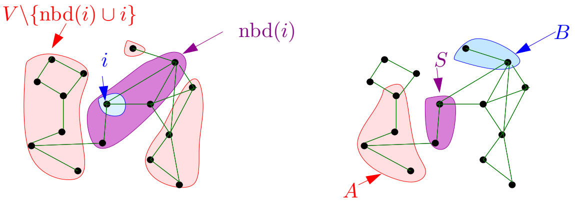
- Intimately related to graphical models in machine learning

Relation to Graphical Models

- Intimately related to graphical models in machine learning
- Graph structure $G = (\mathcal{V}, \mathcal{E})$ represents the joint distribution of a random vector $(X_1, \ldots, X_p)$: $\mathcal{V} = \{1, \ldots, p\}$ and $\mathcal{E} \subset \binom{\mathcal{V}}{2}$
- Node $i \in \mathcal{V}$ corresponds to random variable/process X_i.
- Edge set $\mathcal{E}$ corresponds to conditional independencies.

Relation to Graphical Models

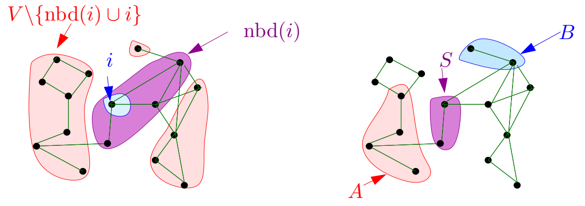
- Intimately related to graphical models in machine learning
- Graph structure $G = (\mathcal{V}, \mathcal{E})$ represents the joint distribution of a random vector $(X_1, \ldots, X_p)$: $\mathcal{V} = \{1, \ldots, p\}$ and $\mathcal{E} \subset \binom{\mathcal{V}}{2}$
- Node $i \in \mathcal{V}$ corresponds to random variable/process X_i.
- Edge set $\mathcal{E}$ corresponds to conditional independencies.



$X_i \perp\!\!\!\perp \mathbf{X}_{V\backslash\{\mathrm{nbd}(i)\cup i\}} | \mathbf{X}_{\mathrm{nbd}(i)}$ $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B | \mathbf{X}_S$

Relation to Graphical Models

- Intimately related to graphical models in machine learning
- Graph structure $G = (\mathcal{V}, \mathcal{E})$ represents the joint distribution of a random vector $(X_1, \ldots, X_p)$: $\mathcal{V} = \{1, \ldots, p\}$ and $\mathcal{E} \subset \binom{\mathcal{V}}{2}$
- Node $i \in \mathcal{V}$ corresponds to random variable/process X_i.
- Edge set $\mathcal{E}$ corresponds to conditional independencies.

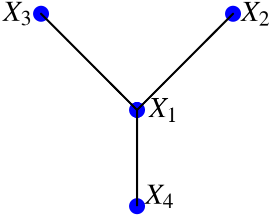


$X_i \perp\!\!\!\perp \mathbf{X}_{V\backslash\{\mathrm{nbd}(i)\cup i\}} | \mathbf{X}_{\mathrm{nbd}(i)}$

$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B | \mathbf{X}_S$

- Kalman filter, hidden Markov models, Bayesian networks...

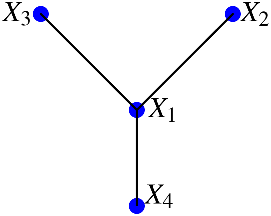
Tree-Structured Graphical Models



$$P(\mathbf{x}) = \prod_{i \in \mathcal{V}} P_i(x_i) \prod_{(i,j) \in \mathcal{E}} \frac{P_{i,j}(x_i, x_j)}{P_i(x_i) P_j(x_j)}$$

$$= P_1(x_1) \frac{P_{1,2}(x_1, x_2)}{P_1(x_1)} \frac{P_{1,3}(x_1, x_3)}{P_1(x_1)} \frac{P_{1,4}(x_1, x_4)}{P_1(x_1)}$$

Tree-Structured Graphical Models

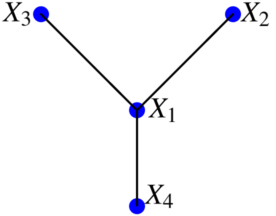


$$P(\mathbf{x}) = \prod_{i \in \mathcal{V}} P_i(x_i) \prod_{(i,j) \in \mathcal{E}} \frac{P_{i,j}(x_i, x_j)}{P_i(x_i) P_j(x_j)}$$

$$= P_1(x_1) \frac{P_{1,2}(x_1, x_2)}{P_1(x_1)} \frac{P_{1,3}(x_1, x_3)}{P_1(x_1)} \frac{P_{1,4}(x_1, x_4)}{P_1(x_1)}$$

For trees, directed model is equal to the undirected model

Tree-Structured Graphical Models



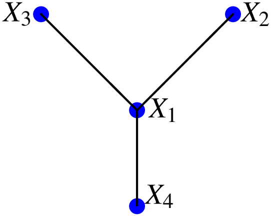
$$P(\mathbf{x}) = \prod_{i \in \mathcal{V}} P_i(x_i) \prod_{(i,j) \in \mathcal{E}} \frac{P_{i,j}(x_i, x_j)}{P_i(x_i) P_j(x_j)}$$

$$= P_1(x_1) \frac{P_{1,2}(x_1, x_2)}{P_1(x_1)} \frac{P_{1,3}(x_1, x_3)}{P_1(x_1)} \frac{P_{1,4}(x_1, x_4)}{P_1(x_1)}$$

For trees, directed model is equal to the undirected model

Tree-structured Graphical Models: Tractable Learning and Inference

Tree-Structured Graphical Models



$$P(\mathbf{x}) = \prod_{i \in \mathcal{V}} P_i(x_i) \prod_{(i,j) \in \mathcal{E}} \frac{P_{i,j}(x_i, x_j)}{P_i(x_i) P_j(x_j)}$$

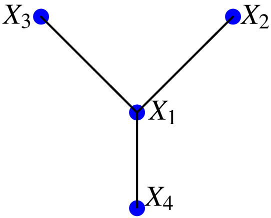
$$= P_1(x_1) \frac{P_{1,2}(x_1, x_2)}{P_1(x_1)} \frac{P_{1,3}(x_1, x_3)}{P_1(x_1)} \frac{P_{1,4}(x_1, x_4)}{P_1(x_1)}$$

For trees, directed model is equal to the undirected model

Tree-structured Graphical Models: Tractable Learning and Inference

- Maximum-Likelihood learning of tree structure is tractable
  - Chow-Liu Algorithm (1968)

Tree-Structured Graphical Models



$$P(\mathbf{x}) = \prod_{i \in \mathcal{V}} P_i(x_i) \prod_{(i,j) \in \mathcal{E}} \frac{P_{i,j}(x_i, x_j)}{P_i(x_i) P_j(x_j)}$$

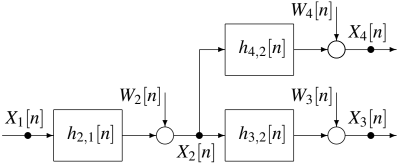
$$= P_1(x_1) \frac{P_{1,2}(x_1, x_2)}{P_1(x_1)} \frac{P_{1,3}(x_1, x_3)}{P_1(x_1)} \frac{P_{1,4}(x_1, x_4)}{P_1(x_1)}$$

For trees, directed model is equal to the undirected model

Tree-structured Graphical Models: Tractable Learning and Inference

- Maximum-Likelihood learning of tree structure is tractable
  - Chow-Liu Algorithm (1968)
- Inference on Trees is tractable
  - Sum-Product Algorithm (Pearl 1988)

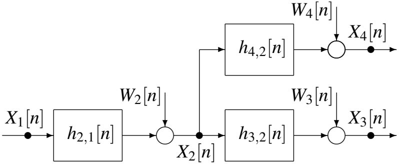
Main Result



Let us observe the p Gaussian WSS processes for N time steps

$$\{X_1[n]\}_{n=0}^{N-1}, \{X_2[n]\}_{n=0}^{N-1}, \ldots, \{X_p[n]\}_{n=0}^{N-1}$$

Main Result



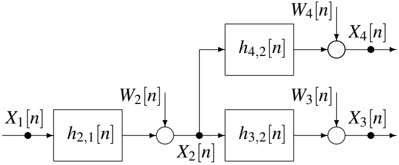
Let us observe the p Gaussian WSS processes for N time steps

$$\{X_1[n]\}_{n=0}^{N-1}, \{X_2[n]\}_{n=0}^{N-1}, \ldots, \{X_p[n]\}_{n=0}^{N-1}$$

Theorem

Let $\varepsilon > 0$. If $\mathcal{T}$ is a tree and mutual information rates on the edges are uniformly bounded away from zero, and if

Main Result



Let us observe the p Gaussian WSS processes for N time steps

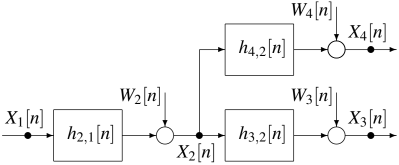
$$\{X_1[n]\}_{n=0}^{N-1}, \{X_2[n]\}_{n=0}^{N-1}, \ldots, \{X_p[n]\}_{n=0}^{N-1}$$

Theorem

Let $\varepsilon > 0$. If $\mathcal{T}$ is a tree and mutual information rates on the edges are uniformly bounded away from zero, and if

$$N = O\left(\log^{1+\varepsilon}\left(\frac{p}{\delta^{1/3}}\right)\right),$$

Main Result



Let us observe the p Gaussian WSS processes for N time steps

$$\{X_1[n]\}_{n=0}^{N-1}, \{X_2[n]\}_{n=0}^{N-1}, \ldots, \{X_p[n]\}_{n=0}^{N-1}$$

Theorem

Let $\varepsilon > 0$. If $\mathcal{T}$ is a tree and mutual information rates on the edges are uniformly bounded away from zero, and if

$$N = O\left(\log^{1+\varepsilon}\left(\frac{p}{\delta^{1/3}}\right)\right), \qquad \textit{then} \qquad \mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T}) < \delta$$

Learning Procedure I

- Given observations $\{X_1[n], \ldots, X_p[n]\}_{n=0}^{N-1}$, learn the tree $\mathcal{T}$

Learning Procedure I

- Given observations $\{X_1[n], \ldots, X_p[n]\}_{n=0}^{N-1}$, learn the tree $\mathcal{T}$
- $\mathcal{M}_{\text{Tree}}$: Set of probability measures associated to tree processes

Learning Procedure I

- Given observations $\{X_1[n], \ldots, X_p[n]\}_{n=0}^{N-1}$, learn the tree $\mathcal{T}$
- $\mathcal{M}_{\text{Tree}}$: Set of probability measures associated to tree processes
- Consider the problem

$$\min_{\nu \in \mathcal{M}_{\text{Tree}}} D(\mu \,||\, \nu), \qquad D(\mu \,||\, \nu) := \int \log \frac{d\mu}{d\mu} \, d\nu$$

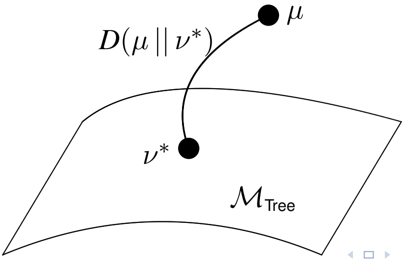
where μ is an estimate of the probability measure of $(X_1, \ldots, X_p)$

Learning Procedure I

- Given observations $\{X_1[n], \ldots, X_p[n]\}_{n=0}^{N-1}$, learn the tree $\mathcal{T}$
- $\mathcal{M}_{\text{Tree}}$: Set of probability measures associated to tree processes
- Consider the problem

$$\min_{\nu \in \mathcal{M}_{\text{Tree}}} D(\mu \,||\, \nu), \qquad D(\mu \,||\, \nu) := \int \log \frac{d\mu}{d\mu} \, d\nu$$

where μ is an estimate of the probability measure of $(X_1, \ldots, X_p)$



Learning Procedure II

$$\min_{\nu \in \mathcal{M}_{\text{Tree}}} D(\mu \,||\, \nu)$$

Learning Procedure II

$$\min_{\nu \in \mathcal{M}_{\text{Tree}}} D(\mu \,||\, \nu)$$

Lemma (Chow-Liu for processes)

The measure that achieves the minimum is Markov on $\hat{\mathcal{T}}_N$ given by

$$\hat{\mathcal{T}}_N = \operatorname*{arg\,max}_{\mathcal{T} \in \textit{Tree}} \sum_{(i,j) \in \mathcal{T}} I_\mu(X_i; X_j),$$

Learning Procedure II

$$\min_{\nu \in \mathcal{M}_{\text{Tree}}} D(\mu \,||\, \nu)$$

Lemma (Chow-Liu for processes)

The measure that achieves the minimum is Markov on $\hat{\mathcal{T}}_N$ given by

$$\hat{\mathcal{T}}_N = \operatorname*{arg\,max}_{\mathcal{T} \in \textit{Tree}} \sum_{(i,j) \in \mathcal{T}} I_\mu(X_i; X_j),$$

- Max-weight spanning tree problem: Kruskal's algorithm

Learning Procedure II

$$\min_{\nu \in \mathcal{M}_{\text{Tree}}} D(\mu \,||\, \nu)$$

Lemma (Chow-Liu for processes)

The measure that achieves the minimum is Markov on $\hat{\mathcal{T}}_N$ given by

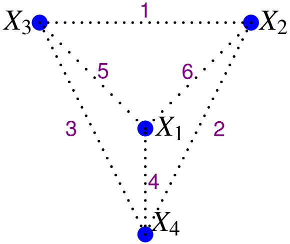
$$\hat{\mathcal{T}}_N = \arg\max_{\mathcal{T} \in \textit{Tree}} \sum_{(i,j) \in \mathcal{T}} I_\mu(X_i; X_j),$$

- Max-weight spanning tree problem: Kruskal's algorithm
- The mutual information rate for Gaussian processes is

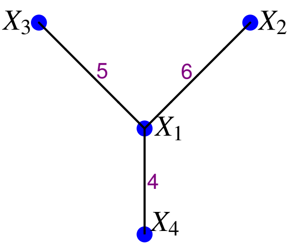
$$I(X_i; X_j) := -\frac{1}{4\pi} \int_0^{2\pi} \log\left(1 - |\gamma_{i,j}(\omega)|^2\right) \, d\omega$$

- $\gamma_{i,j}(\omega)$ is the coherence function between X_i and X_j

Learning Procedure III

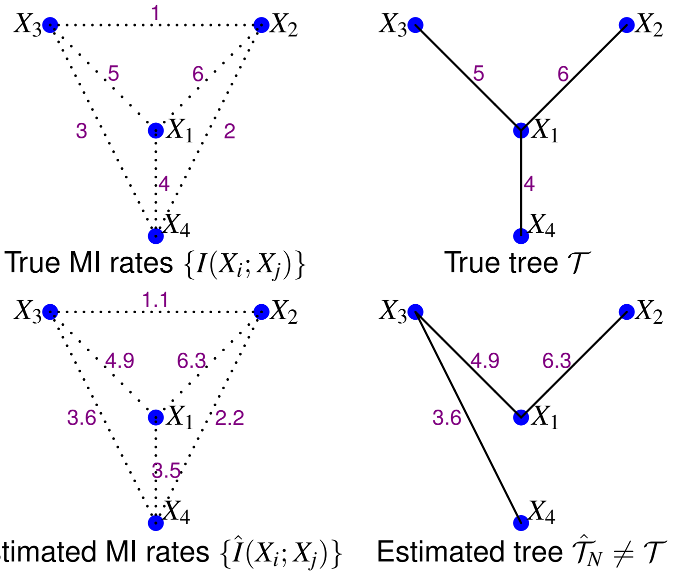


True MI rates $\{I(X_i; X_j)\}$



True tree $\mathcal{T}$

Learning Procedure III

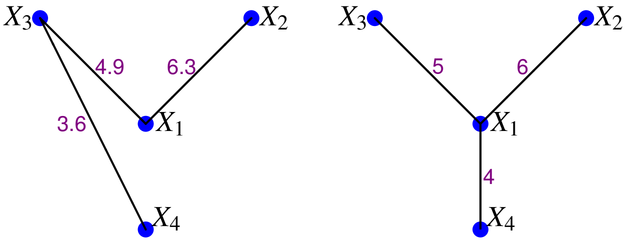


Problem Statement

- Error event is $\left\{\hat{\mathcal{T}}_N \neq \mathcal{T}\right\}$

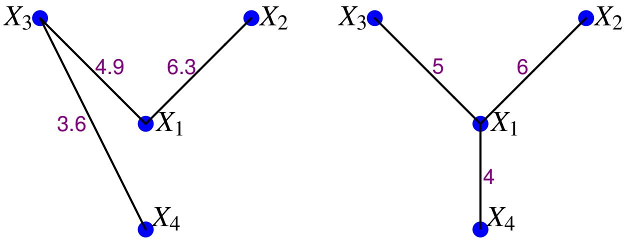
Problem Statement

- Error event is $\left\{\hat{\mathcal{T}}_N \neq \mathcal{T}\right\}$



Problem Statement

- Error event is $\left\{ \hat{\mathcal{T}}_N \neq \mathcal{T} \right\}$

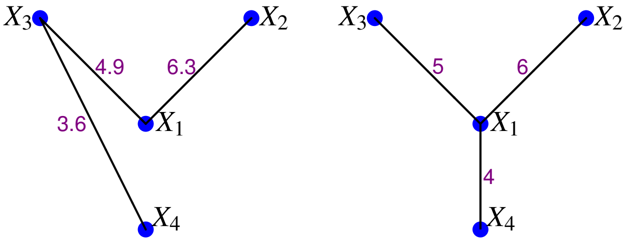


- Bound the error probability

$$\mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T})$$

Problem Statement

- Error event is $\left\{\hat{\mathcal{T}}_N \neq \mathcal{T}\right\}$



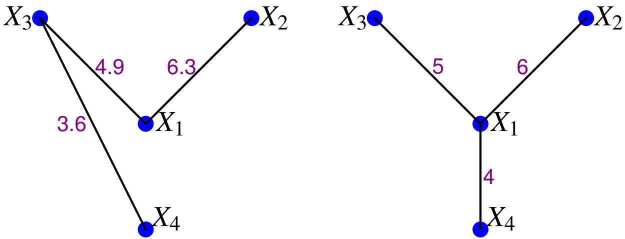
- Bound the error probability

$$\mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T})$$

- How do errors occur?

Problem Statement

- Error event is $\left\{ \hat{\mathcal{T}}_N \neq \mathcal{T} \right\}$



- Bound the error probability

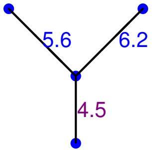
$$\mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T})$$

- How do errors occur?
- The order of the estimated MI rates relative to the true ones is important

Order of MI Rates

Correct Structure

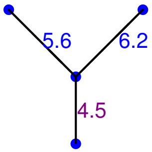
$I(X_i; X_j)$	6	5	4	3	2	1
$\hat{I}(X_i; X_j)$	6.2	5.6	4.5	2.8	2.2	1.1



Order of MI Rates

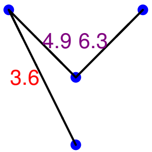
Correct Structure

$I(X_i;X_j)$	6	5	4	3	2	1
$\hat{I}(X_i;X_j)$	6.2	5.6	4.5	2.8	2.2	1.1



Incorrect Structure!

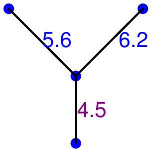
$I(X_i;X_j)$	6	5	4	3	2	1
$\hat{I}(X_i;X_j)$	6.3	4.9	3.5	3.6	2.2	1.1



Order of MI Rates

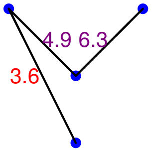
Correct Structure

$I(X_i;X_j)$	6	5	4	3	2	1
$\hat{I}(X_i;X_j)$	6.2	5.6	4.5	2.8	2.2	1.1



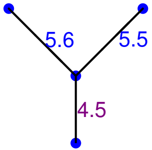
Incorrect Structure!

$I(X_i;X_j)$	6	5	4	3	2	1
$\hat{I}(X_i;X_j)$	6.3	4.9	3.5	3.6	2.2	1.1



Structure Unaffected

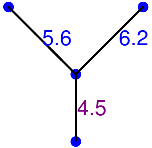
$I(X_i;X_j)$	6	5	4	3	2	1
$\hat{I}(X_i;X_j)$	5.5	5.6	4.5	3.0	2.2	1.1



Order of MI Rates

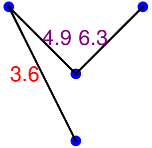
Correct Structure

$I(X_i;X_j)$	6	5	4	3	2	1
$\hat{I}(X_i;X_j)$	6.2	5.6	4.5	2.8	2.2	1.1



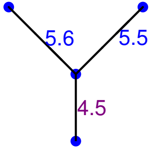
Incorrect Structure!

$I(X_i;X_j)$	6	5	4	3	2	1
$\hat{I}(X_i;X_j)$	6.3	4.9	3.5	3.6	2.2	1.1



Structure Unaffected

$I(X_i;X_j)$	6	5	4	3	2	1
$\hat{I}(X_i;X_j)$	5.5	5.6	4.5	3.0	2.2	1.1



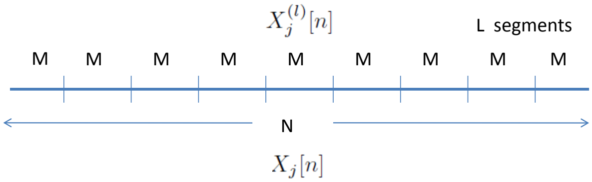
Estimate MI rates and ensure $\hat{I}(X_i;X_j) \approx I(X_i;X_j)$

Estimating the Mutual Information Rates I

- Given $\{X_1[n]\}_{n=0}^{N-1}, \ldots, \{X_p[n]\}_{n=0}^{N-1}$, estimate $\{I(X_i; X_j)\}_{(i,j) \in \mathcal{V}}$ using Bartlett's procedure

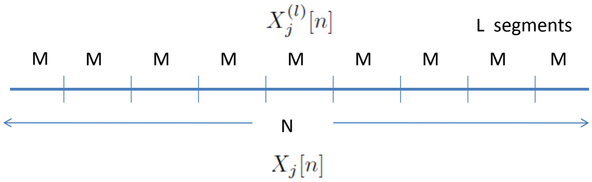
Estimating the Mutual Information Rates I

- Given $\{X_1[n]\}_{n=0}^{N-1}, \ldots, \{X_p[n]\}_{n=0}^{N-1}$, estimate $\{I(X_i;X_j)\}_{(i,j)\in\mathcal{V}}$ using Bartlett's procedure
- Divide each length-N realization $\{X_j[n]\}_{n=0}^{N-1}$ into L non-overlapping segments of length M such that $LM \approx N$



Estimating the Mutual Information Rates I

- Given $\{X_1[n]\}_{n=0}^{N-1}, \ldots, \{X_p[n]\}_{n=0}^{N-1}$, estimate $\{I(X_i; X_j)\}_{(i,j)\in\mathcal{V}}$ using Bartlett's procedure
- Divide each length-N realization $\{X_j[n]\}_{n=0}^{N-1}$ into L non-overlapping segments of length M such that $LM \approx N$



- Compute the length-M discrete Fourier transform for each signal segment, i.e., $\tilde{X}_j^{(l)}[k] = \mathcal{F}(X_j^{(l)}[n])$

Estimating the Mutual Information Rates II

- Estimate the time-averaged periodograms using Bartlett's averaging procedure on the L signal segments, i.e.,

$$\hat{\Phi}_{X_i}[k] := \frac{1}{L}\sum_{l=0}^{L-1}\left|\tilde{X}_i^{(l)}[k]\right|^2, \qquad \hat{\Phi}_{X_i,X_j}[k] := \frac{1}{L}\sum_{l=0}^{L-1}\left(\tilde{X}_i^{(l)}[k]\right)^* \tilde{X}_j^{(l)}[k]$$

Estimating the Mutual Information Rates II

- Estimate the time-averaged periodograms using Bartlett's averaging procedure on the L signal segments, i.e.,

$$\hat{\Phi}_{X_i}[k] := \frac{1}{L}\sum_{l=0}^{L-1}\left|\tilde{X}_i^{(l)}[k]\right|^2, \qquad \hat{\Phi}_{X_i,X_j}[k] := \frac{1}{L}\sum_{l=0}^{L-1}\left(\tilde{X}_i^{(l)}[k]\right)^* \tilde{X}_j^{(l)}[k]$$

- Estimate the magnitude-squared coherences:

$$|\hat{\gamma}_{i,j}[k]|^2 := \frac{|\hat{\Phi}_{X_i,X_j}[k]|^2}{\hat{\Phi}_{X_i}[k]\hat{\Phi}_{X_j}[k]}.$$

Estimating the Mutual Information Rates II

- Estimate the time-averaged periodograms using Bartlett's averaging procedure on the L signal segments, i.e.,

$$\hat{\Phi}_{X_i}[k] := \frac{1}{L}\sum_{l=0}^{L-1}\left|\tilde{X}_i^{(l)}[k]\right|^2, \qquad \hat{\Phi}_{X_i,X_j}[k] := \frac{1}{L}\sum_{l=0}^{L-1}\left(\tilde{X}_i^{(l)}[k]\right)^* \tilde{X}_j^{(l)}[k]$$

- Estimate the magnitude-squared coherences:

$$|\hat{\gamma}_{i,j}[k]|^2 := \frac{|\hat{\Phi}_{X_i,X_j}[k]|^2}{\hat{\Phi}_{X_i}[k]\hat{\Phi}_{X_j}[k]}.$$

- Estimate the MI rates by using the Riemann sum:

$$\hat{I}(X_i;X_j) := -\frac{1}{2M}\sum_{k=0}^{M-1}\log\left(1-|\hat{\gamma}_{i,j}[k]|^2\right).$$

Estimating the Tree

- With the estimated $\{\hat{I}(X_i; X_j)\}_{(i,j)\in\mathcal{V}}$, we run a max-weight spanning tree procedure to learn $\hat{\mathcal{T}}$

Estimating the Tree

- With the estimated $\{\hat{I}(X_i;X_j)\}_{(i,j)\in\mathcal{V}}$, we run a max-weight spanning tree procedure to learn $\hat{\mathcal{T}}$
- This procedure is efficient

$$O(p^2(N\log M + \log p)).$$

Estimating the Tree

- With the estimated $\{\hat{I}(X_i; X_j)\}_{(i,j)\in\mathcal{V}}$, we run a max-weight spanning tree procedure to learn $\hat{\mathcal{T}}$
- This procedure is efficient

$$O(p^2(N\log M + \log p)).$$

- But not clear how to tradeoff between number of signal segments L and length of each segment M

Estimating the Tree

- With the estimated $\{\hat{I}(X_i;X_j)\}_{(i,j)\in\mathcal{V}}$, we run a max-weight spanning tree procedure to learn $\hat{\mathcal{T}}$
- This procedure is efficient

$$O(p^2(N\log M + \log p)).$$

- But not clear how to tradeoff between number of signal segments L and length of each segment M
- For convenience, we let $M = M_L$ be a function of L

Concentration of MI Rate

Lemma (Concentration of MI Rate)

If the number of DFT points M_L satisfies

$$\lim_{L\to\infty} M_L = \infty, \qquad \lim_{L\to\infty} L^{-1}\log M_L = 0,$$

Concentration of MI Rate

Lemma (Concentration of MI Rate)

If the number of DFT points M_L satisfies

$$\lim_{L\to\infty} M_L = \infty, \qquad \lim_{L\to\infty} L^{-1}\log M_L = 0,$$

then for any $\eta > 0$, we have

$$\mathbb{P}\left(|\hat{I} - I| > \eta\right) \leq \exp\left[-(L-1)\min_{\omega\in[0,2\pi)} \varphi(\gamma(\omega);\eta)\right]$$

Concentration of MI Rate

Lemma (Concentration of MI Rate)

If the number of DFT points M_L satisfies

$$\lim_{L\to\infty} M_L = \infty, \qquad \lim_{L\to\infty} L^{-1}\log M_L = 0,$$

then for any $\eta > 0$, we have

$$\mathbb{P}\left(|\hat{I} - I| > \eta\right) \leq \exp\left[-(L-1)\min_{\omega\in[0,2\pi)} \varphi(\gamma(\omega);\eta)\right]$$

We can choose $M = \lceil L^{\varepsilon} \rceil, \varepsilon > 0$ to satisfy above condition

Bounding the Overall Error Probability I

Bounding the Overall Error Probability I

- Recall that the optimization for the tree structure is

$$\hat{\mathcal{T}}_N = \operatorname*{arg\,max}_{\mathcal{T} \in \text{Tree}} \sum_{(i,j) \in \mathcal{T}} \hat{I}(X_i; X_j),$$

where $\hat{I}(X_i; X_j)$ are the estimated mutual information rates

Bounding the Overall Error Probability I

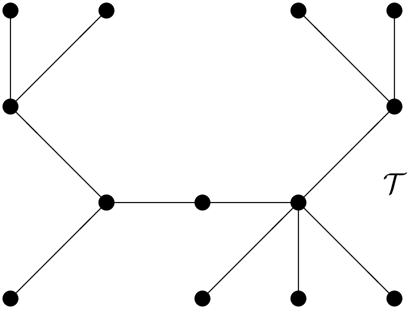
- Recall that the optimization for the tree structure is

$$\hat{\mathcal{T}}_N = \operatorname*{arg\,max}_{\mathcal{T} \in \text{Tree}} \sum_{(i,j) \in \mathcal{T}} \hat{I}(X_i; X_j),$$

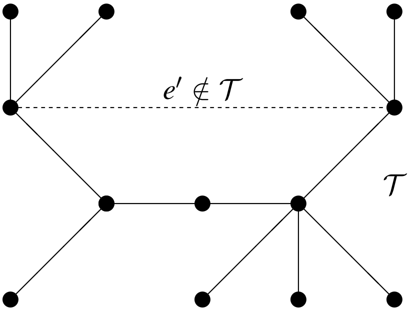
  where $\hat{I}(X_i; X_j)$ are the estimated mutual information rates

- How to get $\mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T})$ from the constituent error probabilities $\mathbb{P}\left(|\hat{I} - I| > \eta\right)$?

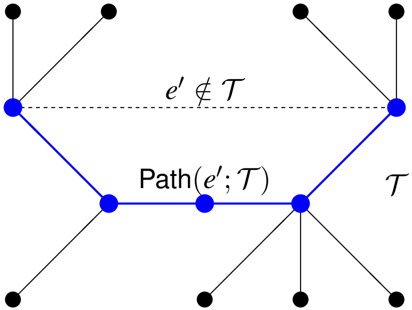
Bounding the Overall Error Probability II



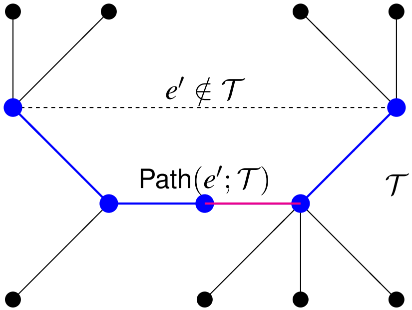
Bounding the Overall Error Probability II



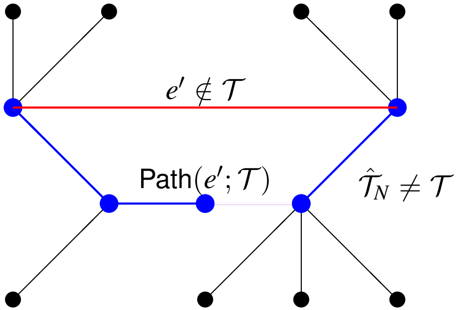
Bounding the Overall Error Probability II



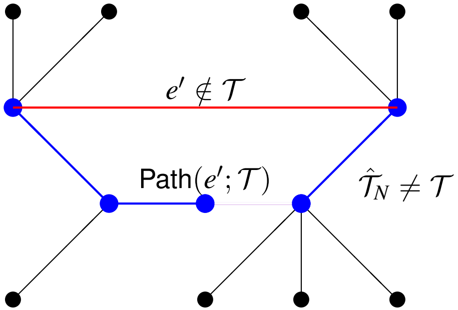
Bounding the Overall Error Probability II



Bounding the Overall Error Probability II



Bounding the Overall Error Probability II



Lemma

$$\left\{\hat{\mathcal{T}}_N \neq \mathcal{T}\right\} = \bigcup_{(k,l)\notin\mathcal{E}} \; \bigcup_{(i,j)\in \mathit{Path}(k,l)} \left\{\hat{I}(X_k; X_l) \geq \hat{I}(X_i; X_j)\right\}$$

Possible to bound $\mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T})$ by using the union bound

Main Result

Theorem

If $\mathcal{T}$ is a tree and mutual information rates on the edges are uniformly bounded away from zero, and if

Main Result

Theorem

If $\mathcal{T}$ is a tree and mutual information rates on the edges are uniformly bounded away from zero, and if

$$N = O\left(\log^{1+\varepsilon} p\right),$$

Main Result

Theorem

If $\mathcal{T}$ is a tree and mutual information rates on the edges are uniformly bounded away from zero, and if

$$N = O\left(\log^{1+\varepsilon} p\right), \qquad \textit{then} \qquad \mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T}) \to 0$$

Conclusion

- We have proposed a framework for learning tree-structured networks comprising dynamical systems and stochastic processes

Conclusion

- We have proposed a framework for learning tree-structured networks comprising dynamical systems and stochastic processes

- We showed that if observation time N is almost logarithmic in the number of nodes p,

$$\lim_{N\to\infty} \mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T}) = 0$$

Conclusion

- We have proposed a framework for learning tree-structured networks comprising dynamical systems and stochastic processes

- We showed that if observation time N is almost logarithmic in the number of nodes p,

$$\lim_{N\to\infty} \mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T}) = 0$$

- Future work I: Is $N = O(\log^{1+\varepsilon} p)$ optimal?

Conclusion

- We have proposed a framework for learning tree-structured networks comprising dynamical systems and stochastic processes
- We showed that if observation time N is almost logarithmic in the number of nodes p,

$$\lim_{N\to\infty} \mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T}) = 0$$

- Future work I: Is $N = O(\log^{1+\varepsilon} p)$ optimal?
- Future work II: Estimate directed tree

Conclusion

- We have proposed a framework for learning tree-structured networks comprising dynamical systems and stochastic processes
- We showed that if observation time N is almost logarithmic in the number of nodes p,

$$\lim_{N\to\infty} \mathbb{P}(\hat{\mathcal{T}}_N \neq \mathcal{T}) = 0$$

- Future work I: Is $N = O(\log^{1+\varepsilon} p)$ optimal?
- Future work II: Estimate directed tree