Likelihood-Free Overcomplete ICA and Applications In Causal Discovery

Chenwei Ding[1] Mingming Gong[2] Kun Zhang[3] Dacheng Tao[1]
[1]the University of Sydney
[2]The University of Melbourne
[3]Carnegie Mellon University

Previous Works on Causal Discovery

- Constraint-based methods, e.g. PC and FCI.
- Score-based methods, e.g. GES
- Functional Causal Models, e.g. LiNGAM (ICA based)

Independent Component Analysis (ICA)

- $\mathbf{x} = \mathbf{A}\mathbf{s}$, where mixtures $\mathbf{x} \in \mathbb{R}^p$, independent components $\mathbf{s} \in \mathbb{R}^d$, mixing matrix $\mathbf{A} \in \mathbb{R}^{p \times d}$.

Overcomplete ICA

- $p < d$
- Some causal discovery problems, e.g. causal discovery from measurement error and causal discovery from missing common causes, can be seen as extension of OICA.

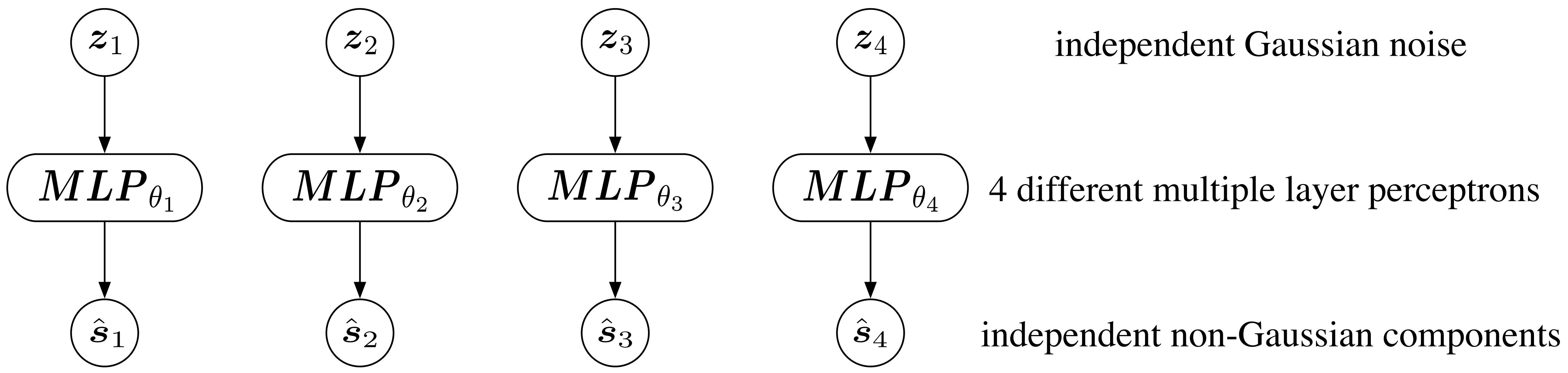
- **Maximum Likelihood Learning Based Solutions for OICA**
    - Assume parametric distribution for the ICs.
    - Significant computational challenges.
    - Restrictive for many real-world applications.
- **Likelihood Free Solution for OICA (Ours)**
    - No explicit assumptions on the density functions of the ICs.
    - Implicitly learn the distribution of ICs.
    - Computationally efficient.

LFOICA Framework

- Sample independently from some easy distribution, i.g. Gaussian.

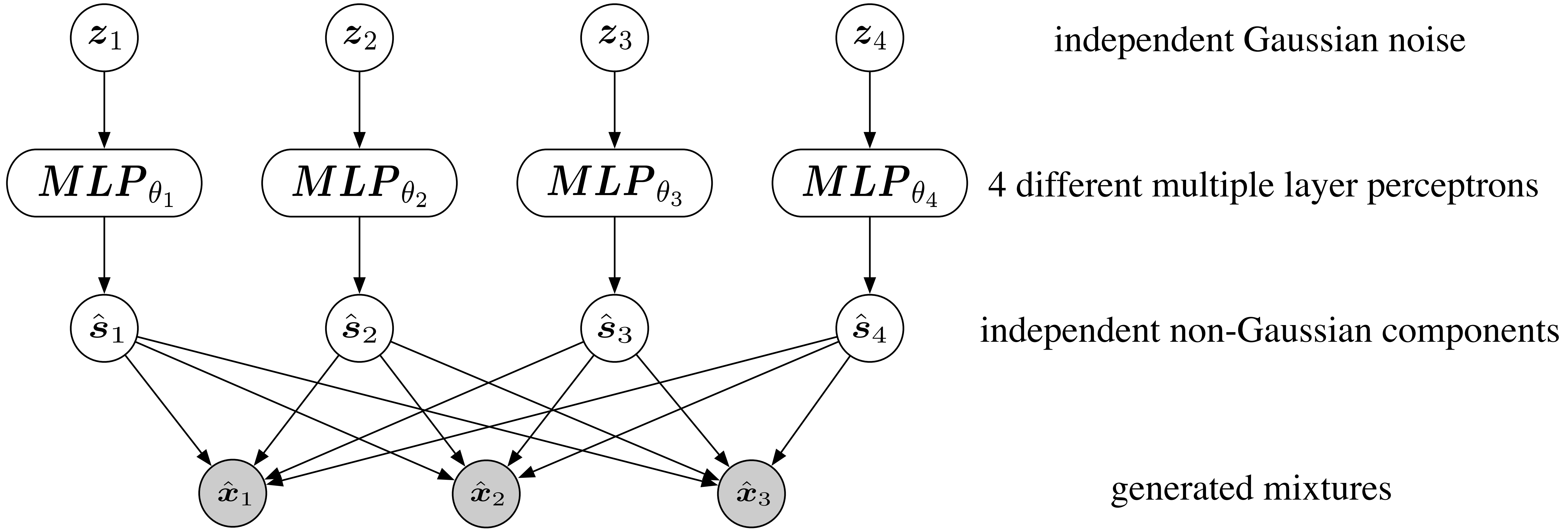
LFOICA Framework

- Sample independently from some easy distribution, i.g. Gaussian.
- For each IC, initialize a separate MLP. Generate corresponding IC.



LFOICA Framework

- Sample independently from some easy distribution, i.g. Gaussian.
- For each IC, initialize a separate MLP. Generate corresponding IC.
- Initialize a mixing matrix, mix the ICs and generate the mixtures.
- Calculate the MMD between the distribution of true mixtures and the distribution of generated mixtures.
- Minimize MMD by updating the mixing matrix and the parameters in MLPs.



Causal Discovery under Measurement Error

- Causal model without measurement error.

$$\tilde{\mathbf{X}} = \mathbf{B}\tilde{\mathbf{X}} + \tilde{\mathbf{E}}$$

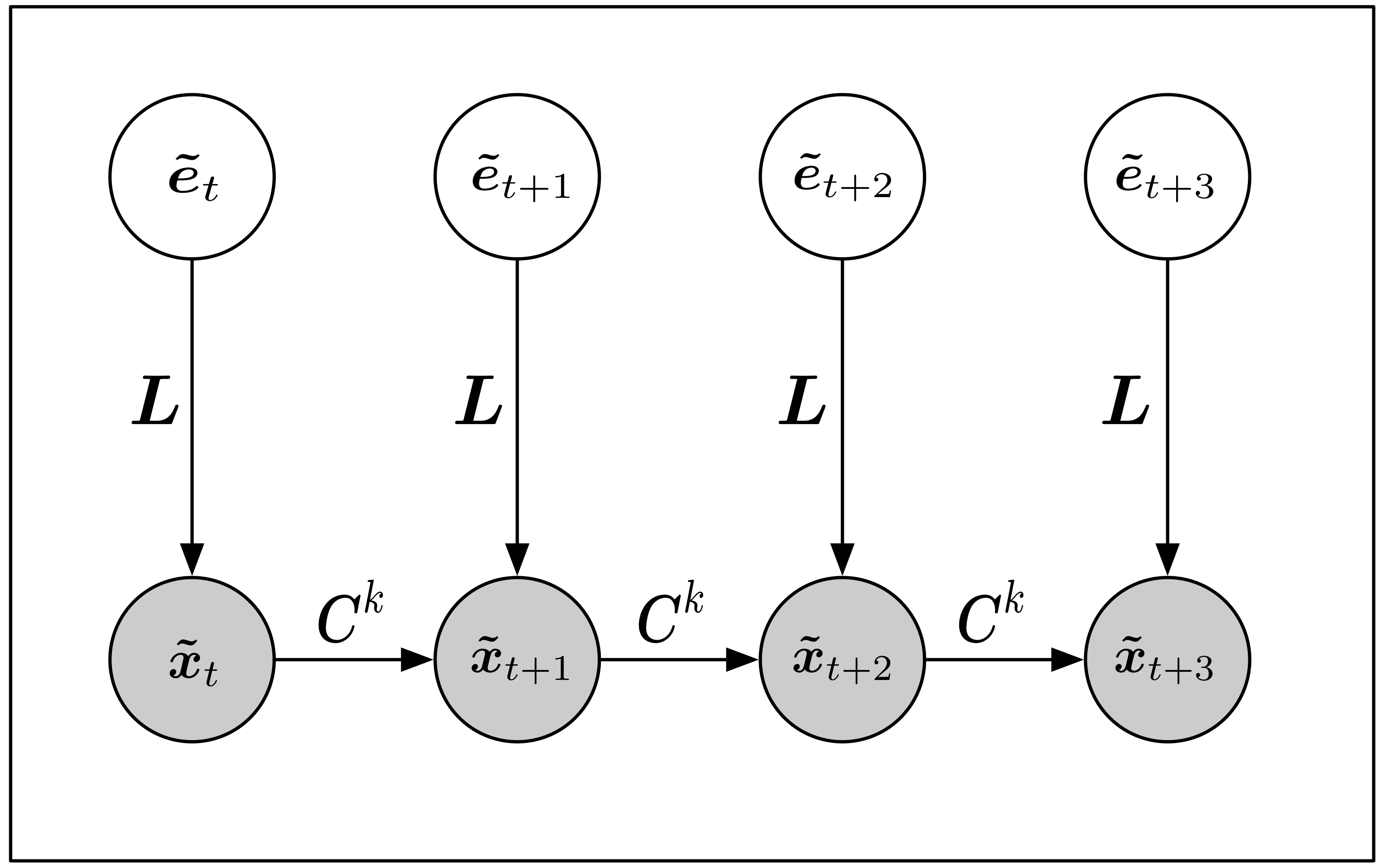
- Add measurement error to the causal model.

$$\mathbf{X} = \tilde{\mathbf{X}} + \mathbf{E} = (\mathbf{I} - \mathbf{B})^{-1}\tilde{\mathbf{E}} + \mathbf{E} = \left[(\mathbf{I} - \mathbf{B})^{-1}\mathbf{I}\right] \begin{bmatrix} \tilde{\mathbf{E}} \\ \mathbf{E} \end{bmatrix}$$

- The causal model with measurement error can be seen as an OICA model and LFOICA can be applied.

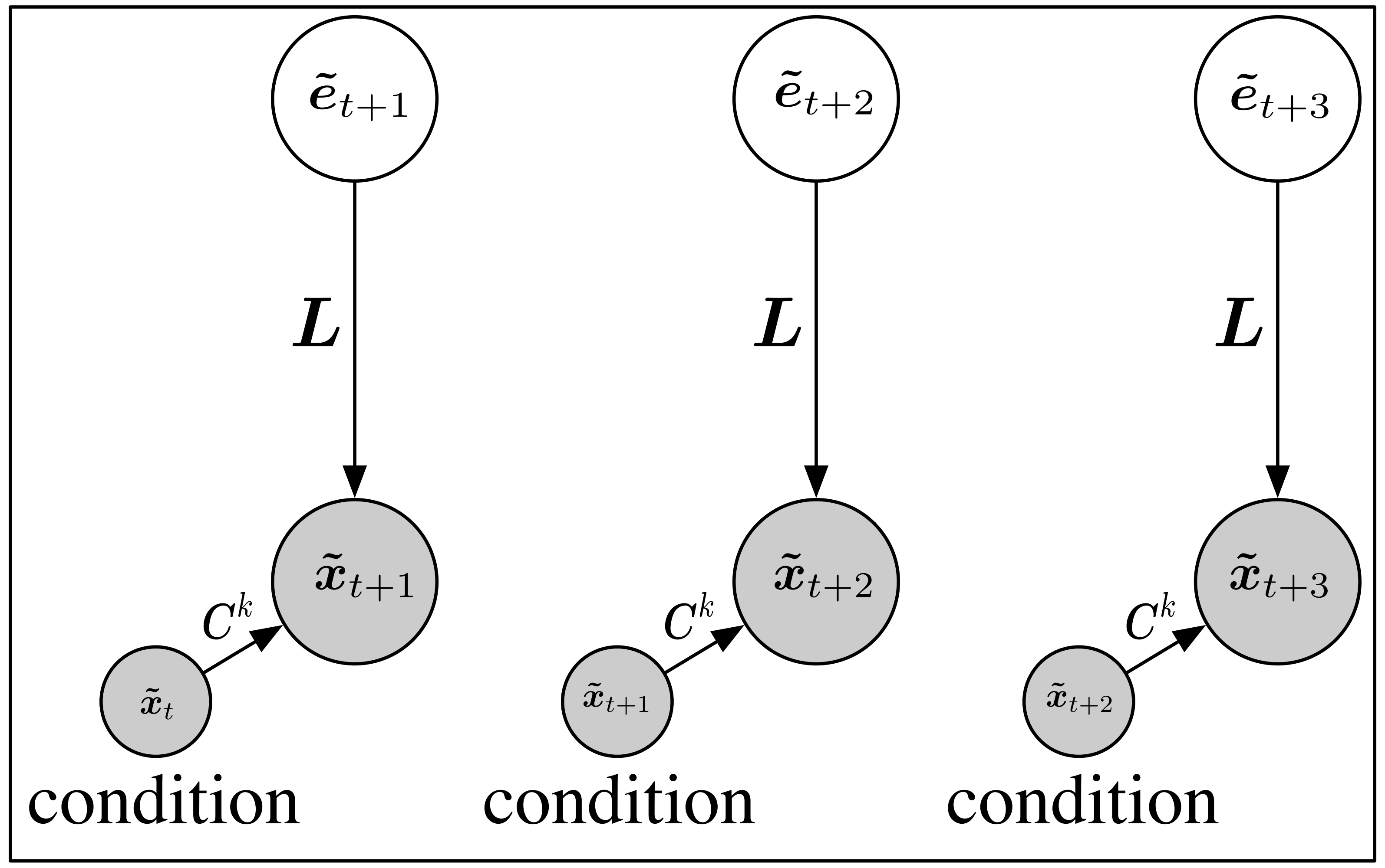
Causal Discovery from Subsampled Time Series

- First assume that data at the original causal frequency follows a VAR(1) process $\mathbf{x}_t = \mathbf{C}\mathbf{x}_{t-1} + \mathbf{e}_t$
- The observed subsampled data with subsampling factor k can be represented as $\tilde{\mathbf{x}}_{t+1} = \mathbf{C}^k \tilde{\mathbf{x}}_t + \mathbf{L}\tilde{\mathbf{e}}_{t+1}$, where $\mathbf{L} = \left[\mathbf{I}, \mathbf{C}, \mathbf{C}^2, \ldots, \mathbf{C}^{k-1}\right]$ and $\tilde{\mathbf{e}}_\mathbf{t} = \left(\mathbf{e}^\top_{1+tk-0}, \mathbf{e}^\top_{1+tk-1}, \ldots, \mathbf{e}^\top_{1+tk-(k-1)}\right)^\top$



Causal Discovery from Subsampled Time Series

- We propose to model the conditional probability $\mathbb{P}\left(\tilde{\mathbf{x}}_{t+1} \,|\, \tilde{\mathbf{x}}_t\right)$



- In this case, the model for subsampled data can be seen as an extension of OICA with $\mathbf{L}$ as mixing matrix and $\tilde{\mathbf{e}}_{t+1}$ as ICs. LFOICA can be applied.

Poster #45

Tue Dec 10, 2019

East Exhibition Hall B + C