Multivariate Causal Discovery with General Nonlinear Relationships






Patrik Reizinger*, Yash Sharma, Matthias Bethge, Bernhard Schölkopf, Ferenc Huszár, Wieland Brendel

1st workshop on Causal Representation Learning @ UAI2022

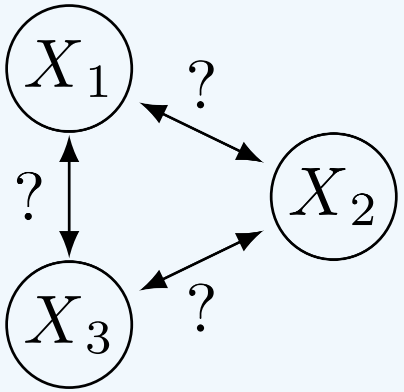
05.08.2022

Our goal is causal discovery for nonlinear models

Input: observational data

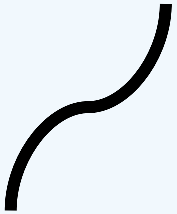
$$X_1 = f_1(N_1)$$
$$X_2 = f_2(X_1, N_2)$$
$$X_3 = f_3(X_1, X_2, N_3)$$

Goal: infer the edges



X_i - observed variables
N_i - exogenous (noise) variables

Desiderata for causal discovery



Nonlinear

$$\frac{\partial}{\partial X_i}$$

End-to-end differentiable

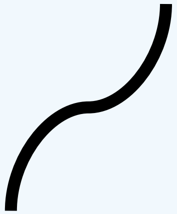
Scalable

Observational

No interventions

?

Desiderata for causal discovery



Nonlinear

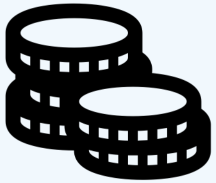
$$\frac{\partial}{\partial X_i}$$

End-to-end Differentiable

Scalable

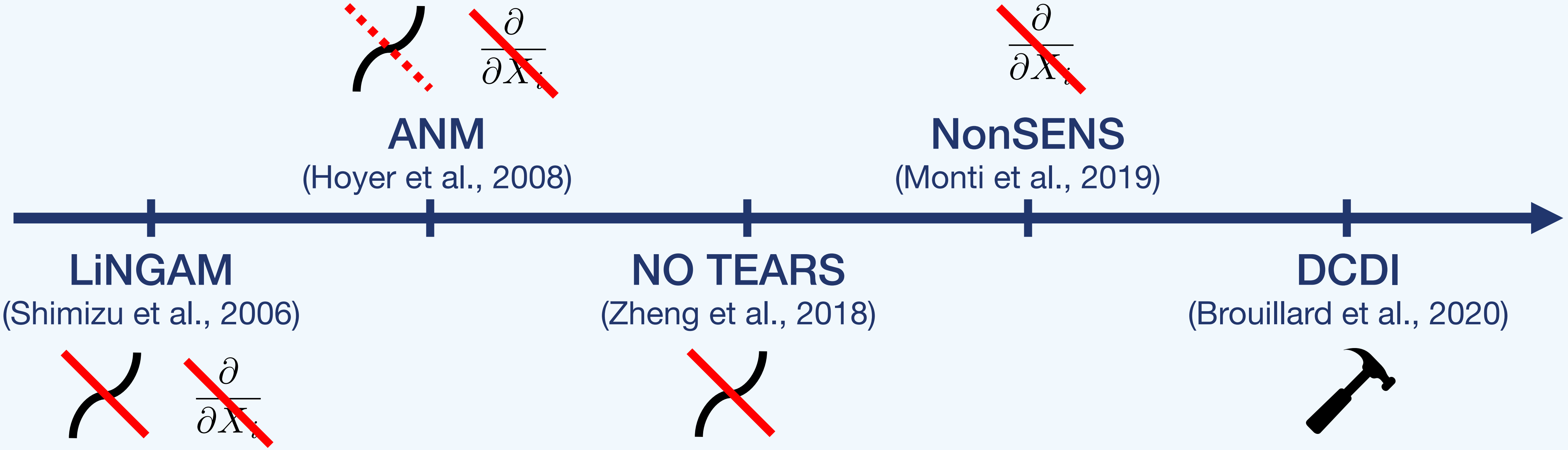
Observational

No interventions



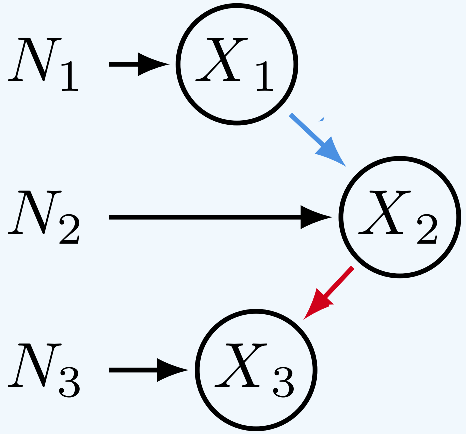
The price we pay
(assumptions)

Related work



Intuition: the inverse DGP captures the DAG in linear SEMs

Example



X_i - observed variables
N_i - exogenous (noise) variables

DGP

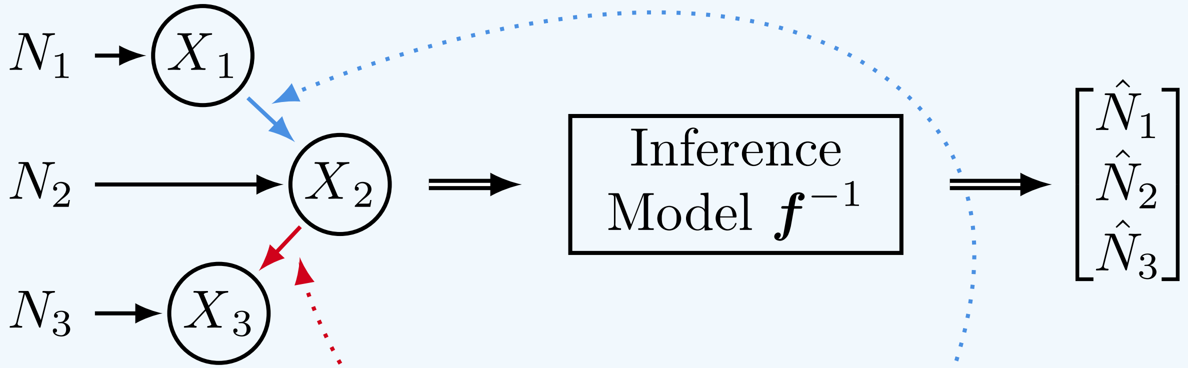
$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ {\color{red}a} & 1 & 0 \\ {\color{red}ab} & {\color{blue}b} & 1 \end{bmatrix} \begin{bmatrix} N_1 \\ N_2 \\ N_3 \end{bmatrix}$$

Inverse DGP

$$\begin{bmatrix} N_1 \\ N_2 \\ N_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ {\color{red}-a} & 1 & 0 \\ 0 & {\color{blue}-b} & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

DGP (Data Generation Process)

The inference Jacobian the DAG in a nonlinear SEM



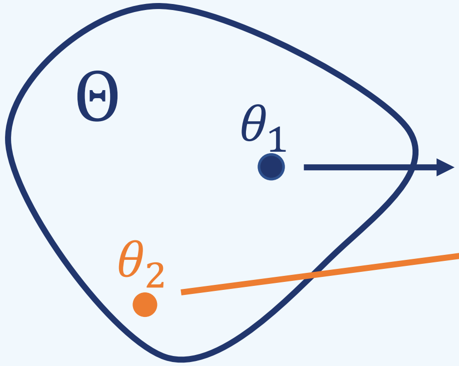
$$X_1 = f_1(N_1)$$
$$X_2 = f_2(N_1, N_2)$$
$$X_3 = f_3(N_1, N_2, N_3)$$

$$\mathbf{J}_{\boldsymbol{f}^{-1}} = \begin{bmatrix} * & 0 & 0 \\ \blacksquare & * & 0 \\ 0 & \blacktriangle & * \end{bmatrix}$$

X_i - observed variables
N_i - exogenous (noise) variables

Identifiable representation learning for causal discovery

Identifiability



$$f_{\theta_1}^{-1}(x) = A f_{\theta_2}^{-1}(x) + c$$

θ_i - parameter
f^{-1} - inference (unmixing) function

How to learn f^{-1} from data?

- Opportunity: recent nonlinear ICA results → indeterminacies
- SEM:
  - Jacobian of f^{-1} is lower-triangular
  - Causal ordering is unknown

Resolving the permutation indeterminacy of ICA

1. ICA indeterminacy + unknown order of $\mathrm{N_i}$

$$\begin{bmatrix} \blacksquare & 0 & * \\ * & 0 & 0 \\ 0 & * & \blacktriangle \end{bmatrix} \overset{?}{\Longleftrightarrow} \begin{bmatrix} * & 0 & 0 \\ \blacksquare & * & 0 \\ 0 & \blacktriangle & * \end{bmatrix}$$

Resolving the permutation indeterminacy of ICA

1. ICA indeterminacy + unknown order of $\mathrm{N_i}$

$$\begin{bmatrix} \blacksquare & 0 & * \\ * & 0 & 0 \\ 0 & * & \blacktriangle \end{bmatrix} \overset{?}{\Longleftrightarrow} \begin{bmatrix} * & 0 & 0 \\ \blacksquare & * & 0 \\ 0 & \blacktriangle & * \end{bmatrix}$$

2. Two trainable permutation networks

$$\begin{bmatrix} \mathbf{S}_{\mathrm{ICA}} \end{bmatrix} \begin{bmatrix} \blacksquare & 0 & * \\ * & 0 & 0 \\ 0 & * & \blacktriangle \end{bmatrix} \begin{bmatrix} \mathbf{S} \end{bmatrix}$$

Resolving the permutation indeterminacy of ICA

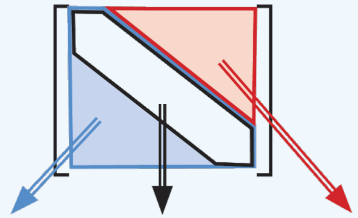
1. ICA indeterminacy + unknown order of $\mathrm{N_i}$

$$\begin{bmatrix} \blacksquare & 0 & * \\ * & 0 & 0 \\ 0 & * & \blacktriangle \end{bmatrix} \overset{?}{\Longleftrightarrow} \begin{bmatrix} * & 0 & 0 \\ \blacksquare & * & 0 \\ 0 & \blacktriangle & * \end{bmatrix}$$

2. Two trainable permutation networks

$$\begin{bmatrix} \mathbf{S}_{\mathrm{ICA}} \end{bmatrix} \begin{bmatrix} \blacksquare & 0 & * \\ * & 0 & 0 \\ 0 & * & \blacktriangle \end{bmatrix} \begin{bmatrix} \mathbf{S} \end{bmatrix}$$

3. Loss: the Jacobian should be lower triangular



$$\min[-\mathrm{tril}(|\bullet|) + \mathrm{diag}(|\bullet|)^{-1} + \mathrm{triu}(|\bullet|)]$$

Results

Validation experiments

DGP	d	MCC	Acc_π	Acc	SHD
LIN. SEM	3	1.	1.	1.	0.
	5	1.	0.966	1.	0.0013
	8	1.	1.	1.	0.
NL. SEM	3	1.	1.	1.	0.
	5	0.971 ± 0.07	0.828	0.974	0.0262
	8	0.987 ± 0.03	0.793	0.968	0.0318

Comparing to NonSENS

# LAYERS	MCC	Acc	SHD
1	1.	1.	0.
2	0.999	1.	0.0056
3	0.932 ± 0.09	0.9	0.1
4	0.833 ± 0.01	0.817	0.1833
5	0.848 ± 0.02	0.839	0.1611

MCC (Mean Correlation Coefficient): measures identifiability (higher is better; range: [0;1])
Acc(uracy): subscript refers to correctly resolving the permutation; π stands for permutations (higher is better; range: [0;1])
SHD (Structural Hamming Distance): ratio of incorrectly inferred edges (lower is better, range: [0;1])

Results

Validation experiments

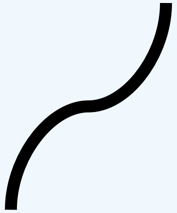
DGP	d	MCC	Acc_π	Acc	SHD
LIN. SEM	3	1.	1.	1.	0.
	5	1.	0.966	1.	0.0013
	8	1.	1.	1.	0.
NL. SEM	3	1.	1.	1.	0.
	5	0.971 ± 0.07	0.828	0.974	0.0262
	8	0.987 ± 0.03	0.793	0.968	0.0318

Comparing to NonSENS

# LAYERS	MCC	Acc	SHD
1	1.	1.	0.
2	0.999	1.	0.0056
3	0.932 ± 0.09	0.9	0.1
4	0.833 ± 0.01	0.817	0.1833
5	0.848 ± 0.02	0.839	0.1611

MCC (Mean Correlation Coefficient): measures identifiability (higher is better; range: [0;1])
Acc(uracy): subscript refers to correctly resolving the permutation; π stands for permutations (higher is better; range: [0;1])
SHD (Structural Hamming Distance): ratio of incorrectly inferred edges (lower is better, range: [0;1])

Conclusion



Nonlinear

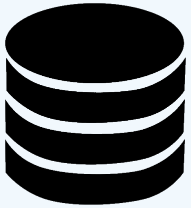
$$\frac{\partial}{\partial X_i}$$

End-to-end differentiable

Scalable

Observational

No interventions



Assumptions

(Zimmermann et al., 2021)

Multivariate Causal Discovery with General Nonlinear Relationships



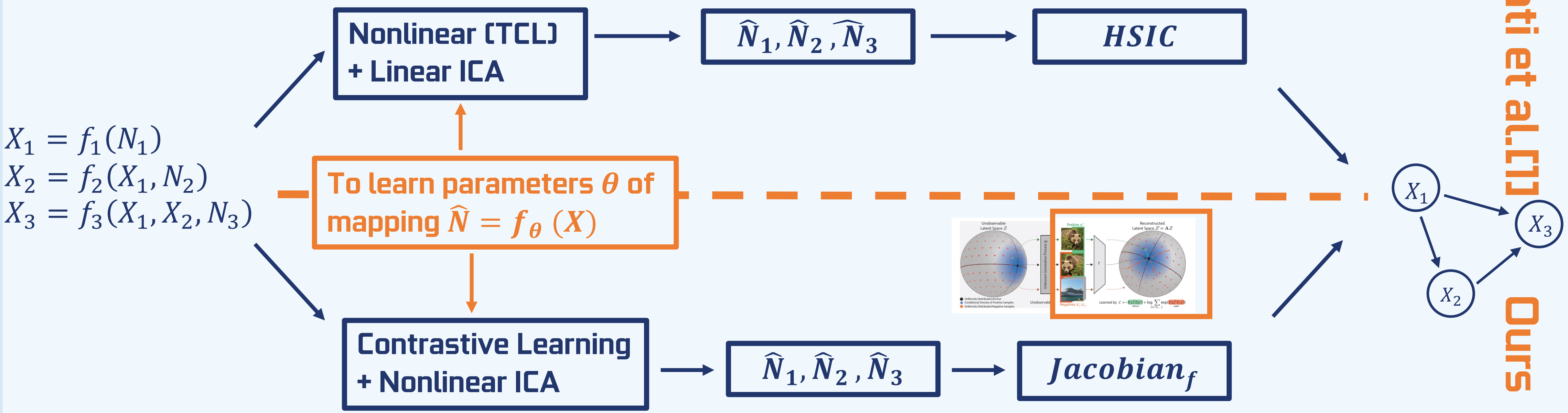



Patrik Reizinger*, Yash Sharma, Matthias Bethge, Bernhard Schölkopf, Ferenc Huszár, Wieland Brendel

1st workshop on Causal Representation Learning @ UAI2022

05.08.2022

We learn the edges of the DAG



[1] Monti et al. Causal Discovery with General Non-Linear Relationships Using Non-Linear ICA

N – exogenous, X – observed

Causal discovery with Nonlinear ICA (Monti et al.)

- NonSENS (Monti et al., [1]) is the most relevant prior work
- Nonlinear ICA + HSIC (independece test between latents and observations)
- Limitations:
  - Not scalable
  - Not differentiable: Cannot be used in a causal representation learning pipeline
  - Yields false positives for indirect causes
- Advantages:
  - Has a significance value

[1] Monti et al. Causal Discovery with General Non-Linear Relationships Using Non-Linear ICA

Multivariate Causal Discovery with General Nonlinear Relationships






Patrik Reizinger*, Yash Sharma, Matthias Bethge, Bernhard Schölkopf, Ferenc Huszár, Wieland Brendel

1st workshop on Causal Representation Learning @ UAI2022

05.08.2022