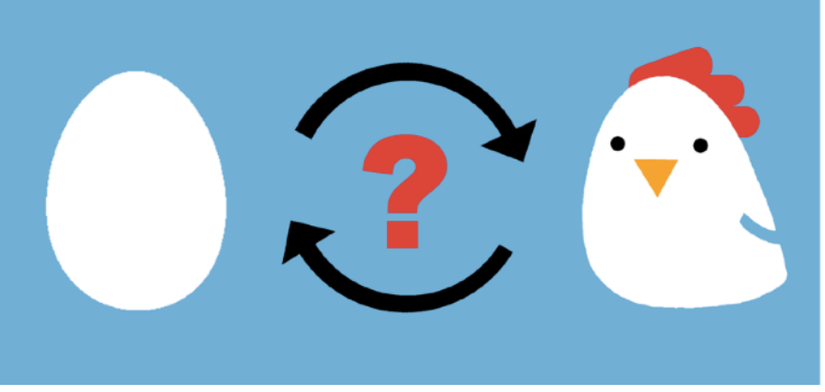


Causality and Machine Learning
(80-816/516)

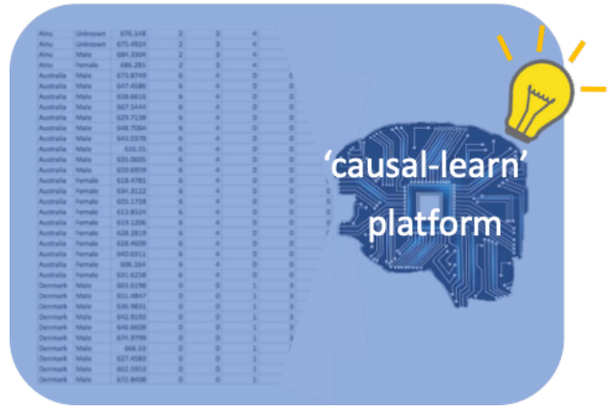
Classes 18 (March 20, 2025)

Causal Representation Learning 1: IID Case

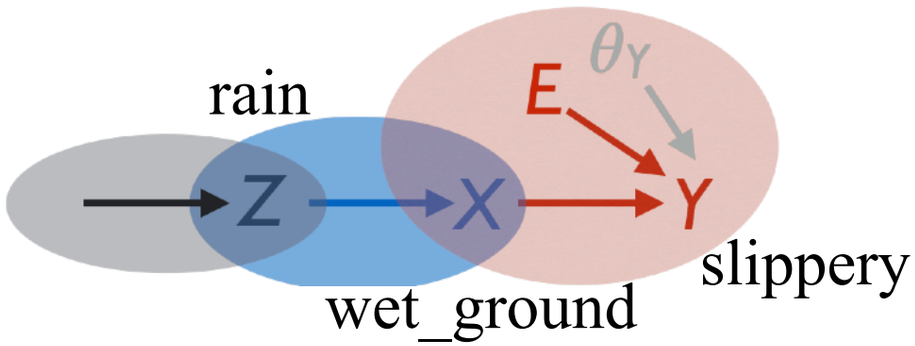
Instructor:

Kun Zhang (kunz1@cmu.edu)
Zoom link: https://cmu.zoom.us/j/8214572323)
Office Hours: W 3:00–4:00PM (on Zoom or in person); other times by appointment

Uncover Causality from Observational Data: How?



- Causal discovery (Spirtes et al., 1993)/ causal representation learning (Schölkopf et al., 2021): find such representations with identifiability guarantees
- Causal system has "irrelevant" modules (Spirtes et al., 1993; Pearl, 2000)



- conditional independence among variables;
- independent noise condition;
- minimal (and independent) changes...

Footprint of causality in data

- Three dimensions of the problem:

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

We Mainly Focused on the IID Case: Recent Advances in Causal Representation Learning

i.i.d. data?	Parametric constraints?	Latent confounders?	What can we get?
Yes	*No*	*No*	*(Different types of) equivalence class*
		Yes	
	Yes	**No**	**Unique identifiability (under structural conditions)**
		Yes	
Non-I, but I.D.	No/Yes	No	(Extended) regression
		Yes	Latent temporal causal processes identifiable!
I., but non-I.D.	No	No	More informative than MEC (CD-NOD)
	Yes		May have unique identifiability
	No	Yes	Changing subspace identifiable
	Yes		Variables in changing relations identifiable

- PC, FCI, etc.

- LiNGAM

- Rank-based, GIN…

CRL in IID Case: How to achieve it?

- Linear-Gaussian case
  - Tetrad conditions
  - Rank deficiency-based method
- Linear, non-Gaussian case
  - Theoretical results
  - GIN-based method
- Nonlinear case
  - Sparsity
- Summary: Why is it possible?

i.i.d. data?	Parametric constraints?	Latent confounders?	What can we get?
Yes	*No*	*No*	*(Different types of) equivalence class*
		Yes	
	Yes	**No**	**Unique identifiability (under structural conditions)**
		Yes	

A Problem in Psychology: Finding Underlying Mental Conditions?

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

- 50 questions for big 5 personality test

race	age	engnat	gender	hand	source	country	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	A1	A2	A3	A4	A5
3	53	1	1	1	1	US	4	2	5	2	5	1	4	3	5	1	1	5	2	5	1	1	1	1	1	1	1	5	1	5	2
13	46	1	2	1	1	US	2	2	3	3	3	3	1	5	1	5	2	3	4	2	3	4	3	2	2	4	1	3	3	4	4
1	14	2	2	1	1	PK	5	1	1	4	5	1	1	5	5	1	5	1	5	5	5	5	5	5	5	5	5	1	5	5	1
3	19	2	2	1	1	RO	2	5	2	4	3	4	3	4	4	5	5	4	4	2	4	5	5	5	4	5	2	5	4	4	3
11	25	2	2	1	2	US	3	1	3	3	3	1	3	1	3	5	3	3	3	4	3	3	3	3	3	4	5	5	3	5	1
13	31	1	2	1	2	US	1	5	2	4	1	3	2	4	1	5	1	5	4	5	1	4	4	1	5	2	2	2	3	4	3
5	20	1	2	1	5	US	5	1	5	1	5	1	5	4	4	1	2	4	2	4	2	2	3	2	2	2	5	5	1	5	1
4	23	2	1	1	2	IN	4	3	5	3	5	1	4	3	4	3	1	4	4	4	1	1	1	1	1	1	2	5	1	4	3
5	39	1	2	3	4	US	3	1	5	1	5	1	5	2	5	3	2	4	5	3	3	5	5	4	3	3	1	5	1	5	1
3	18	1	2	1	5	US	1	4	2	5	2	4	1	4	1	5	5	2	5	2	3	4	3	2	3	4	2	3	1	4	2
3	17	2	2	1	1	IT	1	5	2	5	1	4	1	4	1	5	5	3	5	3	2	5	3	3	4	3	2	4	2	4	1
13	15	2	1	1	1	IN	3	3	5	3	3	3	2	4	3	3	1	5	3	3	2	3	2	3	2	4	4	4	2	2	5
13	22	1	2	1	2	US	3	3	4	2	4	2	2	3	4	3	3	3	3	3	2	2	4	4	2	3	1	4	1	5	1
3	21	1	2	1	5	US	1	3	2	5	1	1	1	5	1	5	5	3	5	2	5	5	3	2	5	3	1	1	1	4	2
3	28	2	2	1	2	US	3	3	3	4	3	2	2	4	3	5	2	4	4	4	4	4	2	2	3	2	1	4	2	4	2
3	21	1	1	1	5	US	2	3	2	3	3	1	1	3	4	4	2	4	2	4	1	2	2	2	2	2	4	2	4	2	5
13	19	1	2	1	2	FR	1	3	2	4	2	4	1	4	3	4	4	2	3	2	1	3	1	2	2	3	4	2	3	1	4
3	21	1	2	1	5	US	4	1	5	2	5	1	5	3	5	1	5	2	5	2	3	3	3	3	4	2	1	5	2	5	2

Learning Hidden Variables & Their Relations

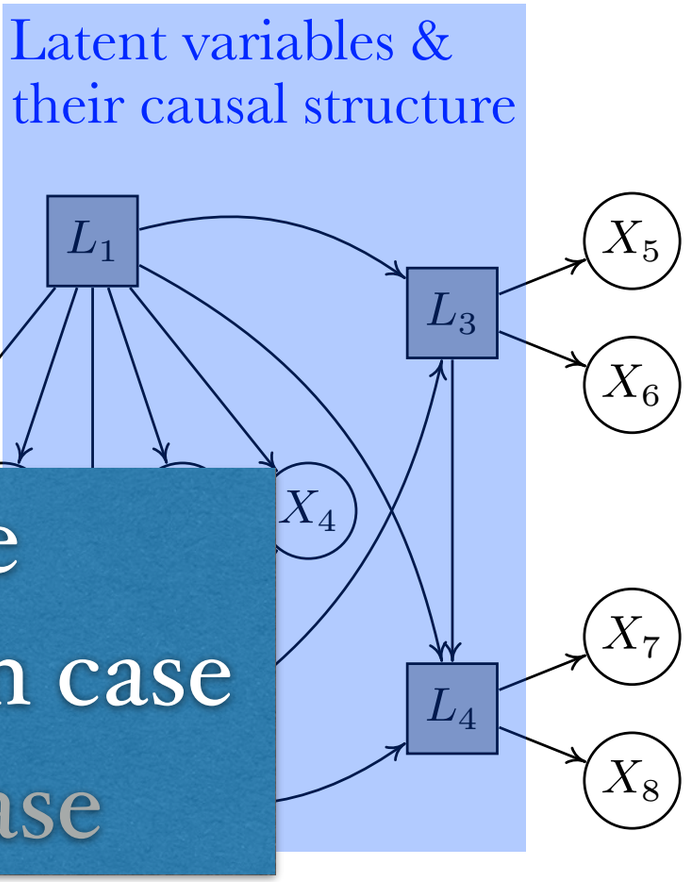
i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

- Measured variables (e.g., answer scores in psychometric questionnaires) were generated by causally related latent variables

X1	X2	X3	X4	X5	X6	X7	X8
4.2	3.6	6.5	6.8	9.6	7.6	2.7	4.8
3.8	1.9	6.5	7.3	8.9	6.9	1.1	4.6
4.2	3.4	6.5	6.9	9.5	7.4	2.5	4.6
4.2	2.2	6.2	6.9	9.6	7.2	1.9	4.8
3.9	1.9	6.5	6.8	9.0	6.8	1.7	4.4
4.0	2.0	6.4	7.2	9			
3.8	1.7	6.4	7.3	9			
4.1	2.8	6.5	6.9	9			
...	...	...	...				

Discovery: How?

Latent variables & their causal structure



- Linear-Gaussian case
- Linear, non-Gaussian case
- General nonlinear case

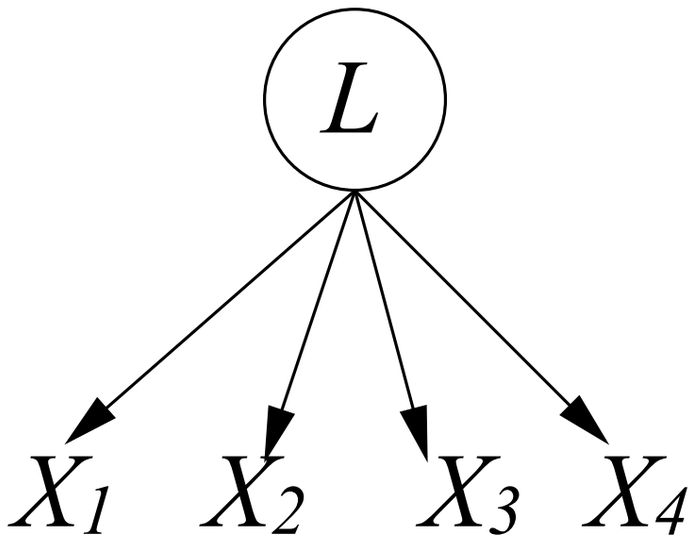
- Find latent variables L_i and their causal relations from measured variables X_i ?

CRL in IID Case: How to achieve it?

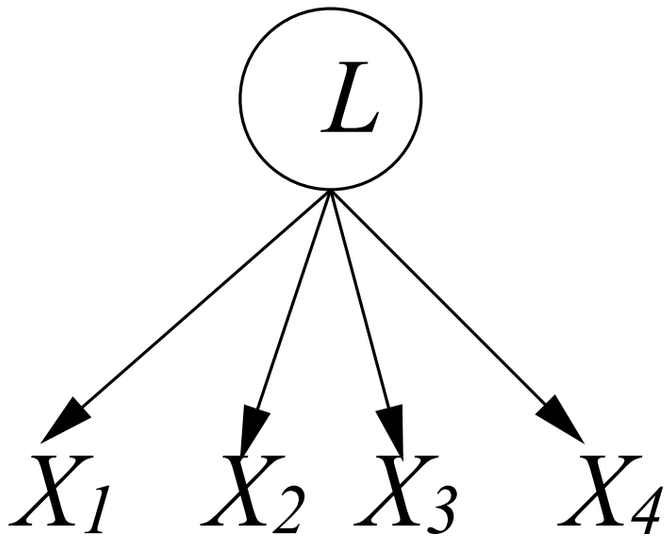
- Linear-Gaussian case
  - Tetrad conditions
  - Rank deficiency-based method
- Linear, non-Gaussian case
  - Theoretical results
  - GIN-based method
- Nonlinear case
  - Sparsity
- Summary: Why is it possible?

i.i.d. data?	Parametric constraints?	Latent confounders?	What can we get?
Yes	*No*	*No*	*(Different types of) equivalence class*
		Yes	
	Yes	**No**	**Unique identifiability (under structural conditions)**
		Yes	

Tetrad Constraints



Tetrad Constraints

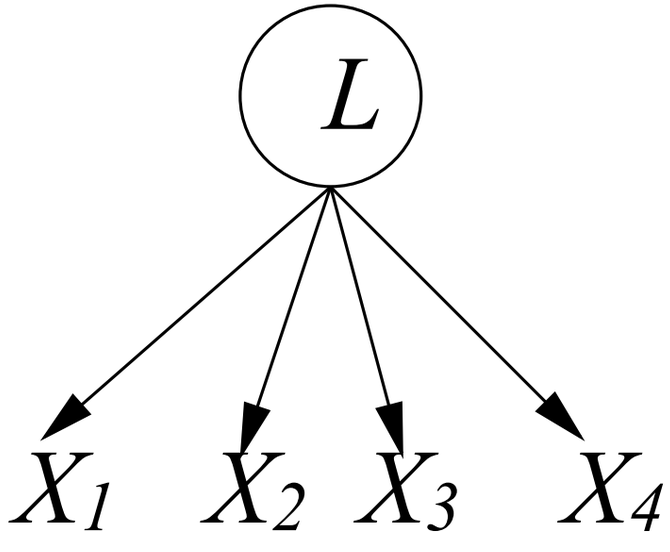


$$
\begin{aligned}
X_1 &= \lambda_1 L + \varepsilon_1 \\
X_2 &= \lambda_2 L + \varepsilon_2 \\
X_3 &= \lambda_3 L + \varepsilon_3 \\
X_4 &= \lambda_4 L + \varepsilon_4
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{X_1X_2}\sigma_{X_3X_4} &= (\lambda_1\lambda_2\sigma_L^2)(\lambda_3\lambda_4\sigma_L^2) = (\lambda_1\lambda_3\sigma_L^2)(\lambda_2\lambda_4\sigma_L^2) = \sigma_{X_1X_3}\sigma_{X_2X_4} \\
&= (\lambda_1\lambda_2\sigma_L^2)(\lambda_3\lambda_4\sigma_L^2) = (\lambda_1\lambda_4\sigma_L^2)(\lambda_2\lambda_3\sigma_L^2) = \sigma_{X_1X_4}\sigma_{X_2X_3}
\end{aligned}
$$

(Intuition! Same for correlation coefficients)

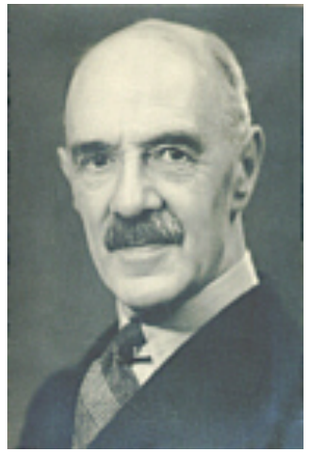
Tetrad Constraints



$$
\begin{aligned}
X_1 &= \lambda_1 L + \varepsilon_1 \\
X_2 &= \lambda_2 L + \varepsilon_2 \\
X_3 &= \lambda_3 L + \varepsilon_3 \\
X_4 &= \lambda_4 L + \varepsilon_4
\end{aligned}
$$

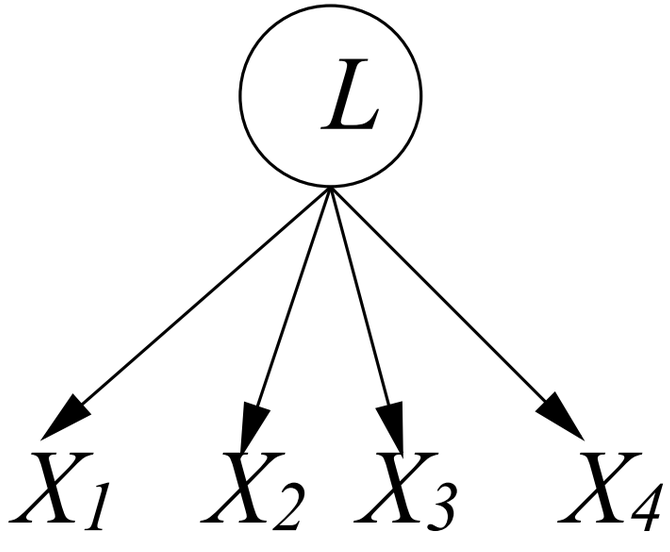
$$
\begin{aligned}
\sigma_{X_1X_2}\sigma_{X_3X_4} &= (\lambda_1\lambda_2\sigma_L^2)(\lambda_3\lambda_4\sigma_L^2) = (\lambda_1\lambda_3\sigma_L^2)(\lambda_2\lambda_4\sigma_L^2) = \sigma_{X_1X_3}\sigma_{X_2X_4} \\
&= (\lambda_1\lambda_2\sigma_L^2)(\lambda_3\lambda_4\sigma_L^2) = (\lambda_1\lambda_4\sigma_L^2)(\lambda_2\lambda_3\sigma_L^2) = \sigma_{X_1X_4}\sigma_{X_2X_3}
\end{aligned}
$$

Charles Spearman (1904)



Statistical Constraints → Measurement Model Structure

Tetrad Constraints



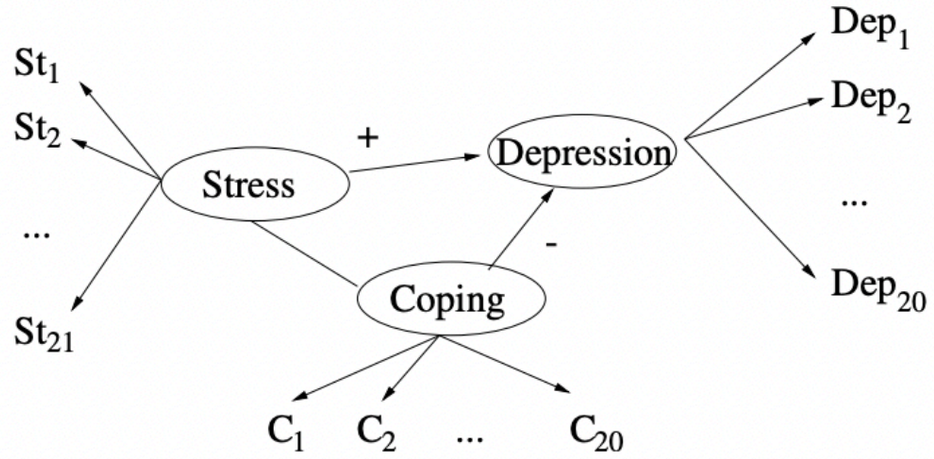
$$
\begin{aligned}
X_1 &= \lambda_1 L + \varepsilon_1 \\
X_2 &= \lambda_2 L + \varepsilon_2 \\
X_3 &= \lambda_3 L + \varepsilon_3 \\
X_4 &= \lambda_4 L + \varepsilon_4
\end{aligned}
$$

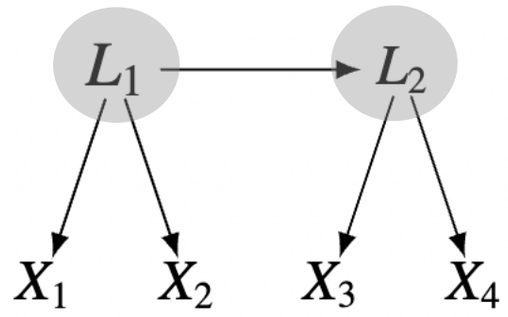
$$
\begin{aligned}
\sigma_{X_1X_2}\sigma_{X_3X_4} &= (\lambda_1\lambda_2\sigma_L^2)(\lambda_3\lambda_4\sigma_L^2) = (\lambda_1\lambda_3\sigma_L^2)(\lambda_2\lambda_4\sigma_L^2) = \sigma_{X_1X_3}\sigma_{X_2X_4} \\
&= (\lambda_1\lambda_2\sigma_L^2)(\lambda_3\lambda_4\sigma_L^2) = (\lambda_1\lambda_4\sigma_L^2)(\lambda_2\lambda_3\sigma_L^2) = \sigma_{X_1X_4}\sigma_{X_2X_3}
\end{aligned}
$$

- Identify the structure from the constraint?
  - Assumptions (LMC, faithfulness, linearity)
  - In this case we can recover the structure if X_i are correlated

An illustration of Tetrad Conditions

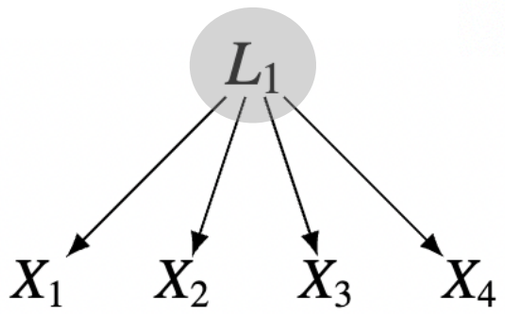
- Tetrad condition [Spearman 1904, Anderson & Rubin 1956]





(a)

$$\rho_{12}\rho_{34} \neq \rho_{23}\rho_{14}$$
$$\rho_{13}\rho_{24} = \rho_{23}\rho_{14}$$
$$\rho_{12}\rho_{34} \neq \rho_{13}\rho_{24}$$



(b)

$$\rho_{12}\rho_{34} = \rho_{23}\rho_{14}$$
$$\rho_{13}\rho_{24} = \rho_{23}\rho_{14}$$
$$\rho_{12}\rho_{34} = \rho_{13}\rho_{24}$$

X_i: measured variables
L_i: latent variables

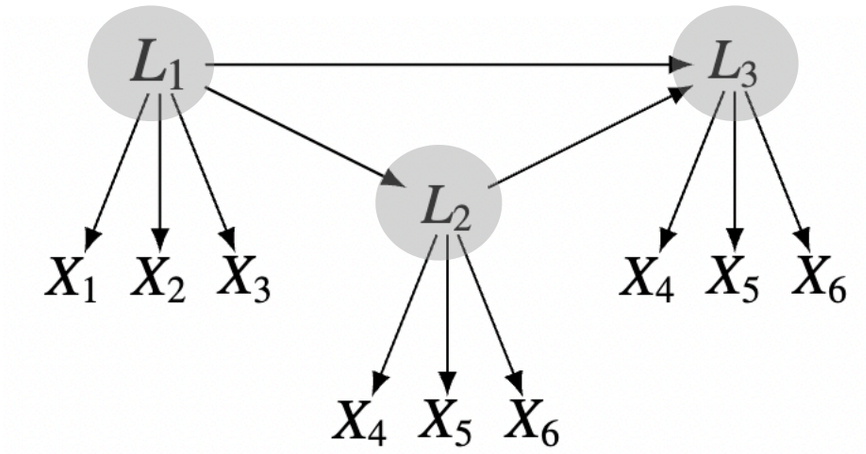
Tetrad condition:
Indicates rank deficiency of 2 x 2 off-diagonal covariance matrices

ρ_{ij} denotes the correlation coefficient between x_i and x_j

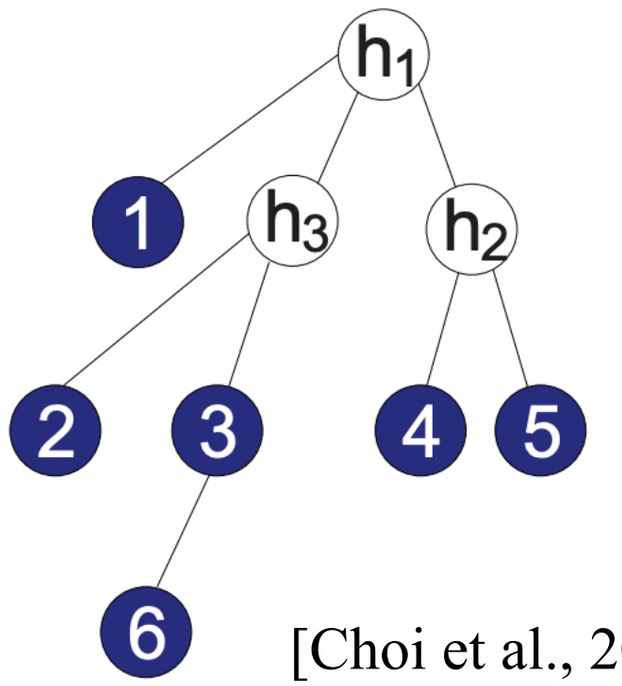
Applications of Tetrad Conditions

- One-factor measurement model [Silva et al., 2006, Kummerfeld et al., 2016]
- Tree structure [Pearl, 1988, Choi et al., 2011]

One-factor measurement model



Tree



[Choi et al., 2011]

CRL in IID Case: How to achieve it?

- Linear-Gaussian case
  - Tetrad conditions
  - Rank deficiency-based method
- Linear, non-Gaussian case
  - Theoretical results
  - GIN-based method
- Nonlinear case
  - Sparsity
- Summary: Why is it possible?

i.i.d. data?	Parametric constraints?	Latent confounders?	What can we get?
Yes	*No*	*No*	*(Different types of) equivalence class*
		Yes	
	Yes	**No**	**Unique identifiability (under structural conditions)**
		Yes	

t-separation

- Trek-separation

Definition 10.2 (Treks (Sullivant et al., 2010))**.** In a DAG $\mathcal{G}$, a trek from node X to node Y is an ordered pair of directed paths (P_1, P_2) where P_1 has a sink X, P_2 has a sink Y, and both P_1 and P_2 have the same source (the source of a directed path is the starting node from which the path originates).

Definition 10.3 (t-separation (Sullivant et al., 2010))**.** Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C_A}$, and $\mathbf{C_B}$ be four subsets of $\mathbf{V}$ in graph $\mathcal{G}$ (not necessarily disjoint). ($\mathbf{C_A}$,$\mathbf{C_B}$) t-separates $\mathbf{A}$ from $\mathbf{B}$ if for every trek (P_1,P_2) from a vertex in $\mathbf{A}$ to a vertex in $\mathbf{B}$, either P_1 contains a vertex in $\mathbf{C_A}$ or P_2 contains a vertex in $\mathbf{C_B}$.

- From the draft of "Causal Representation Learning" (distributed in class)

Why t-separation?

- t-separation and rank constraints:

Theorem 10.4 (Rank and t-separation (Sullivant et al., 2010)). *Given two sets of variables* **A** *and* **B** *from a linear model with graph* $\mathcal{G}$*, we have:*

$$rank(\Sigma_{\mathbf{A},\mathbf{B}}) \leq \min\{|\mathbf{C_A}| + |\mathbf{C_B}| : (\mathbf{C_A}, \mathbf{C_B}) \textit{ t-separates } \mathbf{A} \textit{ from } \mathbf{B} \textit{ in } \mathcal{G}\}, \quad (10.2)$$

where $\Sigma_{\mathbf{A},\mathbf{B}}$ *is the cross-covariance over* **A** *and* **B***, and equality generically holds.*

- Under the so-called rank faithfulness, the inequality relation in Theorem 10.4 becomes equality.
- Check the rank of sub-matrices of the covariance matrix and infer the minimal size of $|\mathbf{C_A}|+|\mathbf{C_B}|$ for **A** and **B** to be t-separated, and infer relations

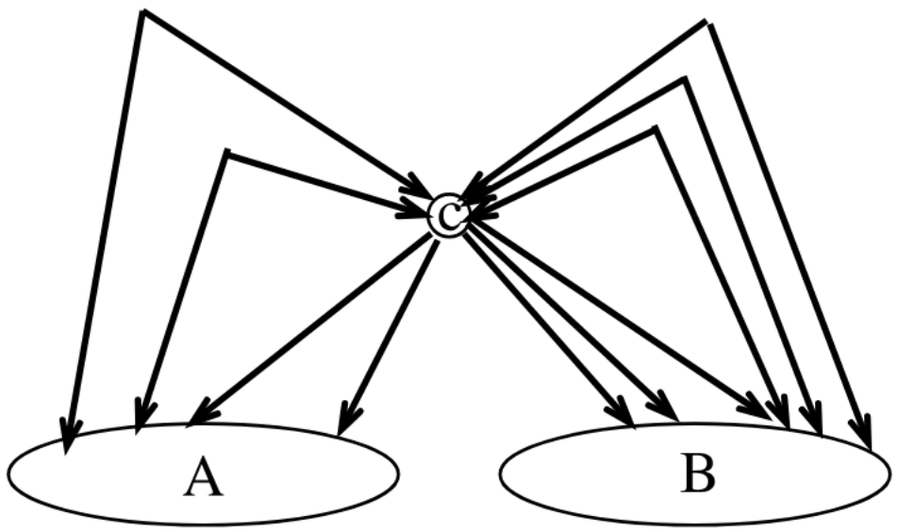
Connection to d-separation

Theorem 10.5 (t- and d-sep (Di, 2009))**.** *For disjoint sets* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$, $\mathbf{C}$ *d-separates* $\mathbf{A}$ *and* $\mathbf{B}$ *in graph* $\mathcal{G}$*, iff there is a partition* $\mathbf{C} = \mathbf{C_A} \cup \mathbf{C_B}$ *such that* $(\mathbf{C_A}, \mathbf{C_B})$ *t-separates* $\mathbf{A} \cup \mathbf{C}$ *from* $\mathbf{B} \cup \mathbf{C}$*.*

- But it can be more informative than d-separation:

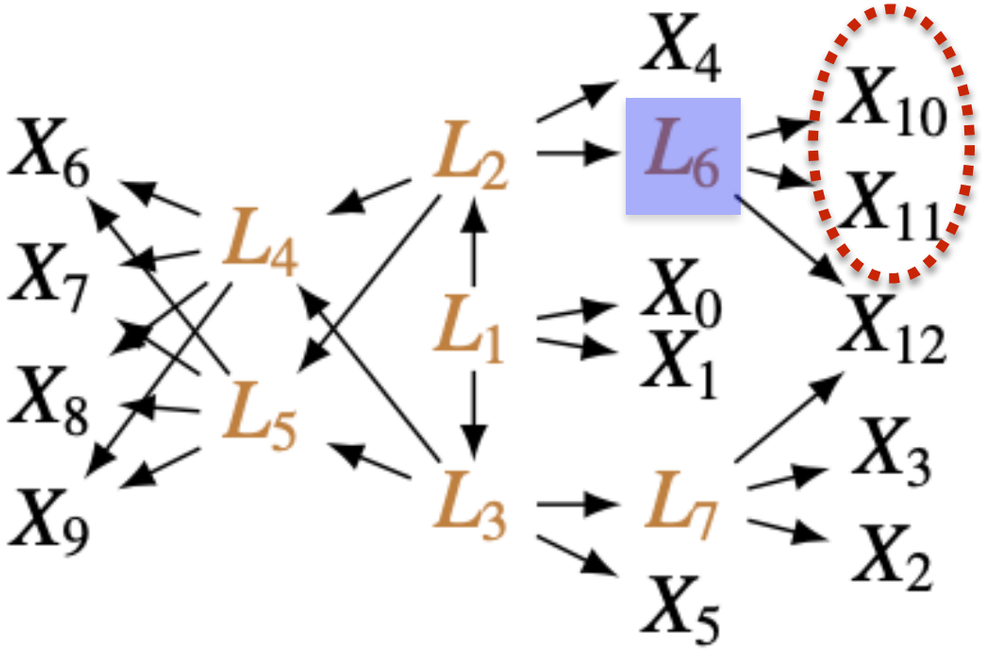
EXAMPLE 2.13 (Spiders). Consider the graph in Figure 2 which we call a *spider*.

Clearly, we have that $(\{c\}, \{c\})$ t-separates A from B, so that the submatrix $\Sigma_{A,B}$ has rank at most 2. Although this rank condition must be implied by CI rank constraints on Σ and the fact that Σ is positive definite, it does not appear to be easily derivable from these constraints.



- S. Sullivant K. Talaska, and J. Draisma, “Trek separation for Gaussian graphical models”, Annals of Statistics, 2008

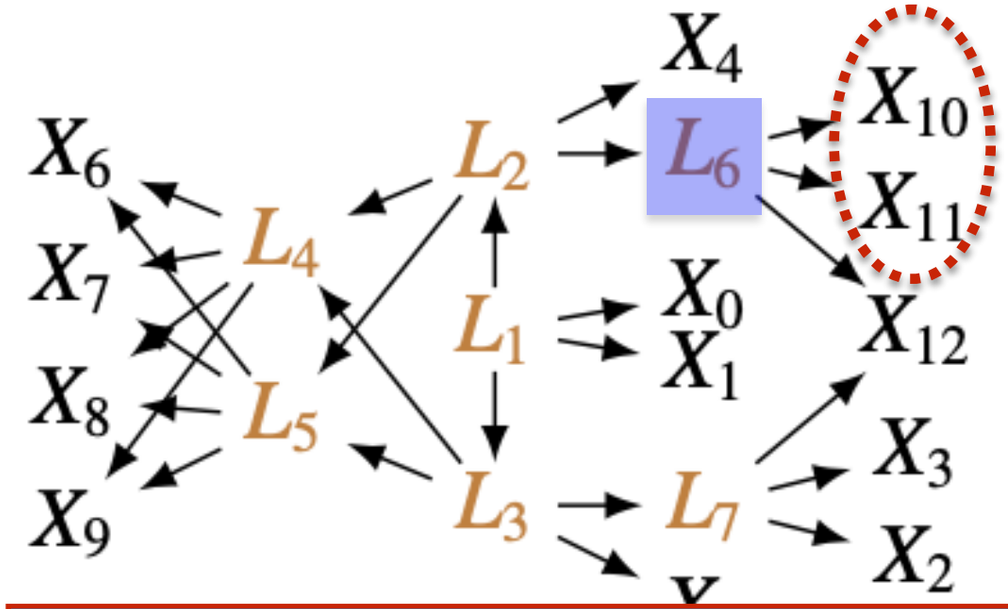
Linear, Gaussian Case: With **Rank Deficiency** Constraints



- Can we find L_6?
  - $\Sigma_{(X_{10},X_{11}),\, \mathbf{X}\setminus\{X_{10},X_{11}\}} = 1$
- Recovering the equivalence class
  - With rank deficiency of cross-covariance matrices
  - recursively and cleverly

- *Huang, Low, Xie, Glymour, Zhang, "Latent Hierarchical Causal Structure Discovery with Rank Constraints," NeurIPS 2022*

Linear, Gaussian Case: With **Rank Deficiency** Constraints



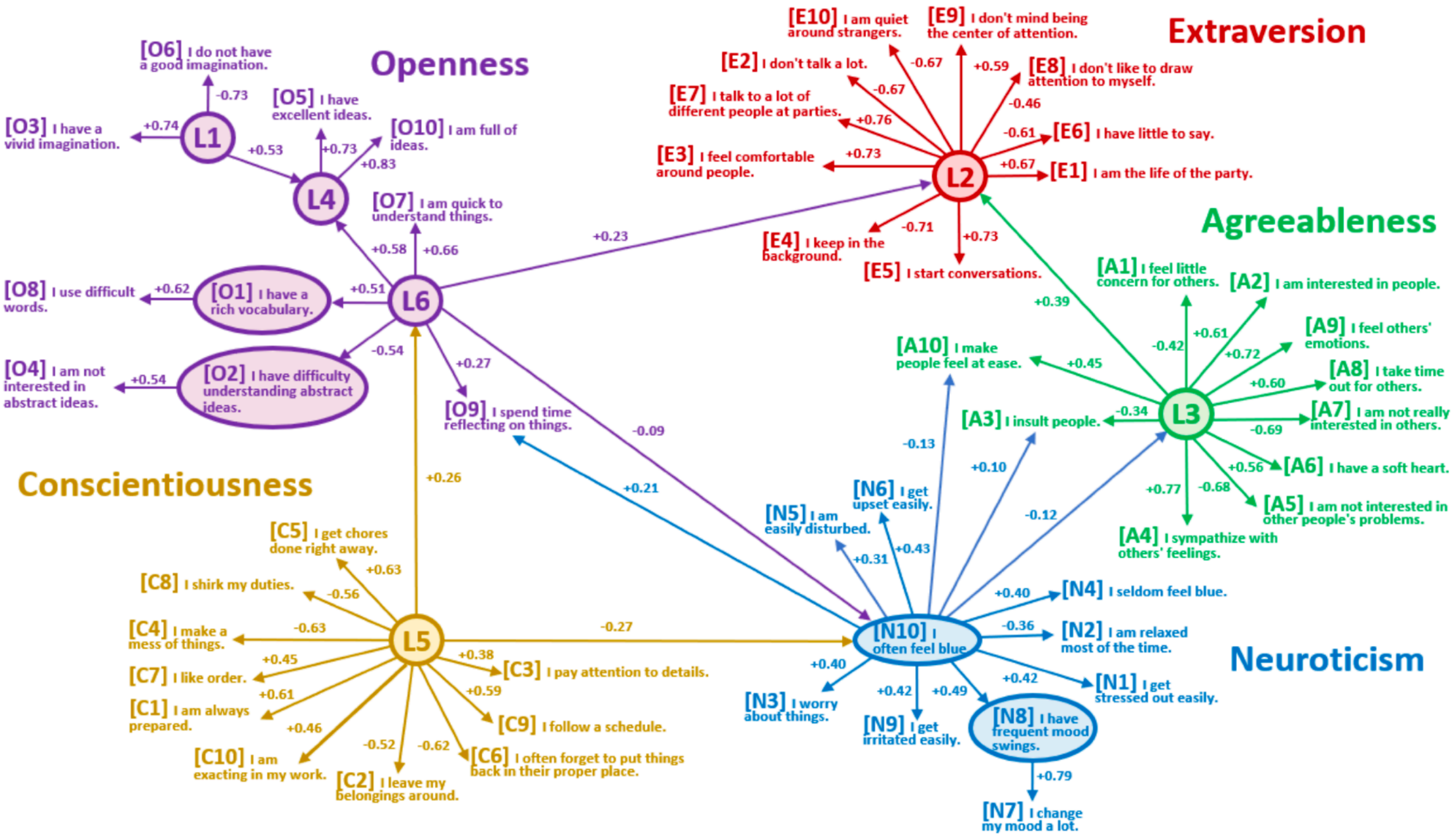
- Can we find L_6?
  - $\Sigma_{(X_{10},X_{11}),\ \mathbf{X}\setminus\{X_{10},X_{11}\}} = 1$
- Recovering the equivalence class
  - With rank deficiency of cross-covariance matrices

- Conditional independence is a special case
  - $\text{rank}(\Sigma_{(X1,\ X2),\ (X2,\ X3)}) = 1 \Leftrightarrow X_1 \perp\!\!\!\perp X_3 \mid X_2$
- Unified causal discovery based on rank deficiency constraints

- Dong, Huang, Ng, Song, Zheng, Jin, Legaspi, Spirtes, Zhang, "A Versatile Causal Discovery Framework to Allow Causally-Related Hidden Variables," ICLR 2024

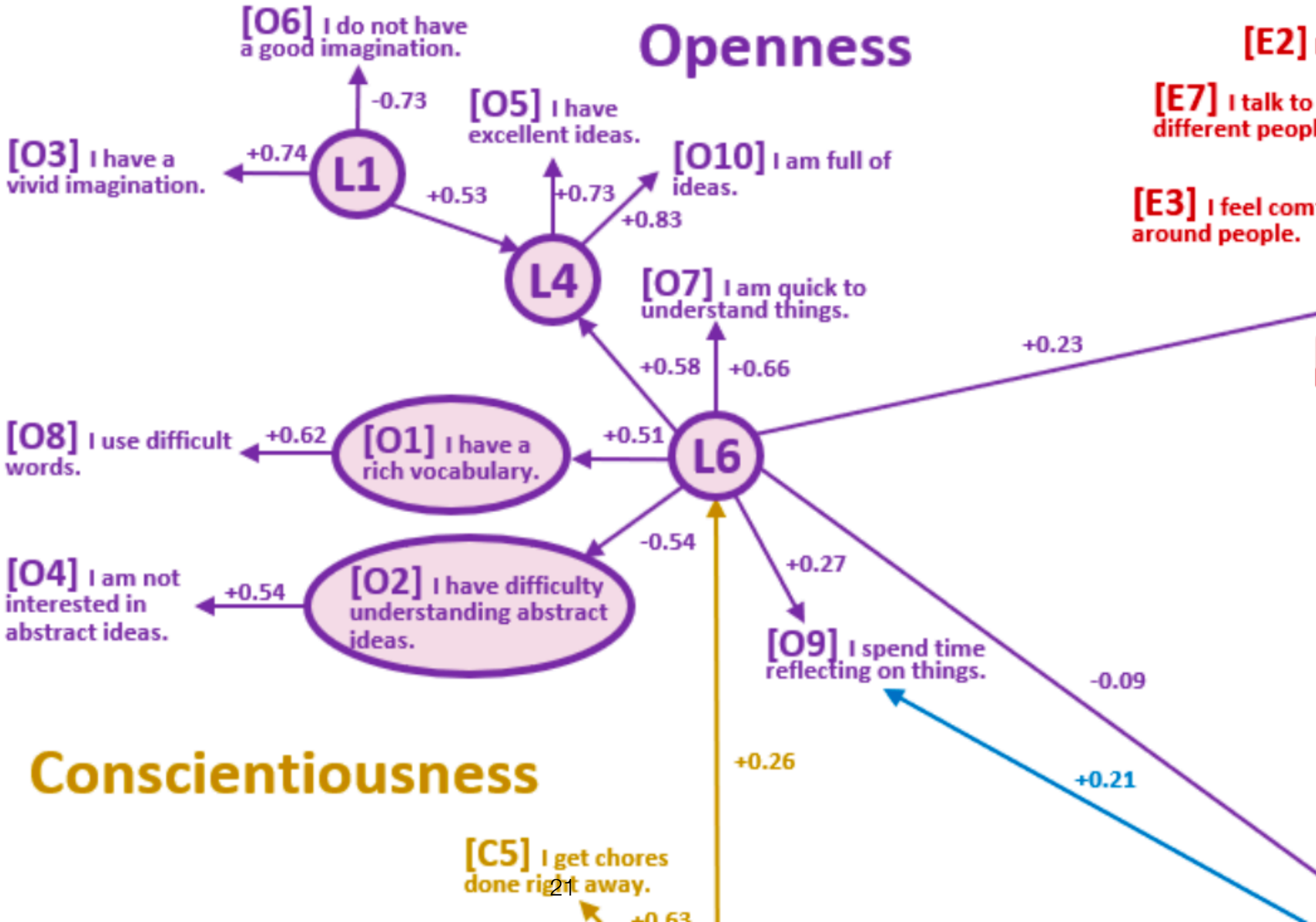
Example: Big 5 Questions Are Well Designed but…

Big 5: openness; conscientiousness; extraversion; agreeableness; neuroticism



- *Dong, Huang, Ng, Song, Zheng, Jin, Legaspi, Spirtes, Zhang, "A Versatile Causal Discovery Framework to Allow Causally-Related Hidden Variables," ICLR 2024*

Example: Big 5 Questions Are Well Designed but…



- *Dong, Huan*

CRL in IID Case: How to achieve it?

- Linear-Gaussian case
  - Tetrad conditions
  - Rank deficiency-based method
- Linear, non-Gaussian case
  - Theoretical results
  - GIN-based method
- Nonlinear case
  - Sparsity
- Summary: Why is it possible?

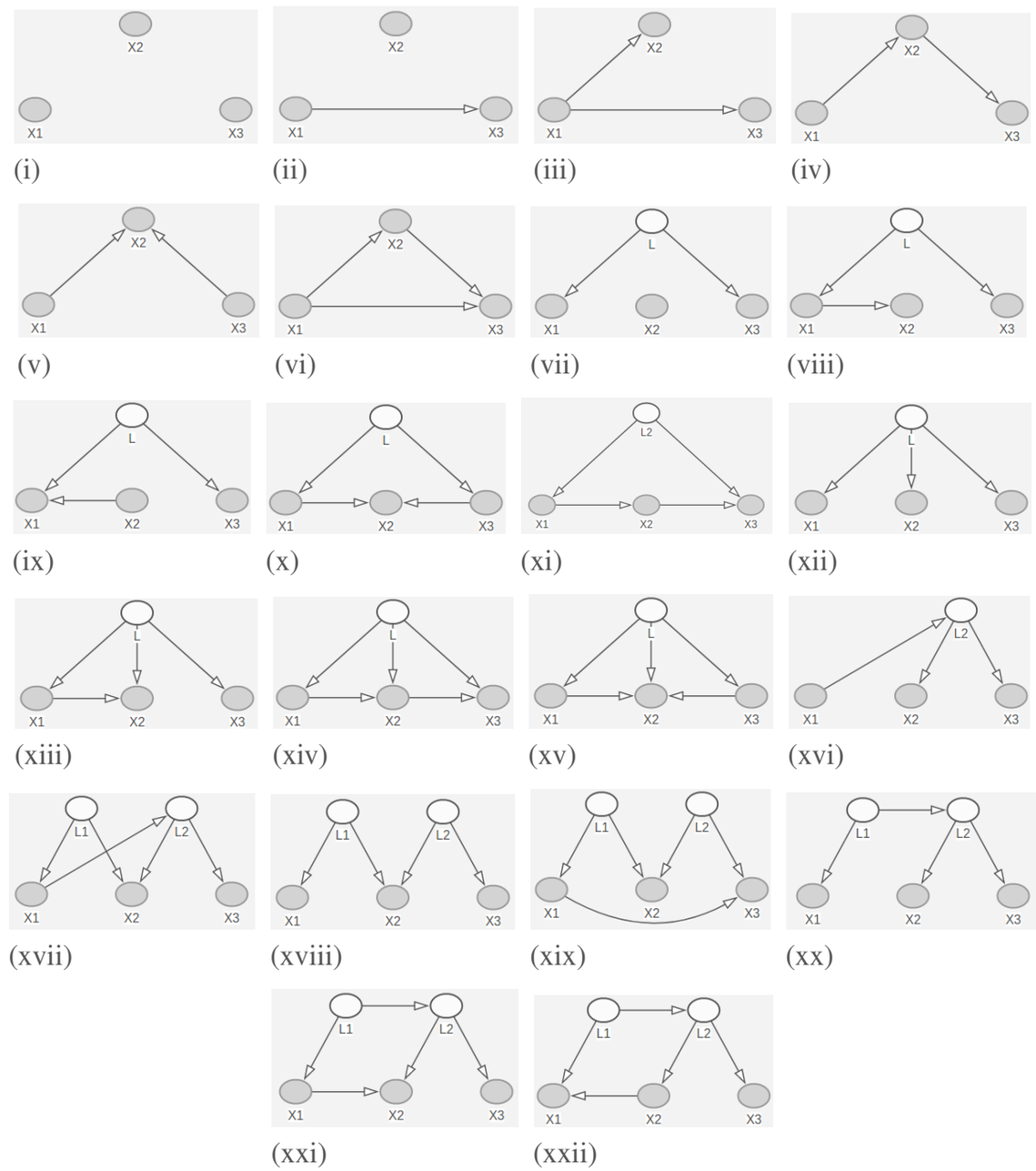
i.i.d. data?	Parametric constraints?	Latent confounders?	What can we get?
Yes	*No*	*No*	*(Different types of) equivalence class*
		Yes	
	Yes	**No**	**Unique identifiability (under structural conditions)**
		Yes	

Necessary & Sufficient Conditions on the Structure: Linear, non-Gaussian case

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

- Allow a large number of latent variables
- Minimality has to be assumed
- Estimation is generally difficult

Identifiable graphs with only 3 measured variables



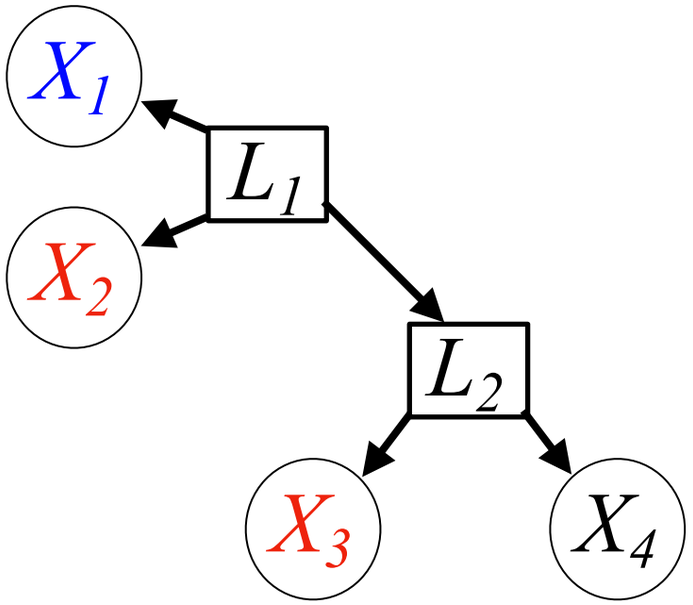
- Adams, Hansen, Zhang, "Identification of Partially Observed Linear Causal Models: Graphical Conditions for the Non-Gaussian and Heterogeneous Cases," NeurIPS 2021

CRL in IID Case: How to achieve it?

- Linear-Gaussian case
  - Tetrad conditions
  - Rank deficiency-based method
- Linear, non-Gaussian case
  - Theoretical results
  - GIN-based method
- Nonlinear case
  - Sparsity
- Summary: Why is it possible?

i.i.d. data?	Parametric constraints?	Latent confounders?	What can we get?
Yes	*No*	*No*	*(Different types of) equivalence class*
		Yes	
	Yes	**No**	**Unique identifiability (under structural conditions)**
		Yes	

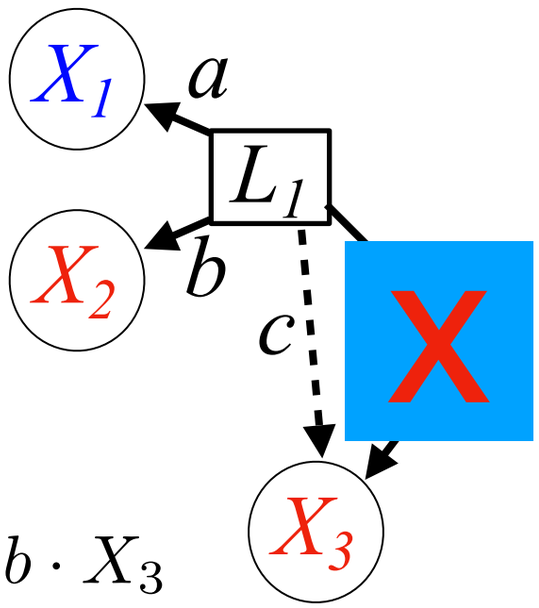
Linear, Non-Gaussian Case: Generalized Independent Noise Condition



- Find direction between latent variables L_1 and L_2?

- *Xie, Cai, Huang, Glymour, Hao, Zhang, "Generalized Independent Noise Condition for Estimating Linear Non-Gaussian Latent Variable Causal Graphs," NeurIPS 2020*
- *Cai, Xie, Glymour, Hao, Zhang, "Triad Constraints for Learning Causal Structure of Latent Variables," NeurIPS 2019*

Linear, Non-Gaussian Case: Generalized Independent Noise Condition



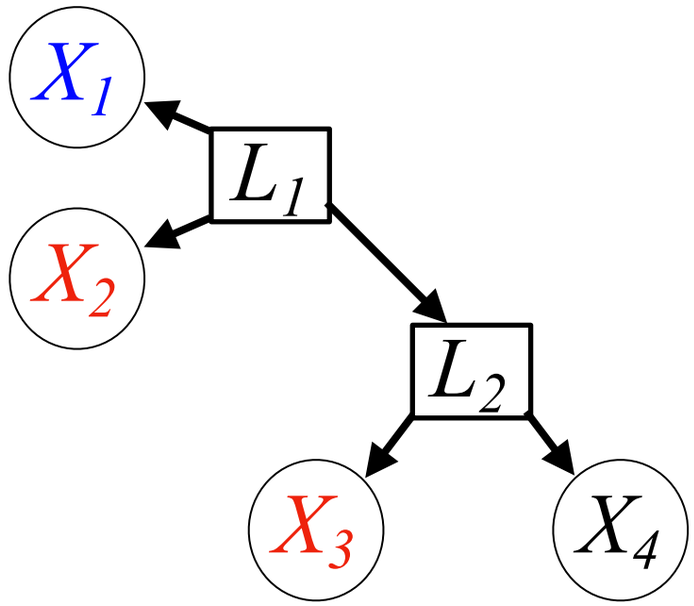
$c \cdot X_2 - b \cdot X_3$

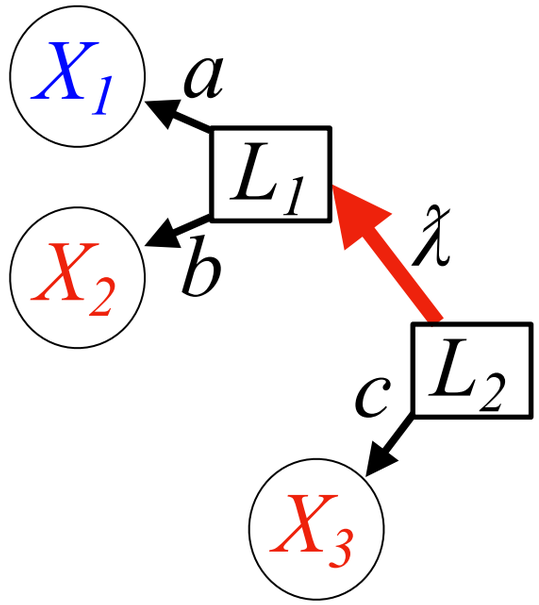
$= c(bL1 + E_2) - b(cL_1 + E_3)$

$= cE_2 - bE_3,$

independent from L_1 and from X_1,

and we know $\frac{b}{c} = \frac{Cov(X_2, X_3)}{Cov(X_1, X_3)}$

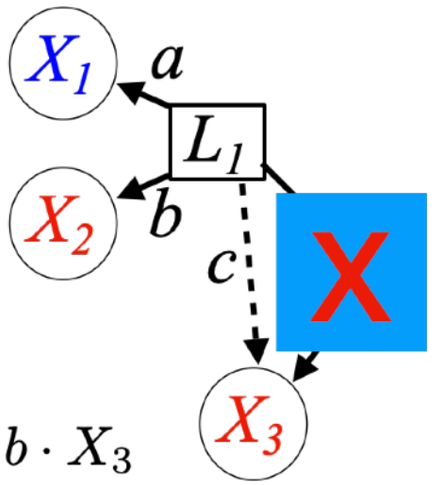




Nontrivial linear combination of X_2 and X_3 will involve the noise term in L_1, hence dependent on X_1

- *Xie, et al., "Generalized Independent Noise Condition for Estimating Linear Non-Gaussian Latent Variable Causal Graphs," NeurIPS 2020*

Linear, Non-Gaussian Case: Generalized Independent Noise Condition



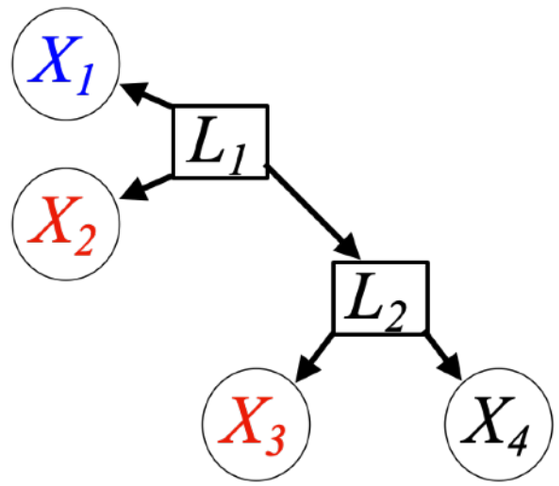
$c \cdot X_2 - b \cdot X_3$

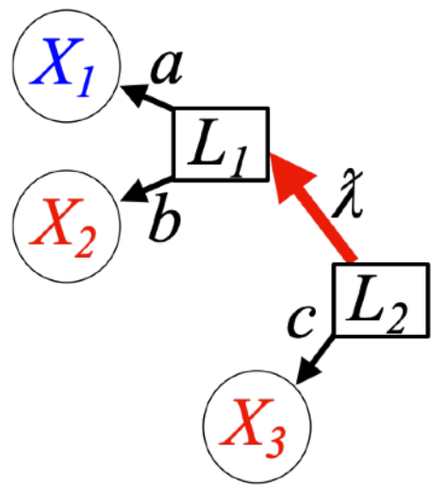
$=c(bL1 + E_2) - b(cL_1 + E_3)$

$=cE_2 - bE_3,$

independent from L_1 and from X_1,

and we know $\frac{b}{c} = \frac{Cov(X_2, X_3)}{Cov(X_1, X_3)}$





Nontrivial linear combination of X_2 and X_3 will involve the noise term in L_1, hence dependent on X_1

Let **Z** = $\{X_1\}$ and **Y** = $\{X_2, X_3\}$, GIN!

- GIN condition: (**Z**, **Y**) follows GIN $\Leftrightarrow w^\intercal \mathbf{Y} \perp\!\!\!\perp \mathbf{Z}$ for nonzero w
  - has graphical implications

Linear, Non-Gaussian Case: Generalized Independent Noise Condition

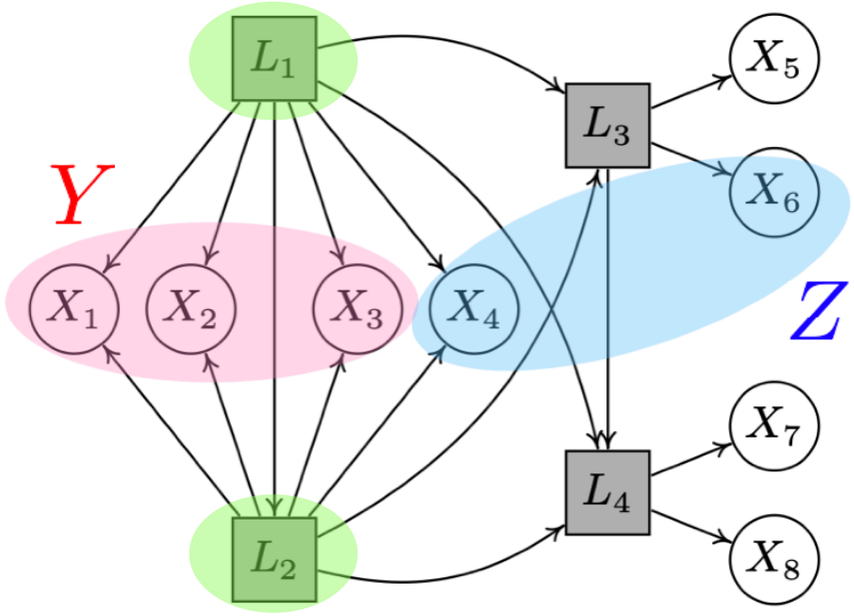
- Generalized Independent Noise (GIN) Condition:

  (Z, Y) follows the GIN condition $\iff \omega^\top Y \perp\!\!\!\perp Z$,

  where $\omega^\top \mathrm{Cov}(Y, Z) = 0$ and $\omega \neq 0$

- Graphical criterion

  (Z, Y) follows the GIN condition iff there is an exogenous set S of PA(Y) that blocks all paths between Y and Z, where 0<=|S|<=min(|Z|, |Y|-1)



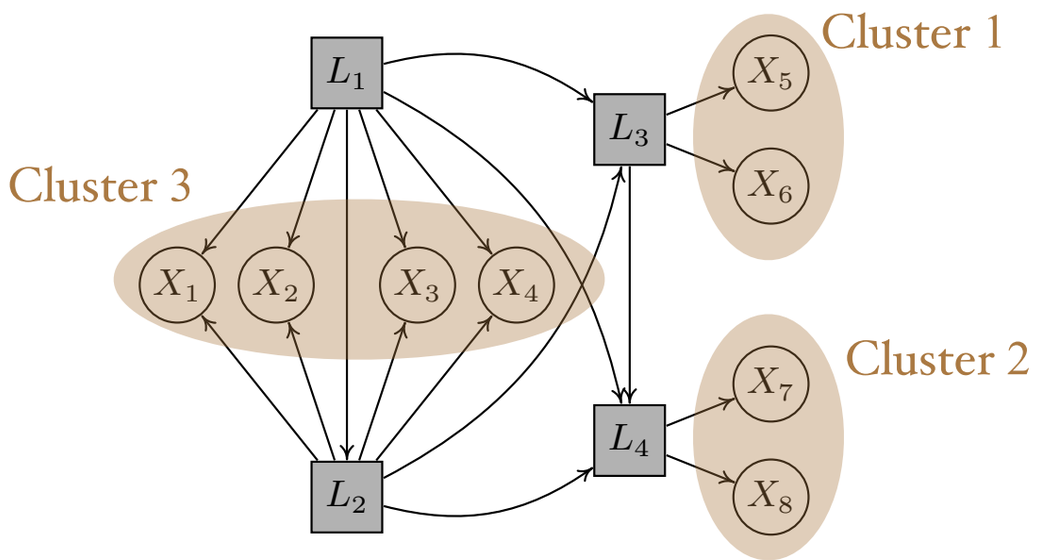
X_i: observed variables
L_i: latent variables

- *Xie, Cai, Huang, Glymour, Hao, Zhang, "Generalized Independent Noise Condition for Estimating Linear Non-Gaussian Latent Variable Causal Graphs," NeurIPS 2020*
- *Cai, Xie, Glymour, Hao, Zhang, "Triad Constraints for Learning Causal Structure of Latent Variables," NeurIPS 2019*

GIN for Estimating Linear, Non-Gaussian LV Model

- A two-step algorithm to identify the latent variable graph
  - By testing for GIN conditions over the input $X_1, \cdots, X_8$

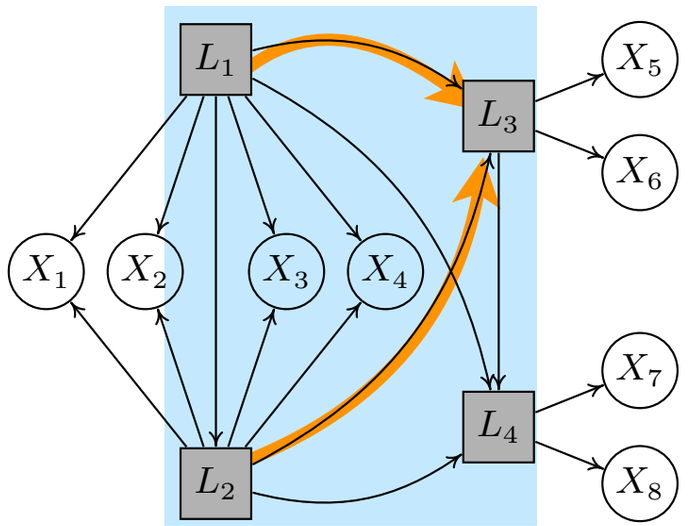
Step 1: find ***causal clusters***



$$(\overbrace{\{X_1, \cdots, X_4, X_7, X_8\}}^{\mathbf{Z}}, \overbrace{\{X_5, X_6\}}^{\mathbf{Y}})$$

satisfies GIN condition

Step 2: determine ***causal structure*** of the latent variables



$$(\overbrace{\{X_3, X_4\}}^{\mathbf{Z}}, \overbrace{\{X_1, X_2, X_5\}}^{\mathbf{Y}})$$

Cluster 3 Cluster 1 & 3

satisfies GIN condition

GIN-Based Method: Application to Teacher's Burnout Data

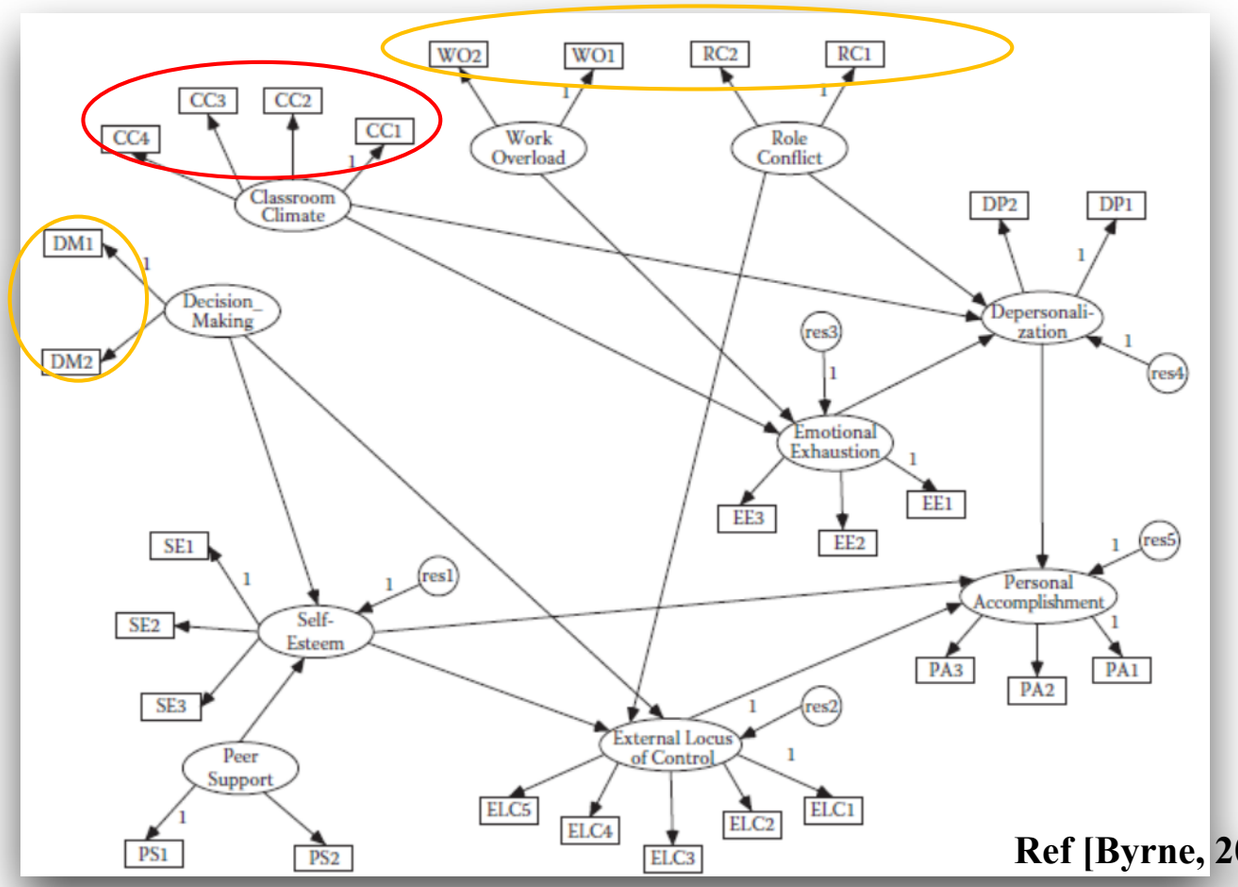
- Contains 28 measured variables
- Discovered clusters and causal order of the latent variables:

Causal Clusters	Observed variables
$\mathcal{S}_1$ (1)	$RC_1, RC_2, WO_1, WO_2, DM_1, DM_2$
$\mathcal{S}_2$ (1)	CC_1, CC_2, CC_3, CC_4
$\mathcal{S}_3$ (1)	PS_1, PS_2
$\mathcal{S}_4$ (1)	$ELC_1, ELC_2, ELC_3, ELC_4, ELC_5$
$\mathcal{S}_5$ (2)	$SE_1, SE_2, SE_3, EE_1, EE_2, EE_3, DP_1, PA_3$
$\mathcal{S}_6$ (3)	DP_2, PA_1, PA_2

$$L(\mathcal{S}_1) \succ L(\mathcal{S}_2) \succ L(\mathcal{S}_3) \succ L(\mathcal{S}_5) \succ L(\mathcal{S}_4) \succ L(\mathcal{S}_6).$$

(from root to leaf)

Hypothesized model by experts



Ref [Byrne, 2010]

- Consistent with the hypothesized model

- *Xie, Cai, Huang, Glymour, Hao, Zhang, "Generalized Independent Noise Condition for Estimating Linear Non-Gaussian Latent Variable Causal Graphs," NeurIPS 2020*
- *Cai, Xie, Glymour, Hao, Zhang, "Triad Constraints for Learning Causal Structure of Latent Variables," NeurIPS 2019*

CRL in IID Case: How to achieve it?

- Linear-Gaussian case
  - Tetrad conditions
  - Rank deficiency-based method
- Linear, non-Gaussian case
  - Theoretical results
  - GIN-based method
- Nonlinear case
  - Sparsity
- Summary: Why is it possible?

i.i.d. data?	Parametric constraints?	Latent confounders?	What can we get?
Yes	*No*	*No*	*(Different types of) equivalence class*
		Yes	
	Yes	**No**	**Unique identifiability (under structural conditions)**
		Yes	

Nonlinear Cases Generally Non-Identifiable

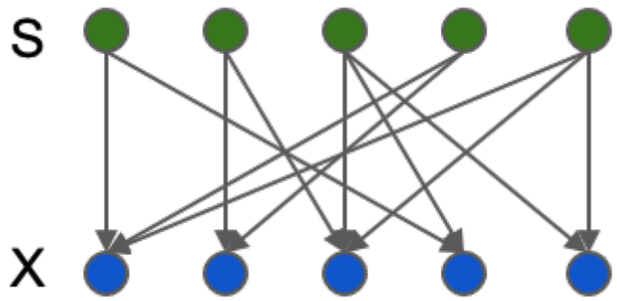
i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

- Nonlinear ICA:
  - Generative model: $\mathbf{X} = f(\mathbf{S})$, where $\mathbf{S}$ has independent components
  - De-mixing procedure: $\mathbf{Y} = g(\mathbf{X})$, where $\mathbf{Y}$ components are as independent as possible
  - Solutions always exist and are highly non-unique: Why?

- *Zheng, Ng, Zhang, On the Identifiability of Nonlinear ICA: Sparsity and Beyond, NeurIPS 2022*

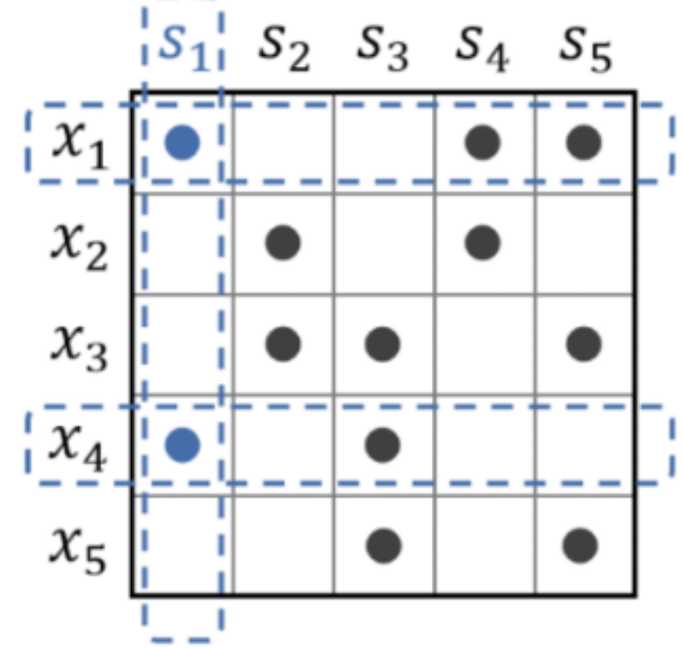
Identifiability of Nonlinear ICA: Structural Sparsity

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes



(Structural Sparsity) For all $k \in \{1, \ldots, n\}$, there exists $\mathcal{C}_k$ such that

$$\bigcap_{i \in \mathcal{C}_k} \operatorname{supp}(\mathbf{J_f}(\mathbf{s})_{i,:}) = \{k\}.$$

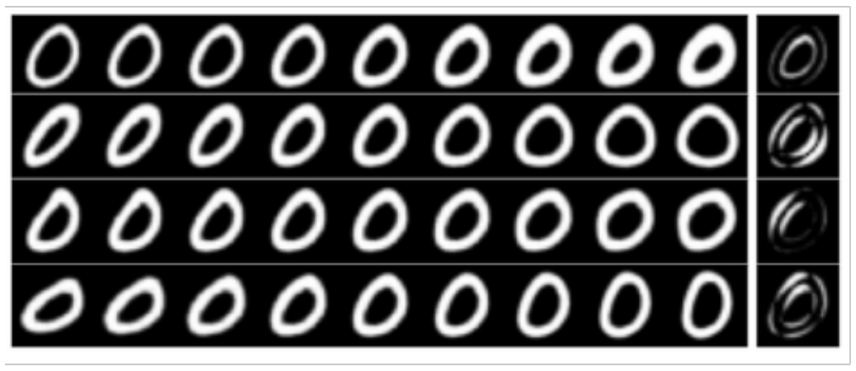


- **Graphically**, for every latent variable S_i, there exists a set of observed variable(s) such that the intersection of their/its parent(s) is S_i
- **Example:** for S_1, there exists X_1 and X_4 such that the intersection of their parents is S_1

- *Zheng, Ng, Zhang, On the Identifiability of Nonlinear ICA: Sparsity and Beyond, NeurIPS 2022*

Further Generalization of Nonlinear ICA

- Undercompleteness
  - More observed variables than latent variables
- Partial sparsity
  - Sparsity is violated for some variables
- Partial source dependence
  - Source independence violated for some variables
- Flexible grouping structures
  - Dependence within each group, independence across groups



Applied to real-world datasets (EMNIST)

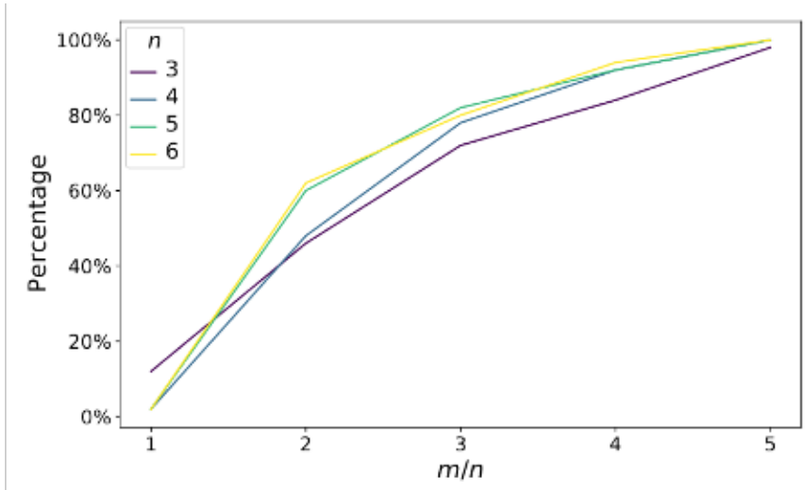


Figure 4: Percentage of random structures satisfying Structural Sparsity w.r.t. different degree of undercompleteness (i.e., m/n).

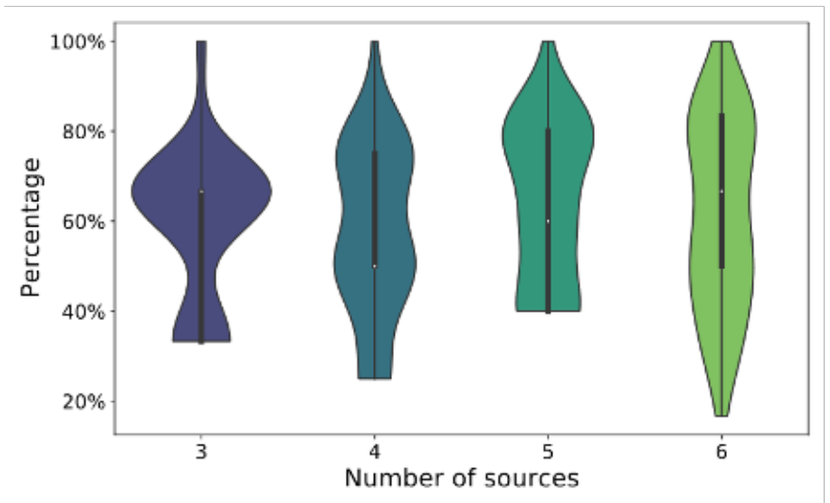
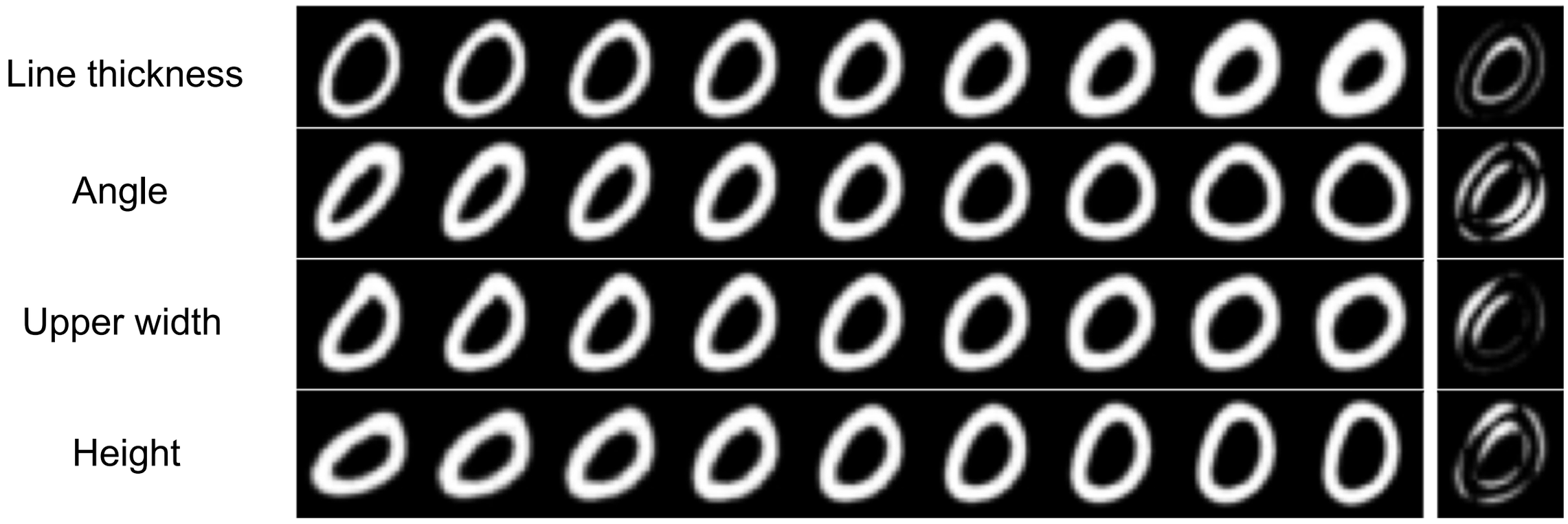


Figure 5: Percentage of sources satisfying *Structural Sparsity* w.r.t. different numbers of sources in the bijective setting ($m/n = 1$).

- *Zheng and Zhang, Generalizing Nonlinear ICA beyond Structural Sparsity, NeurIPS 2023 (oral)*

Identifiability of nonlinear ICA: real-world images



Identification results on EMNIST

Each row represents an identified source with its value varying

Summary: CRL in IID Case

- Traditional causal discovery is a special case
- Why possible?
  - Sparsity! Stronger or weaker…
    - Linear-Gaussian case: strong
    - Nonparametric case (weak parametric constraints): strong
    - Linear, non-Gaussian case (additional parametric constraints): weaker
- How to leverage non-IDD features of the data

Where Are We?

<table>
<tr><th>i.i.d. data?</th><th>Parametric constraint?</th><th>Latent confounders?</th><th>What can we get?</th></tr>
<tr><td rowspan="4">Yes</td><td rowspan="2">No</td><td>No</td><td rowspan="2">(Different types of) equivalence class</td></tr>
<tr><td>Yes</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td rowspan="2">Unique identifiability (under structural conditions)</td></tr>
<tr><td>Yes</td></tr>
<tr><td>Non-I, but I.D.</td><td>No/Yes</td><td>No</td><td rowspan="6">?</td></tr>
<tr><td></td><td></td><td>Yes</td></tr>
<tr><td rowspan="4">I., but non-I.D.</td><td>No</td><td rowspan="2">No</td></tr>
<tr><td>Yes</td></tr>
<tr><td>No</td><td rowspan="2">Yes</td></tr>
<tr><td>Yes</td></tr>
</table>