Identifiability Guarantees for Causal Disentanglement from Purely Observational Data

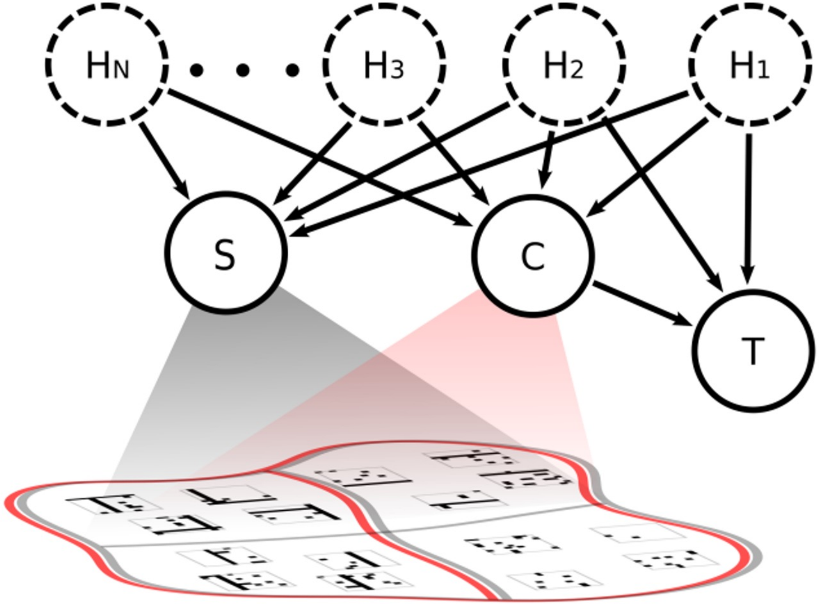
Ryan Welch*, Jiaqi Zhang*, Caroline Uhler

ERIC AND WENDY
SCHMIDT CENTER
AT BROAD INSTITUTE

NEURAL INFORMATION
PROCESSING SYSTEMS

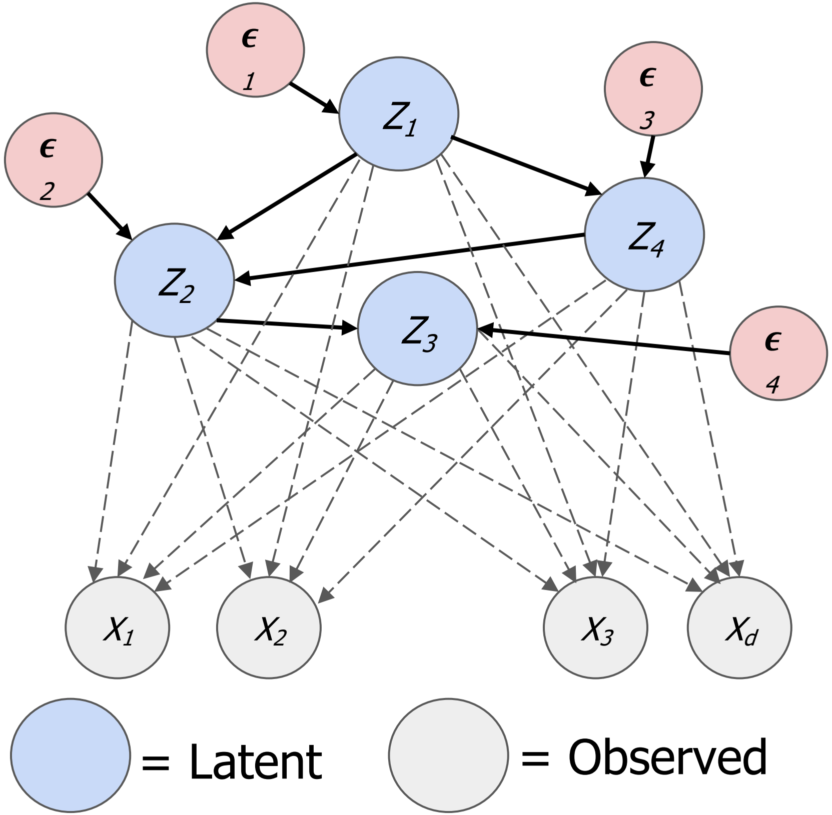
Motivation

Causal disentanglement aims to uncover the underlying causal mechanisms present in complex, unobserved systems.



Particularly useful in learning complicated gene regularly networks

Nonlinear Additive Gaussian Equation Models

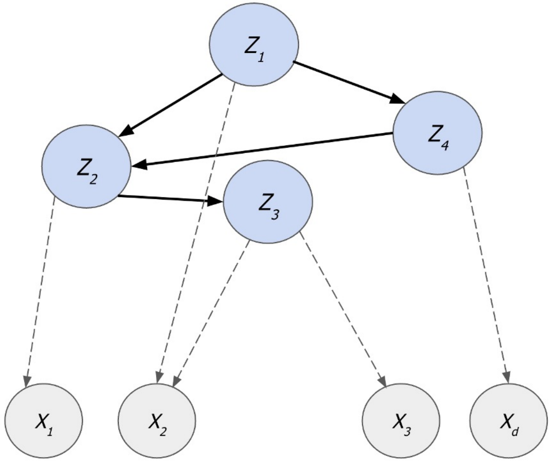


$$Z_i = f_i(Z_{pa(i)}) + \mathcal{E}_i, \qquad \forall i \in [n]$$

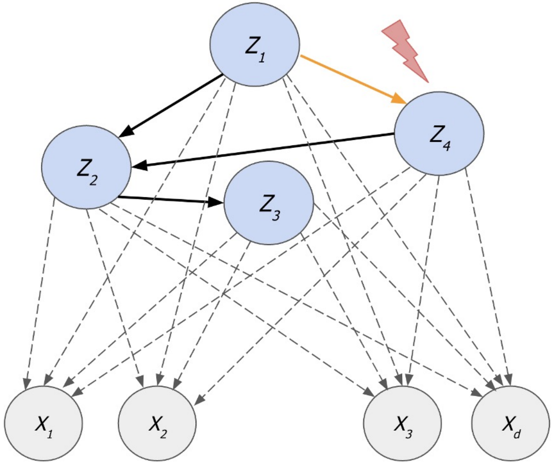
- $\mathcal{E}_i \sim \mathcal{N}(0, \sigma_i^2)$, f_i is nonlinear
- Observed $X = g(Z)$
- $g = H \in \mathbb{R}^{n \times d}$ is linear

Latent factors are identifiable with…

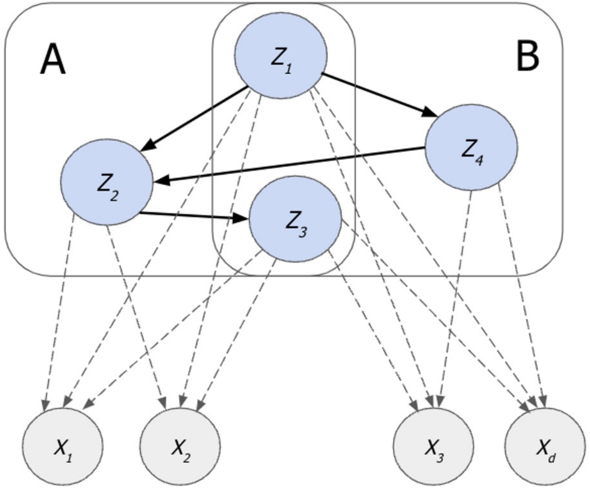
Graphical constraints on the mixing process



Access to atomic interventions

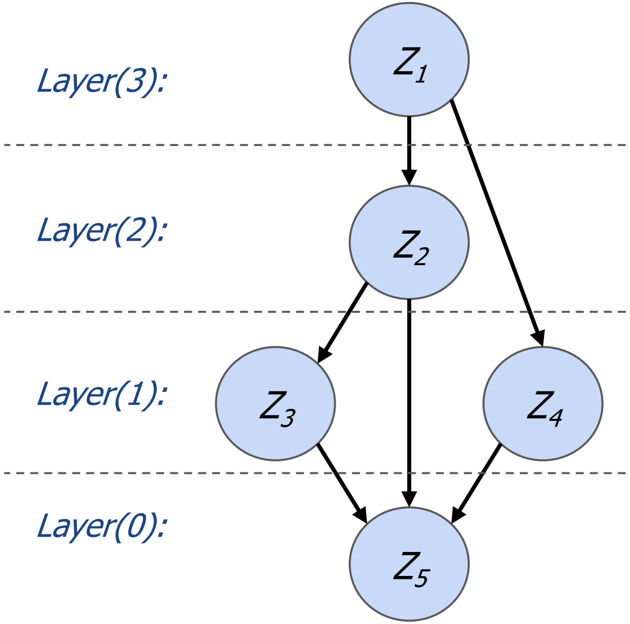


Data from multiple modalities



What is identifiable without any of the above assumptions?

Layer-wise Identifiability



Definition 1 (Identifiability up to upstream layers). *The latent causal variables Z are* identifiable up to upstream layers *if it is possible to learn $\hat{Z}(X)$ from $p_X(\cdot)$ such that:*

$$\hat{Z}(X) = P_\pi \cdot C \cdot Z, \qquad \forall Z \in \mathbb{R}^n,$$

where $P_\pi \in \mathbb{R}^{n\times n}$ is a permutation matrix, and $C \in \mathbb{R}^{n\times n}$ is a constant matrix with non-zero diagonal entries and $[C]_{i,j} = 0$ for all i, j such that $i \in layer(k)$ and $j \in \cup_{l<k} layer(l)$.

Definition 2 (Identifiability up to layers). *The exogenous noise variables $\mathcal{E}$ are* identifiable up to layers *if it is possible to learn $\hat{\mathcal{E}}(X)$ from $p_X(\cdot)$ such that:*

$$\hat{\mathcal{E}}(X) = P_\pi \cdot C \cdot \mathcal{E}, \qquad \forall \mathcal{E} \in \mathbb{R}^n,$$

where $P_\pi \in \mathbb{R}^{n\times n}$ is a permutation matrix, and $C \in \mathbb{R}^{n\times n}$ is a constant matrix with non-zero diagonal entries and $[C]_{i,j} = 0$ for all i, j such that $i \in layer(k)$ and $j \notin layer(k)$.

Layers of a causal DAG. A latent variable is contained in $layer(k)$ if its longest path to a lead node is is length k.

Preview of Main Results

Theorem 1. *Under Assumptions 1 and 2, the latent variables Z are identifiable up to their upstream layers from purely observational data.*

Theorem 2. *Under Assumptions 1 and 2, the exogenous noise variables $\mathcal{E}$ are identifiable up to their layers from purely observational data.*

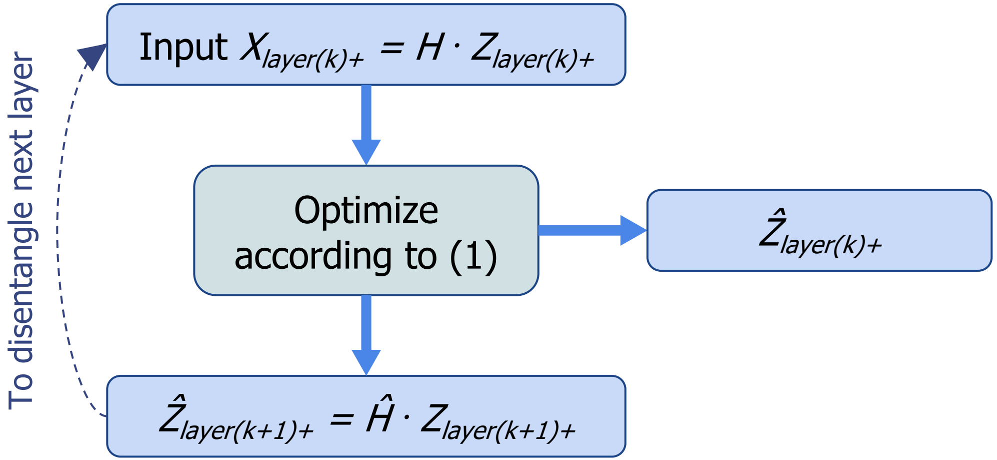
Proposition 1. *Under Assumptions 1 and 2, the exogenous noise variables $\mathcal{E}$ are generally unidentifiable beyond layer-wise transformation from observational data.*

Assumption 1: Linear mixing
Assumption 2: Nonlinear additive Gaussian noise model

Learning Latent Variable Representations

$$\begin{aligned} \min_{\hat{H} \in \mathbb{R}^n} \quad & \|\mathrm{Var}\big(\mathrm{diag}(\hat{H}^\top J_X(\hat{H}^\dagger x)\hat{H})\big)\|_0, \\ \text{such that} \quad & \mathrm{rank}(\hat{H}) = n, \end{aligned}$$

$$\hat{Z}_i = \begin{cases} \mathrm{linear}(Z_{non-leaf}) & \text{if } \mathrm{Var}\left(\left[J_{\hat{Z}}(\hat{z})\right]_{ii}\right) \neq 0, \\ \mathrm{linear}(Z) & \text{if } \mathrm{Var}\left(\left[J_{\hat{Z}}(\hat{z})\right]_{ii}\right) = 0. \end{cases}$$



Quadratic Programming on Estimated Scores

Can solve as a rank-constrained optimization problem:

$$\hat{H} = \arg\min_{\hat{H}\in\mathbb{R}^n} \left\| Var\Big(diag(J_{\hat{Z}}(\hat{H}^\dagger x))\Big)\right\|_0 , \quad \text{such that} \quad \text{rank}(\hat{H}) = n$$

Can solve iteratively by column as a QCQP:

$$\begin{aligned}[\hat{H}]_i = \arg\min_{h\in\mathbb{R}^n} \quad & 0 \\ \text{such that} \quad & h^\top \tilde{J}_X(x^{(m)})h = 0, \quad \forall m \in [N], \\ & h^\top h = 1, \\ & h^\top [\hat{H}]_j = 0, \quad \forall j \in [i-1], \end{aligned}$$

$$\text{where } \tilde{J}_X(x^{(m)}) \triangleq \hat{J}_X(x^{(m)}) - \Big(\tfrac{1}{N}\textstyle\sum_{m=1}^N \hat{J}_X(x^{(m)})\Big)$$

Discontinuous and Non-convex

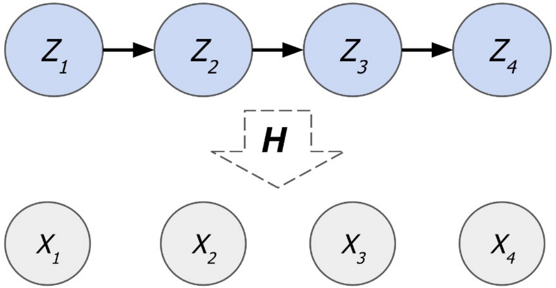
d x n dimensions

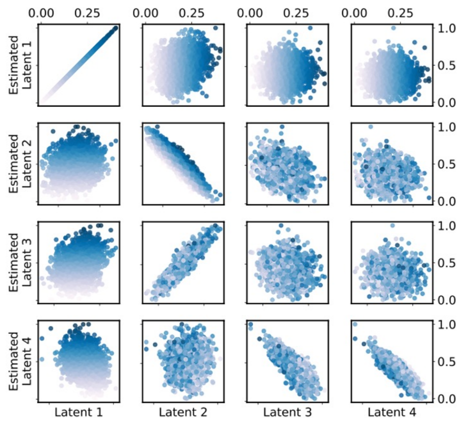
Continuous

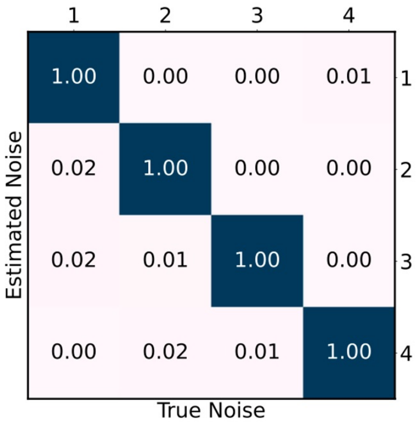
n dimensions

Results on Synthetic Data

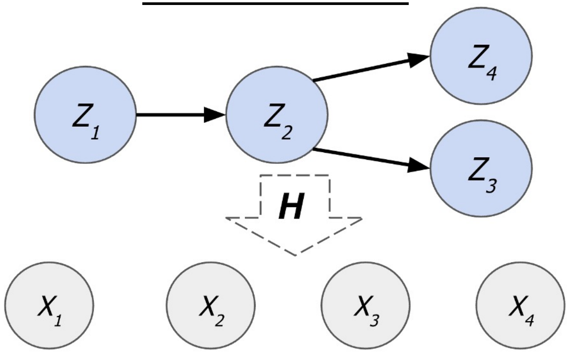
Line Graph

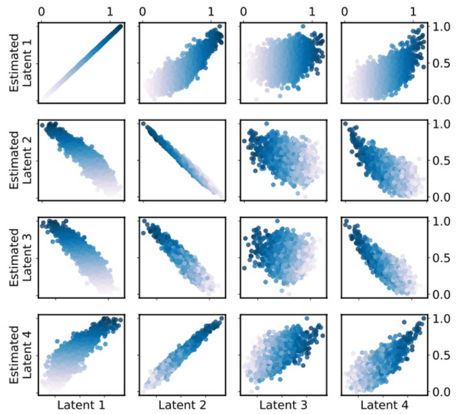


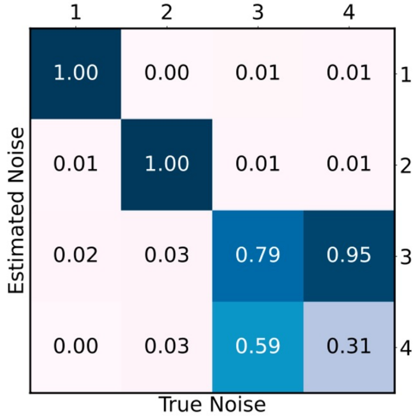




Y-Structure







Summary

- Prove that latent causal variables can be disentangled up to their upstream layer representations
- Present practical algorithm to perform such disentanglement
- Validate our theory and algorithm with experiments on synthetic data