Linear Causal Disentanglement via Interventions

Chandler Squires[13*], Anna Seigal[12*], Salil Bhate[1*], Caroline Uhler[13*]

[1]Broad Institute of Harvard and MIT [2]School of Engineering and Applied Sciences, Harvard, Harvard
[3]Institute for Data, Systems, and Society, MIT *Equal contribution

IDENTIFIABLE REPRESENTATION LEARNING

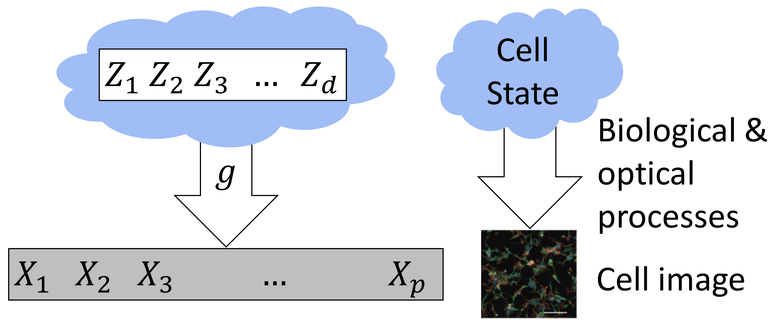
Identifiability: Does a unique **generative model (theory)** explain the **observed data**?

(Latent) *Macro-variables*

(Unknown) *Mixing function*

(Observed) *Micro-variables*

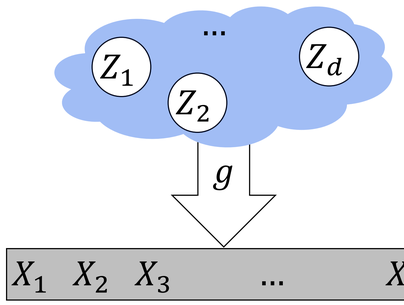
Example



Causal representation learning/causal disentanglement: We assume $\mathbb{P}(Z)$ is Markov to a DAG $\mathcal{G}$ (w.l.o.g., assume $(1, 2, \dots, d)$ is a topological order of $\mathcal{G}$).

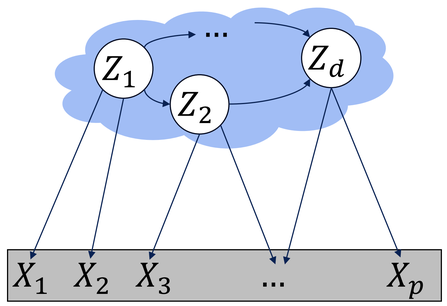
COMMON APPROACHES TO ESTABLISHING IDENTIFIABILITY

Restrict the latent distribution



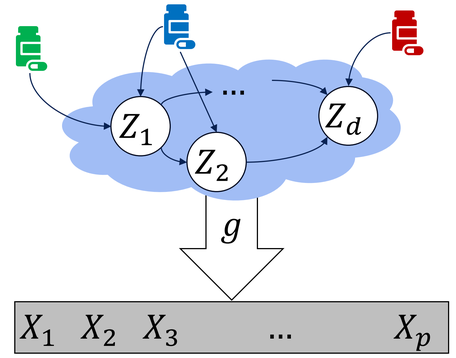
Independent component analysis [4, 5]

Restrict the mixing function



Most prior work on latent DAG recovery [3,6]

Incorporate multiple contexts



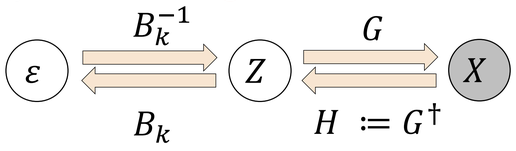
This work and other recent works [1,2,8,9,10]

ASSUMPTIONS IN OUR WORK

(a) Linear latent model: For context k, we have $Z = A_k Z + \Omega_k^{1/2}\varepsilon$. Here, the support of A_k is consistent with $\mathcal{G}$, Ω_k is diagonal, and $\mathrm{Cov}(\varepsilon) = I$. Define $B_k = \Omega_k^{1/2}(I - A_k)$.

(b) Single-node interventions: For context k, there exists an intervention target i_k such that only the i_k-th row of B_k changes.

(c) Linear mixing: In each context, $X = GZ$ for $G \in \mathbb{R}^{p\times d}$ full rank.



Under these assumptions:
$\Theta_k := \mathrm{Cov}_k(X)^\dagger = H^\top B_k^\top B_k H$

OUR IDENTIFIABILITY GUARANTEES

One **perfect** intervention per variable is sufficient, and in the worst case necessary to identify $\mathcal{G}$ and $(B_0, B_1, \dots, B_K, G)$.

One **soft** intervention per variable is sufficient, and in the worst case necessary to identify $\mathcal{G}$ up to transitive closure.

Proof components

Key Identity: $\Theta_k - \Theta_0 = \underbrace{\left(H^\top B_k^\top \boldsymbol{e}_{i_k}\right)^{\otimes 2} - \left(H^\top B_0^\top \boldsymbol{e}_{i_k}\right)^{\otimes 2}}$

Rank one if i_k is a source node, rank two otherwise. Used to find $\boldsymbol{h}_{i_k}$ when i_k is a source node.

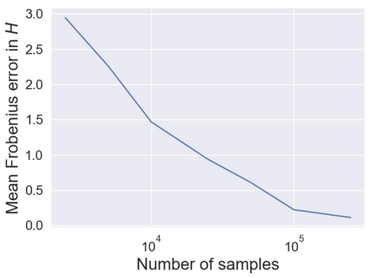
General principle: $\mathrm{rowspan}(\Theta_k - \Theta_0) \subseteq \{\boldsymbol{h}_i : i \in \mathcal{I}\}$ if and only if $\mathcal{I} = \mathrm{pa}(i_k) \cup \{i_k\}$.

This lets us iteratively recover (1) the partial order over i_k's and (2) the corresponding rows of H.

EMPIRICAL RESULTS

Synthetic data

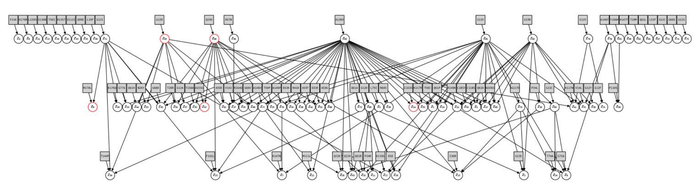
- $d = K = 5$
- $p = 10$
- 500 random models, Erdős-Rényi structure



Lung cancer scRNA-seq data [7]:

- $K = 83$ (mutations of the KRAS oncogene)
- $p = 83$ (most variable genes)

Qualitative findings: Mutations of G12 and G13 are predicted to have widespread effects; these are key functional residues, and their mutations are known drivers of cancer.



EXTENSIONS AND FUTURE WORK

1) Multi-node interventions: For each context k, allow a set I_k of intervention targets.

2) Soft interventions: See poster [10] at the Workshop on Spurious Correlations, Invariance, and Stability.

3) Non-parametric models: Latent model [1,8,10]. Mixing function [2]. Both (when $d = 2$) [9].

4) Partial identifiability: What *is* identifiable when full identifiability is not achievable? Is this sufficient for downstream tasks?

5) Statistically/computationally efficient algorithms.

REFERENCES

[1] Ahuja, K., Mahajan, D., Wang, Y., & Bengio, Y. (2023). *Interventional causal representation learning.*
[2] Buchholz, S., Rajendran, G., Rosenfeld, E., Aragam, B., Schölkopf, B., & Ravikumar, P. (2023). *Learning Linear Causal Representations from Interventions under General Nonlinear Mixing.*
[3] Cai, R., Xie, F., Glymour, C., Hao, Z., & Zhang, K. (2019). *Triad constraints for learning causal structure of latent variables.*
[4] Comon, P. (1994). *Independent component analysis, a new concept?*
[5] Hyvarinen, A., Sasaki, H., & Turner, R. (2019). *Nonlinear ICA using auxiliary variables and generalized contrastive learning.*
[6] Silva, R., Scheines, R., Glymour, C., Spirtes, P., & Chickering, D. M. (2006). *Learning the Structure of Linear Latent Variable Models.*
[7] Ursu, O., Neal, J. T., Shea, E., Thakore, P. I., Jerby-Arnon, L., Nguyen, L., ... & Boehm, J. S. (2022). *Massively parallel phenotyping of coding variants in cancer with Perturb-seq.*
[8] Varici, B., Acarturk, E., Shanmugam, K., Kumar, A., & Tajer, A. (2023). *Score-based causal representation learning with interventions.*
[9] von Kügelgen, J., Besserve, M., Liang, W., Gresele, L., Kekić, A., Bareinboim, E., ... & Schölkopf, B. (2023). *Nonparametric Identifiability of Causal Representations from Unknown Interventions.*
[10] Zhang, J., Squires, C., Greenewald, K., Srivastava, A., Shanmugam, K., & Uhler, C. (2023). *Identifiability Guarantees for Causal Disentanglement from Soft Interventions.*