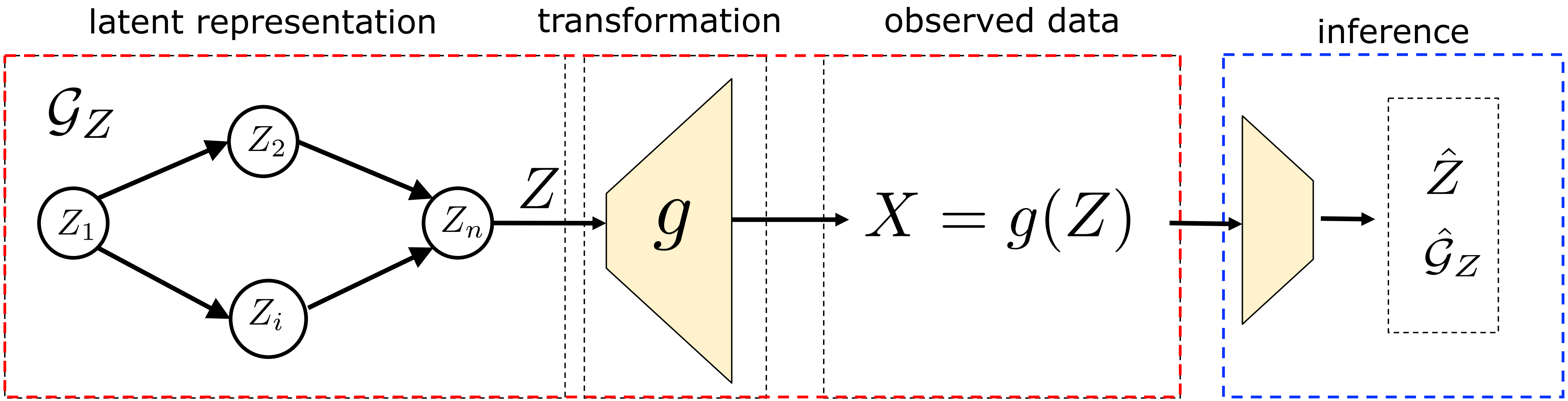
Linear Causal Representation Learning from Unknown Multi-node Interventions

Burak Varıcı

Carnegie Mellon University

Emre Acartürk (RPI) Karthikeyan Shanmugam (Google India) Ali Tajer (RPI)

Causal Representation Learning (CRL)

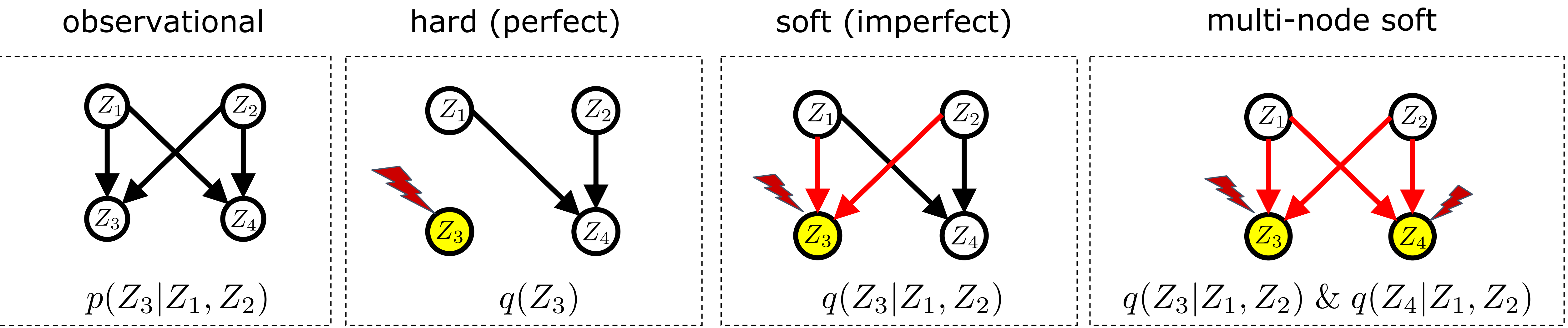


use data X to find:

- Z: latent random variables
- $\mathcal{G}_Z$: latent causal graph

- Identifiability: Uniquely recovering Z and $\mathcal{G}_Z$
- Design provably correct algorithms

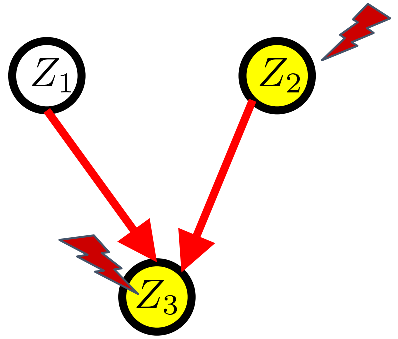
Interventional CRL



- **Single-node**: Perfect identifiability via **hard** [1-5], Identifiability up to ancestors via **soft** [1,3,4,6]
- Missing counterparts for stochastic multi-node interv. (prelim. work on different settings [7,8,9])

This paper: unknown **multi-node soft/hard interv.** on general SCMs + linear transform achieve same guarantees as single-node interventions!

Problem Setup



- Latent $Z = [Z_1, \dots, Z_n]$ and observed $X = [X_1, \dots, X_d]$
- Linear transformation $X = \mathbf{G} \cdot Z,$ where $\mathbf{G} \in \mathbb{R}^{d \times n}$
- Multi-node interventional environments:

$$\text{env. } \mathcal{E}^m \text{ with targets } I^m : \quad p^m(z) = \prod_{i \in I^m} q_i(z_i | z_{\mathrm{pa}(i)}) \prod_{i \notin I^m} p_i(z_i | z_{\mathrm{pa}(i)})$$

Goal: Same identifiability guarantees for unknown multi-node interventions

Score-based Methodology

Key ideas of score-based CRL (our earlier work [1,2])

- Score functions: $s(z) = \nabla_z \log p(z)$ and $s_X(x) = \nabla_x \log p_X(x)$
- Observed to latent scores: $s(z) = \mathbf{G}^\top \cdot s_X(x)$
- Sparse score differences: If $I_m = \{i\}$ yields $s(z) - s^m(z) = \nabla_z \log \frac{p_i}{q_i}(z_i | z_{\text{pa}(i)})$ (nonzero only at indices i and $\text{pa}(i)$)
- Aim to minimize score differences for $\hat{Z}$

Challenges for multi-node interventions

- The intervention targets are fully unknown!
- Latent score differences are no more sparse ($|I^m| + |\cup_{i \in I^m} \text{pa}(i)|$ non-zeros)

Multi-node Interventions

1. Find combinations of multi-node interventions to create sparser interventions

- **Idea:** if intervention targets are diverse, reduce to single-int. problem
- Example: Given $I^1 = \{1\}, I^2 = \{1,3\}, I^3 = \{2,3\}$ and $I^0 = \emptyset$
- $s^2(z) - s^1(z)$ gives $\tilde{I}^3 = \{3\}$, $\quad s^3(z) - \tilde{s}^3(z)$ gives $\tilde{I}^2 = \{2\}$

2. How to do it with unknown intervention targets?

- Consider $s^0_X, s^1_X, \ldots, s^n_X$. Iteratively search for mixing vectors $\mathbf{w} \in \mathbb{N}^{n+1}$ (in a finite search space),

$$\dim\Big(\text{image}\big(\sum_{i=0}^{n} w_i \cdot s^i_X\big)\Big) = 1$$

Results

Theorem (soft): Using diverse, regular unknown multi-node soft interventions, we have **identifiability up to ancestors**:

- $\hat{Z}_i$ is a linear function of $Z_i \cup Z_{\text{ancestors}(i)}$,
- $\hat{\mathcal{G}}_Z$ is transitive closure of $\mathcal{G}_Z$

Theorem (hard): Using diverse, regular unknown multi-node hard interventions and additive noise models, we have **perfect identifiability**:

- $\hat{Z}_i = c_i \times Z_i$ for a constant scalar c_i, and $\hat{\mathcal{G}}_Z = \mathcal{G}_Z$

- Sufficiently diverse interventions: Intervention matrix D is full rank where $D_{i,m} = \mathbb{I}(i \in I^m)$.

- Intervention regularity: effect of a multi-node intervention is not the same on the scores associated with different nodes.

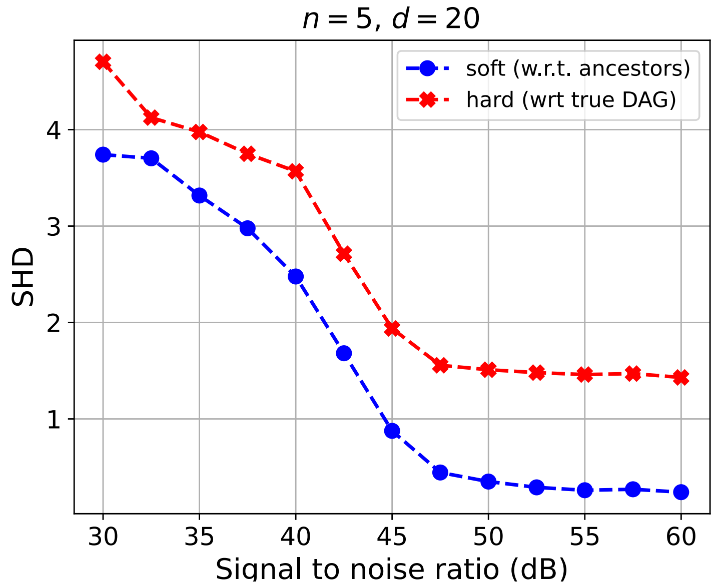
Experiments

- Linear Gaussian SEMs with Erdős–Rényi random graphs (100 runs)
- Scores: $s_X(x) = -\Theta \cdot x$, estimate precision matrix Θ with 10^5 samples
- **Structural Hamming distance (SHD)** for latent graph (ideally 0)
- **Mean correlation coefficient (MCC)** for latent variables (ideally 1)

Observed dimension: d=50

Latent dim.	Soft SHD	Soft MCC	Hard SHD	Hard MCC
4	0.77	0.96	0.66	0.98
5	1.93	0.93	1.80	0.98
6	3.39	0.92	3.05	0.95
7	4.62	0.91	6.12	0.91
8	8.26	0.90	9.01	0.88

Sensitivity analysis for quadratic causal models



Linear Causal Representation Learning from Unknown Multi-node Interventions

Paper: https://arxiv.org/abs/2406.05937

Code: https://github.com/acarturk-e/umni-crl

Conference: Poster session 1, December 11, 11am-2pm

Also see at Poster session 1: Sample Complexity of Interventional CRL

Contact: bvarici@andrew.cmu.edu

project page

[1] B. Varıcı, E. Acartürk, K. Shanmugam, A. Kumar, and A. Tajer. "*Score-based causal representation: Linear and general transformations*". arXiv: 2402.00849

[2] B. Varıcı, E. Acartürk, K. Shanmugam, and A. Tajer. "*General identifiability and achievability for causal representation learning*". AISTATS 2024.

[3] C. Squires, A. Seigal, S. S. Bhate, and C. Uhler. *"Linear causal disentanglement via interventions".* ICML 2023

[4] S. Buchholz, G. Rajendran, E. Rosenfeld, B. Aragam, B. Schölkopf, and P. Ravikumar. *"Learning linear causal representations from interventions under general nonlinear mixing"*. NeurIPS 2023

[5] J. von Kügelgen, M. Besserve, W. Liang, L. Gresele, A. Kekić, E. Bareinboim, D. M. Blei, and B. Schölkopf. *"Nonparametric identifiability of causal representations from unknown interventions"*. NeurIPS 2023

[6] J. Zhang, C. Squires, K. Greenewald, A. Srivastava, K. Shanmugam, and C. Uhler. "*Identifiability guarantees for causal disentanglement from soft interventions".* NeurIPS 2023

[7] S. Bing, U. Ninad, J. Wahl, and J. Runge. "*Identifying linearly-mixed causal representation learning from multi-node interventions*", CLeaR 2024.

[8] J. Jin and V. Syrgkanis. "*Learning Causal Representations from General Environments: Identifiability and Intrinsic Ambiguity*." NeurIPS 2024

[9] K. Ahuja, A. Mansouri, and Y. Wang. *"Multi-domain causal representation learning via weak distributional invariances".* AISTATS 2024