Learning Causal World Models from Acting and Seeing using score functions

B. Varici, E. Acartürk, K.Shanmugam, A. Kumar, A. Tajer [Score-based Causal Representation Learning: Linear and General Transformations]() (JMLR 2025).

Speaker: Karthikeyan Shanmugam, Google Deepmind

Burak Varici
(CMU)

Emre Acarturk
(RPI)

Abhishek Kumar
(ex-GDM,Amazon)

Ali Tajer
(RPI)

Karthikeyan
Shanmugam

Outline

- Causal representation learning - Learn representations that capture cause-effect relationships behind the perceptual data u see

Outline

- Causal representation learning - Learn representations that capture cause-effect relationships behind the perceptual data u see

- **Key Idea:** Score Functions used in Diffusion and connections to CRL

- Our results: Linear and Non-Linear Transforms

Motivating Position Paper

INVITED PAPER

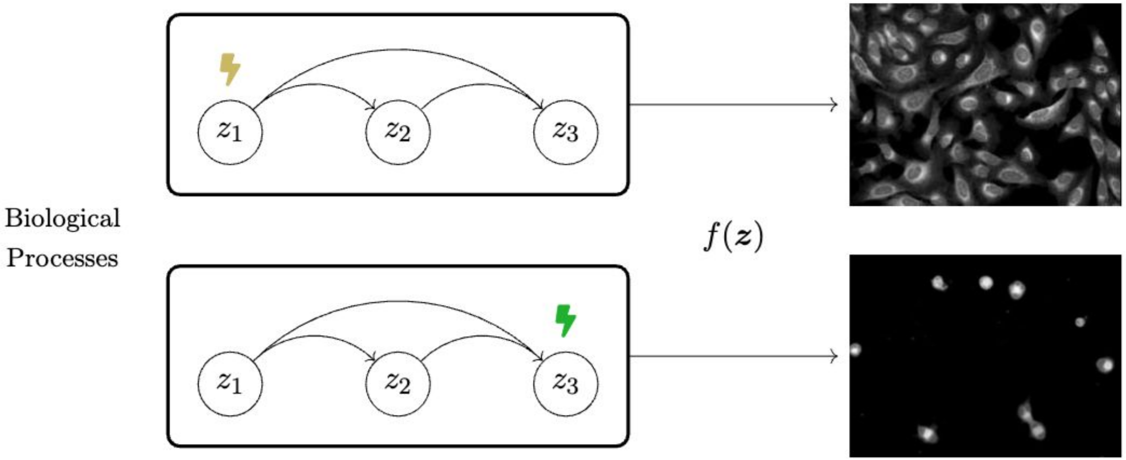
Toward Causal Representation Learning

This article reviews fundamental concepts of causal inference and relates them to crucial open problems of machine learning, including transfer learning and generalization, thereby assaying how causality can contribute to modern machine learning research.

By Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio

Causal Representation Learning: Motivations

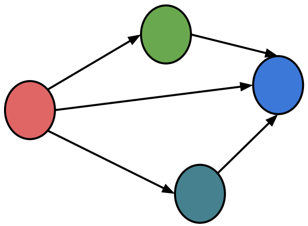
- Gene regulatory mechanisms -> Gene Expression Data captured as images



Moran, Aragam. *Towards Interpretable Deep Generative Models via Causal Representation Learning.* arxiv:2504.11609

Causal Representation Learning: Motivations

- Robotics: Joints are causally related -> image of robot from camera



Positions of
various Joints

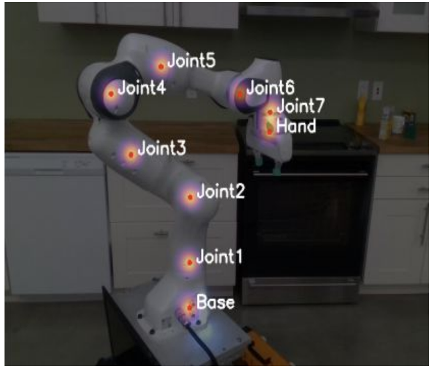
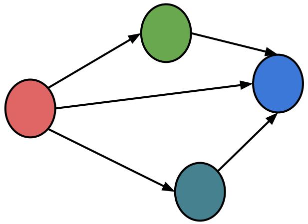


Figure: Camera-to-Robot Pose Estimation from a Single Image ICRA 2020

Challenge: Inferring Latent Causal variables from Data



Positions of various Joints

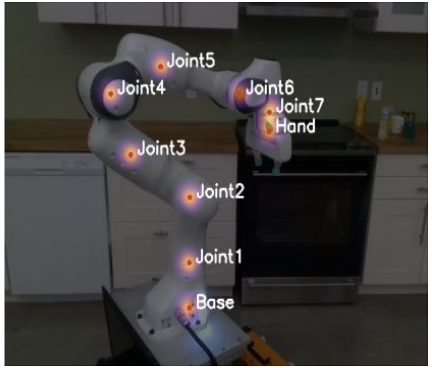
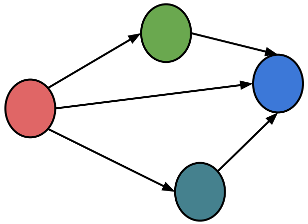


Figure: Camera-to-Robot Pose Estimation from a Single Image ICRA 2020

Goal: To invert this unknown transformation

Challenge: Inferring Latent Causal variables from Data



Positions of various Joints

Unknown Transformation

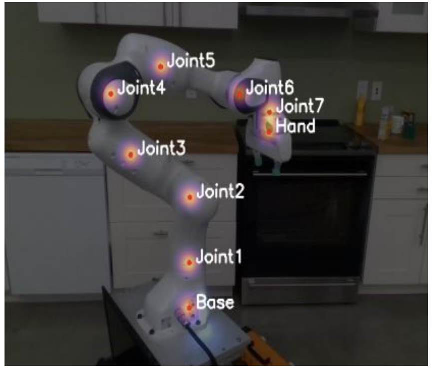


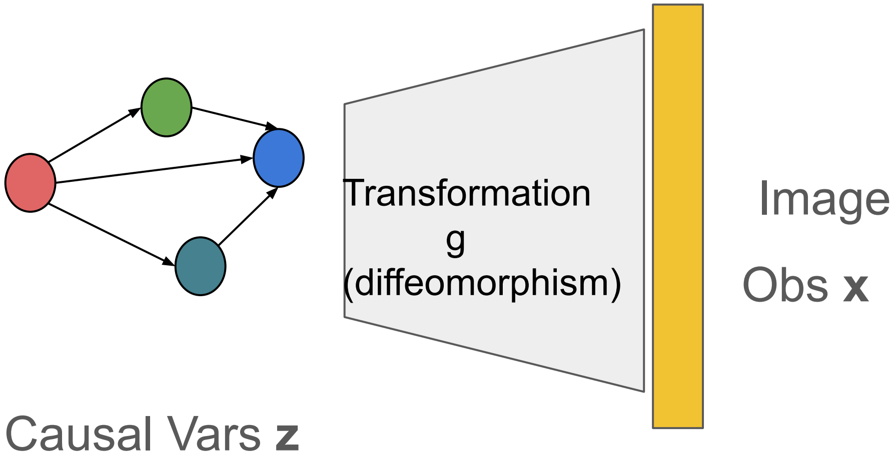
Figure: Camera-to-Robot Pose Estimation from a Single Image ICRA 2020

Goal: To invert this unknown transformation

Causal Variables exhibit sparse changes upon intervention + Conditional Independencies

- Intervention on the shoulder motor changes only the Shoulder -> Elbow relationship.
- Variables are not completely independent

Problem Setting



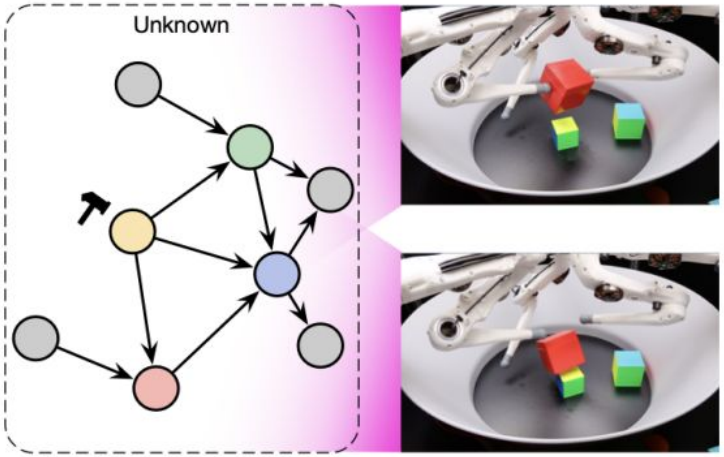
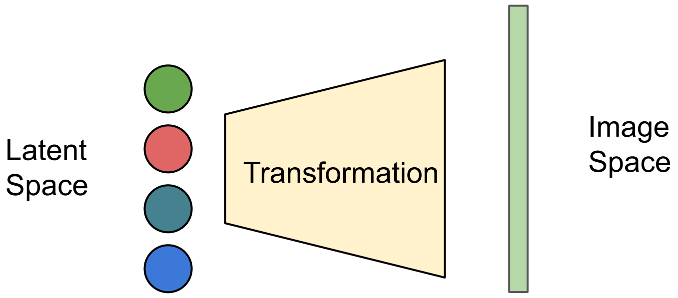


Figure: Towards Causal Representation Learning, Scholkopf et. al. 2021.

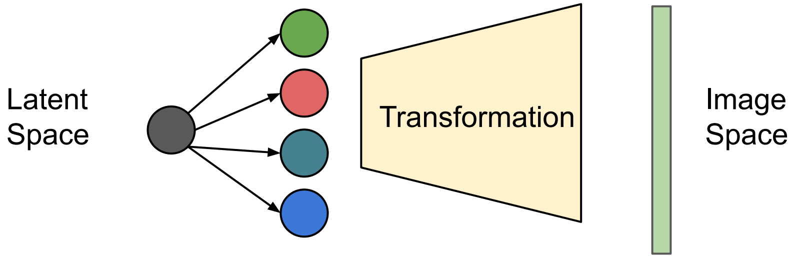
How can you invert g ?

Disentanglement : Story so far .....
[before 2023]

Disentanglement focuses on forcing independence in latent dimensions
[DIP-VAE[2017], beta-VAE [2016], InfoGAN[2016]...]



ICA - Conditioning on a common cause renders latents independent
[Hyvarinen et. al. 2019, Khemakhem et. al. 2019]
(+ other specific independence models)



Primarily independent or conditionally independent variables

Why is CRL hard with only one distribution ?

- Extreme case: **linear** transformation and **independent** latents (**linear ICA**)

$$X = \mathbf{G} \cdot Z \ , \qquad p(Z) = \prod_i p(Z_i)$$

- Is linear ICA solution set unique?

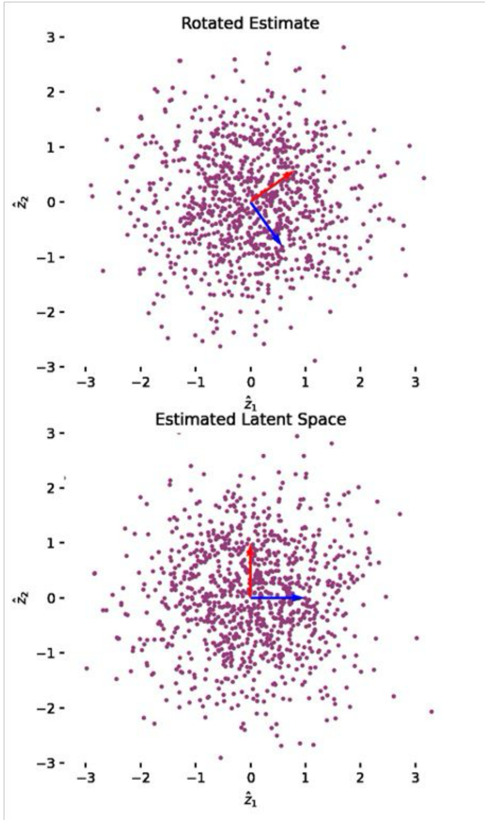
$$\{(\hat{Z}, \hat{\mathbf{G}}) \ : \ X = \hat{\mathbf{G}} \cdot \hat{Z} \ \text{ and } \ \hat{Z}_i \perp\!\!\!\perp \hat{Z}_j \ \ \forall i, j\}$$

- **no** – e.g., Gaussians are rotation invariant

$$X = \mathbf{G} R_\theta^\top R_\theta Z \ , \qquad p(Z) = p(R_\theta Z)$$

- What can be guaranteed? If at most one Z_i is Gaussian: ID up to permutation ($\mathbf{P}_\sigma$) and scaling ($\mathbf{D}$)

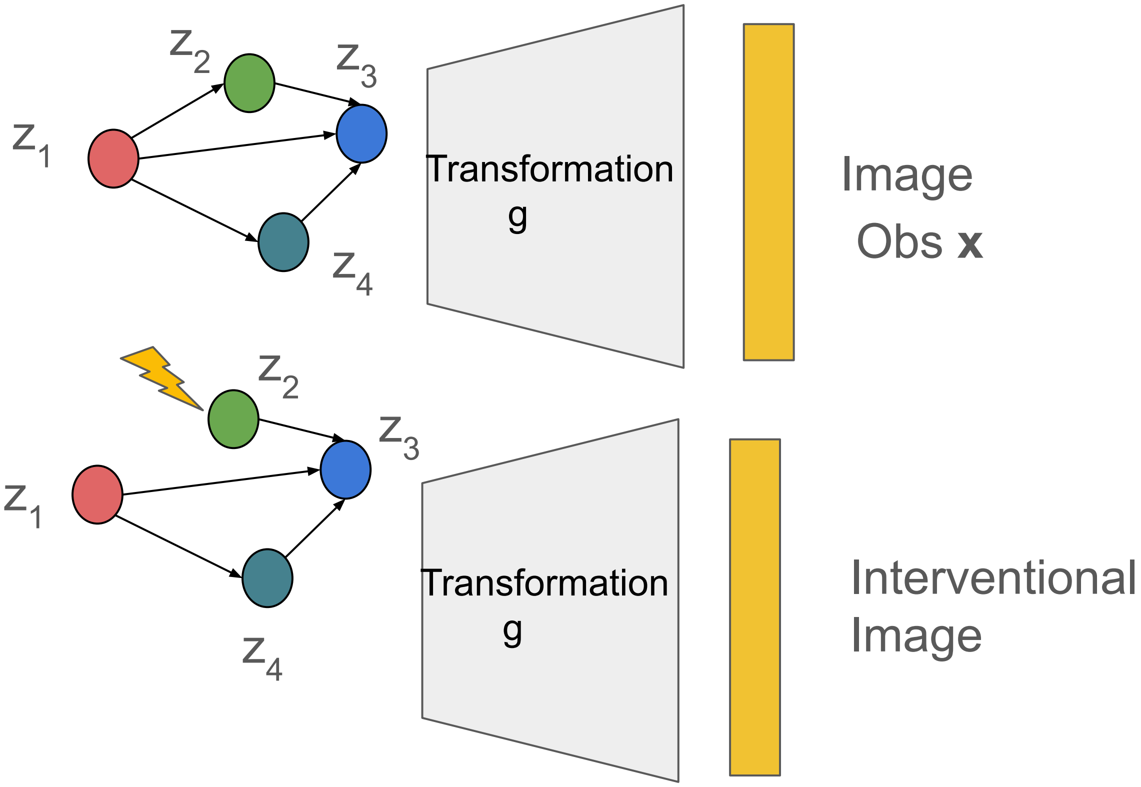
$$\hat{Z} = \mathbf{P}_\sigma \cdot \mathbf{D} \cdot Z$$



Interventional Data is needed

$$p(z_1)p(z_2|z_1)p(z_3|z_1)p(z_4|z_3,z_2,z_1)$$

$$p(z_1)\color{blue}{q}\color{black}(z_2|z_1)p(z_3|z_1)p(z_4|z_3,z_2,z_1)$$

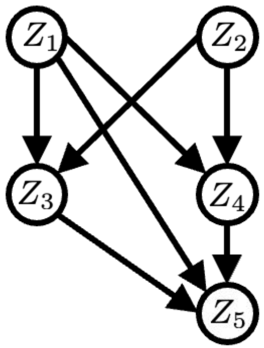


Interventions

CRL is impossible without sufficient statistical diversity

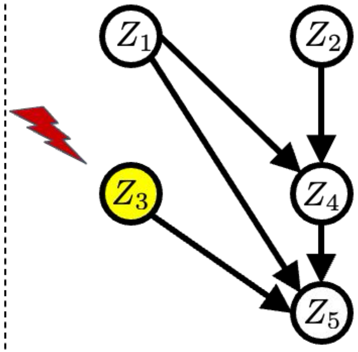
5

observational



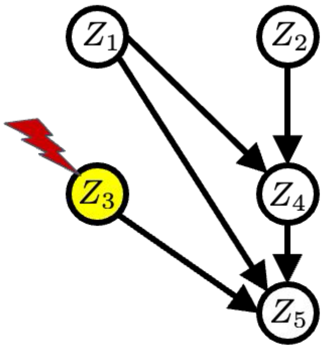
$$p(Z_3|Z_1, Z_2)$$

do



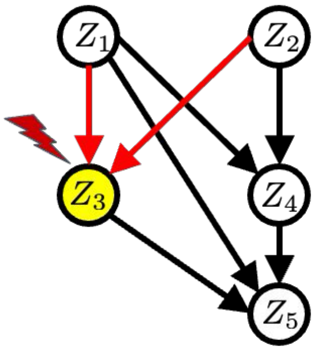
$$\begin{cases} 1 & \text{for} \quad Z_3 = Z \\ 0 & \text{for} \quad Z_3 \neq Z \end{cases}$$

hard (perfect)



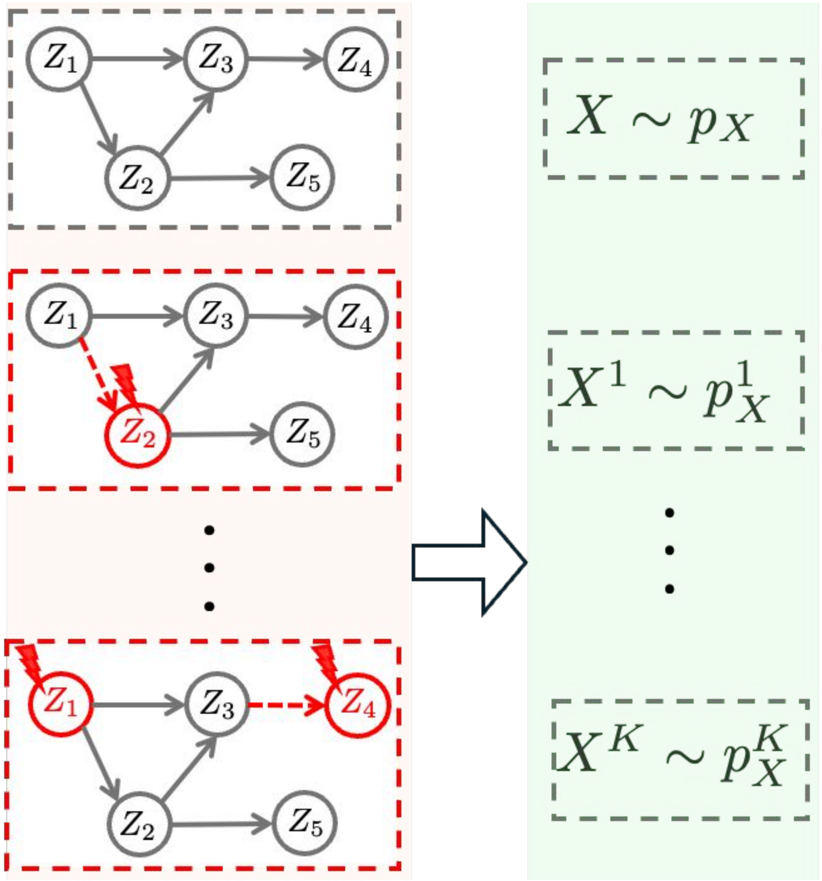
$$q(Z_3)$$

soft (imperfect)



$$q(Z_3|Z_1, Z_2)$$

Inference and Data Generation



Multiple intervention on latent space

Observations only from X space.

How do you learn the inverting transformation ?

What is missing?

	Transform	Latent Model	Intervention / node
Ahuja et al. (2023)	Polynomial	Nonparametric	1 do
Squires et al. (2023)	Linear	Lin. Gaussian	1 hard
Buchholz et al. (2023)	Nonparametric	Lin. Gaussian	1 hard
?	Linear	Nonparametric	1 hard / soft
von Kügelgen et al. (2023)	Nonparametric	Nonparametric + faithfulness	2 hard
?	Nonparametric		2 hard

Provably correct tractable algorithms or differentiable loss functions

Our Contributions

Transformation	Causal Model	Interventions	Identifiability of Transformation	Identifiability of Graph
1-1 Non Linear Transform	Arbitrary	2 Hard/node	Upto monotonic coord. transform	Perfect ID

Varici et. al. “General Identifiability and Achievability for Causal Representation Learning” AISTATS 2024

Differentiable regularizer on Autoencoders
whose global optima provably achieves the ID result

Our Contributions

Transformation	Causal Model	Interventions	Identifiability of Transformation	Identifiability of Graph
1-1 Non Linear Transform	Arbitrary	2 Hard/node	Upto monotonic coord. tx	Perfect ID
Linear Transform	Arbitrary	1 Hard/node	Upto coord scaling	Perfect ID
Linear Transform	Arbitrary	1 Soft/node	Mixing upto ancestors	Ancestral Graph

Sample Complexity Results

https://openreview.net/forum?id=XL9aaXl0u6 NeurIPS 2024

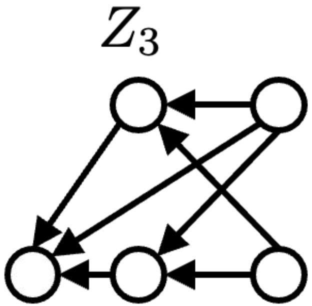
Score Functions

$$\nabla_z \log p(z)$$ Score Function of distribution p(z) (used in diffusion)

Song & Ermon 2019. Generative Modeling by Estimating Gradients of the Data Distribution

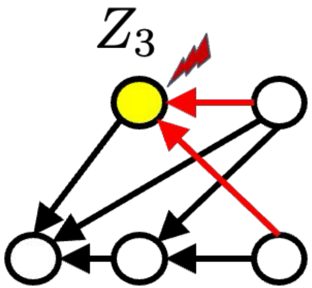
Score Differences in true latent space are sparse

$p_3(z_3 \mid z_{\mathrm{pa}(3)})$



$$p(Z) = p_3(z_3 \mid z_{\mathrm{pa}(3)}) \prod_{i \neq 3} p_i(z_i \mid z_{\mathrm{pa}}(i))$$

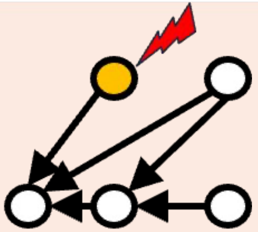
$q_3(z_3 \mid z_{\mathrm{pa}(3)})$



$$p^3(Z) = q_3(z_3 \mid z_{\mathrm{pa}(3)}) \prod_{i \neq 3} p_i(z_i \mid z_{\mathrm{pa}}(i))$$

$$\underbrace{\nabla_z \log p(Z)}_{s(Z)} - \underbrace{\nabla_z \log p^3(Z)}_{s^3(Z)} = \begin{array}{cl} \begin{bmatrix} \mathrm{x} \\ 0 \\ \mathrm{x} \\ 0 \\ 0 \\ \mathrm{x} \\ 0 \end{bmatrix} & \begin{array}{l} \text{coordinates of parents of node } i \\ \\ \\ \\ \\ \text{node } i \\ \\ \end{array} \end{array}$$

Hard interventions have a sparser score imprint

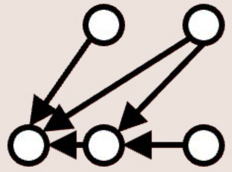


$$p^3(Z) = q_3(z_3) \prod_{i \neq 3} p_i(z_i \mid z_{\mathrm{pa}}(i))$$

Two hard interventions on the same node

$q_i(z_i)$ & $\tilde{q}_i(z_i)$

$\tilde{q}_3(z_3)$

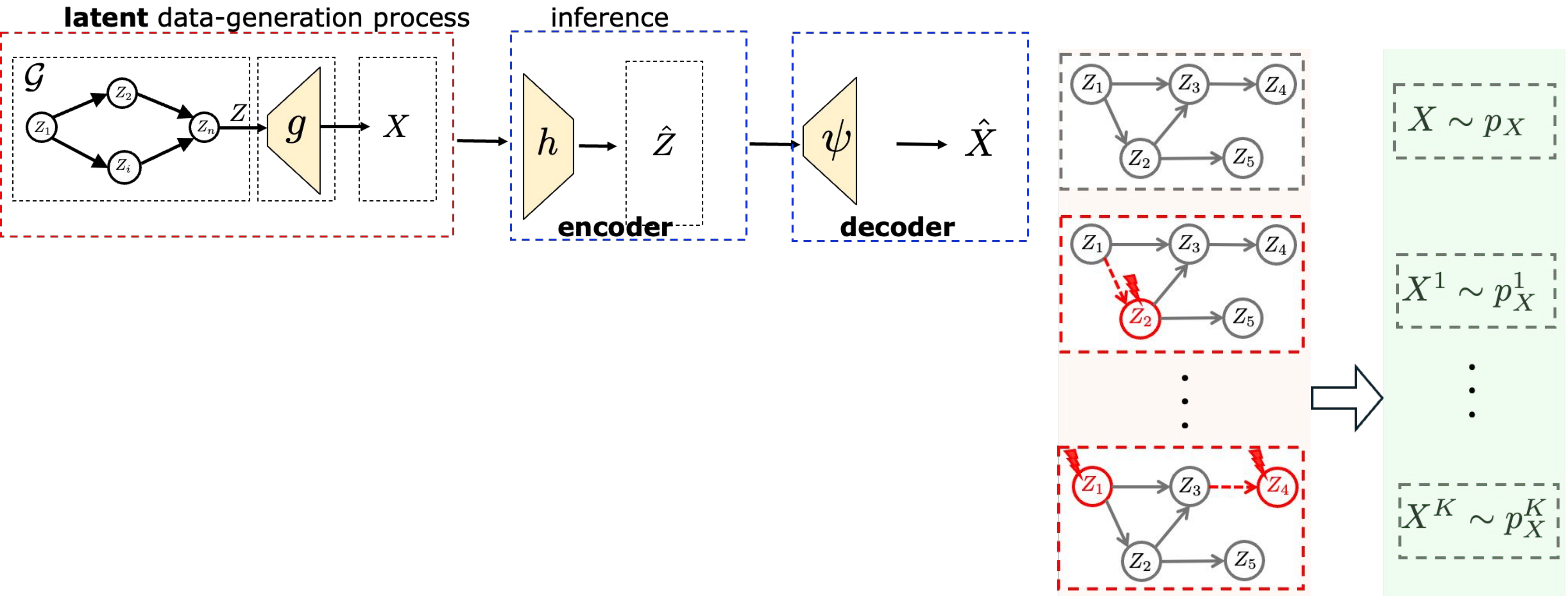


$$\tilde{p}^3(Z) = \tilde{q}_3(z_3) \prod_{i \neq 3} p_i(z_i \mid z_{\mathrm{pa}}(i))$$

$$\log p^i(z) - \log \tilde{p}^i(z) = \log q_i(z_i) - \log \tilde{q}_i(z_i)$$ function of only z_i

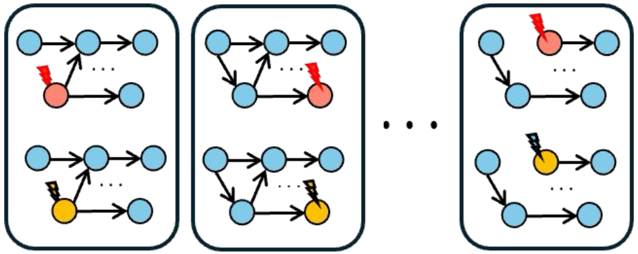
Score functions: $s^i(Z) - \tilde{s}^i(Z) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \times \\ 0 \end{bmatrix}$ ← intervened node

Inference and Data Generation



Our result for Non Linear Transforms

if given two hard interventions per node



$s_X^1 - \tilde{s}_X^1 \quad \cdots \quad s_X^n - \tilde{s}_X^n$

(observational – interventional score):
non-zero at coordinates i and parents of i

Main Result

Solve for the encoder

$$h^* = \arg\min_{h \in \mathcal{H}} \sum_{i=1}^{n} \left\| \mathbb{E}\left[\left| s^i(\hat{z}) - \tilde{s}^i(\hat{z}) \right| \right] - \mathbf{e}_i \right\|^2$$

complete ID guarantee $\hat{Z}_i = \phi_i(Z_i)$ for all $i \in [n]$

Exact graph recovery

$$\mathbb{E}\left[\left| s(\hat{z}) - s^i(\hat{z}) \right| \right]_k \neq 0 \iff k \in \mathrm{pa}(i) \cup i$$

Varici et. al. "General Identifiability and Achievability for Causal Representation Learning" AISTATS 2024

Partial identifiability if only some nodes are intervened

if given **only**

two environments

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix} \quad \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix}$$

Solve for the encoder

$$h_i^* = \arg\min_{h \in \mathcal{H}} \left\| \mathbb{E}\left[\left| s^i(\hat{z}) - \tilde{s}^i(\hat{z}) \right| \right] - \mathbf{e}_i \right\|^2$$

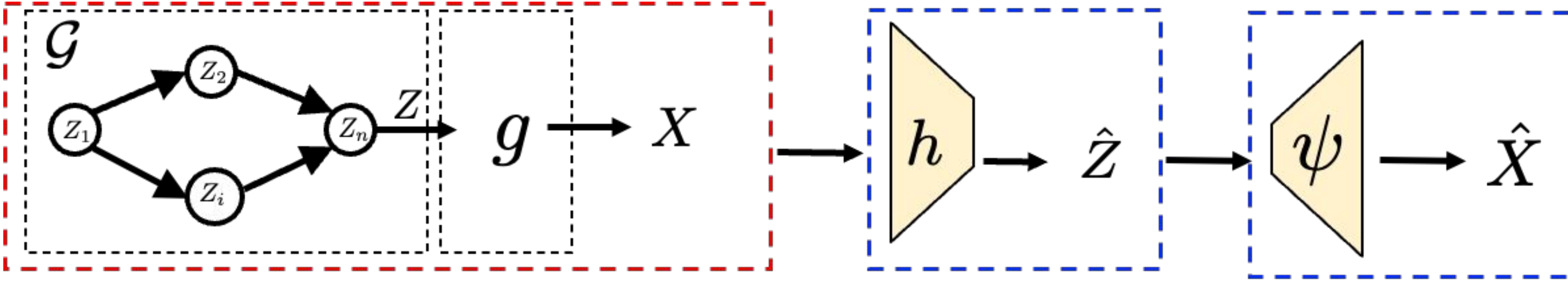
node-level ID guarantee

$$\hat{Z}_i = [h_i^*(X)]_i = \phi_i(Z_i)$$

ϕ_i: diffeomorphism (bijection, differentiable)

JMLR 2025
https://jmlr.org/papers/volume26/24-0194/24-0194.pdf

Differentiable Alg: Regularized Autoencoder Training



$$h^*, \psi^* = \arg\min_{h,\psi} \sum_{i=1}^{n} \left\| \mathbb{E}\left[\left| s^i(\hat{z}) - \tilde{s}^i(\hat{z}) \right| \right] - \mathbf{e}_i \right\|^2 + \|(\psi \circ h)(X) - X\|^2$$

AE reconstruction loss

Score difference in some arbitrary latent space

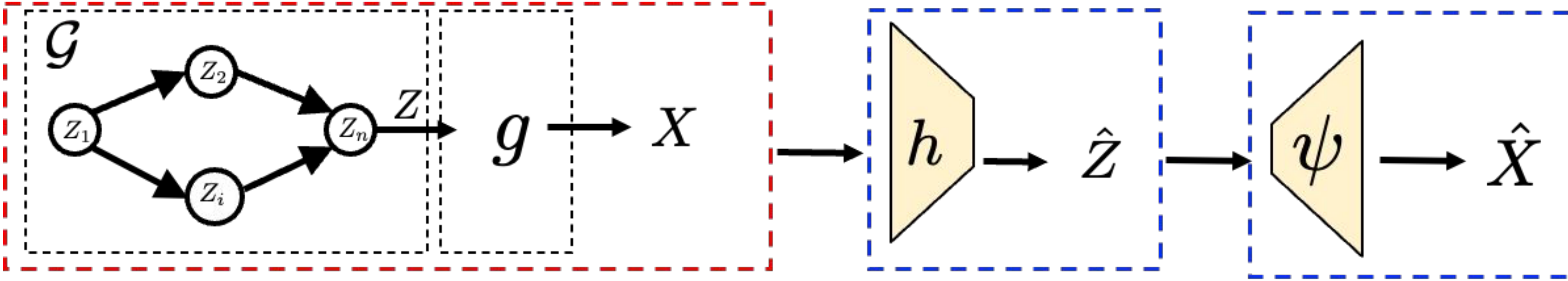
$$X = g(Z) \qquad\qquad p_X(x) \;=\; p_Z(z) \;\times\; \left|\det\left(J_g(z)^\top J_g(z)\right)\right|^{-\frac{1}{2}}$$

$$s_Z(z) - s_Z^m(z) = \left[J_g(z)\right]^\top \cdot \left(s_X(x) - s_X^m(x)\right)$$

$$Z \xrightarrow[\text{true dec.}]{g} X \xrightarrow[\text{cand. enc.}]{h} \hat{Z} \xrightarrow[\text{cand. dec.}]{h^{-1}} X$$

$$s_{\hat{Z}}(\hat{z}) - s_{\hat{Z}}^m(\hat{z}) = \left[J_{h^{-1}}(x)\right]^\top \cdot \left(s(x) - s^m(x)\right)$$

Differentiable Alg: Regularized Autoencoder Training



$$h^*, \psi^* = \arg\min_{h,\psi} \sum_{i=1}^{n} \left\| \mathbb{E}\left[\left| s^i(\hat{z}) - \tilde{s}^i(\hat{z}) \right| \right] - \mathbf{e}_i \right\|^2 + \|(\psi \circ h)(X) - X\|^2$$

AE reconstruction loss

$$s^i(\hat{z}) - \tilde{s}^i(\hat{z}) = [J_{h^{-1}}(\hat{z})]^T [s^i(x) - \tilde{s}^i(x)]$$

Proof Sketch:

$$g \circ h = \phi$$

$$s^i(\hat{z}) - \tilde{s}^i(\hat{z}) = J_\phi^{-\top}(z) \cdot \left(s^i(z) - \tilde{s}^i(z)\right)$$ **where** $\hat{Z} = \phi(Z$

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ \times \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial \phi_1}{\partial Z_1} & \frac{\partial \phi_1}{\partial Z_2} & \cdots & \frac{\partial \phi_1}{\partial Z_n} \\ \frac{\partial \phi_2}{\partial Z_1} & \frac{\partial \phi_2}{\partial Z_2} & \cdots & \frac{\partial \phi_2}{\partial Z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_m}{\partial Z_1} & \frac{\partial \phi_m}{\partial Z_2} & \cdots & \frac{\partial \phi_m}{\partial Z_n} \end{bmatrix} \cdot \begin{bmatrix} 0 \\ \times \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

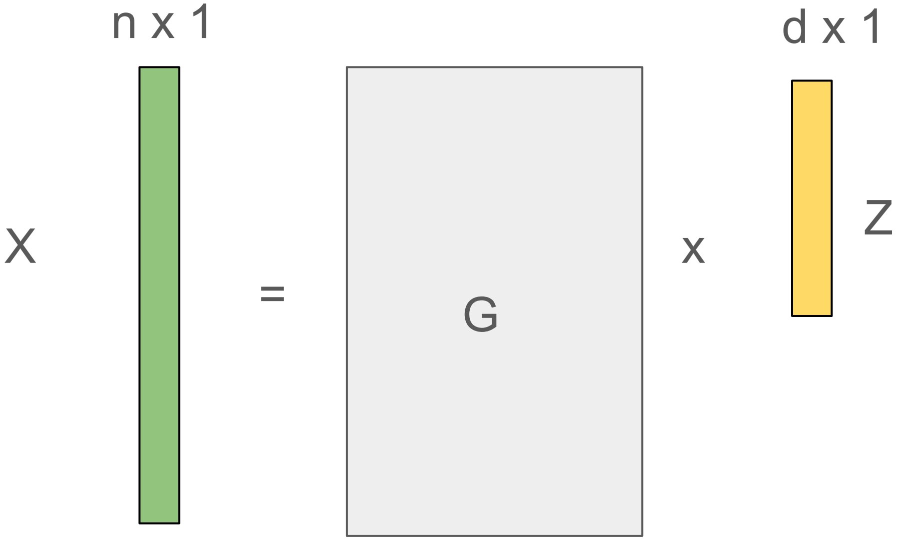
$$\frac{d\Phi_2}{dz_i} = 0, \ i \neq 4$$

Our Contributions

Transformation	Causal Model	Interventions	Identifiability of Transformation	Identifiability of Graph
1-1 Non Linear Transform	Arbitrary	2 Hard/node	Upto monotonic coord. tx	Perfect ID
Linear Transform	Arbitrary	1 Hard/node	Upto coord scaling	Perfect ID
Linear Transform	Arbitrary	1 Soft/node	Mixing upto ancestors	Ancestral Graph

JMLR 2025
https://jmlr.org/papers/volume26/24-0194/24-0194.pdf

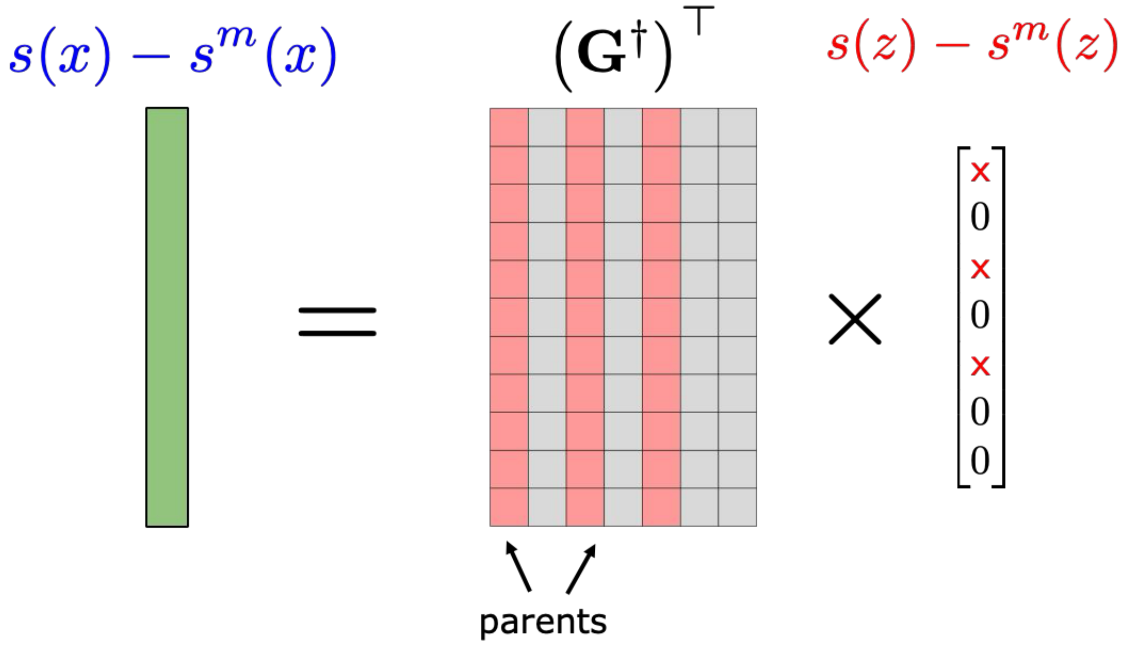
Our result: Linear Transforms



$$(G^\dagger)^T(s_Z(z) - s_Z^i(z)) = s_X(x) - s_X^i(x)$$

Score differences are linearly related

Our result: Linear Transforms



Infer about the inverse transform using observed score difference

$$\left(s(x) - s^m(x)\right) \in \ \mathrm{span}\{\mathbf{G}_j^\dagger : j \in \mathrm{pa}(i) \cup i\}$$

Our result: Linear Transforms and soft interventions

Covariance matrices of score difference

$$R_X^i = E[(s_X(x) - s_X^i(x))(s_X(x) - s_X^i(x))^T],\ R_Z^i = E[(s_Z(z) - s_Z^i(z))(s_Z(z) - s_Z^i(z))^T]$$

Our result: Linear Transforms and soft interventions

Covariance matrices of score difference

$$R_X^i = E[(s_X(x) - s_X^i(x))(s_X(x) - s_X^i(x))^T],\ R_Z^i = E[(s_Z(z) - s_Z^i(z))(s_Z(z) - s_Z^i(z))^T]$$

Guess for the i-th row of the decoder(X): y ~ Unif over Sphere $h_i = R_X^i * y$

$s_Z(z) - s_Z^i(Z)$ is non-zero in i, pa(i) coordinates

Our result: Linear Transforms and soft interventions

Covariance matrices of score difference

$$R^i_X = E[(s_X(x) - s^i_X(x))(s_X(x) - s^i_X(x))^T], \; R^i_Z = E[(s_Z(z) - s^i_Z(z))(s_Z(z) - s^i_Z(z))^T]$$

Guess for the i-th row of the decoder(X): y ~ Unif over Sphere $h_i = R^i_X * y$

$s_Z(z) - s^i_Z(Z)$ is non-zero in i, pa(i) coordinates

Partial Disentanglement

$$\hat{Z}_i = h_i^T X = h_i^T G Z = \sum_{j \in pa(i)} c_j Z_j + c_i Z_i$$

Our result: Linear Transforms and Soft Interventions

$$\hat{\mathrm{pa}}(m) \triangleq \left\{ i \neq m : \mathbb{E}\left[\left| \boldsymbol{s}_{\hat{\mathbf{Z}}}(\hat{\mathbf{Z}}; \hat{\mathbf{H}}) - \boldsymbol{s}^m_{\hat{\mathbf{Z}}}(\hat{\mathbf{Z}}; \hat{\mathbf{H}}) \right|_i \right] \neq 0 \right\} .$$

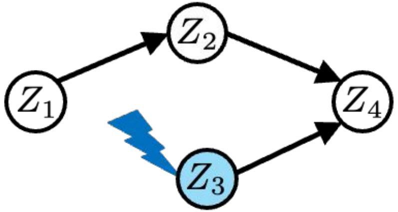
Estimate a graph using non zero score differences of the new estimate

Ancestral graph of this graph = Ancestral graph of the true graph

Estimate the Ancestral Graph and
A representation that mixes **every variable only with its parents**

Our result: Linear Transforms and Hard Interventions

- **One hard int/node:** target node becomes independent of its non-descendants.
- Additional step: use this property to resolve mixing with parents
- **Linear MMSE** estimator to update the encoder (in topological order)



$$Z_3 \perp\!\!\!\perp Z_1, Z_2$$

$$\mathbf{u} \leftarrow \mathrm{Cov}(\hat{Z}_i, \hat{Z}_{\hat{\mathrm{pa}}(i)}) \cdot [\mathrm{Cov}(\hat{Z}_{\hat{\mathrm{pa}}(i)})]^{-1}$$

$$\mathbf{H}_i \leftarrow \mathbf{H}_i - \mathbf{u} \cdot \mathbf{H}_{\hat{\mathrm{pa}}(i)}$$

identifiability up to scaling

$$\boxed{\mathbf{H}_i = c_i \cdot \mathbf{G}_i^{\dagger} \;\; \rightarrow \;\; \hat{Z}_i = c_i \cdot Z_i}$$

Further Results: Linear Transforms

Linear Transforms + Non-linear causal models “of sufficient complexity”

- Under soft interventions, can recover the DAG structure and obtain evern sparser disentanglement - mixing upto a specific subset of parents.

  Score-based Causal Representation Learning: Linear and General Transformations (JMLR 2025)

Further Results: Linear Transforms

Linear Transforms + Non-linear causal models “of sufficient complexity”

- Under soft interventions, can recover the DAG structure and obtain evern sparser disentanglement - mixing upto a specific subset of parents.

  Score-based Causal Representation Learning: Linear and General Transformations (JMLR 2025)

Sample Complexity of Linear Transform Case

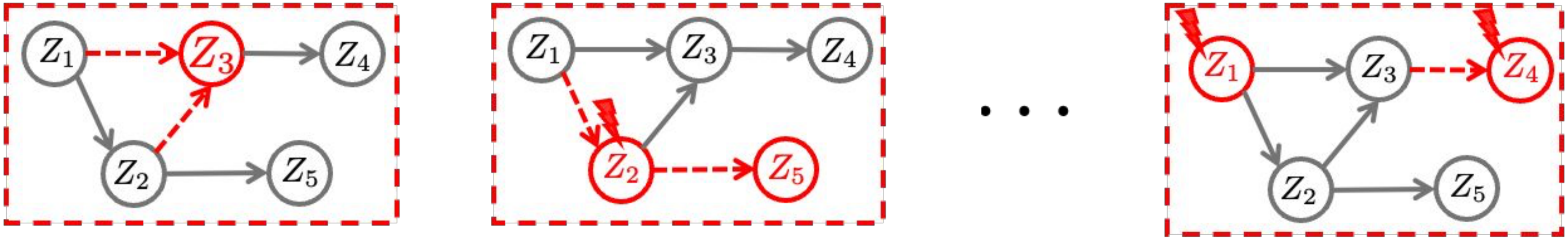
- “Sample Complexity of Interventional Causal Representation Learning”,
  Emre Acartürk, Burak Varıcı, Karthikeyan Shanmugam, Ali Tajer, NeurIPS 2024.

Linear Transforms: When Interventions are unknown and on multiple nodes at once

- “Linear Causal Representation Learning from Unknown Multi-node Interventions”,

  Burak Varıcı, Emre Acartürk, Karthikeyan Shanmugam, Ali Tajer, NeurIPS 2024.

Linear Transforms: Unknown multi node interventions



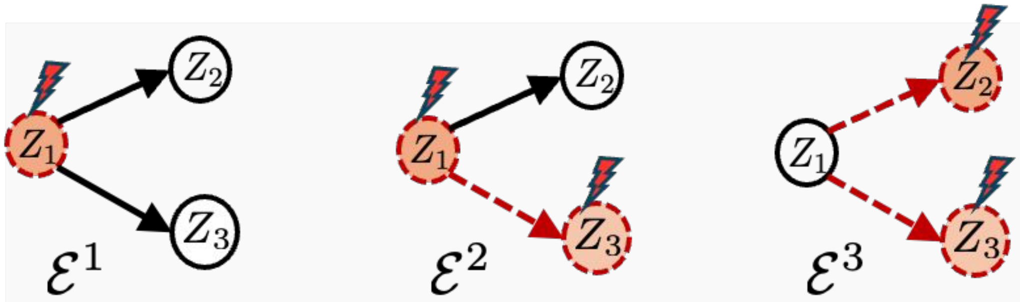
- Multi-node interventional environments:

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

- **Unknown intervention targets**: e.g., passive observations, off-target effects
- **Challenge:** score differences are not sparse anymore
- **Question**: under what conditions the single-node intervention guarantees hold?
- **Idea:** use multi-node score differences to find node-level score differences

Linear Transforms: Score combinations in latent space

Find combinations of multi-node interventions to create sparser interventions

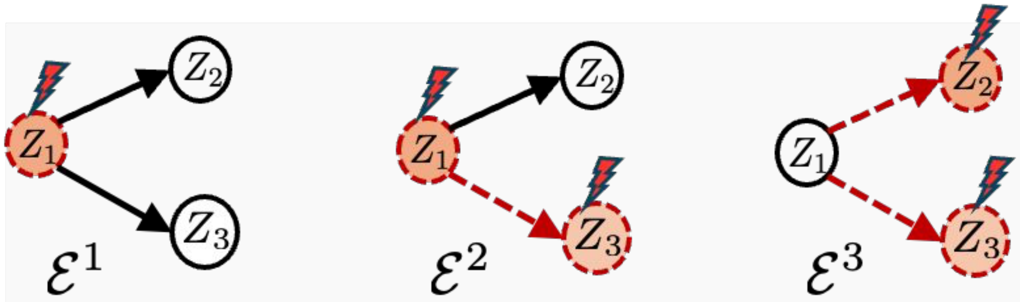


$$\mathbf{D} = \begin{array}{c} \begin{array}{ccc} \mathcal{E}^1 & \mathcal{E}^2 & \mathcal{E}^3 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{array} \begin{array}{l} -Z_1 \\ -Z_2 \\ -Z_3 \end{array}$$

$s^2(z) - s^1(z) + s(z)$ is the score function of distribution when only 3rd node was intervened

Linear Transforms: Score combinations in ambient space space

Find combinations of multi-node interventions to create sparser interventions



$$\begin{array}{c} \\ \mathbf{D} = \end{array} \begin{array}{c} \begin{array}{ccc} \mathcal{E}^1 & \mathcal{E}^2 & \mathcal{E}^3 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{array} \begin{array}{l} \\ -Z_1 \\ -Z_2 \\ -Z_3 \end{array}$$

$s^2(x) - s^1(x) + s(x)$ is the score of interventional distribution with only 3rd node intervened in the ambient space !!

Linear Transforms: Unknown multi node interventions

Need an intervention set where all atomic interventions can be recovered

Iteratively search for mixing vectors $\mathbf{w} \in \mathbb{N}^{n+1}$ (in a finite search space),

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

unknown multi-node **soft interventions**	$\hat{Z}_i = c_i \cdot Z_i + \sum_{k \in \mathrm{an}(i)} c_k \cdot Z_k$	$\hat{\mathcal{G}}_{\text{trans.clos.}} = \mathcal{G}_{\text{trans.clos.}}$
unknown multi-node **hard interventions**	$\hat{Z}_i = c_i \cdot Z_i$	perfect identifiability $\hat{\mathcal{G}} = \mathcal{G}$

Synthetic Data Results: General Transforms

Table 9: GSCALE-I for a quadratic causal model with **two coupled hard** interventions per node. Noisy scores are obtained using SSM-VR with $n_{\text{score}} = 30000$ samples.

			expected num.	**perfect scores**		**noisy scores**	
n	d	n_s	edges in $\mathcal{G}$	MCC	$\text{SHD}(\mathcal{G}, \hat{\mathcal{G}})$	MCC	$\text{SHD}(\mathcal{G}, \hat{\mathcal{G}})$
5	100	200	5	1.00 ± 0.00	0.00 ± 0.00	0.85 ± 0.02	4.50 ± 0.38
8	100	500	14	0.95 ± 0.01	1.50 ± 0.27	0.75 ± 0.02	12.9 ± 0.44

Transform = Tanh activated 1-hidden layer NN
MCC - maximum correlation coefficient
SSM - Sliced Score Matching is used to estimate scores

$$\text{MCC}(Z, \hat{Z}) \triangleq \max_{\pi} \frac{1}{n} \sum_{i \in [n]} \text{corr}(Z_i, \hat{Z}_{\pi(i)}) \, .$$

Synthetic Data Results: Linear Transforms

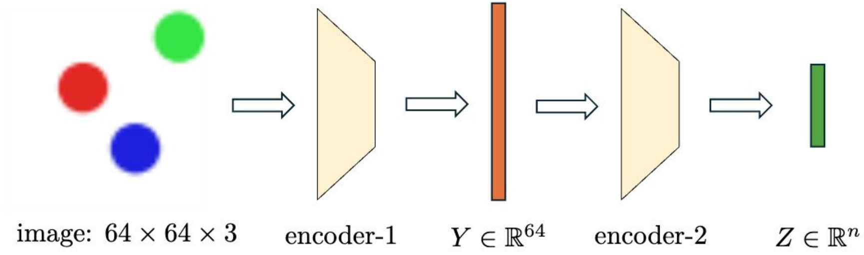
Table 6: LSCALE-I for an MLP causal model with **one hard** intervention per node ($n_s = 50000$).

		perfect scores			noisy scores		
n	d	MCC	ℓ_{scale}	$\text{SHD}(\mathcal{G}, \hat{\mathcal{G}})$	MCC	ℓ_{scale}	$\text{SHD}(\mathcal{G}, \hat{\mathcal{G}})$
5	100	1.00 ± 0.00	0.03 ± 0.00	0.01 ± 0.01	0.94 ± 0.01	0.62 ± 0.02	4.27 ± 0.20

Table 7: LSCALE-I for a linear causal model with **one soft** intervention per node.

			perfect scores			noisy scores		
n	d	n_s	MCC	ℓ_{pa}	$\text{SHD}(\mathcal{G}_{\text{tc}}, \hat{\mathcal{G}}_{\text{tc}})$	MCC	ℓ_{pa}	$\text{SHD}(\mathcal{G}_{\text{tc}}, \hat{\mathcal{G}}_{\text{tc}})$
5	100	5000	0.98 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.98 ± 0.00	0.04 ± 0.00	0.59 ± 0.11
5	100	10000	0.98 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.98 ± 0.00	0.03 ± 0.00	0.36 ± 0.08
5	100	50000	0.98 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.98 ± 0.00	0.01 ± 0.00	0.28 ± 0.06
8	100	5000	0.98 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.98 ± 0.00	0.07 ± 0.00	3.84 ± 0.36
8	100	10000	0.98 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.98 ± 0.00	0.05 ± 0.00	1.23 ± 0.20
8	100	50000	0.98 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.98 ± 0.00	0.02 ± 0.00	0.49 ± 0.10

Simplistic Image Datasets: Image rendering = Transformation



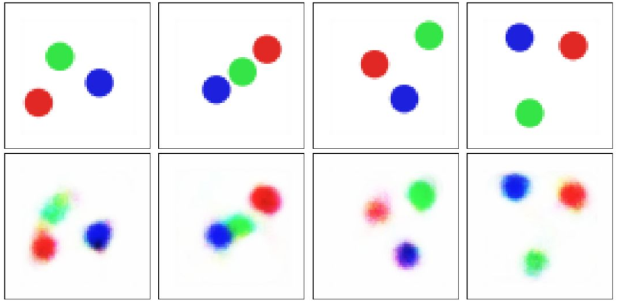


Figure 6: Sample images (top row) versus their reconstructions (bottom row).

Table 14: MCC comparison in image experiments (over 5 runs).

Algorithm	SCM	# balls	# int. / node	int. type	mean (std. error)
GSCALE-I	linear	2	2	hard	0.80 ± 0.03
GSCALE-I	linear	3	2	hard	0.76 ± 0.08
GSCALE-I	nonlinear	2	2	hard	0.93 ± 0.02
GSCALE-I	linear	2	1	hard	0.79 ± 0.03
GSCALE-I	nonlinear	2	1	hard	0.92 ± 0.02
Ahuja et al. (2023)	linear	2	1	do	0.13 ± 0.03
Ahuja et al. (2023)	linear	2	3	do	0.73 ± 0.03
Ahuja et al. (2023)	linear	2	5	do	0.83 ± 0.03
Buchholz et al. (2023)	linear	2	1	hard	0.87 ± 0.03
Buchholz et al. (2023)	linear	5	1	hard	0.94 ± 0.01

Conclusions and Future Work

- Presented a differentiable algorithm with guarantees for CRL with general transforms

- Currently ccaling score based regularizers to large scale setups - robot simulators

- Future Work:
  - Extend our framework by looking at action data from a single long trajectory

  - Can **score difference** estimation in ambient space be done efficiently ?

Thank You

References:

B. Varici, E. *Acartürk, K.Shanmugam, A. Kumar, A. Tajer* Score-based Causal Representation Learning: Linear and General Transformations *JMLR 2025*.

Varici et. al. General Identifiability and Achievability for Causal Representation Learning *AISTATS 2024* (Oral)

E. Acartürk, B. Varıcı, K. Shanmugam, A. Tajer. Sample Complexity of Interventional Causal Representation Learning. *NeurIPS 2024*.

B. Varici,, E. Acartürk, K. Shanmugam, A. Tajer. Linear Causal Representation Learning from Unknown Multi-node Interventions. *NeurIPS 2024*.