Score-based CRL from Interventions

Burak Varıcı

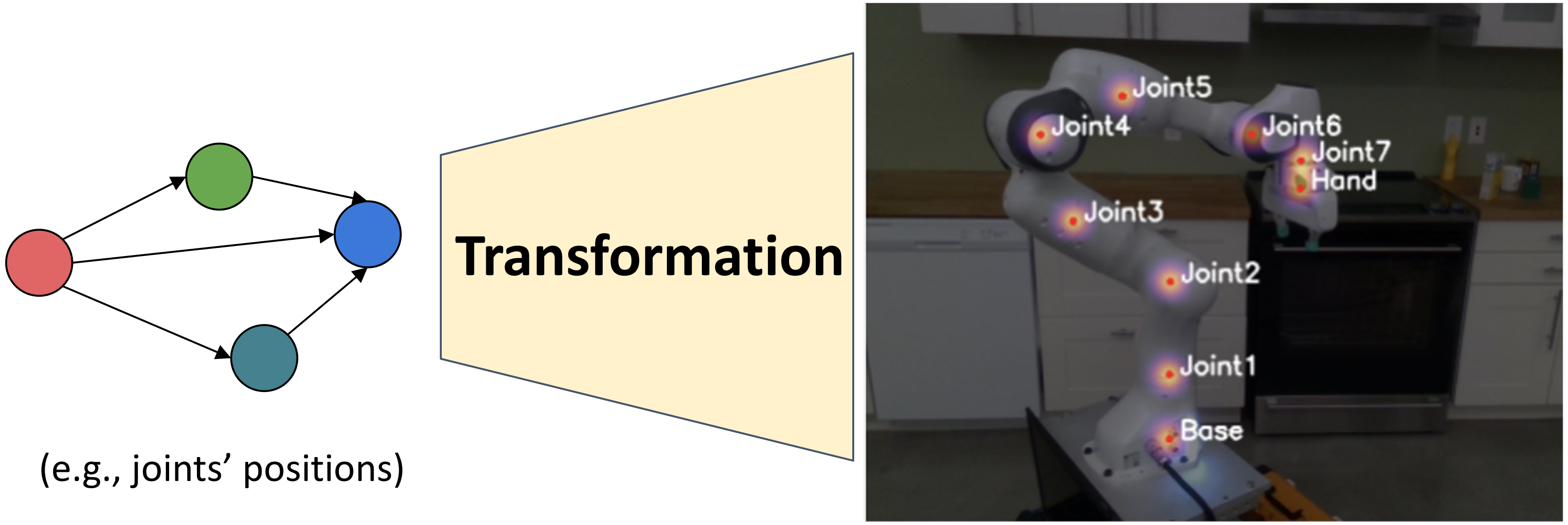
Rensselaer Polytechnic Institute

Emre Acartürk (RPI)

Karthikeyan Shanmugam (Google)

Ali Tajer (RPI)

Causal Representation Learning (CRL)



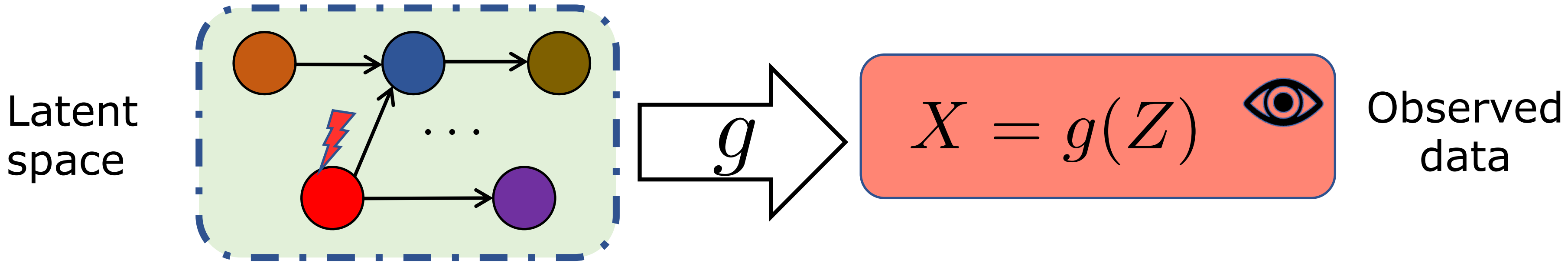
(e.g., joints' positions)

“.. **learn a representation (partially) exposing the unknown causal structure**, e.g., which variables describe the system, and their relations .. “ Schölkopf et al., 2021

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

CRL Objectives

CRL is impossible from only observational data



1. **Identifiability**: Conditions for uniquely recovering Z and $\mathcal{G}_Z$

2. **Achievability**: Provably correct algorithms to recover Z and $\mathcal{G}_Z$

Identifiability Results

Parametric Latent Models

	Latent Model	Transform	Int. / node	ID Results
Squires et al. (2023)	Linear + Gaussian	Linear	1 hard (or soft)	✓ (or ancestors)
Buchholz et al. (2023)	Linear + Gaussian	Nonparametric	1 hard	✓

Nonparametric Latent Models

	Latent Model	Transform	Int. / node	ID Results
	Nonparametric	Linear		
Ahuja et al. (2023)	Nonparametric	Polynomial	1 do (or 1 soft with ind. support)	✓
Zhang et al. (2023)	Nonparametric + nonlinear	Polynomial	1 soft	ancestors
von Kügelgen et al. (2023)	Nonparametric + faithfulness	Nonparametric	2 hard	✓
	Nonparametric			

Nonparametric Latent – What is Missing?

Provably correct algorithms for nonparametric transform

Identifiability guarantees with 1 intervention/node

Nonparametric Latent Models

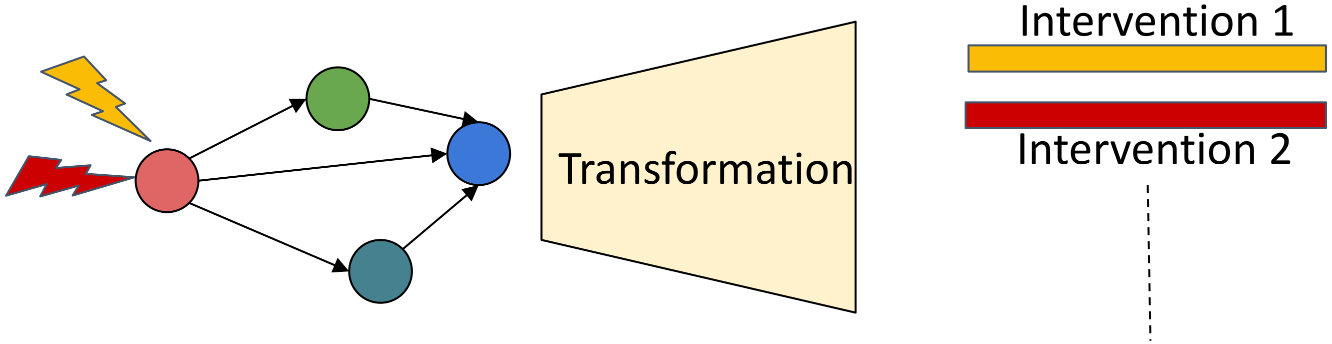
	Latent Model	Transform	Int. / node	ID Results
	Nonparametric	Linear		
Ahuja et al. (2023)	Nonparametric	Polynomial	1 do (or 1 soft with ind. support)	✓
Zhang et al. (2023)	Nonparametric + nonlinear	Polynomial	1 soft	ancestors
von Kügelgen et al. (2023)	Nonparametric + faithfulness	Nonparametric	2 hard	✓
	Nonparametric			

Our Contributions

Latent model		Transform		Interventions		Main results	
Nonparametric	+	Nonparametric	+	Two hard	=	1. perfect ID 2. provably correct algo	CRL@ NeurIPS
Sufficiently nonlinear	+	Linear	+	One hard (soft)	=	1. perfect ID (1. perfect DAG + Markov) 2. provably correct algo	Varıcı et al. 2023
Nonparametric	+	Linear	+	One hard (soft)	=	1. perfect ID (1. ID up to ancestors) 2. provably correct algo	Coming soon

Algorithm Overview

Sufficient Interventional Diversity: 2 different hard interventions per node in the latent space

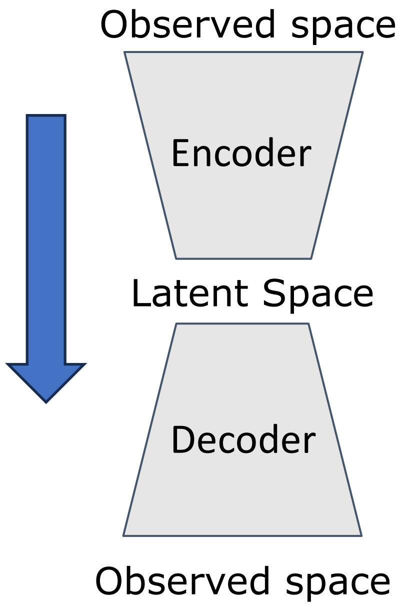


$$\boxed{\nabla \log p^{\mathsf{Int}_1}(x) = s^1(x)}$$

$$\nabla \log p^{\mathsf{Int}_2}(x) = s^2(x)$$

$$\vdots$$

$$\nabla \log p^{\mathsf{Int}_{2n}}(x) = s^{2n}(x)$$



Minimize

$$\left\| \mathbb{E}\left[\left| \mathsf{Jac}_{\mathsf{dec}}(x)(s^1(x) - s^2(x)) \right| \right] \quad \cdots \quad \mathbb{E}\left[\left| \mathsf{Jac}_{\mathsf{dec}}(x)(s^{2n-1}(x) - s^{2n}(x)) \right| \right] \right\|_0 + \text{Reconstruction Loss}$$

Provably correct algorithm for unsupervised learning
(small variations for each setting)

Empirical Results

Non-linear latent model: $Z_i = \sqrt{Z_{\mathrm{pa}(i)}^{\top} A_{p,i} Z_{\mathrm{pa}(i)}} + N_{p,i}$

n=8 latent variables

Input score differences $(s_X - s_X^m)$: Perfect score oracle or learn from data (Sliced Score Matching)

Non-linear transform: $X = \tanh(T\,.\,Z)$

Two hard / node

Obs. dim	Norm. Z error	DAG error (SHD)	Norm. Z error	DAG error (SHD)
8	0.16	1.56	0.70	11.9
25	0.20	1.55	0.68	10.5
40	0.21	1.14	0.71	11.8
	score oracle		noisy scores	

Linear transform: $X = T\,.\,Z$

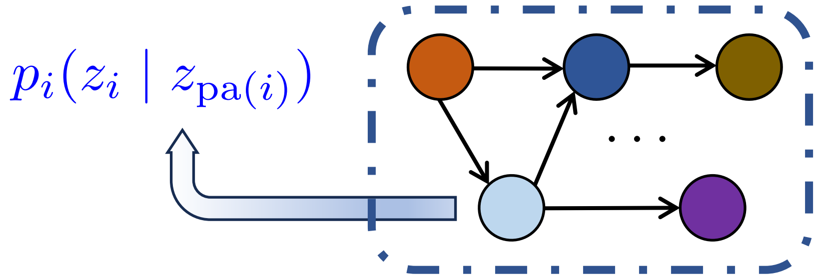
One hard / node

Obs. dim	Norm. Z error	DAG error (SHD)	Norm. Z error	DAG error (SHD)
8	0.50	5.4	0.75	10.3
25	0.51	6.0	0.78	8.9
40	0.50	5.3	0.61	11.9
	score oracle		noisy scores	

Y. Song, S. Garg, J. Shi, and S. Ermon, "Sliced score matching: A scalable approach to density and score estimation," UAI 2020

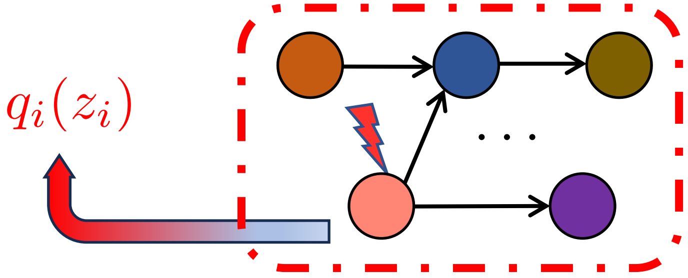
Score-based CRL

Observational Env.



$$s(z) \triangleq \nabla_z \log p(z)$$

Interventional Env.

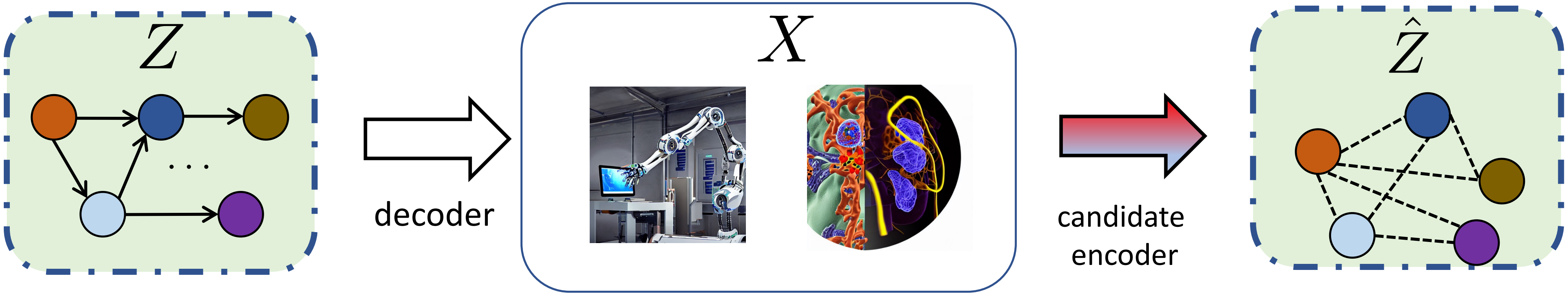


$$s^m(z) \triangleq \nabla_z \log p^m(z)$$

$$s(z) - s^m(z) = \nabla_z \log p_i(z_i \mid z_{\mathrm{pa}(i)}) - \nabla_z \log q_i(z_i)$$

Score functions contain all the information about latent DAGs

Applying an Encoder



$$Z \xrightarrow{g} X \xrightarrow{h} \hat{Z}(h)$$

incorrect encoder $\rightarrow$ $s_{\hat{Z}}(\hat{z}) - s^{m}_{\hat{Z}}(\hat{z})$ **not** a function of only $z_{\overline{\mathrm{pa}}(i)}$

estimated score differences cannot be sparser than true score differences

Minimizing Score Differences

Minimize score variations over environment pairs = correct encoder

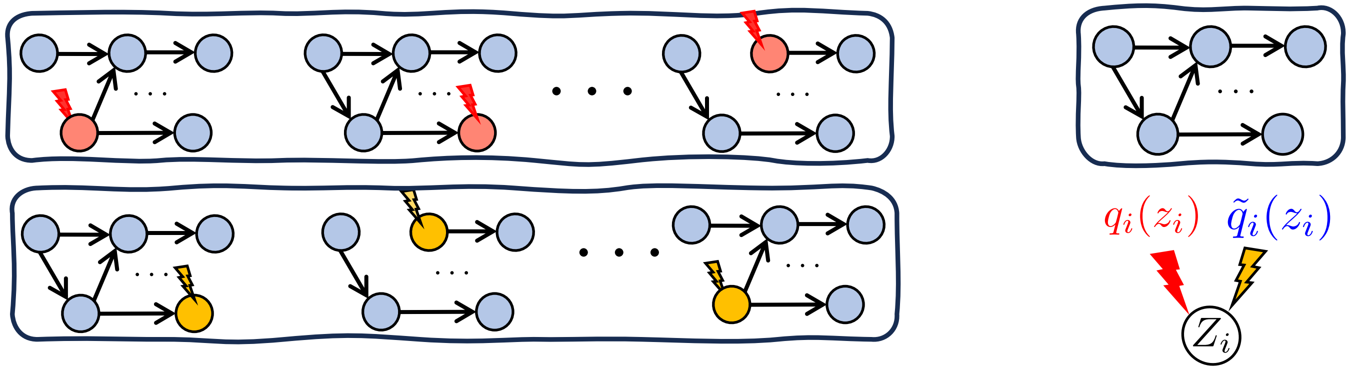
Only need the score differences in observations space

$$s_{\hat{Z}}(\hat{z}) - s^m_{\hat{Z}}(\hat{z}) = [J_{\mathsf{decoder}}(\hat{z})]^\top (s_X(x) - s^m_X(x))$$

Our Contributions

Latent model		Transform		Interventions		Main results	
Nonparametric	+	Nonparametric	+	Two hard	=	1. perfect ID 2. provably correct algo	CRL@ NeurIPS
Sufficiently nonlinear	+	Linear	+	One hard (soft)	=	1. perfect ID (1. perfect DAG + Markov) 2. provably correct algo	Varıcı et al. 2023
Nonparametric	+	Linear	+	One hard (soft)	=	1. perfect ID (1. ID up to ancestors) 2. provably correct algo	Coming soon

Nonparametric transform + two hard



Interventional discrepancy: $\frac{\partial}{\partial z_i}\frac{q_i(z_i)}{\tilde{q}_i(z_i)} \neq 0$ almost everywhere

Theorem : Observational data and **two hard** interventions/node = **Perfect ID**

von Kügelgen et al. (2023): Coupled two hard + faithfulness (for all candidates) = Perfect ID

Linear transform + nonlinear latents + one hard/soft

Theorem : Linear transform + sufficiently nonlinear latent model + **one hard**/node

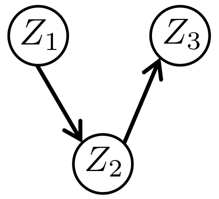
$$X = T \cdot Z$$

Perfect ID

Theorem : Linear transform + sufficiently nonlinear latent model + **one soft**/node

1. Perfect DAG recovery
2. Estimated latents have Markov property

Going beyond mixing with ancestors (nonlinear models = up to ancestors in Zhang'23)



$$\hat{Z}_1 = Z_1$$
$$\hat{Z}_2 = Z_2$$
$$\hat{Z}_3 = \mathrm{lin}(Z_2, Z_3)$$

Linear transform + any latents + one hard/soft

Theorem : Linear transform + **one hard**/node = **Perfect ID**

No parametric restrictions on latents (linear models on Squires'23, Buchholz'23)

Negative results for linear latents + soft (Squires'23): ID up to ancestors is the best one can hope.

Theorem : Linear transform + **one soft**/node = **ID up to ancestors**

$$\hat{Z}_i = \mathrm{lin}(Z_{\mathrm{anc}(i)}) \quad \text{and} \quad \text{tr-closure}(\hat{\mathcal{G}}_Z) = \text{tr-closure}(\mathcal{G}_Z)$$

Summary

- Score functions contain all the information about latent DAG
- Minimizing score variations = constructive proof = provably correct algorithms
- Non-parametric transform with 2 interventions/node: https://arxiv.org/abs/2310.15450
- Liner transform with 1 intervention/node: https://arxiv.org/abs/2301.08230

References

B. Varıcı, E. Acartürk, K. Shanmugam, and A. Tajer. "*Score-based causal representation learning from interventions: Nonparametric identifiability*". Causal Representation Learning Workshop at NeurIPS 2023.

B. Varıcı, E. Acartürk, K. Shanmugam, A. Kumar, and A. Tajer. "*Score-based causal representation with interventions*". arXiv: 2301.08230

K. Ahuja, D. Mahajan, Y. Wang, and Y. Bengio. *"Interventional causal representation learning".* ICML 2023

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

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

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

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

W. Liang, A. Kekić, J. von Kügelgen, S. Buchholz, M. Besserve, L. Gresele, and B. Schölkopf. "*Causal component analysis"*. NeurIPS 2023

Y. Jiang and B. Aragam. "*Learning nonparametric latent causal graphs with unknown interventions*". NeurIPS 2023

S. Bing, U. Ninad, J. Wahl, and J. Runge. "*Identifying Linearly-Mixed Causal Representation Learning from Multi-Node Interventions*", arXiv:2311.02695, 2023

J. Jin and V. Syrgkanis. "*Learning Causal Representations from General Environments: Identifiability and Intrinsic Ambiguity*." arXiv:2311.12267, 2023.

P. Sanchez, X. Liu, A. Q. O'Neil, and S. A. Tsaftaris. "*Diffusion models for causal discovery via topological ordering*". ICLR 2023

P. Rolland, V. Cevher, M. Kleindessner, C. Russell, D. Janzing, B. Schölkopf, and F. Locatello. "*Score matching enables causal discovery of nonlinear additive noise models"*. ICML 2022

F. Montagna, N. Noceti, L. Rosasco, K. Zhang, and F. Locatello. "*Scalable causal discovery with score matching*". CLeaR 2023