Causal Component Analysis

Liang Wendong[1,2],
Armin Kekić[1], Julius von Kügelgen[1,3], Simon Buchholz[1], Michel Besserve[1],
Luigi Gresele[1,†], Bernhard Schölkopf[1,†]

[1] Max Planck Institute for Intelligent Systems, Tübingen
[2] ENS Paris-Saclay, Gif-sur-Yvette, France
[3] University of Cambridge
[†] Joint supervision.

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

école normale supérieure paris—saclay

UNIVERSITY OF CAMBRIDGE

Causal Representation Learning (CRL) & Independent Component Analysis (ICA)

Causal Representation Learning (CRL) (Schölkopf et al., 2021) aims to identify *causally related* latent variables, together with a causal graph encoding their relationships.

CRL provides a principled definition of *disentanglement* (Bengio et al., 2013).

Difficult problem; vast (and growing) literature, often based on strong assumptions.

Counterfactual data (von Kügelgen* et al., 2021; Brehmer et al., 2022); **Temporal structure** (and **graph sparsity**) (Lachapelle and Lacoste-Julien, 2022; Lippe et al., 2022); **Parametric family of latent distributions** (Lachapelle, Rodriguez, et al., 2022; Squires et al., 2023; Buchholz et al., 2023); **Strong restrictions on the mixing function class (e.g., linear, parametric)** (Squires et al., 2023; Varici et al., 2023; Ahuja et al., 2022).

Causal Representation Learning (CRL) & Independent Component Analysis (ICA)

Causal Representation Learning (CRL) (Schölkopf et al., 2021) aims to identify *causally related* latent variables, together with a causal graph encoding their relationships.

CRL provides a principled definition of *disentanglement* (Bengio et al., 2013).

Difficult problem; vast (and growing) literature, often based on strong assumptions.

Counterfactual data (von Kügelgen* et al., 2021; Brehmer et al., 2022); **Temporal structure** (and **graph sparsity**) (Lachapelle and Lacoste-Julien, 2022; Lippe et al., 2022); **Parametric family of latent distributions** (Lachapelle, Rodriguez, et al., 2022; Squires et al., 2023; Buchholz et al., 2023); **Strong restrictions on the mixing function class (e.g., linear, parametric)** (Squires et al., 2023; Varici et al., 2023; Ahuja et al., 2022).

Independent Component Analysis (ICA) (Comon, 1994) aims to recover independent latent variables from observed mixtures thereof.

It is a special case of CRL, where the latent graph is known and empty.

Causal Component Analysis (CauCA)

CRL is challenging: Both the graph and the unmixing are unknown.

Causal Component Analysis (CauCA)

CRL is challenging: Both the graph and the unmixing are unknown.

We introduce *Causal Component Analysis (CauCA)*:

- **Special case of CRL**, presupposes knowledge of the causal graph;
- **Generalization of ICA**, causal components which support interventions.

Causal Bayesian networks (CBNs) and interventions (Pearl, 2009)

In a CBN with graph G, the conditional probabilities $\mathbb{P}_i\left(Z_i \mid \mathbf{Z}_{\mathrm{pa}(i)}\right)$ in the corresponding Markov factorization are called *causal mechanisms*. A CBNs consist of:

- a graph G & a collection of **causal mechanisms** $\{\mathbb{P}_i(Z_i \mid \mathbf{Z}_{\mathrm{pa}(i)})\}_{i\in[d]}$.
- a collection of **(stochastic) interventions** $\{\{\widetilde{\mathbb{P}}_j^k\left(Z_j \mid \mathbf{Z}_{\mathrm{pa}^k(j)}\right)\}_{j\in\tau_k}\}_{k\in[K]}$ across K interventional regimes, with $\tau_k \subseteq V(G)$ *intervention targets*.

The joint probability for interventional regime k is given by:

$$\mathbb{P}^k(\mathbf{Z}) := \begin{cases} \prod_{i=1}^d \mathbb{P}_i(Z_i \mid \mathbf{Z}_{\mathrm{pa}(i)}) & k = 0 \\ \prod_{j\in\tau_k} \widetilde{\mathbb{P}}_j^k\left(Z_j \mid \mathbf{Z}_{\mathrm{pa}^k(j)}\right) \prod_{i\notin\tau_k} \mathbb{P}_i\left(Z_i \mid \mathbf{Z}_{\mathrm{pa}(i)}\right) & \forall k \in [K] \end{cases}$$

where $\mathbb{P}^0$ is the *unintervened* distribution, and $\mathbb{P}^k$ are *interventional* distributions.

Causal Bayesian networks (CBNs) and interventions (Pearl, 2009)

In a CBN with graph G, the conditional probabilities $\mathbb{P}_i\left(Z_i \mid \mathbf{Z}_{\mathrm{pa}(i)}\right)$ in the corresponding Markov factorization are called *causal mechanisms*. A CBNs consist of:

- a graph G & a collection of **causal mechanisms** $\{\mathbb{P}_i(Z_i \mid \mathbf{Z}_{\mathrm{pa}(i)})\}_{i\in[d]}$.
- a collection of **(stochastic) interventions** $\{\{\widetilde{\mathbb{P}}_j^k\left(Z_j \mid \mathbf{Z}_{\mathrm{pa}^k(j)}\right)\}_{j\in\tau_k}\}_{k\in[K]}$ across K interventional regimes, with $\tau_k \subseteq V(G)$ *intervention targets*.

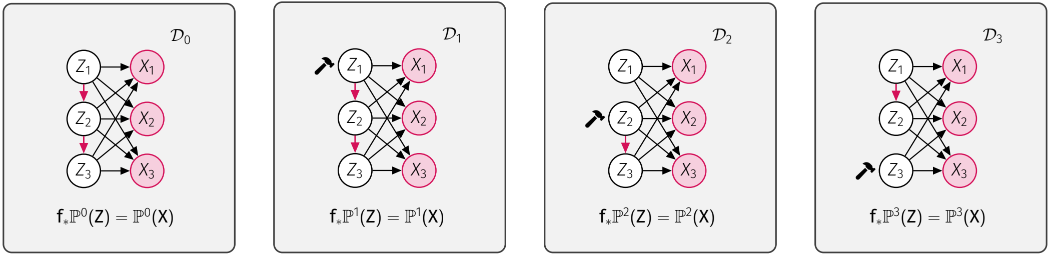
The joint probability for interventional regime k is given by:

$$\mathbb{P}^k(\mathbf{Z}) := \begin{cases} \prod_{i=1}^d \mathbb{P}_i(Z_i \mid \mathbf{Z}_{\mathrm{pa}(i)}) & k=0 \\ \prod_{j\in\tau_k} \widetilde{\mathbb{P}}_j^k\left(Z_j \mid \mathbf{Z}_{\mathrm{pa}^k(j)}\right) \prod_{i\notin\tau_k} \mathbb{P}_i\left(Z_i \mid \mathbf{Z}_{\mathrm{pa}(i)}\right) & \forall k \in [K] \end{cases}$$

where $\mathbb{P}^0$ is the *unintervened* distribution, and $\mathbb{P}^k$ are *interventional* distributions.

Causal mechanisms are *modular*:
We can modify some without affecting the others (Pearl, 2009; Peters et al., 2017).

CauCA based on multiple interventions

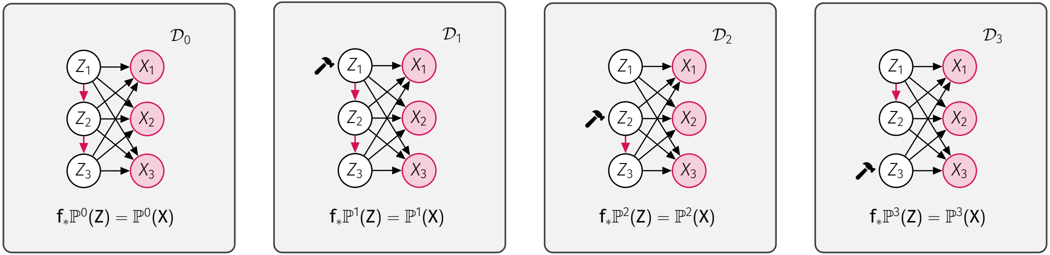


We are given multiple datasets $\mathcal{D}_k$ generated by stochastic interventions on **Z**,

$$\mathcal{D}_k := \left(\tau_k, \left\{\mathbf{x}^{(n,k)}\right\}_{n=1}^{N_k}\right), \quad \text{with} \quad \mathbf{x}^{(n,k)} = \mathbf{f}\left(\mathbf{z}^{(n,k)}\right) \quad \text{and} \quad \mathbf{z}^{(n,k)} \overset{\text{i.i.d.}}{\sim} \mathbb{P}^k,$$

where $\mathbb{P}^k$ are **nonparametric** distributions of **z**, and **f** is a **diffeomorphism**.

CauCA based on multiple interventions



We are given multiple datasets $\mathcal{D}_k$ generated by stochastic interventions on $\mathbf{Z}$,

$$\mathcal{D}_k := \left(\tau_k, \left\{\mathbf{x}^{(n,k)}\right\}_{n=1}^{N_k}\right), \quad \text{with} \quad \mathbf{x}^{(n,k)} = \mathbf{f}\left(\mathbf{z}^{(n,k)}\right) \quad \text{and} \quad \mathbf{z}^{(n,k)} \overset{\text{i.i.d.}}{\sim} \mathbb{P}^k,$$

where $\mathbb{P}^k$ are **nonparametric** distributions of $\mathbf{z}$, and $\mathbf{f}$ is a **diffeomorphism**.

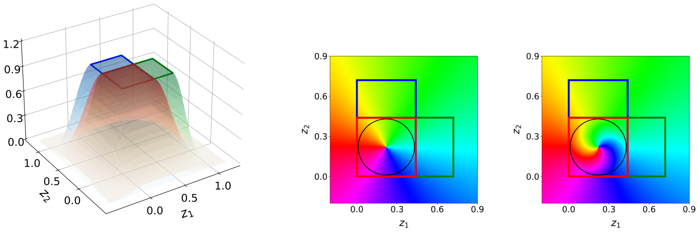
Our goal: Identify both the f and the $\{\mathbb{P}_i^k(z_i|\mathbf{z}_{\text{pa}(i)})\}_{i,k}$.

Identifiability results for a nontrivial latent graph

Consider the example where the latent graph is given by $z_1 \to z_2 \to z_3$:

Requirement of interventions	Learned representation $\hat{z} = \hat{f}^{-1}(x)$	Reference
1 intervention per node	$[h_1(z_1), h_2(z_1, z_2), h_3(z_1, z_2, z_3)]$	Thm. 4.2 *(i)*
1 perfect intervention per node	$[h_1(z_1), h_2(z_2), h_3(z_3)]$	Thm. 4.2 *(ii)*
1 intervention per node for z_1 and z_2, plus $\lvert\overline{\text{pa}}(3)\rvert(\lvert\overline{\text{pa}}(3)\rvert+1) = 2\times 3$ imperfect interventions on z_3 with *"variability"* assumption	$[h_1(z_1), h_2(z_2), h_3(z_2, z_3)]$	Prop. D.1
1 perfect intervention on z_1 and $2+1=3$ perfect fat-hand interventions on (z_2, z_3)	$[h_1(z_1), h_2(z_2, z_3), h_3(z_2, z_3)]$	Thm. 4.5

Negative results: non-identifiability when some assumptions are violated



Our paper also contributes 4 non-identifiability results when:

- The assumption of *interventional discrepancy* is violated;
- At least 1 of the variables is not intervened on (when the graph is not empty);
- At least 2 variables are not intervened on (when the graph is empty, i.e., ICA);
- When the targets of interventions are totally unknown.

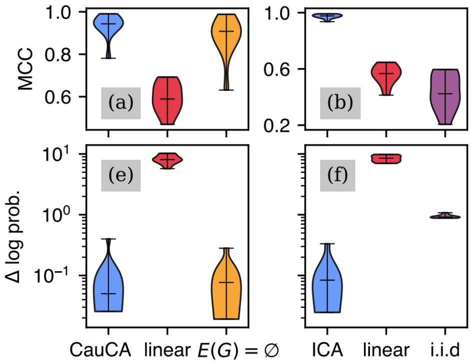
These results show that our assumptions are necessary, even for CRL.

Identifiability results for a trivial latent graph (ICA)

Consider the example when the latent graph is empty (i.e., no arrows):

Requirement of interventions	Learned representation $\hat{\mathbf{z}} = \hat{\mathbf{f}}^{-1}(\mathbf{x})$	Reference
1 intervention on any two nodes respectively	$[h_1(z_1), h_2(z_2), h_3(z_3)]$	Prop. 4.6
1 intervention on z_1 and 2 fat-hand interventions on (z_2, z_3)	$[h_1(z_1), h_2(z_2, z_3), h_3(z_2, z_3)]$	Corollary 4.8
1 intervention on z_1 and 4 fat-hand interventions on (z_2, z_3) with "variability" assumption	$\left[h_1(z_1), \pi[h_2(z_2), h_3(z_3)]\right]$	Prop. 4.9
1 intervention per node on any two nodes respectively with unknown order	$\pi\,[h_1(z_1), h_2(z_2), h_3(z_3)]$	Prop. E.6
6 fat-hand interventions on (z_1, z_2, z_3) with "variability" assumption	$\pi\,[h_1(z_1), h_2(z_2), h_3(z_3)]$	Hyvärinen et al., 2019, Thm. 1

Summary of our main contributions



We introduce a likelihood-based approach using normalizing flows to estimate both the unmixing function and the causal mechanisms.

Our method effectively recovers the latent causal components in synthetic experiments.

Conclusion

CauCA simplifies CRL by assuming knowledge of the causal graph.

- *Impossibility* results for CauCA also apply for CRL.
- *Possibility* results can inspire corresponding results in CRL.

Conclusion

CauCA simplifies CRL by assuming knowledge of the causal graph.

- *Impossibility* results for CauCA also apply for CRL.
- *Possibility* results can inspire corresponding results in CRL.

By studying CauCA, we gain insights into the minimal assumptions required for CRL.

Thank you for your attention!

Poster Session: Wed 13 Dec 8:45 a.m. PST — 10:45 a.m. PST

Location: Great Hall & Hall B1+B2 #827

Paper: https://arxiv.org/abs/2305.17225

Code: https://github.com/akekic/causal-component-analysis

References

- Ahuja, Kartik et al. (2022). **"Interventional Causal Representation Learning".** In: *arXiv preprint arXiv:2209.11924*.
- Bengio, Yoshua et al. (2013). **"Representation learning: A review and new perspectives".** In: *IEEE transactions on pattern analysis and machine intelligence* 35.8, pp. 1798–1828.
- Brehmer, Johann et al. (2022). **"Weakly supervised causal representation learning".** In: *arXiv preprint arXiv:2203.16437*.
- Buchholz, Simon et al. (2023). **"Learning Linear Causal Representations from Interventions under General Nonlinear Mixing".** In: *arXiv preprint arXiv:2306.02235*.
- Comon, Pierre (1994). **"Independent component analysis, a new concept?"** In: *Signal processing* 36.3, pp. 287–314.

References ii

- Hyvärinen, Aapo et al. (2019). **"Nonlinear ICA using auxiliary variables and generalized contrastive learning".** In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 859–868.
- Lachapelle, Sébastien and Simon Lacoste-Julien (2022). **"Partial Disentanglement via Mechanism Sparsity".** In: *arXiv preprint arXiv:2207.07732*.
- Lachapelle, Sébastien, Pau Rodriguez, et al. (2022). **"Disentanglement via mechanism sparsity regularization: A new principle for nonlinear ICA".** In: *Conference on Causal Learning and Reasoning*. PMLR, pp. 428–484.
- Lippe, Phillip et al. (2022). **"Citris: Causal identifiability from temporal intervened sequences".** In: *International Conference on Machine Learning*. PMLR, pp. 13557–13603.
- Pearl, Judea (2009). ***Causality.*** Cambridge university press.
- Peters, Jonas et al. (2017). ***Elements of causal inference: foundations and learning algorithms.*** The MIT Press.
- Schölkopf, Bernhard et al. (2021). **"Toward causal representation learning".** In: *Proceedings of the IEEE* 109.5, pp. 612–634.

Squires, Chandler et al. (2023). *Linear Causal Disentanglement via Interventions.*

Varici, Burak et al. (2023). **"Score-based Causal Representation Learning with Interventions".** In: *arXiv preprint arXiv:2301.08230*.

von Kügelgen*, Julius et al. (2021). **"Self-Supervised Learning with Data Augmentations Provably Isolates Content from Style".** In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 34.