Abstracting Causal Models

Fabio Massimo Zennaro
fabio.zennaro@warwick.ac.uk

University of Warwick
February 17th, 2022

Outline

1. Introduction

Problem definition

Systems may be represented at different **levels of abstraction** (LoA).

Thermodynamics example:

Low-level / Base model:
Microscopic description $\mathbf{p}, \dot{\mathbf{p}}$.

High-level / Abstracted model:
Macroscopic description P, T, V.

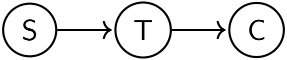
LoA may be inaccessible, so we may want to *shift* among LoAs.

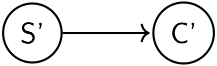
- We need a *mapping* between LoAs.
- We want the mapping to be *consistent*.
  - Ideally consistency is not only *observational*, but *interventional* too.

Problem definition

SCMs are becoming more popular for encoding causal models.

Lung cancer scenario example:





- How do we find a *mapping*?
- How do we define and guarantee some form of *consistency*?

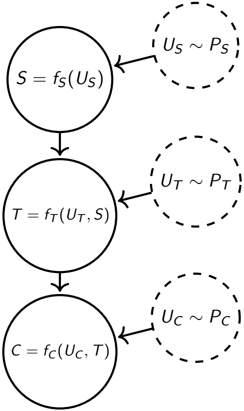
This could allow us to shift between LoAs of SCMs, taking advantage of data and computational resources.

2. Background

SCMs

We express a causal model as a **structural causal model** $\mathcal{M}$ [5, 6]:

- $\mathcal{X}$: set of *endogenous nodes* (S, T, C) representing variables of interest
- $\mathcal{E}$: Set of *exogenous nodes* (U_S, U_T, U_C) representing stochastic factors
- $\mathcal{F}$: Set of *structural functions* (f_S, f_T, f_C) describing the dynamics of each variable
- $\mathcal{P}$: Set of *distributions* (P_S, P_T, P_C) describing the random factors



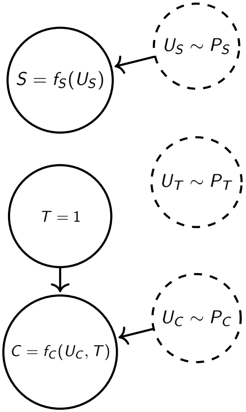
Every SCM $\mathcal{M}$ implies a (joint) **distribution** $P_\mathcal{M}$: $P_\mathcal{M}(S, T, C)$

Interventions

We can perform **interventions** on a causal model [5, 6]:

$do(T = 1)$

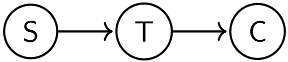
1. Remove incoming edges in the intervened node
2. Set the value of the intervened node

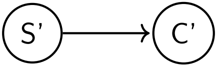


An intervention ι_1 effectively defines a new **intervened model** $\mathcal{M}_{\iota_1}$ such that $P_{\mathcal{M}}(S, T, C) \neq P_{\mathcal{M}_{\iota_1}}(S, T, C)$

Consistency

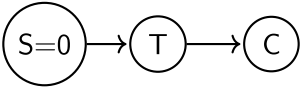
- **Observational consistency:** sampling the two models I obtain the same (observational) distributions of interest.

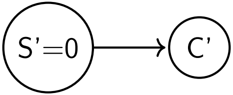




$$P_{\mathcal{M}}(S, C) = P_{\mathcal{M}'}(S', C')$$

- **Interventional consistency:** under an intervention the two models produce the same (interventional) distributions of interest.

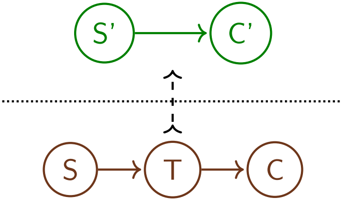




$$P_{\mathcal{M}}(C|do(S = 0)) = P_{\mathcal{M}'}(C'|do(S' = 0))$$

Two approaches

Lung cancer scenario example:



$$\mathcal{M}'[S'] = \mathcal{M}'[C'] = \{0,1\}$$

$$\mathcal{M}[S] = \mathcal{M}[T] = \mathcal{M}[C] = \{0,1\}$$

- The **transformation** approach [9]
- The **abstraction** approach [8]

3. Transformation approach [9]

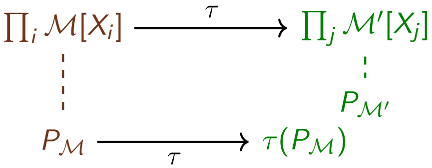
The *transformation* approach: transformation

Given two SCMs $\mathcal{M}$ and $\mathcal{M}'$, let us consider the **transformation**:

$$\tau : \prod_i \mathcal{M}[X_i] \rightarrow \prod_j \mathcal{M}'[X_j]$$

τ : domain of the variables of $\mathcal{M}$ $\rightarrow$ domain of the variables of $\mathcal{M}'$.
τ : an output/configuration of $\mathcal{M}$ $\mapsto$ an output/configuration of $\mathcal{M}'$.

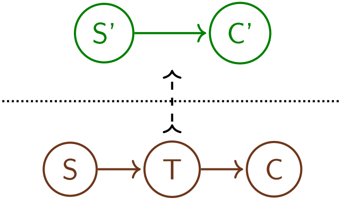
This implies a (pushforwarded) distribution on $\mathcal{M}'$:



If $\tau(P_{\mathcal{M}}) = P_{\mathcal{M}'}$ we have *observational consistency*.

The *transformation* approach: an example (I)

Lung cancer scenario example:



$\tau : \mathcal{M}[S] \times \mathcal{M}[T] \times \mathcal{M}[C] \rightarrow \mathcal{M}'[S'] \times \mathcal{M}'[C']$

$\tau : \{0,1\}^3 \rightarrow \{0,1\}^2$

$\tau : (s,t,c) \mapsto (s,c)$
$\tau : (0,1,1) \mapsto (0,1)$

Observational consistency condition:

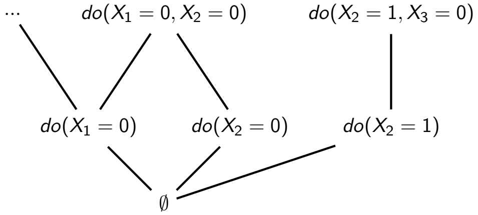
$$\{0,1\}^3 \xrightarrow{\tau} \{0,1\}^2$$

$$P_{\mathcal{M}} \xrightarrow[\tau]{} \tau(P_{\mathcal{M}}) = P_{\mathcal{M}'}$$

The *transformation* approach: poset of interventions

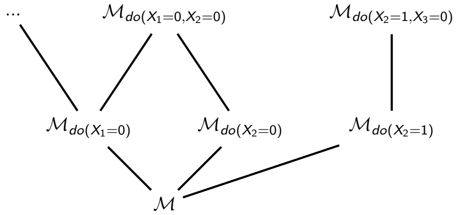
Let us now consider a set of interventions of interest $\mathcal{I}$ on $\mathcal{M}$.

The set of interventions has a *partially ordered set* structure wrt inclusion.



The *transformation* approach: poset of interventions

The poset of interventions induces a *partially ordered set* structure over SCMs.



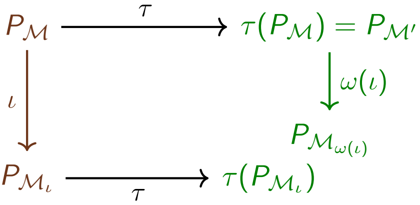
The *transformation* approach: exact transformation

Let us consider a mapping between interventions:

$$\omega : \mathcal{I} \to \mathcal{I}'$$

$$\omega : \text{an intervention on } \mathcal{M} \mapsto \text{an intervention on } \mathcal{M}'.$$

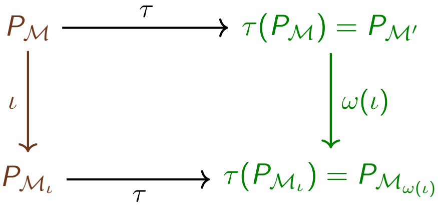
A transformation is an *exact transformation* if there exist a surjective order-preserving ω such that:



where $\tau(P_{\mathcal{M}_\iota}) = P_{\mathcal{M}_{\omega(\iota)}}$ for every $\iota \in \mathcal{I}$.

The *transformation* approach: consistency

(Interventional) consistency is the commutativity of the diagram:

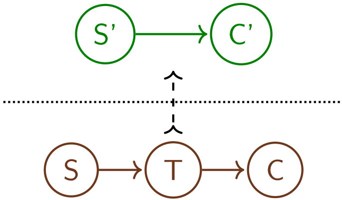


It produces the same result to:

- abstract, then intervene $(\omega(\iota) \circ \tau)$
- intervene, then abstract $(\tau \circ \iota)$

The *transformation* approach: an example (II)

Lung cancer scenario example:



$\tau : (s, t, c) \mapsto (s, c)$

Set of interventions: $\mathcal{I} = \{\emptyset, do(S = 0)\}$

$$\omega : \begin{cases} \emptyset \mapsto \emptyset \\ do(S = 0) \mapsto do(S' = 0) \end{cases}$$

Consistency condition:

$$P_{\mathcal{M}}(S, T, C) \xrightarrow{\tau} P_{\mathcal{M}'}(S', C')$$

$$P_{\mathcal{M}}(S, T, C) \xrightarrow{\iota} P_{\mathcal{M}}(T, C | do(S = 0)) \qquad P_{\mathcal{M}'}(S', C') \xrightarrow{\omega(\iota)} P_{\mathcal{M}'}(C' | do(S' = 0))$$

$$P_{\mathcal{M}}(T, C | do(S = 0)) \not\mapsto P_{\mathcal{M}'}(C' | do(S' = 0))$$

The *transformation* approach: summary

Given:

- A low-level model $\mathcal{M}$ with a set of interventions of interest $\mathcal{I}$;
- A high-level model $\mathcal{M}'$;
- A surjective order-preserving $\omega : \mathcal{I} \rightarrow \mathcal{I}'$

an **exact transformation** τ guarantees that if I:

- work (intervene) at low-level and then switch (abstract) to high-level,
- or, switch first to high-level and then work there,

I will observe the same statistical behavior in the two models.

The *transformation* approach: a few observations

- A *coarse-grained* description of abstraction.
- *Structural information* mediated only through interventions.
- Consistency wrt to a limited *set of interventions*.
- Work with *continuous models*.

4. Abstraction approach [8]

The *abstraction* approach: abstraction

Let $\mathcal{M}$ and $\mathcal{M}'$ be two finite SCMs with finite domains. An **abstraction** is a tuple

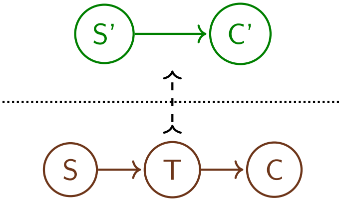
$$(R, a, \alpha)$$

where

- $R \subseteq \mathcal{X}_{\mathcal{M}}$ is a subset of *relevant nodes* among the endogenous nodes of $\mathcal{M}$.
- $a : R \rightarrow \mathcal{X}_{\mathcal{M}'}$ is a *surjective function* mapping a low-level node in $\mathcal{M}$ to a high-level node in $\mathcal{M}'$.
- α is a *collection of surjective functions*, one for each high-level node X', defined as $\alpha_{X'} : \mathcal{M}[a^{-1}(X')] \rightarrow \mathcal{M}'[X']$.
  α'_X maps an output of the low-level nodes sent onto X' by a onto an output of X'.

The *abstraction* approach: an example (I)

Lung cancer scenario example:



$$R = \{S, C\} \subseteq \mathcal{X}_{\mathcal{M}}$$

$$a : R \rightarrow \mathcal{X}_{\mathcal{M}'}$$

$$a : \begin{cases} S \mapsto S' \\ C \mapsto C' \end{cases}$$

$$\alpha : \begin{cases} \alpha_{S'} : \{0,1\} \rightarrow \{0,1\} \\ \alpha_{S'} : s \mapsto s \\ \alpha_{C'} : \{0,1\} \rightarrow \{0,1\} \\ \alpha_{C'} : c \mapsto c \end{cases}$$

The *abstraction* approach: consistency

We have *(interventional) consistency* if the following diagram commutes for all the disjoint subsets $X', Y' \in \mathcal{X}_{\mathcal{M}'}$ for every value in $\mathcal{M}[a^{-1}(X')]$:

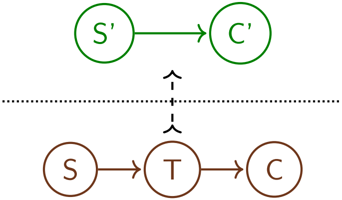
$$\begin{array}{ccc} \mathcal{M}[a^{-1}(X')] & \xrightarrow{\mathcal{M}[\phi_{a^{-1}(Y')}]} & \mathcal{M}[a^{-1}(Y')] \\ \downarrow \alpha_{X'} & & \downarrow \alpha_{Y'} \\ \mathcal{M}'[X'] & \xrightarrow[\mathcal{M}'[\phi_{Y'}]]{} & \mathcal{M}'[Y'] \end{array}$$

It produces the same result to:

- mechanism, then abstract $(\alpha_{Y'} \circ \mathcal{M}[\phi_{a^{-1}(Y')}])$
- abstract, then mechanism $(\mathcal{M}'[\phi_{Y'}] \circ \alpha_{X'})$

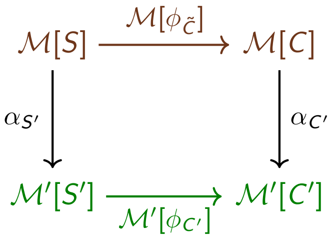
The *abstraction* approach: an example (II)

Lung cancer scenario example:



Disjoint subsets in
$\mathcal{X}_{\mathcal{M}'} = \{\{S'\}, \{C'\}\}$

Consistency condition:



The *abstraction* approach: abstraction error

If the diagram does not commute for $X', Y' \in \mathcal{X}_{\mathcal{M}'}$:

$$
\begin{array}{ccc}
\mathcal{M}[a^{-1}(X')] & \xrightarrow{\mathcal{M}[\phi_{a^{-1}(Y')}]} & \mathcal{M}[a^{-1}(Y')] \\
\Big\downarrow \alpha_{X'} & & \Big\downarrow \alpha_{Y'} \\
\mathcal{M}'[X'] & \xrightarrow[\mathcal{M}'[\phi_{Y'}]]{} & \mathcal{M}'[Y']
\end{array}
$$

I can compute the *abstraction error* for X', Y':

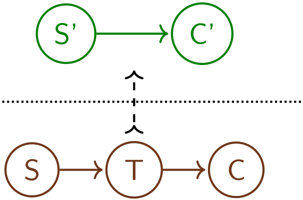
$$E_\alpha(X', Y') = D_{JSD}(\alpha_{Y'} \circ \mathcal{M}[\phi_{a^{-1}(Y')}], \mathcal{M}'[\phi_{Y'}] \circ \alpha_{X'})$$

I can compute the *overall abstraction error* as the worst-case:

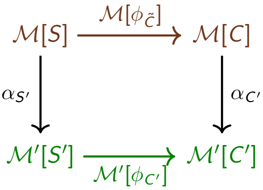
$$e(\alpha) = \sup_{X', Y' \in \mathcal{X}_{\mathcal{M}'}} E_\alpha(X', Y')$$

The *abstraction* approach: an example (II)

Lung cancer scenario example:



Assuming no commutativity



I can compute *abstraction error*:
$E_\alpha(S', C') = D_{JSD}(\alpha_{C'} \circ \mathcal{M}[\phi_{\tilde{C}}], \mathcal{M}'[\phi_{C'}] \circ \alpha_{S'})$

Since there are not other subsets this is also the *overall abstraction error*:
$e_\alpha = E_\alpha(S', C')$

The *abstraction* approach: summary

Given:

- A low-level model $\mathcal{M}$;
- A high-level model $\mathcal{M}'$;
- An abstraction (R, a, α)

a **zero-error abstraction** guarantees that, under intervention, if I:

- work (mechanism) at low-level and then switch (abstract) to high-level,
- or, switch first to high-level and then work there,

I will observe the same statistical behavior in the two models.

The *abstraction* approach: a few observations

- A *fine-grained* description of abstraction.
- *Structure* defines abstraction.
- Consistency wrt to *all interventions* (in a finite set).
- Work with *finite models*.
- Finiteness reduces SCMs to sets and stochastic matrices.
- Commuting diagram grounded in category theory.

5. Conclusions

Properties of abstraction

We discussed:

- Observational consistency
- Interventional consistency

We have not dealt with:

- Compositionality [9, 8, 7]
- Counterfactual consistency
- Locality
- Other formalizations [2, 1, 4]

Learning/discovery/search aspects

We discussed:

- Formal setup of abstraction
- Well-defined models
- Verification of properties of abstractions

We have not dealt with:

- Learning abstractions
  - Learning causal features [3]
- Transferring knowledge between models
  - Homogeneity of abstractions and interventions

Thanks!

Thank you for listening!

References I

[1] Sander Beckers, Frederick Eberhardt, and Joseph Y Halpern. Approximate causal abstractions. In *Uncertainty in Artificial Intelligence*, pages 606–615. PMLR, 2020.

[2] Sander Beckers and Joseph Y Halpern. Abstracting causal models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2678–2685, 2019.

[3] Krzysztof Chalupka, Pietro Perona, and Frederick Eberhardt. Visual causal feature learning. *arXiv preprint arXiv:1412.2309*, 2014.

[4] Jun Otsuka and Hayato Saigo. On the equivalence of causal models: A category-theoretic approach. *arXiv preprint arXiv:2201.06981*, 2022.

[5] Judea Pearl. *Causality*. Cambridge University Press, 2009.

[6] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: Foundations and learning algorithms*. MIT Press, 2017.

References II

[7] Eigil F Rischel and Sebastian Weichwald. Compositional abstraction error and a category of causal models. *arXiv preprint arXiv:2103.15758*, 2021.

[8] Eigil Fjeldgren Rischel. The category theory of causal models. 2020.

[9] Paul K Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. In *33rd Conference on Uncertainty in Artificial Intelligence (UAI 2017)*, pages 808–817. Curran Associates, Inc., 2017.