Causal Optimal Transport of Abstractions

Yorgos Felekis, Fabio M. Zennaro, Nicola Branchini, Theodoros Damoulas

May 17, 2024

Outline

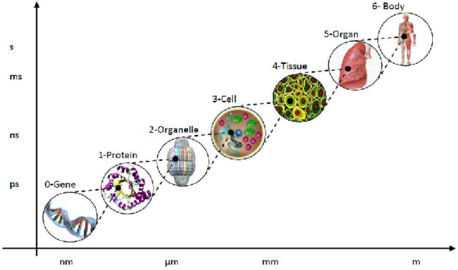
Motivation

Motivation

Complex systems can be represented at **different levels of abstraction**!

Example: Biology

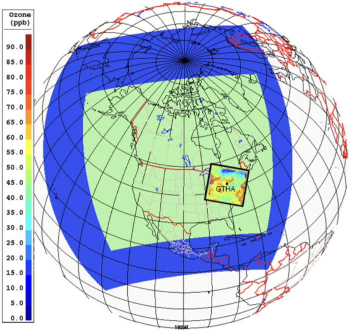
- Micro/low-level model: Focuses on cellular processes within organs; provides insights into the intricate mechanisms that govern cellular behavior within specific organs.
- Macro/high-level model: Describes the overall functionality and interactions of organs within the body; provides a holistic view of how organs collaborate to sustain life at the body level.



Credit: Barbulescu and Ioan 2015

Example: Climate

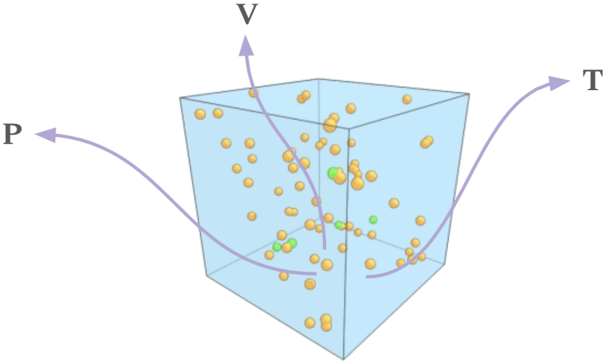
- Micro/low-level model: describes local phenomena with high resolution.
- Macro/high-level model: describes meteorological events at a regional scale.



Credit: Stroud et al. 2020

Example: Physics

- Micro/low-level model: Statistical mechanics study the behaviour of molecules.
- Macro/high-level model: Thermodynamics (P, V, T).



Credit: Sean Kelley/NIST

Motivation

Learning relations between models and underlying representations at different levels is a key challenge across sciences and especially AI as it can enable:

- Aggregation of information
- Transfer learning
- Emulation via surrogate models
- Multi-scale estimation and reasoning

Motivation

Learning relations between models and underlying representations at different levels is a key challenge across sciences and especially AI as it can enable:

- Aggregation of information
- Transfer learning
- Emulation via surrogate models
- Multi-scale estimation and reasoning

Motivation

Learning relations between models and underlying representations at different levels is a key challenge across sciences and especially AI as it can enable:

- Aggregation of information
- Transfer learning
- Emulation via surrogate models
- Multi-scale estimation and reasoning

Motivation

Learning relations between models and underlying representations at different levels is a key challenge across sciences and especially AI as it can enable:

- Aggregation of information
- Transfer learning
- Emulation via surrogate models
- Multi-scale estimation and reasoning

Motivation

- In causal modeling, such models at different levels of abstraction should be **consistent** $\implies$ agree in their predictions of the effects of interventions!
- e.g. if we were to observe the evolution of the climate micro-model under a reduction of CO_2 and then coarsen our result to a regional scale, we would like to obtain the same result as directly observing the evolution of the macro-model under the same intervention of reduction of CO_2.

Motivation

- In causal modeling, such models at different levels of abstraction should be **consistent** $\implies$ agree in their predictions of the effects of interventions!
- e.g. if we were to observe the evolution of the climate micro-model under a reduction of CO_2 and then coarsen our result to a regional scale, we would like to obtain the same result as directly observing the evolution of the macro-model under the same intervention of reduction of CO_2.

Motivation

Goal: Learn a map between two causal models of varying degrees of granularity that describe the same system such that the aforementioned property of consistency holds!

- Causal evidence synthesis
- Causally consistent representations at different resolutions.
- Interventions alignment across models.

Motivation

Goal: Learn a map between two causal models of varying degrees of granularity that describe the same system such that the aforementioned property of consistency holds!

- Causal evidence synthesis
- Causally consistent representations at different resolutions.
- Interventions alignment across models.

Motivation

Goal: Learn a map between two causal models of varying degrees of granularity that describe the same system such that the aforementioned property of consistency holds!

- Causal evidence synthesis
- Causally consistent representations at different resolutions.
- Interventions alignment across models.

Structural Causal Models

Causality

"X is a cause of Y, if Y listens to X and decides its value in response to what it hears.", Judea Pearl

We assume causality to be directed and mechanistic.

Structural Causal Models

A **structural causal model** (SCM) is defined as a tuple:

$$M = \langle \mathbf{X}, \mathbf{U}, \mathcal{F}, \mathbb{P}(\mathbf{U}) \rangle$$

where:

- $\mathbf{X}$ is a set of endogenous variables (*variables of interest*);
- $\mathbf{U}$ is a set of exogenous variables (*noise*);
- $\mathcal{F}$ is a set of structural functions, one for each endogenous node;

$$f_i : \text{dom}[\text{PA}(X_i)] \times \text{dom}[U_i] \to \text{dom}[X_i]$$

where $\text{PA}(X_i) \subseteq \mathbf{X} \setminus X_i$ is the set of parent nodes of X_i in the underlying DAG $\mathcal{G}_M$.
- $\mathbb{P}(\mathbf{U})$ is a set of probability distributions, one for each exogenous node.

Structural Causal Models

A **structural causal model** (SCM) is defined as a tuple:

$$M = \langle \mathbf{X}, \mathbf{U}, \mathcal{F}, \mathbb{P}(\mathbf{U}) \rangle$$

where:

- $\mathbf{X}$ is a set of endogenous variables (*variables of interest*);
- $\mathbf{U}$ is a set of exogenous variables (*noise*);
- $\mathcal{F}$ is a set of structural functions, one for each endogenous node;

$$f_i : \text{dom}[\text{PA}(X_i)] \times \text{dom}[U_i] \to \text{dom}[X_i]$$

  where $\text{PA}(X_i) \subseteq \mathbf{X} \setminus X_i$ is the set of parent nodes of X_i in the underlying DAG $\mathcal{G}_M$.
- $\mathbb{P}(\mathbf{U})$ is a set of probability distributions, one for each exogenous node.

Structural Causal Models

A **structural causal model** (SCM) is defined as a tuple:

$$M = \langle \mathbf{X}, \mathbf{U}, \mathcal{F}, \mathbb{P}(\mathbf{U}) \rangle$$

where:

- $\mathbf{X}$ is a set of endogenous variables (*variables of interest*);
- $\mathbf{U}$ is a set of exogenous variables (*noise*);
- $\mathcal{F}$ is a set of structural functions, one for each endogenous node;

$$f_i : \text{dom}[\text{PA}(X_i)] \times \text{dom}[U_i] \to \text{dom}[X_i]$$

  where $\text{PA}(X_i) \subseteq \mathbf{X} \setminus X_i$ is the set of parent nodes of X_i in the underlying DAG $\mathcal{G}_{\boldsymbol{M}}$.
- $\mathbb{P}(\mathbf{U})$ is a set of probability distributions, one for each exogenous node.

Structural Causal Models

A **structural causal model** (SCM) is defined as a tuple:

$$M = \langle \mathbf{X}, \mathbf{U}, \mathcal{F}, \mathbb{P}(\mathbf{U}) \rangle$$

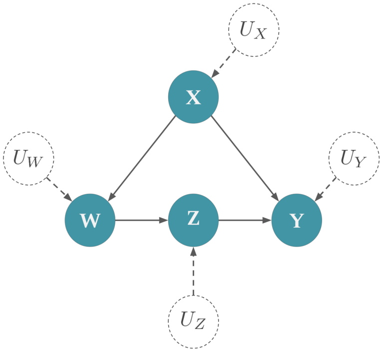
where:

- $\mathbf{X}$ is a set of endogenous variables (*variables of interest*);
- $\mathbf{U}$ is a set of exogenous variables (*noise*);
- $\mathcal{F}$ is a set of structural functions, one for each endogenous node;

  $$f_i : \text{dom}[\text{PA}(X_i)] \times \text{dom}[U_i] \to \text{dom}[X_i]$$

  where $\text{PA}(X_i) \subseteq \mathbf{X} \setminus X_i$ is the set of parent nodes of X_i in the underlying DAG $\mathcal{G}_{\boldsymbol{M}}$.
- $\mathbb{P}(\mathbf{U})$ is a set of probability distributions, one for each exogenous node.

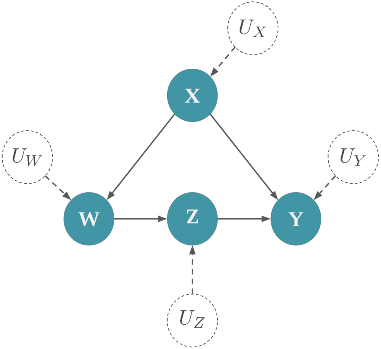
Structural Causal Models



We assume:

- **acyclicity** $\implies$ DAG $\mathcal{G}_{\boldsymbol{M}}$;
- faithfulness $\implies$ independencies in the data are captured in $\mathcal{G}_{\boldsymbol{M}}$;
- causal sufficiency $\implies$ no unobserved confounders.

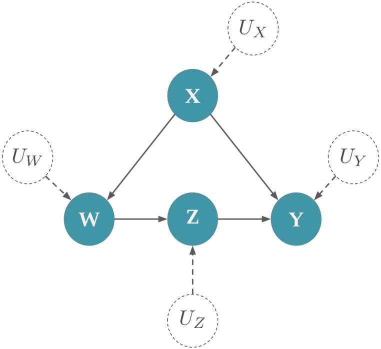
Structural Causal Models



We assume:

- **acyclicity** $\implies$ DAG $\mathcal{G}_{\boldsymbol{M}}$;
- **faithfulness** $\implies$ independencies in the data are captured in $\mathcal{G}_{\boldsymbol{M}}$;
- causal sufficiency $\implies$ no unobserved confounders.

Structural Causal Models



We assume:

- **acyclicity** $\implies$ DAG $\mathcal{G}_{\boldsymbol{M}}$;
- **faithfulness** $\implies$ independencies in the data are captured in $\mathcal{G}_{\boldsymbol{M}}$;
- **causal sufficiency** $\implies$ no unobserved confounders.

Structural Causal Models

The probability distribution $\mathbb{P}(\mathbf{U})$ over the exogenous variables can be push-forwarded over the endogenous variables and define a probability distribution over them $\mathbb{P}(\mathbf{X}) = \mathbb{P}_{\#}(\mathbf{U})$.

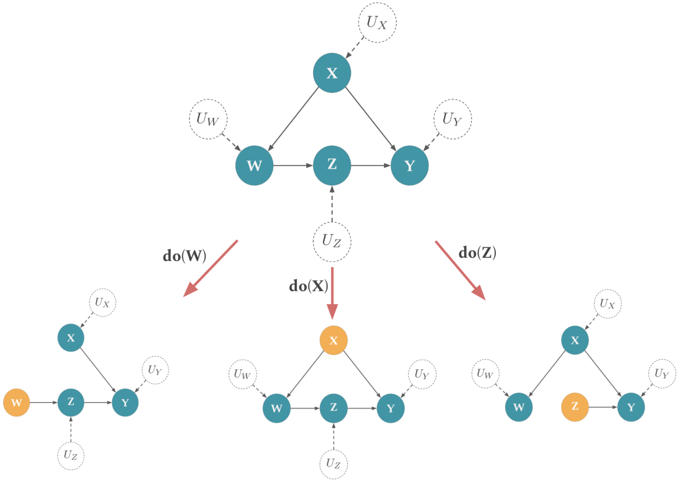
Bayesian Factorization

Given a probability distribution P and a DAG G, P factorizes according to G by the product decomposition rule:

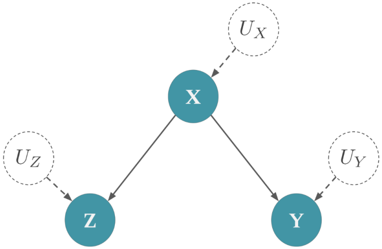
$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i|\text{PA}_i)$$

Structural Causal Models: Interventions

We define an **intervention** operator do(**S**=**s**) on M as the one that replaces the structural function f_i of every $X_i \in S$ with the respective constant s_i. An intervention on M defines a new **post-intervention model** $M_{\text{do}(\mathbf{s})}$.

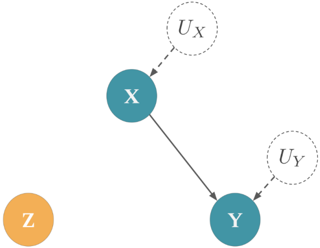


Conditioning $\neq$ Intervening



$\mathbb{P}(Y|Z)$

Seeing Z allows inference on distribution of X and then Y.



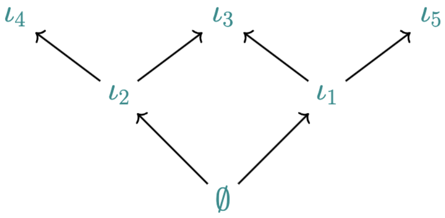
$\mathbb{P}(Y|\text{do}(Z))$

Doing Z does not affect the distribution of X and as a result of Y.

Structural Causal Models: Interventions

Sets of interventions are equipped with a natural partially-ordered set (poset) structure with respect to *containment*. An intervention $\iota = \mathrm{do}(\mathsf{A} = \mathsf{a})$ precedes $\eta = \mathrm{do}(\mathsf{B} = \mathsf{b})$ and we write $\iota \preceq \eta$ iff:

$$\mathsf{A} \subseteq \mathsf{B} \text{ and } \mathsf{a} = \mathrm{proj}(\mathsf{b}, \mathsf{A}) \iff \mathsf{A} \subseteq \mathsf{B} \text{ and for } B_j = A_i \implies b_j = a_i$$



Structural Causal Models: Interventions

Given an SCM $M = \langle \mathbf{X}, \mathbf{U}, \mathscr{F}, \mathbb{P}(\mathbf{U}) \rangle$ and a set of variables $\mathbf{V} \subseteq \mathbf{X}$:

- We call $\mathbf{v} \in \mathbf{V}$ a *partial setting* and $\mathbf{x} \in \mathbf{X}$ a *total setting*.
- The restriction of $\mathbf{x}$ to $\mathbf{V}$ is the projection $\text{proj}(\mathbf{x}, \mathbf{V}) \in \text{dom}[\mathbf{V}]$.
- The restriction $\text{Rst}(M_\iota)$ of an intervention $\iota = \text{do}(\mathbf{V} = \mathbf{v})$ on a model M is the subset of total settings on $\mathbf{X}$ matching the partial setting $\mathbf{v}$.

$$\text{Rst}(M_\iota) = \{\mathbf{x} \in \text{dom}[X] \mid \mathbf{v} = \text{proj}(\mathbf{x}, \mathbf{V})\}$$

- We say that a total setting $\mathbf{x}$ is *compatible* with an intervention $\iota = \text{do}(\mathbf{v})$ and we write $\text{Cmp}(\mathbf{x}, \iota)$ if $\mathbf{x} \in \text{Rst}(M_\iota)$.

  Given a simple SCM which consists of three binary variables X, Y, Z and an intervention $\iota = \text{do}(X = 0, Y = 1)$ then the total settings that are compatible with ι are $(0, 1, 0)$ and $(0, 1, 1)$.

Structural Causal Models: Interventions

Given an SCM $M = \langle \mathbf{X}, \mathbf{U}, \mathcal{F}, \mathbb{P}(\mathbf{U}) \rangle$ and a set of variables $\mathbf{V} \subseteq \mathbf{X}$:

- We call $\mathbf{v} \in \mathbf{V}$ a *partial setting* and $\mathbf{x} \in \mathbf{X}$ a *total setting*.
- The restriction of $\mathbf{x}$ to $\mathbf{V}$ is the projection $\text{proj}(\mathbf{x}, \mathbf{V}) \in \text{dom}[\mathbf{V}]$.
- The restriction $\text{Rst}(M_\iota)$ of an intervention $\iota = \text{do}(\mathbf{V} = \mathbf{v})$ on a model M is the subset of total settings on $\mathbf{X}$ matching the partial setting $\mathbf{v}$.

$$\text{Rst}(M_\iota) = \{\mathbf{x} \in \text{dom}[X] \mid \mathbf{v} = \text{proj}(\mathbf{x}, \mathbf{V})\}$$

- We say that a total setting $\mathbf{x}$ is *compatible* with an intervention $\iota = \text{do}(\mathbf{v})$ and we write $\text{Cmp}(\mathbf{x}, \iota)$ if $\mathbf{x} \in \text{Rst}(M_\iota)$.

Given a simple SCM which consists of three binary variables X, Y, Z and an intervention $\iota = \text{do}(X = 0, Y = 1)$ then the total settings that are compatible with ι are $(0, 1, 0)$ and $(0, 1, 1)$.

Structural Causal Models: Interventions

Given an SCM $M = \langle \mathbf{X}, \mathbf{U}, \mathcal{F}, \mathbb{P}(\mathbf{U}) \rangle$ and a set of variables $\mathbf{V} \subseteq \mathbf{X}$:

- We call $\mathbf{v} \in \mathbf{V}$ a *partial setting* and $\mathbf{x} \in \mathbf{X}$ a *total setting*.
- The restriction of $\mathbf{x}$ to $\mathbf{V}$ is the projection $\text{proj}(\mathbf{x}, \mathbf{V}) \in \text{dom}[\mathbf{V}]$.
- The restriction $\text{Rst}(M_\iota)$ of an intervention $\iota = \text{do}(\mathbf{V} = \mathbf{v})$ on a model M is the subset of total settings on $\mathbf{X}$ matching the partial setting $\mathbf{v}$.

$$\text{Rst}(M_\iota) = \{\mathbf{x} \in \text{dom}[X] \mid \mathbf{v} = \text{proj}(\mathbf{x}, \mathbf{V})\}$$

- We say that a total setting $\mathbf{x}$ is *compatible* with an intervention $\iota = \text{do}(\mathbf{v})$ and we write $\text{Cmp}(\mathbf{x}, \iota)$ if $\mathbf{x} \in \text{Rst}(M_\iota)$.

Given a simple SCM which consists of three binary variables X, Y, Z and an intervention $\iota = \text{do}(X = 0, Y = 1)$ then the total settings that are compatible with ι are $(0, 1, 0)$ and $(0, 1, 1)$.

Structural Causal Models: Interventions

Given an SCM $\boldsymbol{M} = \langle \mathbf{X}, \mathbf{U}, \mathscr{F}, \mathbb{P}(\mathbf{U}) \rangle$ and a set of variables $\mathbf{V} \subseteq \mathbf{X}$:

- We call $\mathbf{v} \in \mathbf{V}$ a *partial setting* and $\mathbf{x} \in \mathbf{X}$ a *total setting*.
- The restriction of $\mathbf{x}$ to $\mathbf{V}$ is the projection $\mathrm{proj}(\mathbf{x}, \mathbf{V}) \in \mathrm{dom}[\mathbf{V}]$.
- The restriction $\mathrm{Rst}(\boldsymbol{M}_\iota)$ of an intervention $\iota = \mathrm{do}(\mathbf{V} = \mathbf{v})$ on a model $\boldsymbol{M}$ is the subset of total settings on $\mathbf{X}$ matching the partial setting $\mathbf{v}$.

$$\mathrm{Rst}(\boldsymbol{M}_\iota) = \{\mathbf{x} \in \mathrm{dom}[X] \mid \mathbf{v} = \mathrm{proj}(\mathbf{x}, \mathbf{V})\}$$

- We say that a total setting $\mathbf{x}$ is *compatible* with an intervention $\iota = \mathrm{do}(\mathbf{v})$ and we write $\mathrm{Cmp}(\mathbf{x}, \iota)$ if $\mathbf{x} \in \mathrm{Rst}(\boldsymbol{M}_\iota)$.

  Given a simple SCM which consists of three binary variables X, Y, Z and an intervention $\iota = \mathrm{do}(X = 0, Y = 1)$ then the total settings that are compatible with ι are $(0, 1, 0)$ and $(0, 1, 1)$.

Structural Causal Models: Interventions

Truncated Factorization

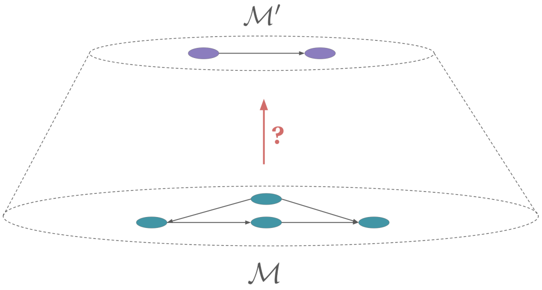
$$P(X_1, X_2, ..., X_n | \text{do}(\mathbf{S}{=}\mathbf{s})) = \prod_{X_i \notin S}^{n} P(X_i | \text{PA}_i), \ \forall i \text{ with } X_i \text{ not in } \mathbf{S}.$$

Pre/Post Interventional relation

$$P(X_1, X_2, ..., X_n | \text{do}(\mathbf{S}{=}\mathbf{s})) = \begin{cases} \frac{P(X_1, X_2, ..., X_n)}{P(S_i | \text{PA}_i)} & \text{if } \text{Cmp}(x, \text{do}(\mathbf{s})) \\ 0 & \text{otherwise} \end{cases}$$

Causal Abstractions

Causal Abstractions



Examples of applications:

- *complex physical systems* in which micro-level descriptors are abstracted into high-level statistics
- *social systems* where individual preferences and behaviours are coarsened into classes.

Causal Abstractions: Main works

- P. K. Rubenstein, S. Weichwald, S. Bongers, J. M. Mooij, D. Janzing, M. Grosse-Wentrup, and B. Schölkopf. *"Causal consistency of structural equation models"*, 2017.
- S. Beckers and J. Y. Halpern. "Abstracting causal models", 2019
- E. F. Rischel. *"The category theory of causal models"*, 2020
- S. Beckers, F. Eberhardt, and J. Y. Halpern. *"Approximate causal abstractions"*, 2020
- F. M. Zennaro, M. Drávucz, G. Apachitei, W. D. Widanage, and T. Damoulas. *"Jointly learning consistent causal abstractions over multiple interventional distributions"*, 2022
- J. Otsuka and H. Saigo. *"On the equivalence of causal models: A category-theoretic approach"*, 2022

Exact Transformations: Mapping SCMs

Let two SCMs $\boldsymbol{M} =< \mathbf{X}, \mathbf{U}, \mathscr{F}, \mathbb{P}(\mathbf{U}) >$ and $\boldsymbol{M} =< \mathbf{X}', \mathbf{U}', \mathscr{F}', \mathbb{P}(\mathcal{U}') >$ equipped with posets of interventions $\mathscr{I}, \mathscr{I}'$ respectively.

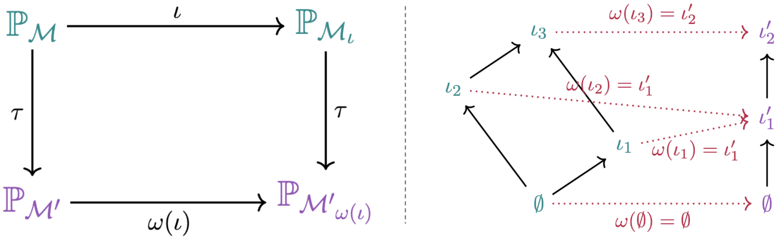
A function $\tau : \text{dom}[\mathbf{X}] \to \text{dom}[\mathbf{X}']$ is called an **exact** (τ, ω)-**transformation** of $\boldsymbol{M}$ to $\boldsymbol{M}'$ if there exists a surjective and order preserving map $\omega : \mathscr{I} \mapsto \mathscr{I}'$ such that:

$$\tau_{\#}(\mathbb{P}^{\iota}_{\boldsymbol{M}}(\mathbf{X})) = \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X}'), \ \ \forall \iota \in \mathscr{I}$$

A τ-ω transformation is a form of abstraction between causal models!

Exact Transformations: Consistency of mapping

Given a mapping $\omega : \mathscr{I} \mapsto \mathscr{I}'$ between the interventions of the low-level and the high-level model (right) then a transformation τ : dom[**X**] $\to$ dom[**X'**] is exact if the diagram on the left **commutes**:



Roughly speaking, if you start from the low-level model you can move up to the high-level one by following two distinct routes, either:

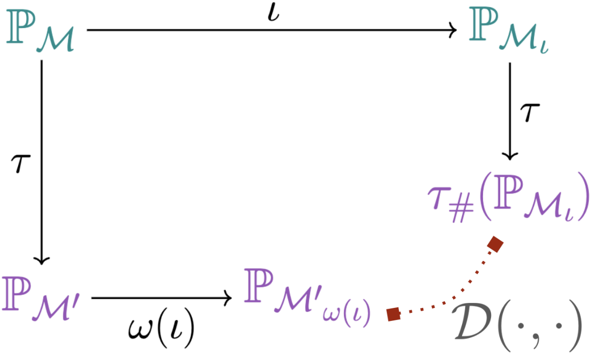
- **intervene** (ι) and then **transform** (τ), or
- **transform** (τ) and then **intervene** ($\omega(\iota)$)

Abstraction Error

Let τ be a τ-ω transformation between $\texttt{SCM}$ $\boldsymbol{M}$ and $\boldsymbol{M}'$ wrt $\mathscr{I}$ and ω. Given a discrepancy measure $\mathscr{D}$ between distributions, and a distribution q over the intervention set $\mathscr{I}$, we evaluate the approximation introduced by τ as the **abstraction error**:

$$e(\tau) = \mathbb{E}_{\iota \sim q}\left[\mathscr{D}\left(\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_\iota}),\ \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}\right)\right]$$

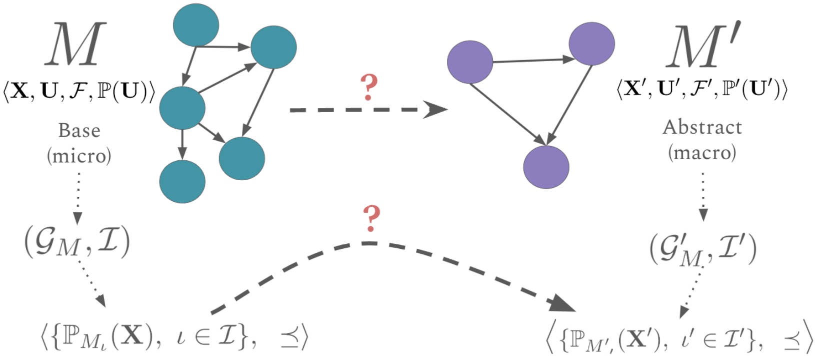
Abstraction Error



- We compute the distance between $\tau_{\#}(\mathbb{P}_{\mathbf{M}_\iota})$ and $\mathbb{P}_{\mathbf{M}'_{\omega(\iota)}}$ using $\mathscr{D}$.
- $\mathscr{D} = 0 \implies$ exact τ-ω abstraction.

Problem Statement

Problem Statement



Problem Statement

Given:

- Two DAGs: $\mathcal{G}_{\boldsymbol{M}}$ (base) and $\mathcal{G}_{\boldsymbol{M'}}$ (abstracted).
- The posets of interventions $(\mathcal{I}, \mathcal{I}')$ for both models.
- The mapping $\omega : \mathcal{I} \rightarrow \mathcal{I}'$.
- Samples, from the pre-interventional and post-interventional distributions for both models for all $\iota \in \mathcal{I}$.

Problem Statement

Given:

- Two DAGs: $\mathcal{G}_{\boldsymbol{M}}$ (base) and $\mathcal{G}_{\boldsymbol{M'}}$ (abstracted).
- The posets of interventions $(\mathcal{I}, \mathcal{I}')$ for both models.
- The mapping $\omega : \mathcal{I} \to \mathcal{I}'$.
- Samples, from the pre-interventional and post-interventional distributions for both models for all $\iota \in \mathcal{I}$.

Problem Statement

Given:

- Two DAGs: $\mathcal{G}_{\boldsymbol{M}}$ (base) and $\mathcal{G}_{\boldsymbol{M}'}$ (abstracted).
- The posets of interventions $(\mathcal{I}, \mathcal{I}')$ for both models.
- The mapping $\omega : \mathcal{I} \to \mathcal{I}'$.
- Samples, from the pre-interventional and post-interventional distributions for both models for all $\iota \in \mathcal{I}$.

Problem Statement

Given:

- Two DAGs: $\mathcal{G}_{\boldsymbol{M}}$ (base) and $\mathcal{G}_{\boldsymbol{M'}}$ (abstracted).
- The posets of interventions $(\mathcal{I}, \mathcal{I}')$ for both models.
- The mapping $\omega : \mathcal{I} \to \mathcal{I}'$.
- Samples, from the pre-interventional and post-interventional distributions for both models for all $\iota \in \mathcal{I}$.

Problem Statement

Given:

- Two DAGs: $\mathcal{G}_{\boldsymbol{M}}$ (base) and $\mathcal{G}_{\boldsymbol{M}'}$ (abstracted).
- The posets of interventions $(\mathcal{I}, \mathcal{I}')$ for both models.
- The mapping $\omega : \mathcal{I} \to \mathcal{I}'$.
- Samples, from the pre-interventional and post-interventional distributions for both models for all $\iota \in \mathcal{I}$.

We seek to learn an exact transformation $\tau : \text{dom}[\mathbf{X}] \to \text{dom}[\mathbf{X}']$ such that:

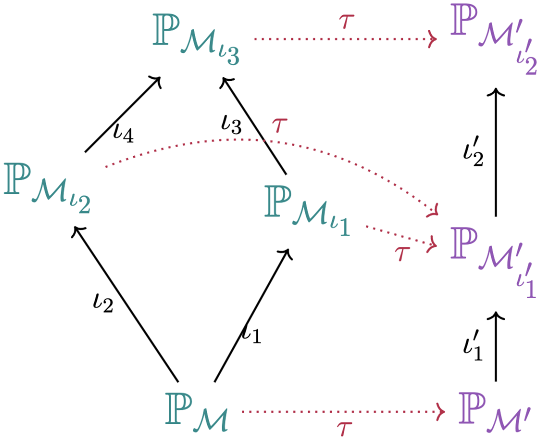
$$\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_{\iota}}(\mathbf{X})) = \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X}'), \;\; \forall \iota \in \mathcal{I}$$

Problem Statement

In other words, we seek to find a *single* function $\tau : \mathrm{dom}[\mathbf{X}] \rightarrow \mathrm{dom}[\mathbf{X}']$ such that:

$$
\begin{aligned}
\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_{\varnothing}}(\mathbf{X})) &= \mathbb{P}_{\boldsymbol{M}'_{\varnothing}}(\mathbf{X}') \\
\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_{\iota_1}}(\mathbf{X})) &= \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota_1)}}(\mathbf{X}') \\
\vdots \quad &= \quad \vdots \\
\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_{\iota_k}}(\mathbf{X})) &= \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota_k)}}(\mathbf{X}')
\end{aligned}
$$

Problem Statement

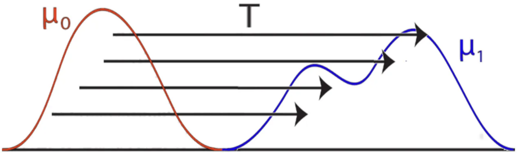


But we need a tool to learn such a map τ !

Optimal Transport

Optimal Transport

Optimal transport provides a general mathematical way of moving one distribution of mass to another as efficiently as possible. Specifically, by looking amongst the set of all possible ways to transport the mass from the one distribution to the other it selects the one which minimizes a cost function, evaluating the cost of moving the mass.



source: "Optimal Transport for Image Processing", Papadakis, 2017

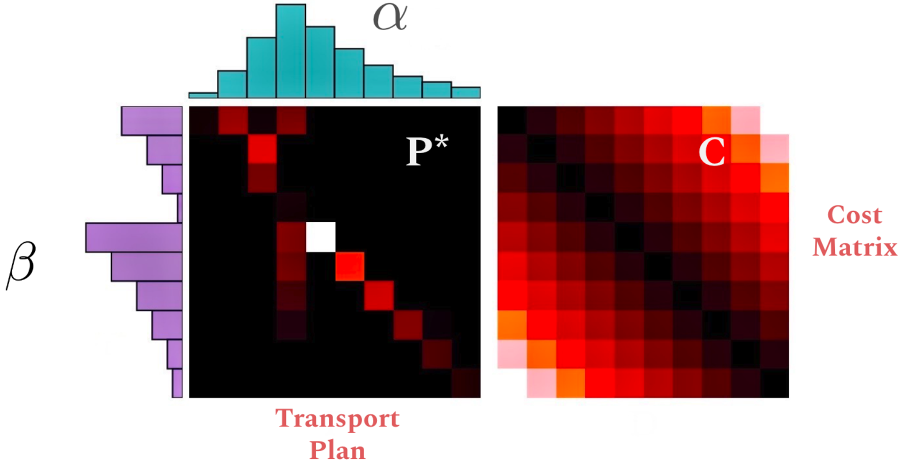
Optimal Transport

Consider $\mathcal{X} = \{x_i\}_{i=1}^{n} \subset \mathbb{R}^d$ and $\mathcal{Y} = \{y_j\}_{j=1}^{m} \subset \mathbb{R}^d$ with respective (probability) weights α, β. Thus, we have the discrete probability measures:

$$\alpha = \sum_{i=1}^{n} \alpha_i \delta_{x_i} \text{ and } \beta = \sum_{j=1}^{m} \beta_i \delta_{y_j}$$

Finally, assuming that the cost of transporting a unit of mass from x_i to y_j is $c(x_i, y_j)$ where $c : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ is the *cost function*, which induces the *cost matrix* $C_{ij} = c(x_i, y_j)$.

Optimal Transport



Optimal Transport

Kantorovic formulation

The (Entropic) Kantorovich problem for discrete measures solves the following optimization problem:

$$\mathrm{OT}_C^{\epsilon}(\alpha, \beta) = \min_{P \in \mathcal{U}(\alpha,\beta)} \left\{ \left\langle C, P \right\rangle - \epsilon \mathcal{H}(P) \right\}$$

$$= \min_{P \in \mathcal{U}(\alpha,\beta)} \left\{ \sum_{i,j} C_{ij} P_{ij} - \epsilon \mathcal{H}(P) \right\}$$

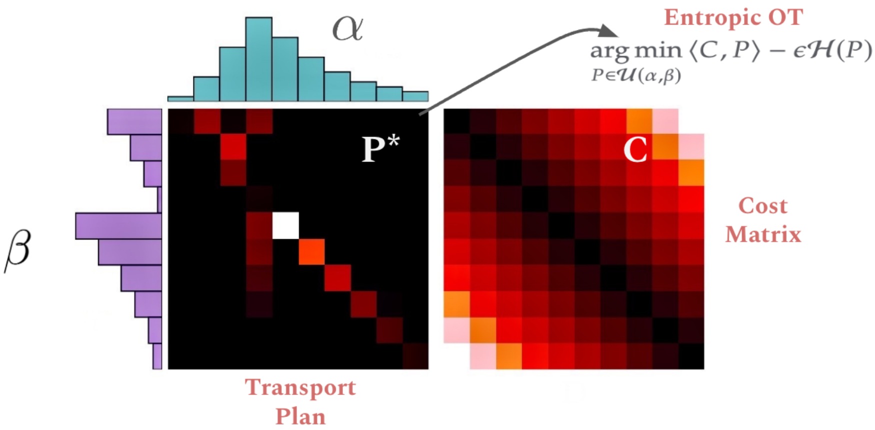
where the Frobenius inner product $\left\langle C, P \right\rangle$ gives the total transportation cost, $\mathcal{H}(P)$ is the discrete entropy of the coupling matrix P and $\mathcal{U}(\alpha, \beta)$ is the set of joint probability measures with marginals α and β which is a convex polytope, called the *transport polytope* or *coupling set*.

Optimal Transport

The transport polytope imposes the marginal constraints of the OT optimisation problem

$$\mathcal{U}(\alpha, \beta) = \left\{ P \in \mathbb{R}^{n \times m} : \sum_{j=1}^{m} P_{ij} = \alpha, \sum_{i=1}^{n} P_{ij} = \beta, \sum_{j=1}^{m} \sum_{i=1}^{n} P_{ij} = 1 \right\}$$

Optimal Transport



So, now we have a tool!

Problem Statement revised

We want to learn an exact τ-ω-transformation $\tau : \text{dom}[\mathbf{X}] \rightarrow \text{dom}[\mathbf{X}']$ s.t.:

$$\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_{\iota}}(\mathbf{X})) = \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X}), \ \ \forall \iota \in \mathcal{I}$$

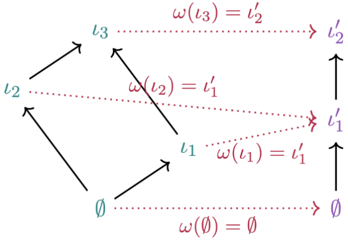
Problem Statement revised

Clearly, $\omega : \mathscr{I} \mapsto \mathscr{I}'$ induces a set of pairs between the distributions of $\boldsymbol{M}$ and $\boldsymbol{M}'$. We denote this as:

$$\Pi_\omega(\mathscr{I}) = \{\pi_i : i = 1, ..., |\mathscr{I}|\}$$

where $\forall \iota \in \mathscr{I} : \pi_\iota = (\pi_{\iota,s}, \pi_{\iota,t}) = \left(\widehat{\mathbb{P}}_{\boldsymbol{M}_\iota}(\mathbf{X}), \widehat{\mathbb{P}}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X})\right)$.

- $\pi_{i,s}$ expresses the base model's distribution of the i-th pair.
- $\pi_{i,t}$ expresses the abstracted model's distribution of the i-th pair.



Abstraction Learning as Multi-marginal OT

- We address the problem by viewing each pair π_ι as marginals in an Entropic OT problem within the Kantorovich formulation for discrete measures.

- We compute a plan P^ι for each pair π_ι, thereby leading to a *multi-marginal* optimization problem, made up of $|\Pi_\omega(\mathscr{I})|$ independent OT problems:

$$\mathbf{P}^\star = \mathtt{OT}_c(\Pi_\omega(\mathscr{I})) = \underset{\{P^\iota \in \mathcal{U}(\pi_\iota)\}_{\iota \in \mathscr{I}}}{\arg\min} \left\{ \sum_{\iota \in \mathscr{I}} \left\langle C, P^\iota \right\rangle - \epsilon \mathscr{H}(P^\iota) \right\}$$

where $\mathcal{U}(\pi_\iota)$ is the transport polytope of each pair π_ι.

Abstraction Learning as Multi-marginal OT

- We address the problem by viewing each pair π_ι as marginals in an Entropic OT problem within the Kantorovich formulation for discrete measures.
- We compute a plan P^ι for each pair π_ι, thereby leading to a *multi-marginal* optimization problem, made up of $|\Pi_\omega(\mathcal{I})|$ independent OT problems:

$$\mathbf{P}^\star = \mathtt{OT}_c(\Pi_\omega(\mathcal{I})) = \underset{\{P^\iota \in \mathcal{U}(\pi_\iota)\}_{\iota\in\mathcal{I}}}{\arg\min} \left\{ \sum_{\iota\in\mathcal{I}} \left\langle C, P^\iota \right\rangle - \epsilon\mathcal{H}(P^\iota) \right\}$$

where $\mathcal{U}(\pi_\iota)$ is the transport polytope of each pair π_ι.

Abstraction Learning as Multi-marginal OT

- Previously $\mathbf{P}^{\star}$ was a vector of $|\Pi_{\omega}(\mathscr{I})|$ optimal independent plans $P_{\star}^{\iota}$.
- Since we are looking for a single transformation τ, we aggregate those into a single average plan $\widehat{\mathscr{P}} = \frac{1}{|\mathbf{P}^{\star}|} \sum_{\iota \in \mathscr{I}} P_{\star}^{\iota}$, from which the map τ can be derived as a stochastic mapping $\tau = f_s(\widehat{\mathscr{P}})$ where $f_s : \mathrm{dom}[\mathrm{X}] \to \mathscr{A}^{|\mathrm{dom}[\mathrm{X}']|}$ and $\mathscr{A}^n = \{p \in \mathbb{R}^n, : \ p_i \geq 0, \sum_i p_i = 1\}$ the simplex in $\mathbb{R}^n$.
- The stochastic mapping converts the mass allocation, induced by $\widehat{\mathscr{P}}$, by assigning each base sample to a probability vector, depicting a distribution over the abstracted samples.

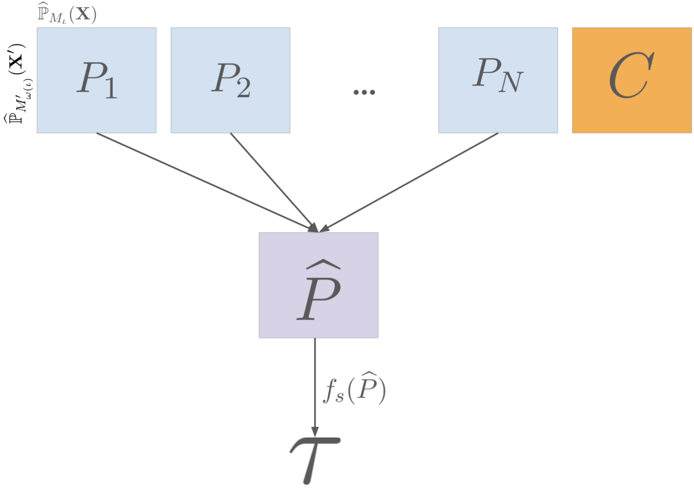
Abstraction Learning as Multi-marginal OT

- Previously $\mathbf{P}^\star$ was a vector of $|\Pi_\omega(\mathcal{I})|$ optimal independent plans $P^\iota_\star$.
- Since we are looking for a single transformation τ, we aggregate those into a single average plan $\widehat{\mathscr{P}} = \frac{1}{|\mathbf{P}^\star|} \sum_{\iota \in \mathcal{I}} P^\iota_\star$, from which the map τ can be derived as a stochastic mapping $\tau = f_s(\widehat{\mathscr{P}})$ where $f_s : \mathrm{dom}[\mathrm{X}] \to \mathcal{A}^{|\mathrm{dom}[\mathrm{X}']|}$ and $\mathcal{A}^n = \{p \in \mathbb{R}^n, :\; p_i \geq 0, \sum_i p_i = 1\}$ the simplex in $\mathbb{R}^n$.
- The stochastic mapping converts the mass allocation, induced by $\widehat{\mathscr{P}}$, by assigning each base sample to a probability vector, depicting a distribution over the abstracted samples.

Abstraction Learning as Multi-marginal OT

- Previously $\mathbf{P}^\star$ was a vector of $|\Pi_\omega(\mathcal{I})|$ optimal independent plans $P_\star^\iota$.
- Since we are looking for a single transformation τ, we aggregate those into a single average plan $\widehat{\mathscr{P}} = \frac{1}{|\mathbf{P}^\star|} \sum_{\iota \in \mathcal{I}} P_\star^\iota$, from which the map τ can be derived as a stochastic mapping $\tau = f_s(\widehat{\mathscr{P}})$ where $f_s : \mathrm{dom}[\mathrm{X}] \to \mathcal{A}^{|\mathrm{dom}[\mathrm{X}']|}$ and $\mathcal{A}^n = \{p \in \mathbb{R}^n, : \; p_i \geq 0, \sum_i p_i = 1\}$ the simplex in $\mathbb{R}^n$.
- The stochastic mapping converts the mass allocation, induced by $\widehat{\mathscr{P}}$, by assigning each base sample to a probability vector, depicting a distribution over the abstracted samples.

Abstraction Learning as Multi-marginal OT



Causal Optimal Transport of Abstractions

- The previous optimization problem is a collection of independent OT problems
- We incorporate causal knowledge by:
  - A causally informed cost function derived from the interventional information induced by the ω map.
  - Causal/*do-calculus* constraints linking the different transport plans.
- Thus, we transform the initial problem into a **joint** multi-marginal OT problem integrated with causal knowledge from different sources.

COTA

- The previous optimization problem is a collection of independent OT problems
- We incorporate causal knowledge by:
  - A causally informed cost function derived from the interventional information induced by the ω map.
  - Causal/*do-calculus* constraints linking the different transport plans.
- Thus, we transform the initial problem into a **joint** multi-marginal OT problem integrated with causal knowledge from different sources.

- The previous optimization problem is a collection of independent OT problems
- We incorporate causal knowledge by:
    - A causally informed cost function derived from the interventional information induced by the ω map.
    - Causal/*do-calculus* constraints linking the different transport plans.
- Thus, we transform the initial problem into a **joint** multi-marginal OT problem integrated with causal knowledge from different sources.

- The previous optimization problem is a collection of independent OT problems
- We incorporate causal knowledge by:
  - A causally informed cost function derived from the interventional information induced by the ω map.
  - Causal/*do-calculus* constraints linking the different transport plans.

- Thus, we transform the initial problem into a **joint** multi-marginal OT problem integrated with causal knowledge from different sources.

- The previous optimization problem is a collection of independent OT problems
- We incorporate causal knowledge by:
  - A causally informed cost function derived from the interventional information induced by the ω map.
  - Causal/*do-calculus* constraints linking the different transport plans.

- Thus, we transform the initial problem into a **joint** multi-marginal OT problem integrated with causal knowledge from different sources.

The ω-cost

- In order to compute a distance between samples $x \in \text{dom}[X]$ of the base and $x' \in \text{dom}[X']$ of the abstracted model, given interventions $\iota = \text{do}(a)$ and $\omega(\iota) = \text{do}(a')$, we exploit ω to discount the cost of transporting sample a to a'.
- We define $c_\omega : \text{dom}[X] \times \text{dom}[X'] \to \mathbb{R}_{\geq 0}$:

$$c_\omega(x, x') = |\mathcal{I}| - \sum_{\iota \in \mathcal{I}} \mathbf{1}\left[\text{Cmp}(x, \iota) \wedge \text{Cmp}(x', \omega(\iota))\right],$$

- c_ω discounts the cost of transporting the sample x to x' proportionally to the number of pairs $(\iota, \omega(\iota))$ w.r.t. which x and x' are compatible.

The ω-cost

- In order to compute a distance between samples $\mathsf{x} \in \mathrm{dom}[\mathsf{X}]$ of the base and $\mathsf{x}' \in \mathrm{dom}[\mathsf{X}']$ of the abstracted model, given interventions $\iota = \mathrm{do}(\mathsf{a})$ and $\omega(\iota) = \mathrm{do}(\mathsf{a}')$, we exploit ω to discount the cost of transporting sample a to a'.
- We define $c_\omega : \mathrm{dom}[\mathsf{X}] \times \mathrm{dom}[\mathsf{X}'] \to \mathbb{R}_{\geq 0}$:

$$c_\omega(\mathsf{x}, \mathsf{x}') = |\mathcal{I}| - \sum_{\iota \in \mathcal{I}} \mathbf{1}\left[\mathrm{Cmp}(\mathsf{x}, \iota) \wedge \mathrm{Cmp}(\mathsf{x}', \omega(\iota))\right],$$

- c_ω discounts the cost of transporting the sample x to x' proportionally to the number of pairs $(\iota, \omega(\iota))$ w.r.t. which x and x' are compatible.

The ω-cost

- In order to compute a distance between samples $\mathrm{x} \in \mathrm{dom}[\mathrm{X}]$ of the base and $\mathrm{x}' \in \mathrm{dom}[\mathrm{X}']$ of the abstracted model, given interventions $\iota = \mathrm{do}(\mathrm{a})$ and $\omega(\iota) = \mathrm{do}(\mathrm{a}')$, we exploit ω to discount the cost of transporting sample a to a'.
- We define $c_\omega : \mathrm{dom}[\mathrm{X}] \times \mathrm{dom}[\mathrm{X}'] \to \mathbb{R}_{\geq 0}$:

$$c_\omega(\mathrm{x}, \mathrm{x}') = |\mathscr{I}| - \sum_{\iota \in \mathscr{I}} \mathbf{1}\left[\mathrm{Cmp}(\mathrm{x}, \iota) \wedge \mathrm{Cmp}(\mathrm{x}', \omega(\iota))\right],$$

- c_ω discounts the cost of transporting the sample x to x' proportionally to the number of pairs $(\iota, \omega(\iota))$ w.r.t. which x and x' are compatible.

The ω-cost

$$c_\omega(\mathbf{x}, \mathbf{x}') = |\mathcal{I}| - \sum_{\iota \in \mathcal{I}} \mathbb{1}\left[\mathrm{Cmp}(\mathbf{x}, \iota) \wedge \mathrm{Cmp}(\mathbf{x}', \omega(\iota))\right]$$

	$\mathbf{x}_1 = 000$	$\mathbf{x}_2 = 001$	$\mathbf{x}_3 = 010$	$\mathbf{x}_4 = 011$	$\mathbf{x}_5 = 110$	$\mathbf{x}_6 = 110$
$\mathbf{x}'_1 = 00$	0	0	1	1	2	2
$\mathbf{x}'_2 = 01$	0	0	1	1	2	2
$\mathbf{x}'_3 = 10$	2	2	2	2	2	2
$\mathbf{x}'_4 = 11$	2	2	2	2	2	2

$\iota_1 : \mathrm{do}(\emptyset) \overset{\omega}{\mapsto} \mathrm{do}(\emptyset)$

$\iota_2 : \mathrm{do}(\mathbf{X}[1] = 0) \overset{\omega}{\mapsto} \mathrm{do}(\mathbf{X}'[1] = 0)$

$\iota_3 : \mathrm{do}(\mathbf{X}[1] = 0, \mathbf{X}[2] = 0) \overset{\omega}{\mapsto} \mathrm{do}(\mathbf{X}'[1] = 0)$

do-calculus constraints

Let $\iota = \text{do(a)}, \omega(\iota) = \text{do(a')}$ and $\eta = \text{do(b)}, \omega(\eta) = \text{do(b')}$, s.t. $\iota \preceq \eta$

The *mass conservation constraints* $\mathcal{U}(\pi_\iota)$ on P^ι induced by `OT` guarantee:

$$\overbrace{\widehat{\mathbb{P}}_{\boldsymbol{M}_\iota}(X_j) = \left(\sum_i P^\iota_{i,j}\right)_j}^{\text{Base}} \quad \forall j \in \text{dom[X]} \qquad \overbrace{\widehat{\mathbb{P}}_{\boldsymbol{M}'_{\omega(\iota)}}(X'_i) = \left(\sum_j P^\iota_{i,j}\right)_i}^{\text{Abstracted}} \quad \forall i \in \text{dom[X']}$$

do-calculus constraints

Without loss of generality, let π_ι be the pair of observational distributions, where $\iota, \omega(\iota)$ are the null interventions. Then, from the *truncated factorisation*, it holds that:

$$\mathbb{P}_{\boldsymbol{M}_{\mathrm{do(b)}}}(\mathrm{X}) = \left.\begin{cases} \frac{\mathbb{P}_{\boldsymbol{M}}(\mathrm{X})}{\prod_i \mathbb{P}_{\boldsymbol{M}}(\mathrm{B}_i = \mathrm{b}_i \mid \mathrm{PA}(\mathrm{B}_i))} & \text{if } \mathrm{Cmp}(\mathrm{x}, \mathrm{do(b)}) \\ 0 & \text{otherwise} \end{cases}\right\} \text{Base}$$

do-calculus constraints

In our empirical setup, we express this through the minimization of a statistical divergence $d : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}_{\geq 0}$, where D is $|\mathrm{dom}[\mathsf{X}]|$ for the base and $|\mathrm{dom}[\mathsf{X}']|$ for the abstracted model, as follows:

$$d\left(\widehat{\mathbb{P}}_{\boldsymbol{M}_{\mathrm{do(b)}}}(\mathsf{X}),\ \frac{1}{\prod_i \widehat{\mathbb{P}}_{\boldsymbol{M}}(\mathsf{B}_i = \mathsf{b}_i \ \mid \ \mathsf{PA}(\mathsf{B}_i))}\widehat{\mathbb{P}}_{\boldsymbol{M}}(\mathsf{X})\right) \quad \text{if } \mathrm{Cmp}(\mathsf{x}, \mathrm{do}(\mathsf{b}))$$

do-calculus constraints

Finally, we substitute in the mass conservation constraints for both the base and the abstracted models:

$$\left.\begin{aligned} \delta_{\iota,\eta}(P^{\iota}, P^{\eta}) &:= d\left(\left(\sum_i P^{\eta}_{i,j}\right)_j, \ \tfrac{1}{(\mathscr{Z}^{\eta})_j}\left(\sum_i P^{\iota}_{i,j}\right)_j\right) \quad \text{if } \operatorname{Cmp}(x_j, \eta). \\ \delta'_{\iota,\eta}(P^{\iota}, P^{\eta}) &:= d\left(\left(\sum_j P^{\eta}_{i,j}\right)_i, \ \tfrac{1}{(\mathscr{Z}^{\omega(\eta)})_i}\left(\sum_j P^{\iota}_{i,j}\right)_i\right) \quad \text{if } \operatorname{Cmp}(x'_i, \omega(\eta)). \end{aligned}\right.\begin{aligned} &\}\text{Base} \\ &\}\text{Abstracted} \end{aligned}$$

where $\mathscr{Z}^{\eta}, \mathscr{Z}^{\omega(\eta)}$ are the normalizing vectors for the base and the abstracted distributions respectively.

Instead of independently computing the `OT` plans we can jointly learn plans that preserve causal relations by incorporating the base and abstracted model distances $\mathscr{D}(P^{\iota}, P^{\eta}) = [\delta_{\iota,\eta}, \ \delta'_{\iota,\eta}]^{\top}$ defined over the marginals of two plans.

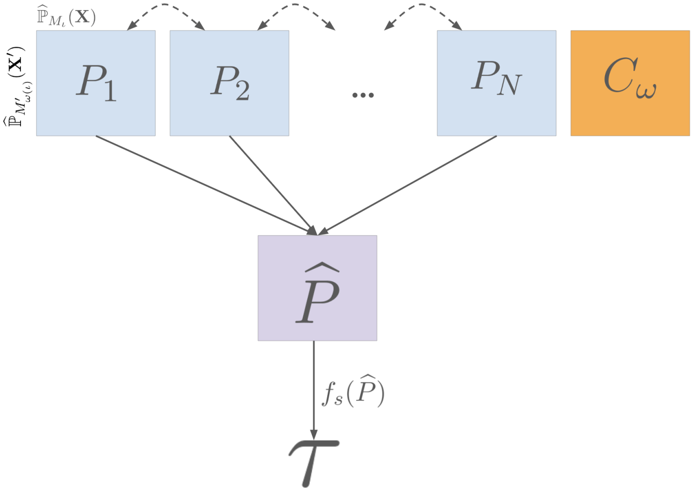
The COTA objective

For a given set of pairs $\Pi_\omega(\mathcal{I}) = \{\pi_{\iota_1}, \cdots \pi_{\iota_N} \mid \iota_n \in \mathcal{I}\}$, we define the objective function of COTA as the following OT problem:

$$P_k^\star = \texttt{COTA}_c\left(\Pi_\omega\left(\mathcal{I}\right)\right)$$

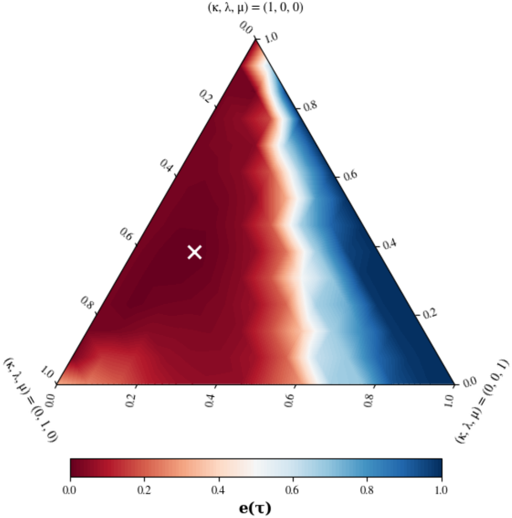
$$= \underset{\{P^{\iota_n} \in \mathcal{U}(\pi_{\iota_n})\}_{\iota_n \in \mathcal{I}}}{\arg\min} \left\{ \kappa \cdot \sum_{\iota_n \in \mathcal{I}} \underbrace{\langle C_\omega, P^{\iota_n} \rangle}_{\texttt{OT}} + \boldsymbol{\lambda}^\top \underbrace{\mathcal{D}(P^{\iota_n}, P^{\iota_{n+1}})}_{\text{do-calculus constraints}} - \mu \cdot \underbrace{\mathcal{H}(P^{\iota_n})}_{\text{entropy}} \right\}$$

COTA

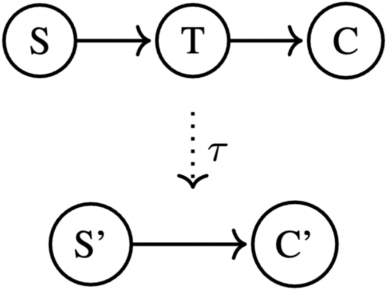

$\widehat{\mathbb{P}}_{M_\iota}(\mathbf{X})$
$\widehat{\mathbb{P}}_{M'_{\omega(\iota)}}(\mathbf{X}')$
P_1
P_2
...
P_N
C_ω
$\widehat{P}$
$f_s(\widehat{P})$
$\mathcal{T}$

Experimental Results

Results: The benefit of *do-calculus* constraints



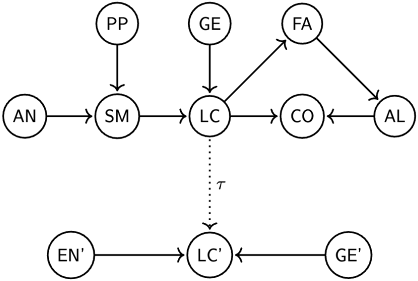
Results: Abstraction Error Evaluation



Synthetic: Simple Lung Cancer with "rich" intervention set

Method	$\mathcal{D}$	$\mathcal{C}$	$e_{\texttt{JSD}}(\tau)$	$e_{\texttt{WASS}}(\tau)$
COTA	FRO	c_ω	**0.010 ± 0.005**	**0.011 ± 0.003**
		$c_{\mathcal{H}}$	0.087 ± 0.007	0.025 ± 0.001
	JSD	c_ω	0.012 ± 0.006	0.012 ± 0.003
		$c_{\mathcal{H}}$	0.087 ± 0.006	0.025 ± 0.001
Pwise OT	–	c_ω	0.013 ± 0.002	0.011 ± 0.002
	–	$c_{\mathcal{H}}$	0.093 ± 0.004	0.039 ± 0.002
Map OT	–	c_ω	0.023 ± 0.022	0.147 ± 0.001
	–	$c_{\mathcal{H}}$	0.169 ± 0.022	0.156 ± 0.001
Bary OT	–	c_ω	0.233 ± 0.142	0.067 ± 0.042
	–	$c_{\mathcal{H}}$	0.323 ± 0.074	0.095 ± 0.039

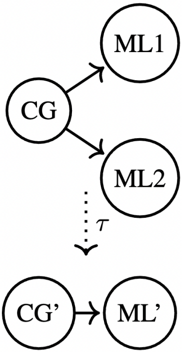
Results: Abstraction Error Evaluation



Synthetic: LUng CAncer Set (LUCAS)

Method	$\mathcal{D}$	$\mathcal{C}$	$e_{\texttt{JSD}}(\tau)$	$e_{\texttt{WASS}}(\tau)$
COTA	FRO	c_ω	$\mathbf{0.263 \pm 0.005}$	0.061 ± 0.001
		$c_\mathcal{H}$	$\mathbf{0.263 \pm 0.006}$	0.061 ± 0.001
Pwise OT	–	c_ω	0.306 ± 0.009	0.045 ± 0.001
	–	$c_\mathcal{H}$	0.387 ± 0.002	0.047 ± 0.001
Map OT	–	c_ω	0.294 ± 0.008	0.054 ± 0.001
	–	$c_\mathcal{H}$	0.350 ± 0.005	0.054 ± 0.001
Bary OT	–	c_ω	0.294 ± 0.047	0.044 ± 0.003
	–	$c_\mathcal{H}$	0.414 ± 0.040	0.046 ± 0.010

Results: COTA as a data augmentation tool



Real-world data: Electric Battery Manufacturing

Training set	Test set	Zennaro et al. [2023]	COTA
LRCS[$CG \neq k$]	LRCS[$CG = k$]	1.86 ± 1.75	$\mathbf{1.40 \pm 1.39}$
LRCS[$CG \neq k$] +WMG	LRCS[$CG = k$]	0.22 ± 0.26	$\mathbf{0.13 \pm 0.07}$
LRCS[$CG \neq k$] +WMG[$CG \neq k$]	LRCS[$CG = k$] WMG[$CG = k$]	1.22 ± 0.95	$\mathbf{0.85 \pm 0.81}$

Summary

Summary

- We wanted to learn a map between causal models $(\boldsymbol{M}, \boldsymbol{M}')$ that describe the same system at different levels of abstraction;
- Learn an exact transformation $\tau : \text{dom}[\mathbf{X}] \to \text{dom}[\mathbf{X}']$ s.t.:

$$\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_\iota}(\mathbf{X})) = \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X}'), \quad \forall \iota \in \mathcal{I}$$

- We addressed the problem by viewing each pair π_ι induced by the ω map of the τ-ω framework as marginals in an Entropic OT problem for discrete measures;
- We incorporated causal knowledge into this OT problem by defining:
  - An interventionally-informed cost function c_ω.
  - Causal constraints linking the different transport plans.
- We showed the superiority of COTA against non causal baselines and prior state-of-the-art.

Summary

- We wanted to learn a map between causal models $(\boldsymbol{M}, \boldsymbol{M}')$ that describe the same system at different levels of abstraction;
- Learn an exact transformation $\tau : \text{dom}[\mathbf{X}] \to \text{dom}[\mathbf{X}']$ s.t.:

$$\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_\iota}(\mathbf{X})) = \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X}'), \quad \forall \iota \in \mathcal{I}$$

- We addressed the problem by viewing each pair π_ι induced by the ω map of the τ-ω framework as marginals in an Entropic `OT` problem for discrete measures;
- We incorporated causal knowledge into this `OT` problem by defining:
  - An interventionally-informed cost function c_ω.
  - Causal constraints linking the different transport plans.
- We showed the superiority of `COTA` against non causal baselines and prior state-of-the-art.

Summary

- We wanted to learn a map between causal models $(\boldsymbol{M}, \boldsymbol{M}')$ that describe the same system at different levels of abstraction;
- Learn an exact transformation $\tau : \mathrm{dom}[\mathbf{X}] \to \mathrm{dom}[\mathbf{X}']$ s.t.:

$$\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_\iota}(\mathbf{X})) = \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X}'), \quad \forall \iota \in \mathcal{I}$$

- We addressed the problem by viewing each pair π_ι induced by the ω map of the τ-ω framework as marginals in an Entropic `OT` problem for discrete measures;
- We incorporated causal knowledge into this `OT` problem by defining:
  - An interventionally-informed cost function c_ω.
  - Causal constraints linking the different transport plans.
- We showed the superiority of `COTA` against non causal baselines and prior state-of-the-art.

Summary

- We wanted to learn a map between causal models $(\boldsymbol{M}, \boldsymbol{M}')$ that describe the same system at different levels of abstraction;
- Learn an exact transformation $\tau : \text{dom}[\mathbf{X}] \rightarrow \text{dom}[\mathbf{X}']$ s.t.:

$$\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_\iota}(\mathbf{X})) = \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X}'), \quad \forall \iota \in \mathcal{I}$$

- We addressed the problem by viewing each pair π_ι induced by the ω map of the τ-ω framework as marginals in an Entropic `OT` problem for discrete measures;
- We incorporated causal knowledge into this `OT` problem by defining:
  - An interventionally-informed cost function c_ω.
  - Causal constraints linking the different transport plans.
- We showed the superiority of `COTA` against non causal baselines and prior state-of-the-art.

Summary

- We wanted to learn a map between causal models $(\boldsymbol{M}, \boldsymbol{M}')$ that describe the same system at different levels of abstraction;
- Learn an exact transformation $\tau : \text{dom}[\mathbf{X}] \to \text{dom}[\mathbf{X}']$ s.t.:

$$\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_\iota}(\mathbf{X})) = \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X}'), \quad \forall \iota \in \mathcal{I}$$

- We addressed the problem by viewing each pair π_ι induced by the ω map of the τ-ω framework as marginals in an Entropic `OT` problem for discrete measures;
- We incorporated causal knowledge into this `OT` problem by defining:
  - An interventionally-informed cost function c_ω.
  - Causal constraints linking the different transport plans.
- We showed the superiority of `COTA` against non causal baselines and prior state-of-the-art.

Summary

- We wanted to learn a map between causal models $(\boldsymbol{M}, \boldsymbol{M}')$ that describe the same system at different levels of abstraction;
- Learn an exact transformation $\tau : \mathrm{dom}[\mathbf{X}] \to \mathrm{dom}[\mathbf{X}']$ s.t.:

$$\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_\iota}(\mathbf{X})) = \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X}'), \quad \forall \iota \in \mathcal{I}$$

- We addressed the problem by viewing each pair π_ι induced by the ω map of the τ-ω framework as marginals in an Entropic `OT` problem for discrete measures;
- We incorporated causal knowledge into this `OT` problem by defining:
  - An interventionally-informed cost function c_ω.
  - Causal constraints linking the different transport plans.
- We showed the superiority of `COTA` against non causal baselines and prior state-of-the-art.

Summary

- We wanted to learn a map between causal models $(\boldsymbol{M}, \boldsymbol{M}')$ that describe the same system at different levels of abstraction;
- Learn an exact transformation $\tau : \mathrm{dom}[\mathbf{X}] \to \mathrm{dom}[\mathbf{X}']$ s.t.:

$$\tau_{\#}(\mathbb{P}_{\boldsymbol{M}_\iota}(\mathbf{X})) = \mathbb{P}_{\boldsymbol{M}'_{\omega(\iota)}}(\mathbf{X}'), \quad \forall \iota \in \mathcal{I}$$

- We addressed the problem by viewing each pair π_ι induced by the ω map of the τ-ω framework as marginals in an Entropic OT problem for discrete measures;
- We incorporated causal knowledge into this OT problem by defining:
  - An interventionally-informed cost function c_ω.
  - Causal constraints linking the different transport plans.
- We showed the superiority of COTA against non causal baselines and prior state-of-the-art.

Future Work

- Extend to continuous settings;
- Lift the causal sufficiency assumption;
- Further theoretical guarantees for the existence/uniqueness of the estimated map, especially from the OT perspective.

Future Work

- Extend to continuous settings;
- Lift the causal sufficiency assumption;
- Further theoretical guarantees for the existence/uniqueness of the estimated map, especially from the OT perspective.

Future Work

- Extend to continuous settings;
- Lift the causal sufficiency assumption;
- Further theoretical guarantees for the existence/uniqueness of the estimated map, especially from the OT perspective.

Thank you!

Paper