Triangular Monotonic Generative Models Can Perform Causal Discovery

Johnny Xi, Sebastian Gonzalez and Benjamin Bloem-Reddy
Department of Statistics, UBC

Triangular Monotonic (TMI) Maps

$$T(x) = \begin{bmatrix} T_1(x_1) \\ T_2(x_1, x_2) \\ \vdots \\ T_d(x_1, \ldots, x_d) \end{bmatrix}$$

$T_1(x_1)$ ← Monotone

$T_2(x_1, x_2)$ ← Monotone in x_2, for all x_1

$P_\epsilon = \mathcal{U}[0,1]^d$ $\quad T_\# P_x = P_\epsilon$, full support ★

T_i is the cond. CDF $T_i(x_{\le i}) = F_{X_i | X_{<i}}(x_i \mid x_{<i})$

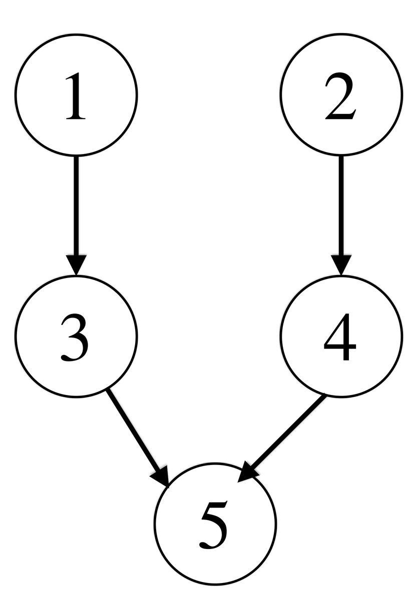
Sparsity of T_i implies conditional independence

$$F_{X_i | X_{<i}} = F_{X | X_{<i} \setminus X_j} \iff X_i \perp\!\!\!\perp X_j \mid X_{<i} \setminus X_j \quad ★$$

Still holds if P_ϵ is merely independent

CI-based Structure Learning

Given DAG $\mathcal{G}$ and faithful P_x :

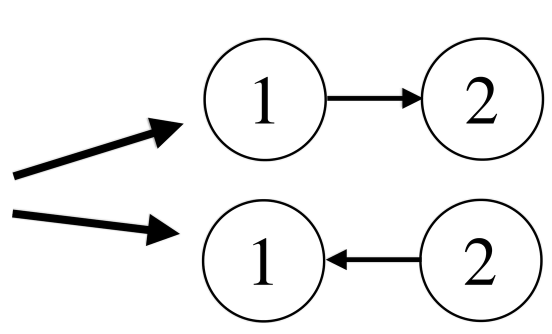


$I_1 \perp\!\!\!\perp_{\mathcal{G}} I_2 \mid I_3$ (d-sep)

$\iff X_1 \perp\!\!\!\perp X_2 \mid X_3$ (CI)

Recover $\mathcal{G}$ with $\{X_n\} \sim P_x$ using CI signatures

Only id. up to MEC: $X_1 \not\perp\!\!\!\perp X_2$



Use a causal generative model (SCM) for P_x:

$x_i = f(x_{pa(i)}, \epsilon_i)$, parents from $\mathcal{G}$

Asymmetries in f can refine the MEC

TMI Maps for Causal Discovery

TMI maps combine structure learning + SCM!

1) Permuting TMI maps can identify the MEC:

Theorem (Raskutti + Uhler 2018): The sparsest permutation of X, in terms of the number of CI's discovered in the sequence ★, identifies the MEC of $\mathcal{G}$.

I.e., finding π with the sparsest Jacobian gives adjacency matrix $J_{T,\pi} \overset{Markov}{\sim} \mathcal{G}$

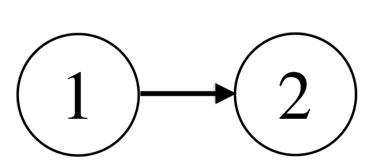
Novelty: don't need to assume an underlying SCM to do structure learning with Jacobian

2) TMIs can also fit SCMs:

If $x_i = f(x_{pa(i)}, \epsilon_i)$ monotone in ϵ_i, can write

$(x_1, \ldots, x_d) = T(\epsilon_1, \ldots, \epsilon_d)$ (in practice, fit the abductive T^{-1}, which is also TMI)

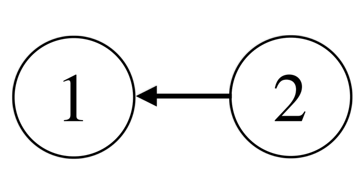
TMIs infer latent variables (★) if order correct



: $\hat{T}(x_1, x_2) = T^{-1}(x_1, x_2) = (\epsilon_1, \epsilon_2)$

$\implies X_1 \perp\!\!\!\perp \epsilon_2$

Incorrect order refines the MEC:



: true SCM not in model class!

$$\hat{T}(x_1, x_2) = \begin{bmatrix} \hat{T}_1(x_1) \\ \hat{T}_2(x_1, x_2) \end{bmatrix} = (\hat{\epsilon}_1, \hat{\epsilon}_2) \implies X_1 \not\perp\!\!\!\perp \hat{\epsilon}_2$$

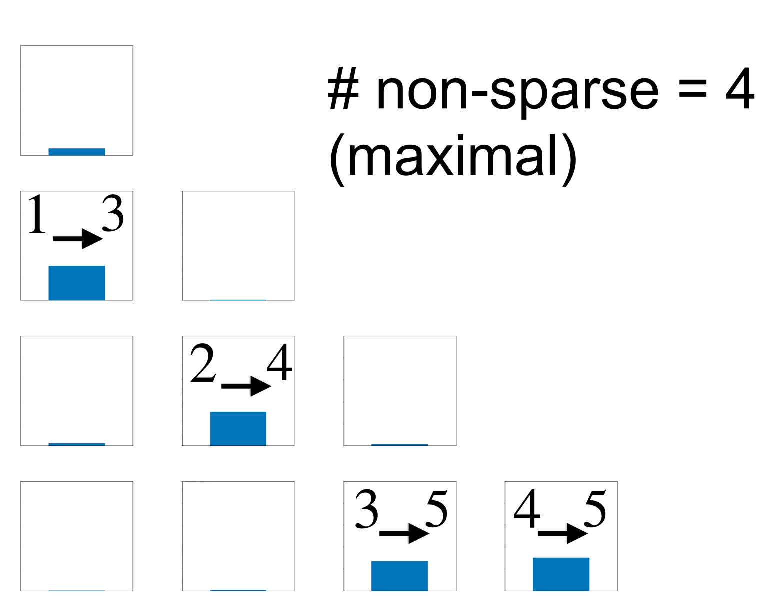
Novelty: TMI SCMs are less restrictive than the additive noise models typically used for CD

Computational Challenges Ahead

Hard computational problems:

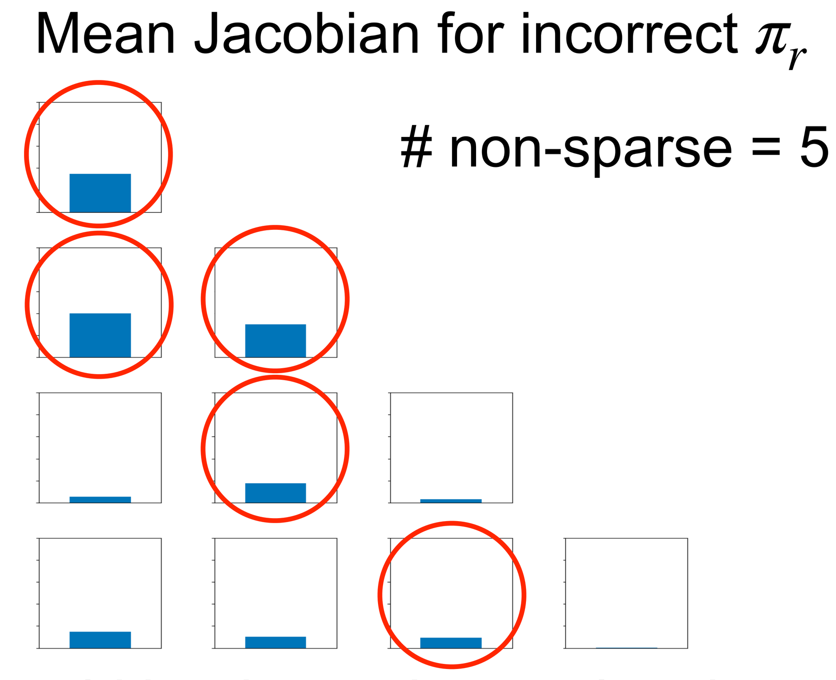
- Search over permutations 🫠
- How to threshold for sparsity 😵‍💫
- Doing it for each permutation 😵

Mean Jacobian for true π^*



Clear sparsity pattern

Mean Jacobian for incorrect π_r



○ should be dense, but unclear in practice!

THE UNIVERSITY OF BRITISH COLUMBIA

NSERC CRSNG

Raskutti, G., & Uhler, C. (2018). Learning directed acyclic graph models based on sparsest permutations. *Stat*, *7*(1).