Differentiable Cyclic Causal Discovery Under Unmeasured Confounders

Muralikrishnna G. Sethuraman, Faramarz Fekri
Georgia Institute of Technology

Motivation

- Causal understanding of real-world systems is crucial for prediction under unseen perturbations
- With a few notable exceptions, most existing work rely on the following assumptions which are often violated in practice:
  Acyclicity: no directed cycles
  Causal sufficiency: no unmeasured confounders
- Assumptions simplify search space; Often unrealistic in practice
  Economics: Supply→Demand, and Demand→Supply
  Biology: Several biological systems (like gene regulatory) are known to exhibit *feedback loops, and unmeasured variables*
- We propose DCCD-CONF: a novel differentiable causal discovery framework that handles *feedback loops, nonlinearity, and hidden confounding*

Problem Setup

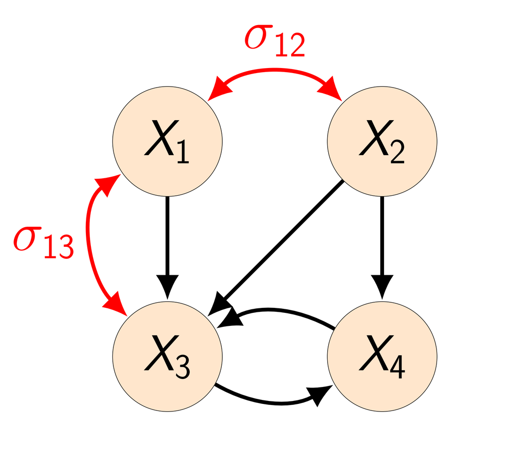
Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ represent a cyclic *Directed Mixed Graph* (DMG)

Structural Equations Model (SEM)

Let Z_i's denote the exogenous variables. Then,

$$X_i = F_i(\mathbf{X}_{\mathrm{pa}_{\mathcal{G}}(i)}, Z_i), \quad i = 1, \ldots, d.$$

Let $\mathbf{Z} = (Z_1, \ldots, Z_d) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_Z)$, $(\Sigma_Z)_{ij} \neq 0 \Rightarrow i \leftrightarrow j$.

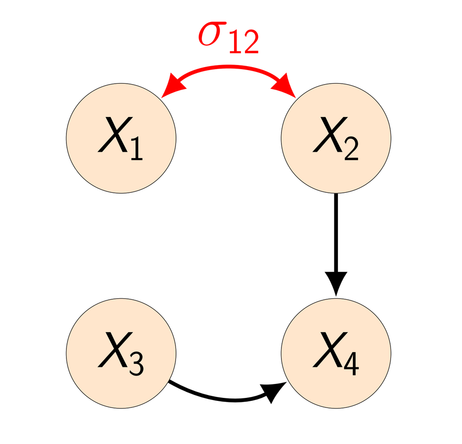


(a) Causal Graph $\mathcal{G}$

Vectorization: $\mathbf{X} = \mathbf{F}(\mathbf{X}, \mathbf{Z})$

Interventions

Hard interventions. All incoming edges to intervened nodes are removed. $\mathbf{U} \in \mathbb{R}^{d \times d}$ interventional mask matrix, then: $\mathbf{X} = \mathbf{U}\mathbf{F}(\mathbf{X}, \mathbf{Z}) + \mathbf{C}$



(b) Intervention on X_3

Let $\mathbf{f}_x^{(I)} : \mathbf{X} \mapsto \mathbf{Z}$ be *forward map* under intervention, $\mathcal{U} = \mathcal{V} \setminus I$. The data likelihood is given by

$$p_{\mathrm{do}(I)(\mathcal{G})}(\mathbf{X}) = p_I(\mathbf{C}) p_Z\left(\left[\mathbf{f}_x^{(I)}(\mathbf{X})\right]_{\mathcal{U}}\right) \left| \det\left(\mathbf{J}_{\mathbf{f}_x^{(I)}}(\mathbf{X})\right) \right|$$

Likelihood depends on *determinant of Jacobian*.

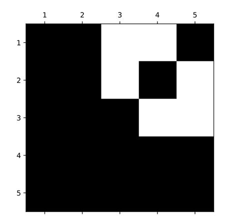
DCCD-CONF

Goal. Given interventions $\mathcal{I} = \{I_k\}_{k \in [K]}$, we would like to learn the SEM by maximizing **regularized log-data likelihood**:

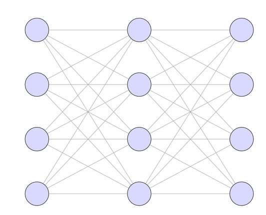
$$\mathcal{S}_{\mathcal{I}}(\mathcal{G}) := \sup_{\theta, \mathbf{\Sigma}_Z} \sum_{k=1}^{K} \mathop{\mathbb{E}}_{\mathbf{X} \sim p^{(k)}} \log p_{\mathrm{do}(I_k)(\mathcal{G})}(\mathbf{X}) - \lambda \|\mathcal{G}\|_1$$

Challenge 1. Modeling Causal Mechanism

Causal Mechanism: $\mathbf{F}(\mathbf{x}, \mathbf{z}) = -\mathbf{g}_x(\mathbf{x}) + \mathbf{g}_z(\mathbf{z}) + \mathbf{z}$ (allows nonlinear interaction between $\mathbf{x}$ and $\mathbf{z}$)

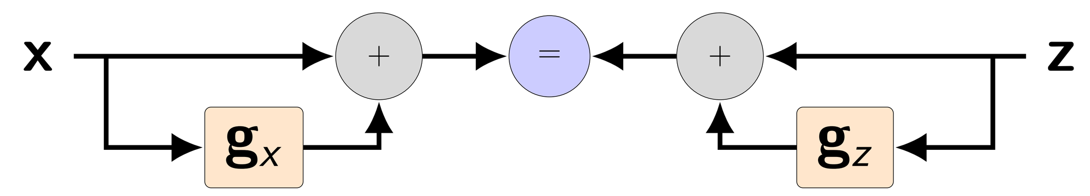


Dependency mask for graph adjacency
(Can model directed cycles)



$\mathbf{g}_x, \mathbf{g}_z$ are modeled using *contractive NN*

The SEM forms an *implicit layer*



Challenge 2. Computing Log-det-Jacobian

Power series expansion of log-det-Jacobian utilizing

$$\log \left| \det\left(\mathbf{J}_{\mathbf{f}_x^{(I_k)}}(\mathbf{X})\right) \right| = \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \left[\mathrm{Tr}\{\mathbf{J}^m_{\mathbf{U}_k \mathbf{g}_x}(\mathbf{X})\} - \mathrm{Tr}\{\mathbf{J}^m_{\mathbf{U}_k \mathbf{g}_z}(\mathbf{Z})\} \right]$$

Reduces gradient calls from $\mathcal{O}(d^3)$ to $\mathcal{O}(d^2)$.

Hutchinson Trace Estimator (even more reduction)

$$\mathrm{Tr}\{\mathbf{A}\} = \mathbb{E}_{\mathbf{W}}[\mathbf{W}^\top \mathbf{A} \mathbf{W}], \quad \mathbb{E}[\mathbf{W}] = 0, \ \mathbb{E}[\mathbf{W}^2] = 1$$

Challenge 3. Parameter Update

NN and graph parameters: Using *implicit function theorem*, gradients can be efficiently backpropagated

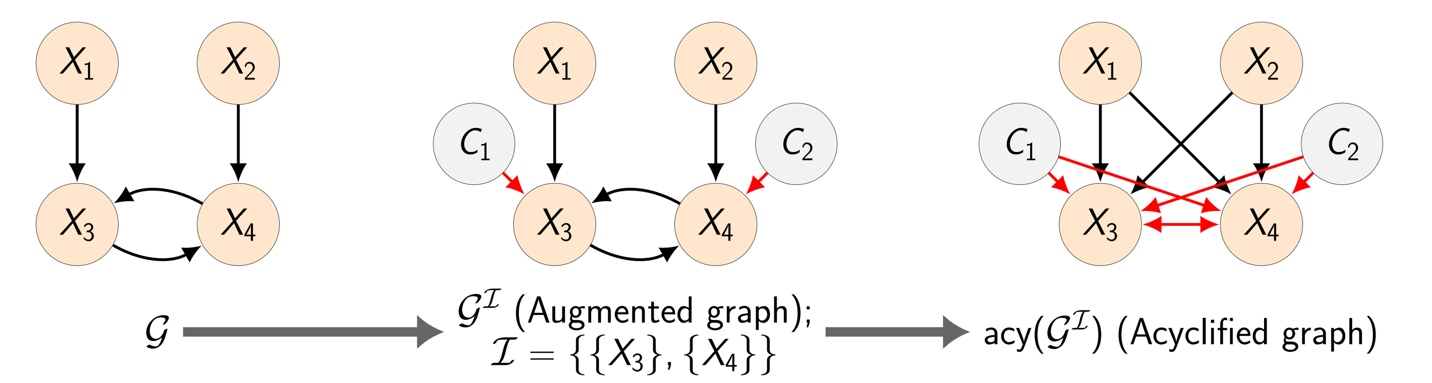
Exogenous noise covariance: Let $\mathbf{z}^{(i)} = \mathbf{f}_x(\mathbf{x}^{(i)})$, and $\mathbf{S}$ be the sample covariance of $\mathbf{Z}$. $\mathbf{\Sigma}_Z$ is obtained by solving the following convex problem:

$$\tilde{\mathcal{L}}(I_k) = \sup_{\mathbf{\Sigma}_Z} -\mathrm{Tr}(\mathbf{S}\mathbf{\Sigma}_Z^{-1}) - \log |\mathbf{\Sigma}_Z|$$

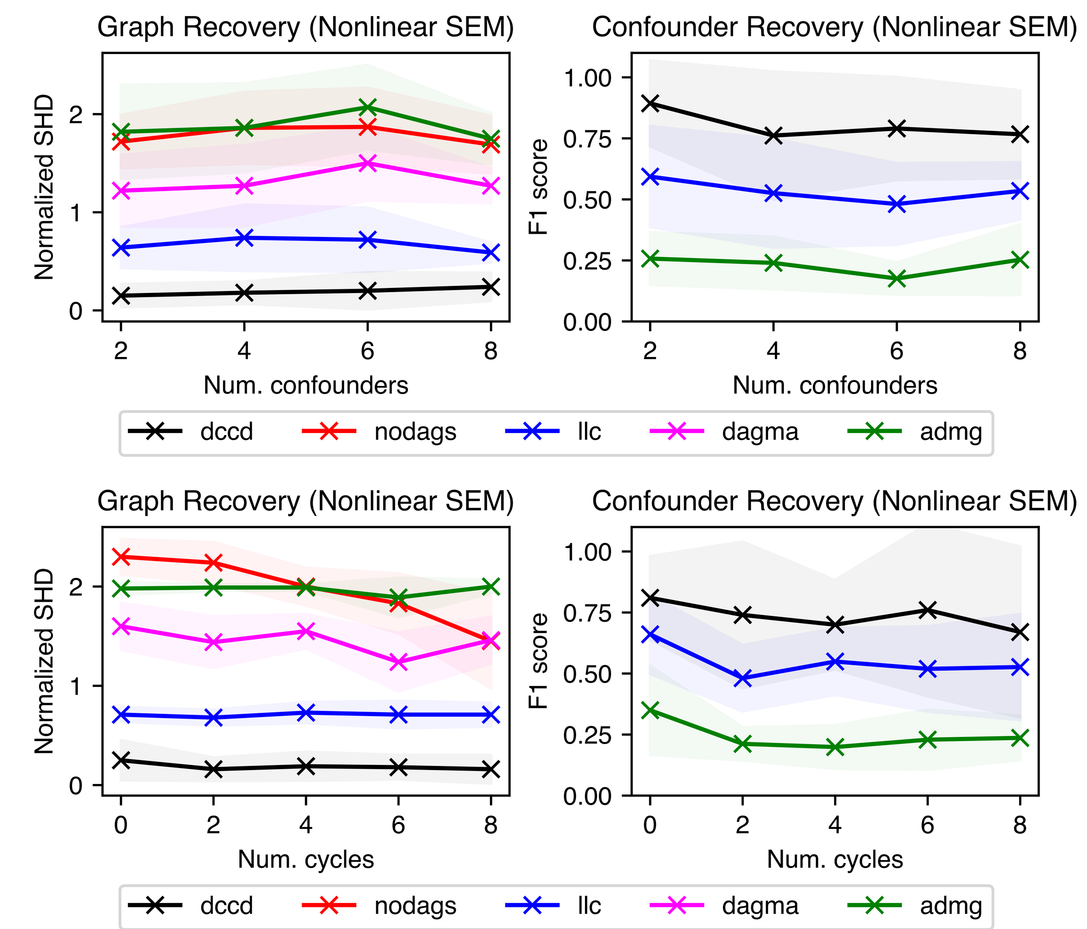
Results

Theorem. Let $\mathcal{I}$ be a family of interventions, let $\mathcal{G}^*$ denote the ground truth DMG, and $\hat{\mathcal{G}} := \arg\max_{\mathcal{G}} \mathcal{S}(\mathcal{G})$. Then, under suitable assumptions and $\lambda > 0$, we have that $\hat{\mathcal{G}} \equiv_{\mathcal{I}} \mathcal{G}^*$.

The proof of the theorem relies on the following transformations:



Ablation Study



Performance comparison between DCCD-CONF and baselines by varying number of confounders (top plot) and the number of cycles in the graph (bottom plot).

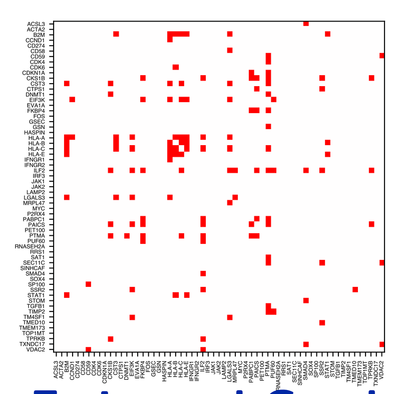
Graph recovery w.r.t nSHD (left, lower the better), and confounder recovery w.r.t F1-score (right, higher the better)

Gene Regulatory Network

Contains gene expressions taken from gene knockout experiments on over 200,000 melanoma cells. Choose 61 genes from around 20,000.

Performance comparison with respect to NLL
(Lower the better)

Method	Control	Co-Culture	IFN-γ
DCCD-CONF	**1.375** (0.103)	**1.245** (0.039)	**1.235** (0.338)
NODAGS	1.465 (0.015)	1.406 (0.012)	1.504 (0.009)
LLC	1.385 (0.039)	1.325 (0.029)	1.430 (0.048)
DCDI	1.523 (0.036)	1.367 (0.018)	1.517 (0.041)



Estimated Graph

DCCD-CONF outperforms the baselines showcasing the importance of handling confounders and cycles in practice.

Acknowledgment

This work was supported by funding from the National Science Foundation.