Differentiable Cyclic Causal Discovery Under Unmeasured Confounders

Murali G Sethuraman   Faramarz Fekri

School of Electrical & Computer Engineering
Georgia Institute of Technology

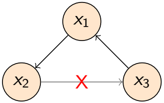
Based on work supported by NSF

Motivation

- Causal understanding of real-world systems is crucial for prediction under unseen perturbations

Motivation

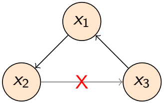
- Causal understanding of real-world systems is crucial for prediction under unseen perturbations
- With a few notable exceptions, most existing work rely on the following assumptions which are often violated in practice:



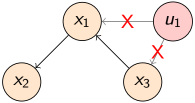
Acyclicity: No directed cycles

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:



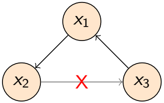
Acyclicity: No directed cycles



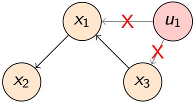
Causal sufficiency: No unmeasured confounders

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:

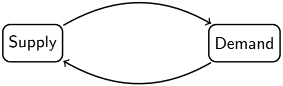


Acyclicity: No directed cycles



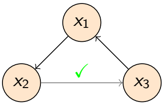
Causal sufficiency: No unmeasured confounders

- Assumptions simplify search space; Often unrealistic in practice

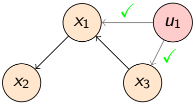


Contributions

- **DCCD-CONF**: novel differentiable causal discovery framework that handles *feedback loops, nonlinearity, and hidden confounding*



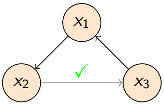
Feedback loops



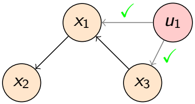
Hidden confounders

Contributions

- **DCCD-CONF**: novel differentiable causal discovery framework that handles *feedback loops, nonlinearity, and hidden confounding*

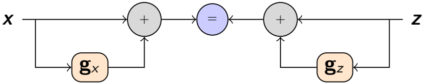


Feedback loops



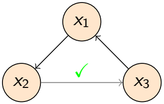
Hidden confounders

- DCCD-CONF performs maximum likelihood-based graph recovery utilizing *implicit normalizing flows*

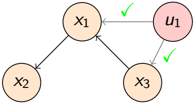


Contributions

- **DCCD-CONF**: novel differentiable causal discovery framework that handles *feedback loops, nonlinearity, and hidden confounding*

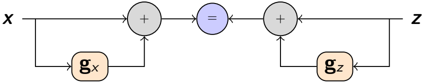


Feedback loops



Hidden confounders

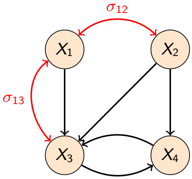
- DCCD-CONF performs maximum likelihood-based graph recovery utilizing *implicit normalizing flows*



- We show *consistency* in infinite sample regime, and showcase its practical use through synthetic and real-world experiments

Problem setup

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



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

Problem setup

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

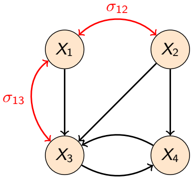
Structural Equations Model

Let Z_i's denote the exogenous variables. Then,

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

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

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



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

Problem setup

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

Structural Equations Model

Let Z_i's denote the exogenous variables. Then,

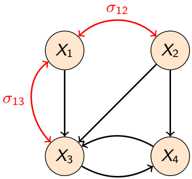
$$X_i = F_i(\boldsymbol{X}_{\mathrm{pa}_{\mathcal{G}}(i)}, Z_i), \quad i = 1, \ldots, d.$$

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

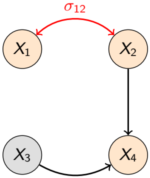
Vectorization: $\boldsymbol{X} = \mathbf{F}(\boldsymbol{X}, \boldsymbol{Z})$

Interventions

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



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



(b) Intervention on X_3

Problem setup

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

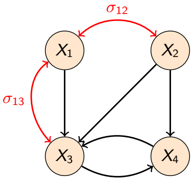
Structural Equations Model

Let Z_i's denote the exogenous variables. Then,

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

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

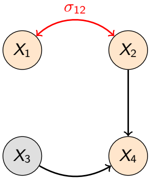
Vectorization: $\boldsymbol{X} = \mathbf{F}(\boldsymbol{X}, \boldsymbol{Z})$



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

Interventions

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



(b) Intervention on X_3

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

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

DCCD-CONF: Diff. Cyclic Causal Discovery Under Confounders

Objective Function

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

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

Three main challenges:

1. Modeling the causal mechanism
2. Computing Log-determinant of the Jacobian
3. Updating the model parameters

DCCD-CONF: Diff. Cyclic Causal Discovery Under Confounders

Objective Function

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

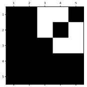
$$\mathcal{S}_{\mathcal{I}}(\mathcal{G}) := \sup_{\boldsymbol{\theta},\boldsymbol{\Sigma}_Z} \sum_{k=1}^{K} \mathop{\mathbb{E}}_{\boldsymbol{X}\sim p^{(k)}} \log p_{\mathrm{do}(I_k)(\mathcal{G})}(\boldsymbol{X}) - \lambda\|\mathcal{G}\|_1.$$

Three main challenges:

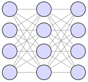
1. Modeling the causal mechanism
2. Computing Log-determinant of the Jacobian
3. Updating the model parameters

Modeling Causal Mechanism

Causal Mechanism: $\mathbf{F}(\boldsymbol{x},\boldsymbol{z}) = -\mathbf{g}_x(\boldsymbol{x}) + \mathbf{g}_z(\boldsymbol{z}) + \boldsymbol{z}$

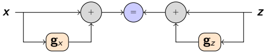


Dependency mask for graph adjacency



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

The SEM forms an *implicit layer*



DCCD-CONF: Diff. Cyclic Causal Discovery Under Confounders

Objective Function

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

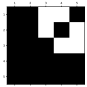
$$\mathcal{S}_{\mathcal{I}}(\mathcal{G}) := \sup_{\boldsymbol{\theta},\boldsymbol{\Sigma}_Z} \sum_{k=1}^{K} \mathbb{E}_{\boldsymbol{X}\sim p^{(k)}} \log p_{\mathrm{do}(I_k)(\mathcal{G})}(\boldsymbol{X}) - \lambda\|\mathcal{G}\|_1.$$

Three main challenges:

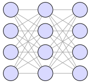
1. Modeling the causal mechanism
2. Computing Log-determinant of the Jacobian
3. Updating the model parameters

Modeling Causal Mechanism

Causal Mechanism: $\mathbf{F}(\boldsymbol{x},\boldsymbol{z}) = -\mathbf{g}_x(\boldsymbol{x}) + \mathbf{g}_z(\boldsymbol{z}) + \boldsymbol{z}$

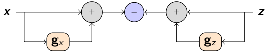


Dependency mask for graph adjacency



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

The SEM forms an *implicit layer*



Computing Log-det-Jacobian

Computing Log-det-Jacobian naively - $O(d^3)$

Power series expansion of $\log\left|\det\left(\mathbf{J}_{\mathbf{f}_x^{(I_k)}}(\boldsymbol{X})\right)\right|$ utilizing $\mathrm{Tr}\left(\mathbf{J}^m_{\mathbf{f}_x^{(I_k)}}(\boldsymbol{X})\right)$ - $O(d^2)$

Hutchinson Trace Estimator (even more reduction)

$$\mathrm{Tr}\{\mathbf{A}\} = \mathbb{E}_{\boldsymbol{W}}[\boldsymbol{W}^\top \mathbf{A}\boldsymbol{W}],$$

where $\mathbb{E}\boldsymbol{W} = 0$, and $\mathbb{E}\boldsymbol{W}^2 = 1$.

DCCD-CONF: Diff. Cyclic Causal Discovery Under Confounders

Objective Function

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

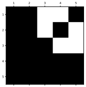
$$\mathcal{S}_{\mathcal{I}}(\mathcal{G}) := \sup_{\boldsymbol{\theta}, \boldsymbol{\Sigma}_Z} \sum_{k=1}^{K} \mathop{\mathbb{E}}_{\boldsymbol{X} \sim p^{(k)}} \log p_{\mathrm{do}(I_k)(\mathcal{G})}(\boldsymbol{X}) - \lambda \|\mathcal{G}\|_1 .$$

Three main challenges:

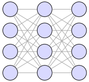
1. Modeling the causal mechanism
2. Computing Log-determinant of the Jacobian
3. Updating the model parameters

Modeling Causal Mechanism

Causal Mechanism: $\mathbf{F}(\boldsymbol{x}, \boldsymbol{z}) = -\mathbf{g}_x(\boldsymbol{x}) + \mathbf{g}_z(\boldsymbol{z}) + \boldsymbol{z}$

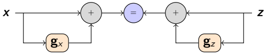


Dependency mask for graph adjacency



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

The SEM forms an *implicit layer*



Computing Log-det-Jacobian

Computing Log-det-Jacobian naively - $O(d^3)$

Power series expansion of $\log \left| \det \left(\mathbf{J}_{\mathbf{f}_x^{(I_k)}}(\boldsymbol{X}) \right) \right|$ utilizing $\mathrm{Tr}\left(\mathbf{J}^m_{\mathbf{f}_x^{(I_k)}}(\boldsymbol{X}) \right)$ - $O(d^2)$

Hutchinson Trace Estimator (even more reduction)

$$\mathrm{Tr}\{\mathbf{A}\} = \mathbb{E}_{\boldsymbol{W}}[\boldsymbol{W}^\top \mathbf{A} \boldsymbol{W}],$$

where $\mathbb{E}\boldsymbol{W} = 0$, and $\mathbb{E}\boldsymbol{W}^2 = 1$.

Parameter Update

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

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

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

which can be solved one column at a time as a series of LASSO regressions.

Results

Theorem. Let $\mathcal{I} = \{I_k\}_{k=1}^{K}$ be a family of interventional targets, let $\mathcal{G}^*$ denote the ground truth directed mixed graph, let $p^{(k)}$ denote the data generating distribution for I_k, and $\hat{\mathcal{G}} := \arg\max_{\mathcal{G}} \mathcal{S}(\mathcal{G})$. Then, under suitable assumptions and a suitably chosen $\lambda > 0$, we have that $\hat{\mathcal{G}}$ is $\mathcal{I}$-Markov equivalent to $\mathcal{G}^*$.

[1] Frangieh, C., et al. "Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion."

Results

Theorem. Let $\mathcal{I} = \{I_k\}_{k=1}^{K}$ be a family of interventional targets, let $\mathcal{G}^*$ denote the ground truth directed mixed graph, let $p^{(k)}$ denote the data generating distribution for I_k, and $\hat{\mathcal{G}} := \arg\max_{\mathcal{G}} \mathcal{S}(\mathcal{G})$. Then, under suitable assumptions and a suitably chosen $\lambda > 0$, we have that $\hat{\mathcal{G}}$ is $\mathcal{I}$-Markov equivalent to $\mathcal{G}^*$.

Gene regulatory network: Data was taken from Frangieh et al (2021)[1]. Contains gene expressions taken from 218,331. Choose 61 genes from around 20,000.

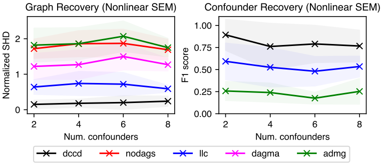
Table: Performance comparison with respect to I-NLL

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)

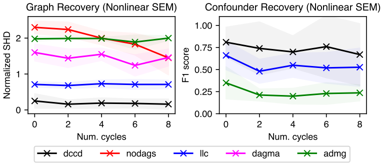
[1] Frangieh, C., et al. "Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion."

Results - Ablation Study

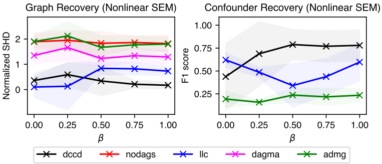
Synthetic Experiments: All experiments were performed on $d = 10$ node graphs. DCCD-CONF was compared with NODAGS-Flow, LLC, DAGMA, DCD



(a) Ablation - Confounders



(b) Ablation - Cycles



(c) Ablation - Nonlinearity