Causal Identification with Matrix Equations

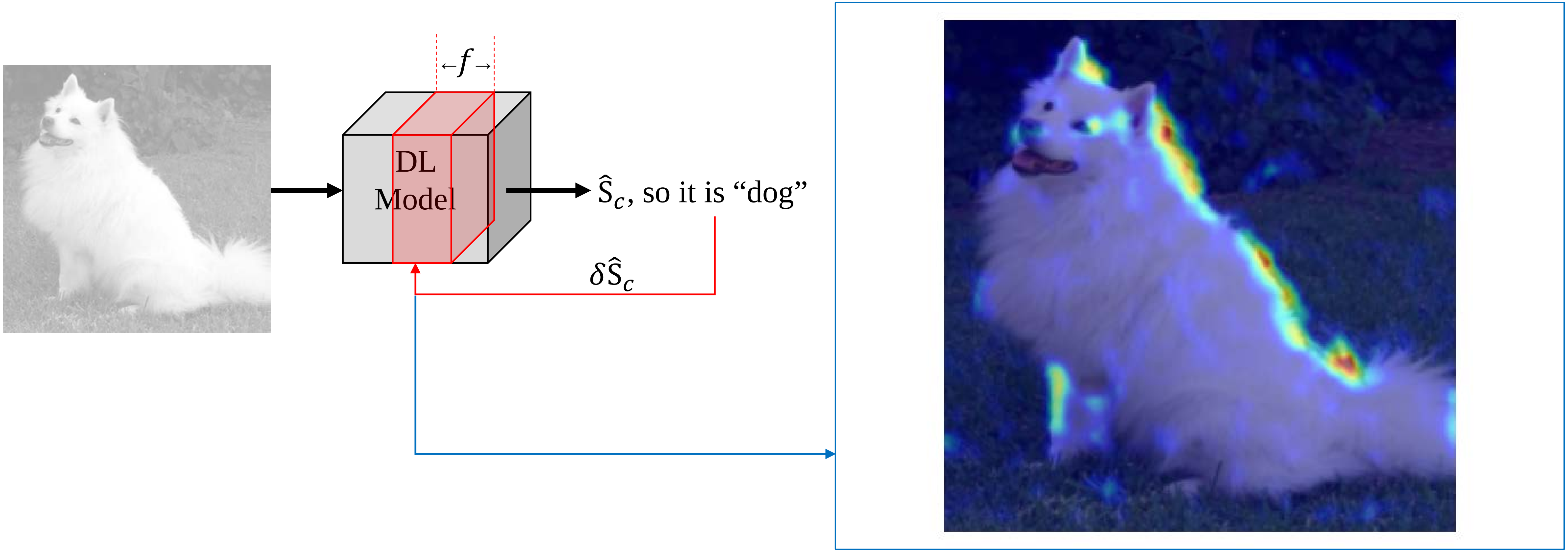
Sanghack Lee and Elias Bareinboin

Presenter: Changha Lee

NeurIPS 2021 Oral Paper

Lab Background

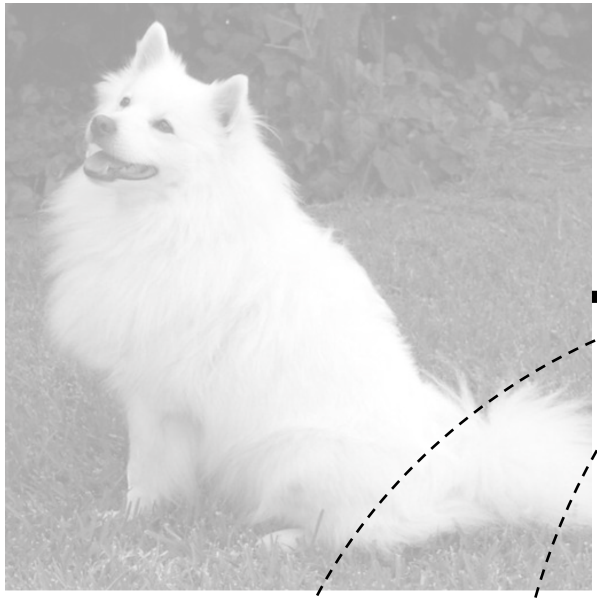
- Class Activation Mapping (CAM)[1] and its variance Grad-CAM[2] gives importance pixels by estimating relations between a DL layer(=activation map f) and a output(=class $\hat{S}_c$).

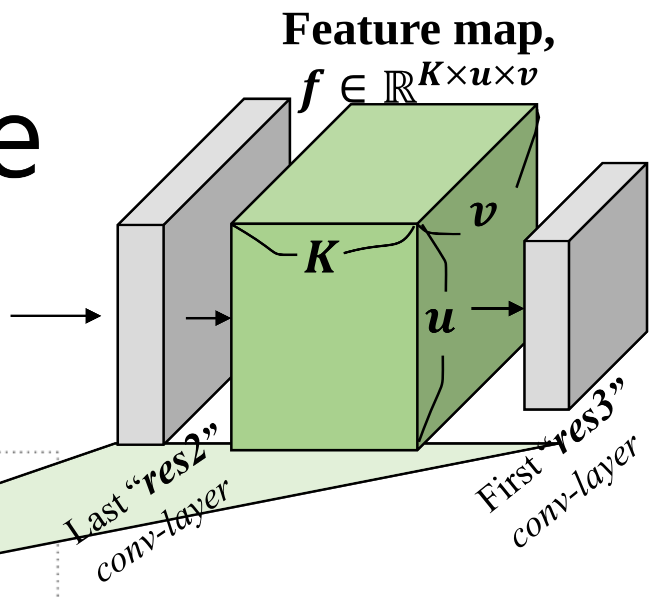


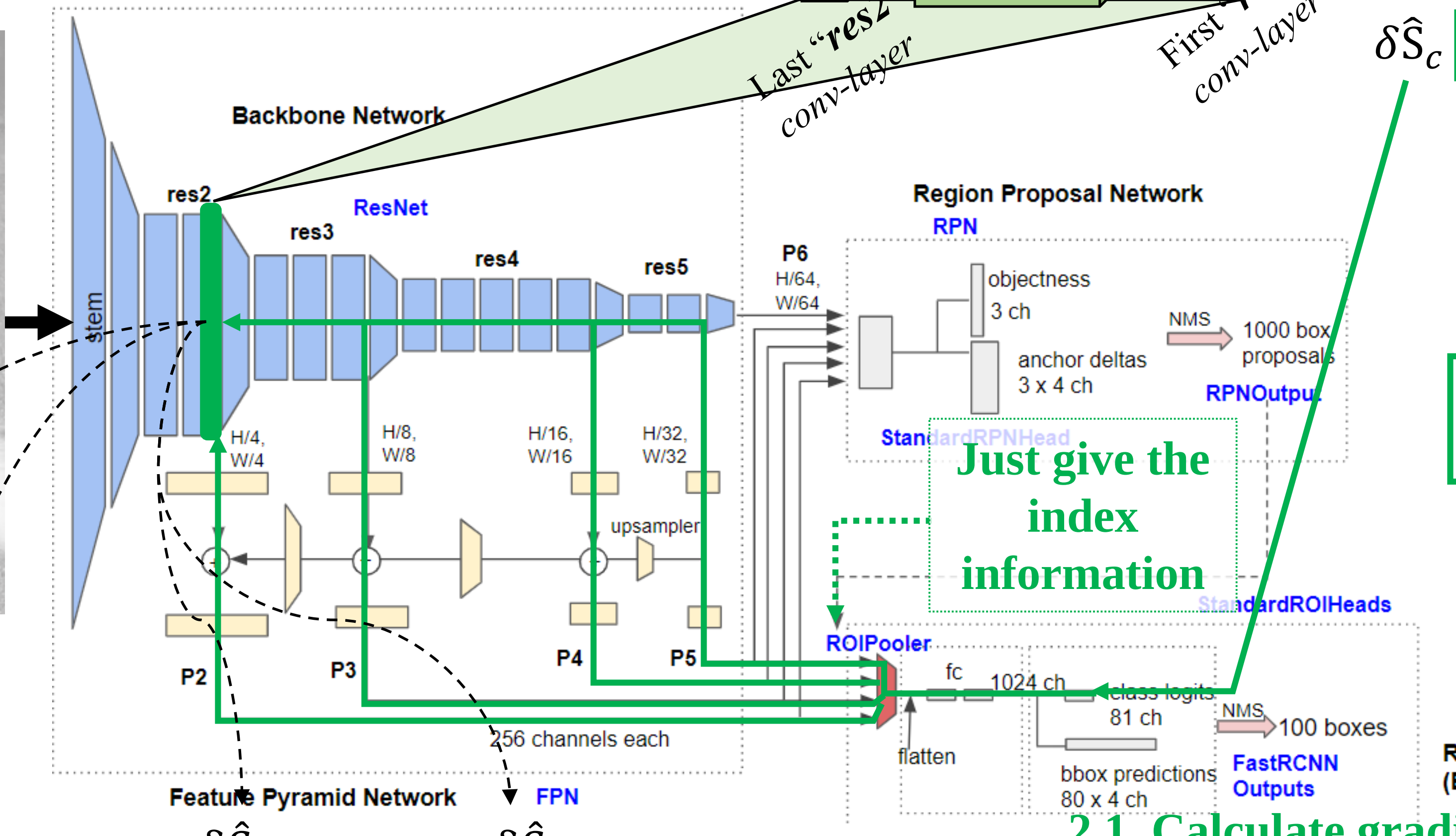
Lab Background Example

Example) GradCAM at ***"res2"***

1. Measure Class Activation Gradients (Backward Processing)







Prediction Score (unsorted)
1.4E^{-5}
2.3E^{-6}
…
$\boxed{\mathbf{0.95}}$
…

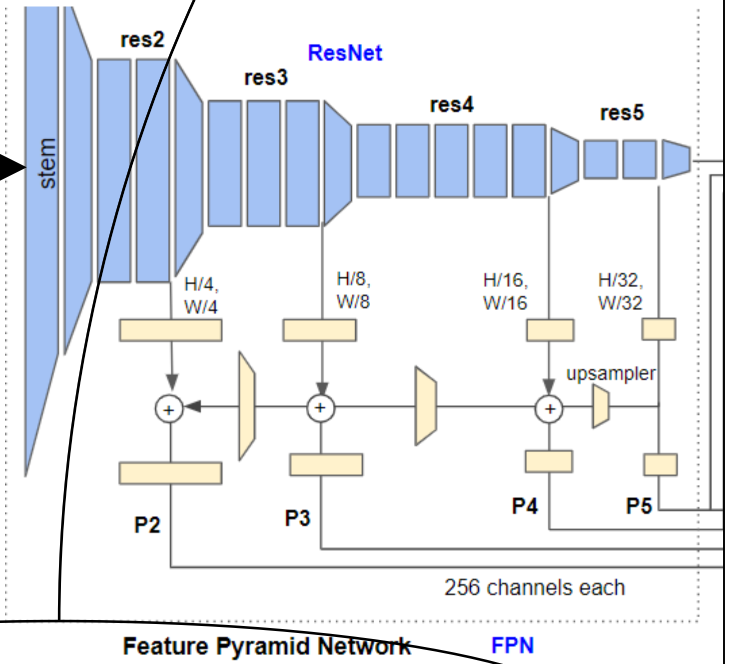
object #1
{ class: "dog"
score: **0.95**,
all_scores: [1.4E^{-5}, 2.3E^{-6} …, **0.95**, …]
…
boxes: [x_1, x_2, y_1, y_2]
}

$$\text{Results: } \frac{\delta \hat{S}_c}{\delta f_K(i,j)}, \frac{\delta \hat{S}_c}{\delta f_{K-1}(i,j)}, \dots, \frac{\delta \hat{S}_c}{\delta f_k(i,j)}, \dots, \frac{\delta \hat{S}_c}{\delta f_1(i,j)}, where\ i \in [u], j \in [v]$$

Lab Background Example

Example) GradCAM at ***"res2"***

2. Generate "Localization Score(=Class Activation Mapping)"



Results: $\dfrac{\delta \hat{S}_c}{\delta f_K(i,j)}, \dfrac{\delta \hat{S}_c}{\delta f_{K-1}(i,j)}, \dots, \dfrac{\delta \hat{S}_c}{\delta f_k(i,j)}, \dots, \dfrac{\delta \hat{S}_c}{\delta f_1(i,j)}$

3.1. Estimate 'importance' of feature map f_k for class c

$$a_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial S^c}{\partial f_k(i,j)}$$ - Eq. (1)

3.2. Weighted Sum-up and ReLU

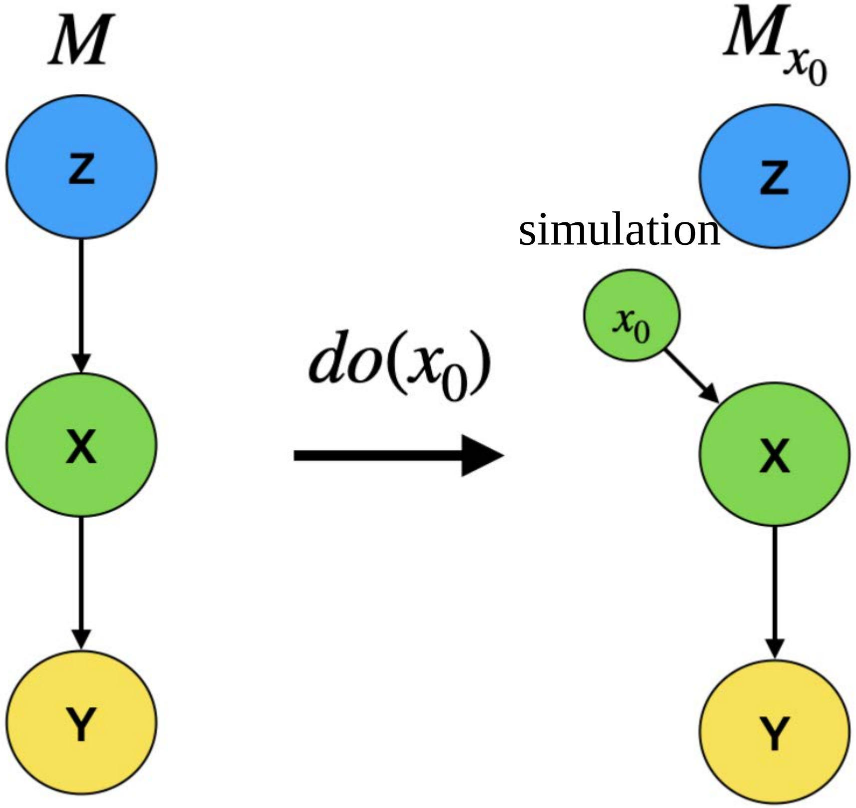
$$L^c_{Grad-CAM}(i,j) = ReLU\left(\sum_k a_k^c f_k(i,j)\right)$$ - Eq. (2)

3.3. Visualization

Then, Scale-up $\boldsymbol{L}^c \in \mathbb{R}^{u \times v}$ to the input image size, $\boldsymbol{I} \in \mathbb{R}^{H \times W}$

Background

- Basically,
- the inferential target usually appears as the effect of a set variables $do(\boldsymbol{X} = x)$ on another set of variables $\boldsymbol{Y}$, which is written as $P(y|do(x))$ or $P_x(y)$ [3].



- Assumptions about the underlying DL processes are commonly expressed as a **causal graph** $\mathcal{G}$ over variables $\boldsymbol{V}$.

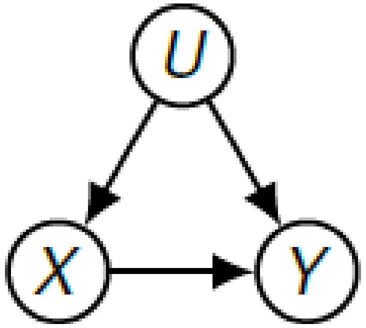
[3] J. Pearl. Causal diagrams for empirical research. Biometrika, 82(4):669–688, 1995.

Background

How to estimate causality: The Baseline, ***factorization-based approach*.**

- If there is ***non-parametric constraints*** U encoded in $\mathcal{G}$,
- A number of necessary and sufficient conditions were developed for determining the identifiability status of the query from observational data [3,4,5,6].

- For example, in Figure 1, the causal effect $P_x(y)$ is identified by a *back-door criterion* as $P_x(y) = \sum_{u \in U} P(y|x,u)P(u)$.



Assumption: all data (i.e., U, X, and Y) can be observed.

Figure 1. An example of back-door condition.

Background (About **Confounder, exogeneous**)

- Simple Example

	No Medicine (x_0)		**Medicine (x_1)**	
Gender (U)	**Heart Attack (y_0)**	**No Heart Attack (y_1)**	**Heart Attack (y_0)**	**No Heart Attack (y_1)**
Female (u_0)	1(5%)	19	3 (7.5%)	37
Male (u_1)	12 (30%)	28	8 (40%)	12
Total	13 (27.7%)	47	11 (18.3%)	49

$P_X(Y|U)$: Heart Attack Probability at given Gender According to Medicine:
Female: No Medicine (5%) → With Medicine (7.5%) ➔ Wrong Interpretation: Increase Heart Attack by Medicine
Male: No Medicine (30%) → With Medicine (40%) ➔ Wrong Interpretation: Increase Heart Attack by Medicine
Total: No Medicine (27.7%) → With Medicine (18.3%) ➔ Right Interpretation: Decrease Heart Attack by Medicine

In this case, we call the variable, Gender={Female, Male}, is a **confounder $\boldsymbol{U}$** (i.e. non-parametric constraints in previous).

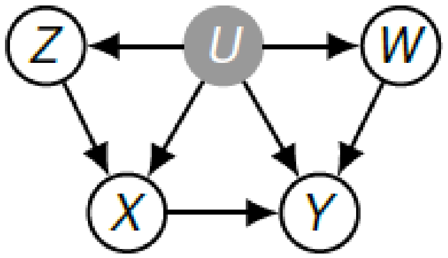
Problem in Unobserved Confounders

How to estimate causality: ***proxy approach*.**

- Exploit assumptions about the relationship between *unobservable confounders U* and the observable variables, which will possibly lead to the invertibility of certain matrices, through the idea of *proxy variables* [8,9].

- For example, in Figure 2, there are an input X, outcome Y, and the proxies W, Z for the set of unobserved confounders U.

"Note that since U is unobserved, the effect of X on Y, i.e., $P_x(y)$ is provably not identifiable by previous approach"



If U cannot be observed, …

Figure 2. An example of proxy approach.

Appendix (***Conditional Independencies***)

This paper author refer [10]. and call Figure 2. as the ***identification condition MGT criterion.***

- [10] **M**iao, W., **G**eng, Z., & **T**chetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. Biometrika, 105(4), 987-993.

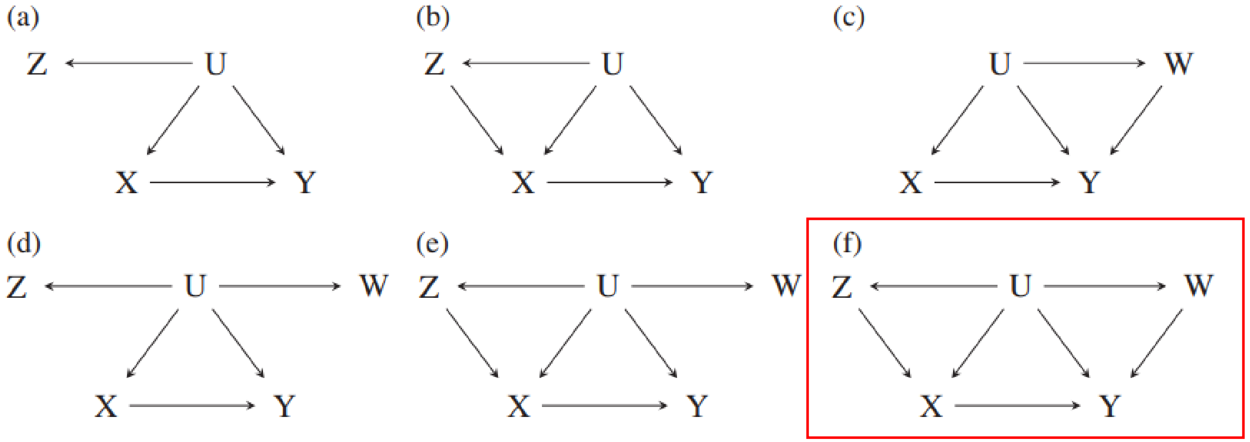


Fig. 1. Causal diagrams with confounder proxies.

MGT criterion
(comes from **M**iao, **G**eng, **T**chetgen **criterion**)

Table 1. *Conditional independencies of the causal diagrams in Fig.* 1

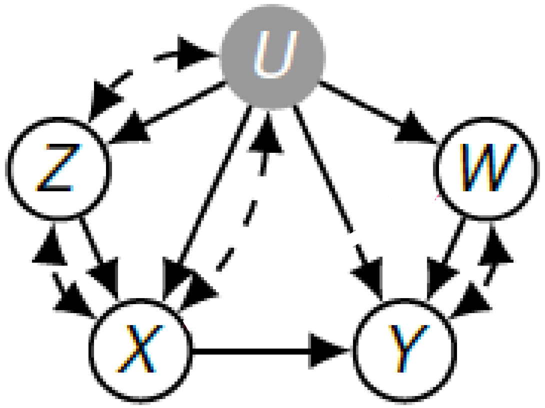
(a) $Z \perp\!\!\!\perp (X, Y) \mid U$	(b) $Z \perp\!\!\!\perp Y \mid (U, X)$
(c) $W \perp\!\!\!\perp X \mid U$	(d) $W \perp\!\!\!\perp (Z, X, Y) \mid U, Z \perp\!\!\!\perp (X, Y) \mid U$
(e) $W \perp\!\!\!\perp (Z, X, Y) \mid U, Z \perp\!\!\!\perp Y \mid (U, X)$	(f) $W \perp\!\!\!\perp (Z, X) \mid U, Z \perp\!\!\!\perp Y \mid (U, X)$

Preliminary (1)

Causality (cambridge.org) (You can check book pdf)

- denote **a variable** by an upper case letter X, and its value is denoted by its corresponding lower case letter x.
- denote **a set of variables** by a bold capital letter $\boldsymbol{X}$, with its value $\boldsymbol{x}$.
- the union of *disjoint* sets; $\dot{\cup}$ not $\cup$
- Given $\boldsymbol{Z} \subseteq \boldsymbol{W}$, $\boldsymbol{W} \backslash \boldsymbol{Z}$ means **all other variables excluding $\boldsymbol{Z}$ in $\boldsymbol{W}$**.
- Let $\boldsymbol{a}/\boldsymbol{B} = (\boldsymbol{A} \cap \boldsymbol{B}, \boldsymbol{a} \backslash \mathrm{B})$ which retains $\boldsymbol{B}$ as a set of variables and values of $\boldsymbol{a}$ excluding $\boldsymbol{B}$.
- *observational*: $P(y|x)$
- *experimental* or *interventional*: $P(y|do(x))$

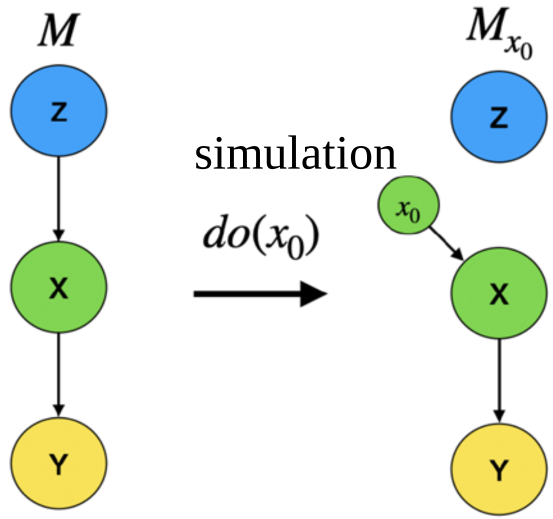
❖They may employ
conditional $\boldsymbol{W} \supseteq \emptyset$,
experimental $\boldsymbol{Z} \supseteq \emptyset$ compared to *observational* $\boldsymbol{Z} = \emptyset$,
and *marginal* if $\boldsymbol{Z} \cup \boldsymbol{W} \cup \boldsymbol{V}' \subsetneq \boldsymbol{V}$ ($\boldsymbol{V}'$ is a subset of $\boldsymbol{V}$)



❖ double-proxy with surrogate examples

simple example

$\mathcal{M} = \langle U, V, F \rangle$



Preliminary (2)

To solve problem, use Structural Causal Models (SCMs)

- $\mathcal{M}$ is a quadruple $\langle \mathbf{U}, \mathbf{V}, P(\mathbf{U}), \mathbf{F} \rangle$.

where

(i) $\mathbf{U}$ is a set of **exogenous(or background, confounder)**, that is determined by factors outside the model, which follows $P(\mathbf{U})$.

(ii) $\boldsymbol{V}$ is a set $\{V_1, \ldots, V_n\}$ of **endogenous** variables, that are determined by variables in the model (or functions) $\mathbf{F} = \{f_{\mathrm{i}}\}_{V_i \in \mathbf{V}}$ such that $V_i \leftarrow f_i(\mathbf{pa}_{\mathrm{i}}, \mathbf{u}_{\mathrm{i}})$ *where* $\mathbf{PA}_{\mathrm{i}} \subseteq \mathbf{V} \backslash \{V_i\}$ *and* $\mathbf{U}_i \subseteq \mathbf{U}$.

…*(omitted)*

$\mathbf{PA}_{\mathrm{i}}$ means a parent set of V_i and its variables are $\mathbf{pa}_i$.

Method (review the part of the paper)

Goal 1. Characterizations of matrix equations with graphical constraints in a given causal diagram $\mathcal{G}$.

- denote $\mathbf{P}$ the matrix notation of a distribution P where free outcomes are row and free condition or intervention are columns
- use $'$ and $''$ to represent two disjoint subsets such that $\mathbf{B} = \mathbf{B}' \dot{\cup} \mathbf{B}''$

Common Characterization.
Let $Q = P_{\mathbf{r}}$ be an arbitrary interventional distribution. Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *and* $\mathbf{R}$ be *disjoint*, a marginal probability is expressed as,

$$Q(\mathbf{a}, \mathbf{b}'|\mathbf{c}) = \sum_{\mathbf{b}''} Q(\mathbf{a}|\mathbf{b}, \mathbf{c})Q(\mathbf{b}|\mathbf{c}) = \mathbf{Q}(\mathbf{a}|\mathbf{b}', \mathbf{B}'', \mathbf{c})\mathbf{Q}(\mathbf{B}'', \mathbf{b}'|\mathbf{c}).$$

Method (review the part of the paper)

Chain Rule with Conditional Independence

- Considering conditional independence, we can enrich previous common characterization.
- **Lemma 1.** *Given a causal diagram* $\mathcal{G}$, *let* $Q = P_{\mathbf{r}}$ *for some* $\mathbf{r} \in \mathfrak{X}_{\mathbf{R}}$ *where* $\mathbf{R} \subsetneq \mathbf{V}$. *Let* $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$ *be disjoint subsets of* $\mathbf{V} \setminus \mathbf{R}$. *If* $(\mathbf{D} \perp\!\!\!\perp \mathbf{A} \mid \mathbf{B}, \mathbf{C}, \mathbf{E})$ *and* $(\mathbf{E} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C}, \mathbf{D})$ *in* $\mathcal{G} \setminus \mathbf{R}$, *then,*

$$\mathbf{Q}(\mathbf{A}, \mathrm{b}' | \mathrm{c}, \mathbf{D}, \mathrm{e}) = \mathbf{Q}(\mathbf{A} | \mathrm{b}', \mathbf{B}'', \mathrm{c}, \mathrm{e}) \mathbf{Q}(\mathbf{B}'', \mathrm{b}' | \mathrm{c}, \mathbf{D}).$$

Adjustment Criteria

- It generalizes back-door criterion [3].
- Its matricized expression with employing $Q = P_{\mathbf{r}}$ is

$$Q_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{z}} Q(\mathbf{y} | \mathbf{x}, \mathbf{z}) Q(\mathbf{z}) = \mathbf{Q}(\mathbf{y} | \mathbf{x}, \mathbf{Z}) \mathbf{Q}(\mathbf{Z}).$$

Method (review the part of the paper)

C-Factorization

- also we did matricize C-Factorization [3].
- In Figure 3a. $\mathbf{P}_{\mathbf{X}_{ij}\backslash\mathbf{Y}_{ij}}(\mathbf{Y}_{ij}\backslash\mathbf{Z})$ can be obtained by multiplication of submatrices $\mathbf{P}_{\mathbf{X}_i}(\mathbf{Y}_i)$ and $\mathbf{P}_{\mathbf{X}_j}(\mathbf{Y}_j)$

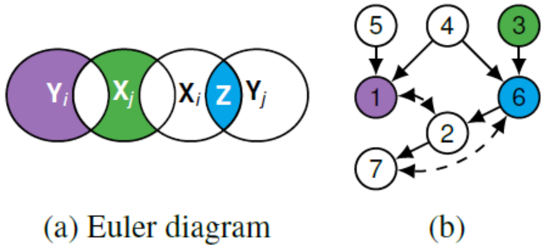


Figure 3: (a) set relationships, (b) a causal graph where the numbers correspond to the regions of (a) from the left to right.

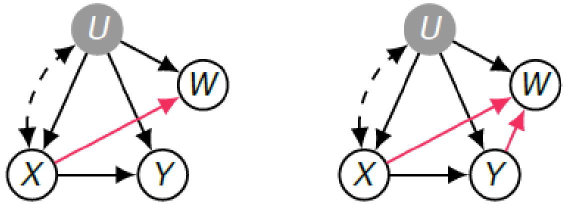
- **Lemma 2 (Matrix Equation of C-Factorization with Two-Factors)**

Given a causal diagram $\mathcal{G}$ and an experimental distribution $Q = P_{\mathbf{r}}$, if $\mathbf{Z} \subseteq \mathbf{X}_i \cap \mathbf{Y}_j$, a *causal effect* which is c-factorized as $Q_x(y) = \sum_z Q_{x_i}(y_i)Q_{x_j}(y_j)$ can be matricized by

$$\mathbf{Q}_{(\mathbf{x}_{ij}\backslash\mathbf{y}_{ij})/(\mathbf{X}_j\backslash\mathbf{X}_i\backslash\mathbf{Y}_i)}((\mathbf{y}_{ij}\backslash\mathbf{Z})/(\mathbf{Y}_i\backslash\mathbf{X}_j)) = \mathbf{Q}_{\mathbf{x}_i/\mathbf{Z}}(\mathbf{y}_i/(\mathbf{Y}_i\backslash\mathbf{X}_j))\mathbf{Q}_{\mathbf{x}_j/(\mathbf{X}_j\backslash\mathbf{X}_i\backslash\mathbf{Y}_i)}(\mathbf{y}_j/\mathbf{Z}).$$

Method (review the part of the paper)

Goal 2. With the characterization from previous goal, the next goal is the <u>generalization of single-</u> and double-<u>proxy settings</u> [11].



(a) $P_x(W)$ (b) $P_{x,y}(W)$

Single-Proxy Setting

Figure 4. illustrate the available distributions and unknown distributions, considered in Figure 4b.

$$\overbrace{\Big(\underbrace{(\mathbf{P}(W|U)^{-1}\mathbf{P}(y,W|x))}_{\mathbf{P}(y,U|x)} \oslash \underbrace{(\mathbf{P}(W|U)^{-1}\mathbf{P}(W|x))}_{\mathbf{P}(U|x)}\Big)^{\top}}^{\mathbf{P}(y|U,x)} \underbrace{\mathbf{P}(W|U)^{-1}\mathbf{P}(W)}_{\mathbf{P}(U)}.$$

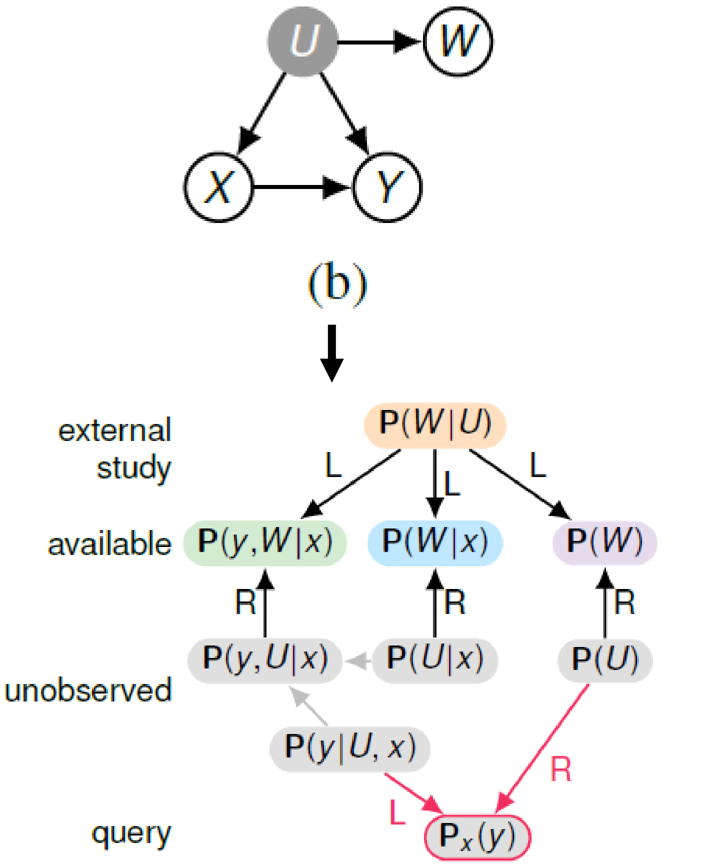
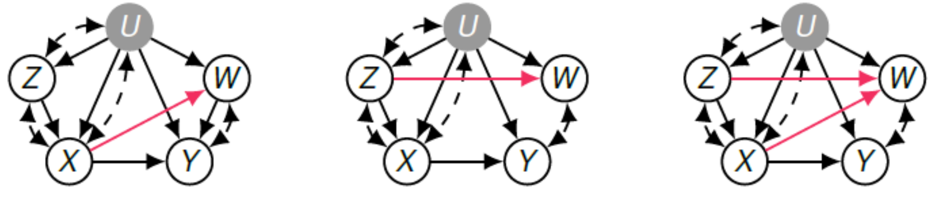


Figure 4: Schematic for a single proxy setting. Gray and red lines are for elementwise matrix multiplication and adjustment criterion, respectively.

Method (review the part of the paper)

Goal 2. With the characterization from previous goal, the next goal is the generalization of single- and double-proxy settings [11].

Double-Proxy Setting



(c) $P_x(W)$ (d) $P_z(W)$ (e) $P_{x,z}(W)$

Figure 5. illustrate the available distributions and unknown distributions, considered ***MGT criterion*** in Figure 2.

Theorem 2 (Generalized MGT Criterion). *Given a causal graph $\mathcal{G}$, let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{U}, \mathbf{R} \subset \mathbf{V}$ be disjoint sets of variables where $\mathbf{R}$ can be empty. Let $Q = P_{\mathbf{r}}$ for some $\mathbf{r} \in \mathfrak{X}_{\mathbf{R}}$ and $\mathcal{H} = \mathcal{G} \backslash \mathbf{R}$. Let $\mathbf{S} \subseteq \mathbf{X} \cup \mathbf{Z}$. A causal effect $Q_{\mathbf{x}}(\mathbf{y}) = P_{\mathbf{x},\mathbf{r}}(\mathbf{y})$ is identifiable in $\mathcal{G}$ if, for some $\mathbf{z}$, (1) $\mathbf{U}$ is an adjustment set for $Q_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{H}$; (2) $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{U}, \mathbf{X})_{\mathcal{H}}$; (3) $(\mathbf{W} \perp\!\!\!\perp \mathbf{S}' \mid \mathbf{U}, \mathbf{S})_{\mathcal{H}}$ where $\mathbf{S}' = (\mathbf{X} \cup \mathbf{Z}) \backslash \mathbf{S}$; (4) $\mathbf{U}$ is an adjustment set for $Q_{\mathbf{s}'}(\mathbf{W})$ in $\mathcal{H}$; (5) $\mathbf{Q}(\mathbf{W}|\mathbf{U}, \mathbf{s}')$ is invertible; (6) $\mathbf{Q}(\mathbf{U}|\mathbf{z}/(\mathbf{S} \cup \mathbf{Z}'), \mathbf{x})$ is invertible for some $\mathbf{Z}' \subseteq \mathbf{Z} \backslash \mathbf{S}$ and $Q(\mathbf{y}|\mathbf{z}/(\mathbf{S} \cup \mathbf{Z}'), \mathbf{x})$, $Q(\mathbf{W}|\mathbf{z}/(\mathbf{S} \cup \mathbf{Z}'), \mathbf{x})$, and $Q_{\mathbf{s}'}(\mathbf{W})$ are available. Additionally, $Q(\mathbf{W}|\mathbf{U}, \mathbf{s}'/\mathbf{Z}')$ is available if $\mathbf{Z}' \neq \emptyset$.*

Remind Figure 2.

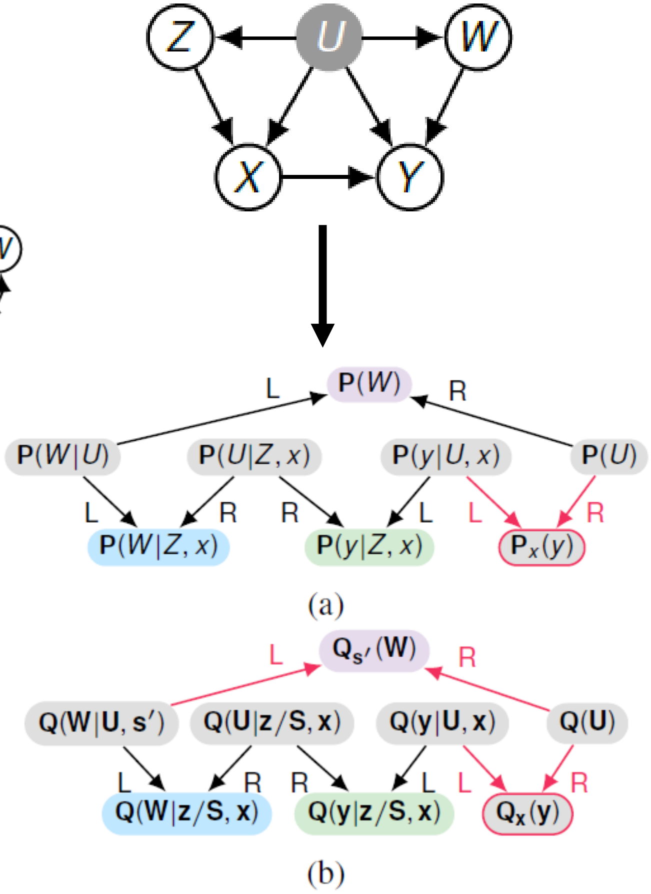


Figure 5: Schematics of (a) MGT criterion and (b) its generalization (simplified)

Method (review the part of the paper)

The paper suggest **a novel identification condition(=intermediary criteria)** for a challenging setting that neither previous matrical nor graphical approaches could handle.

Pseudoinverse

- First, they prove that the probability of interest is ***identificable*** through the ***pseudoinverse*** without invertibility assumption from proxy settings

Lemma 3 (Base Intermediary Criterion). *Let* $\{P_1, P_2, P_3, P_4\}$ *be distributions. Let* $\{\mathbf{P}_i\}_{i=1}^4$ *be their matrix representations. If submatrices* $\{\mathbf{P}'_i\}_{i=1}^4$ *of* $\{\mathbf{P}_i\}_{i=1}^4$ *satisfy* $\mathbf{P}'_1 = \mathbf{P}'_2\mathbf{P}'_3\mathbf{P}'_4$ *and* $\mathbf{P}'_2\mathbf{P}'_3$, $\mathbf{P}'_3\mathbf{P}'_4$, *and* $\mathbf{P}'_3$ *are given, then,* $\mathbf{P}'_1 = (\mathbf{P}'_2\mathbf{P}'_3)\mathbf{P}'^{\dagger}_3(\mathbf{P}'_3\mathbf{P}'_4)$.

Proof. By the given condition, the associativity of matrix multiplications, and the property of pseudoinverse $\mathbf{P}\mathbf{P}^{\dagger}\mathbf{P} = \mathbf{P}$, $\mathbf{P}'_1 = \mathbf{P}'_2\mathbf{P}'_3\mathbf{P}'_4 = \mathbf{P}'_2(\mathbf{P}'_3\mathbf{P}'^{\dagger}_3\mathbf{P}'_3)\mathbf{P}'_4 = (\mathbf{P}'_2\mathbf{P}'_3)\mathbf{P}'^{\dagger}_3(\mathbf{P}'_3\mathbf{P}'_4)$. □

Method (review the part of the paper)

Intermediary Criterion

- Second, they characterize intermediary criterion from **Chain-Rule** to **C-Factorization.**

Lemma 4 (Chain-Rule Intermediary Criterion). *Given a causal diagram* $\mathcal{G}$*, let* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, *and* $\mathbf{R}$ *be disjoint subsets of* $\mathbf{V}$ *with* $\mathbf{D}$ *and* $\mathbf{R}$ *can be empty. Let* $\mathbf{B} = \mathbf{B}' \dot{\cup} \mathbf{B}''$ *and* $\mathbf{C} = \mathbf{C}' \dot{\cup} \mathbf{C}''$ *where* $\mathbf{B}'$ *and* $\mathbf{C}'$ *are not empty. Given an interventional distribution* $Q = P_{\mathbf{r}}$*, if* $(\mathbf{C}' \perp\!\!\!\perp \mathbf{A} \mid \mathbf{B}, \mathbf{C}''\mathbf{D})$ *in* $\mathcal{G}\backslash\mathbf{R}$ *and* $Q(\mathbf{a}, \mathbf{b}'', \mathbf{c}''|\mathbf{d}) = \sum_{\mathbf{b}',\mathbf{c}'} Q(\mathbf{a}|\mathbf{b}, \mathbf{c}, \mathbf{d}) Q(\mathbf{b}|\mathbf{c}, \mathbf{d}) Q(\mathbf{c}|\mathbf{d})$*, then,*

$$\mathbf{Q}(\mathbf{A}, \mathbf{b}'', \mathbf{c}''|\mathbf{d}) = \mathbf{Q}(\mathbf{A}, \mathbf{b}''|\mathbf{C}', \mathbf{c}'', \mathbf{d}) \cdot \mathbf{Q}(\mathbf{B}', \mathbf{b}''|\mathbf{C}', \mathbf{c}'', \mathbf{d})^{\dagger} \cdot \mathbf{Q}(\mathbf{B}', \mathbf{b}'', \mathbf{c}''|\mathbf{d}).$$

Theorem 3 (C-Factorization Intermediary Criterion). *Let* $\mathcal{G}$ *be a causal diagram and* $Q = P_{\mathbf{r}}$*. Let* $Q_{\mathbf{x}_\ell}(\mathbf{y}_\ell)$ *be c-factorized as Eq.* ($\boxed{4}$)*. Let* $\mathbf{X}_k^+$ *be a subset of* $\mathbf{X}_k$ *excluding the rest five sets,* $\{\mathbf{Y}_i, \mathbf{Y}_j, \mathbf{Y}_k, \mathbf{X}_i, \mathbf{X}_j\}$*.* $\mathbf{Y}_i^+$ *is similarly defined. If* $\mathbf{Z} \subseteq (\mathbf{X}_i \cap \mathbf{Y}_j)\backslash\mathbf{X}_k$ *and* $\mathbf{W} \subseteq \mathbf{X}_j \cap \mathbf{Y}_k$*, then* $\mathbf{Q}_{\mathbf{x}_\ell/\mathbf{X}_k^+}(\mathbf{y}_\ell/\mathbf{Y}_i^+)$*, a submatrix of* $\mathbf{Q}_{\mathbf{X}_\ell}(\mathbf{Y}_\ell)$*, becomes*

$$\mathbf{Q}_{\mathbf{x}_\ell/\mathbf{X}_k^+}(\mathbf{y}_\ell/\mathbf{Y}_i^+) = \mathbf{Q}_{(\mathbf{x}_{ij}\backslash\mathbf{y}_{ij})/\mathbf{W}}((\mathbf{y}_{ij}\backslash\mathbf{Z})/\mathbf{Y}_i^+) \cdot \mathbf{Q}_{\mathbf{x}_j/\mathbf{W}}(\mathbf{y}_j/\mathbf{Z})^{\dagger} \cdot \mathbf{Q}_{(\mathbf{x}_{jk}\backslash\mathbf{y}_{jk})/\mathbf{X}_k^+}((\mathbf{y}_{jk}\backslash\mathbf{W})/\mathbf{Z}).$$

Method (review the part of the paper)

- Then, they suggest **ID-ME** (Identification with Matrix Equation) algorithm

Algorithm 1 ID-ME

function ID-ME$(\mathbf{x}, \mathbf{y}, \mathcal{G}, \mathbb{D})$
  Input: $\mathbf{x}, \mathbf{y}$ value assignments for a query $P_{\mathbf{x}}(\mathbf{y})$; $\mathcal{G}$ a causal diagram; $\mathbb{D}$ a collection of distributions
  Output: a formula for $P_{\mathbf{x}}(\mathbf{y})$ made with $\mathbb{D}$ or FAIL.
    $\mathbf{X}, \mathcal{G}' \leftarrow \mathrm{an}^{\mathcal{G}_{\overline{\mathbf{X}}}}(\mathbf{Y}) \cap \mathbf{X}, \mathcal{G}[\mathrm{An}^{\mathcal{G}}(\mathbf{Y})]$
    $\mathbb{D} \leftarrow$ expand$(\mathcal{G}, \mathbb{D})$ **unless** $\mathbb{D}$ is unconditional
    $\mathbb{F}^{+} \leftarrow \mathbb{F}_{\mathcal{G}'}$ **if** $\mathbb{D}$ is unconditional and nonmarginal **else** $\bigcup\{\mathbb{F}_{\mathcal{G}'\langle\mathbf{V}'\rangle} \mid \mathbf{X} \cup \mathbf{Y} \subseteq \mathbf{V}' \subseteq \mathbf{V}\}$
    $\mathbb{F}', \mathbb{F}'' \leftarrow$ empty dictionary, $\mathbb{F}^{+}$
    for all $P_{\mathbf{X}_i}(\mathbf{Y}_i) \in \mathbb{F}''$ and $\{P_{\mathbf{Z}}(\mathbf{V}'|\mathbf{T}) \in \mathbb{D} \mid \mathbf{Y}_i \subseteq \mathbf{V}', \mathbf{X}_i \subseteq \mathbf{Z} \cup \mathbf{V}' \cup \mathbf{T}\}$ **do**
      update $\mathbb{F}', \mathbb{F}''$ with ID-RC$(P_{\mathbf{X}_i}(\mathbf{Y}_i), \mathcal{G}', P_{\mathbf{Z}}(\mathbf{V}'|\mathbf{T}))$
    for all $P_{\mathbf{X}_i}(\mathbf{Y}_i) \in \mathbb{F}''$ **do** update $\mathbb{F}', \mathbb{F}''$ with PROXY$(P_{\mathbf{X}_i}(\mathbf{Y}_i), \mathcal{G}, \mathbb{D} \cup \mathbb{F}')$
    repeat update $\mathbb{F}', \mathbb{F}''$ with CF-INT$(\mathcal{G}', \mathbb{F}', \mathbb{F}'')$ and CF-INV$(\mathcal{G}', \mathbb{F}', \mathbb{F}'')$ **until** $\mathbb{F}'$ not changed
    return exact-cover$(\mathcal{G}', P_{\mathbf{x}}(\mathbf{y}), \mathbb{F}')$

expand: a procedure of chain rule closure with *chain-rule matrix inversion* and *chain-rule intermediary criterion*
ID-RC: an identification module modified to hand a conditional distribution [12,13]
PROXY: generalized proxy criteria
CF-INT: Lemma 2
CF-INV: Theorem 3

Conclusion

- Characterize matrix equations made of distributions driven by graphical constraints
- Generalize proxy-based identification methods
- Develop novel intermediary criteria by utilizing pseudoinverse
- Suggest a causal identification algorithm for a causal query given a causal graph and distributions that can be marginal, experimental, and conditional harnessing the power of matrix equations and (pseudo)inverse.

References

[1] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). *Learning Deep Features for Discriminative Localization* (pp. 2921–2929). http://cnnlocalization.csail.mit.edu

[2] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). *Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization* (pp. 618–626). http://gradcam.cloudcv.org

[3] J. Pearl. Causal diagrams for empirical research. Biometrika, 82(4):669–688, 1995.

[4] J. Tian and J. Pearl. A general identification condition for causal effects. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002), pages 567–573, MenloPark, CA, 2002. AAAI Press/The MIT Press.

[5] I. Shpitser and J. Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI 2006), pages 1219–1226. AAAI Press, Menlo Park, CA, 2006.

[6] Y. Huang and M. Valtorta. Pearl's calculus of intervention is complete. In R. Dechter and T. Richardson, editors, Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI 2006), pages 217–224. AUAI Press, Corvallis, OR, 2006.

[7] J. Tian and J. Pearl. A general identification condition for causal effects. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002), pages 567–573, MenloPark, CA, 2002. AAAI Press/The MIT Press.

[8] R. Carroll, D. Ruppert, L. Stefanski, and C. Crainiceanu. Measurement Error in Nonlinear Models: A Modern Perspective. Boca Raton: Chapman & Hall/CRC, 2nd edition, 2006.

[9] S. Greenland and T. Lash. Bias analysis. In K. Rothman, S. Greenland, and T. Lash, editors, Modern Epidemiology, pages 345–380. Lippincott Williams & Wilkins, Philadelphia, PA, 3rd edition, 2008.

[10] W. Miao, Z. Geng, and E. J. Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. Biometrika, 105(4):987–993, 08 2018. ISSN 0006-3444.

[11] Miao, Wang, Zhi Geng, and Eric J. Tchetgen Tchetgen. "Identifying causal effects with proxy variables of an unmeasured confounder." *Biometrika* 105.4 (2018): 987-993.

[12] E. Bareinboim and J. Tian. Recovering causal effects from selection bias. In S. Koenig and B. Bonet, editors, Proceedings of the 29th AAAI Conference on Artificial Intelligence, pages 3475–3481, Palo Alto, CA, 2015. AAAI Press.

[13] J. J. R. Lee and I. Shpitser. Identification methods with arbitrary interventional distributions as inputs, 2020. arXiv:2004.01157.