Enumerating Optimal Cost-Constrained Adjustment Sets

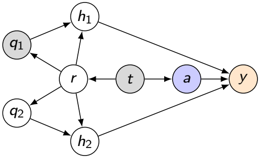
Batya Kenig
Technion, Israel Institute of Technology

41st Conference on Uncertainty in Artificial Intelligence (UAI),
July 21-25, 2025

Measuring the Causal Effect

Observational Dist. $+$ Causal Graph $\Rightarrow$ Interventional Dist.

$P(q_2, h_1, r, h_2, a, y)$ $P(Y \mid do(a))$



[1]

From $P(Y \mid do(a))$ we can compute quantities of interest:

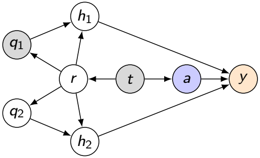
ATE: $\mathbb{E}_P(Y \mid do(a)) - \mathbb{E}_P(Y \mid do(a'))$

[1] Illustration based on Rotnitzky 2021.

Measuring the Causal Effect

Observational Dist. $+$ Causal Graph $\Rightarrow$ Interventional Dist.

$P(q_2, h_1, r, h_2, a, y)$ $P(Y \mid do(a))$



1

From $P(Y \mid do(a))$ we can compute quantities of interest:

ATE: $\mathbb{E}_P(Y \mid do(a)) - \mathbb{E}_P(Y \mid do(a'))$

Can the causal effect of a on y be identified from the observational dist. ?

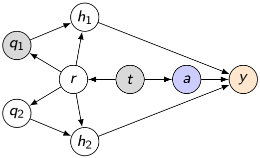
Yes. $P(y \mid do(a)) = \sum_r P(y \mid a, r) \cdot P(r)$

[1] Illustration based on Rotnitzky 2021.

Measuring the Causal Effect

Observational Dist. $\quad + \quad$ Causal Graph $\quad \Rightarrow \quad$ Interventional Dist.

$P(q_2, h_1, r, h_2, a, y)$ $\qquad\qquad$ $P(Y \mid do(a))$



[1]

From $P(Y \mid do(a))$ we can compute quantities of interest:

$$\text{ATE: } \mathbb{E}_P(Y \mid do(a)) - \mathbb{E}_P(Y \mid do(a'))$$

Can the causal effect of a on y be identified from the observational dist. ?

$$\text{Yes.} \; P(y \mid do(a)) = \sum_r P(y \mid a, r) \cdot P(r)$$

In this example, $\{r\}$ is an **adjustment set** for estimating the causal effect of a on y. Since r is observed, then $\{r\}$ is a **valid adjustment set**, and hence the causal effect of a on y is **identifiable**.

[1] Illustration based on Rotnitzky 2021.

Adjustment Sets

Definition (Adjustment Sets)

Given a DAG G, and pairwise disjoint $X, Y, Z \subseteq V(G)$, Z is an *adjustment set* for estimating the causal effect of X on Y if, for every distribution P that factorizes according to G:

$$P(\boldsymbol{Y} \mid do(\boldsymbol{X})) = \begin{cases} P(\boldsymbol{Y} \mid \boldsymbol{X}) & \text{if } Z = \emptyset \\ \sum_{\boldsymbol{Z}} P(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}) P(\boldsymbol{Z}) & \text{otherwise} \end{cases}$$

The choice of adjustment set Z affects:

1. Time to compute interventional dist. $P(\boldsymbol{Y} \mid do(\boldsymbol{X}))$ and the derived quantities of interest. Takes time exponential in $|\boldsymbol{Z}|$.
2. The variance $\sigma^2_{\boldsymbol{Z}}(P)$ of the (asymptotically) normally-distributed estimator $\hat{\mathbb{E}}_P(\boldsymbol{Y} \mid do(\boldsymbol{X}))$ Smucler et al. 2021, Smucler and Rotnitzky 2022, Runge 2021.

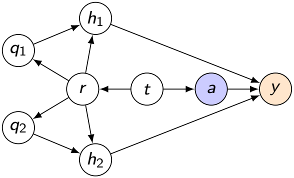
Natural Requirements from Adjustment Sets

Let G be a DAG with observable variables $R \subseteq V(G)$. We wish to find adjustment sets $Z \subseteq V(G)$ that are:

1. Valid : $Z \subseteq R$.
2. Non-Redundant (Minimal) : there is no $Z' \subsetneq Z$ that is an adjustment set for estimating the causal effect.
3. Small size (or weight); of size at most k.
4. Yields estimator with low variance $\sigma^2_{\boldsymbol{Z}}(P)$:

$$\hat{\mathbb{E}}_P(\boldsymbol{Y} \mid do(\boldsymbol{X})) = \sum_{\boldsymbol{Z}} \hat{\mathbb{E}}_P(\boldsymbol{Y} \mid do(\boldsymbol{X}), \boldsymbol{Z}) \cdot P(\boldsymbol{Z}).$$

Examples of Adjustment Sets



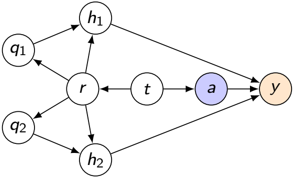
2

Minimal adjustment sets for estimating the causal effect of a on y: $\{t\}$, $\{r\}$, $\{h_1, q_2\}$, $\{h_2, q_1\}$, $\{h_1, h_2\}$, $\{q_1, q_2\}$.

- Adjustment set(s) that optimize for size: $\{t\}$, $\{r\}$.
- Adjustment set(s) that optimize for variance: $\{h_1, h_2\}$ Smucler et al. 2021.

[2] Illustration based on Rotnitzky 2021.

Examples of Adjustment Sets



[2]

Minimal adjustment sets for estimating the causal effect of a on y: $\{t\}$, $\{r\}$, $\{h_1, q_2\}$, $\{h_2, q_1\}$, $\{h_1, h_2\}$, $\{q_1, q_2\}$.

- Adjustment set(s) that optimize for size: $\{t\}$, $\{r\}$.
- Adjustment set(s) that optimize for variance: $\{h_1, h_2\}$ Smucler et al. 2021.

Problem: Size and variance are at odds!

[2] Illustration based on Rotnitzky 2021.

Domination Among Adjustment Sets

- Let $Z_1, Z_2 \subseteq V(G)$ be two valid adjustment sets for estimating the causal effect of X on Y in G.
- We write $Z_1 \leq^{\sigma}_{G} Z_2$ if $\sigma^2_{Z_1}(P) \leq \sigma^2_{Z_2}(P)$ for every joint probability dist. P that factorizes according to G.
  - Induces a partial order. There may exist two adjustment sets S_1, S_2 where $\sigma^2_{S_1}(P) < \sigma^2_{S_2}(P)$ and $\sigma^2_{S_2}(P') < \sigma^2_{S_1}(P')$ where P, P' both factorize according to G.
- We say that Z_1 **dominates** Z_2 if:

$$Z_1 \leq^{\sigma}_{G} Z_2 \text{ and } |Z_1| \leq |Z_2|$$

and one of these inequalities is proper.

Pareto-Optimal Frontier of Adjustment Sets

Goal: Return all non-dominated adjustment sets.

Pareto-Optimal Frontier of Adjustment Sets

Goal: Return all non-dominated adjustment sets.

Problems:

1. Too many.
2. Too big/expensive to be practical.

Refined Goal: Return all non-dominated adjustment sets whose size (weight) is at most k.

Pareto-Optimal Frontier of Adjustment Sets

Goal: Return all non-dominated adjustment sets.

Problems:

1. Too many.
2. Too big/expensive to be practical.

Refined Goal: Return all non-dominated adjustment sets whose size (weight) is at most k.

Theorem

There is an algorithm that lists the Pareto-Optimal minimal adjustment sets for computing an unbiased estimator of the interventional mean of outcomes Y under interventions on X, of size at most k, in total time $O(4^k \cdot k \cdot (n+m))$.

Generating the Pareto-Optimal Frontier of Adjustment Sets

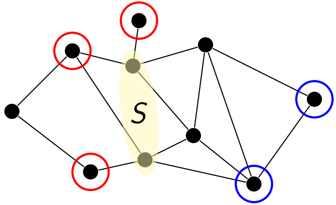
Fixed setting: DAG G; obtain unbiased estimator of the interventional mean of outcomes Y under interventions on X.

Ingredients:

1. Translate problem of finding adjustment sets in DAG G to finding *separators* in an undirected graph $\mathcal{H}$.
   - Validity constraints - we want only adjustment sets $Z \subseteq R$.
   - Inclusion constraints - we want only adjustment sets $I \subseteq Z$.
2. Translate critierion for (partially) ordering adjustment sets based on the variance they yield ($Z_1 \leq_G^{\sigma} Z_2$) to a graphical criterion of separators.

Ingredient #1: From Adjustment Sets to Separators

- Let $\mathcal{H}$ be an undirected graph, and $X, Y \subseteq V(\mathcal{H})$.
- $S \subseteq V(\mathcal{H})$ is an X, Y-separator of $\mathcal{H}$ if in the graph $\mathcal{H} - S$ there is no path from X to Y.
- S is a *minimal* X, Y-separator if none of its proper subsets are.
- S is a *minimum* X, Y-separator of $\mathcal{H}$ if $|S| \leq |S'|$ for every X, Y-separator S'.



Graph $\mathcal{H}$, vertex sets X, $Y \subseteq V(\mathcal{H})$.

We denote by $\mathcal{S}_{X,Y}(\mathcal{H})$ the minimal X, Y-separators in $\mathcal{H}$.

Ingredient #1: From Adjustment Sets to Separators

We denote by $\mathcal{A}_{X,Y}(I, R, G)$ all subsets $I \subseteq Z \subseteq R$ that are adjustment sets for computing an unbiased estimator of the interventional mean of outcomes Y under interventions X. We denote by $\mathcal{A}^{\mathrm{MIN}}_{X,Y}(I, R, G)$ the minimal adjustment sets.

Theorem (based on van der Zander et al. 2019, Smucler et al. 2021)

There is an undirected graph $\mathcal{H}$ derived from G where :

1. $\mathcal{A}_{X,Y}(I, R, G) \neq \emptyset$ *if and only if* X *and* Y *are non-adjacent in* $\mathcal{H}$.
2. $S \in \mathcal{A}^{\mathrm{MIN}}_{X,Y}(I, R, G)$ *if and only if* $S \in \mathcal{S}_{X,Y}(\mathcal{H})$.

Ingredient #2: Graphical Criterion for $\leq_G^\sigma$

- Smucler et al. 2021 presented a graphical criterion for the univariate case - single treatment variable X and single outcome variable Y.
- We generalize to the case where X and Y are variable-sets, and show an equivalent criteria.
- For an X, Y-separator S in $\mathcal{H}$, we denote by:
  - $\mathcal{C}_X(\mathcal{H}-S) \stackrel{\text{def}}{=} \{v \in V(\mathcal{H}) : v \text{ is reachable from } X \text{ in } \mathcal{H}-S\}$.
  - $\mathcal{C}_Y(\mathcal{H}-S) \stackrel{\text{def}}{=} \{v \in V(\mathcal{H}) : v \text{ is reachable from } Y \text{ in } \mathcal{H}-S\}$.

Theorem

Let $S_1, S_2 \in \mathcal{A}_{X,Y}^{\mathrm{MIN}}(I, R, G)$, then:

$$S_1 \leq_G^\sigma S_2 \text{ if and only if } \mathcal{C}_Y(\mathcal{H}-S_1) \subseteq \mathcal{C}_Y(\mathcal{H}-S_2).$$

Graphical Criteria for Domination of Adjustment Sets

Theorem

Let $S_1, S_2 \in \mathcal{A}^{\mathrm{MIN}}_{X,Y}(I, R, G)$*, then:*

$$S_1 \leq^{\sigma}_{G} S_2 \text{ if and only if } \mathcal{C}_Y(\mathcal{H} - S_1) \subseteq \mathcal{C}_Y(\mathcal{H} - S_2).$$

- Causal model G; $R \subseteq V(G)$ are observed; treatment variables X and outcome variables Y.
- Let $\mathcal{H}$ be the undirected graph derived from G where $\mathcal{A}^{\mathrm{MIN}}_{X,Y}(I, R, G) = \mathcal{S}_{X,Y}(\mathcal{H})$ van der Zander et al., Smucler et al. 2019, 2021.
- Let $Z \in \mathcal{S}_{X,Y}(\mathcal{H})$.
- By applying the theorem, we have that $Z \in \mathcal{A}^{\mathrm{MIN}}_{X,Y}(I, R, G)$ belongs to the Pareto frontier if and only if

$$\mathcal{C}_Y(\mathcal{H} - Z') \subseteq \mathcal{C}_Y(\mathcal{H} - Z) \implies |Z| < |Z'|.$$

for every $Z' \in \mathcal{A}_{X,Y}(I, R, G)$.

Pareto Frontier of Adjustment Sets $\equiv$ *Important* Separators

Definition (Important Separator Marx 2011)

Let $S \subseteq V(\mathcal{H})$. We say that S is an important X, Y-separator if $S \in \mathcal{S}_{X,Y}(\mathcal{H})$, and for any other $S' \in \mathcal{S}_{X,Y}(\mathcal{H})$ it holds that:

$$\mathcal{C}_Y(\mathcal{H}-S') \subset \mathcal{C}_Y(\mathcal{H}-S) \implies |S'| > |S|$$

$$\Big\downarrow \text{Ingredients } 1 + 2$$

$S \in \mathcal{A}^{\mathrm{MIN}}_{X,Y}(I, R, G)$ is Pareto optimal $\Leftrightarrow$ $S \in \mathcal{S}_{X,Y}(\mathcal{H})$ and important.

Pareto Frontier of Adjustment Sets $\equiv$ *Important* Separators

$S \in \mathcal{A}^{\mathrm{MIN}}_{X,Y}(I, R, G)$ is Pareto optimal $\Leftrightarrow$ $S \in \mathcal{S}_{X,Y}(\mathcal{H})$ and important.

Theorem (Marx 2011)

There are at most 4^k important X, Y-separators of $\mathcal{H}$ whose size is at most k, and they can all be generated in time $O(4^k \cdot k \cdot (n + m))$.

Pareto Frontier of Adjustment Sets $\equiv$ *Important* Separators

$S \in \mathcal{A}^{\mathrm{MIN}}_{X,Y}(I, R, G)$ is Pareto optimal $\Leftrightarrow S \in \mathcal{S}_{X,Y}(\mathcal{H})$ and important.

Theorem (Marx 2011)

There are at most 4^k important X, Y-separators of $\mathcal{H}$ whose size is at most k, and they can all be generated in time $O(4^k \cdot k \cdot (n + m))$.

$\downarrow$

Theorem

The Pareto-Optimal minimal adjustment sets of size at most k, can be generated in total time $O(4^k \cdot k \cdot (n + m))$.

Is This Enough ?

- This approach works only if k is small.
- Variable weights bounded by constant c.
- Not really ranked: if $k = 10$, then adjustment sets of size 10 may be returned before those of size 3.

Is This Enough ?

- This approach works only if k is small.
- Variable weights bounded by constant c.
- Not really ranked: if $k = 10$, then adjustment sets of size 10 may be returned before those of size 3.

Theorem

Let G be a causal DAG with weight function $w : V(G) \to \mathbb{N}_{\geq 1}$, and $X, Y \subseteq V(G)$ disjoint. There is an algorithm that enumerates all valid adjustment sets for (X, Y) in order of **non-decreasing total weight**, *breaking ties by $\leq_G^\sigma$ (i.e., proximity to Y). The delay is $O(Kn \cdot m^{1+o(1)})$, where K is the size of the largest set listed.*

Conclusion & Outlook

- Our enumeration algorithms generate a comprehensive (pareto-optimal) set of valid, cost-constrained adjustment sets.
- These outputs have the potential to serve as **training data for ML models** that predict high-quality adjustment sets.
- This approach enables data-driven discovery of patterns connecting graphical structure to estimator variance.
  - By analyzing these patterns, we may gain new insights into how the position and composition of adjustment sets affect estimator variance.

Thank you!

References I

[1] Dániel Marx. Important separators and parameterized algorithms. In Petr Kolman and Jan Kratochvíl, editors, *Graph-Theoretic Concepts in Computer Science - 37th International Workshop, WG 2011, Teplá Monastery, Czech Republic, June 21-24, 2011. Revised Papers*, volume 6986 of *Lecture Notes in Computer Science*, pages 5–10. Springer, 2011. doi: 10.1007/978-3-642-25870-1_2. URL https://doi.org/10.1007/978-3-642-25870-1_2.

[2] Andrea Rotnitzky. Optimal adjustment sets in non-parametric graphical causal models. Talk presented at the Online Causal Inference Seminar, apr 2021. URL https://drive.google.com/file/d/1WDfsU2dyaVrodyYi91nljBiwQogd1uAZ/view. Slide 27.

References II

[3] Jakob Runge. Necessary and sufficient graphical conditions for optimal adjustment sets in causal graphical models with hidden variables. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 15762–15773, 2021. URL `https://proceedings.neurips.cc/paper/2021/hash/8485ae387a981d783f8764e508151cd9-Abstract.html`.

[4] E Smucler, F Sapienza, and A Rotnitzky. Efficient adjustment sets in causal graphical models with hidden variables. *Biometrika*, 109(1):49–65, 03 2021. ISSN 1464-3510. doi: 10.1093/biomet/asab018. URL `https://doi.org/10.1093/biomet/asab018`.

References III

[5] Ezequiel Smucler and Andrea Rotnitzky. A note on efficient minimum cost adjustment sets in causal graphical models. *CoRR*, abs/2201.02037, 2022. URL https://arxiv.org/abs/2201.02037.

[6] Benito van der Zander, Maciej Liskiewicz, and Johannes Textor. Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework. *Artif. Intell.*, 270:1–40, 2019. doi: 10.1016/J.ARTINT.2018.12.006. URL https://doi.org/10.1016/j.artint.2018.12.006.