From Non-Identification to Causal Answers

Designing the Right Experiments

Sina Akbari

27 October 2026

Statistical Laboratory, University of Cambridge

Foundations of AI Rising Stars Workshop, Warwick

Four Papers, One Research Programme

Optimal experiment design

ICML 2022 Outstanding Paper Runner-Up **JMLR 2025**

Graphical characterization and algorithms for minimum-cost experiments.

with Jalal Etesami and Negar Kiyavash

Fast proxy experiment design

NeurIPS 2024

SAT/ILP reformulations and a polynomial-time generalized-adjustment surrogate.

with Sepehr Elahi, Jalal Etesami, Negar Kiyavash, and Patrick Thiran

Sharp analytical causal bounds

UAI 2026 Best Paper Award

Sharp IV bounds, fundamental complexity, and an output-optimal construction.

with Arefe Boushehrian, Mohammad Reza Badri, and Negar Kiyavash

Causal bounds with neural generative models

New work

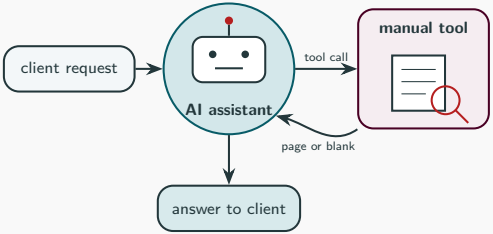
Differentiable neural causal models and extrapolation for accurate causal bounds.

with Max Zhu, Srivathsan Amruth, and Qingyuan Zhao

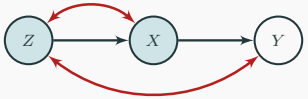
Motivation

An AI Support Assistant Under Time Pressure

Helping clients troubleshoot software and devices



Conditioning on the client request

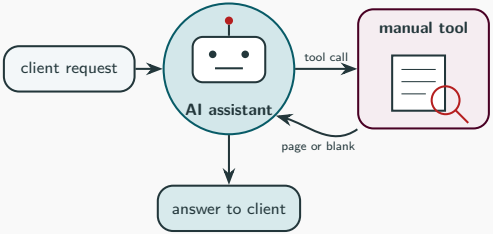


$Z \leftrightarrow X$ congestion / time pressure

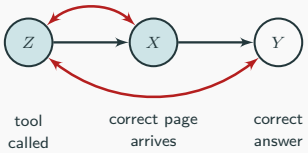
$Z \leftrightarrow Y$ model uncertainty

An AI Support Assistant Under Time Pressure

Helping clients troubleshoot software and devices



Conditioning on the client request



$Z \leftrightarrow X$ congestion / time pressure

$Z \leftrightarrow Y$ model uncertainty

Expensive ideal experiment

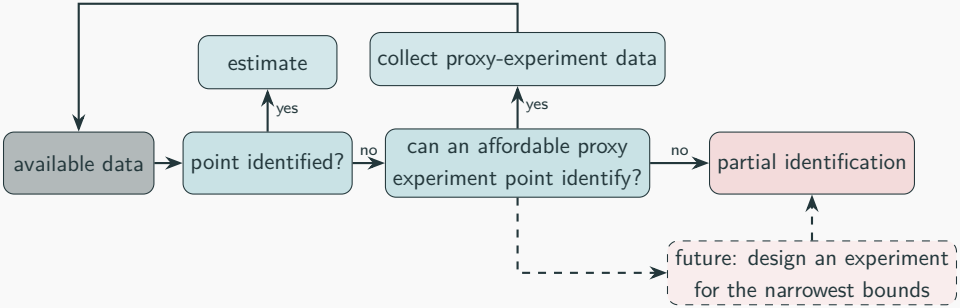
Guarantee the correct manual page.

Cheap proxy experiment

Randomly force or disable the tool call.

Stylized AI application of Bareinboim and Pearl (2012); adaptive retrieval motivation from Ni et al. (2024).

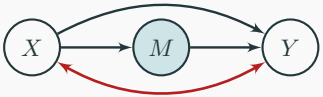
Identification Determines What to Optimize



Technical Setup

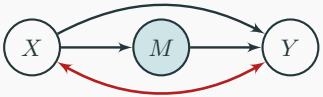
Counterfactual Targets and Experimental Data

Let $\mathcal{G} = (V, E_{\rightarrow}, E_{\leftrightarrow})$ be an ADMG: $u \rightarrow v$ is a causal edge and $u \leftrightarrow v$ represents latent confounding.



Counterfactual Targets and Experimental Data

Let $\mathcal{G} = (V, E_{\rightarrow}, E_{\leftrightarrow})$ be an ADMG: $u \rightarrow v$ is a causal edge and $u \leftrightarrow v$ represents latent confounding.



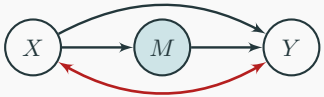
For an intervention $A = a$ and outcome vector B,

$$P_a(b) := P(B(a) = b)\,.$$

The target is $P_x(y) = P(Y(x) = y)$.

Counterfactual Targets and Experimental Data

Let $\mathcal{G} = (V, E_{\rightarrow}, E_{\leftrightarrow})$ be an ADMG: $u \rightarrow v$ is a causal edge and $u \leftrightarrow v$ represents latent confounding.



For an intervention $A = a$ and outcome vector B,

$$P_a(b) := P(B(a) = b)\,.$$

The target is $P_x(y) = P(Y(x) = y)$.

Available information

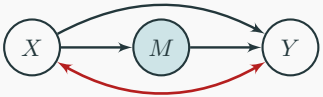
P_A denotes the full experimental kernel generated by intervening on A.

$$\boxed{\{P_A : A \in \mathcal{A}\}}$$

with $P_{\varnothing} = P(V)$.

Counterfactual Targets and Experimental Data

Let $\mathcal{G} = (V, E_{\rightarrow}, E_{\leftrightarrow})$ be an ADMG: $u \rightarrow v$ is a causal edge and $u \leftrightarrow v$ represents latent confounding.



For an intervention $A = a$ and outcome vector B,

$$P_a(b) := P(B(a) = b)\,.$$

The target is $P_x(y) = P(Y(x) = y)$.

Available information

P_A denotes the full experimental kernel generated by intervening on A.

$$\boxed{\{P_A : A \in \mathcal{A}\}}$$

with $P_{\varnothing} = P(V)$.

Identification

$P_x(y)$ is identified if it has the same value in every causal model compatible with $\mathcal{G}$ and the available information.

Minimum-Cost Intervention Design

Assign each variable an intervention cost $C(v) \in [0, \infty]$; $C(v) = \infty$ rules out intervention on v.

Minimum-Cost Intervention Design

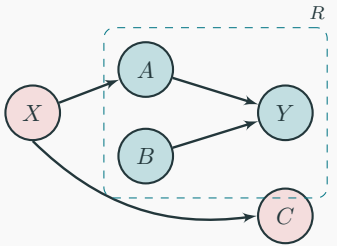
Assign each variable an intervention cost $C(v) \in [0, \infty]$; $C(v) = \infty$ rules out intervention on v.

Design problem

$$\min_{\mathcal{A}} \sum_{A \in \mathcal{A}} \sum_{v \in A} C(v) \qquad \text{s.t.} \qquad P_x(y) \text{ is identified from } \{P_A : A \in \mathcal{A}\}.$$

Akbari, Etesami and Kiyavash, JMLR 2025.

Reduce the Target to Its Relevant Ancestors



$R = \{A, B, Y\}$ $\bar{R} = \{X, C\}$

Remove X, then retain every remaining ancestor of Y:

$$R = \mathrm{An}_{\mathcal{G}[V \setminus X]}(Y), \qquad \bar{R} = V \setminus R.$$

The corresponding canonical kernel is

$$Q[R](r \mid v_{\bar{R}}) := P_{v_{\bar{R}}}(r).$$

Identifying $P_x(y)$ reduces to identifying $Q[R]$ and then marginalizing.

Target reduction: Shpitser and Pearl (2006); Jaber et al. (2019); Lee et al. (2020); Kivva et al. (2022).

Districts Localize the Identification Problem

A **district** is a maximal set of vertices connected by bidirected paths. Let

$$R = S_1 \dot{\cup} \cdots \dot{\cup} S_r$$

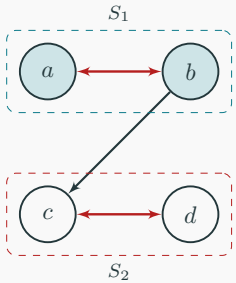
be the districts of the canonical target graph $\mathcal{G}[R]$.

Districts Localize the Identification Problem

A **district** is a maximal set of vertices connected by bidirected paths. Let

$$R = S_1 \dot{\cup} \cdots \dot{\cup} S_r$$

be the districts of the canonical target graph $\mathcal{G}[R]$.



Q-decomposition

The canonical kernel factorizes district by district:

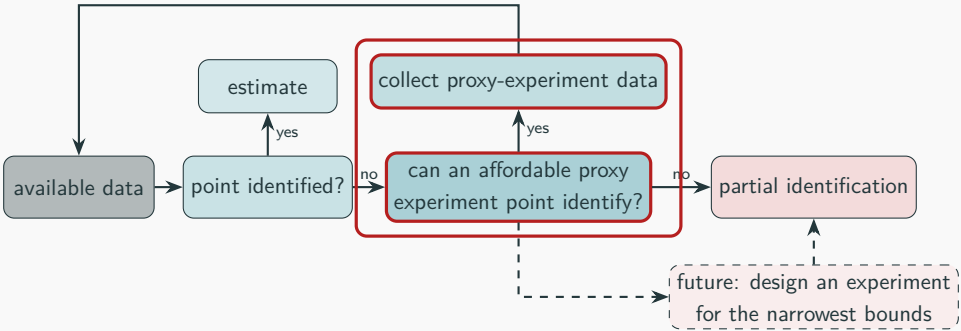
$$Q[R] = \prod_{\ell=1}^{r} Q[S_\ell].$$

Thus $Q[R]$ is identified iff every $Q[S_\ell]$ is.

Q-decomposition: Tian and Pearl, UAI 2002.

Optimal Experiment Design

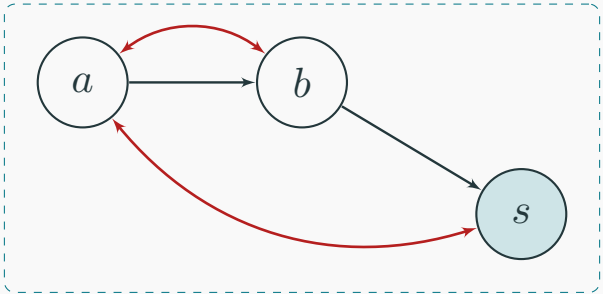
Where We Are: Designing Proxy Experiments



Design question

Which affordable interventions make the target point identified?

A Hedge Is a Graphical Obstruction



$$F = \{a, b, s\}$$

The Design Problem, in Graphical Terms

Minimum-cost intervention design

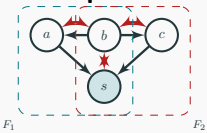
$$\min_{\mathcal{A}} \sum_{A \in \mathcal{A}} \sum_{v \in A} C(v) \qquad \text{s.t.} \qquad \boxed{\text{no hedge for } Q[S] \text{ remains under } \mathcal{A}}.$$

Hedge criterion: Shpitser and Pearl (2006); design formulation: Akbari, Etesami and Kiyavash (2022, 2025).

How Experiments Remove Graphical Obstructions

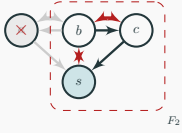
Two hedges may overlap, so one intervention can sometimes break several obstructions at once.

No experiment



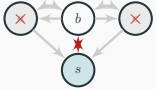
Two overlapping obstructions: $F_1 = \{a, b, s\}$ and $F_2 = \{b, c, s\}$.

Intervene on a



F_1 is broken, but F_2 remains.

Intervene on a and c



No hedge remains: the red edge alone is not a complete obstruction.

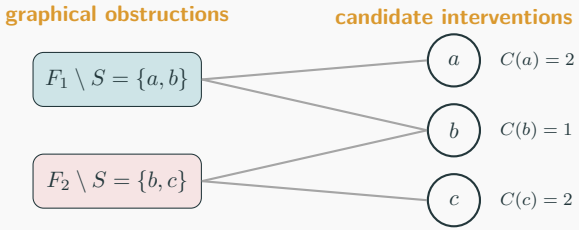
Experiment-design question

We could intervene on a and c separately—or break both obstructions at once by intervening on their shared variable b. Which choice is cheaper?

Hedge criterion: Shpitser and Pearl (2006); intervention-design form: Akbari, Etesami and Kiyavash (2022, 2025).

Hitting Set: Break Every Obstruction at Minimum Cost

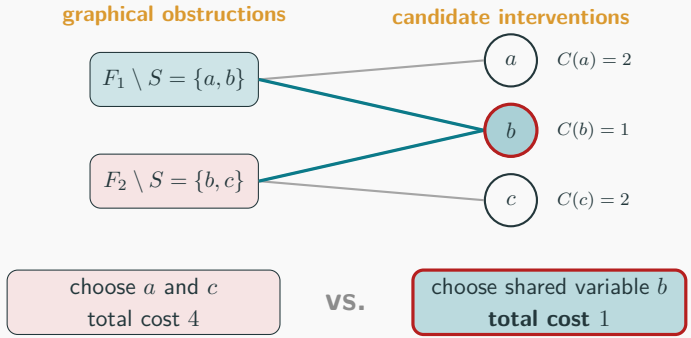
Each hedge contributes a set of variables that can break it. A valid design must touch every set.



Akbari, Etesami and Kiyavash, ICML 2022; JMLR 2025.

Hitting Set: Break Every Obstruction at Minimum Cost

Each hedge contributes a set of variables that can break it. A valid design must touch every set.



Akbari, Etesami and Kiyavash, ICML 2022; JMLR 2025.

The Combinatorial Difficulty Is Fundamental

Exact computation

The optimization problem is **NP-hard**. Its cost-threshold decision version is **NP-complete**, even when all intervention costs are equal.

Approximation

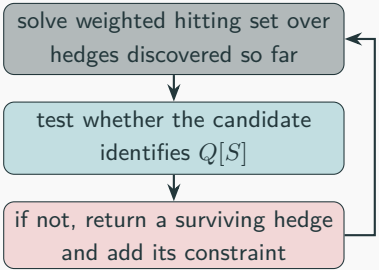
No $(1 - o(1)) \log |V|$ approximation exists under standard complexity assumptions.

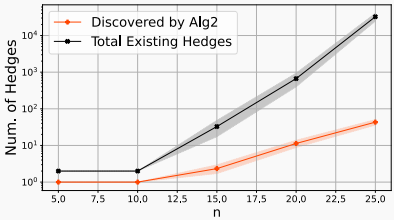
A second source of complexity

An ADMG may contain exponentially many hedges, so generating every constraint before optimization can be prohibitive.

Akbari, Etesami and Kiyavash, JMLR 2025.

Solve One District Without Listing All Hedges

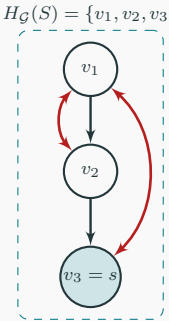




Often only a small fraction of all hedges is discovered (log scale).

Fast Proxy Experiment Design

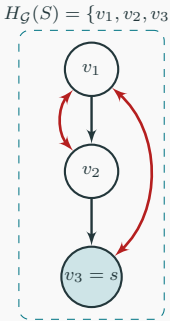
A Three-Node SAT Instance



$$j = 1 \qquad (\neg\rho_{1,0} \vee \rho_{1,1} \vee \neg\rho_{2,1})$$
$$\wedge\, (\neg\rho_{2,0} \vee \rho_{2,1} \vee \neg\rho_{3,1}) \implies (\neg\rho_{2,0} \vee \rho_{2,1})$$

$$S = \{v_3\}, \qquad m = 2,$$
$$\rho_{3,j} \equiv 1.$$

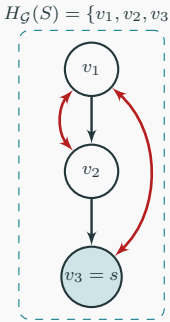
A Three-Node SAT Instance



$$
\begin{aligned}
j = 1 \quad & (\neg\rho_{1,0} \vee \rho_{1,1} \vee \neg\rho_{2,1}) \\
& \wedge (\neg\rho_{2,0} \vee \rho_{2,1} \vee \neg\rho_{3,1}) \implies (\neg\rho_{2,0} \vee \rho_{2,1}) \\
j = 2 \quad & (\neg\rho_{1,1} \vee \rho_{1,2} \vee \neg\rho_{2,2}) \\
& \wedge (\neg\rho_{2,1} \vee \rho_{2,2} \vee \neg\rho_{1,2}) \\
& \wedge (\neg\rho_{1,1} \vee \rho_{1,2} \vee \neg\rho_{3,2}) \implies (\neg\rho_{1,1} \vee \rho_{1,2}) \\
& \wedge (\neg\rho_{3,1} \vee \rho_{3,2} \vee \neg\rho_{1,2}) \implies 1
\end{aligned}
$$

$$S = \{v_3\}, \qquad m = 2,$$
$$\rho_{3,j} \equiv 1.$$

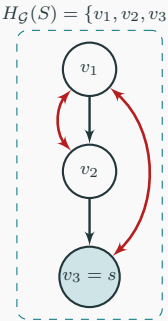
A Three-Node SAT Instance



$$
\begin{aligned}
j = 1 \quad & (\neg\rho_{1,0} \vee \rho_{1,1} \vee \neg\rho_{2,1}) \\
& \wedge (\neg\rho_{2,0} \vee \rho_{2,1} \vee \neg\rho_{3,1}) \implies (\neg\rho_{2,0} \vee \rho_{2,1}) \\
j = 2 \quad & (\neg\rho_{1,1} \vee \rho_{1,2} \vee \neg\rho_{2,2}) \\
& \wedge (\neg\rho_{2,1} \vee \rho_{2,2} \vee \neg\rho_{1,2}) \\
& \wedge (\neg\rho_{1,1} \vee \rho_{1,2} \vee \neg\rho_{3,2}) \implies (\neg\rho_{1,1} \vee \rho_{1,2}) \\
& \wedge (\neg\rho_{3,1} \vee \rho_{3,2} \vee \neg\rho_{1,2}) \implies 1 \\
j = 3 \quad & (\neg\rho_{1,2} \vee \rho_{1,3} \vee \neg\rho_{2,3}) \\
& \wedge (\neg\rho_{2,2} \vee \rho_{2,3} \vee \neg\rho_{3,3}) \implies (\neg\rho_{2,2} \vee \rho_{2,3})
\end{aligned}
$$

$$S = \{v_3\}, \qquad m = 2,$$
$$\rho_{3,j} \equiv 1.$$

A Three-Node SAT Instance



$$
\begin{aligned}
j = 1 \quad & (\neg\rho_{1,0} \vee \rho_{1,1} \vee \neg\rho_{2,1}) \\
& \wedge (\neg\rho_{2,0} \vee \rho_{2,1} \vee \neg\rho_{3,1}) \implies (\neg\rho_{2,0} \vee \rho_{2,1}) \\
j = 2 \quad & (\neg\rho_{1,1} \vee \rho_{1,2} \vee \neg\rho_{2,2}) \\
& \wedge (\neg\rho_{2,1} \vee \rho_{2,2} \vee \neg\rho_{1,2}) \\
& \wedge (\neg\rho_{1,1} \vee \rho_{1,2} \vee \neg\rho_{3,2}) \implies (\neg\rho_{1,1} \vee \rho_{1,2}) \\
& \wedge (\neg\rho_{3,1} \vee \rho_{3,2} \vee \neg\rho_{1,2}) \implies 1 \\
j = 3 \quad & (\neg\rho_{1,2} \vee \rho_{1,3} \vee \neg\rho_{2,3}) \\
& \wedge (\neg\rho_{2,2} \vee \rho_{2,3} \vee \neg\rho_{3,3}) \implies (\neg\rho_{2,2} \vee \rho_{2,3}) \\
j = 4 \quad & \neg\rho_{1,3} \wedge \neg\rho_{2,3}
\end{aligned}
$$

$$S = \{v_3\}, \qquad m = 2,$$
$$\rho_{3,j} \equiv 1.$$

Max-SAT Adds Costs; ILP Reuses the Same Clauses

Weighted partial Max-SAT

Keep the reachability clauses hard. Add the soft clause $\rho_{i,0}$ with weight $C(v_i)$, so violating it means paying to intervene:

$$\min \sum_{i=1}^{m} (1 - \rho_{i,0}) C(v_i)$$

subject to the hard 3-SAT clauses.

Max-SAT Adds Costs; ILP Reuses the Same Clauses

Weighted partial Max-SAT

Keep the reachability clauses hard. Add the soft clause $\rho_{i,0}$ with weight $C(v_i)$, so violating it means paying to intervene:

$$\min \sum_{i=1}^{m} (1 - \rho_{i,0}) C(v_i)$$

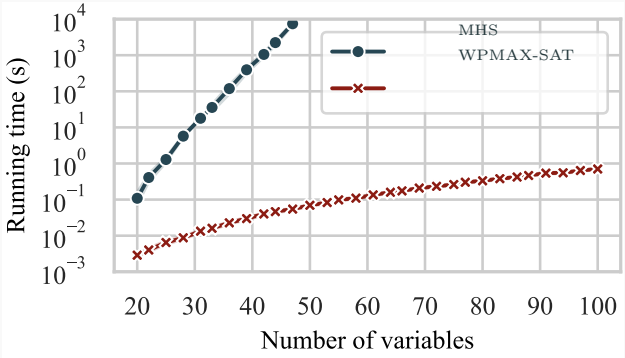
subject to the hard 3-SAT clauses.

Equivalent ILP

Only after forming the logical constraints, map each clause to an inequality; for example,

$$(\neg p \vee q \vee \neg r) \quad \mapsto \quad (1-p) + q + (1-r) \geq 1.$$

The Reformulation Changes the Practical Scale

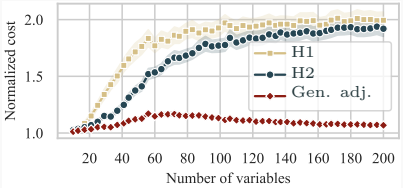


Increasing number of variables

100 random ADMGs per setting; shaded regions are 99% confidence intervals.

A Polynomial-Time Generalized-Adjustment Surrogate

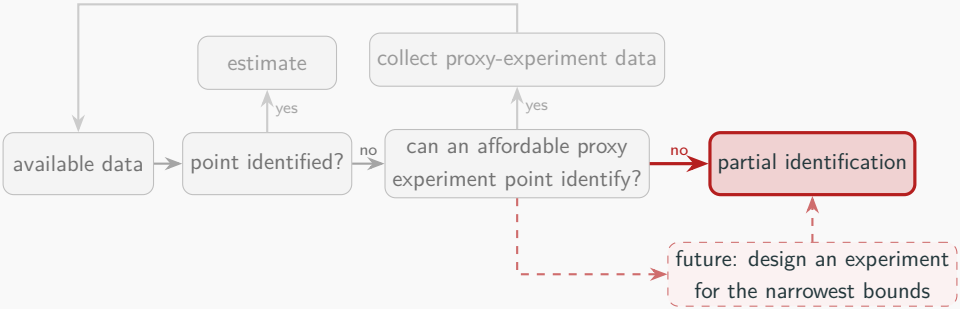
$$P_x(y) = \mathbb{E}_{P_A}[P_A(y \mid x, W)]$$



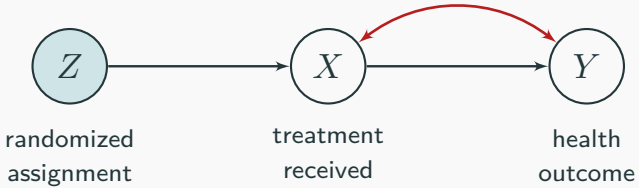
Cost relative to the exact SAT solution

Partial Identification Through Instruments

Where We Are: Point Identification Has Failed



Randomized Trials with Imperfect Adherence



Sharp Bounds Are Linear Programs

Let $\mathbf{q}$ collect the probabilities of all latent response types. Compatibility with the observed distribution is linear:

$$M^\top \mathbf{q} = \mathbf{p}, \qquad \mathbf{q} \geq 0, \qquad \theta = \mathbf{c}^\top \mathbf{q}.$$

Sharp Bounds Are Linear Programs

Let $\mathbf{q}$ collect the probabilities of all latent response types. Compatibility with the observed distribution is linear:

$$M^\top \mathbf{q} = \mathbf{p}, \qquad \mathbf{q} \geq 0, \qquad \theta = \mathbf{c}^\top \mathbf{q}.$$

Primal lower bound

$$L(\mathcal{P}) = \min_{\mathbf{q} \geq 0} \ \mathbf{c}^\top \mathbf{q} \quad \text{s.t.} \quad M^\top \mathbf{q} = \mathbf{p}.$$

Dual representation

$$L(\mathcal{P}) = \max_{\mathbf{v}} \ \mathbf{v}^\top \mathbf{p} \quad \text{s.t.} \quad M\mathbf{v} \leq \mathbf{c}.$$

Sharp Bounds Are Linear Programs

Let $\mathbf{q}$ collect the probabilities of all latent response types. Compatibility with the observed distribution is linear:

$$M^\top \mathbf{q} = \mathbf{p}, \qquad \mathbf{q} \geq 0, \qquad \theta = \mathbf{c}^\top \mathbf{q}.$$

Primal lower bound

$$L(\mathcal{P}) = \min_{\mathbf{q} \geq 0} \ \mathbf{c}^\top \mathbf{q} \quad \text{s.t.} \quad M^\top \mathbf{q} = \mathbf{p}.$$

Dual representation

$$L(\mathcal{P}) = \max_{\mathbf{v}} \ \mathbf{v}^\top \mathbf{p} \quad \text{s.t.} \quad M\mathbf{v} \leq \mathbf{c}.$$

If $\mathcal{V}$ denotes the M-distinct dual vertices, the symbolic sharp bounds are

$$\max_{\mathbf{v} \in \mathcal{V}} \mathbf{v}^\top \mathbf{p} \leq \theta \leq -\max_{\mathbf{v} \in \mathcal{V}} \mathbf{v}^\top \bar{\mathbf{p}}.$$

Boushehrian, Badri, Akbari and Kiyavash, 2026.

Example: Balke–Pearl Bounds for Binary Z, X, Y

Let $p_{yx,z} = P(Y = y, X = x \mid Z = z)$ and $\bar{p}_{yx,z} = p_{y(1-x),z}$. The sharp ATE bounds are

$$L_{\mathrm{BP}}(\mathbf{p}) = \max_{k \le 8} \ell_k \le \theta \le -L_{\mathrm{BP}}(\bar{\mathbf{p}}) = U_{\mathrm{BP}}(\mathbf{p}).$$

$$\begin{aligned}
\ell_1 &= -p_{10,0} - p_{01,0} + p_{11,0} - p_{10,1} - p_{11,1}, \\
\ell_2 &= -p_{10,0} - p_{11,0} - p_{10,1} - p_{01,1} + p_{11,1}, \\
\ell_3 &= -p_{00,0} - p_{10,0} - p_{01,0} - p_{11,0} + p_{00,1} + p_{11,1}, \\
\ell_4 &= -p_{10,0} - p_{01,0} - p_{11,0} + p_{11,1}.
\end{aligned}$$

$$\begin{aligned}
\ell_5 &= -p_{00,0} - p_{10,0} - p_{01,0} + p_{00,1}, \\
\ell_6 &= -p_{10,0} - p_{01,0}, \\
\ell_7 &= -p_{00,0} - p_{01,0} + p_{00,1} - p_{10,1} - p_{01,1}, \\
\ell_8 &= p_{00,0} - p_{10,0} - p_{01,0} - p_{00,1} - p_{01,1}.
\end{aligned}$$

Balke (1995); Balke and Pearl (1997).

Two Open Problems, Solved

1. **Analytical complexity.**

   Any sharp linear representation requires exponentially many terms. For binary Z and an n-level outcome,

   $$|\mathcal{V}| = 5 \cdot 4^{n-1} - 2^{n+2} + 4.$$

   A matching output-optimal algorithm generates all terms in $O(n|\mathcal{V}|)$ time.

Two Open Problems, Solved

1. **Analytical complexity.**

   Any sharp linear representation requires exponentially many terms. For binary Z and an n-level outcome,

   $$|\mathcal{V}| = 5 \cdot 4^{n-1} - 2^{n+2} + 4.$$

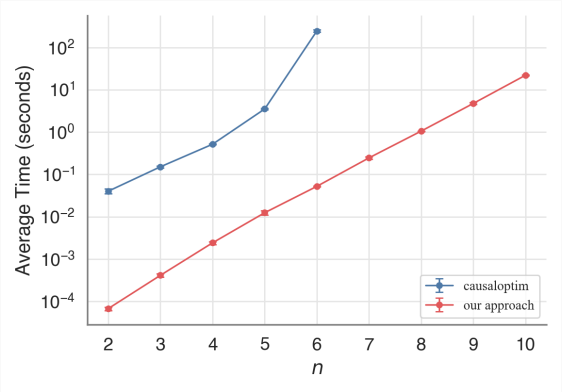
   A matching output-optimal algorithm generates all terms in $O(n|\mathcal{V}|)$ time.

2. **Numerical computation.**

   There is no need to solve a linear program for each observed distribution: evaluate the generated terms directly and take their maximum or minimum, in time linear in the representation size.

Boushehrian, Badri, Akbari and Kiyavash, 2026.

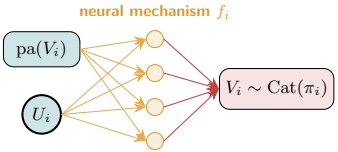
Structure Gives an Output-Optimal Algorithm



Neural Partial Identification

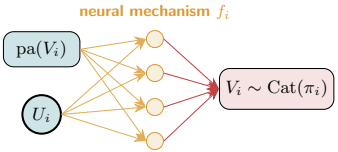
Beyond Analytical Bounds

Flexible, differentiable route



Beyond Analytical Bounds

Flexible, differentiable route



The causal target becomes a constrained learning problem

$$\underline{\theta} = \min_{\eta:\, P_\eta = P_{\mathrm{obs}}} \theta(\eta), \qquad \overline{\theta} = \max_{\eta:\, P_\eta = P_{\mathrm{obs}}} \theta(\eta).$$

Zhu, Amruth, Akbari and Zhao, 2026.

Regularization Traces Two Boundary Branches

Replace the exact compatibility constraint by an energy-distance penalty:

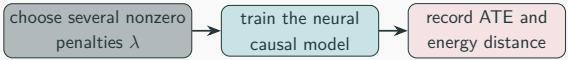
$$J_\lambda(\eta) = L_{\mathrm{ED}}\big(P_\eta, P_{\mathrm{obs}}\big) + \lambda\theta(\eta),$$

Lower branch: $\lambda > 0$

Minimizing the objective penalizes larger ATE values and traces the lower boundary.

Upper branch: $\lambda < 0$

Minimizing the objective rewards larger ATE values and traces the upper boundary.



What regularization gives us

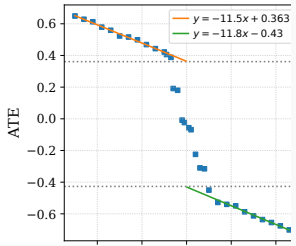
Each nonzero λ gives a point on a lower or upper solution path. The desired bounds are the one-sided endpoints as $\lambda \to 0$.

Zhu, Amruth, Akbari and Zhao, 2026.

The Extrapolation Idea

Near $\lambda = 0$, smooth one-sided solution paths satisfy

$$\theta(\eta_\lambda) = \theta(\eta_0) + c_\pm \lambda + O(\lambda^2).$$



The estimand is locally linear in λ.

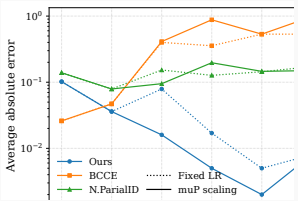
Zhu, Amruth, Akbari and Zhao, 2026.

Extrapolation Improves Bound Accuracy

Compare direct regularized fits with the extrapolated intercepts and the analytical benchmark.

Ternary-IV instance

	lower	upper
analytical	-0.427	0.361
direct fit	-0.450	0.388
extrapolated	$\mathbf{-0.430}$	$\mathbf{0.363}$



Average absolute error across both bounds.

Across the benchmark suite

On four discrete causal settings and ten datasets per setting, extrapolation achieves the lowest absolute error among the neural methods in every upper- and lower-bound comparison.

Zhu, Amruth, Akbari and Zhao, 2026.

References

References

Identification and experiment design

Tian and Pearl (2002). *A General Identification Condition for Causal Effects*.

Shpitser and Pearl (2006). *Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models*.

Bareinboim and Pearl (2012). *Causal Inference by Surrogate Experiments: z-Identifiability*.

Lee, Correa and Bareinboim (2020). *General Identifiability with Arbitrary Surrogate Experiments*.

Jung, Tian and Bareinboim (2023). *Estimating Joint Treatment Effects by Combining Multiple Experiments*.

Akbari, Etesami and Kiyavash (2022, 2025). *Minimum Cost Intervention Design* and *Optimal Experiment Design for Causal Effect Identification*.

Elahi et al. (2024). *Fast Proxy Experiment Design for Causal Effect Identification*.

Partial identification

Balke (1995); Balke and Pearl (1997). *Bounds on Treatment Effects from Studies with Imperfect Compliance*.

Sachs et al. (2023). *A General Method for Deriving Tight Symbolic Bounds on Causal Effects*.

Boushehrian et al. (2026). *Fundamental Limits and Optimal Methods for Sharp Analytical Causal Bounds in Instrumental Variable Models*.

Zhu et al. (2026). *Estimating Causal Bounds from Observational Data with Neural Generative Models*.

Adaptive retrieval motivation

Ni et al. (2024). *When Do LLMs Need Retrieval Augmentation?*

Labruna, Campos and Azkune (2025). *When to Retrieve: Teaching LLMs to Utilize Information Retrieval Effectively*.

Distributional distance

Rizzo and Székely (2016). *Energy Distance*.

Papers and resources

Take Away

1. **Identification is a decision point:** it tells us whether to estimate, experiment, or bound.

Take Away

1. **Identification is a decision point:** it tells us whether to estimate, experiment, or bound.
2. **Experiment design is combinatorial but structured:** graphical obstructions yield exact optimization problems, and SAT/ILP makes them practical at much larger scales.

Take Away

1. **Identification is a decision point:** it tells us whether to estimate, experiment, or bound.
2. **Experiment design is combinatorial but structured:** graphical obstructions yield exact optimization problems, and SAT/ILP makes them practical at much larger scales.
3. **Partial identification admits complementary routes:** exact structure yields sharp analytical bounds, while neural extrapolation provides a flexible alternative.

Papers and resources

Take Away

1. **Identification is a decision point:** it tells us whether to estimate, experiment, or bound.
2. **Experiment design is combinatorial but structured:** graphical obstructions yield exact optimization problems, and SAT/ILP makes them practical at much larger scales.
3. **Partial identification admits complementary routes:** exact structure yields sharp analytical bounds, while neural extrapolation provides a flexible alternative.

Papers and resources

Research programme

Build systems that automatically return the strongest causal conclusion supported by the data, graph, and experimental constraints.