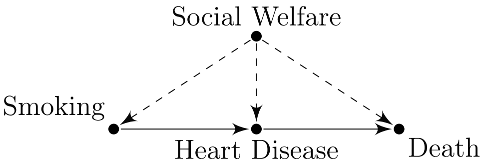
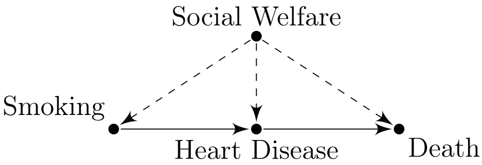
Minimum Cost Intervention Design for Causal Effect Identification

Jalal Etesami,
Sina Akbari, Negar Kiyavash

ICML 2022


Social Welfare
Smoking
Heart Disease
Death

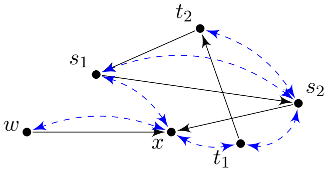


Intervene on Smoking:

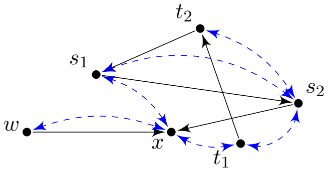
$$P(Death|do(Smoking))$$

do(Smoking=1): Force to smoke
do(Smoking=0): Force to stop smoking

- Causal graph $\mathcal{G}$:



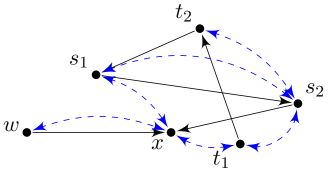
- Causal graph $\mathcal{G}$:



Identifiability

- $\mathcal{G}$ and $P(V)$ are given.
- **Goal:** uniquely compute a causal query $P(S|do(T))$.

- Causal graph $\mathcal{G}$:



Identifiability

- $\mathcal{G}$ and $P(V)$ are given.
- **Goal:** uniquely compute a causal query $P(S|do(T))$.

General identifiability

- $\mathcal{G}$ and a set of distributions $\{P(Y_1|do(A_1)), ..., P(Y_k|do(A_k))\}$ are given.
- **Goal:** uniquely compute a causal query $P(S|do(T))$.

EPFL

What is known?[1,2]

- Let $Y_i := V \setminus A_i$, $\forall i$.
- $\mathcal{G}$ and a set of intervention sets $A_1, ..., A_k$ are given.
- Whether a causal query $P(S|do(T))$ is identifiable from

$$\mathbf{P} := \{P(Y_1|do(A_1)), ..., P(Y_k|do(A_k))\}.$$

[1] Lee et. al., "General identifiability with arbitrary surrogate experiments," UAI 2020.

[2] Kivva et. al., "Revisiting the general identifiability problem," UAI 2022.

What is known?[1,2]

- Let $Y_i := V \setminus A_i, \forall i$.
- $\mathcal{G}$ and a set of intervention sets $A_1, ..., A_k$ are given.
- Whether a causal query $P(S|do(T))$ is identifiable from

$$\mathbf{P} := \{P(Y_1|do(A_1)), ..., P(Y_k|do(A_k))\}.$$

The intervention design problem

- Having $\mathbf{P}$ is costly.
- What is the set $\mathbf{P}$ with minimum cost that identifies $P(S|do(T))$?

[1] Lee et. al., "General identifiability with arbitrary surrogate experiments," UAI 2020.
[2] Kivva et. al., "Revisiting the general identifiability problem," UAI 2022.

Minimum cost intervention design problem

- Causal graph $\mathcal{G}$ is known.
- Two disjoint subsets $S, T \subseteq V$ are given.

Minimum cost intervention design problem

- Causal graph $\mathcal{G}$ is known.
- Two disjoint subsets $S, T \subseteq V$ are given.
- **Goal:** Find a collection $\mathbf{A}^* = \{A_1, ..., A_m\}$ of subsets of V s.t.
  - cost of $\mathbf{A}^*$ is minimum,
  - $P(S|do(T))$ is identifiable form

$$\left\{P\big(V \setminus A_1|do(A_1)\big), ..., P\big(V \setminus A_m|do(A_m)\big)\right\}.$$

EPFL

Minimum cost intervention design problem

- Causal graph $\mathcal{G}$ is known.
- Two disjoint subsets $S, T \subseteq V$ are given.
- **Goal:** Find a collection $\mathbf{A}^* = \{A_1, ..., A_m\}$ of subsets of V s.t.
    - cost of $\mathbf{A}^*$ is minimum,
    - $P(S|do(T))$ is identifiable form

$$\big\{P\big(V \setminus A_1|do(A_1)\big), ..., P\big(V \setminus A_m|do(A_m)\big)\big\}.$$

Assumption

Cost function is additive, i.e., $\mathbf{C}(\cdot) : V \to \mathbb{R}^{\geq 0}$, *and*

$$\mathbf{C}(A_i) = \sum_{a \in A_i} \mathbf{C}(a).$$

EPFL

Definition: Let $\mathbf{ID}_{\mathcal{G}}(S,T)$ denote the set of all collections of subsets of V, e.g., $\mathbf{A} = \{A_1, ..., A_m\}$, where $A_i \subseteq V$, s.t.

- $P(S|do(T))$ is identifiable in $\mathcal{G}$ from

$$\left\{P\big(V \setminus A_1|do(A_1)\big), ..., P\big(V \setminus A_m|do(A_m)\big)\right\}.$$

Note: $|\mathbf{ID}_{\mathcal{G}}(S,T)| \leq 2^{2^{|V|}}$.

Definition: Let $\mathbf{ID}_{\mathcal{G}}(S,T)$ denote the set of all collections of subsets of V, e.g., $\mathbf{A} = \{A_1, ..., A_m\}$, where $A_i \subseteq V$, s.t.

- $P(S|do(T))$ is identifiable in $\mathcal{G}$ from

$$\left\{P\big(V \setminus A_1|do(A_1)\big), ..., P\big(V \setminus A_m|do(A_m)\big)\right\}.$$

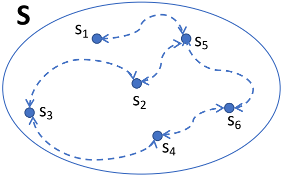
Note: $|\mathbf{ID}_{\mathcal{G}}(S,T)| \leq 2^{2^{|V|}}$.

Problem:

$$\mathbf{A}^*_{S,T} \in \arg\min_{\mathbf{A}\in\mathbf{ID}_{\mathcal{G}}(S,T)} \sum\nolimits_{A\in\mathbf{A}} \mathbf{C}(A). \tag{1}$$

Definition (C-component)

$\mathcal{G}_{[S]}$ is a c-component.



Theorem

- *Let* $\mathbf{A} = \{A_1, ..., A_m\}_{m>1}$ *be a member of* $\boldsymbol{ID}_{\mathcal{G}}(S, T)$.
- *Suppose* S *is a subset of variables s.t.* $\mathcal{G}_{[S]}$ *is a c-component.*

Then,

$$\exists \tilde{A} \subseteq V \ \ s.t. \ \ \tilde{\mathbf{A}} = \{\tilde{A}\} \in \boldsymbol{ID}_{\mathcal{G}}(S, T) \ \ and \ \mathbf{C}(\tilde{\mathbf{A}}) \leq \mathbf{C}(\mathbf{A}).$$

Theorem

- *Let* $\mathbf{A} = \{A_1, ..., A_m\}_{m>1}$ *be a member of* $\boldsymbol{ID}_{\mathcal{G}}(S,T)$.
- *Suppose* S *is a subset of variables s.t.* $\mathcal{G}_{[S]}$ *is a c-component.*

Then,

$$\exists \tilde{A} \subseteq V \ \ s.t. \ \ \tilde{\mathbf{A}} = \{\tilde{A}\} \in \boldsymbol{ID}_{\mathcal{G}}(S,T) \ \ and \ \mathbf{C}(\tilde{\mathbf{A}}) \leq \mathbf{C}(\mathbf{A}).$$

- *More precisely,* $\tilde{A} := \bigcup_{i=1}^{m} A_i \setminus S$.

EPFL

Theorem

- *Let* $\mathbf{A}=\{A_1,...,A_m\}_{m>1}$ *be a member of* $\boldsymbol{ID}_{\mathcal{G}}(S,T)$.
- *Suppose* S *is a subset of variables s.t.* $\mathcal{G}_{[S]}$ *is a c-component.*

Then,

$$\exists \tilde{A} \subseteq V \ \ s.t. \ \ \tilde{\mathbf{A}} = \{\tilde{A}\} \in \boldsymbol{ID}_{\mathcal{G}}(S,T) \ \ and \ \mathbf{C}(\tilde{\mathbf{A}}) \leq \mathbf{C}(\mathbf{A}).$$

- *More precisely,* $\tilde{A} := \bigcup_{i=1}^{m} A_i \setminus S$.

Exponential Formulation

$$A^*_{S,T} \in \arg \min_{A \in \mathbf{ID}_1(S,T)} \sum_{a \in A} \mathbf{C}(a). \qquad (2)$$

Theorem

- *Let* $\mathbf{A} = \{A_1, ..., A_m\}_{m>1}$ *be a member of* $\boldsymbol{ID}_{\mathcal{G}}(S,T)$.
- *Suppose* S *is a subset of variables s.t.* $\mathcal{G}_{[S]}$ *is a c-component.*

Then,

$$\exists \tilde{A} \subseteq V \;\; s.t. \;\; \tilde{\mathbf{A}} = \{\tilde{A}\} \in \boldsymbol{ID}_{\mathcal{G}}(S,T) \;\; and \;\; \mathbf{C}(\tilde{\mathbf{A}}) \leq \mathbf{C}(\mathbf{A}).$$

- *More precisely,* $\tilde{A} := \bigcup_{i=1}^{m} A_i \setminus S$.

Exponential Formulation

$$A^*_{S,T} \in \arg \min_{A \in \mathbf{ID}_1(S,T)} \sum_{a \in A} \mathbf{C}(a). \tag{2}$$

Note: $|\mathbf{ID}_1(S,T)| \leq 2^{|V|}$.

Can we do better?

Theorem

There exists a polynomial-time reduction from the minimum vertex cover problem to the minimum cost intervention design problem.

Theorem

There exists a polynomial-time reduction from the minimum vertex cover problem to the minimum cost intervention design problem.

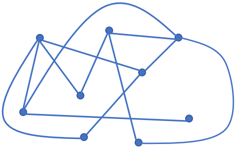


Figure: Minimum Vertex Cover (MVC)

Theorem

There exists a polynomial-time reduction from the minimum vertex cover problem to the minimum cost intervention design problem.

Theorem

There exists a polynomial-time reduction from the minimum vertex cover problem to the minimum cost intervention design problem.

Corollary

Minimum cost intervention design problem is NP-hard.

Theorem

There exists a polynomial-time reduction from the minimum vertex cover problem to the minimum cost intervention design problem.

Corollary

Minimum cost intervention design problem is NP-hard.

Remark

- *Let* $\mathbf{C}(v) = 1, \; \forall v \in V$.
- *The problem is still NP-hard.*

Definition: Let $Q[S]$ denotes the causal effect of $do(V \setminus S)$ on S,

$$Q[S] := P(S|do(V \setminus S)).$$

[3] Shpitser and Pearl 2006, Jaber et. al. 2019, Kivva et. al. 2022, etc.

Definition: Let $Q[S]$ denotes the causal effect of $do(V \setminus S)$ on S,

$$Q[S] := P(S|do(V \setminus S)).$$

- $P(S|do(T))$ is identifiable in $\mathcal{G}$ iff $Q[\mathbf{Anc}_{\mathcal{G}\setminus T}(S)]$ is identifiable[3].
  - $\mathbf{Anc}_{\mathcal{G}\setminus T}(S)$ are ancestors of S in $\mathcal{G}$ after deleting T.

[3] Shpitser and Pearl 2006, Jaber et. al. 2019, Kivva et. al. 2022, etc.

Definition: Let $Q[S]$ denotes the causal effect of $do(V \setminus S)$ on S,

$$Q[S] := P(S|do(V \setminus S)).$$

- $P(S|do(T))$ is identifiable in $\mathcal{G}$ iff $Q[\mathbf{Anc}_{\mathcal{G}\setminus T}(S)]$ is identifiable[3].
  - $\mathbf{Anc}_{\mathcal{G}\setminus T}(S)$ are ancestors of S in $\mathcal{G}$ after deleting T.

Original problem:

$$A^*_{S,T} \in \arg \min_{A \in \mathbf{ID}_1(S,T)} \sum_{a \in A} \mathbf{C}(a).$$

[3] Shpitser and Pearl 2006, Jaber et. al. 2019, Kivva et. al. 2022, etc.

Definition: Let $Q[S]$ denotes the causal effect of $do(V \setminus S)$ on S,

$$Q[S] := P(S|do(V \setminus S)).$$

- $P(S|do(T))$ is identifiable in $\mathcal{G}$ iff $Q[\mathbf{Anc}_{\mathcal{G}\setminus T}(S)]$ is identifiable[3].
  - $\mathbf{Anc}_{\mathcal{G}\setminus T}(S)$ are ancestors of S in $\mathcal{G}$ after deleting T.

Original problem:

$$A^*_{S,T} \in \arg \min_{A \in \mathbf{ID}_1(S,T)} \sum_{a \in A} \mathbf{C}(a).$$

Simplified problem: We can assume $T = V \setminus S$.

$$A^*_{S} \in \arg \min_{A \in \mathbf{ID}_1(S,V\setminus S)} \sum_{a \in A} \mathbf{C}(a).$$

[3] Shpitser and Pearl 2006, Jaber et. al. 2019, Kivva et. al. 2022, etc.

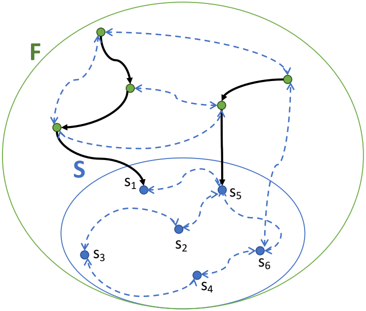
$$\arg\min_{\mathbf{A}\in\mathbf{ID}_{\mathcal{G}}(S,T)} \sum_{A\in\mathbf{A}} \mathbf{C}(A).$$

$$\arg\min_{A\in\mathbf{ID}_1(S,T)} \sum_{a\in A} \mathbf{C}(a).$$

$$\arg\min_{A\in\mathbf{ID}_1(S,V\setminus S)} \sum_{a\in A} \mathbf{C}(a).$$

Definition (Hedge)

- Let S be a subset of V s.t. $\mathcal{G}_{[S]}$ is a c-component in $\mathcal{G}$. Subset $F \subseteq V$ forms a hedge for $Q[S]$ if
  - $S \subsetneq F$,
  - F is the set of ancestors of S in $\mathcal{G}_{[F]}$,
  - $\mathcal{G}_{[F]}$ is a c-component.



EPFL

Definition

- Let S be a subset of V s.t. $\mathcal{G}_{[S]}$ is a c-component in $\mathcal{G}$. Subset $F \subseteq V$ forms a hedge for $Q[S]$ if
    - $S \subsetneq F$,
    - F is the set of ancestors of S in $\mathcal{G}_{[F]}$,
    - $\mathcal{G}_{[F]}$ is a c-component.

EPFL

Definition

- Let S be a subset of V s.t. $\mathcal{G}_{[S]}$ is a c-component in $\mathcal{G}$. Subset $F \subseteq V$ forms a hedge for $Q[S]$ if
  - $S \subsetneq F$,
  - F is the set of ancestors of S in $\mathcal{G}_{[F]}$,
  - $\mathcal{G}_{[F]}$ is a c-component.

Fact

- *Let S be a subset of V s.t. $\mathcal{G}_{[S]}$ is a c-component.*
- *Let $\{F_1, ..., F_m\}$ denotes the set of all hedges of $Q[S]$ in $\mathcal{G}$.*

Definition

- Let S be a subset of V s.t. $\mathcal{G}_{[S]}$ is a c-component in $\mathcal{G}$. Subset $F \subseteq V$ forms a hedge for $Q[S]$ if
  - $S \subsetneq F$,
  - F is the set of ancestors of S in $\mathcal{G}_{[F]}$,
  - $\mathcal{G}_{[F]}$ is a c-component.

Fact

- *Let S be a subset of V s.t. $\mathcal{G}_{[S]}$ is a c-component.*
- *Let $\{F_1, ..., F_m\}$ denotes the set of all hedges of $Q[S]$ in $\mathcal{G}$.*
- *Then, $A_S \in \boldsymbol{ID}_1(S, V \setminus S)$ iff*

$$A_S \cap (F_i \setminus S) \neq \emptyset, \ \forall i.$$

EPFL

Definition

- Let S be a subset of V s.t. $\mathcal{G}_{[S]}$ is a c-component in $\mathcal{G}$. Subset $F \subseteq V$ forms a hedge for $Q[S]$ if
  - $S \subsetneq F$,
  - F is the set of ancestors of S in $\mathcal{G}_{[F]}$,
  - $\mathcal{G}_{[F]}$ is a c-component.

Fact

- *Let S be a subset of V s.t. $\mathcal{G}_{[S]}$ is a c-component.*
- *Let $\{F_1, ..., F_m\}$ denotes the set of all hedges of $Q[S]$ in $\mathcal{G}$.*

Definition

- Let S be a subset of V s.t. $\mathcal{G}_{[S]}$ is a c-component in $\mathcal{G}$. Subset $F \subseteq V$ forms a hedge for $Q[S]$ if
  - $S \subsetneq F$,
  - F is the set of ancestors of S in $\mathcal{G}_{[F]}$,
  - $\mathcal{G}_{[F]}$ is a c-component.

Fact

- *Let S be a subset of V s.t. $\mathcal{G}_{[S]}$ is a c-component.*
- *Let $\{F_1, ..., F_m\}$ denotes the set of all hedges of $Q[S]$ in $\mathcal{G}$.*

- *Then A_S^* is a solution to the simplified problem iff it is a solution to the MWHS problem for the sets $\{F_1 \setminus S, ..., F_m \setminus S\}$, with the weight function $\omega(\cdot) := \mathbf{C}(\cdot)$.*

(I) Enumerate the hedges F_i, (II) Solve the MWHS problem.

Algorithm 1: Min-cost intervention$(S, \mathcal{G})$.

F $\leftarrow \emptyset, \quad H \leftarrow Hhull(S, \mathcal{G}_{[V \setminus \mathbf{pa}^{\leftrightarrow}(S)]})$
if $Q[S]$ is ID **return** $\mathbf{pa}^{\leftrightarrow}(S)$
while True **do**
  while True **do**
    $a \leftarrow \arg\min_{a \in H \setminus S} \mathbf{C}(a)$
    if $Q[S]$ is ID in $\mathcal{G}_{[H \setminus \{a\}]}$ **then**
      $\mathbf{F} \leftarrow \mathbf{F} \cup \{H\}$
      break
    else
      $H \leftarrow Hhull(S, \mathcal{G}_{[H \setminus \{a\}]})$
  $A \leftarrow$ solve min hitting set for $\{F \setminus S | F \in \mathbf{F}\}$
  if $A \cup \mathbf{pa}^{\leftrightarrow}(S) \in \mathbf{ID_1}(S)$ **then**
    return $(A \cup \mathbf{pa}^{\leftrightarrow}(S))$
  $H \leftarrow Hhull(S, \mathcal{G}_{[V \setminus (A \cup \mathbf{pa}^{\leftrightarrow}(S))]})$

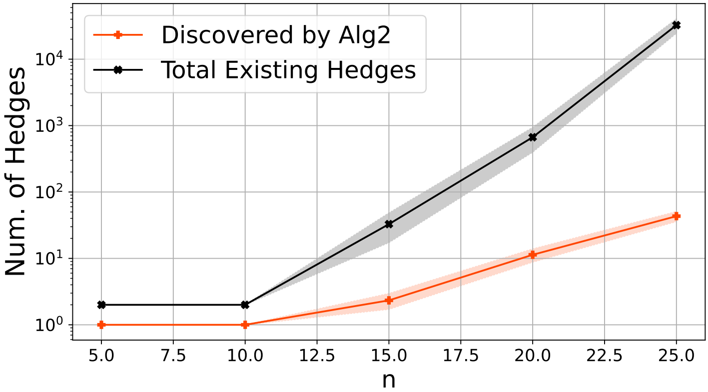
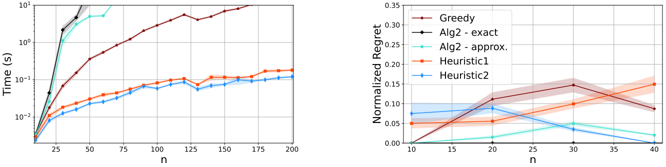
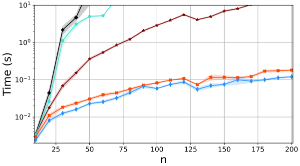


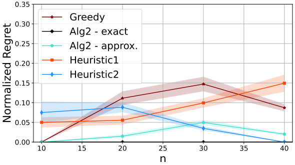
Figure: Number of hedges formed for $Q[S]$.

- Heuristic algorithms: A few of them discussed in the paper.



- Heuristic algorithms: A few of them discussed in the paper.

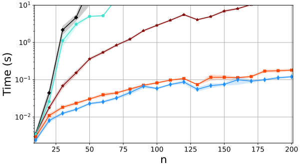


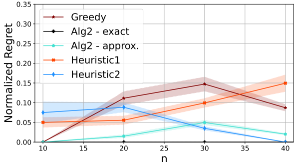


- Structural side information:
  - If $\mathcal{G}$ is tree, polynomial algorithms exist.

Tackling the Computational Complexity

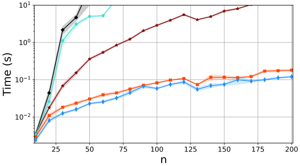
- Heuristic algorithms: A few of them discussed in the paper.

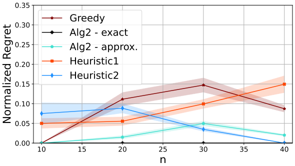




- Structural side information:
  - If $\mathcal{G}$ is tree, polynomial algorithms exist.
- Additional side information:
  - Under certain properties of the $\mathbf{C}(\cdot)$, the problem can be simplified.

- Heuristic algorithms: A few of them discussed in the paper.





- Structural side information:
  - If $\mathcal{G}$ is tree, polynomial algorithms exist.
- Additional side information:
  - Under certain properties of the $\mathbf{C}(\cdot)$, the problem can be simplified.
- *Future work:* Approximation algorithms.

THANK YOU...