RCPC: A Sound Causal Discovery Algorithm under Orientation Unfaithfulness

Kenneth Lee, Murat Kocaoglu

Elmore Family School of Electrical and Computer Engineering, Purdue University

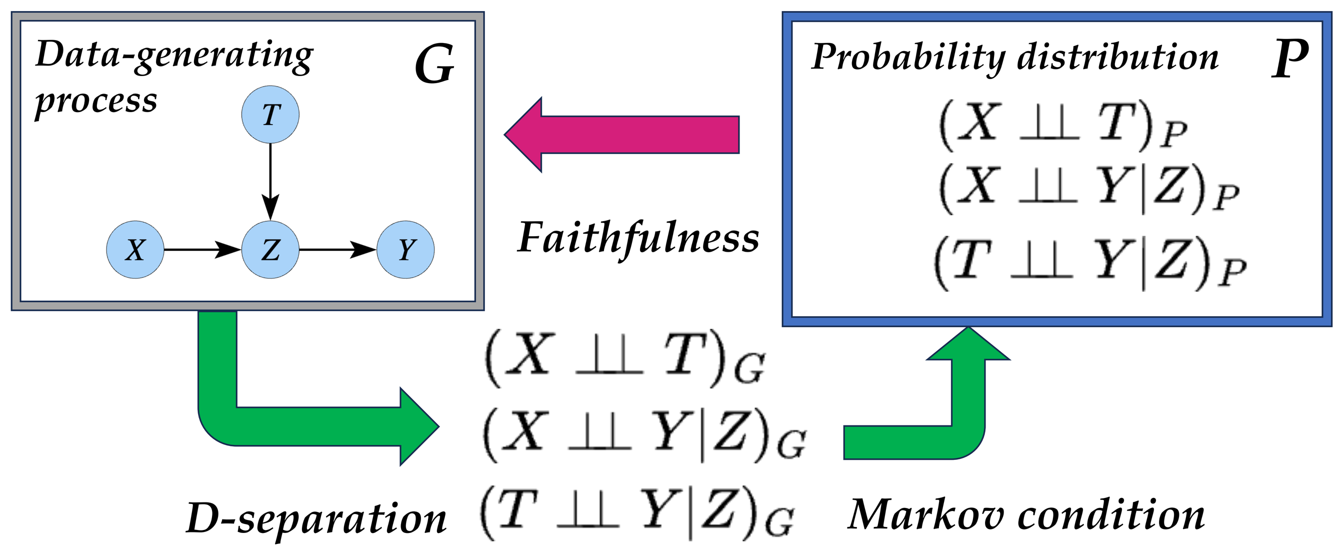
Constraint-based causal discovery approaches often rely on a strong assumption known as **faithfulness.** A conservative PC (CPC) that relies on a weaker assumption called adjacency faithfulness has been proposed. **CPC is conjectured to be complete**. We show that the CPC algorithm is not complete and propose two additional sound orientation rules.

Probabilistic Causal Inference Fundamentals

- *Pearlian framework* [Pearl' 09]: **Directed acyclic graphs (DAGs)** encode causal relation between variables.
- **Arrows:** Deterministic functional relations called *structural equations.*

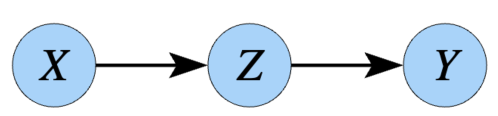
$$X_i = f(Pa_{X_i}, E_{X_i}),\ E_i \perp\!\!\!\perp E_j$$

- **D-separation**
  - In a DAG, a path p between vertices X and Y is **active (d-connecting)** relative to a set of vertices Z if
    - Every non-collider on p is not a member of Z
    - Every collider on p is an ancestor of some members of Z
- **D-separation and Conditional Independence**



- **Adjacency faithfulness**
  - If X and Y are adjacent in G, they are conditionally dependent given any subsets of V.
- **Orientation faithfulness**
  - Let $\langle X, Y, Z\rangle$ be any unshielded triple in G.
    - If $X \rightarrow Y \leftarrow Z$, then X and Z are dependent given any subset that contains Y;
    - Otherwise, X and Z are dependent conditional on any subset that does not contain Y

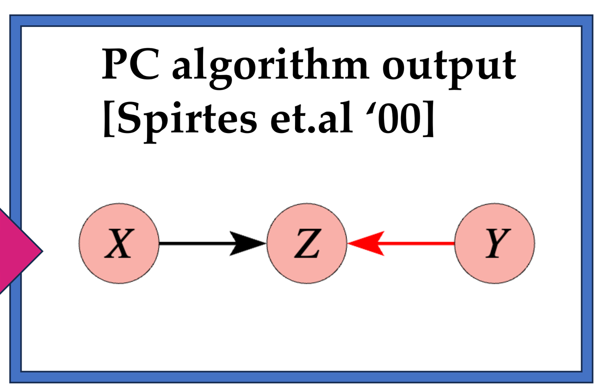
Example (Adj. faithfulness holds with orientation unfaithfulness)



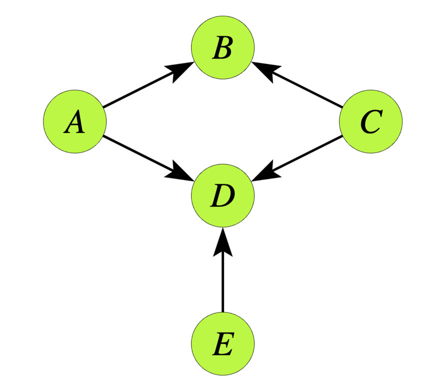
Observed from data:

$(X \perp\!\!\!\perp Y)_P$

$(X \perp\!\!\!\perp Y|Z)_P$



CPC Algorithm [Ramsey et. al'12]



True DAG

Observed CI statements:x

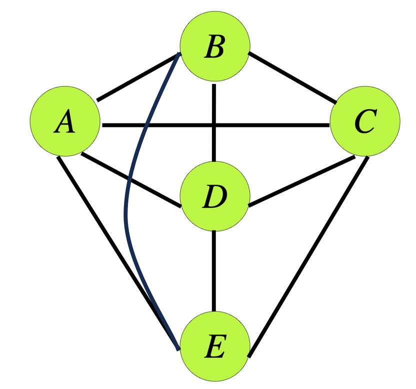
$(A \perp\!\!\!\perp E)_P \quad (E \perp\!\!\!\perp B)_P$

$(A \perp\!\!\!\perp C)_P \quad (E \perp\!\!\!\perp C)_P$

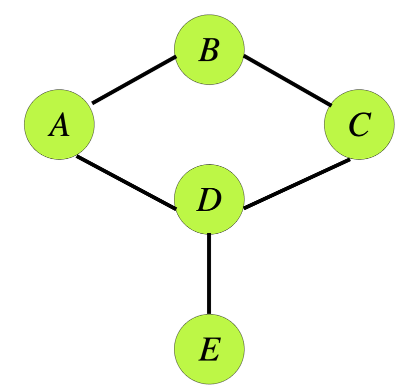
$(A \not\perp\!\!\!\perp C|B)_P$

$(A \perp\!\!\!\perp C|B, D)_P$

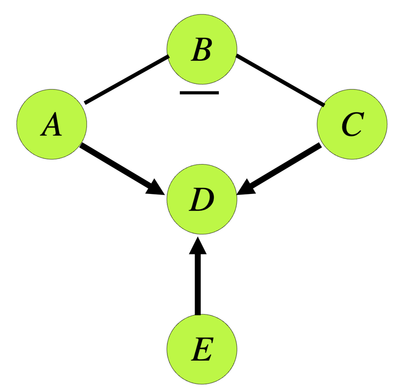
$(B \perp\!\!\!\perp D|A, C)_P$



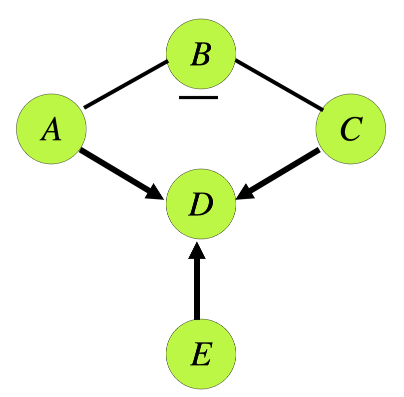
Step 1: Start from a complete graph



Step 2: Remove edges based on conditional independence



Step 3: For each unshielded triple $\langle X, Y, Z\rangle$,
- if Y is not in any separating set, orient $X \rightarrow Y \leftarrow Z$.
- If Y is in all separating sets, orient $X - Y - Z$.
- Otherwise, mark $X - \underline{Y} - Z$



Step 4: Apply Meek rules to unshielded non-colliders, not including triples that are marked unfaithful

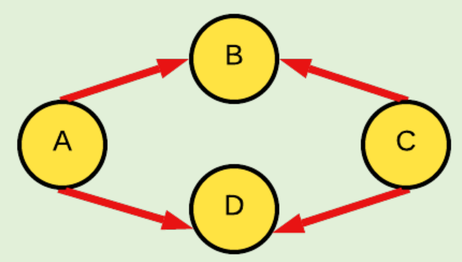
CPC Is Not Complete

A critical observation for the example above:
- By **Markov condition**, $(A \not\perp\!\!\!\perp C|B)_P$ implies the triple $\langle A, B, C\rangle$ cannot be a non-collider.
- Hence, $\langle A, B, C\rangle$ should be oriented as $A \rightarrow B \leftarrow C$.

Sources of Unfaithfulness

Cancelled Paths: in a DAG G=(V, E) with any unfaithful distribution p compatible with G, we say the active paths q between a set of variables **X** and another set of variables **Y** are cancelled relative to a set of vertices $\boldsymbol{Z} \subseteq V, (\boldsymbol{X}, \boldsymbol{Y} \not\subseteq \boldsymbol{Z})$ if

- $(\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_G$ and
- $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_P$



When **all cancelled paths** from A to C relative to B **are along all the d-connecting paths** from X to Y relative to J, we denote it as $Path(A, C, B) \subseteq_C Path(X, Y, J)$

Revised CPC (RCPC) Algorithm

Following step 4 of CPC, we recursively apply R5 and R6 until there is no more edges that can be oriented by them. Let G be the resulting graph after step 4.

R5: For every unshielded triple $\langle A, B, C\rangle$ that has been marked unfaithful,

a) if $A \rightarrow \underline{B} \leftarrow C$, unmark $A \rightarrow \underline{B} \leftarrow C$ as $A \rightarrow B \leftarrow C$.
   - I. Mark all CIs $(A \perp\!\!\!\perp C|W)_P$ as NM (non-Markov) statement for any W that contains B and
   - II. If $Path(A, C, B) \subseteq_C Path(X, Y, J)$, mark $(X \perp\!\!\!\perp Y|J)_P$ as NM statement for any J that contains B

b) else, mark $(A \perp\!\!\!\perp C|S)_P$ as NM statement for any S that does not contain B and unmark the triple.
   - I. If $Path(A, C, \emptyset) \subseteq_C Path(X, Y, D)$, mark $(X \perp\!\!\!\perp Y|D)_P$ as NM statement for any D that does not contain B.

Then, **excluding all NM statements**, for each unshielded triple $\langle A, T, C\rangle$ that **is marked as unfaithful:**
- If T is not in any set conditional on which A and C are independent, orient $A - \underline{T} - C$ as $A \rightarrow \underline{T} \leftarrow C$.
- If T is in all such sets conditional on which A and C are independent, unmark $\langle A, T, C\rangle$.

R6: Recursively apply R1, R3, and R4 of Meek rules **[Meek'95]** to unshielded non-colliders that are not marked as unfaithful except that R2 can be applied to any triple.
- **Additionally,** for any unshielded triple $\langle A, T, C\rangle$ that is oriented as as $A \rightarrow \underline{T} - C$, if there is an undirected path p e.g. $Q - \cdots - C$ and no triples along p has been marked unfaithful and there is a directed path q e.g. $Q \rightarrow \cdots \rightarrow A$. Then, we orient $A \rightarrow \underline{T} \leftarrow C$ (keeping the underline for marking NM statement).

Theorem: Under the causal Markov and Adjacency-Faithfulness assumptions, the RCPC algorithm is correct in the sense that given a perfect conditional independence oracle, the algorithm returns an extended pattern that represents the true causal DAG.

Example 1 (Application of R5):

$(A \perp\!\!\!\perp E)_P$

$(A \perp\!\!\!\perp C)_P$

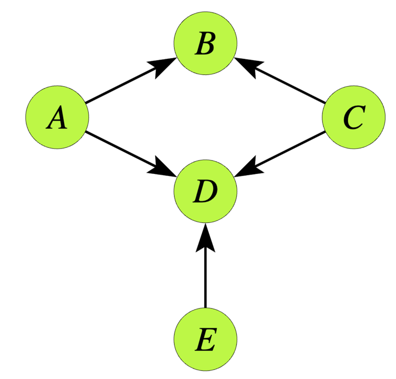
$(E \perp\!\!\!\perp B)_P$

$(E \perp\!\!\!\perp C)_P$

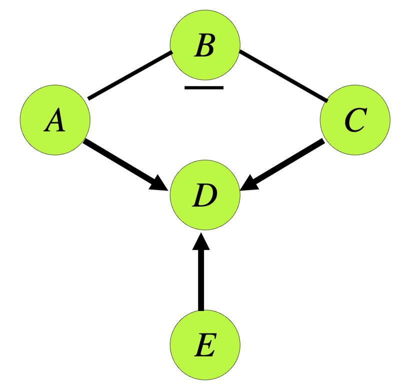
$(A \not\perp\!\!\!\perp C|B)_P$

$(A \perp\!\!\!\perp C|B, D)_P$

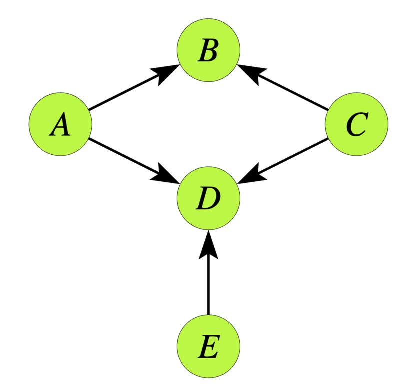
$(B \perp\!\!\!\perp D|A, C)_P$



True DAG



CPC output



RCPC output

Example 2 (Application of R5 and R6):

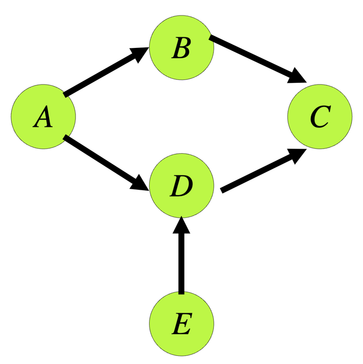
$(A \perp\!\!\!\perp C)_P$

$(B \perp\!\!\!\perp D|A, C)_P$

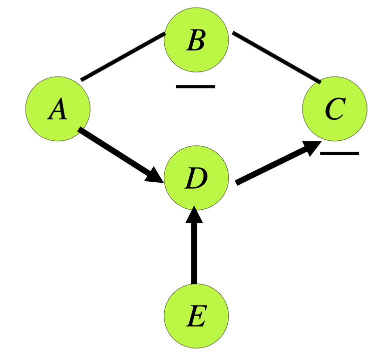
$(A \perp\!\!\!\perp C|B, D)_P$

$(A \not\perp\!\!\!\perp C|B)_P$

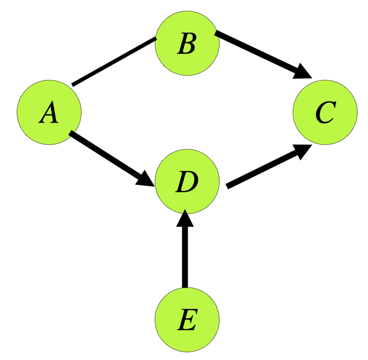
$(A \not\perp\!\!\!\perp C|D)_P$



True DAG



CPC output



RCPC output

R5: unmark $\langle A, B, C\rangle$.

R6: orient $\langle B, C, D\rangle$ as an unshielded collider

Reference

J. Pearl, Causality: Models, Reasoning and Inference. Cambridge University Press, 2009

Spirtes, Peter, Clark Glymour, and Richard Scheines. Causation, prediction, and search. MIT press, 2001.

Ramsey, Joseph, Peter Spirtes, and Jiji Zhang. "Adjacency-faithfulness and conservative causal inference." Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence. 2006.

Meek, Christopher. "Causal inference and causal explanation with background knowledge." Proceedings of the Eleventh conference on Uncertainty in artificial intelligence. 1995.