Constraint-based Causal Discovery from a Collection of Conditioning Sets

Kenneth Lee[1] **Bruno Ribeiro**[2] **Murat Kocaoglu**[1]

[1]Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA
[2]Department of Computer Science, Purdue University, West Lafayette, Indiana, USA

Abstract

In constraint-based causal discovery, existing algorithms systematically use a series of conditional independence (CI) relations observed in the data to recover an equivalence class of causal graphs in the large sample limit. One limitation of these algorithms, such as the PC algorithm, is the reliance on CI tests, which can quickly lose statistical power due to finite samples as the conditioning set size increases or the support of the conditioning set is large. The idea of bounding the size of conditioning sets has been proposed for robust causal discovery. However, the existing algorithms require exhaustive testing of all CI relations with conditioning set sizes up to a certain integer k. To further relax this restriction, we propose using CI tests where the conditioning sets are restricted to a given set of conditioning sets including the empty set. We call this set a *conditionally closed* set $\mathcal{C}$. We define the notion of $\mathcal{C}$-Markov equivalence. We propose a graphical representation to characterize $\mathcal{C}$-Markov equivalence between two causal graphs. We propose a sound constraint-based algorithm called the $\mathcal{C}$-PC algorithm for learning the $\mathcal{C}$-Markov equivalence class. We demonstrate the utility of the proposed algorithm via experiments in scenarios where high-dimensional variables and spurious correlations are present in the data.

1 INTRODUCTION

Constraint-based methods struggle with limited data as CI tests are prone to have a high false positive rate, especially when the conditioning set is large [Wille and Bühlmann, 2006, Shah and Peters, 2020]. Due to the sensitivity of CI tests to sample noise, several ideas have been proposed to enhance the accuracy of the algorithm outputs, including ensuring the output's independence from the sequence of CI tests conducted [Colombo et al., 2014], relaxing model assumptions [Ramsey et al., 2006], ensuring the consistency of conditioning sets used in CI tests [Li et al., 2019], and characterizing and learning of causal graphs based on small conditioning sets [Wienöbst and Liskiewicz, 2020, Kocaoglu, 2024]. Building on previous research that explores learning causal graphs with small conditioning sets, our paper aims to further relax the requirement of exhausting all CI relations up to degree k. We achieve this by employing CI tests with a *conditionally closed* set, which comprises a specified collection of conditioning sets, including the empty set.

Q1	Q2	Q3	Q4	Q5
0	10	1	5	1
1	10	1	5	0
1	10	0	5	0
1	10	0	5	1
0	10	1	5	0
0	10	1	5	0
...	...	...	...	...

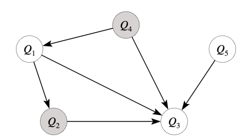
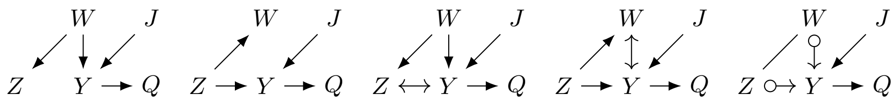


Figure 1: An example where learning causal graphs from a collection of conditioning sets is desired. A survey dataset consists of some categorical variables, e.g. $Q2, Q4$, that show most participants respond in the same way, creating highly correlated variables that pose challenges in testing conditional independence.

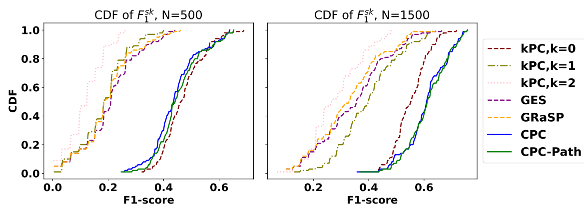
1.1 MOTIVATING EXAMPLE

Consider a setting where one has collected some survey results and wants to understand the causal relations among the variables observed from the survey questions. As shown by Figure 1, the survey mostly consists of categorical variables. Some of the survey questions, i.e., $Q2, Q4$, involve many choices for participants to choose from and the majority of the participants tend to select the same answer. This results



(a) D (b) D' (c) $\mathcal{S}_{\mathcal{C}}(D)$ (d) $\mathcal{S}_{\mathcal{C}}(D')$ (e) $\varepsilon_{\mathcal{C}}(D)$

Figure 2: An example shows the difference between a Markov equivalence class and a $\mathcal{C}$-Markov equivalence class, where $\mathcal{C} = \{\emptyset, \{Y\}\}$. (a)-(b): Two $\mathcal{C}$-Markov equivalent DAGs, i.e., $D \sim_{\mathcal{C}} D'$. However, D and D' are not Markov equivalent. (c)-(d): Two Markov equivalent $\mathcal{C}$-closure graphs $\mathcal{S}_{\mathcal{C}}(D), \mathcal{S}_{\mathcal{C}}(D')$. (e): The $\mathcal{C}$-essential graph of D.



(a) $F1_{\text{skeleton}}, N = 500$ (b) $F1_{\text{skeleton}}, N = 1500$

Figure 3: Empirical cumulative distribution function of various F1 scores on 100 random DAGs with 60 edges and 30 nodes where the number of states is assigned to be 2 or 30 with probability 0.7 and 0.3 respectively. The lower and farther from the left side the better. k-PC ($k = 0$) performs comparably with $\mathcal{C}$-PC at $N = 500$. $\mathcal{C}$-PC outperforms all baselines at $N = 1500$.

in a dataset that has highly correlated variables, which poses challenges in testing conditional independence. Hence, it is desirable to choose which variable to condition on for constraint-based causal discovery.

2 RESULTS

In this section, we provide the definition of conditionally closed sets $\mathcal{C}$ and the graphical representation of $\mathcal{C}$-Markov equivalence. Figure 2 illustrates all the relevant concepts.

Definition 2.1 (Conditionally Closed Sets)**.** For a DAG $D = (\mathbf{V}, \mathbf{E})$, let $\mathcal{I} = \{I_i\}$ be a set of CI statements of the form $I_i = (X, \mathbf{Z}, Y), i.e., (X \perp\!\!\!\perp Y|\mathbf{Z})$ or $(X \not\perp\!\!\!\perp Y|\mathbf{Z})$, where $X, Y \in \mathbf{V}, \mathbf{Z} \subset \mathbf{V}$. A set $\mathcal{C}$ is called *conditionally closed* if the following holds

1. $\emptyset \in \mathcal{C}$ and
2. $\exists X, Y \in \mathbf{V}, (X, \mathbf{C}, Y) \in \mathcal{I} \Rightarrow (A, \mathbf{C}, B) \in \mathcal{I}$ for all $A, B \in \mathbf{V}$ and for all $\mathbf{C} \in \mathcal{C}$.

Definition 2.2 ($\mathcal{C}$-Markov equivalence)**.** Two DAGs D_1, D_2 are *$\mathcal{C}$-Markov equivalent* if for any three disjoint subsets $\mathbf{X} \subset \mathbf{V}$, $\mathbf{Y} \subset \mathbf{V}, \mathbf{Z} \in \mathcal{C}$, $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ in D_1 if and only if $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ in D_2, where $\mathcal{C}$ is conditionally closed. The set of DAGs that encode the same set of conditional independence induced only by the causal Markov assumption with conditioning sets from $\mathcal{C}$ is called the *$\mathcal{C}$-Markov equivalence class*. We denote two DAGs D_1, D_2 that are $\mathcal{C}$-Markov equivalent as $D_1 \sim_{\mathcal{C}} D_2$.

Definition 2.3 ($\mathcal{C}$-closure)**.** For a DAG D and a conditionally closed set $\mathcal{C}$, the $\mathcal{C}$-closure of D, denoted as $\mathcal{S}_{\mathcal{C}}(D)$, is a graph that has the following properties:

1. If: $\nexists \mathbf{C} \in \mathcal{C}$ s.t. $\mathbf{C}$ d-separates X and Y in D
   (i) if $X \in An_D(Y)$, then $X \to Y$ in $\mathcal{S}_{\mathcal{C}}(D)$, (ii) if $Y \in An_D(X)$, then $Y \to X$ in $\mathcal{S}_{\mathcal{C}}(D)$, (iii) else $X \leftrightarrow Y$ in $\mathcal{S}_{\mathcal{C}}(D)$.
2. Else: X, Y are not adjacent in $\mathcal{S}_{\mathcal{C}}(D)$.

Lemma 2.1. *$\mathcal{C}$-closure graph $\mathcal{S}_{\mathcal{C}}(D)$ of a DAG D entails the same d-separation statements conditioned any $\mathbf{C} \in \mathcal{C}$ as the DAG, i.e., $(X \perp\!\!\!\perp Y|\mathbf{C})_D \Leftrightarrow (X \perp\!\!\!\perp Y|\mathbf{C})_{\mathcal{S}_{\mathcal{C}}(D)}, \forall \mathbf{C} \in \mathcal{C}$.*

Theorem 2.4. *Two DAGs D_1, D_2 are $\mathcal{C}$-Markov equivalent if and only if $\mathcal{S}_{\mathcal{C}}(D_1)$ and $\mathcal{S}_{\mathcal{C}}(D_2)$ are Markov equivalent.*

Definition 2.5 (edge unions: —, o—o, $o\!\!\to$)**.** The edge union operations of a set of $\mathcal{C}$-closure graphs are defined as: (i) $X — Y := X \to Y \cup X \leftarrow Y$, (ii) $X\, o\!\!—\!\!o\, Y := X \to Y \cup X \leftarrow Y \cup X \leftrightarrow Y$, (iii) $X\, o\!\!\to Y := X \to Y \cup X \leftrightarrow Y$.

Definition 2.6 ($\mathcal{C}$-essential graph)**.** For any DAG D, the edge union of all $\mathcal{C}$-closure graphs that are Markov equivalent to $\mathcal{S}_{\mathcal{C}}(D)$ is called the $\mathcal{C}$-essential graph of D, denoted as $\varepsilon_{\mathcal{C}}(D)$.

Theorem 2.7. *$\mathcal{C}$-PC algorithm is sound for learning $\mathcal{C}$-essential graph given a conditional independence oracle under the causal Markov and $\mathcal{C}$-faithfulness assumptions, i.e., if $\mathcal{C}$-PC returns M, we have $\varepsilon_{\mathcal{C}}(D) \subseteq M \subseteq PAG(\mathcal{S}_{\mathcal{C}}(D))$.*

The pseudocode of $\mathcal{C}$-PC is in Appendix 1. We conduct synthetic experiments to evaluate the performance of $\mathcal{C}$-PC when high-dimensional variables are present in data. We include the baselines: k-PC [Kocaoglu, 2024], GES [Chickering, 2002], and GRaSP [Lam et al., 2022]. The results are shown in Figure 3.

3 CONCLUSION

In conclusion, we propose a sound algorithm called $\mathcal{C}$-PC for learning causal graphs from a collection of conditioning sets known as conditionally closed sets.

References

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Nov):507–554, 2002.

Diego Colombo, Marloes H Maathuis, et al. Order-independent constraint-based causal structure learning. *J. Mach. Learn. Res.*, 15(1):3741–3782, 2014.

Murat Kocaoglu. Characterization and learning of causal graphs with small conditioning sets. *Advances in Neural Information Processing Systems*, 36, 2024.

Wai-Yin Lam, Bryan Andrews, and Joseph Ramsey. Greedy relaxations of the sparsest permutation algorithm. In *Uncertainty in Artificial Intelligence*, pages 1052–1062. PMLR, 2022.

Honghao Li, Vincent Cabeli, Nadir Sella, and Hervé Isambert. Constraint-based causal structure learning with consistent separating sets. *Advances in Neural Information Processing Systems*, 32, 2019.

Joseph Ramsey, Peter Spirtes, and Jiji Zhang. Adjacency-faithfulness and conservative causal inference. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 401–408, 2006.

Rajen D Shah and Jonas Peters. The hardness of conditional independence testing and the generalised covariance measure. 2020.

Marcel Wienöbst and Maciej Liskiewicz. Recovering causal structures from low-order conditional independencies. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10302–10309, 2020.

Anja Wille and Peter Bühlmann. Low-order conditional independence graphs for inferring genetic networks. *Statistical applications in genetics and molecular biology*, 5 (1), 2006.

Jiji Zhang. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9(7), 2008.

Constraint-based Causal Discovery from a Collection of Conditioning Sets (Supplementary Material)

Kenneth Lee[1] **Bruno Ribeiro**[2] **Murat Kocaoglu**[1]

[1]Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA
[2]Department of Computer Science, Purdue University, West Lafayette, Indiana, USA

A DISCUSSION

A.1 DIFFERENCES BETWEEN A MARKOV EQUIVALENCE CLASS AND A $\mathcal{C}$-MARKOV EQUIVALENCE CLASS

In this section, we give an example to show that two DAGs that share the same d-separation statements based on a conditionally closed set $\mathcal{C}$ and do not have the same skeleton and the same set of unshielded colliders. We use Figure 2 to illustrate the difference between the Markov equivalence class and the $\mathcal{C}$-Markov equivalence class. As shown by Figures 2(a) and 2(b), both D_1 and D_2 have different skeletons and different sets of unshielded colliders. Nonetheless, D_1 and D_2 are $\mathcal{C}$-Markov equivalent when $\mathcal{C} = \{\emptyset, \{Y\}\}$ because both D_1 and D_2 entail the same CI statement of degree 0 and the same CI statements with conditioning set $\{Y\}$.

A.2 INTUITION ABOUT $\mathcal{C}$-CLOSURE GRAPHS

When two variables X, Y are $\mathcal{C}$-covered and none of them is an ancestor of another, it is as if there exists a confounder between them that is not observed, resulting in $X \leftrightarrow Y$. When one is an ancestor of another, $\mathcal{C}$-closure graphs preserve this ancestral relationship. Figures 2(c) and 2(d) show an example of $\mathcal{C}$-closure graphs. Lemma 2.1 gives the relationship between the CI statements entailed by a DAG and the CI statements entailed by a $\mathcal{C}$-closure graph based on a conditionally closed set.

A.3 INTUITION ABOUT EDGE UNION AND C-ESSENTIAL GRAPH

To understand the edge union operation in Definition 2.5 and its connection to the construction of the $\mathcal{C}$-essential graph, consider two Markov equivalent $\mathcal{C}$-closure graphs $\mathcal{S}_\mathcal{C}(D), \mathcal{S}_\mathcal{C}(D')$ shown by Figures 2(c) and 2(d). We can take the union of the edge $Z \leftrightarrow Y$ in $\mathcal{S}_\mathcal{C}(D)$ and the edge $Z \rightarrow Y$ in $\mathcal{S}_\mathcal{C}(D')$ to derive the edge $Zo\!\rightarrow Y$ in the $\mathcal{C}$-essential graph shown in Figure 2(e). From both Definition 2.5 and Definition 2.6, we see that a directed edge $\rightarrow$ appears in a $\mathcal{C}$-essential graph only if such directed edge appears in all $\mathcal{C}$-closure graphs that are Markov equivalent. It is important to know the difference between the edges o—o and — from a causal viewpoint. The former says that there exists a $\mathcal{C}$-closure graph in the equivalence class where two variables cannot be a cause of each other. The latter indicates that there exists a $\mathcal{C}$-closure graph in the equivalence class where one variable is a cause of another variable.

A.4 FUTURE WORK

For future work, we want to further relax the restriction of a conditionally closed set and investigate whether one can systematically leverage arbitrary CI statements on top of all marginal independence relations for learning causal graphs. This feature is desirable because one can then directly avoid testing certain CI statements that are very likely to be unreliable when spurious correlations are present in practice. Another interesting direction is to study the orientation rules with respect to the conditionally closed set. It is also worthwhile to explore how side information and data preprocessing techniques can be used to construct the conditionally closed set for $\mathcal{C}$-PC to increase the robustness of the proposed method.

B ALGORITHMS

Algorithm 1 $\mathcal{C}$-PC

input Observational data $\mathbf{V}$, a conditionally closed set $\mathcal{C}$, CI tester

Initiate a complete graph M among the set of observed variables with circle edge o—o.
Find separating sets $S_{X,Y}$ for every pair $X, Y \in \mathbf{V}$ by conditioning on $\mathbf{C} \in \mathcal{C}$.
Update M by removing the edges between pairs that are separable.
Orient unshielded colliders of M: For any induced subgraph $Xo{—}oZo{—}oY$, set $Xo{\rightarrow}\ Z \leftarrow oY$ for any non-adjacent pair X, Y where $S_{X,Y}$ does not contain Z.
$M \leftarrow$ **kPC_Orient**(M) {see Algorithm 3}
return M

Algorithm 2 k-PC Kocaoglu [2024]

input Observational data $\mathbf{V}$, k, CI tester

Initiate a complete graph M among the set of observed variables with circle edge o—o.
Find separating sets $S_{X,Y}$ for every pair $X, Y \in V$ by conditioning on subsets $\mathbf{Z} \subset \mathbf{V}$ of size at most k.
Update M by removing the edges between pairs that are separable.
Orient unshielded colliders of M: For any induced subgraph $Xo{—}oZo{—}oY$, set $Xo{\rightarrow}\ Z \leftarrow oY$ for any non-adjacent pair X, Y where $S_{X,Y}$ does not contain Z.
$M \leftarrow$ **kPC_Orient**(M) {See Algorithm 3}
return M

Algorithm 3 kPC_Orient Kocaoglu [2024]

input Mixed graph M

$M \leftarrow$ FCI_Orient(M) {See Algorithm 4}
For any variable X that has no incoming edges, construct the sets $\mathcal{B}, \mathcal{Q}$:

$$\mathcal{B} = \{Y \in Ne(X) : Xo{\rightarrow}\ Y\}, \mathcal{Q} = \{Z \in Ne(X) : Xo{—}oZ\}$$

and define sets $\mathcal{B}^\star$ as the set of variables that are non-adjacent to any of the nodes in $\mathcal{Q}$ and $\mathcal{Q}^\star$ as the set of variables that are non-adjacent to other variables in $\mathcal{Q}$:

$$\mathcal{B}^\star = \{Y \in \mathcal{B} : Y, Z \text{ are non-adjacent } \forall Z \in \mathcal{Q}\}, \mathcal{Q}^\star = \{Z' \in \mathcal{Q} : Z', Z \text{ are non-adjacent } \forall Z' \neq Z, Z' \in \mathcal{Q}\}$$

$\mathcal{R}11$: Orient $Xo{\rightarrow}\ Y$ as $X \rightarrow Y, \forall Y \in \mathcal{B}^\star$
$\mathcal{R}12$: Orient $Xo{—}oY$ as $X{—}Y, \forall Z \in \mathcal{Q}^\star$
return M

Algorithm 4 FCI_Orient Kocaoglu [2024]

Apply the orientation rules of $\mathcal{R}1, \mathcal{R}2, \mathcal{R}3$ of Zhang [2008] to M until none applies.
Apply the orientation rules of $\mathcal{R}8, \mathcal{R}9, \mathcal{R}10$ of [Zhang, 2008] to M.
return M