Faithfulness and Intervention-Only Causal Discovery

Bijan Mazaheri [1 2] **Jiaqi Zhang** [3 2] **Caroline Uhler** [3 2]

Abstract

Causal discovery seeks to learn a network describing the causal dependencies between observed variables. Constraint-based causal discovery makes use of conditional independence properties to narrow the space of possible causal networks down to a Markov equivalence class, which consists of adjacency information (e.g., A causes B or B causes A, but we might not know the direction). Score-based causal discovery differs algorithmically, but also relies on statistical properties of the observed distribution to determine adjacency. A critical assumption for both approaches is faithfulness — a requirement that causally linked variables exhibit statistical dependence. Previous works have shown faithfulness to be a strong and restrictive assumption, especially in the finite sample regime. While interventions are usually utilized to orient causal edges, the results of these orientations also contain adjacency-specific information that is generally not utilized. In particular, we show that faithfulness violations can be resolved using interventions. To formalize this notion, we provide a mild assumption that we call intervention-adjacency (IA) faithfulness and build intervention-only causal discovery algorithms that are provably consistent under this assumption. We also specify equivalence classes when the identification criteria are not met due to limitations in the scope of interventions, which may be further resolved via conditional independence testing. Our results provide new insights into the power of online learning and learning by doing.

[1]Thayer School of Engineering, Dartmouth, Hanover, NH, USA [2]Eric and Wendy Schmidt Center, Broad Institute of MIT and Harvard, Cambridge, MA, USA [3]Laboratory for Information and Decision Systems, MIT, Cambridge, MA, USA. Correspondence to: Bijan Mazaheri <bijan.h.mazaheri@dartmouth.edu>, Jiaqi Zhang <viczhang@mit.edu>, Caroline Uhler <cuhler@mit.edu>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

1. Introduction

Structural Causal Models Structural Causal Models (SCMs), popularized by Pearl (1998; 2009), graphically describe causal networks. Causal discovery is the task of recovering the underlying causal structure from data in the form of a *directed acyclic graph* (DAG) or a representation of an equivalence class of DAGs (see Squires & Uhler (2023) for a review). One approach to causal discovery involves a *constraint-based* search guided by conditional independence ("CI-tests"), e.g., the PC-algorithm from Spirtes et al. (2000). Constraint-based approaches use the observation that variables without a direct causal link can be made independent by conditioning on intermediary causal paths. This is known as the causal Markov condition (Pearl, 2009).

The formulation of structural causal models under causal sufficiency (no unobserved confounding) implies the causal Markov condition, but the converse is not necessarily true (Ramsey et al., 2012). A causal path between two variables does not necessarily require that they be statistically dependent under all condition sets. "Faithfulness" is the assumption that causal links imply statistical dependencies, which gives a two-way correspondence between CI-tests and graphical properties.

Motivation The set of unfaithful distributions has Lebesgue measure 0 (Meek, 2013). Hence, faithfulness is a mild assumption under exact statistics. However, success under finite sample uncertainty requires a stronger notion, referred to as λ-strong faithfulness (Zhang & Spirtes, 2012), to ensure that stochastic deviations towards faithfulness violations are unlikely. Under this perspective, faithfulness becomes a strong assumption due the manifold of violations behaving like a "space filling curve." In particular, Uhler et al. (2013) showed that the majority of distributions generated by linear structural equations with additive Gaussian noise are "close"[1] to a faithfulness violation.

Evidently, λ-strong faithfulness violations are highly likely in both real-world and simulated datasets. To resolve these violations, we turn to interventions. Information about a causal structure can be decomposed into *skeletal (adjacency)*

[1]A distribution is "close" to a violation of faithfulness if, e.g., two causally connected variables are almost independent. For Gaussian noise, this corresponds to a near-zero covariance.

information and *orientation information*. Causal adjacency corresponds to the existence of a direct causal connection, agnostic of the direction. For example, A, B are causally adjacent if $A \to B$ or $B \to A$. Orientation then specifies the direction of these causal adjacencies.

Constraint-based causal discovery recovers a Markov Equivalence Class (MEC) corresponding to all of the adjacency information, but generally only part of the orientation information. Further information about causal orientation is often obtained by utilizing asymmetric intervention responses, i.e., interventions on causes change effects, but interventions on effects do not change causes.

To resolve adjacency errors due to strong-faithfulness violations, we will study the use of interventions beyond edge orientation. For example, if X, Y are not causally adjacent and also not connected by any directed paths, then intervening on X does not change Y, and intervening on Y does not change X. If such interventions must be performed to orient causal graphs after determining Markov equivalence, this information might as well be incorporated at earlier points. A concrete goal of this work is to quantify the skeletal information encoded in interventions relative to how many variables are intervened on at once.

1.1. Contributions

In Section 3, we provide a definition of intervention-adjacency (IA) faithfulness that is milder than the standard (conditional independence/CI) faithfulness used for constraint-based causal discovery. IA faithfulness only requires nontrivial coefficients, i.e., the structural equations used to generate each variable from its causes have nontrivial partial derivatives (see Definition 3.1). This notion of faithfulness is also milder than previous notions of intervention faithfulness, which require changes in *all of the causal descendants* of an intervention. We argue the mildness of this condition by proving that the volume of distributions that are "close" to violating IA faithfulness is significantly less than the volume of distributions that are "close" to violating CI faithfulness.

In Section 4, we develop "change sets," which form the building blocks of causal discovery algorithms that we summarize in Section 4.1 and present in Appendix C. These algorithms succeed under our new, milder, λ-strong IA faithfulness. In this setting, identifiability depends on the number of variables on which we can intervene simultaneously (k) relative to the minimum vertex cut κ of the graph.

Theorem 1.1 (informal)**.** *A causal graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ with vertex connectivity κ is identifiable using $\mathcal{O}(|\mathbf{V}|^2|\mathbf{E}|)$ do-interventions on up to $k = \kappa + 1$ nodes, so long as λ-IA faithfulness is paired with "significant"[2] interventions.*

We also define the "k-robust transitive closure" of $\mathcal{G}$, which corresponds to adding edges that follow topological order that cross any two vertices that cannot be separated into two connected components using a vertex cut of $< k$. Under limited cardinality of interventions, DAGs that have the same k-robust transitive closure form an equivalence class, which we cannot distinguish between using $\leqslant k$-node interventions.

Theorem 1.2 (informal)**.** *A causal graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ is identifiable up to its k-robust transitive closure using do-interventions on up to k nodes, so long as λ-IA faithfulness is paired with sufficiently "significant" interventions.*

The equivalence classes provide insight into the adjacency information held by do-interventions. In Section 5, we perform an empirical study on synthetic data to verify the relative mildness of our new type of faithfulness and the relative robustness of intervention-only causal discovery. In the absence of real datasets with large-scale interventions, our results serve as motivation to develop new ways to perform these multi-node interventions to help improve the accuracy of causal discovery.

2. Preliminaries

2.1. Notation

We will use the capital Roman alphabet to denote random variables (e.g., A, B, C, V) and the lowercase Roman alphabet to denote assignments to those random variables (e.g., $A = a$ or just a). Bold will indicate a set of random variables, e.g., $\mathbf{V} = (V_1, V_2, \ldots)^\top$, and $\mathbf{v}$ is an assignment to $\mathbf{V}$. Parents ($\mathbf{PA}$), children ($\mathbf{CH}$), ancestors ($\mathbf{AN}$), and descendants ($\mathbf{DE}$) in graphs will also follow these conventions, e.g., $\mathbf{PA}(V) = \mathbf{pa}^{\mathbf{v}}(v)$, where the assignments to those parents come from values specified in $\mathbf{v}$. We use subscripts to indicate the relevant graph structure for the parents, e.g., $\mathbf{PA}_{\mathcal{G}}(V)$. We will generally use the Greek alphabet (e.g., α, β) to represent parameters for structural equations and thresholds to quantify faithfulness.

2.2. Conditional Independence Faithfulness

D-separation provides a graphical criterion that constitutes a necessary but not sufficient condition for conditional independence under the causal Markov condition, which is satisfied in standard structural equation models that exclude unobserved confounding. We say that two variables V_i, V_j are d-separated, denoted $V_i \perp\!\!\!\perp_d V_j$ if there is no active path between V_i and V_j in the causal graph. See Pearl (2009) for a more detailed description of active paths and d-separation.

Faithfulness generally refers to the assumption that causal

[2] In Theorem 1.1 and the following Theorem 1.2, "significant" interventions perturb variables some minimum amount from their expectation (see Definition 4.2 in Section 4).

linkage always carries statistical dependence, making d-separation both necessary and sufficient for conditional independence. To emphasize the difference between this form of faithfulness and others, we will call (regular) faithfulness conditional independence (CI) faithfulness.

For linear structural equation models with additive Gaussian noise, CI faithfulness with respect to a causal graph $\mathcal{G}$ corresponds to nonzero covariance/correlation, e.g., $\mathrm{Cov}(V_i, V_j) \neq 0$ for all $V_i \not\perp\!\!\!\perp_d V_j$. Consider the following example with N_1, N_2, N_3 independent unbiased Gaussians with variance 1.

$$\begin{aligned} X_1 &= N_1 \\ X_2 &= \alpha_{12} X_1 + N_2 \\ X_3 &= \alpha_{13} X_1 + \alpha_{23} X_2 + N_3 \end{aligned} \tag{1}$$

If we write X_3 in terms of N_1, we can express the covariance between X_1, X_3 as a function of coefficients

$$\mathrm{Cov}(X_1, X_3) = \alpha_{13} + \alpha_{23}\alpha_{12}.$$

Notice that a monomial of parameters emerges from each active path between X_1 and X_3. When $\alpha_{13} + \alpha_{23}\alpha_{12} = 0$ we have $\mathrm{Cov}(X_1, X_3) = 0$, which corresponds to a CI faithfulness violation. Notice that this is also a violation of "adjacency faithfulness" as defined in Zhang & Spirtes (2008) because X_1, X_3 are adjacent. CI-faithfulness violations occur in two ways: (1) trivial coefficients, e.g., $\alpha_{ij} = 0$, and (2) the "cancelation" of monomial terms from multiple paths (an example of which is given above).

λ-strong faithfulness (Uhler et al., 2013) requires $|\mathrm{Cov}(V_i, V_j)| > \lambda\sqrt{\mathrm{Var}(V_i)\mathrm{Var}(V_j)}$ for all V_i, V_j that are d-connected. Geometrically, violations in λ-strong faithfulness correspond to distributions that are close to the hypersurfaces formed by the polynomial equations describing regular CI faithfulness.

The argument for λ-strong faithfulness primarily stems from uncertainty in finite-sample settings, where conditional independence tests are done on empirical covariance matrix estimates. As such, even when $\mathrm{Cov}(X_1, X_3) \neq 0$, it is unlikely that its empirical estimate $\widetilde{\mathrm{Cov}}(X_1, X_3)$ will also be nonzero. λ-faithfulness ensures that causally linked variables can be distinguished from unlinked ones exhibiting this "accidental" dependence by forcing true causal links to have some minimum strength. This, in turn, allows conditional independence tests to be "thresholded" to handle this noise. In general, Uhler et al. (2013) showed that λ-strong CI faithfulness is a relatively strong assumption.

3. Intervention Adjacency Faithfulness

Chevalley et al. (2025) introduced a notion of "ε-strong intervention faithfulness" that formalizes the requirement that intervening on a variable must induce a change on all of its descendants. Such an assumption is still vulnerable to a "cancellation of paths" or decay of dependence between variables that are linked by a long causal chain. For example, notice that the model described by Equations 1 exhibits both a CI faithfulness violation and an intervention faithfulness violation when $\alpha_{13} + \alpha_{23}\alpha_{12} = 0$, because pertubing X_1 will elicit no change in X_3. In this section, we will develop λ-strong intervention adjacentcy faithfulness (IA faithfulness) as a milder and more robust assumption for causal discovery in finite sample settings.

We will weaken other notions of intervention faithfulness to only require nontrivial coefficients. As with conditional independence faithfulness, we can introduce a stronger notion of such faithfulness that ensures this change is detectable.

Definition 3.1 (λ-strong IA faithfulness)**.** For a linear structural equation model with additive Gaussian noise, we say that the model is λ-strong IA faithful if all coefficients α_{ij} for $V_i \to V_j$ are lower bounded by $|\alpha_{ij}| > \lambda$.

Return to the example given in Equation (1), where λ-strong IA faithfulness requires $|\alpha_{12}|, |\alpha_{13}|, |\alpha_{23}| > \lambda$. In contrast, λ-strong CI faithfulness requires six correlation terms to be greater than λ, three of which simplify to $\alpha_{12}, \alpha_{13}, \alpha 23$. Notice that setting any of the α_{ij} coefficients to 0 automatically results in a λ (and regular) CI-faithfulness violation. The converse is not true, as exemplified by setting the coefficient to large values with $\alpha_{12}(\alpha_{13} + \alpha_{12}\alpha_{23}) + \alpha_{23} = 0$, e.g. $\alpha_{12} = \alpha_{23} = \lambda = 1$ and $\alpha_{13} = -2\lambda$. Such an assignment of coefficients gives $\mathrm{Cov}(X_2, X_3) = 0$.

While this model violates CI faithfulness, interventions can still provide some information about the adjacency between X_2 and X_3. Notice that shifting X_2 by δ will elicit a proportional change in X_3 by $\delta\alpha_{23}$. In fact, nontrivial coefficients guarantee a resulting change in all descendants that have no additional directed paths (which could cause path cancellations). Conveniently, there is always at least one such descendant or no descendants at all, since additional directed paths must also go through descendants. This will be used to develop an algorithm for learning causal structures.

4. Intervention-Only Models

In this section, we will formalize the notion of interventions that we will use to develop our algorithm. We are motivated by the experimental study of gene regulatory networks, on which it is possible to perform a "knockout" intervention (Guan et al., 2010). A knockout fixes the value of a variable a to be 0. The resulting distribution on $\mathcal{G}, \mathrm{Pr}(\cdot)$ is given by a do-intervention (Pearl, 2009),

$$\mathrm{Pr}(\mathbf{v} \backslash \mathbf{a} | \mathrm{do}(\mathbf{a})) = \prod_{v \in \mathbf{v} \backslash \mathbf{a}} \mathrm{Pr}(v | \, \mathbf{pa}^{\mathcal{G}}_{\mathbf{v}}(v) \backslash \mathbf{A}, \mathbf{PA}^{\mathcal{G}}(V) \cap \mathbf{a}). \tag{2}$$

Here, $\mathbf{pa}^{\mathcal{G}}_{\mathbf{v}}(v) \setminus \mathbf{A}$ and $\mathbf{PA}^{\mathcal{G}}(V) \cap \mathbf{a}$ abuse notation. The first takes the assignments $\mathbf{pa}^{\mathcal{G}}_{\mathbf{v}}(v)$ and removes $\mathbf{A}$ and the second takes the variables in $\mathbf{A} \cap \mathbf{PA}^{\mathcal{G}}(V)$ and gives them assignments from $\mathbf{a}$. A knockout replaces the structural equations for $\mathbf{A}$ with $\mathbf{a} = \mathbf{0}$. We will utilize the following lemma that follows from λ-strong IA faithfulness.

Lemma 4.1. *If a probability distribution* $\Pr(\cdot)$ *is* λ*-strong IA faithful to its DAG* $\mathcal{G}$*, then for all* $V_i \in \mathbf{V}$*,* $V_j \in \mathbf{CH}^{\mathcal{G}}(V_i)$ *such that there are no other directed paths from* $V_i \to \ldots \to V_j$*, we have*

$$|\mathbb{E}[V_j] - \mathbb{E}[V_j \mid do(v_i)]| > \lambda|\mathbb{E}[V_i] - v_i|. \tag{3}$$

Lemma 4.1 follows from $\mathbb{E}[V_j] - \mathbb{E}[V_j \mid \mathrm{do}(v_i)] = \alpha_{ij}(\mathbb{E}[V_i] - v_i)$ and applying λ-strong IA faithfulness to lower bound α_{ij}. To ensure that an intervention changes an outcome significantly, we must quantify its strength.

Definition 4.2. We say that an intervention $\mathrm{do}(v_i)$ is γ-significant if $|\mathbb{E}[V_i] - v_i| > \gamma$.

These notions are used to formally develop change sets in Appendix B, as well as a few of their important properties.

Definition 4.3 (Change Sets)**.** We define the β-change set for $\mathrm{do}(\mathbf{a})$ to be

$$\mathrm{CHG}_\beta(\mathbf{a}) \coloneqq \{V \in \mathbf{V} \;:\; |\mathbb{E}[V] - \mathbb{E}[V \mid \mathrm{do}(\mathbf{a})]| > \beta\}.$$

Definition 4.4 (Conditional Change Set)**.** We define the conditional change set for disjoint $A \in \mathbf{V}$ and $\mathbf{C} \subseteq \mathbf{V}$ for do/knockout interventions to be

$$\mathrm{CHG}_\beta(a \mid \mathrm{do}(\mathbf{c})) \coloneqq \{V \in \mathbf{V} \setminus \mathbf{C} \;: \\ |\mathbb{E}[V \mid \mathrm{do}(a), \mathrm{do}(\mathbf{c})] - \mathbb{E}[V \mid \mathrm{do}(\mathbf{c})]| > \beta\}.$$

4.1. Algorithm

We provide two algorithms for causal discovery using only interventions that use IA-faithfulness: (1) UIC, which does not limit the cardinality of an intervention, and (2) k-RIC, which does limit the cardinality. For both approaches, we start by using the change-sets on single-node interventions to find some of the causal relationships. Unfortunately, change-sets need not contain every child of the intervened node, since children with multiple causal paths may have their effects canceled out. However, the change sets do contain at least one child with only a single causal pathway. By taking the transitive closure of our output, we obtain the transitive closure of the true graph.

Both algorithms then proceed by refining this transitive closure with edge-removals by searching for redundancy between intervention sets that differ in one intervened vertex (using conditional change-sets). While UIC is able to efficiently compare interventions on large potential parent sets, k-RIC must be slightly more efficient to refine certain parts of the graph to reduce the required intervention size. Further details are given in Appendix C.

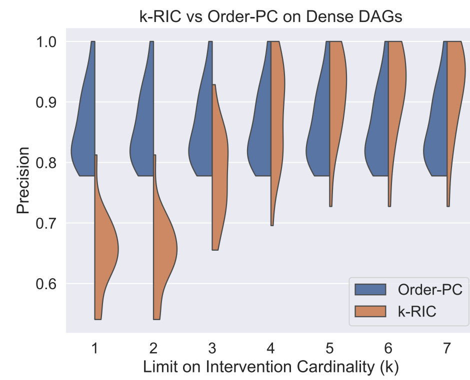


Figure 1. The precision of Order-PC vs k-RIC.

5. Empirical Verification

We use a two-population Student-t test to detect a change in the distribution mean. We use the implementation from `scipy.stats` with a significance level of .005 for step 1 (to avoid starting with a cyclic $\mathcal{G}'$), and .05 for further pruning of edges. This is compared to an implementation of the PC-algorithm (Spirtes et al., 2000) that starts with knowledge of the correct topological ordering and utilizes the Fisher-Z conditional independence tests implemented by Chandler Squires (2018). We call this comparison "Order-PC." Our k-RIC algorithm surpasses the topological-ordering-informed PC algorithm at around $k = 3$-node interventions. Figure 1 shows how the precision of k-RIC scales relative to k. Full results are given in Appendix D.

6. Conclusion

While many algorithms for causal discovery use different faithfulness assumptions whose failures correspond to a Lebesgue-measure-0 set, these assumptions have different geometries that give rise to varying levels of robustness to finite-sample noise. This paper shows that we can still recover causal structure with a much milder and well-behaved notion of intervention faithfulness.

In practice, ignoring CI testing is unnecessary. However, because CI and "change-set" detection have fundamentally different data requirements, it is worth studying their relative equivalence classes to isolate points of uncertainty that may be resolved by the other. Future work should optimize a balanced integration of both observational CI tests and interventional change-sets.

Intervention-based discovery shares many similarities with information-theoretic causal discovery (Janzing & Schölkopf, 2010; Xu et al., 2025). Our results may shed light on the relative stability of these approaches, and future work should look into further developing this connection.

Acknowledgements

CU was partially supported by NCCIH/NIH (1DP2AT012345), ONR (N00014-22-1-2116 and N00014-24-1-2687), DOE-ASCR (DE-SC0023187), the MIT-IBM Watson AI Lab, the Eric and Wendy Schmidt Center at the Broad Institute, and a Simons Investigator Award. JZ was partially supported by an Apple AI/ML PhD Fellowship. BM was supported by a postdoctoral fellowship at the Eric and Wendy Schmidt Center at the Broad Institute of MIT and Harvard. BM is now a Visiting Affiliate Scientist at the Schmidt Center.

References

Chandler Squires. `causaldag`: *creation, manipulation, and learning of causal models*, 2018. URL https://github.com/uhlerlab/causaldag.

Chevalley, M., Schwab, P., and Mehrjou, A. Deriving causal order from single-variable interventions: Guarantees & algorithm. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=u63OVngeSp.

Eberhardt, F. and Scheines, R. Interventions and causal inference. *Philosophy of science*, 74(5):981–995, 2007.

Guan, C., Ye, C., Yang, X., and Gao, J. A review of current large-scale mouse knockout efforts. *genesis*, 48(2):73–85, 2010.

Hauser, A. and Bühlmann, P. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. *The Journal of Machine Learning Research*, 13(1):2409–2464, 2012.

Hauser, A. and Bühlmann, P. Jointly interventional and observational data: estimation of interventional markov equivalence classes of directed acyclic graphs. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77(1):291–318, 2015.

Janzing, D. and Schölkopf, B. Causal inference using the algorithmic markov condition. *IEEE Transactions on Information Theory*, 56(10):5168–5194, 2010.

Kocaoglu, M. Characterization and learning of causal graphs with small conditioning sets. *Advances in Neural Information Processing Systems*, 36, 2023.

Meek, C. Strong completeness and faithfulness in bayesian networks. *arXiv preprint arXiv:1302.4973*, 2013.

Pearl, J. Graphs, causality, and structural equation models. *Sociological Methods & Research*, 27(2):226–284, 1998.

Pearl, J. *Causality*. Cambridge university press, 2009.

Ramsey, J., Zhang, J., and Spirtes, P. L. Adjacency-faithfulness and conservative causal inference. *arXiv preprint arXiv:1206.6843*, 2012.

Shanmugam, K., Kocaoglu, M., Dimakis, A. G., and Vishwanath, S. Learning causal graphs with small interventions. *Advances in Neural Information Processing Systems*, 28, 2015.

Shiragur, K., Zhang, J., and Uhler, C. Causal discovery with fewer conditional independence tests. *arXiv preprint arXiv:2406.01823*, 2024.

Spirtes, P., Glymour, C. N., and Scheines, R. *Causation, prediction, and search*. MIT press, 2000.

Squires, C. and Uhler, C. Causal structure learning: A combinatorial perspective. *Foundations of Computational Mathematics*, 23(5):1781–1815, 2023.

Squires, C., Wang, Y., and Uhler, C. Permutation-based causal structure learning with unknown intervention targets. In *Conference on Uncertainty in Artificial Intelligence*, pp. 1039–1048. PMLR, 2020.

Uhler, C., Raskutti, G., Bühlmann, P., and Yu, B. Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics*, pp. 436–463, 2013.

Wang, Y., Solus, L., Yang, K., and Uhler, C. Permutation-based causal inference algorithms with interventions. *Advances in Neural Information Processing Systems*, 30, 2017.

Xu, S., Mameche, S., and Vreeken, J. Information-theoretic causal discovery in topological order. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025. URL https://openreview.net/forum?id=9pjJXQWYXc.

Yang, K., Katcoff, A., and Uhler, C. Characterizing and learning equivalence classes of causal dags under interventions. In *International Conference on Machine Learning*, pp. 5541–5550. PMLR, 2018.

Zhang, J. and Spirtes, P. Detection of unfaithfulness and robust causal inference. *Minds and Machines*, 18:239–271, 2008.

Zhang, J. and Spirtes, P. L. Strong faithfulness and uniform consistency in causal inference. *arXiv preprint arXiv:1212.2506*, 2012.

Zhang, K., Zhu, S., Kalander, M., Ng, I., Ye, J., Chen, Z., and Pan, L. gcastle: A python toolbox for causal discovery, 2021.

.1. Related Works

Interventions in Causal Discovery Seminal work by Eberhardt & Scheines (2007) studied the integration of interventions into causal discovery, particularly on orienting Markov equivalence classes. Shanmugam et al. (2015) studied the effect of limiting the size of these interventions. Shanmugam et al. (2015) explored limited intervention size when orienting a Markov equivalence class, but not learning the graph. Hauser & Bühlmann (2015) gave Markov equivalence classes after changes from an intervention set. This provides a critical framework for utilizing an intervention-first approach to causal discovery. Interventional data has also been incorporated to inform direction in causal discovery algorithms (Hauser & Bühlmann, 2012; Wang et al., 2017; Yang et al., 2018). Squires et al. (2020) utilized interventions to first establish a permutation ordering before performing causal discovery.

Limiting Conditioning and Conditional Independence Shiragur et al. (2024) showed that incorporating topological information in adjacency search can greatly reduces the number of conditional independencies needed to recover a DAG. Kocaoglu (2023) characterized the equivalence classes of graphs with conditional independence tests with restricted conditioning sets.

A. Faithfulness Volumes

A.1. CI Faithfulness

Uhler et al. (2013) showed that λ-strong CI faithfulness is a relatively strong assumption. Theorems 5.2 and 5.3 from Uhler et al. (2013) are summarised in Theorem A.1.

Theorem A.1 (informal, (Uhler et al., 2013))**.** *The volume of λ-strong CI faithfulness violations with linear structural equations and additive Gaussian noise with all coefficients in $[-1, 1]$ and $\lambda \in (0, 1)$ is at least $\omega(\lambda^{poly(n)} 2^{|\mathbf{E}|})$ in the worst case.*

A.2. IA Faithfulness

To formalize that IA faithfulness is milder than CI faithfulness, we show that the growth of λ-strong IA faithfulness violations is linear in λ.

Theorem A.2. *The volume of λ-strong intervention faithfulness violations with all coefficients in $[-1, 1]$ and $\lambda \in (0, 1)$ is $\mathcal{O}(\lambda 2^{|\mathbf{E}|})$.*

Proof. Observe that λ-strong intervention faithfulness violations correspond to the set of distributions with at least one parameter $< \lambda$. For each parameter, this corresponds to the volume between two hyperplanes that are separated by a distance 2λ. Hence, the region of λ-strong intervention faithfulness violations corresponds to the sum of these regions for each edge's parameter. □

This illustrates that using interventional information requires much milder assumptions than utilizing CI tests. Since some skeletal information can be learned from interventions, we argue that interventions should precede the results of CI tests in mixed intervention and observational data settings. In order to understand the limitations of the skeletal information from interventions, we study intervention-only models.

B. Change Sets

The graphical information gained from knockout interventions comes from the observation that "changing a variable will change its effects, but not its causes." For example, if we have $A \to B$, then knocking out A will correspond to some change in B, but knocking out B will correspond to no change in A. We will abstract the graphical information gained from a knockout on $\mathbf{a}$ under IA faithfulness as a subset $\mathbf{C} \subseteq \mathbf{V}$ of the variables for which $\mathbb{E}[C] \neq \mathbb{E}[C \mid \mathtt{do}(\mathbf{a})]$ for all $C \in \mathbf{C}$.

Definition B.1 (Change Sets)**.** We define the β-change set for $\mathtt{do}(\mathbf{a})$ to be

$$\text{CHG}_\beta(\mathbf{a}) \coloneqq \{V \in \mathbf{V} \ : \ |\mathbb{E}[V] - \mathbb{E}[V \mid \mathtt{do}(\mathbf{a})]| > \beta\}.$$

β-change sets always return some subset of the descendants, since do-interventions only change causally downstream variables. It is not hard to see that 0-change sets are a query of descendants under stronger notions of intervention faithfulness that require changes in all descendants. Without this stronger notion of intervention faithfulness, β-change sets do not give quite as much information, since there may be a decay in the change in the expected value of farther downstream variables, or even canceling out of multiple paths.

Lemma B.2. *Under λ-strong IA faithfulness and a γ-significant do-intervention on $A \in \mathbf{V}$, if $A \to V$ with no other directed paths from A to V then $V \in \mathrm{CHG}_{\lambda\gamma}(a)$.*

Lemma B.2 is limited because of our mild faithfulness requirement, which does not require distance > 1 paths to exhibit significant change, and does not even require distance 1 paths to exhibit change if other paths cancel out.

B.1. Conditional Change Sets

The nature of do interventions allows us to query slightly more than change sets because they can set variables to *specific* values (e.g. 0, for knockout interventions).

This allows us to compare the distributions of two overlapping interventions to capture redundancies. For example, if

$$\begin{aligned} X_1 &= N_1 \\ X_2 &= \alpha_{12} X_1 + N_2 \\ X_3 &= \alpha_{13} X_1 A + \alpha_{23} X_2 + N_3 \end{aligned}$$

then if we compare a knockout on X_2 and a knockout on X_1, X_2, the change X_3 will only be different if $\alpha_{13} \neq 0$.

Definition B.3 (Conditional Change Set)**.** We define the conditional change set for disjoint $A \in \mathbf{V}$ and $\mathbf{C} \subseteq \mathbf{V}$ for do/knockout interventions to be

$$\mathrm{CHG}_\beta(a \mid \mathtt{do}(\mathbf{c})) \coloneqq \{V \in \mathbf{V} \setminus \mathbf{C} \ : \ |\mathbb{E}[V \mid \mathtt{do}(a), \mathtt{do}(\mathbf{c})] - \mathbb{E}[V \mid \mathtt{do}(\mathbf{c})]| > \beta\}.$$

As we have hinted at with our notation, one can think of a $\mathrm{CHG}(a \mid \mathtt{do}(\mathbf{c}))$ as a "conditional intervention," i.e., testing the outcome of intervening on A after conditioning on $\mathbf{C}$. Use $\mathcal{G} \setminus \mathbf{C}$ to denote the graph with all $\mathbf{C}$ removed as well as edges to and from $\mathbf{C}$, we have the following.

Lemma B.4. *For DAG $\mathcal{G}$, remove $A \to V$ (if it is there) to get $\mathcal{G}^-$ and let $\mathbf{C}$ be a vertex cut that separates $A, V \in \mathbf{V}$ into two different connected components in $\mathcal{G}^-$. Under λ-strong IA faithfulness and a γ-significant do-intervention on A, $V \in \mathrm{CHG}_{\lambda\gamma}(a \mid do(\mathbf{c}))$ if and only if $A \to V$ is in $\mathcal{G}$.*

Proof. Let $\mathcal{G}^{\text{int}}$ be $\mathcal{G}$ with all arrows coming in or out of $\mathbf{C}$ removed. The intervened distribution $\mathtt{do}(\mathbf{c})$ is Markovian in $\mathcal{G}^{\text{int}}$. If $A \to V$ is not in $\mathcal{G}$, then V is not a descendant of A in $\mathcal{G}^{\text{int}}$, meaning V will not change when adding an intervention in A. If $A \to V$ is in $\mathcal{G}$, then there are no other paths from A to V in $\mathcal{G}^{\text{int}}$, so by Lemma B.2, V is in the change set created by intervening on A and $\mathbf{C}$. □

Lemma B.4 allows us to verify and disprove an adjacency between A and V by looking for a vertex cut $\mathbf{C}$ that separates them. While this information may seem mild, conditional change-sets are sufficient to recover the complete causal structure

C. Algorithm

We will now provide an algorithm that makes use of IA faithfulness and change sets to learn a causal DAG. For this entire section, $\mathcal{G}$ will denote the true graph with n vertices, and we will assume that the empirical distribution is λ-strong IA faithful and all interventions are γ-significant.

A critical level of complexity is the "cardinality" of the interventions that we can perform, $|\mathbf{A}|$. We can be very efficient if the cardinality is unlimited, but real-world applications are rarely so flexible. As such, we will give less efficient algorithms as we limit the cardinality of interventions, eventually proceeding into the regime where the full graph can only be narrowed into equivalence classes, which may be further resolved by conditional independence testing.

C.0.1. Unrestricted Intervention Cardinality

When the cardinality of interventions is not limited, we can construct a very efficient set of interventions according to what we will call the Unrestricted Intervention Cardinality (UIC) algorithm. The correct graph is $\mathcal{G} = (\mathbf{V}, \mathbf{E})$.

1. For all $V_i \in \mathbf{V}$, intervene on $V_i \in \mathbf{V}$ and add $V_i \to V_j$ if $V_j \in \mathrm{CHG}_{\gamma\lambda}(v_i)$. Find the transitive closure and call the resulting graph $\mathcal{G}'$.
2. Start with $\mathcal{G}'' = \mathcal{G}'$ and iterate through each $V_j \in \mathbf{V}$ according to a topological order and then iterate through $V_i \in \mathbf{PA}_{\mathcal{G}'}(V_j)$, remove $V_i \to V_j$ from $\mathcal{G}''$ if $V_j \notin \mathrm{CHG}_{\gamma\lambda}(v_i \mid \mathtt{do}(\mathbf{pa}_{\mathcal{G}''}(V_j) \setminus \{v_i\}))$.

To prove correctness, we will first show that step one gives us the transitive closure of the true graph.

Lemma C.1. *$\mathcal{G}'$ is the transitive closure of $\mathcal{G}$ at the end of step 1.*

Proof. Every vertex V_i has at least one $V_j \in \mathbf{CH}_{\mathcal{G}}(V_i)$ with no other directed paths between V_i, V_j (the next vertex in the topological order). These edges are detected by Lemma B.4, which are then chained together by the transitive closure. □

We now only need to show that step 2 removes all incorrect edges and does not remove correct edges.

Lemma C.2. *For every V_i, V_j with $V_j \notin \mathrm{CHG}_{\gamma\lambda}(v_i \mid \mathbf{PA}_{\mathcal{G}''}(V_j) \setminus \{V_i\})$ if and only if $V_i \to V_j \notin \mathbf{E}$.*

Proof. We know that $\mathbf{PA}_{\mathcal{G}}(V_i) \subseteq \mathbf{PA}_{\mathcal{G}''}(V_i)$ and $\mathbf{PA}_{\mathcal{G}}(V_j) \setminus \{V_i\}$ is a vertex-cut for V_i, V_j if $V_i \to V_j$ is not in $\mathcal{G}$. This allows us to apply Lemma B.4. □

Theorem C.3. *For a distribution that is λ-strong IA faithful in $\mathcal{G}$, the UIC algorithm learns $\mathcal{G}$ in using $\mathcal{O}(n^2)$ γ-significant interventions of cardinality of up to $n-1$.*

Step 1 does one intervention per vertex, and step 2 does a maximum of one intervention per parent in $\mathcal{G}'$ per vertex, which is $\mathcal{O}(n^2)$.

C.0.2. Minimum Intervention Cardinality

The UIC algorithm potentially requires simultaneous interventions on $n-1$ nodes of the graph, which might be unrealistic in applied settings. Hence, we will now limit $|\mathbf{A}| \leqslant k$. The following k-Restricted Intervention Cardinality (k-RIC) algorithm shows that we can still get identifiability when $\kappa \leqslant k < n-1$, where κ is the vertex connectivity of the true DAG $\mathcal{G}$. The first step is the same as UIC, with the following modification for the second step:

2. Start with $\mathcal{G}'' = \mathcal{G}'$ and iterate through each $V_j \in \mathbf{V}$ according to a topological order and then iterate through $V_i \in \mathbf{PA}_{\mathcal{G}'}(V_j)$:
   - Now let $\mathcal{G}^-$ be $\mathcal{G}''$ with $V_i \to V_j$ removed. Find the smallest $\mathbf{Z} \subset \mathbf{PA}_{\mathcal{G}'}(V_i) \setminus \{V_i\}$ with $|\mathbf{Z}| < k$ that forms a vertex cut in $\mathcal{G}^-$ separating V_i and V_j.
   - Test whether $V_j \in \mathrm{CHG}_{\gamma\lambda}(v_i \mid \mathtt{do}(\mathbf{z}))$ and remove $V_i \to V_j$ if not.

The k-RIC algorithm is identical to the UIC algorithm, but searches for a smaller vertex cut that separates V_i, V_j instead of using $\mathbf{PA}_{\mathcal{G}''}(V_j) \setminus \{V_i\}$. Such a set must be in $\mathbf{AN}_{\mathcal{G}}(V_j) = \mathbf{PA}_{\mathcal{G}'}(V_j)$, so the algorithm will return the correct $\mathcal{G}$ so long as a small-enough $\mathbf{Z}$ exists.

Theorem C.4. *For a distribution that is λ-strong IA faithful in $\mathcal{G}$ with vertex connectivity κ, the k-RIC algorithm with $k \geqslant \kappa + 1$ learns $\mathcal{G}$ using $\mathcal{O}(n^2)$ γ-significant interventions.*

Theorem 1.1 is a direct corollary of Theorem C.4.

Proof. First, we observe that the first V_j considered correctly is a source (no incoming edges), since we know from Lemma C.1 that the topological ordering is correct. We will now induct on the iteration through V_js, assuming all ancestors have correctly specified all incoming edges (and lack of incoming edges) in $\mathcal{G}''$. Notice that this assumption means that the edges between vertices in $\mathbf{PA}(V_j)$ are the same in both $\mathcal{G}''$ and $\mathcal{G}$.

We know that there exists some minimum vertex cut with cardinality $|\mathbf{Z}| = \kappa$ that separates V_i from V_j in $\mathcal{G}$ with $V_i \to V_j$ removed. In order for $\mathbf{Z}$ to not be a vertex cut in $\mathcal{G}''$ at this step, we would need to have a directed path from V_i through some vertices in $\mathbf{PA}(V_j) \cap \mathbf{DE}(V_i)$ that ends at V_j. By the inductive assumption, there are no additional edges that can create such a path. We deduce that the minimum vertex cut that separates V_i and V_j in $\mathcal{G}$ (with $V_i \to V_j$ removed) is also a minimum vertex cut in $\mathcal{G}''$ (with $V_i \to V_j$ removed). This ensures that $V_i \to V_j$ edges are correctly removed or added for all V_i. □

C.0.3. Partial Identification

When restricting $|\mathbf{Z}| < k$, we lose identifiability. To see this, consider the following two graphs: (1) $A \to B \to C$ and (2) $A \to B \to C$, $A \to C$. Suppose we run the 1-RIC algorithm on single-node interventions. For both causal graphs, (i) $\mathrm{do}(a)$ changes B, C, (ii) $\mathrm{do}(b)$ changes C, and (iii) $\mathrm{do}(c)$ changes nothing. Hence, we cannot recover the difference between these two causal graphs without a CI test or a double-node (i.e., cardinality 2) intervention.

Generally, any two graphs with the same transitive closure are (potentially) indistinguishable under single-node interventions. With higher cardinality interventions, we can partially prune extra children using conditional change sets. However, with a limitation on the cardinality of interventions, all pairs of non-adjacent vertices V_i, V_j that cannot be separated using a vertex-cut of cardinality $k-1$ could be connected with orientation determined by the topological order.

Definition C.5. The k-robust transitive closure of $\mathcal{G}$ involves adding edges $V_i \to V_j$ whenever the minimum vertex cut that separates V_i and V_j has cardinality greater than $k-1$.

Theorem C.6. *For a distribution that is λ-strong IA faithful in $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, the k-RIC algorithm recovers a $\mathcal{G}$ up to its k-robust transitive closure using $\mathcal{O}(n^2)$ γ-significant interventions.*

Proof. For all $V_i \to V_j$ not in $\mathcal{G}$ that can be separated by a cardinality $k-1$ vertex cut, the k-RIC algorithm will find that vertex cut during step 2 and remove that edge. For all other non-adjacencies, the edge will not be removed. $\square$

Theorem 1.2 is a corollary of Theorem C.6.

D. Empirical Study Details and Results

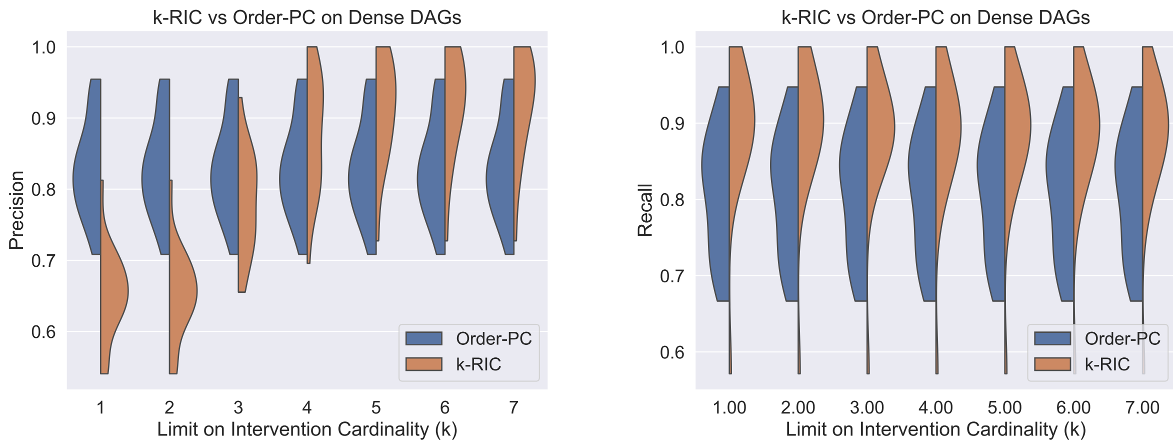
When implementing intervention-based causal discovery, we can use a two-population Student-t test to detect a change in the distribution mean. We use the implementation from `scipy.stats` with a significance level of .005 for step 1 (to avoid starting with a cyclic $\mathcal{G}'$), and .05 for further pruning of edges. This is compared to an implementation of the PC-algorithm (Spirtes et al., 2000) that starts with knowledge of the correct topological ordering and utilizes the Fisher-Z conditional independence tests implemented by Chandler Squires (2018).

We used `g-castle` (Zhang et al., 2021) to generate 20 random DAGs that are expected to have 25 edges on 10 nodes via Erdős-Reñyi with random edge weights between 0 and 1. We report both the precision and recall of the recovered edges. We sample 1000 points from the observational distribution and 1000 points for each knockout to the value 0.

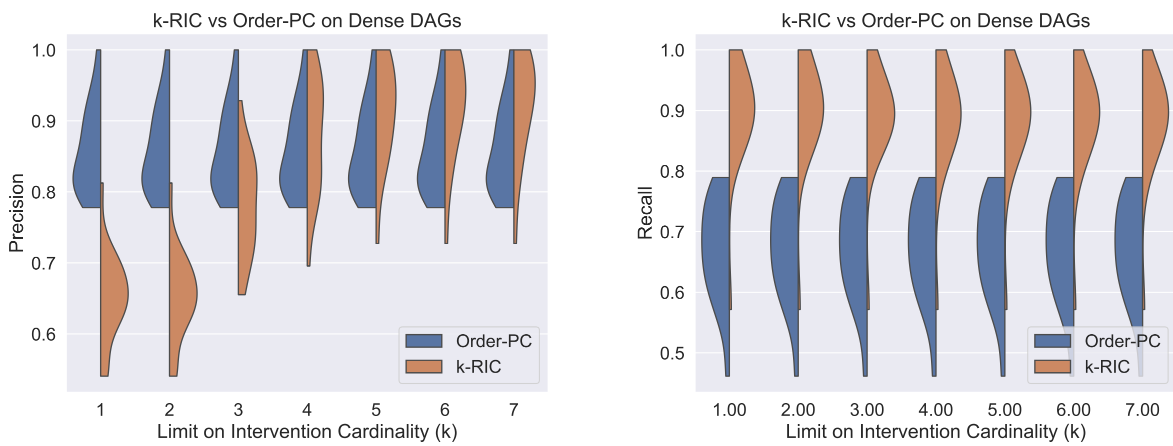
Intervention-based causal discovery is a type of online learning that requires gathering additional (perturbed) data. With this in mind, we first run the k-RIC algorithm and keep track of how many samples it requires. We then compare k-RIC to the ordered PC algorithm on the total number of data points used by the k-RIC algorithm. Note that this provides significantly more data for each conditional independence test than is obtained for each perturbation. Furthermore, the PC algorithm is given a significant advantage by receiving a perfect topological ordering that would normally be obtained by interventions, while k-RIC receives noisy topological ordering information and no conditional independence information.

The results are shown in Figure 2. Despite the advantages given to the order-informed PC-algorithm, the intervention-based k-RIC performs significantly better with respect to recall. For precision, the intervention-based approach begins slightly worse than PC when we are limited to single-node interventions. However, with $k \geqslant 3$-node interventions, the k-RIC algorithm matches and then exceeds the performance of the PC algorithm.

An arguably fairer comparison would be to give order-informed PC the same amount of observational data that k-RIC utilizes for its empty-set interventions. In this setting, k-RIC vastly outperforms PC, which struggles in recall without sufficient data to resolve violations in strong faithfulness. PC appears to perform well in precision with less data because it has removed too many edges.



(a) Precision and recall for 20 random DAGs when giving Order-PC the same amount of data as k-RIC utilizes in total.



(b) Precision and recall for 20 random DAGs when comparing equal amounts of observational data.

Figure 2. Each choice of k is tested on the same 20 randomly drawn DAGs with random edge weights between 0 and 1 and additive noise in $N(1, .5)$. Experiments take a few minutes to run on one CPU.