Characterizing and Learning Equivalence Classes of Causal DAGs under Interventions

Karren D. Yang [1] **Abigail Katcoff** [1] **Caroline Uhler** [1]

Abstract

We consider the problem of learning causal DAGs in the setting where both observational and interventional data is available. This setting is common in biology, where gene regulatory networks can be intervened on using chemical reagents or gene deletions. Hauser & Bühlmann (2012) previously characterized the identifiability of causal DAGs under perfect interventions, which eliminate dependencies between targeted variables and their direct causes. In this paper, we extend these identifiability results to *general interventions*, which may modify the dependencies between targeted variables and their causes without eliminating them. We define and characterize the *interventional Markov equivalence class* that can be identified from general (not necessarily perfect) intervention experiments. We also propose the first provably consistent algorithm for learning DAGs in this setting and evaluate our algorithm on simulated and biological datasets.

1. Introduction

The problem of learning a causal *directed acyclic graph* (DAG) from observational data over its nodes is important across disciplines such as computational biology, sociology, and economics (Friedman et al., 2000; Pearl, 2003; Robins et al., 2000; Spirtes et al., 2000). A causal DAG imposes conditional independence (CI) relations on its node variables that can be used to infer its structure. Since multiple DAGs can encode the same CI relations, a causal DAG is generally only identifiable up to its *Markov equivalence class* (MEC) (Verma & Pearl, 1991; Andersson et al., 1997).

The identifiability of causal DAGs can be improved by performing *interventions* on the variables. Interventions that eliminate the dependency between targeted variables and their causes are known as *perfect* (or *hard*) interventions (Eberhardt et al., 2005). Under perfect interventions, the identifiability of causal DAGs improves to a smaller equivalence class called the *perfect-$\mathcal{I}$-MEC*[1] (Hauser & Bühlmann, 2012). Recently, Wang et al. (2017) proposed the first provably consistent algorithm for recovering the perfect-$\mathcal{I}$-MEC and successfully applied it towards learning regulatory networks from interventional data.

However, only considering perfect interventions is restrictive: in practice, many interventions are *non-perfect* (or *soft*) and modify the causal relations between targeted variables and their direct causes without eliminating them (Eberhardt et al., 2005). In genomics, for example, interventions such as RNA interference or CRISPR-mediated gene activation often have only modest effects on gene suppression and activation respectively (Dominguez et al., 2016). Even interventions meant to be perfect, such as CRISPR/Cas9-mediated gene deletions, may not be uniformly successful across a cell population (Dixit et al., 2016). Although non-perfect interventions may be considered inefficient from an engineering perspective, they may still provide valuable information about regulatory networks. The identifiability of causal DAGs in this setting needs to be formally analyzed to develop maximally effective algorithms for learning from these types of interventions.

In this paper, we define and characterize *$\mathcal{I}$-Markov equivalence classes ($\mathcal{I}$-MECs)* of causal DAGs that can be identified from *general* interventions that are not assumed to be perfect, thus extending the results of Hauser & Bühlmann (2012) (Section 3). We show that under reasonable assumptions on the experiments, general interventions provide the same causal information as perfect interventions. These insights allow us to develop the first *provably consistent algorithm* for learning the $\mathcal{I}$-MEC from data from general interventions (Section 4), which we evaluate on synthetic and biological datasets (Section 5).

[1]Massachusetts Institute of Technology, Cambridge, MA. Correspondence to: Caroline Uhler <cuhler@mit.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

[1]In Hauser & Bühlmann (2012), they call this the interventional MEC ($\mathcal{I}$-MEC). We call it the perfect-$\mathcal{I}$-MEC to avoid confusion with the equivalence class for DAGs under general interventions that we characterize in this paper, which we call the $\mathcal{I}$-MEC.

2. Related Work

2.1. Identifiability of causal DAGs

Given only observational data and without further distributional assumptions[2], the identifiability of a causal DAG is limited to its MEC (Verma & Pearl, 1991). Hauser & Bühlmann (2012) proved that a smaller class of DAGs, the perfect-$\mathcal{I}$-MEC, can be identified given data from perfect interventions. They conjectured but did not prove that their results extend to soft interventions. For general interventions, Tian & Pearl (2001) presented a graph-based criterion for two DAGs being indistinguishable under single-variable interventions. Their criterion is consistent with Hauser and Bühlmann's perfect-$\mathcal{I}$-MEC, but they did not discuss equivalence classes, nor did they consider multi-variable interventions. Eberhardt & Scheines (2007) and Eberhardt (2008) provided results on the number of single-target interventions required for full identifiability of the causal DAG. However, their work does not characterize equivalence classes for when the DAG is only partially identifiable.

2.2. Causal inference algorithms

There are two main categories of algorithms for learning causal graphs from observational data: *constraint-based* and *score-based* (Brown et al., 2005; Murphy, 2001). Constraint-based algorithms, such as the prominent PC algorithm (Spirtes et al., 2000), view causal inference as a constraint satisfaction problem based on CI relations inferred from data. Score-based algorithms, such as greedy equivalence search (GES) (Chickering, 2002), maximize a particular score function over the space of graphs. Hybrid algorithms such as greedy sparsest permutation (GSP) combine elements of both methods (Solus et al., 2017).

Algorithms have also been developed to learn causal graphs from both observational and interventional data. GIES is an extension of GES that incorporates interventional data into the score function it uses to search over the space of DAGs (Hauser & Bühlmann, 2012), but it is in general not consistent (Wang et al., 2017). *Perfect interventional GSP* (perfect-IGSP) is a provably consistent extension of GSP that uses interventional data to reduce the search space and orient edges, but it requires perfect interventions (Wang et al., 2017). Methods that allow for latent confounders and unknown intervention targets include Eaton & Murphy (2007), JCI (Magliacane et al., 2016), HEJ (Hyttinen et al., 2014), CombINE (Triantafillou & Tsamardinos, 2015), and ICP (Peters et al., 2016), but they do not have consistency guarantees for returning a DAG in the correct class.

[2]See Shimizu et al. (2006), Hoyer et al. (2009), Peters et al. (2014) for identifiability results for non-Gaussian or nonlinear structural equation models.

3. Identifiability under general interventions

In this section, we characterize the $\mathcal{I}$-MEC: a smaller equivalence class than the MEC that can be identified under general interventions with known targets. The main result is a graphical criterion for determining whether two DAGs are $\mathcal{I}$-Markov equivalent, which extends the identifiability results of Hauser & Bühlmann (2012) from perfect interventions to general interventions.

3.1. Preliminaries

Let the causal DAG $\mathcal{G} = ([p], E)$ represent a causal model in which every node $i \in [p]$ is associated with a random variable X_i, and let f denote the joint probability distribution over $X = (X_1, \cdots, X_p)$. Under the causal Markov assumption, f satisfies the *Markov property* (or *is Markov*) with respect to $\mathcal{G}$, i.e., $f(X) = \prod_i f(X_i|X_{\mathrm{pa}_\mathcal{G}(i)})$, where $\mathrm{pa}_\mathcal{G}(i)$ denotes the parents of node i in $\mathcal{G}$ (Lauritzen, 1996).

Let $\mathcal{M}(\mathcal{G})$ denote the set of strictly positive densities that are Markov with respect to $\mathcal{G}$. Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ for which $\mathcal{M}(\mathcal{G}_1) = \mathcal{M}(\mathcal{G}_2)$ are said to be *Markov equivalent* and belong to the same MEC (Andersson et al., 1997). Verma & Pearl (1991) gave a graphical criterion for Markov equivalence: two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ belong to the same MEC if and only if they have the same skeleta (i.e., underlying undirected graph) and v-structures (i.e., induced subgraphs $i \to j \leftarrow k$).

Under perfect interventions, the identifiability of $\mathcal{G}$ improves from its MEC to its perfect-$\mathcal{I}$-MEC, which has the following graphical characterization (Hauser & Bühlmann, 2012).

Theorem 3.1. *Let $\mathcal{I} \subset \mathfrak{P}([p])$[3] be a conservative (multi)-set of intervention targets, i.e. $\forall j \in [p]$, $\exists I \in \mathcal{I}$ s.t. $j \notin I$. Two DAGs $\mathcal{G}, \mathcal{H}$ belong to the same perfect-$\mathcal{I}$-MEC if and only if $\mathcal{G}_{(I)}$, $\mathcal{H}_{(I)}$ are in the same MEC for all $I \in \mathcal{I}$, where $\mathcal{G}_{(I)}$ denotes the sub-DAG of $\mathcal{G}$ with vertex set $[p]$ and edge set $\{(a \to b)|(a \to b) \in E, b \notin I\}$ and similarly for $\mathcal{H}_{(I)}$.*

In this work, we extend this result to general interventions.

Definition 3.2. Under a (general) *intervention* on target $I \subset [p]$, the *interventional distribution* $f^{(I)}$ can be factorized as

$$f^{(I)}(X) = \prod_{i \in I} f^{(I)}(X_i|X_{\mathrm{pa}_\mathcal{G}(i)}) \prod_{j \notin I} f^{(\emptyset)}(X_j|X_{\mathrm{pa}_\mathcal{G}(j)}) \tag{1}$$

where $f^{(I)}$ and $f^{(\emptyset)}$ denote the interventional and observational distributions over X respectively. Note that $f^{(I)}(X_j|X_{\mathrm{pa}_\mathcal{G}(i)}) = f^{(\emptyset)}(X_j|X_{\mathrm{pa}_\mathcal{G}(i)})$, $\forall j \notin I$, i.e. the conditional distributions of non-targeted variables are *invariant* to the intervention.

[3]power set of $[p]$

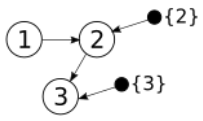


Figure 1. Let $\mathcal{G}$ be the DAG $1 \to 2 \to 3$ and let $\mathcal{I} = \{\emptyset, \{2\}, \{3\}\}$. The interventional DAG $\mathcal{G}^{\mathcal{I}}$ is shown above. Solid circles represent the $\mathcal{I}$-vertices, which are parameters indicating the intervention, and open circles represent random variables.

3.2. Main Results

Let $\{f^{(I)}\}_{I\in\mathcal{I}}$ denote a collection of distributions over X indexed by $I \in \mathcal{I}$.

Definition 3.3. For a DAG $\mathcal{G}$ and interventional target set $\mathcal{I}$, define

$$\mathcal{M}_{\mathcal{I}}(\mathcal{G}) := \{\, \{f^{(I)}\}_{I\in\mathcal{I}} \mid \forall I, J \in \mathcal{I} : f^{(I)} \in \mathcal{M}(\mathcal{G}) \text{ and } f^{(I)}(X_j|X_{\mathrm{pa}_{\mathcal{G}}(j)}) = f^{(J)}(X_j|X_{\mathrm{pa}_{\mathcal{G}}(j)}), \forall j \notin I \cup J\}$$

$\mathcal{M}_{\mathcal{I}}(\mathcal{G})$ contains exactly the sets of interventional distributions (Definition 3.2) that can be generated from a causal model with DAG $\mathcal{G}$ by intervening on $\mathcal{I}$ (see Supplementary Material for details). We therefore use $\mathcal{M}_{\mathcal{I}}(\mathcal{G})$ to formally define equivalence classes of DAGs under interventions.

Definition 3.4 ($\mathcal{I}$-Markov Equivalence Class)**.** Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ for which $\mathcal{M}_{\mathcal{I}}(\mathcal{G}_1) = \mathcal{M}_{\mathcal{I}}(\mathcal{G}_2)$ belong to the same *$\mathcal{I}$-Markov equivalence class ($\mathcal{I}$-MEC).*

From here, we extend the Markov property to the interventional setting to establish a graphical criterion for $\mathcal{I}$-MECs. We start by introducing the following graphical framework for representing DAGs under interventions.

Definition 3.5. Let $\mathcal{G} = ([p], E)$ be a DAG and let $\mathcal{I}$ be a collection of intervention targets. The *interventional DAG*[4] *($\mathcal{I}$-DAG)* $\mathcal{G}^{\mathcal{I}}$ is the graph $\mathcal{G}$ augmented with $\mathcal{I}$-vertices $\{\zeta_I\}_{I\in\mathcal{I}, I\neq\emptyset}$ and $\mathcal{I}$-edges $\{\zeta_I \to i\}_{i\in I\in\mathcal{I}, I\neq\emptyset}$.

Figure 1 gives a concrete example of an $\mathcal{I}$-DAG. Note that each $\mathcal{I}$-vertex represents an intervention, and an $\mathcal{I}$-edge from an $\mathcal{I}$-vertex to a regular node i indicates that i is targeted under that intervention. Next, we define the $\mathcal{I}$-Markov property for $\mathcal{I}$-DAGs, analogous to the Markov property based on d-separation for DAGs. For now, we make the simplifying assumption that $\emptyset \in \mathcal{I}$; in Section 3.3, we will show that this assumption can be made without loss of generality.

[4]In some previous work, interventions have been treated as additional variables of the causal system, which at first glance results in a DAG similar to the $\mathcal{I}$-DAG. The challenge then is that the new variables are deterministically related to each other, which leads to faithfulness violations (see Magliacane et al. (2016)). We have avoided this problem by treating the interventions as parameters instead of variables.

Definition 3.6 ($\mathcal{I}$-Markov Property)**.** Let $\mathcal{I}$ be a set of intervention targets such that $\emptyset \in \mathcal{I}$, and suppose $\{f^{(I)}\}_{I\in\mathcal{I}}$ is a set of (strictly positive) probability distributions over $X_1, \cdots, X_p$ indexed by $I \in \mathcal{I}$. $\{f^{(I)}\}_{I\in\mathcal{I}}$ satisfies the *$\mathcal{I}$-Markov property* with respect to the $\mathcal{I}$-DAG $\mathcal{G}^{\mathcal{I}}$ iff

1. $X_A \perp\!\!\!\perp X_B \mid X_C$ for any $I \in \mathcal{I}$ and any disjoint $A, B, C \subset [p]$ such that C d-separates A and B in $\mathcal{G}$.
2. $f^{(I)}(X_A|X_C) = f^{(\emptyset)}(X_A|X_C)$ for any $I \in \mathcal{I}$ and any disjoint $A, C \subset [p]$ such that $C \cup \zeta_{\mathcal{I}\setminus I}$ d-separates A and ζ_I in $\mathcal{G}^{\mathcal{I}}$, where $\zeta_\emptyset := \emptyset$ and $\zeta_{\mathcal{I}\setminus I} := \{\zeta_J \mid J \in \mathcal{I}, J \neq I\}$.

The first condition is simply the Markov property for DAGs based on d-separation. The second condition generalizes this property to $\mathcal{I}$-DAGs by relating d-separation between $\mathcal{I}$-vertices and regular vertices to the *invariance* of conditional distributions across interventions. We note that the $\mathcal{I}$-Markov property is very similar to the "missing-link compatibility" by Bareinboim et al. (2012)

Example 3.7. *Consider again the augmented graph $\mathcal{G}^{\mathcal{I}}$ from Figure 1, and suppose $\{f^{(I)}\}_{I\in\mathcal{I}}$ satisfies the $\mathcal{I}$-Markov property with respect to $\mathcal{G}^{\mathcal{I}}$. Then $\{f^{(I)}\}_{I\in\mathcal{I}}$ satisfies the following invariance relations based on d-separation: (1) $f^{(\emptyset)}(X_1) = f^{(\{2\})}(X_1) = f^{(\{3\})}(X_1)$; (2) $f^{(\emptyset)}(X_3|X_2) = f^{(\{2\})}(X_3|X_2)$; (3) $f^{(\emptyset)}(X_2|X_1) = f^{(\{3\})}(X_2|X_1)$.*

Having defined the $\mathcal{I}$-Markov property, we now formalize its relationship to $\mathcal{I}$-MECs.

Proposition 3.8. *Suppose $\emptyset \in \mathcal{I}$. Then $\{f^{(I)}\}_{I\in\mathcal{I}} \in \mathcal{M}_{\mathcal{I}}(\mathcal{G})$ if and only if $\{f^{(I)}\}_{I\in\mathcal{I}}$ satisfies the $\mathcal{I}$-Markov property with respect to $\mathcal{G}^{\mathcal{I}}$.*

This result states that DAGs are in the same $\mathcal{I}$-MEC if and only if the d-separation statements of their $\mathcal{I}$-DAGs imply the same conditional invariances and independences based on the $\mathcal{I}$-Markov property. We now state the main result of this section: the graphical characterization of $\mathcal{I}$-MECs.

Theorem 3.9. *Suppose $\emptyset \in \mathcal{I}$. Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ belong to the same $\mathcal{I}$-MEC if and only if their $\mathcal{I}$-DAGs $\mathcal{G}_1^{\mathcal{I}}$ and $\mathcal{G}_2^{\mathcal{I}}$ have the same skeleta and v-structures.*

The proof of this theorem uses the following weak completeness result for the $\mathcal{I}$-Markov property.

Lemma 3.10. *For any disjoint $A, C \subset [p]$ and any $J \in \mathcal{I}$ such that $C \cup \zeta_{\mathcal{I}\setminus J}$ does not d-separate A and ζ_J in $\mathcal{G}^{\mathcal{I}}$, there exists some $\{f^{(I)}\}_{I\in\mathcal{I}}$ that satisfies the $\mathcal{I}$-Markov property with respect to $\mathcal{G}^{\mathcal{I}}$ with $f^{(\emptyset)}(X_A|X_C) \neq f^{(J)}(X_A|X_C)$.*

Proof of Theorem 3.9. If $\mathcal{G}_1^{\mathcal{I}}$ and $\mathcal{G}_2^{\mathcal{I}}$ have the same skeleta and v-structures, then they satisfy the same d-separation

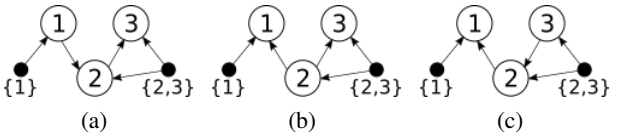


Figure 2. Example of $\mathcal{I}$-DAGs for 3-node graphs with $\mathcal{I} = \{\emptyset, \{1\}, \{2,3\}\}$

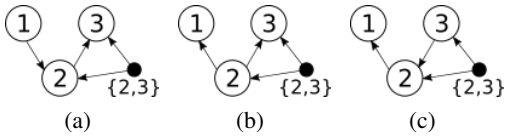


Figure 3. Example of $\tilde{\mathcal{I}}_{\{2\}}$-DAGs for 3-node graphs with $\mathcal{I} = \{\{2\}, \{3\}\}$. Note that the $\tilde{\mathcal{I}}_{\{3\}}$-DAGs are identical since $\tilde{\mathcal{I}}_{\{2\}} = \tilde{\mathcal{I}}_{\{3\}} = \{\emptyset, \{2,3\}\}$ in this case.

statements, and hence $\mathcal{M}_{\mathcal{I}}(\mathcal{G}_1) = \mathcal{M}_{\mathcal{I}}(\mathcal{G}_2)$ by Proposition 3.8. If $\mathcal{G}_1^{\mathcal{I}}$ and $\mathcal{G}_2^{\mathcal{I}}$ do not have the same skeleta or v-structures, then (a) $\mathcal{G}_1$ and $\mathcal{G}_2$ do not have the same skeleta or v-structures, or (b) there exists $I \in \mathcal{I}$ and $j \in [p]$ such that $\zeta_I \to j$ is part of a v-structure in one $\mathcal{I}$-DAG and not the other. In case (a), $\mathcal{G}_1$ and $\mathcal{G}_2$ do not belong to the same MEC (Verma & Pearl, 1991), so they also cannot belong to the same $\mathcal{I}$-MEC by the first condition in Definition 3.6. In case (b), suppose without loss of generality that $\zeta_I \to j$ is part of a v-structure in $\mathcal{G}_1^{\mathcal{I}}$ but not in $\mathcal{G}_2^{\mathcal{I}}$ for some $I \in \mathcal{I}$ and some $j \in [p]$. Then j has a neighbor $k \in [p] \setminus \{j\}$ with orientation $k \to j$ in $\mathcal{G}_1^{\mathcal{I}}$ and $j \to k$ in $\mathcal{G}_2^{\mathcal{I}}$. Thus, k and ζ_I are d-separated in $\mathcal{G}_1^{\mathcal{I}}$ given $pa_{\mathcal{G}_1^{\mathcal{I}}}(k)$ but d-connected in $\mathcal{G}_2^{\mathcal{I}}$ given $pa_{\mathcal{G}_2^{\mathcal{I}}}(k)$. Hence by Lemma 3.10, there exists some $\{f^{(I)}\}_{I \in \mathcal{I}}$ that satisfies the $\mathcal{I}$-Markov property with respect to $\mathcal{G}_1^{\mathcal{I}}$ but not $\mathcal{G}_2^{\mathcal{I}}$ and thus $\mathcal{M}_{\mathcal{I}}(\mathcal{G}_1) \neq \mathcal{M}_{\mathcal{I}}(\mathcal{G}_2)$. □

Example 3.11. *The three DAGs in Figure 2 belong to the same MEC. Given interventions on $\mathcal{I} = \{\emptyset, \{1\}, \{2,3\}\}$, by Theorem 3.9, DAG (a) is not in the same $\mathcal{I}$-MEC as DAGs (b-c) due to its lack of v-structure $\zeta_{\{1\}} \to 1 \leftarrow 2$. The intervention improves the identifiability of these structures.*

It is straightforward to show that our graphical criterion of $\mathcal{I}$-MECs when $\emptyset \in \mathcal{I}$ is equivalent to the characterization of perfect-$\mathcal{I}$-MECs by Hauser & Bühlmann (2012) for perfect interventions, which proves their conjecture.

Corollary 3.12. *When $\emptyset \in \mathcal{I}$, two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are in the same $\mathcal{I}$-MEC iff they are in the same perfect-$\mathcal{I}$-MEC.*

3.3. Extension to $\emptyset \notin \mathcal{I}$

The identifiability results for perfect-$\mathcal{I}$-MECs by Hauser & Bühlmann (2012) hold for conservative $\mathcal{I}$, while our results for $\mathcal{I}$-MECs requires a stronger assumption, namely that $\emptyset \in \mathcal{I}$ (i.e. observational data is available). While this assumption is not restrictive in practice, it raises the question of whether our results can be extended to conservative sets of targets when $\emptyset \notin \mathcal{I}$. The following example shows that our current graphical characterization of $\mathcal{I}$-MECs (Theorem 3.9) does not generally hold under this weaker assumption.

Example 3.13. *Let $\mathcal{G}$ be the causal DAG $1 \to 2$ and let $\mathcal{I} = \{\{1\}, \{2\}\}$. The interventional distributions have the factorization $f^{(\{1\})}(X) = f^{(\{1\})}(X_1) f^{(\emptyset)}(X_2|X_1)$ and $f^{(\{2\})}(X) = f^{(\emptyset)}(X_1) f^{(\{2\})}(X_2|X_1)$ respectively, according to Definition 3.2. Any distributions with this factorization can also be written as $f^{(\{1\})}(X) = g^{(\emptyset)}(X_2) g^{(\{1\})}(X_1|X_2)$ and $f^{(\{2\})}(X) = g^{(\{2\})}(X_2) g^{(\emptyset)}(X_1|X_2)$ for an appropriate choice of $g^{(\emptyset)}$, $g^{(\{1\})}$ and $g^{(\{2\})}$. Thus, $\mathcal{G}_1$ and $\mathcal{G}_2$ belong to the same $\mathcal{I}$-MEC (i.e., $\mathcal{M}_{\mathcal{I}}(\mathcal{G}_1) = \mathcal{M}_{\mathcal{I}}(\mathcal{G}_2)$). But $\mathcal{G}_1^{\mathcal{I}}$ and $\mathcal{G}_2^{\mathcal{I}}$ do not have the same v-structures, contradicting the graphical criterion of Theorem 3.9.*

The following theorem extends our graphical characterization of $\mathcal{I}$-MECs to conservative sets of intervention targets when we don't necessarily have $\emptyset \in \mathcal{I}$. The proof of this result is provided in the Supplementary Material.

Theorem 3.14. *Let $\mathcal{I} \subset \mathfrak{P}([p])$ be a conservative set of intervention targets. Two causal DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ belong to the same $\mathcal{I}$-MEC if and only if for all $I \in \mathcal{I}$ the interventional DAGs $\mathcal{G}_1^{\tilde{\mathcal{I}}_I}$ and $\mathcal{G}_2^{\tilde{\mathcal{I}}_I}$ have the same skeletons and v-structures, where*

$$\tilde{\mathcal{I}}_I := \{\emptyset, \{I \cup J\}_{J \in \mathcal{I}, J \neq I}\}$$

The proof formalizes the following intuition: in the absence of an observational dataset, we can relabel one of the interventional datasets (i.e. from intervening on I) as the observational one; or equivalently, we "pretend" that our datasets are obtained under interventions on $\tilde{\mathcal{I}}_I$ instead of $\mathcal{I}$. Then two DAGs cannot be distinguished under interventions on $\mathcal{I}$ if and only if this also holds for $\tilde{\mathcal{I}}_I$, for all $I \in \mathcal{I}$. Note that if $\emptyset \in \mathcal{I}$, then this statement is equivalent to Theorem 3.9. Hence the assumption $\emptyset \in \mathcal{I}$ in Section 3.2 can be made without loss of generality and our identifiability results extend to all conservative sets of intervention targets.

Example 3.15. *The three DAGs in Figure 3 belong to the same MEC. Given interventions on $\mathcal{I} = \{\{2\}, \{3\}\}$, by Theorem 3.14, DAG (a) is not in the same $\mathcal{I}$-MEC as DAGs (b-c) due to its v-structure $\zeta_{\{2,3\}} \to 2 \leftarrow 1$. The intervention improves the identifiability of these structures.*

4. Consistent algorithm for learning $\mathcal{I}$-MECs

Having shown that the $\mathcal{I}$-MEC of a causal DAG can be identified from general interventions, we now propose a permutation-based algorithm for learning the $\mathcal{I}$-MEC. The

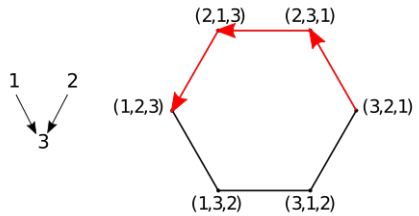


Figure 4. Left: DAG corresponding to permutations (1,2,3) or (2,1,3). Right: Illustration of greedy search over the space of permutations for $p = 3$, starting at (3,2,1). The space of permutations is represented by a polytope known as the *permutahedron* in which each node corresponds to a permutation and edges connect neighboring transpositions. A greedy search corresponds to a greedy edge walk (red arrows) over the permutahedron.

algorithm takes interventional datasets obtained under *general interventions* with *known targets* $\mathcal{I}$ and returns a DAG in the correct $\mathcal{I}$-MEC.

4.1. Preliminaries

Permutation-based causal inference algorithms search for a permutation π^* that is consistent with the topological order of the true causal DAG $\mathcal{G}^*$, i.e. if (i, j) is an edge in $\mathcal{G}^*$ then $i < j$ in π^* (Figure 4, left). Given π^*, $\mathcal{G}^*$ can then be determined by learning an undirected graph over the nodes and orienting the edges according to the order π^*.

To find π^*, one option is to do a greedy search over the space of permutations by tranposing neighboring nodes and optimizing a score function (Figure 4, right). In Solus et al. (2017), the authors propose an algorithm called *Greedy Sparsest Permutations (GSP)* that uses a score function based on CI relations. Specifically, the score of a given permutation π is the number of edges in its *minimal I-map* $\mathcal{G}_\pi = ([p], E_\pi)$, which is the sparsest DAG consistent with π such that $f^{(\emptyset)}$ is Markov with respect to $\mathcal{G}_\pi$. Since the score is only guaranteed to be weakly decreasing on any path from π to π^*, the algorithm iteratively uses a depth-first-search. Additionally, instead of considering all neighboring transpositions of π in the search, GSP only transposes neighboring nodes in the permutation that are connected by *covered* edges[5] in $\mathcal{G}_\pi$, which improves the efficiency of the algorithm. Under the assumptions of causal sufficiency and faithfulness[6], GSP is *consistent* in that it returns a permutation τ where $\mathcal{G}_\tau$ is in the same MEC as the true DAG $\mathcal{G}^*$ (Solus et al., 2017; Mohammadi et al., 2018). However, GSP does not use data from interventions, so it is not guaranteed to return a DAG in the correct $\mathcal{I}$-MEC.

Perfect-IGSP extends GSP to incorporate data from interventions (Wang et al., 2017). However, the consistency result of perfect-IGSP requires the interventional data to come from perfect interventions. This motivates our development of a new algorithm, *IGSP (or general-IGSP)*, which is provably consistent for finding the $\mathcal{I}$-MEC of $\mathcal{G}^*$ when the data come from general interventions.

[5] An edge (i, j) in a DAG $\mathcal{G}$ is *covered* if $\mathrm{pa}_\mathcal{G}(i) = \mathrm{pa}_\mathcal{G}(j)\backslash\{i\}$.

[6] *Causal sufficiency* is the assumption that there are no hidden latent confounders, and *faithfulness* implies that all CI relations of the observational distribution $f^\emptyset$ are implied by d-separation in $\mathcal{G}$.

4.2. Main Results

In Algorithm 1, we present *IGSP*, a greedy permutation-based algorithm for recovering the $\mathcal{I}$-MEC of $\mathcal{G}^*$ from $\{f^{(I)}\}_{I\in\mathcal{I}}$ for general interventions with known targets $\mathcal{I}$.

Similar to GSP, IGSP starts with a permutation π and implements depth-first-search to look for a permutation τ such that $|\mathcal{G}_\tau| < |\mathcal{G}_\pi|$, where $\mathcal{G}_\tau$ and $\mathcal{G}_\pi$ are the minimal I-maps of τ and π respectively; and iterates until no such permutation can be found. One difference from GSP is that in each step of the search, IGSP only transposes neighboring nodes that are connected by $\mathcal{I}$-*covered* edges in the corresponding minimal I-map.

Definition 4.1. A covered edge $i \to j$ in a DAG $\mathcal{G}$ is $\mathcal{I}$-*covered* if it satisfies the following two conditions:

(1) if $\{i\} \in \mathcal{I}$, then $f^{(\{i\})}(X_j) = f^{(\emptyset)}(X_j)$;

(2) $f^{(I)}(X_i) \neq f^{(\emptyset)}(X_i)$ for any $I \in \mathcal{I}_{j\backslash i}$, where $\mathcal{I}_{j\backslash i} := \{I \in \mathcal{I} \mid j \in I, i \notin I\}$.

The use of $\mathcal{I}$-covered edges restricts the search space and ensures that we do not consider permutations that contradict order relations derived from the intervention experiments. Furthermore, the tranposition of neighboring nodes connected by $\mathcal{I}$-covered edges that are also $\mathcal{I}$-*contradictory* edges is prioritized during the search.

Definition 4.2. Let $\mathrm{ne}_\mathcal{G}(i)$ denote the neighbors of node i in a DAG $\mathcal{G}$. An edge $i \to j$ in $\mathcal{G}$ is $\mathcal{I}$-*contradictory* if at least one of the following two conditions hold:

(1) There exists a set $S \subset \mathrm{ne}_\mathcal{G}(j)\backslash\{i\}$ such that $f^{(\emptyset)}(X_j|X_S) = f^{(I)}(X_j|X_S)$ for all $I \in \mathcal{I}_{i\backslash j}$;

(2) $f^{(\emptyset)}(X_i|X_S) \neq f^{(I)}(X_i|X_S)$ for some $I \in \mathcal{I}_{j\backslash i}$, for all $S \subset \mathrm{ne}_\mathcal{G}(i)\backslash\{j\}$.

$\mathcal{I}$-contradictory edges are prioritized because they violate the $\mathcal{I}$-Markov property (Definition 3.6). Thus, a DAG in the correct $\mathcal{I}$-MEC should minimize the number of $\mathcal{I}$-contradictory edges. Evaluating whether edges are $\mathcal{I}$-contradictory requires invariance tests that grow with the maximum degree of $\mathcal{G}_\pi$. When $\mathcal{I}$ consists of only single-node interventions, a modified definition of $\mathcal{I}$-contradictory edges can be used to reduce the number of tests.

Definition 4.3. Let $\mathcal{I}$ be a set of intervention targets such that $\{i\} \in \mathcal{I}$ or $\{j\} \in \mathcal{I}$. The edge $i \to j$ is $\mathcal{I}$-*contradictory* if either of the following is true:

Algorithm 1 IGSP for general interventions

Input: A collection of intervention targets $\mathcal{I}$ with $\emptyset \in \mathcal{I}$, samples from distributions $\{f^{(I)}\}_{I\in\mathcal{I}}$, and a starting permutation π_0.
Output: A permutation τ and associated I-map $\mathcal{G}_\tau$
Set $\pi = \pi_0$, $\mathcal{G}_\pi :=$ minimal I-map of π.
repeat
Using a depth-first-search with root π, search for a permutation τ with minimal I-map $\mathcal{G}_\tau$ such that $|\mathcal{G}_\pi| > |\mathcal{G}_\tau|$ that is connected to $\mathcal{G}_\pi$ by a sequence of $\mathcal{I}$-covered edge reversals, with priority given to $\mathcal{I}$-contradictory edge reversals. If τ exists, set $\pi = \tau$, $\mathcal{G}_\pi = \mathcal{G}_\tau$.
until No such τ can be found.
Return the permutation τ and the associated I-map $\mathcal{G}_\tau$ with $|\mathcal{G}_\tau| = |\mathcal{G}_\pi|$ that minimizes the number of $\mathcal{I}$-contradicting edges.

(1) $\{i\} \in \mathcal{I}$ and $f^{\{i\}}(X_j) = f^{\emptyset}(X_j)$; or

(2) $\{j\} \in \mathcal{I}$ and $f^{\{j\}}(X_i) \neq f^{\emptyset}(X_i)$.

In the special case where we only have single-node interventions, the number of invariance tests no longer depends on the maximum degree of $\mathcal{G}_\pi$ under this simplification.

Unlike perfect-IGSP, which is consistent only under perfect interventions, our method is consistent for general interventions under the following two assumptions:

Assumption 4.4. *Let $I \in \mathcal{I}$ with $i \in I$. Then $f^{(I)}(X_j) \neq f^{(\emptyset)}(X_j)$ for all descendents j of i.*

Assumption 4.5. *Let $I \in \mathcal{I}$ with $i \in I$. Then $f^{(I)}(X_j|X_S) \neq f^{(\emptyset)}(X_j|X_S)$ for any child j of i such that $j \notin I$ and for all $S \subset \mathrm{ne}_{\mathcal{G}^*}(j) \setminus \{i\}$, where $\mathrm{ne}_{\mathcal{G}^*}(j)$ denotes the neighbors of node j in $\mathcal{G}^*$.*

Both assumptions are strictly weaker than the faithfulness assumption on the $\mathcal{I}$-DAG. Assumption 4.4 extends the assumption by Tian & Pearl (2001) to interventions on multiple nodes. It essentially requires interventions on upstream nodes to affect downstream nodes. Assumption 4.5 is similarly intuitive and requires the distribution of X_j to change under an intervention on its parent X_i as long as X_i is not part of the conditioning set.

The main result of this section is the following theorem, which states the consistency of IGSP.

Theorem 4.6. *Algorithm 1 is consistent under assumptions 4.4 and 4.5, faithfulness of $f^{(\emptyset)}$ with respect to $\mathcal{G}$, and causal sufficiency. When $\mathcal{I}$ only contains single-variable interventions, assumption 4.5 is not required for the correctness of the algorithm.*

4.3. Implementation of Algorithm 1

Testing for invariance: To test whether a (conditional) distribution $f^{(I)}(X_i|X_S)$ is invariant, we used a method proposed by Heinze-Deml et al. (2017) that we found to work well in practice. Briefly, we test whether X_i is independent of the *index* of the interventional dataset given X_S, using the kernel-based CI test of Zhang et al. (2012).

Data pooling for CI testing: Let $\mathrm{an}_{\mathcal{G}_\pi}(i)$ denote the ancestors of node i in $\mathcal{G}_\pi$. After reversing an $\mathcal{I}$-covered edge (i, j), updating $\mathcal{G}_\pi$ requires testing if $X_i \perp\!\!\!\perp X_k \mid X_{\mathrm{an}_{\mathcal{G}_\pi}(i)\setminus\{k\}}$ for $k \in \mathrm{pa}_{\mathcal{G}_\pi}(i)$ under the observational distribution $f^{(\emptyset)}$. By combining the interventional data with the observational data in a provably correct manner, we can increase the power of the CI tests, which is useful when the sample sizes are limited. In the Supplementary Material, we present a proposition giving sufficient conditions under which CI relations hold when the data come from a mixture of interventional distributions, and use this to derive a set of checkable conditions on $\mathcal{G}_\pi$ for determining which datasets can be combined to test $X_i \perp\!\!\!\perp X_k \mid X_{\mathrm{an}_{\mathcal{G}_\pi}(i)\setminus\{k\}}$ for $k \in \mathrm{pa}_{\mathcal{G}_\pi}(i)$.

5. Empirical Results

5.1. Experiments on simulated datasets

IGSP vs perfect-IGSP: We compared Algorithm 1 to perfect-IGSP on the task of recovering the correct $\mathcal{I}$-MEC under three types of interventions: perfect, *inhibiting*, and *imperfect*. By an *inhibiting* intervention, we mean an intervention that reduces the effect of the parents of the target node. This simulates a biological intervention such as a small-molecule inhibitor with a modest effect. By an *imperfect* intervention, we mean an intervention that is perfect with probability α and ineffective with probability $1 - \alpha$ for some $\alpha \in (0, 1)$. This simulates biological experiments such as gene deletions that might not work in all cells.

For each simulation, we sampled 100 DAGs from an Erdös-Renyi random graph model with an average neighborhood size of 1.5 and $p \in \{10, 20\}$ nodes. The data for each causal DAG $\mathcal{G}$ was generated using a linear structural equation model with independent Gaussian noise: $X = AX + \epsilon$, where A is an upper-triangular matrix with edge weights $A_{ij} \neq 0$ if and only if $i \to j$, and $\epsilon \sim \mathcal{N}(0, \mathrm{I}_p)$. For $A_{ij} \neq 0$, the edge weights were sampled uniformly from $[-1, -0.25] \cup [0.25, 1]$. We simulated perfect interventions on i by setting the column $A_{\cdot i} = 0$; inhibiting interventions by decreasing $A_{\cdot i}$ by a factor of 10; and imperfect interventions with a success rate of $\alpha = 0.5$. Interventions were performed on all single-variable targets or all pairs of multiple-variable targets to maximally illuminate the difference between IGSP and perfect-IGSP.

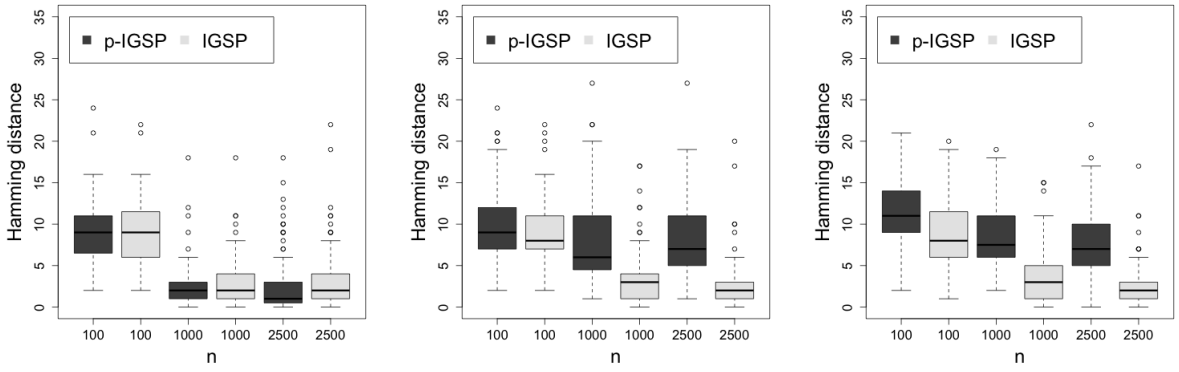


Figure 5. Distributions of Hamming distances of recovered DAGs using IGSP and perfect-IGSP (p-IGSP) for 20-node graphs with single-variable (a) perfect, (b) imperfect, and (c) inhibitory interventions

Figure 5 shows that IGSP outperforms perfect-IGSP on data from inhibiting and imperfect interventions and that the algorithms perform comparably on data from perfect interventions (see also the Supplementary Material for further figures). These empirical comparisons corroborate our theoretical results that IGSP is consistent for general types of interventions, while perfect-IGSP is only consistent for perfect interventions. Consistency for general interventions is particularly important for applications to genomics, where it is usually not known a priori whether an intervention will be perfect; these results suggest we can use IGSP regardless of the type of intervention.

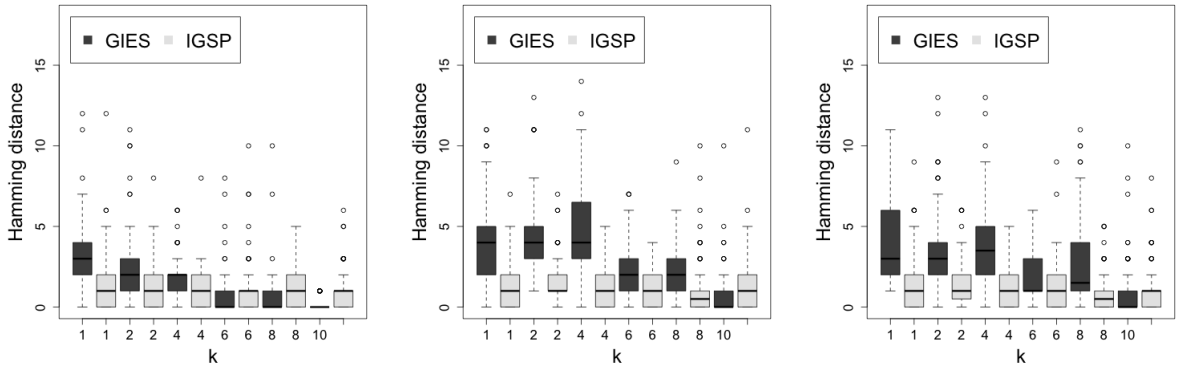
IGSP vs GIES: GIES is an extension of the score-based causal inference algorithm, *Greedy Equivalence Search (GES)*, to the interventional setting. Its score function incorporates the log-likelihood of the data based on the interventional distribution of Equation (1), making it appropriate for learning DAGs under general interventions. Although GIES is not consistent in general (Wang et al., 2017), it has performed well in previous empirical studies (Hauser & Bühlmann, 2012; 2015). Additionally, both IGSP and GIES assume causal sufficiency and output DAGs, while the other methods mentioned in Section 2 do not output a DAG or use different assumptions. We therefore used GIES as a baseline for comparison.

We evaluated IGSP and GIES on learning DAGs from different types of interventions, varying the number of interventional datasets ($|\mathcal{I}| = k \in \{1, 2, 4, 6, 8, 10\}$). The synthetic data was otherwise generated as described above. Figure 6 shows that IGSP in general significantly outperforms GIES. However, GIES performs better when the number of interventional datasets is large, i.e. for $|\mathcal{I}| = 10$. This performance increase can be credited to the GIES score function which efficiently pools the interventional datasets.

5.2. Experiments on Biological Datasets

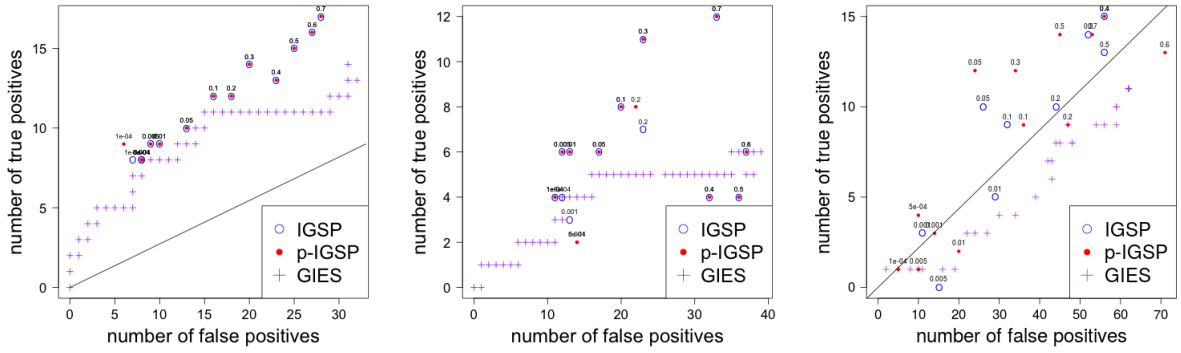
Protein Expression Dataset: We evaluated our algorithm on the task of learning a protein network from a protein mass spectroscopy dataset (Sachs et al., 2005). The processed dataset consists of 5846 measurements of phosphoprotein and phospholipid levels from primary human immune system cells. Interventions on the network were perfect interventions corresponding to chemical reagents that strongly inhibit or activate certain signaling proteins. Figures 7(a) and 7(b) illustrate the ROC curves of IGSP, perfect-IGSP (Wang et al., 2017) and GIES (Hauser & Bühlmann, 2015) on learning the skeleton and DAG of the ground-truth network respectively. We found that IGSP and perfect-IGSP performed comparably well on this dataset, which is consistent with our theoretical results. As expected, both IGSP and perfect-IGSP outperform GIES at recovering the true DAG, since the former two algorithms have consistency guarantees in this regime while GIES does not.

Gene Expression Dataset: We also evaluated IGSP on a single-cell gene expression dataset (Dixit et al., 2016). The processed dataset contains 992 observational and 13,435 interventional measurements of gene expression from bone marrow-derived dendritic cells. There are eight interventions in total, each corresponding to a targeted gene deletion using the CRISPR/Cas9 system. Since this dataset introduced the perturb-seq technique and was meant as a demonstration, we expected the interventions to be of high-quality and close to perfect. We applied IGSP, perfect-IGSP, and GIES to learn causal DAGs over 24 transcription factors that modulate each other and play a critical role in regulating downstream genes. Since the ground-truth DAG is not available, we evaluated each learned DAG on its accuracy in predicting the effect of an intervention that was left out during inference, as described by Wang et al. (2017). Figure 7(c) shows that IGSP is competitive with perfect-IGSP,



(a) (b) (c)

Figure 6. Distributions of Hamming distances of recovered DAGs using IGSP and GIES for 10-node graphs with average edge density of 1.5 and single-node (a) perfect, (b) imperfect, and (c) inhibitory interventions on k nodes



(a) (b) (c)

Figure 7. ROC plots evaluating IGSP, perfect-IGSP (p-IGSP) and GIES on learning the (a) skeleton and (b) DAG of the protein network from Sachs et al. (2005) and on (c) predicting the causal effects of interventions on a gene network from (Dixit et al., 2016)

which suggests that the gene deletion interventions were close to perfect. Once again, both IGSP and perfect-IGSP outperform GIES on this dataset.

6. Discussion

In this paper, we studied $\mathcal{I}$-MECs, the equivalence classes of causal DAGs that can be identified from a set of general (not necessarily perfect) intervention experiments. In particular, we provided a graphical characterization of $\mathcal{I}$-MECs and proved a conjecture of Hauser & Bühlmann (2012) showing that $\mathcal{I}$-MECs are equivalent to perfect-$\mathcal{I}$-MECs under basic assumptions. This result has important practical consequences, since it implies that general interventions provide similar causal information as perfect interventions despite being less invasive. An interesting problem for future research is to extend these identifiability results to the setting where the intervention targets are unknown. Such results would have wide-ranging implications, such as in genomics, where the interventions can have off-target effects.

We also propose the first provably consistent algorithm, IGSP, for learning the $\mathcal{I}$-MEC from observational and general interventional data and apply it to protein and gene perturbation experiments. IGSP extends perfect-IGSP (Wang et al., 2017), which is only consistent for perfect interventions. In agreement with the theory, IGSP outperforms perfect-IGSP on data from non-perfect interventions and is competitive with perfect-IGSP on data from perfect interventions, thereby demonstrating the flexibility of IGSP to learn from different types of interventions. A challenge for future research is to scale algorithms like IGSP up to thousands of nodes, which would allow learning the entire gene network of a cell. The main bottleneck for scaling IGSP and an important area for future research is the development of accurate and fast conditional independence tests that can be applied under general distributional assumptions.

Acknowledgements

Karren D. Yang was partially supported by an NSF Graduate Fellowship. Caroline Uhler was partially supported by NSF (DMS-1651995), ONR (N00014-17-1-2147), and a Sloan Fellowship.

References

Andersson, S. A., Madigan, D., and Perlman, M. D. A characterization of markov equivalence classes for acyclic digraphs. *Annals of Statistics*, 25(2):505–541, 1997.

Bareinboim, E., Brito, C., and Pearl, J. Local characterizations of causal bayesian networks. In *Graph Structures for Knowledge Representation and Reasoning*, pp. 1–17, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

Brown, L. E., Tsamardinos, I., and Aliferis, C. F. A comparison of novel and state-of-the-art polynomial bayesian network learning algorithms. In *AAAI*, volume 2005, pp. 739–745, 2005.

Chickering, D. M. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3 (Nov):507–554, 2002.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., et al. Perturb-seq: dissecting molecular circuits with scalable single-cell rna profiling of pooled genetic screens. *Cell*, 167(7):1853–1866, 2016.

Dominguez, A. A., Lim, W. A., and Qi, L. S. Beyond editing: repurposing crispr–cas9 for precision genome regulation and interrogation. *Nature Reviews: Molecular Cell Biology*, 17(1):5, 2016.

Eaton, D. and Murphy, K. Exact bayesian structure learning from uncertain interventions. In *Artificial Intelligence and Statistics*, pp. 107–114, 2007.

Eberhardt, F. Almost optimal intervention sets for causal discovery. In *Uncertainty in Artificial Intelligence*, pp. 161–168, 2008.

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

Eberhardt, F., Glymour, C., and Scheines, R. On the number of experiments sufficient and in the worst case necessary to identify all causal relations among n variables. In *Uncertainty in Artificial Intelligence*, pp. 178–184, 2005.

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. Using bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3-4):601–620, 2000.

Hauser, A. and Bühlmann, P. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. *Journal of Machine Learning Research*, 13(Aug):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.

Heinze-Deml, C., Peters, J., and Meinshausen, N. Invariant causal prediction for nonlinear models. *arXiv preprint arXiv:1706.08576*, 2017.

Hoyer, P. O., Dominik, J., Mooij, J. M., Peters, J., and Schölkopf, B. Nonlinear causal discovery with additive noise models. In *Advances in Neural Information Processing Systems 21*, pp. 689–696. Curran Associates, Inc., 2009.

Hyttinen, A., Eberhardt, F., and Järvisalo, M. Constraint-based causal discovery: Conflict resolution with answer set programming. In *Uncertainty in Artificial Intelligence*, pp. 340–349, 2014.

Lauritzen, S. L. *Graphical Models*, volume 17. Clarendon Press, 1996.

Magliacane, S., Claassen, T., and Mooij, J. M. Joint causal inference on observational and experimental datasets. *arXiv preprint arXiv:1611.10351*, 2016.

Mohammadi, F., Uhler, C., Wang, C., and Yu, J. Generalized permutohedra from probabilistic graphical models. *SIAM Journal on Discrete Mathematics*, 32(1):64–93, 2018.

Murphy, K. The bayes net toolbox for matlab. *Computing Science and Statistics*, 33(2):1024–1034, 2001.

Pearl, J. Causality: Models, reasoning, and inference. *Econometric Theory*, 19(675-685):46, 2003.

Peters, J., Mooij, J., Janzing, D., and Schölkopf, B. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15(1):2009–2053, 2014.

Peters, J., Bühlmann, P., and Meinshausen, N. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):947–1012, 2016.

Robins, J. M., Hernan, M. A., and Brumback, B. Marginal structural models and causal inference in epidemiology, 2000.

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A., and Nolan, G. P. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721): 523–529, 2005.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(Oct): 2003–2030, 2006.

Solus, L., Wang, Y., Matejovicova, L., and Uhler, C. Consistency guarantees for permutation-based causal inference algorithms. *ArXiv preprint. arXiv:1702.03530*, 2017.

Spirtes, P., Glymour, C., and Scheines, R. *Causation, Prediction, and Search*. MIT press, 2000.

Tian, J. and Pearl, J. Causal discovery from changes. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 512–521. Morgan Kaufmann Publishers Inc., 2001.

Triantafillou, S. and Tsamardinos, I. Constraint-based causal discovery from multiple interventions over overlapping variable sets. *Journal of Machine Learning Research*, 16:2147–2205, 2015.

Verma, T. S. and Pearl, J. Equivalence and synthesis of causal models. In *Uncertainty in Artificial Intelligence*, volume 6, pp. 255, 1991.

Wang, Y., Solus, L., Yang, K., and Uhler, C. Permutation-based causal inference algorithms with interventions. In *Advances in Neural Information Processing Systems*, pp. 5824–5833. Curran Associates, Inc., 2017.

Zhang, K., Peters, J., Janzing, D., and Schölkopf, B. Kernel-based conditional independence test and application in causal discovery. *arXiv preprint arXiv:1202.3775*, 2012.