LOCAL CONSTRAINT-BASED CAUSAL DISCOVERY IN THE PRESENCE OF CYCLES AND LATENT VARIABLES

Anonymous authors
Paper under double-blind review

ABSTRACT

Local causal discovery is of practical importance because many practical needs call only for identifying the causal structure surrounding a single target variable, making recovery of the entire global structure unnecessary. However, existing local methods typically assume acyclicity or the absence of latent variables. We study local causal discovery when cycles and latent variables are both present. We first characterize a local region sufficient for target-specific structure learning and establish which locally learned causal information is consistent with that obtained from global discovery. Building on these results, we propose a local constraint-based algorithm and prove that, under standard assumptions, it is sound and complete, recovering the same identifiable structure around the target as global causal discovery. Extensive experiments on random and real-world structures demonstrate the effectiveness and efficiency of our method.

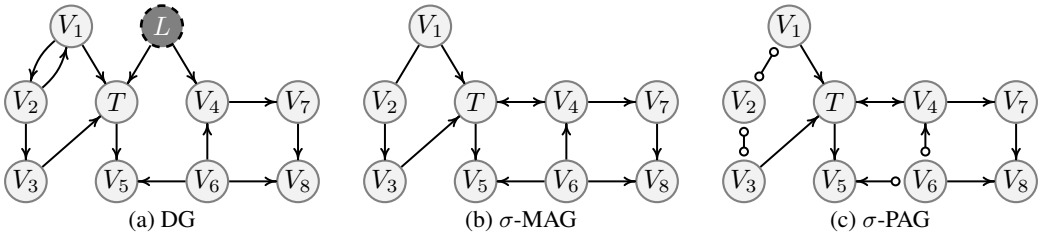
1 INTRODUCTION

Causal discovery aims to identify causal relations among variables from observational data and has attracted increasing attention across a wide range of fields, including computer science (Peters et al., 2017; Pearl, 2018; Schölkopf, 2022), social science (Spirtes et al., 2000), epidemiology (Hernán & Robins, 2020), biology (Glymour et al., 2019), and neuroscience (Smith et al., 2011; Sanchez-Romero et al., 2019). In many scientific problems, however, the goal is not to recover every causal relation in a system, but rather to understand the causal structure surrounding a particular variable of interest (Gao & Ji, 2015; Geng et al., 2019; Dai et al., 2024; Li et al., 2026). Examples include identifying factors directly related to a disease outcome, uncovering regulators of a particular gene, or determining the variables that causally influence a specific response. Local causal discovery addresses such target-specific questions by learning the causal structure around a designated target variable T without reconstructing the entire causal graph.

Motivated by such target-specific tasks, a substantial body of work has been developed for local causal discovery. Early approaches such as PCD-by-PCD (Yin et al., 2008; Zhou et al., 2010) and MB-by-MB (Wang et al., 2014) progressively expand local parents-and-children sets or Markov blankets to determine the causal neighborhood of a target. CMB (Gao & Ji, 2015), ELCS (Yang et al., 2021), PSL (Ling et al., 2022), and GraN-LCS (Liang et al., 2024) further improve the efficiency or orientation capability of local structure learning. These methods, however, are developed under causal sufficiency and therefore do not account for latent common causes. More recently, this restriction has been relaxed by methods that study local causal structure learning in the presence of latent variables (Xie et al., 2024; Li et al., 2025; Ling et al., 2025). Despite these advances, these methods still assume that the underlying causal system is acyclic.

Acyclicity, however, can also be restrictive in practice. Feedback mechanisms naturally arise in biological (Benito et al., 2007), economic (Haavelmo, 1943), chemical, physical, and control systems (Mason, 1953; Mooij & Claassen, 2020). Cyclic structural causal models (SCM) provide a natural framework for representing such feedback systems (Forré & Mooij, 2017; Bongers et al., 2021). Figure 1 illustrates a setting in which both feedback and latent confounding are present. The variables V_1 and V_2 form a feedback loop $V_1 \to V_2 \to V_1$, while an unobserved variable L affects both T and V_4. Such systems violate both acyclicity and causal sufficiency and therefore fall outside the scope of conventional local causal discovery methods. For constraint-based causal discovery in cyclic systems, σ-separation provides an appropriate graphical criterion for character-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) DG (b) σ-MAG (c) σ-PAG

Figure 1: A directed graph with feedback and latent confounding, its corresponding σ-MAG, and the σ-PAG representing its Markov equivalence class. In (a), L is latent, and $V_1 \to V_2 \to V_1$ forms a feedback loop.

izing conditional independencies (CIs) (Forré & Mooij, 2017; Bongers et al., 2021). Building on this framework, Mooij & Claassen (2020) showed that FCI remains sound and complete for simple, σ-Markov, and σ-faithful structural causal models with cycles and latent confounding. We refer to FCI in this setting as σ-FCI. Although this result provides a principled solution for global constraint-based causal discovery, recovering the causal structure around a target variable still requires learning over the entire set of observed variables. A local constraint-based method that avoids global structure learning while simultaneously allowing cycles and latent variables remains lacking.

To address these challenges, we study local constraint-based causal discovery in the simultaneous presence of directed cycles and latent variables. Our main contributions are summarized as follows:

- We provide graphical and identifiable characterizations of the Markov blanket in the presence of directed cycles and latent variables. Building on these characterizations, we establish a theoretical connection between the structure learned from the marginal distribution over the Markov-blanket region and the corresponding global causal structure.
- Building on these theoretical results, we develop a local constraint-based causal discovery algorithm for systems with both directed cycles and latent variables. Under standard assumptions, we prove that the proposed method is sound and complete and recovers the same identifiable local causal structure around the target as global σ-FCI.
- We conduct extensive experiments on random and real-world structures. The results show that the proposed method achieves competitive or improved structural accuracy while substantially reducing computational cost and the number of CI tests compared with global causal discovery methods.

2 Related Work

We briefly review related work on global and local causal discovery, with a focus on constraint-based methods. For comprehensive surveys of causal structure learning, see (Spirtes & Zhang, 2016; Heinze-Deml et al., 2018; Glymour et al., 2019; Yu et al., 2020; Kitson et al., 2023; Zanga et al., 2022).

Global causal discovery. Under causal sufficiency and acyclicity, the IC algorithm (Verma & Pearl, 1990) and the PC algorithm (Spirtes & Glymour, 1991) use conditional independence (CI) relations to recover the Markov equivalence class of the underlying DAG. When latent confounding is allowed, FCI (Spirtes et al., 1995; Zhang, 2008) recovers a PAG representing the Markov equivalence class of the underlying maximal ancestral graph (MAG) and is sound and complete under standard assumptions. RFCI (Colombo et al., 2012) and FCI$^+$ (Claassen et al., 2013) improve the efficiency of constraint-based discovery in sparse or high-dimensional settings, while ICD (Rohekar et al., 2021) and L-MARVEL (Akbari et al., 2021) provide alternative strategies for learning the global causal structure. Despite their different computational and theoretical properties, these methods all aim to recover causal information over the full set of observed variables. Directed cycles require a different treatment because standard d-separation is no longer sufficient to characterize conditional independence relations. Using σ-separation, Mooij & Claassen (2020) showed that FCI remains sound and complete for simple structural causal models in the presence of cycles and latent confounding. More recently, Yao & Mooij (2025) introduced σ-MAGs to represent causal structures with directed cycles and latent variables. These developments provide principled solutions for global causal discovery in cyclic systems, but still require learning over the entire observed variable set.

Local causal discovery. Local causal discovery aims to recover causal relations around a target variable without reconstructing the entire causal graph. LCD (Cooper, 1997) and its variants (Silverstein et al., 2000; Mani & Cooper, 2004; Mani et al., 2006; Versteeg et al., 2022) infer causal relations from local graphical patterns such as Y-structures, but do not generally recover the complete causal structure surrounding a target. Other methods directly learn the neighborhood of a target by progressively identifying and orienting its parents, children, or Markov blanket. Representative methods include PCD-by-PCD (Yin et al., 2008; Zhou et al., 2010), MB-by-MB (Wang et al., 2014), CMB (Gao & Ji, 2015), ELCS (Yang et al., 2021), PSL (Ling et al., 2022), and GraN-LCS (Liang et al., 2024). These methods are mainly developed for acyclic and causally sufficient systems. Recent studies have relaxed these assumptions in different directions. The methods of Xie et al. (2024); Li et al. (2025), together with LatentLCD (Ling et al., 2025), consider local causal discovery in the presence of latent variables, while Li et al. (2026) further considers selection bias. However, these methods still assume that the underlying causal structure is acyclic. In contrast, Dai et al. (2024) studies local causal discovery with cycles, but is restricted to a linear non-Gaussian model and assumes the absence of latent confounding. Therefore, a general local constraint-based method that simultaneously accommodates directed cycles and latent variables remains lacking.

3 BACKGROUND

Throughout the paper, individual variables (or vertices) are denoted by uppercase letters, e.g., X, variable sets by bold uppercase letters, e.g., $\mathbf{X}$, and graphs by calligraphic letters. We write $\mathrm{An}(X, \mathcal{G})$ and $\mathrm{De}(X, \mathcal{G})$ for the ancestors and descendants of X in $\mathcal{G}$, extending these operators to sets by union, and define $\mathrm{An}^+(\mathbf{A}, \mathcal{G}) := \mathbf{A} \cup \mathrm{An}(\mathbf{A}, \mathcal{G})$. The main definitions are given below; additional terminology and a notation summary appear in Appendix A and Table 2.

Definition 1 (Directed graph)**.** *A* directed graph *(DG) is a graph $\mathcal{D} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is the vertex set and $\mathbf{E}$ is the set of directed edges. Directed cycles are allowed. The vertex set is partitioned into observed variables $\mathbf{O}$ and latent variables $\mathbf{L}$, such that $\mathbf{V} = \mathbf{O} \dot\cup \mathbf{L}$.*

For $X \in \mathbf{V}$, the *strongly connected component* (SCC) of X, denoted by $\mathrm{SCC}(X, \mathcal{D})$, is the maximal set containing X such that every pair of vertices in the set is mutually reachable by directed paths.

Definition 2 (σ-separation (Forré & Mooij, 2017; Mooij & Claassen, 2020))**.** *Let $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ be a DG, let $X, Y \in \mathbf{V}$ be distinct, and let $\mathbf{Z} \subseteq \mathbf{V} \setminus \{X, Y\}$. A path π between X and Y is* σ-blocked *by $\mathbf{Z}$ if at least one of the following holds: (i) π contains a collider $V_i \notin \mathrm{An}^+(\mathbf{Z}, \mathcal{D})$; or (ii) π contains a non-collider $V_i \in \mathbf{Z}$ with an outgoing edge along π to a neighboring vertex $V_j \notin \mathrm{SCC}(V_i, \mathcal{D})$.*

The vertices X and Y are *σ-separated* by $\mathbf{Z}$ in $\mathcal{D}$, denoted by $X \perp\!\!\!\perp_\sigma Y \mid \mathbf{Z}$, if every path between X and Y is σ-blocked by $\mathbf{Z}$. When $\mathcal{D}$ is acyclic, every SCC is a singleton. Hence, every conditioned non-collider satisfies condition (ii), and σ-separation reduces to standard d-separation (Pearl, 2000).

Definition 3 (σ-MAG (Yao & Mooij, 2025))**.** *A* σ-maximal ancestral graph *(σ-MAG) $\mathcal{H}$ over $\mathbf{O}$ is a mixed graph with edge types $\{\rightarrow, \leftarrow, \leftrightarrow, -\}$ satisfying:*

(i) between any two distinct vertices there is at most one edge, and no vertex is adjacent to itself;
(ii) $\mathcal{H}$ is ancestral: if it contains an anterior path from A to B, then it does not contain an edge $A \leftarrow\!* B$;
(iii) $\mathcal{H}$ is σ-maximal:
  (a) it is maximal: there is no inducing path between any two distinct non-adjacent vertices;
  (b) it is σ-complete: if $\mathcal{H}$ contains $A* \rightarrow B - C$, then A and C are adjacent. Moreover, if $B - D$ is also present, then C and D are adjacent.

Given an underlying DG $\mathcal{D}$ over $\mathbf{V} = \mathbf{O} \dot\cup \mathbf{L}$, the corresponding σ-MAG $\mathcal{H}$ over $\mathbf{O}$ represents the σ-separation and ancestral relations among the observed variables induced by $\mathcal{D}$. The corresponding σ-MAG can be constructed from $\mathcal{D}$ as follows:

(i) for any distinct $X, Y \in \mathbf{O}$, X and Y are adjacent in $\mathcal{H}$ if and only if there is no $\mathbf{Z} \subseteq \mathbf{O} \setminus \{X, Y\}$ such that $X \perp\!\!\!\perp_\sigma Y \mid \mathbf{Z}$ in $\mathcal{D}$;

(ii) for every adjacent pair X, Y in $\mathcal{H}$, the edge is oriented according to their ancestral relations in $\mathcal{D}$:

$$\begin{cases} X \to Y, & X \in \mathrm{An}(Y, \mathcal{D}), \quad Y \notin \mathrm{An}(X, \mathcal{D}), \\ X \leftarrow Y, & X \notin \mathrm{An}(Y, \mathcal{D}), \quad Y \in \mathrm{An}(X, \mathcal{D}), \\ X \leftrightarrow Y, & X \notin \mathrm{An}(Y, \mathcal{D}), \quad Y \notin \mathrm{An}(X, \mathcal{D}), \\ X - Y, & X \in \mathrm{An}(Y, \mathcal{D}), \quad Y \in \mathrm{An}(X, \mathcal{D}). \end{cases}$$

We provide the following characterization, which avoids searching over all possible conditioning sets when determining adjacency.

Lemma 1. *Let $\mathcal{D}$ be a DG over $\mathbf{V} = \mathbf{O} \dot{\cup} \mathbf{L}$, and let $\mathcal{H}$ be the corresponding σ-MAG over $\mathbf{O}$. For any distinct $X, Y \in \mathbf{O}$, the following are equivalent:*

(i) X and Y are adjacent in $\mathcal{H}$;
(ii) X and Y are not σ-separated in $\mathcal{D}$ by $\mathrm{An}^+(\{X, Y\}, \mathcal{D}) \setminus \big(\mathbf{L} \cup \{X, Y\}\big)$.

Therefore, determining whether X and Y are adjacent in $\mathcal{H}$ requires only a single σ-separation test, rather than an exhaustive search over subsets of $\mathbf{O} \setminus \{X, Y\}$. Consequently, the adjacency can be determined in polynomial time.

Definition 4 (m-separation in σ-MAGs (Yao & Mooij, 2025))**.** *Let $\mathcal{H}$ be a σ-MAG, and let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint vertex sets. A path $\pi = \langle X, V_1, \ldots, V_n, Y \rangle$ is* m-open *given $\mathbf{Z}$ if*

(i) every non-collider on π lies outside $\mathbf{Z}$;
(ii) every collider on π belongs to $\mathrm{An}^+(\mathbf{Z}, \mathcal{H})$; and
(iii) π contains no subpath of the form $V_{k-1} {*\!\!\to} V_k - V_{k+1}$ or $V_{k-1} - V_k {\leftarrow\!\!*} V_{k+1}$.

The sets $\mathbf{X}$ and $\mathbf{Y}$ are m-separated *by $\mathbf{Z}$ in $\mathcal{H}$ if no m-open path given $\mathbf{Z}$ connects a vertex in $\mathbf{X}$ to a vertex in $\mathbf{Y}$.*

For ordinary MAGs, an arrowhead into an undirected edge is prohibited by definition, so condition (iii) is vacuous and the criterion reduces to standard m-separation (Richardson & Spirtes, 2002). Two σ-MAGs over the same vertex set are *m-Markov equivalent* if they entail the same m-separation relations. We denote the m-Markov equivalence class of $\mathcal{H}$ by $[\mathcal{H}]_m$.

Definition 5 (σ-PAG)**.** *A σ-partial ancestral graph (σ-PAG) $\mathcal{P}$ represents an m-Markov equivalence class $[\mathcal{H}]_m$ of σ-MAGs. It has the common skeleton of all members of $[\mathcal{H}]_m$. At each edge, $\mathcal{P}$ contains a tail ('$-$') or an arrowhead ('$>$') if that mark is invariant across $[\mathcal{H}]_m$; otherwise, the endpoint is represented by a circle $\circ$. For convenience, we use an asterisk (*) to denote any possible mark ($\circ, >, -$).*

Remark 1. *Definition 5 defines the global graphical representation recoverable from observational data under the σ-Markov and σ-faithfulness conditions. While Mooij & Claassen (2020) established the soundness and completeness of σ-FCI with respect to directed mixed graphs, in Appendix D we further establish its soundness and completeness with respect to σ-MAGs. In particular, σ-FCI recovers the σ-PAG representing the m-Markov equivalence class of the underlying σ-MAG.*

Problem Formulation: We consider a simple SCM over $\mathbf{V} = \mathbf{O} \dot{\cup} \mathbf{L}$ (Bongers et al., 2021; Mooij & Claassen, 2020), with an associated possibly cyclic causal DG $\mathcal{D}$, where $\mathbf{O}$ and $\mathbf{L}$ denote the observed and latent variables, respectively. Marginalizing $\mathbf{L}$ from $\mathcal{D}$ yields a σ-MAG over $\mathbf{O}$. Let $\mathcal{P}$ denote the σ-PAG representing its m-Markov equivalence class. Let $P_{\mathbf{O}}$ denote the observational distribution over $\mathbf{O}$. Under the σ-Markov and σ-faithfulness conditions, given $P_{\mathbf{O}}$ and a target $T \in \mathbf{O}$, our goal is to recover the local structure of T in $\mathcal{P}$, namely, all adjacencies of T together with the invariant endpoint marks on its incident edges.

4 Theoretical Foundations

Learning the global causal structure from the full observational distribution $P_{\mathbf{O}}$ can be computationally expensive, especially when the number of observed variables is large. A natural alternative is to identify a compact local region around a target variable and perform causal structure learning only within this region. In this work, we use the *Markov blanket* (MB) as such a local region.

In acyclic and causally sufficient systems, the MB of a vertex in a directed acyclic graph (DAG) consists of its parents, children, and the other parents of its children. When directed cycles and latent variables are allowed, this characterization no longer applies. The following lemma provides both a graphical characterization of the MB in a σ-MAG and an identifiable characterization from the observational distribution.

Lemma 2 (Markov blanket in a σ-MAG)**.** *Let $\mathcal{H}$ be a σ-MAG over $\mathbf{O}$. Under the σ-faithfulness condition, the Markov blanket of $X \in \mathbf{O}$ is the set of vertices $Y \in \mathbf{O} \setminus \{X\}$ that are connected to X by a collider path in $\mathcal{H}$, including paths of length one. We denote this set by $\mathrm{MB}(X, \mathcal{H})$. Moreover, for any distinct $X, Y \in \mathbf{O}$,*

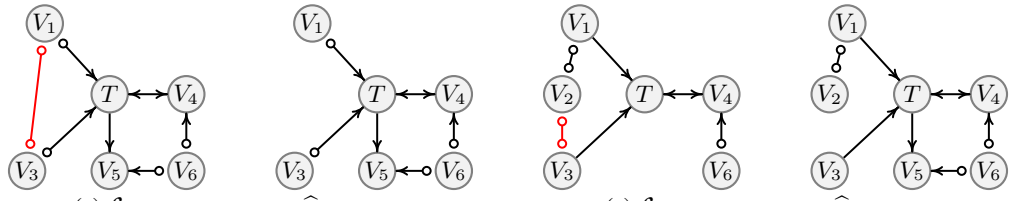
$$Y \in \mathrm{MB}(X, \mathcal{H}) \quad \Longleftrightarrow \quad X \not\perp\!\!\!\perp Y \mid \mathbf{O} \setminus \{X, Y\} \text{ under } P_{\mathbf{O}}. \tag{1}$$

As an illustration, consider the σ-MAG in Figure 1b. The vertices connected to T by collider paths form $\mathrm{MB}(T, \mathcal{H}) = \{V_1, V_3, V_4, V_5, V_6\}$, which m-separates T from all remaining observed vertices and is minimal with this property.

Lemma 2 further shows that the MB is identifiable from $P_{\mathbf{O}}$ through CI information. In practice, methods such as the TC algorithm (Pellet & Elisseeff, 2008) can identify the MB with a linear number of CI tests with respect to $|\mathbf{O}|$, without first learning the global causal structure. We therefore write $\mathrm{MB}(X)$ for $\mathrm{MB}(X, \mathcal{H})$ and define $\mathrm{MB}^+(X) = \mathrm{MB}(X) \cup \{X\}$ as the local region for subsequent structure learning.

In the remainder of this section, we study which structural information learned from the marginal distribution $P_{\mathrm{MB}^+(X)}$ is guaranteed to be consistent with global causal discovery. Let $\mathcal{L}_X$ denote the local σ-PAG learned from $P_{\mathrm{MB}^+(X)}$.

4.1 Local–Global Consistency



(a) $\mathcal{L}_T$ (b) $\widehat{\mathcal{P}}$ after learned $\mathcal{L}_T$ (c) $\mathcal{L}_{V_1}$ (d) $\widehat{\mathcal{P}}$ after learned $\mathcal{L}_{V_1}$

Figure 2: Illustration of the sequential learning process in Example 5 for recovering the target-specific structure around T. The underlying causal DG and the corresponding ground-truth σ-PAG are shown in Figures 1a and 1c, respectively. Red edges indicate locally learned structures that are not guaranteed to be consistent with the global σ-PAG.

We first establish a localization property of m-separation in σ-MAGs. This property is the key to relating causal structure learning over $\mathrm{MB}^+(X)$ to global structure learning over $\mathbf{O}$.

Lemma 3. *Given a σ-MAG $\mathcal{H}$ over $\mathbf{O}$, for any distinct $X, Y \in \mathbf{O}$, the following statements are equivalent:*

(i) there exists $\mathbf{Z} \subseteq \mathbf{O} \setminus \{X, Y\}$ that m-separates X and Y in $\mathcal{H}$;
(ii) there exists $\mathbf{Z}' \subseteq \mathrm{MB}^+(X, \mathcal{H}) \setminus \{X, Y\}$ that m-separates X and Y in $\mathcal{H}$.

Lemma 3 states that whenever X and Y can be m-separated using variables from the full observed set, they can also be m-separated using only variables in the local region $\mathrm{MB}^+(X)$. Equivalently, if X and Y remain m-connected given every subset of $\mathrm{MB}^+(X) \setminus \{X, Y\}$, then they remain m-connected given every subset of $\mathbf{O} \setminus \{X, Y\}$.

Lemma 3 above immediately yields consistency of adjacencies incident to the target variable.

Theorem 1. *Let $\mathcal{P}$ be the global σ-PAG over $\mathbf{O}$, let $X \in \mathbf{O}$, and let $\mathcal{L}_X$ be the local σ-PAG learned from $P_{\mathrm{MB}^+(X)}$. Then, for any $Y \in \mathbf{O} \setminus \{X\}$, Y is adjacent to X in $\mathcal{L}_X$ if and only if Y is adjacent to X in $\mathcal{P}$.*

Thus, Theorem 1 guarantees that restricting structure learning to the marginal distribution over $\mathrm{MB}^+(X)$ neither introduces nor removes adjacencies incident to X relative to global structure learning.

Example 1. Consider the locally learned structure $\mathcal{L}_T$ over $P_{\mathrm{MB}^+(T)}$, shown in Figure 2a, and the global σ-PAG $\mathcal{P}$, shown in Figure 1c. By Theorem 1, the adjacencies between T and $\{V_1, V_3, V_4, V_5\}$ in $\mathcal{L}_T$ coincide with those in $\mathcal{P}$. Likewise, every variable in $\mathbf{O} \setminus \{T, V_1, V_3, V_4, V_5\}$ is non-adjacent to T in $\mathcal{P}$.

Beyond adjacencies, the local region also preserves certain orientation information. We first consider uncovered collider paths originating from X.

Theorem 2. *Let $\mathcal{P}$ be the global σ-PAG over $\mathbf{O}$, let $X \in \mathbf{O}$, and let $\mathcal{L}_X$ be the local σ-PAG learned from $P_{\mathrm{MB}^+(X)}$. For any distinct vertices $V_1, \ldots, V_k, Y \in \mathrm{MB}(X)$ with $k \geq 1$, an uncovered collider path $X* \to V_1 \leftrightarrow \cdots \leftrightarrow V_k \leftarrow* Y$ is present in $\mathcal{L}_X$ if and only if the same uncovered collider path is present in the global σ-PAG $\mathcal{P}$.*

Theorem 2 shows that uncovered collider paths identified from the marginal distribution over $\mathrm{MB}^+(X)$ are neither spurious nor missing relative to the global σ-PAG. In particular, this result provides a local–global correspondence for a class of orientation information that plays a central role in constraint-based causal discovery.

Example 2. Consider the locally learned structure $\mathcal{L}_{V_1}$ over $P_{\mathrm{MB}^+(V_1)}$, shown in Figure 2c, and the global σ-PAG $\mathcal{P}$, shown in Figure 1c. By Theorem 2, the uncovered collider paths originating at V_1, $V_1 *\!\!\to T \leftrightarrow V_4 \leftarrow\!\!* V_6$ and $V_1 *\!\!\to T \leftarrow\!\!* V_3$, identified in $\mathcal{L}_{V_1}$ are also present in $\mathcal{P}$. Conversely, these are the only uncovered collider paths originating at V_1 in $\mathcal{P}$.

We next establish a more general soundness result for definite endpoint marks on edges in the local learned structure.

Theorem 3. *Let $\mathcal{P}$ be the global σ-PAG over $\mathbf{O}$, let $X \in \mathbf{O}$, and let $\mathcal{L}_X$ be the local σ-PAG learned from $P_{\mathrm{MB}^+(X)}$. For any adjacent pair $A, B \in \mathrm{MB}^+(X)$ in $\mathcal{L}_X$, if the edge between A and B is also present in $\mathcal{P}$, then every definite endpoint mark (i.e., a tail or an arrowhead) on this edge in $\mathcal{L}_X$ is also present at the corresponding endpoint in $\mathcal{P}$.*

Combining Theorem 1 and Theorem 3, every definite endpoint mark learned locally on an edge incident to X is sound with respect to the global σ-PAG. Likewise, combining Theorem 2 and Theorem 3, every definite endpoint mark learned locally on an uncovered collider path originating at X is sound with respect to the global σ-PAG.

Example 3. Consider the locally learned structure $\mathcal{L}_{V_1}$ over $P_{\mathrm{MB}^+(V_1)}$, shown in Figure 2c, and the global σ-PAG $\mathcal{P}$, shown in Figure 1c. In $\mathcal{L}_{V_1}$, consider the V-structure $V_1 \to T \leftarrow V_3$. By Theorem 1, the adjacency between V_1 and T agrees with that in $\mathcal{P}$. Moreover, by Theorem 2, the uncovered collider $V_1 *\!\!\to T \leftarrow\!\!* V_3$ is also present in $\mathcal{P}$, which in particular guarantees the adjacency between T and V_3 and the collider at T. Finally, by Theorem 3, the tails at V_1 and V_3 are also globally valid. Therefore, the entire V-structure $V_1 \to T \leftarrow V_3$ identified in $\mathcal{L}_{V_1}$ is consistent with the global σ-PAG $\mathcal{P}$.

A direct consequence of Theorem 3 is that unshielded colliders centered at the target variable are also globally valid.

Corollary 1. *Let $\mathcal{P}$ be the global σ-PAG over $\mathbf{O}$, let $X \in \mathbf{O}$, and let $\mathcal{L}_X$ be the local σ-PAG learned from $P_{\mathrm{MB}^+(X)}$. Let $A, B \in \mathrm{MB}(X)$ be distinct vertices. If $A *\!\!\to X \leftarrow\!\!* B$ is an unshielded collider in $\mathcal{L}_X$, then it is also an unshielded collider in $\mathcal{P}$.*

4.2 Limitations of a Single Markov-Blanket Region

The above results establish a strong local–global correspondence within a single MB region. In particular, learning from $P_{\mathrm{MB}^+(X)}$ exactly recovers all adjacencies incident to X, preserves uncovered collider paths originating at X, and yields only globally valid definite endpoint marks on edges that are also present in the global σ-PAG. Thus, a substantial amount of global causal information can be reliably recovered without learning the full structure.

However, a single MB region may not contain complete information to identify all invariant endpoint marks relevant to the target. In particular, some orientations in the global σ-PAG may depend on

structural information involving variables outside $\mathrm{MB}^+(X)$ and therefore cannot be determined from $P_{\mathrm{MB}^+(X)}$ alone. The following example illustrates this limitation.

Example 4. Consider the locally learned structure $\mathcal{L}_T$ over $P_{\mathrm{MB}^+(T)}$, shown in Figure 2a, and the global σ-PAG $\mathcal{P}$, shown in Figure 1c. In $\mathcal{L}_T$, V_1 and V_3 are adjacent, whereas they are non-adjacent in $\mathcal{P}$. This discrepancy arises because the separating variable V_2 lies outside $\mathrm{MB}^+(T)$ and is therefore unavailable when learning from $P_{\mathrm{MB}^+(T)}$ alone. Consequently, the local structure cannot exploit this non-adjacency information, leaving the endpoints at V_1 and V_3 on the edges incident to T unoriented. In contrast, these endpoints are definite tails in the global σ-PAG.

This example shows that, although a single MB region provides reliable structural information, it may not be complete for target-specific discovery. Recovering the complete target-specific structure may therefore require incorporating additional structural information beyond a single local region.

5 Algorithm

Building on the local–global consistency results in Section 4, we develop LoCL, a local causal discovery algorithm for recovering the target-specific structure identifiable by global causal discovery in the presence of latent variables and directed cycles. The procedure is summarized in Algorithm 1.

Algorithm 1 **Lo**cal Causal Discovery in the Presence of **C**ycles and **L**atent Variables (LoCL)

Input: Target variable T, observed distribution $P_{\mathbf{O}}$
Output: Learned target-specific causal structure $\widehat{\mathcal{P}}$

Initialize $\widehat{\mathcal{P}} \leftarrow \varnothing$, $\mathbf{Sepsets} \leftarrow \varnothing$, $\mathrm{Next} \leftarrow \{T\}$, and $\mathrm{Done} \leftarrow \varnothing$
repeat
  $X \leftarrow$ the first variable in Next.
  $\mathrm{MB}^+(X) \leftarrow \textsc{MBLearn}(X, P_{\mathbf{O}}, \mathbf{O})$
  $(\mathcal{L}_X, \mathbf{Sepsets}_X) \leftarrow \textsc{LocalLearn}(P_{\mathrm{MB}^+(X)})$
  $\mathbf{Sepsets} \leftarrow \mathbf{Sepsets} \cup \mathbf{Sepsets}_X$
  Update $\widehat{\mathcal{P}}$ using $\mathcal{L}_X$ according to Theorems 1 to 3
  $\widehat{\mathcal{P}} \leftarrow \textsc{Orient}(\widehat{\mathcal{P}}, \mathbf{Sepsets})$
  $\mathrm{Done} \leftarrow \mathrm{Done} \cup \{X\}$
  $\mathrm{Next} \leftarrow \left\{V \in \mathbf{V}(\widehat{\mathcal{P}}) \setminus \mathrm{Done} : V \text{ has a potentially anterior path to } T \text{ in } \widehat{\mathcal{P}}\right\}$
until all endpoint marks incident to T are determined or $\mathrm{Next} = \varnothing$
return $\widehat{\mathcal{P}}$

The algorithm maintains an accumulated partial graph $\widehat{\mathcal{P}}$, a collection of separating sets $\mathbf{Sepsets}$, a set Next of variables to be processed, and a set Done of processed variables. For each selected variable X, MBLearn (Pellet & Elisseeff, 2008) identifies its MB region $\mathrm{MB}^+(X)$, and LocalLearn applies a constraint-based causal discovery (Spirtes et al., 1995; Zhang, 2008) procedure to $P_{\mathrm{MB}^+(X)}$ to obtain $\mathcal{L}_X$ and the corresponding separating sets. Only the structural information guaranteed to be globally valid by Theorems 1 to 3 is incorporated into $\widehat{\mathcal{P}}$. The accumulated separating sets are then used to further orient the graph.

As discussed in Section 4.2, a single MB region may not contain complete information to determine all invariant endpoint marks incident to the target. Therefore, LoCL continues to explore unprocessed variables that remain potentially anterior to T in the current graph, until the target-specific structure is determined or no such variable remains.

Example 5. We illustrate LoCL using Figure 2. Let T be the target variable, and let the global σ-PAG be shown in Figure 1c. Initially, $\mathrm{Next} = \{T\}$, $\mathrm{Done} = \varnothing$, and $\widehat{\mathcal{P}} = \varnothing$.

- **Processing** T**.** The algorithm first identifies $\mathrm{MB}^+(T) = \{V_1, V_3, V_4, V_5, V_6, T\}$ and learns $\mathcal{L}_T$ from $P_{\mathrm{MB}^+(T)}$, as shown in Figure 2a. The globally valid local information is incorporated into $\widehat{\mathcal{P}}$, yielding Figure 2b. The candidate set is then updated to $\mathrm{Next} = \{V_1, V_3\}$.
- **Processing** V_1**.** The algorithm next identifies $\mathrm{MB}^+(V_1) = \{V_1, V_2, V_3, V_4, V_6, T\}$ and learns $\mathcal{L}_{V_1}$ from $P_{\mathrm{MB}^+(V_1)}$, as shown in Figure 2c. After incorporating the newly obtained information

and applying the orientation rules, $\widehat{\mathcal{P}}$ is updated to Figure 2d. At this point, all endpoint marks incident to T are determined as $V_1 \to T$, $V_3 \to T$, $T \to V_5$, and $T \leftrightarrow V_4$.

The algorithm therefore terminates and returns the learned target-specific structure around T.

Theorem 4 (Soundness and completeness of LoCL)**.** *Assume that $P_{\mathbf{O}}$ satisfies the σ-Markov and σ-faithfulness conditions with respect to the underlying causal DG. Let $\mathcal{P}$ be the global σ-PAG over $\mathbf{O}$. Then LoCL exactly recovers the adjacencies incident to the target T, together with all invariant endpoint marks on these edges encoded by $\mathcal{P}$.*

Complexity Analysis. Let $|\mathbf{O}|$ denote the number of observed variables, let r denote the number of local structures learned by LoCL, and let $b = \max_{X \in \mathrm{Done}} |\mathrm{MB}^+(X)|$ denote the maximum local search-space size encountered during the procedure. The computational cost mainly consists of MB discovery and local causal structure learning. Using an MB learner such as TC (Pellet & Elisseeff, 2008), the cost of identifying the MBs of the r processed variables is $\mathcal{O}(r|\mathbf{O}|)$. If a constraint-based method such as σ-FCI (Mooij & Claassen, 2020) is used as the local learner, the total cost of the r local subproblems is $\mathcal{O}(rL(b))$, where $L(\cdot)$ denotes the complexity of the local learner as a function of the local search-space size. Therefore, the overall complexity of LoCL is bounded by $\mathcal{O}(r\,(|\mathbf{O}| + L(b)))$. Since typically $b \ll |\mathbf{O}|$, LoCL can reduce the computational cost compared with global causal discovery over all observed variables.

6 Experiments

Baselines. We compare LoCL with two representative global causal discovery methods and three local causal discovery methods. For global discovery, we consider σ-FCI (Mooij & Claassen, 2020), which extends FCI to settings with directed cycles and latent confounding, and the recent L-MARVEL (Akbari et al., 2021) method.[1] For local discovery, we compare with MB-by-MB (Wang et al., 2014), CMB (Gao & Ji, 2015), and the recent LatentLCD (Ling et al., 2025), which further allows for latent variables. Our source code is included in the supplementary materials.

Experimental setup and metrics. Following the convention in (Akbari et al., 2021; Mooij & Claassen, 2020; Garcez Vicente et al., 2026), the underlying cyclic causal structures are parameterized as linear Gaussian simple structural causal models (Bongers et al., 2021). Causal coefficients are sampled uniformly from $\pm(0.5, 1)$, and all noise terms are independently drawn from a standard Gaussian distribution. We evaluate all methods in terms of target-specific structural accuracy and computational efficiency. Structural accuracy is evaluated only on the edge marks incident to the target T, using mark-level *Precision*, *Recall*, *F1*, and *Local-SHD*. For efficiency, we report runtime and the number of conditional independence (CI) tests, with lower values indicating better efficiency. Detailed definitions of the evaluation metrics are provided in Appendix E.2.

6.1 Random Structures

We generate random directed cyclic causal structures using a fixed-edge variant of the Erdős–Rényi (ER) model (Erdős & Rényi, 1960). We vary the graph dimension as $d \in \{50, 100, 150\}$. For $d = 50$, we enforce four directed cycles of length 3 and randomly designate 4 variables as latent; for $d = 100$, we enforce five directed cycles of length 4 and designate 5 variables as latent; and for $d = 150$, we enforce six directed cycles of length 5 and designate 6 variables as latent. Across all settings, the resulting ground-truth σ-PAGs have an average degree between 3.0 and 3.5. All reported results are averaged over 20 independently generated datasets for each setting. Target variables are selected from vertices with relatively large neighborhoods.

Figure 3 reports Local SHD, runtime, and the number of CI tests for $d = 100$ and $d = 150$ across sample sizes. Complete results for all graph sizes and evaluation metrics are provided in Appendix E. The results show that LoCL outperforms the baselines on nearly all metrics across graph sizes and sample sizes. As expected, it requires substantially fewer CI tests than the global methods FCI and L-MARVEL. The local methods MB-by-MB, CMB, and LatentLCD perform less well, suggesting limitations in handling latent confounding and directed cycles simultaneously.

[1] L-MARVEL is theoretically developed under the acyclicity assumption. However, since it follows an FCI-style constraint-based framework, and FCI remains sound and complete under σ-separation in the presence of directed cycles (Mooij & Claassen, 2020), L-MARVEL may also retain similar properties in the cyclic setting.

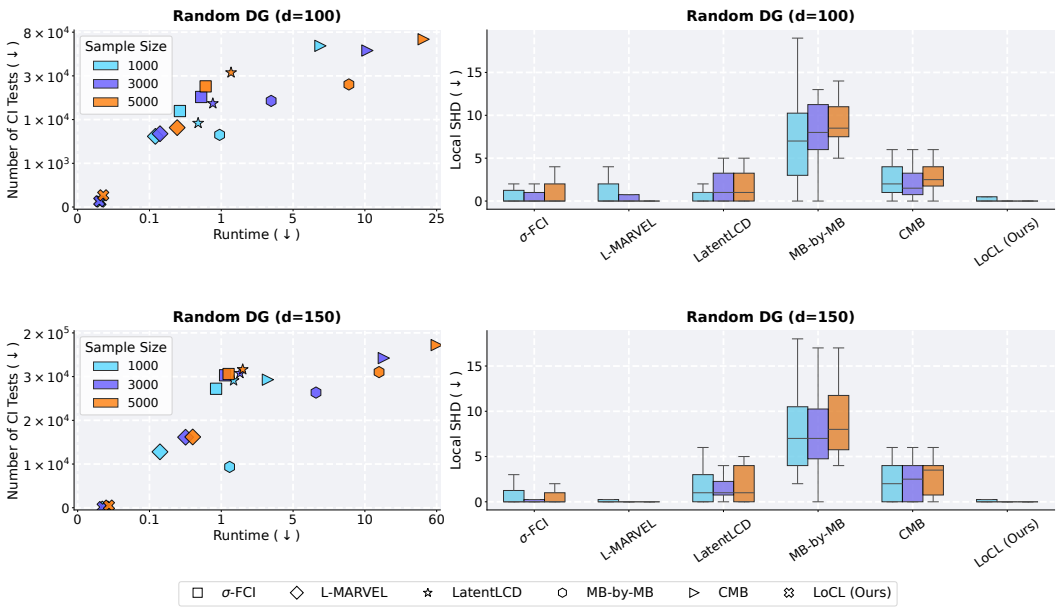


Figure 3: Results on random directed cyclic graphs with $d = 100$ (top row) and $d = 150$ (bottom row) across sample sizes. Lower values are better for all three metrics (↓).

6.2 Real-World Application

We further evaluate the compared methods on a biologically grounded gene regulatory network from the DREAM4 In Silico Network Challenge (Marbach et al., 2010). The network contains 100 vertices and 176 directed edges, with four nontrivial strongly connected components of sizes 4, 3, 2, and 2. We use a sample size of 1,000. In each repetition, we randomly designate 5 vertices as latent and select one observed vertex as the target. Results are averaged over 20 independent repetitions.

Table 1: Results on the DREAM4 gene regulatory network.

	Structural Accuracy				Efficiency	
Method	Precision ↑	Recall ↑	F1 ↑	Local SHD ↓	Runtime ↓	CI tests ($\times 10^3$) ↓
FCI	0.25 ± 0.24	0.23 ± 0.23	0.23 ± 0.23	3.30 ± 1.26	108.68 ± 164.32	2560 ± 4510
L-MARVEL	0.39 ± 0.30	0.39 ± 0.27	0.38 ± 0.27	$\mathbf{2.65 \pm 1.09}$	0.26 ± 0.32	7.73 ± 5.37
MB-by-MB	0.03 ± 0.05	0.15 ± 0.24	0.05 ± 0.08	13.90 ± 5.41	7.17 ± 4.76	63.4 ± 48.9
CMB	0.27 ± 0.23	0.30 ± 0.24	0.28 ± 0.23	4.20 ± 3.16	86.98 ± 119.42	821 ± 1250
LatentLCD	0.29 ± 0.29	0.29 ± 0.28	0.28 ± 0.29	3.45 ± 2.31	1.35 ± 1.10	35.3 ± 38.1
LoCL (ours)	$\mathbf{0.59 \pm 0.31}$	$\mathbf{0.47 \pm 0.11}$	$\mathbf{0.50 \pm 0.16}$	$\underline{2.70 \pm 1.63}$	$\mathbf{0.11 \pm 0.14}$	$\mathbf{4.21 \pm 3.40}$

Note: ↑/↓ indicates that higher/lower values are better. Bold marks the best mean in each column; the underlined entry marks the second-best mean Local SHD. CI counts are reported in thousands.

As shown in Table 1, LoCL achieves the highest mean precision, recall, and F1, and the second-lowest mean Local SHD, indicating strong target-specific structural recovery. It also achieves the lowest mean runtime and CI test count, extending the efficiency gains observed on random structures to this real biological system.

7 Discussion and Conclusion

We proposed LoCL for local causal discovery with latent variables and directed cycles. By connecting local and global causal information, LoCL recovers the target-specific structure without learning the full graph, with soundness and completeness guarantees. Experiments demonstrate strong structural accuracy with reduced runtime and fewer CI tests than both existing global and local methods. Future work includes using interventional data (Baldo & Assaad, 2025; Zhou et al., 2025) or background knowledge (Wang et al., 2023; Zheng et al., 2026) to resolve remaining ambiguities, and extend the framework to multiple targets.

REFERENCES

Sina Akbari, Ehsan Mokhtarian, AmirEmad Ghassami, and Negar Kiyavash. Recursive causal structure learning in the presence of latent variables and selection bias. *Advances in Neural Information Processing Systems*, 34:10119–10130, 2021.

Federico Baldo and Charles K Assaad. Regret-based federated causal discovery with unknown interventions. *arXiv preprint arXiv:2512.23626*, 2025.

Juliana Benito, Hao Zheng, Fanny S Ng, and Paul E Hardin. Transcriptional feedback loop regulation, function, and ontogeny in drosophila. In *Cold Spring Harbor symposia on quantitative biology*, volume 72, pp. 437–444. Cold Spring Harbor Laboratory Press, 2007.

Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M Mooij. Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5):2885–2915, 2021.

Tom Claassen, Joris M Mooij, and Tom Heskes. Learning sparse causal models is not np-hard. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 172. Citeseer, 2013.

Diego Colombo, Marloes H Maathuis, Markus Kalisch, and Thomas S Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pp. 294–321, 2012.

Gregory F Cooper. A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Mining and Knowledge Discovery*, 1:203–224, 1997.

Haoyue Dai, Ignavier Ng, Yujia Zheng, Zhengqing Gao, and Kun Zhang. Local causal discovery with linear non-gaussian cyclic models. In *International Conference on Artificial Intelligence and Statistics*, pp. 154–162. PMLR, 2024.

Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences,*, 5:17–61, 1960.

Patrick Forré and Joris M Mooij. Markov properties for graphical models with cycles and latent variables. *arXiv preprint arXiv:1710.08775*, 2017.

Tian Gao and Qiang Ji. Local causal discovery of direct causes and effects. *Advances in Neural Information Processing Systems*, 28, 2015.

Ana Leticia Garcez Vicente, Gijs van Seeventer, and Saber Salehkaleybar. Data-driven covariate selection for nonparametric and cycle-agnostic causal effect estimation. *arXiv e-prints*, pp. arXiv–2605, 2026.

Zhi Geng, Yue Liu, Chunchen Liu, and Wang Miao. Evaluation of causal effects and local structure learning of causal networks. *Annual review of statistics and its application*, 6(1):103–124, 2019.

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10:524, 2019.

Trygve Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica, Journal of the Econometric Society*, pp. 1–12, 1943.

Christina Heinze-Deml, Marloes H Maathuis, and Nicolai Meinshausen. Causal structure learning. *Annual Review of Statistics and Its Application*, 5:371–391, 2018.

Miguel A Hernán and James M Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.

Neville Kenneth Kitson, Anthony C Constantinou, Zhigao Guo, Yang Liu, and Kiattikun Chobtham. A survey of bayesian network structure learning. *Artificial Intelligence Review*, pp. 1–94, 2023.

Zheng Li, Zeyu Liu, Feng Xie, Hao Zhang, Chunchen Liu, and Zhi Geng. Local identifying causal relations in the presence of latent variables. In *International Conference on Machine Learning*, pp. 35390–35418. PMLR, 2025.

Zheng Li, Hao Zhang, Ruxin Wang, Ruichu Cai, Kun Zhang, and Feng Xie. Local causal structure learning in the presence of latent variables and selection bias. *arXiv preprint arXiv:2607.19866*, 2026.

Jiaxuan Liang, Jun Wang, Guoxian Yu, Carlotta Domeniconi, Xiangliang Zhang, and Maozu Guo. Gradient-based local causal structure learning. *IEEE Transactions on Cybernetics*, 2024.

Zhaolong Ling, Kui Yu, Lin Liu, Jiuyong Li, Yiwen Zhang, and Xindong Wu. Psl: an algorithm for partial bayesian network structure learning. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16(5):1–25, 2022.

Zhaolong Ling, Jiale Yu, Yiwen Zhang, Debo Cheng, Peng Zhou, Xingyu Wu, Bingbing Jiang, and Kui Yu. Local causal discovery without causal sufficiency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18737–18745, 2025.

Subramani Mani and Gregory F Cooper. Causal discovery using a bayesian local causal discovery algorithm. In *MEDINFO 2004*, pp. 731–735. IOS Press, 2004.

Subramani Mani, Peter Spirtes, and Gregory F Cooper. A theoretical study of y structures for causal discovery. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pp. 314–323, 2006.

Daniel Marbach, Robert J Prill, Thomas Schaffter, Claudio Mattiussi, Dario Floreano, and Gustavo Stolovitzky. Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the national academy of sciences*, 107(14):6286–6291, 2010.

Samuel J Mason. Feedback theory-some properties of signal flow graphs. *Proceedings of the IRE*, 41(9):1144–1156, 1953.

Ehsan Mokhtarian, Sepehr Elahi, Sina Akbari, and Negar Kiyavash. Recursive causal discovery. *Journal of Machine Learning Research*, 26(61):1–65, 2025.

Joris M Mooij and Tom Claassen. Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. In *Conference on Uncertainty in Artificial Intelligence*, pp. 1159–1168. Pmlr, 2020.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

Judea Pearl. Theoretical impediments to machine learning with seven sparks from the causal revolution. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 3–3, 2018.

Jean-Philippe Pellet and André Elisseeff. Using markov blankets for causal structure learning. *Journal of Machine Learning Research*, 9(7), 2008.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, 2017.

Thomas Richardson and Peter Spirtes. Ancestral graph markov models. *The Annals of Statistics*, 30 (4):962–1030, 2002.

Raanan Y Rohekar, Shami Nisimov, Yaniv Gurwicz, and Gal Novik. Iterative causal discovery in the possible presence of latent confounders and selection bias. *Advances in Neural Information Processing Systems*, 34:2454–2465, 2021.

Ruben Sanchez-Romero, Joseph D Ramsey, Kun Zhang, Madelyn RK Glymour, Biwei Huang, and Clark Glymour. Estimating feedforward and feedback effective connections from fmri time series: Assessments of statistical methods. *Network Neuroscience*, 3(2):274–306, 2019.

Bernhard Schölkopf. Causality for machine learning. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pp. 765–804. 2022.

Craig Silverstein, Sergey Brin, Rajeev Motwani, and Jeff Ullman. Scalable techniques for mining causal structures. *Data Mining and Knowledge Discovery*, 4:163–192, 2000.

Stephen M Smith, Karla L Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F Beckmann, Thomas E Nichols, Joseph D Ramsey, and Mark W Woolrich. Network modelling methods for fmri. *Neuroimage*, 54(2):875–891, 2011.

Peter Spirtes and Clark Glymour. An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, 9(1):62–72, 1991.

Peter Spirtes and Kun Zhang. Causal discovery and inference: concepts and recent methodological advances. *Applied Informatics*, 3(1):1–28, 2016.

Peter Spirtes, Christopher Meek, and Thomas Richardson. Causal inference in the presence of latent variables and selection bias. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence (UAI)*, pp. 499–506, 1995.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT press, 2000.

Thomas Verma and Judea Pearl. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, pp. 255–270, 1990.

Philip Versteeg, Joris Mooij, and Cheng Zhang. Local constraint-based causal discovery under selection bias. In *Conference on Causal Learning and Reasoning*, pp. 840–860. PMLR, 2022.

Changzhang Wang, You Zhou, Qiang Zhao, and Zhi Geng. Discovering and orienting the edges connected to a target variable in a dag via a sequential local learning approach. *Computational Statistics & Data Analysis*, 77:252–266, 2014.

Tian-Zuo Wang, Tian Qin, and Zhi-Hua Zhou. Sound and complete causal identification with latent variables given local background knowledge. *Artificial Intelligence*, pp. 103964, 2023.

Feng Xie, Zheng Li, Peng Wu, Yan Zeng, Chunchen Liu, and Zhi Geng. Local causal structure learning in the presence of latent variables. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 54511–54530, 2024.

Shuai Yang, Hao Wang, Kui Yu, Fuyuan Cao, and Xindong Wu. Towards efficient local causal structure learning. *IEEE Transactions on Big Data*, 8(6):1592–1609, 2021.

Binghua Yao and Joris M. Mooij. σ-maximal ancestral graphs. In *Proceedings of the Forty-First Conference on Uncertainty in Artificial Intelligence*, 2025.

Jianxin Yin, You Zhou, Changzhang Wang, Ping He, Cheng Zheng, and Zhi Geng. Partial orientation and local structural learning of causal networks for prediction. In *Causation and Prediction Challenge*, pp. 93–105. PMLR, 2008.

Kui Yu, Xianjie Guo, Lin Liu, Jiuyong Li, Hao Wang, Zhaolong Ling, and Xindong Wu. Causality-based feature selection: Methods and evaluations. *ACM Computing Surveys (CSUR)*, 53(5):1–36, 2020.

Alessio Zanga, Elif Ozkirimli, and Fabio Stella. A survey on causal discovery: Theory and practice. *International Journal of Approximate Reasoning*, 151:101–129, 2022.

Jiji Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17):1873–1896, 2008.

Qingyuan Zheng, Yue Liu, and Yangbo He. Local causal discovery with background knowledge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

You Zhou, Changzhang Wang, Jianxin Yin, and Zhi Geng. Discover local causal network around a target to a given depth. In *Causality: Objectives and Assessment*, pp. 191–202. PMLR, 2010.

Zihan Zhou, Muhammad Qasim Elahi, and Murat Kocaoglu. Characterization and learning of causal graphs from hard interventions. *Advances in Neural Information Processing Systems*, 38:77620–77663, 2025.

APPENDIX CONTENTS

A Background and Notation **14**
A.1 Graphical Definitions 14
A.2 σ-Markov and σ-Faithfulness 16

B Proofs of the Main Results **16**
B.1 Proof of Lemma 1 16
B.2 Proof of Lemma 2 18
B.3 Proof of Lemma 3 18
B.4 Proof of Theorem 2 21
B.5 Proofs of Theorem 3 and Corollary 1 24
B.6 Proof of Theorem 4 24

C Algorithmic Details **25**
C.1 Orientation Rules 25
C.2 MBLEARN: Markov Blanket Discovery 27

D Soundness and Completeness of σ-FCI for σ-MAGs **27**

E Experimental Details and Additional Results **29**
E.1 Baseline Implementations and Experimental Setup 29
E.2 Evaluation Metrics 29
E.3 Additional Results on Random Structures 30

F AI Use Statement **34**

A Background and Notation

This appendix collects additional graphical notions used primarily in the theoretical analysis. We keep these definitions here to streamline the main exposition.

Table 2: Main symbols and notation used in this paper.

Symbol	Description
$\mathbf{V}$, $\mathbf{O}$, $\mathbf{L}$	Full, observed, and latent variable sets, with $\mathbf{V} = \mathbf{O} \dot\cup \mathbf{L}$.
X, $\mathbf{X}$, $\lvert\mathbf{X}\rvert$	An individual variable, a variable set, and its cardinality, respectively.
$T \in \mathbf{O}$	Target observed variable whose incident adjacencies and invariant endpoint marks are to be recovered.
$\mathcal{D} = (\mathbf{V}, \mathbf{E})$	Underlying directed graph, possibly cyclic, with vertex set $\mathbf{V}$ and directed edge set $\mathbf{E}$.
$\mathcal{H}$	The σ-MAG over $\mathbf{O}$ representing the observed separation relations induced by $\mathcal{D}$.
$[\mathcal{H}]_m$	The m-Markov equivalence class of $\mathcal{H}$.
$\mathcal{P}$	Global σ-PAG representing $[\mathcal{H}]_m$.
$\widehat{\mathcal{P}}$	Accumulated graph estimate maintained and updated by LoCL.
$\mathcal{L}_X$	Local σ-PAG learned from the marginal distribution $P_{\mathrm{MB}^+(X)}$.
$P_{\mathbf{O}}$, $P_{\mathbf{W}}$	Observational distribution over $\mathbf{O}$ and its marginal over $\mathbf{W} \subseteq \mathbf{O}$.
$\mathrm{MB}(X, \mathcal{H})$, $\mathrm{MB}(X)$	Markov blanket of X: vertices connected to X by collider paths in $\mathcal{H}$, including paths of length one.
$\mathrm{MB}^+(X)$	Local search region $\mathrm{MB}(X) \cup \{X\}$.
$\mathrm{An}(X, \mathcal{G})$, $\mathrm{De}(X, \mathcal{G})$	Ancestors and descendants of X in $\mathcal{G}$; extended to sets by union.
$\mathrm{An}^+(\mathbf{A}, \mathcal{G})$	Ancestors including the set itself: $\mathbf{A} \cup \mathrm{An}(\mathbf{A}, \mathcal{G})$.
$\mathrm{Ant}(X, \mathcal{G})$, $\mathrm{Ant}^+(X, \mathcal{G})$	Anterior vertices of X in $\mathcal{G}$, excluding and including X, respectively; extended to sets by union.
$\mathrm{SCC}(X, \mathcal{D})$	Strongly connected component containing X in $\mathcal{D}$.
$\mathrm{ArrColl}(X, \mathcal{H})$	Vertices reachable from X through an arrow-collider path in $\mathcal{H}$.
$\mathrm{Sepset}(X, Y)$	A conditioning set that separates X and Y.
$\mathbf{Sepsets}_X$, $\mathbf{Sepsets}$	Separating sets returned by local learning at X, and the accumulated collection, respectively.
Next, Done	Variables awaiting local learning and variables already processed by LoCL, respectively.
$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$	Conditional independence of $\mathbf{X}$ and $\mathbf{Y}$ given $\mathbf{Z}$ in the distribution; $\not\perp\!\!\!\perp$ denotes conditional dependence.
$X \perp\!\!\!\perp_\sigma Y \mid \mathbf{Z}$	σ-separation of X and Y given $\mathbf{Z}$ in the underlying DG; $\not\perp\!\!\!\perp_\sigma$ denotes failure of σ-separation.

A.1 Graphical Definitions

Definition 6 (Walks, paths, and colliders)**.** *Let $\mathcal{G}$ be a mixed graph. A* walk *from V_0 to V_k is a finite sequence $\langle V_0, \ldots, V_k \rangle$ such that every consecutive pair is adjacent. A walk is a* path *if no vertex occurs more than once.*

A non-endpoint vertex V_i on a walk is a collider *if both incident edges have arrowheads at V_i, i.e.,*

$$V_{i-1} {*\!\!\rightarrow} V_i \leftarrow\!\!* V_{i+1}.$$

Otherwise, V_i is a non-collider.

Definition 7 (Unshielded and uncovered paths)**.** *A triple $\langle A, B, C \rangle$ is* unshielded *if A and C are non-adjacent while both are adjacent to B. A path is* uncovered *if every consecutive triple on the path is unshielded. An* uncovered collider path *is an uncovered path whose non-endpoint vertices are all colliders.*

Definition 8 (Anterior path)**.** *Let $\mathcal{G}$ be a mixed graph. A path $\langle V_0, \ldots, V_k \rangle$ is an* anterior path *from V_0 to V_k if every edge is out of the preceding vertex, i.e.,*

$$V_0 \mathrel{-\!\!*} V_1 \mathrel{-\!\!*} \cdots \mathrel{-\!\!*} V_k.$$

A vertex A is anterior *to B if there exists an anterior path from A to B. We denote the set of anteriors of B by $\mathrm{Ant}(B, \mathcal{G})$ and define*

$$\mathrm{Ant}^+(\mathbf{A}, \mathcal{G}) := \mathbf{A} \cup \mathrm{Ant}(\mathbf{A}, \mathcal{G}).$$

For a vertex set $\mathbf{A}$, $\mathrm{Ant}(\mathbf{A}, \mathcal{G})$ is defined by union over its elements.

Definition 9 (Potentially Anterior Path)**.** *A path π between X and Y in a σ-PAG is a* potentially anterior path *from X to Y if there is no arrowhead on the path pointing toward X.*

Definition 10 (Inducing walk and path (Yao & Mooij, 2025))**.** *Let $\mathcal{H}$ be a σ-MAG. A walk between distinct vertices X and Y is an* inducing walk *if every non-endpoint vertex is a collider and belongs to $\mathrm{An}(\{X, Y\}, \mathcal{H})$. If no vertex is repeated, it is called an* inducing path*.*

Definition 11 (Strongly Connected Component)**.** *Let $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ be a directed graph, possibly containing directed cycles. For any $X \in \mathbf{V}$, the* strongly connected component *(SCC) of X in $\mathcal{D}$, denoted by $\mathrm{SCC}(X, \mathcal{D})$, is defined as*

$$\mathrm{SCC}(X, \mathcal{D}) = \big\{V \in \mathbf{V} : V \in \mathrm{An}^+(X, \mathcal{D}) \text{ and } X \in \mathrm{An}^+(V, \mathcal{D})\big\}.$$

Equivalently, $\mathrm{SCC}(X, \mathcal{D})$ consists of all vertices that are mutually reachable with X by directed paths in $\mathcal{D}$.

Definition 12 (Directed mixed graph)**.** *A* directed mixed graph *(DMG) is a graph $\mathcal{G} = (\mathbf{V}, \mathbf{E}, \mathbf{F})$ containing directed edges $X \to Y$ and bidirected edges $X \leftrightarrow Y$. Directed cycles are allowed. A DG is a DMG with no bidirected edges.*

Definition 13 (Marginalization (Bongers et al., 2021))**.** *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E}, \mathbf{F})$ be a DMG and let $\mathbf{W} \subseteq \mathbf{V}$. The marginalization of $\mathcal{G}$ over $\mathbf{W}$, denoted by*

$$\mathrm{marg}_{\mathbf{W}}(\mathcal{G}),$$

is the DMG over $\mathbf{V} \setminus \mathbf{W}$ containing:

(i) a directed edge $X \to Y$ whenever $\mathcal{G}$ contains a directed walk

$$X \to W_1 \to \cdots \to W_k \to Y$$

whose non-endpoint vertices all belong to $\mathbf{W}$; and

(ii) a bidirected edge $X \leftrightarrow Y$ whenever $\mathcal{G}$ contains a bifurcation

$$X \leftarrow W_1 \leftarrow \cdots \leftarrow * W_k \to \cdots \to W_m \to Y$$

whose non-endpoint vertices all belong to $\mathbf{W}$.

For the underlying DG $\mathcal{D}$ considered in this paper, we write

$$\mathcal{G}_{\mathbf{O}} := \mathrm{marg}_{\mathbf{L}}(\mathcal{D})$$

for the latent projection onto the observed variables.

Definition 14 (Blockable and unblockable non-colliders (Yao & Mooij, 2025))**.** *Let $\mathcal{G}$ be a DMG and let $\pi = \langle V_0, \ldots, V_k \rangle$ be a walk. A non-endpoint non-collider V_i is* unblockable *if every outgoing edge from V_i along π points to a neighboring vertex in $\mathrm{SCC}(V_i, \mathcal{G})$. Otherwise, V_i is* blockable*. In particular, every endpoint is blockable, and an internal non-collider is blockable whenever it has an outgoing path edge $V_i \to V_j$ with $V_j \notin \mathrm{SCC}(V_i, \mathcal{G})$.*

Definition 15 (σ-open walk)**.** *Let $\mathcal{G}$ be a DMG, let $\mathbf{Z}$ be a vertex set, and let π be a walk whose endpoints lie outside $\mathbf{Z}$. The walk π is* σ-open *given $\mathbf{Z}$ if*

(i) every collider on π belongs to $\mathrm{An}^+(\mathbf{Z}, \mathcal{G})$; and
(ii) every blockable non-collider on π lies outside $\mathbf{Z}$.

Otherwise, π is σ-blocked by $\mathbf{Z}$.

Definition 16 (σ-inducing walk and path (Mooij & Claassen, 2020; Yao & Mooij, 2025))**.** *Let $\mathcal{G}$ be a DMG and let X and Y be distinct vertices. A walk π between X and Y is* σ-inducing *if every collider on π belongs to $\mathrm{An}^+(\{X, Y\}, \mathcal{G})$ and every non-endpoint non-collider is unblockable. If π is a path, it is called a* σ-inducing path*.*

Definition 17 (σ-acyclification)**.** *Let $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ be a directed graph, possibly cyclic, where $\mathbf{V} = \mathbf{O} \dot{\cup} \mathbf{L}$. A* σ-acyclification *of $\mathcal{D}$ is a directed acyclic graph $\mathcal{D}^{\mathrm{acy}} = (\mathbf{V}, \mathbf{E}^{\mathrm{acy}})$ on the same vertex set such that, for any distinct $X, Y \in \mathbf{V}$:*

1. *if $X \notin \mathrm{SCC}(Y, \mathcal{D})$, then $X \to Y \in \mathbf{E}^{\mathrm{acy}}$ if and only if there exists $W \in \mathrm{SCC}(Y, \mathcal{D})$ such that $X \to W \in \mathbf{E}$;*
2. *if $X \in \mathrm{SCC}(Y, \mathcal{D})$, then exactly one of $X \to Y$ and $Y \to X$ is included in $\mathbf{E}^{\mathrm{acy}}$, with the orientations chosen such that $\mathcal{D}^{\mathrm{acy}}$ is acyclic.*

Definition 18 (Markov blanket)**.** *Let P be a distribution over $\mathbf{O}$ and $X \in \mathbf{O}$. A set $\mathrm{MB}(X) \subseteq \mathbf{O} \setminus \{X\}$ is a* Markov blanket *of X if*

$$X \perp\!\!\!\perp \mathbf{O} \setminus \mathrm{MB}^{+}(X) \mid \mathrm{MB}(X), \qquad \mathrm{MB}^{+}(X) := \mathrm{MB}(X) \cup \{X\},$$

and no proper subset of $\mathrm{MB}(X)$ satisfies this property.

A.2 σ-Markov and σ-Faithfulness

Let $\mathcal{G}_{\mathbf{O}} = \mathrm{marg}_{\mathbf{L}}(\mathcal{D})$ denote the latent projection of the underlying DG onto the observed variables. Following Mooij & Claassen (2020), let $\mathrm{IM}_{\sigma}(\mathcal{G}_{\mathbf{O}})$ denote the independence model entailed by σ-separation in $\mathcal{G}_{\mathbf{O}}$, and let $\mathrm{IM}(P_{\mathbf{O}})$ denote the set of conditional independence relations in the observed distribution.

Definition 19 (σ-Markov condition (Mooij & Claassen, 2020))**.** *The observed distribution $P_{\mathbf{O}}$ is* σ-Markov *with respect to $\mathcal{D}$ if $\mathrm{IM}_{\sigma}(\mathcal{G}_{\mathbf{O}}) \subseteq \mathrm{IM}(P_{\mathbf{O}})$; that is, every σ-separation among the observed variables implies the corresponding conditional independence.*

Definition 20 (σ-faithfulness (Mooij & Claassen, 2020))**.** *The observed distribution $P_{\mathbf{O}}$ is* σ-faithful *to $\mathcal{D}$ if $\mathrm{IM}(P_{\mathbf{O}}) \subseteq \mathrm{IM}_{\sigma}(\mathcal{G}_{\mathbf{O}})$; that is, every conditional independence among the observed variables is entailed by σ-separation.*

Under the σ-Markov and σ-faithfulness assumptions, $\mathrm{IM}(P_{\mathbf{O}}) = \mathrm{IM}_{\sigma}(\mathcal{G}_{\mathbf{O}}) = \mathrm{IM}_{m}(\mathcal{H})$, where $\mathrm{IM}_{m}(\mathcal{H})$ denotes the independence model entailed by m-separation in the corresponding observed-variable σ-MAG $\mathcal{H}$.

B Proofs of the Main Results

B.1 Proof of Lemma 1

Lemma 4 (Theorem 4.6 in Forré & Mooij (2017))**.** *Let $\mathcal{G}$ be a DMG over $\mathbf{V}$, let $\mathbf{W} \subseteq \mathbf{V}$, and let*

$$\mathcal{G}' := \mathrm{marg}_{\mathbf{W}}(\mathcal{G})$$

be the marginalization of $\mathcal{G}$ over $\mathbf{W}$, with vertex set $\mathbf{V} \setminus \mathbf{W}$. Then, for any pairwise disjoint sets $\mathbf{A}, \mathbf{B}, \mathbf{C} \subseteq \mathbf{V} \setminus \mathbf{W}$,

$$\mathbf{A} \perp\!\!\!\perp_{\sigma} \mathbf{B} \mid \mathbf{C} \text{ in } \mathcal{G} \quad \Longleftrightarrow \quad \mathbf{A} \perp\!\!\!\perp_{\sigma} \mathbf{B} \mid \mathbf{C} \text{ in } \mathcal{G}'.$$

Lemma 5 (Yao & Mooij (2025))**.** *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DMG with*

$$\mathbf{V} = \mathbf{O} \dot{\cup} \mathbf{S},$$

and let $\mathcal{H}$ be a σ-MAG over $\mathbf{O}$ that represents $\mathcal{G}$ given $\mathbf{S}$. For any distinct $X, Y \in \mathbf{O}$, X and Y are adjacent in $\mathcal{H}$ if and only if $\mathcal{G}$ contains a σ-inducing path between X and Y given $\mathbf{S}$.

Lemma 6 (Proposition 5 in Yao & Mooij (2025))**.** *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DMG with*

$$\mathbf{V} = \mathbf{O} \dot{\cup} \mathbf{S},$$

and let $X, Y \in \mathbf{O}$ be distinct. Then $\mathcal{G}$ contains a σ-inducing path between X and Y given $\mathbf{S}$ if and only if

$$X \not\perp\!\!\!\perp_{\sigma} Y \mid \mathrm{An}^{+}(\{X, Y\} \cup \mathbf{S}, \mathcal{G}) \setminus \{X, Y\}$$

in $\mathcal{G}$.

Lemma 7 (Adjacency characterization in a σ-MAG)**.** *Let $\mathcal{D}$ be a DG, possibly containing directed cycles, with vertex set*

$$\mathbf{V} = \mathbf{O} \dot{\cup} \mathbf{L} \dot{\cup} \mathbf{S},$$

where $\mathbf{O}$, $\mathbf{L}$, and $\mathbf{S}$ are the observed, latent, and selection vertex sets, respectively. Let $\mathcal{H}$ the σ-MAG over $\mathbf{O}$ that represents $\mathcal{D}$. For distinct $X, Y \in \mathbf{O}$, the following are equivalent:

(i) There is an edge between X and Y in $\mathcal{H}$.
(ii) The vertices in $\mathrm{An}^+(\{X,Y\} \cup \mathbf{S}, \mathcal{D})$ that are not in $\mathbf{L} \cup \{X,Y\}$ do not σ-separate X and Y in $\mathcal{D}$, i.e.,

$$X \not\perp\!\!\!\perp_\sigma Y \mid \mathrm{An}^+(\{X,Y\} \cup \mathbf{S}, \mathcal{D}) \setminus \big(\mathbf{L} \cup \{X,Y\}\big) \text{ in } \mathcal{D}.$$

Proof of Lemma 7. Let

$$\mathcal{G} := \mathrm{marg}_{\mathbf{L}}(\mathcal{D})$$

be the latent projection of $\mathcal{D}$ onto $\mathbf{V} \setminus \mathbf{L}$. Thus, the vertex set of $\mathcal{G}$ is $\mathbf{O} \dot{\cup} \mathbf{S}$.

We first relate the relevant ancestral sets in $\mathcal{D}$ and $\mathcal{G}$. By the definition of latent projection, directed paths among the retained vertices are preserved after marginalizing $\mathbf{L}$: every directed path in $\mathcal{D}$ whose internal marginalized vertices lie in $\mathbf{L}$ induces a directed path in $\mathcal{G}$, and conversely every directed edge in $\mathcal{G}$ can be expanded into such a directed path in $\mathcal{D}$. Consequently,

$$\mathrm{An}^+(\{X,Y\} \cup \mathbf{S}, \mathcal{G}) = \mathrm{An}^+(\{X,Y\} \cup \mathbf{S}, \mathcal{D}) \setminus \mathbf{L}. \tag{2}$$

Hence,

$$\begin{aligned} &\mathrm{An}^+(\{X,Y\} \cup \mathbf{S}, \mathcal{G}) \setminus \{X,Y\} \\ &\quad = \mathrm{An}^+(\{X,Y\} \cup \mathbf{S}, \mathcal{D}) \setminus \big(\mathbf{L} \cup \{X,Y\}\big). \end{aligned} \tag{3}$$

By Lemmas 5 and 6, X and Y are adjacent in $\mathcal{H}$ if and only if

$$X \not\perp\!\!\!\perp_\sigma Y \mid \mathrm{An}^+(\{X,Y\} \cup \mathbf{S}, \mathcal{G}) \setminus \{X,Y\} \tag{4}$$

in $\mathcal{G}$.

Let

$$\mathbf{C} := \mathrm{An}^+(\{X,Y\} \cup \mathbf{S}, \mathcal{D}) \setminus \big(\mathbf{L} \cup \{X,Y\}\big).$$

By eq. (3),

$$\mathbf{C} = \mathrm{An}^+(\{X,Y\} \cup \mathbf{S}, \mathcal{G}) \setminus \{X,Y\}.$$

Moreover,

$$\{X,Y\} \cup \mathbf{C} \subseteq \mathbf{V} \setminus \mathbf{L}.$$

Therefore, applying Lemma 4 to the marginalization

$$\mathcal{G} = \mathrm{marg}_{\mathbf{L}}(\mathcal{D})$$

gives

$$X \not\perp\!\!\!\perp_\sigma Y \mid \mathbf{C} \text{ in } \mathcal{G} \quad \Longleftrightarrow \quad X \not\perp\!\!\!\perp_\sigma Y \mid \mathbf{C} \text{ in } \mathcal{D}.$$

Combining this equivalence with eq. (4) yields X and Y are adjacent in $\mathcal{H}$ if and only if

$$X \not\perp\!\!\!\perp_\sigma Y \mid \mathrm{An}^+(\{X,Y\} \cup \mathbf{S}, \mathcal{D}) \setminus \big(\mathbf{L} \cup \{X,Y\}\big)$$

in $\mathcal{D}$, which proves the result. □

Proof of Lemma 1. Setting $\mathbf{S} = \emptyset$ in Lemma 7 immediately yields Lemma 1. □

B.2 Proof of Lemma 2

Proof. Let

$$\mathbf{B} = \{Y \in \mathbf{O} \setminus \{X\} : \text{a collider path connects } X \text{ and } Y \text{ in } \mathcal{H}\},$$

where paths of length one are included.

Separation. Fix $Y \in \mathbf{O} \setminus (\mathbf{B} \cup \{X\})$ and consider any path π from X to Y. Since $Y \notin \mathbf{B}$, π is not a collider path and therefore has an internal non-collider. Let V be the first such vertex encountered from X. Every internal vertex on the prefix from X to V is a collider, so this prefix is a collider path and $V \in \mathbf{B}$. Thus, V blocks π given $\mathbf{B}$. As this holds for every such path and every Y outside $\mathbf{B} \cup \{X\}$, we obtain

$$X \perp\!\!\!\perp \mathbf{O} \setminus (\mathbf{B} \cup \{X\}) \mid \mathbf{B}.$$

Minimality and uniqueness. Let $\mathbf{C} \subseteq \mathbf{O} \setminus \{X\}$ be any set satisfying the blanket separation property. Suppose that some $Y \in \mathbf{B}$ is not in $\mathbf{C}$, and choose a collider path from X to Y. Let W be the first vertex after X on this path that lies outside $\mathbf{C}$. Every internal vertex on the prefix from X to W is a collider in $\mathbf{C}$, and is therefore activated by conditioning on $\mathbf{C}$. Moreover, a collider path has no internal non-collider and cannot contain either of the forbidden arrowhead–undirected configurations in Definition 4; for a path of length one, these conditions are vacuous. Hence, the prefix is m-open given $\mathbf{C}$, implying $X \not\perp\!\!\!\perp W \mid \mathbf{C}$. Since $W \notin \mathbf{C} \cup \{X\}$, this contradicts the blanket separation property. Thus, every blanket separating set contains $\mathbf{B}$. Together with the separation property of $\mathbf{B}$, this proves that $\mathbf{B}$ is the unique minimal Markov blanket of X.

Full-conditioning characterization. For distinct $X, Y \in \mathbf{O}$, let $\mathbf{Z} = \mathbf{O} \setminus \{X, Y\}$. If $Y \in \mathbf{B}$, a collider path from X to Y has all its internal vertices in $\mathbf{Z}$ and contains no forbidden arrowhead–undirected configuration. It is therefore m-open given $\mathbf{Z}$. Conversely, any m-open path from X to Y given $\mathbf{Z}$ must have only colliders as internal vertices, since every internal non-collider would belong to $\mathbf{Z}$ and block the path. Such a path is a collider path, so $Y \in \mathbf{B}$. Translating m-connection into conditional dependence gives

$$Y \in \mathrm{MB}(X, \mathcal{H}) \quad \Longleftrightarrow \quad X \not\perp\!\!\!\perp Y \mid \mathbf{O} \setminus \{X, Y\} \quad \text{under } P_{\mathbf{O}},$$

which proves eq. (1). □

B.3 Proof of Lemma 3

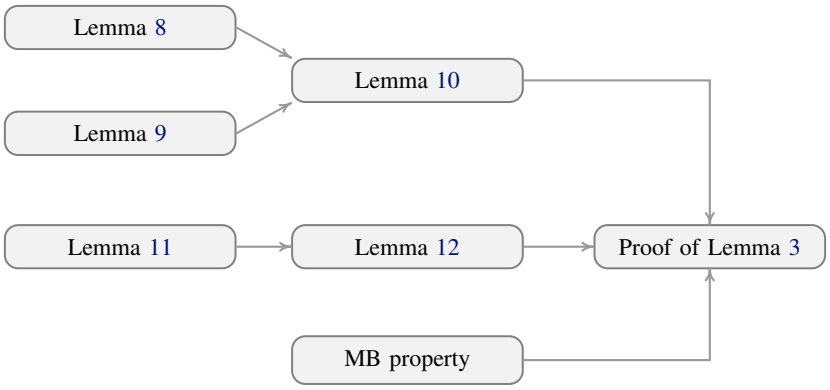


Figure 4: Proof structure of Lemma 3. Arrows indicate dependencies: each source is used in the proof of its target.

Lemma 8. *Let $\mathcal{H}$ be a σ-MAG over $\mathbf{O}$, let $\mathbf{S} \subseteq \mathbf{O}$, and let $X, Y \in \mathbf{S}$ be distinct vertices. Define $\mathbf{S}' = \mathrm{Ant}(\{X, Y\}, \mathcal{H}) \cap \mathbf{S}$. If π is an m-open path between X and Y given $\mathbf{S}'$, then every non-endpoint vertex on π belongs to $\mathrm{Ant}(\{X, Y\}, \mathcal{H})$.*

Proof of Lemma 8. Let V_k be an arbitrary non-endpoint vertex on $\pi = \langle X, V_1, \ldots, V_n, Y \rangle$.

Suppose first that V_k is a collider on π. Since π is m-open given $\mathbf{S}'$, we have $V_k \in \mathrm{An}^+(\mathbf{S}', \mathcal{H})$. Since $\mathbf{S}' \subseteq \mathrm{Ant}(\{X, Y\}, \mathcal{H})$, there exists an anterior path from V_k to either X or Y. Hence, $V_k \in \mathrm{Ant}(\{X, Y\}, \mathcal{H})$.

Now suppose that V_k is a non-collider on π. Since V_k is not a collider, at least one of its two incident edges on π has a tail at V_k. We follow π in such a direction. By symmetry, we may assume that this direction is toward V_{k+1}, so the first edge is either $V_k \to V_{k+1}$ or $V_k - V_{k+1}$. The case where the tail is on the edge toward V_{k-1} is symmetric.

At any subsequent non-collider along this traversal, if the preceding edge has an arrowhead at the current vertex, the next edge must have a tail there; otherwise the current vertex would be a collider. Furthermore, the next edge cannot be undirected, since this would form an m-blocked subpath of the form $A* \to B - C$. If the preceding edge is undirected, the next edge cannot have an arrowhead at the current vertex, since this would form the m-blocked subpath $A - B \leftarrow\!* C$. Therefore, the traversal continues along edges that are out of each current vertex.

Therefore, the path can be followed anteriorly until it reaches either X, Y, or a collider V_c. If it reaches X or Y, then $V_k \in \mathrm{Ant}(\{X, Y\}, \mathcal{H})$. If it reaches a collider V_c, the first part of the proof gives $V_c \in \mathrm{Ant}(\{X, Y\}, \mathcal{H})$. By transitivity of the anterior relation, we again obtain $V_k \in \mathrm{Ant}(\{X, Y\}, \mathcal{H})$.

Since V_k was arbitrary, every non-endpoint vertex on π belongs to $\mathrm{Ant}(\{X, Y\}, \mathcal{H})$. □

Lemma 9 (Fundamental Property of σ-MAGs, Lemma 1 in (Yao & Mooij, 2025))**.** *Let $\mathcal{H}$ be a σ-MAG. If $\mathcal{H}$ contains a triple of the form $A* \to B - C$, then the edge between A and C is of the same type as the edge between A and B, and the neighborhoods of B and C are both complete.*

Lemma 10. *Let $\mathbf{S} \subseteq \mathbf{O}$, and let $X, Y \in \mathbf{S}$ be distinct vertices in a σ-MAG $\mathcal{H}$. Then X and Y are m-separated by some $\mathbf{Z} \subseteq \mathbf{S} \setminus \{X, Y\}$ if and only if they are m-separated by $\mathrm{Ant}(\{X, Y\}, \mathcal{H}) \cap \mathbf{S}$.*

Proof of Lemma 10. Let $\mathbf{S}' = \mathrm{Ant}(\{X, Y\}, \mathcal{H}) \cap \mathbf{S}$. We prove the two directions separately.

($\Rightarrow$) We prove the contrapositive. Suppose that X and Y are not m-separated by $\mathbf{S}'$. Then there exists an m-open path π between X and Y given $\mathbf{S}'$. By Lemma 8, every non-endpoint vertex on π belongs to $\mathrm{Ant}(\{X, Y\}, \mathcal{H})$. Since every non-collider on π lies outside $\mathbf{S}'$, and $\mathbf{S}' = \mathrm{Ant}(\{X, Y\}, \mathcal{H}) \cap \mathbf{S}$, every non-collider on π must lie outside $\mathbf{S}$.

Now let $\mathbf{Z} \subseteq \mathbf{S} \setminus \{X, Y\}$ be arbitrary. We show that X and Y are not m-separated by $\mathbf{Z}$. Since every non-collider on π lies outside $\mathbf{S}$, it also lies outside $\mathbf{Z}$. Moreover, because π is m-open given $\mathbf{S}'$, it contains no blocking subpath of the form $A* \to B - C$ or $A - B \leftarrow\!* C$. Thus, the only reason that π may fail to be m-open given $\mathbf{Z}$ is that some collider may not belong to $\mathrm{An}^+(\mathbf{Z}, \mathcal{H})$.

Consider all walks between X and Y such that every non-endpoint vertex belongs to $\mathrm{Ant}(\{X, Y\}, \mathcal{H})$, every non-collider lies outside $\mathbf{Z}$, and no blocking configuration involving an undirected edge occurs. The path π belongs to this class. Choose such a walk with the minimum number of colliders outside $\mathrm{An}^+(\mathbf{Z}, \mathcal{H})$. Subject to this, choose one that minimizes the total length of the initial undirected segments of shortest anterior paths from these colliders to $\{X, Y\}$.

Suppose, for contradiction, that the chosen walk contains a collider $C \notin \mathrm{An}^+(\mathbf{Z}, \mathcal{H})$. Since $C \in \mathrm{Ant}(\{X, Y\}, \mathcal{H})$, there exists an anterior path from C to either X or Y. Let ρ be a shortest such anterior path.

First, suppose that ρ starts with an undirected edge $C - D$. Let A and B be the vertices adjacent to C on the chosen walk, so that it contains $A* \to C \leftarrow\!* B$. By Lemma 9, A and D, as well as B and D, are adjacent, with the corresponding edge types preserved. Hence, $A* \to D \leftarrow\!* B$ can replace $A* \to C \leftarrow\!* B$.

If $D \in \mathrm{An}^+(\mathbf{Z}, \mathcal{H})$, this replacement reduces the number of colliders outside $\mathrm{An}^+(\mathbf{Z}, \mathcal{H})$. Otherwise, that number is unchanged, but the initial undirected segment of the corresponding anterior path is shortened. Either case contradicts the choice of the walk.

Therefore, ρ cannot start with an undirected edge and must start with a directed edge. By Lemma 9, C is then an ancestor of the corresponding endpoint, say Y. Hence, there exists a directed path from C to Y. Since $C \notin \mathrm{An}^+(\mathbf{Z}, \mathcal{H})$, no vertex on this directed path belongs to $\mathbf{Z}$; otherwise, C would be an ancestor of a vertex in $\mathbf{Z}$.

Replace the subwalk of the chosen walk from C to Y by this directed path. Then C becomes a non-collider, no new collider is introduced along the directed path, and all its non-endpoint vertices lie outside $\mathbf{Z}$. Moreover, no blocking configuration involving an undirected edge is introduced. Therefore, the number of colliders outside $\mathrm{An}^+(\mathbf{Z}, \mathcal{H})$ is strictly reduced, again contradicting the choice of the walk.

Hence, the chosen walk contains no collider outside $\mathrm{An}^+(\mathbf{Z}, \mathcal{H})$. Every non-collider lies outside $\mathbf{Z}$, and no blocking configuration involving an undirected edge occurs. Therefore, the walk is m-open given $\mathbf{Z}$. By Proposition 1 of Yao & Mooij (2025), there exists an m-open path between X and Y given $\mathbf{Z}$.

Since $\mathbf{Z} \subseteq \mathbf{S} \setminus \{X, Y\}$ was arbitrary, no subset of $\mathbf{S} \setminus \{X, Y\}$ m-separates X and Y. This proves the contrapositive.

($\Leftarrow$) The reverse direction is immediate, since $\mathbf{S}' \subseteq \mathbf{S} \setminus \{X, Y\}$. □

Lemma 11. *Let π be a path that is m-open X and Y given $\mathbf{Z}$ in a σ-MAG $\mathcal{H}$. Then every vertex on π belongs to*

$$\mathrm{Ant}^+(\{X, Y\} \cup \mathbf{Z}, \mathcal{H}).$$

Proof of Lemma 11. Let $\pi = \langle X, V_1, \ldots, V_n, Y \rangle$. The result is immediate for the endpoints X and Y. Let V_k be an arbitrary non-endpoint vertex on π.

Suppose first that V_k is a collider on π. Since π is m-open given $\mathbf{Z}$, we have $V_k \in \mathrm{An}^+(\mathbf{Z}, \mathcal{H})$. Since every directed path is also an anterior path, it follows that $V_k \in \mathrm{Ant}^+(\mathbf{Z}, \mathcal{H})$, and hence $V_k \in \mathrm{Ant}^+(\{X, Y\} \cup \mathbf{Z}, \mathcal{H})$.

Now suppose that V_k is a non-collider on π. Since V_k is not a collider, at least one of the two edges incident to V_k on π has a tail at V_k. We follow π in such a direction. By symmetry, suppose this direction is toward V_{k+1}.

At any subsequent non-collider along this traversal, if the preceding edge has an arrowhead at the current vertex, the next edge must have a tail there; otherwise, the current vertex would be a collider. Moreover, the next edge cannot be undirected, since this would form an m-blocked subpath of the form $A* \to B - C$. If the preceding edge is undirected, the next edge cannot have an arrowhead at the current vertex, since this would form the m-blocked subpath $A - B \leftarrow\!* C$. Therefore, the traversal continues along an anterior subpath until it reaches either X, Y, or a collider V_c.

If it reaches X or Y, then $V_k \in \mathrm{Ant}(\{X, Y\}, \mathcal{H})$. If it reaches a collider V_c, the first part of the proof gives $V_c \in \mathrm{Ant}^+(\mathbf{Z}, \mathcal{H})$. By transitivity of the anterior relation, $V_k \in \mathrm{Ant}^+(\mathbf{Z}, \mathcal{H})$.

Thus, in either case, $V_k \in \mathrm{Ant}^+(\{X, Y\} \cup \mathbf{Z}, \mathcal{H})$. Since V_k was arbitrary, every vertex on π belongs to $\mathrm{Ant}^+(\{X, Y\} \cup \mathbf{Z}, \mathcal{H})$. □

Lemma 12. *Let π be a path between distinct vertices X and Y in a σ-MAG $\mathcal{H}$. If π is not contained completely in $\mathrm{Ant}^+(\{X, Y\}, \mathcal{H})$, then π is m-blocked by every $\mathbf{Z} \subseteq \mathrm{Ant}(\{X, Y\}, \mathcal{H})$.*

Proof of Lemma 12. Let $\mathbf{Z} \subseteq \mathrm{Ant}(\{X, Y\}, \mathcal{H})$ be arbitrary. By transitivity of the anterior relation, $\mathrm{Ant}(\{X, Y\} \cup \mathbf{Z}, \mathcal{H}) = \mathrm{Ant}(\{X, Y\}, \mathcal{H})$. If π were m-connecting given $\mathbf{Z}$, Lemma 11 would imply that every vertex on π belongs to $\mathrm{Ant}^+(\{X, Y\}, \mathcal{H})$, contradicting the assumption. Hence π is m-blocked by $\mathbf{Z}$. □

Proof of Lemma 3. ($\Rightarrow$) Suppose that X and Y are m-separated by some subset of $\mathbf{O} \setminus \{X, Y\}$.

If $Y \notin \mathrm{MB}^+(X)$, the result follows directly from the Markov blanket property, since $\mathrm{MB}(X)$ m-separates X and Y, and $\mathrm{MB}(X) \subseteq \mathrm{MB}^+(X) \setminus \{Y\}$.

Now suppose that $Y \in \mathrm{MB}(X)$. By Lemma 10, X and Y are m-separated by $\mathrm{Ant}(\{X, Y\}, \mathcal{H})$. Define $\mathbf{S} = \mathrm{Ant}(\{X, Y\}, \mathcal{H}) \cap \mathrm{MB}(X)$ and $\mathbf{B} = \mathbf{O} \setminus \mathrm{MB}^+(X)$. We show that $\mathbf{S}$ m-separates X and Y. Consider an arbitrary path π between X and Y.

If π is not completely contained in $\mathrm{Ant}^+(\{X, Y\}, \mathcal{H})$, then by Lemma 12, π is m-blocked by $\mathbf{S}$.

Now suppose that every vertex on π belongs to $\mathrm{Ant}^+(\{X, Y\}, \mathcal{H})$. Assume, for contradiction, that π is m-open given $\mathbf{S}$. Since $\mathrm{Ant}(\{X, Y\}, \mathcal{H})$ m-separates X and Y, whereas π is m-open given $\mathbf{S}$. Hence, π contains at least one vertex in $\mathrm{Ant}(\{X, Y\}, \mathcal{H}) \setminus \mathbf{S}$.

Let V be the first such vertex encountered when traversing π from X. Since $\mathbf{S} = \mathrm{Ant}(\{X, Y\}, \mathcal{H}) \cap \mathrm{MB}(X)$ and $V \neq Y$, we have $V \in \mathbf{B}$. Moreover, every non-endpoint vertex on the subpath $\pi(X, V)$ belongs to $\mathbf{S}$ by the choice of V. Hence, $\pi(X, V)$ is also m-open given $\mathrm{MB}(X)$: $\mathbf{S} \subseteq \mathrm{MB}(X)$, and the blocking configurations involving undirected edges are unchanged.

Thus, X and V are m-connected given $\mathrm{MB}(X)$. However, $V \in \mathbf{B} = \mathbf{O} \setminus \mathrm{MB}^+(X)$, while the Markov blanket property implies that X is m-separated from every vertex in $\mathbf{B}$ given $\mathrm{MB}(X)$. This is a contradiction. Therefore, every path between X and Y is m-blocked by $\mathbf{S}$. Hence, $\mathbf{S} \subseteq \mathrm{MB}(X) \setminus \{Y\}$ m-separates X and Y.

($\Leftarrow$) The reverse direction is immediate, since $\mathrm{MB}^+(X) \subseteq \mathbf{O}$. □

B.4 Proof of Theorem 2

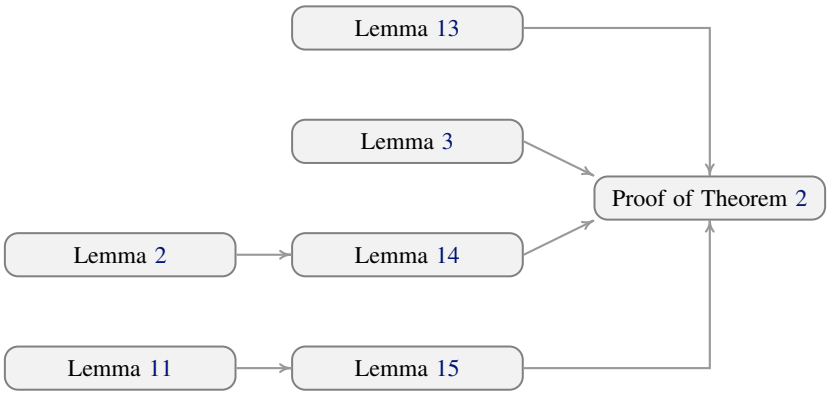


Figure 5: Proof structure of Theorem 2. Arrows indicate dependencies: each source is used in the proof of its target.

Definition 21 (Arrow-collider path)**.** *In a σ-MAG $\mathcal{H}$, a path $\pi = \langle V_0, \ldots, V_k \rangle$ is an* arrow-collider path *from V_0 to V_k if every non-endpoint vertex is a collider on π and the first edge has an arrowhead at V_0. We denote by $\mathrm{ArrColl}(V_0, \mathcal{H})$ the set of vertices reachable from V_0 through an arrow-collider path.*

Lemma 13. *Let $\mathcal{D}$ be the underlying directed graph over $\mathbf{V} = \mathbf{O} \cup \mathbf{L}$, and let $\mathcal{H}$ be the corresponding σ-MAG over $\mathbf{O}$. Let $\mathcal{L}_X$ be the local σ-PAG learned from the marginal distribution over $\mathrm{MB}^+(X)$. For any adjacent vertices $A, B \in \mathrm{MB}^+(X)$, every non-circle endpoint mark in $\mathcal{L}_X$ has the same ancestral interpretation with respect to $\mathcal{D}$. In particular,*

$$A \leftarrow\!* B \quad \Longrightarrow \quad A \notin \mathrm{An}(B, \mathcal{D}),$$

whereas

$$A \twoheadrightarrow B \quad \Longrightarrow \quad A \in \mathrm{An}(B, \mathcal{D}).$$

Consequently, if A and B are adjacent in the global σ-PAG $\mathcal{P}$, every non-circle endpoint mark identified in $\mathcal{L}_X$ is also present at the corresponding endpoint in $\mathcal{P}$.

Proof of Lemma 13. Let $\mathbf{W} = \mathrm{MB}^+(X)$, and let $\mathcal{H}_{\mathbf{W}}$ denote the σ-MAG over $\mathbf{W}$ obtained by marginalizing all variables in $\mathbf{V} \setminus \mathbf{W}$ from the underlying DG $\mathcal{D}$. Thus, $\mathcal{H}_{\mathbf{W}}$ represents the marginal distribution over $\mathbf{W}$.

Suppose first that $\mathcal{L}_X$ contains $A \leftarrow\!* B$. By the definition of a σ-PAG, the endpoint at A is an arrowhead in every σ-MAG in the local Markov equivalence class. In particular, $\mathcal{H}_{\mathbf{W}}$ contains an arrowhead at A. By the endpoint semantics of σ-MAGs, this implies that A is not an ancestor of B in the underlying DG $\mathcal{D}$. Hence, $A \notin \mathrm{An}(B, \mathcal{D})$.

Similarly, if $\mathcal{L}_X$ contains $A \multimap\!\!* B$, then the endpoint at A is a tail in every σ-MAG in the local equivalence class. The endpoint semantics of σ-MAGs therefore imply $A \in \mathrm{An}(B, \mathcal{D})$.

It remains to relate these local endpoint marks to the global σ-PAG $\mathcal{P}$. Consider any DG that is σ-Markov equivalent to $\mathcal{D}$. Its marginal model over $\mathbf{W}$ belongs to the same local Markov equivalence class represented by $\mathcal{L}_X$. Therefore, an arrowhead (respectively, a tail) identified in $\mathcal{L}_X$ implies the corresponding non-ancestral (respectively, ancestral) relation in every globally equivalent DG.

Hence, whenever A and B are adjacent in the global σ-PAG $\mathcal{P}$, the same invariant endpoint mark must appear at the corresponding endpoint in $\mathcal{P}$. □

Lemma 14. *Suppose that a σ-MAG $\mathcal{H}$ contains an arrow-collider path from X to A that is into A. Then*

$$\mathrm{ArrColl}(A, \mathcal{H}) \subseteq \mathrm{MB}(X).$$

Proof of Lemma 14. Let $V \in \mathrm{ArrColl}(A, \mathcal{H})$. Concatenating the arrow-collider path from X to A with an arrow-collider path from A to V yields a path from X to V whose non-endpoint vertices are all colliders. By the Markov blanket characterization (Lemma 2), implies $V \in \mathrm{MB}(X)$. □

Lemma 15. *Let $\mathcal{H}$ be a σ-MAG, and let A and B be non-adjacent vertices such that $A \notin \mathrm{Ant}(B, \mathcal{H})$. Define*

$$\mathbf{Z} = \big(\mathrm{ArrColl}(A, \mathcal{H}) \cap \mathrm{Ant}(\{A, B\}, \mathcal{H})\big) \setminus \{A, B\}.$$

Then $\mathbf{Z}$ m-separates A and B in $\mathcal{H}$.

Proof of Lemma 15. Suppose, for contradiction, that $\mathbf{Z}$ does not m-separate A and B. Then there exists an m-open path

$$\pi = \langle A = V_0, V_1, \ldots, V_n = B \rangle$$

given $\mathbf{Z}$.

By Lemma 11, every vertex on π belongs to

$$\mathrm{Ant}^+(\{A, B\} \cup \mathbf{Z}, \mathcal{H}).$$

Since $\mathbf{Z} \subseteq \mathrm{Ant}(\{A, B\}, \mathcal{H})$ and the anterior relation is transitive, every non-endpoint vertex on π belongs to $\mathrm{Ant}(\{A, B\}, \mathcal{H})$.

We first show that the first edge on π must be into A. Suppose otherwise that the first edge is out of A. Starting from A, follow π toward B. If no collider is encountered, this gives an anterior path from A to B, contradicting $A \notin \mathrm{Ant}(B, \mathcal{H})$. Hence, let C be the first collider encountered. The subpath from A to C is an anterior path. Moreover, its last edge is directed into C, so this anterior path contains at least one directed edge and hence implies $A \in \mathrm{Ant}(C, \mathcal{H})$. By Definition 3,

$$C \notin \mathrm{Ant}(A, \mathcal{H}).$$

Since C is a collider on an m-open path given $\mathbf{Z}$, we have $C \in \mathrm{An}^+(\mathbf{Z}, \mathcal{H})$. By the definition of $\mathbf{Z}$, every vertex in $\mathbf{Z}$ is anterior to either A or B. Hence,

$$C \in \mathrm{Ant}(\{A, B\}, \mathcal{H}).$$

Since $C \notin \mathrm{Ant}(A, \mathcal{H})$, it follows that $C \in \mathrm{Ant}(B, \mathcal{H})$. Together with $A \in \mathrm{Ant}(C, \mathcal{H})$, this gives $A \in \mathrm{Ant}(B, \mathcal{H})$, again a contradiction. Therefore, the first edge on π must be into A.

We next show that every non-endpoint vertex on π is a collider. If the first edge were $A \leftarrow V_1$, then V_1 would be a non-collider on π. Moreover, the edge $A \leftarrow V_1$ itself is an arrow-collider path from A to V_1, so $V_1 \in \mathrm{ArrColl}(A, \mathcal{H})$. As shown above, $V_1 \in \mathrm{Ant}(\{A, B\}, \mathcal{H})$, and therefore $V_1 \in \mathbf{Z}$. This contradicts the m-openness of π. Hence, the first edge must be $A \leftrightarrow V_1$.

Now suppose that π contains a non-endpoint non-collider, and let V_k be the first such vertex encountered from A. Then every vertex preceding V_k on π is a collider. Therefore, the subpath from A to V_k is an arrow-collider path, and hence

$$V_k \in \mathrm{ArrColl}(A, \mathcal{H}).$$

Since also $V_k \in \mathrm{Ant}(\{A, B\}, \mathcal{H})$, we obtain $V_k \in \mathbf{Z}$. But V_k is a non-collider on π, contradicting that π is m-open given $\mathbf{Z}$.

Thus, every non-endpoint vertex on π is a collider. Consequently, π is an arrow-collider path into A, and, as established above, every non-endpoint vertex on π belongs to $\mathrm{Ant}(\{A, B\}, \mathcal{H})$. Hence, there exists an inducing walk between A and B. By Proposition 2 of Yao & Mooij (2025), this contradicts the maximality of $\mathcal{H}$, since A and B are non-adjacent.

Therefore, $\mathbf{Z}$ m-separates A and B in $\mathcal{H}$. □

Proof of Theorem 2. We prove the two directions separately.

($\Rightarrow$) Suppose that the uncovered collider path

$$\pi = \langle V_0(= X), V_1, \ldots, V_k, V_{k+1}(= Y)\rangle$$

with

$$X *\!\!\rightarrow V_1 \leftrightarrow \cdots \leftrightarrow V_k \leftarrow\!\!* Y$$

is present in $\mathcal{L}_X$.

We show sequentially that the same adjacencies and endpoint marks are present in the global σ-PAG $\mathcal{P}$.

First, $V_0 = X$ and V_1 are adjacent in $\mathcal{L}_X$. By Lemma 3, they are also adjacent in $\mathcal{P}$. Since $\mathcal{L}_X$ has an arrowhead at V_1, Lemma 13 implies that the global edge also has an arrowhead at V_1. Hence, $X *\!\!\rightarrow V_1$ is present in $\mathcal{P}$.

Now suppose inductively that

$$X *\!\!\rightarrow V_1 \leftrightarrow \cdots \leftrightarrow V_j$$

has been established in $\mathcal{P}$, where $1 \le j \le k$. We show that V_j and V_{j+1} are also adjacent globally.

Since $\mathcal{L}_X$ contains $V_j \leftarrow\!\!* V_{j+1}$, Lemma 13 implies that V_j is not an ancestor of V_{j+1} in the underlying global DG. Consequently, $V_j \notin \mathrm{Ant}(V_{j+1}, \mathcal{H})$.

Suppose, for contradiction, that V_j and V_{j+1} are non-adjacent in the global σ-MAG. By Lemma 15, they are then m-separated by

$$\mathbf{Z} = \big(\mathrm{ArrColl}(V_j, \mathcal{H}) \cap \mathrm{Ant}(\{V_j, V_{j+1}\}, \mathcal{H})\big) \setminus \{V_j, V_{j+1}\}.$$

By the induction hypothesis, there is an arrow-collider path from X to V_j that is into V_j. Hence, by Lemma 14, $\mathrm{ArrColl}(V_j, \mathcal{H}) \subseteq \mathrm{MB}(X)$, and therefore $\mathbf{Z} \subseteq \mathrm{MB}(X)$.

Thus, V_j and V_{j+1} can be m-separated using only variables in $\mathrm{MB}^+(X)$, contradicting their adjacency in $\mathcal{L}_X$. Hence, V_j and V_{j+1} must be adjacent globally.

Moreover, the arrowhead at V_j in the local edge is globally sound by Lemma 13. Repeating this argument along the path establishes all consecutive adjacencies and their corresponding arrowheads.

It remains to verify that the path is uncovered globally. For every $0 \le j \le k-1$, the vertices V_j and V_{j+2} are non-adjacent in $\mathcal{L}_X$. Hence, there exists a separating set contained in $\mathrm{MB}^+(X)$. The same conditional independence holds in the global distribution, so V_j and V_{j+2} cannot be adjacent in the global σ-MAG. Therefore, every consecutive triple on π remains unshielded globally.

Consequently, the same uncovered collider path

$$X *\!\!\rightarrow V_1 \leftrightarrow \cdots \leftrightarrow V_k \leftarrow\!\!* Y$$

is present in $\mathcal{P}$.

($\Leftarrow$) Conversely, suppose that the uncovered collider path

$$\pi = \langle V_0(= X), V_1, \ldots, V_k, V_{k+1}(= Y)\rangle$$

is present in the global σ-PAG $\mathcal{P}$.

Every consecutive pair V_j, V_{j+1} is adjacent in the global σ-MAG. Hence, by maximality, no conditioning set in $\mathbf{O} \setminus \{V_j, V_{j+1}\}$ m-separates them. In particular, no subset of $\mathrm{MB}^+(X)$ separates them. Therefore, every consecutive pair on π remains adjacent in $\mathcal{L}_X$.

We next show that every consecutive triple remains unshielded and collider-oriented locally.

For the first triple $\langle X, V_1, V_2\rangle$, X and V_2 are non-adjacent globally. By Lemma 3, there exists a set $\mathbf{S}$ contained in $\mathrm{MB}(X)$ that m-separates X and V_2. Hence, X and V_2 are non-adjacent in $\mathcal{L}_X$. Furthermore, so V_1 does not belong to the corresponding separating set $\mathbf{S}$, otherwise, $\mathbf{S}$ cannot m-separates X and V_2. Therefore, $\langle X, V_1, V_2\rangle$ is identified as an unshielded collider in $\mathcal{L}_X$.

Now consider an interior triple $\langle V_{j-1}, V_j, V_{j+1}\rangle$. Since the path is uncovered globally, V_{j-1} and V_{j+1} are non-adjacent. At least one of these two vertices is not anterior to the other; denote it by A and the other by B. Applying Lemma 15 gives a separating set

$$\mathbf{Z} \subseteq \mathrm{ArrColl}(A, \mathcal{H}).$$

Since both V_{j-1} and V_{j+1} lie on the bidirected portion of the collider path, the prefix of π from X to A is an arrow-collider path that is into A. Hence, Lemma 14 gives $\mathbf{Z} \subseteq \mathrm{MB}(X)$. Thus, V_{j-1} and V_{j+1} are non-adjacent in $\mathcal{L}_X$, and their separating set $\mathbf{Z}$ does not contain V_j. Hence, $\langle V_{j-1}, V_j, V_{j+1}\rangle$ is identified as an unshielded collider in $\mathcal{L}_X$.

The last triple is handled analogously. If the last edge is bidirected, the same argument applies directly. If it is of the form $V_k \leftarrow Y$, then $V_{k-1} \notin \mathrm{Ant}(Y, \mathcal{H})$; otherwise, concatenating an anterior path from V_{k-1} to Y with $Y \rightarrow V_k$ would contradict the bidirected edge $V_{k-1} \leftrightarrow V_k$. Thus, Lemma 15 can again be applied with $A = V_{k-1}$.

Therefore, every consecutive triple on π appears as the same unshielded collider in $\mathcal{L}_X$. Hence,

$$X *\!\!\rightarrow V_1 \leftrightarrow \cdots \leftrightarrow V_k \leftarrow\!\!* Y$$

is identified in $\mathcal{L}_X$. □

B.5 Proofs of Theorem 3 and Corollary 1

Proof. For any adjacent pair $A, B \in \mathrm{MB}^+(X)$ in $\mathcal{L}_X$, if the edge between A and B is also present in $\mathcal{P}$, the claim on endpoint marks follows directly from Lemma 13. □

Proof. By Theorem 3, both edges incident to X and their arrowheads at X are present in $\mathcal{P}$. Since A and B are non-adjacent in $\mathcal{L}_X$, there exists a separating set contained in $\mathrm{MB}^+(X) \setminus \{A, B\}$. The same conditional independence holds in the global distribution, so A and B are also non-adjacent in $\mathcal{P}$. Thus, the collider remains unshielded. □

B.6 Proof of Theorem 4

Lemma 16 (Acyclic representation of the global σ-MAG)**.** *Let $\mathcal{D}$ be the underlying directed graph over $\mathbf{V} = \mathbf{O} \dot\cup \mathbf{L}$, and let $\mathcal{H}$ be its corresponding σ-MAG over $\mathbf{O}$. Then there exists a DAG $\mathcal{D}^{\mathrm{acy}}$ on the same vertex set and with the same observed–latent partition such that, let $\mathcal{M}^{\mathrm{acy}}$ denotes the MAG induced by $\mathcal{D}^{\mathrm{acy}}$ over $\mathbf{O}$, then*

$$\mathrm{IM}_m(\mathcal{H}) = \mathrm{IM}_m(\mathcal{M}^{\mathrm{acy}}).$$

Moreover, $\mathcal{M}^{\mathrm{acy}}$ contains no undirected edges.

Proof. Let $\mathcal{D}^{\mathrm{acy}}$ be a DAG σ-acyclification of $\mathcal{D}$ (Definition 17). By Lemma 22, $\mathrm{IM}_\sigma(\mathcal{D}) = \mathrm{IM}_d(\mathcal{D}^{\mathrm{acy}})$. Restricting both models to $\mathbf{O}$ gives the same conditional independence relations among the observed variables.

Let $\mathcal{M}^{\mathrm{acy}}$ be the MAG induced by $\mathcal{D}^{\mathrm{acy}}$ over $\mathbf{O}$. By the standard latent-projection property, $\mathrm{IM}_m(\mathcal{M}^{\mathrm{acy}})$ equals the d-separation model of $\mathcal{D}^{\mathrm{acy}}$ restricted to $\mathbf{O}$. Likewise, since $\mathcal{H}$ represents $\mathcal{D}$ over $\mathbf{O}$, its m-separation model equals the σ-separation model of $\mathcal{D}$ restricted to $\mathbf{O}$. Hence $\mathrm{IM}_m(\mathcal{H}) = \mathrm{IM}_m(\mathcal{M}^{\mathrm{acy}})$.

Finally, since $\mathcal{D}^{\mathrm{acy}}$ is a DAG and selection bias is absent, its induced MAG contains only directed and bidirected edges, and therefore has no undirected edges. □

Since $\mathcal{H}$ and $\mathcal{M}^{\mathrm{acy}}$ induce the same independence model, FCI receives exactly the same oracle CI information in the two representations and therefore returns the same partial graph. By Appendix D, this common output is the global σ-PAG $\mathcal{P}$ associated with $\mathcal{H}$.

Proof of Theorem 4. We first establish soundness. For each processed vertex X, Theorem 1 guarantees that the preserved adjacencies incident to X agree with those in the global σ-PAG $\mathcal{P}$, while Theorems 2 and 3 guarantee the soundness of the preserved endpoint marks and uncovered collider information. Moreover, all separating sets supplied to ORIENTPAG correspond to valid CI relations in $P_{\mathbf{O}}$. Since the orientation rules are sound, every newly oriented endpoint is also consistent with $\mathcal{P}$. Therefore, $\widehat{\mathcal{P}}$ is sound with respect to $\mathcal{P}$.

For completeness, let $\mathcal{M}^{\mathrm{acy}}$ be the MAG given by Lemma 16. By construction, $\mathcal{M}^{\mathrm{acy}}$ and the underlying σ-MAG $\mathcal{H}$ are m-Markov equivalent and therefore correspond to the same PAG. Moreover, $\mathcal{M}^{\mathrm{acy}}$ contains no undirected edges, so the problem reduces to the acyclic setting considered in (Xie et al., 2024; Li et al., 2026).

In particular, the MB-based sequential expansion together with the selective UPDATENEXT, which retains only unprocessed vertices having a potentially anterior path to T, preserves all target-specific information needed for global discovery. Hence, by Theorem 4 of (Li et al., 2026) and Theorem 3 of (Xie et al., 2024), LoCL with exactly this potentially-anterior update rule is complete; thus, the sequential procedure recovers the target-specific structure returned by global FCI for $\mathcal{M}^{\mathrm{acy}}$.

Finally, since $\mathcal{M}^{\mathrm{acy}}$ and $\mathcal{H}$ are m-Markov equivalent, FCI receives the same oracle CI model and therefore returns the same partial graph for both. By Appendix D, this graph is precisely the global σ-PAG $\mathcal{P}$. Thus, the output of LoCL is complete with respect to $\mathcal{P}$.

□

C ALGORITHMIC DETAILS

C.1 ORIENTATION RULES

Let $\mathbf{Sepsets}$ denote the collection of all conditional independence relations identified during the local learning process. A separator $\mathrm{Sepset}(X, Y)$ between two variables X and Y recorded in $\mathbf{Sepsets}$ implies that there is no edge between X and Y in the learned graph.

(Zhang, 2008) presented sound and complete orientation rules for learning a PAG. These rules are summarized as Algorithm 2.

Algorithm 2 ORIENTPAG

Input: A PAG $\mathcal{P}$ and a collection of separating sets $\mathbf{Sepsets}$.
Output: The oriented PAG $\mathcal{P}$.

Initialize $\mathcal{P}$ with the current skeleton and existing edge marks.
repeat
  Rule 1. If $X*\!\!\rightarrow Y \circ\!\!-\!\!* Z$, $\mathrm{Sepset}(X, Z) \in \mathbf{Sepsets}$, and $Y \in \mathrm{Sepset}(X, Z)$, then orient the triple as $X*\!\!\rightarrow Y \rightarrow Z$.
  Rule 2. If $X \rightarrow Y*\!\!\rightarrow Z$ or $X*\!\!\rightarrow Y \rightarrow Z$, and $X \,*\!\!-\!\!\circ Z$, then orient $X \,*\!\!-\!\!\circ Z$ as $X*\!\!\rightarrow Z$.
  Rule 3. If $X*\!\!\rightarrow Y \leftarrow\!\!* Z$, $X \,*\!\!-\!\!\circ W \circ\!\!-\!\!* Z$, $W \,*\!\!-\!\!\circ Y$, $\mathrm{Sepset}(X, Z) \in \mathbf{Sepsets}$, and $W \in \mathrm{Sepset}(X, Z)$, then orient $W \,*\!\!-\!\!\circ Y$ as $W*\!\!\rightarrow Y$.
  Rule 4.
  if there exists a discriminating path $\pi = \langle W, \ldots, X, Y, Z\rangle$ between W and Z for Y, $Y \circ\!\!-\!\!* Z$, and $\mathrm{Sepset}(W, Z) \in \mathbf{Sepsets}$ **then**
    if $Y \in \mathrm{Sepset}(W, Z)$ **then**
      Orient $Y \circ\!\!-\!\!* Z$ as $Y \rightarrow Z$.
    else
      Orient the triple $\langle X, Y, Z\rangle$ as $X \leftrightarrow Y \leftrightarrow Z$.
    end if
  end if
  Rule 8b. If $X \rightarrow Y \rightarrow Z$, and $X \circ\!\!\rightarrow Z$, then orient $X \circ\!\!\rightarrow Z$ as $X \rightarrow Z$.
  Rule 9. If $X \circ\!\!\rightarrow Z$, and $p = \langle X, Y, W, \ldots, Z\rangle$ is an uncovered potentially directed path from X to Z such that $\mathrm{Sepset}(Z, Y) \in \mathbf{Sepsets}$ and $X \in \mathrm{Sepset}(Z, Y)$, then orient $X \circ\!\!\rightarrow Z$ as $X \rightarrow Z$.
  Rule 10. Suppose $X \circ\!\!\rightarrow Z$, $Y \rightarrow Z \leftarrow W$, π_1 is an uncovered potentially directed path from X to Y, and π_2 is an uncovered potentially directed path from X to W. Let U be the vertex adjacent to X on π_1 (U could be Y) and let V be the vertex adjacent to X on π_2 (V could be W). If U and V are distinct, $\mathrm{Sepset}(U, V) \in \mathbf{Sepsets}$, and $X \in \mathrm{Sepset}(U, V)$, then orient $X \circ\!\!\rightarrow Z$ as $X \rightarrow Z$.
until no rule can further orient any edge mark in $\mathcal{P}$
return $\mathcal{P}$.

C.2 MBLEARN: MARKOV BLANKET DISCOVERY

For Markov blanket discovery, we use the Total Conditioning (TC) algorithm (Pellet & Elisseeff, 2008), implemented in the *rcd* package (Mokhtarian et al., 2025); see Algorithm 3.

Algorithm 3 : MBLEARN

Input: Observed distribution $P_{\mathbf{O}}$; target variable X.
1: Initialize: $\mathrm{MB}(X, P_{\mathbf{O}}) \leftarrow \emptyset$
2: **for** each $Y \in \mathbf{O} \setminus \{X\}$ **do**
3: **if** $X \not\perp\!\!\!\perp Y \mid \mathbf{O} \setminus \{X, Y\}$ **then**
4: $\mathrm{MB}(X, P_{\mathbf{O}}) \leftarrow \mathrm{MB}(X, P_{\mathbf{O}}) \cup \{Y\}$
5: **end if**
6: **end for**
7: $\mathrm{MB}^{+}(X, P_{\mathbf{O}}) \leftarrow \{X\} \cup \mathrm{MB}(X, P_{\mathbf{O}})$
Output: $\mathrm{MB}^{+}(X, P_{\mathbf{O}})$

D SOUNDNESS AND COMPLETENESS OF σ-FCI FOR σ-MAGS

In this section, we establish the soundness and completeness of σ-FCI with respect to the m-Markov equivalence class of σ-MAGs. Specifically, we show that the output of σ-FCI is exactly the σ-PAG representing the m-Markov equivalence class of the corresponding σ-MAG.

For the theoretical analysis, we work at the level of the underlying DMG introduced in Appendix A, rather than directly with a DG containing latent variables. This is because marginalizing latent variables in a DG yields a DMG through latent projection, so the relevant graphical relations among the observed variables can be studied directly on the resulting DMG. Moreover, the soundness and completeness results of Mooij & Claassen (2020), as well as the representation and Markov equivalence results of Yao & Mooij (2025), are formulated in terms of DMGs.

Here, σ-FCI refers to FCI applied to a σ-independence model. Mooij & Claassen (2020) established the soundness and completeness of FCI in the σ-separation setting for DMGs under the absence of selection bias, while Yao & Mooij (2025) characterized the representation and m-Markov equivalence of σ-MAGs. Our analysis connects these two results to establish the corresponding soundness and completeness guarantees for σ-MAGs.

For a σ-MAG $\mathcal{H}$, let

$$[\mathcal{H}]_m = \{\mathcal{H}' : \mathcal{H}' \text{ is } m\text{-Markov equivalent to } \mathcal{H}\}.$$

For a DMG $\mathcal{G}$, let

$$[\mathcal{G}]_\sigma = \{\mathcal{G}' : \mathcal{G}' \text{ is } \sigma\text{-Markov equivalent to } \mathcal{G}\}.$$

Lemma 17 (Theorems 3 and 4 in Yao & Mooij (2025))**.** *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two DMGs over $\mathbf{O}$, and let $\mathcal{H}_1$ and $\mathcal{H}_2$ be their corresponding σ-MAGs. Then*

$$\mathcal{G}_1 \text{ is } \sigma\text{-Markov equivalent to } \mathcal{G}_2 \iff \mathcal{H}_1 \text{ is } m\text{-Markov equivalent to } \mathcal{H}_2.$$

Lemma 18 (Theorem 1 in Mooij & Claassen (2020))**.** *Let $\mathcal{G}$ be a DMG over $\mathbf{O}$, and let*

$$\mathcal{P} = \mathcal{P}_{\mathrm{FCI}}\big(\mathrm{IM}_\sigma(\mathcal{G})\big),$$

where $\mathrm{IM}_\sigma(\mathcal{G})$ denotes the set of all σ-separations entailed by $\mathcal{G}$. Then FCI is sound, arrowhead complete, tail complete, and Markov complete in the σ-separation setting. In particular, for any distinct $X, Y \in \mathbf{O}$:

(i) if $\mathcal{P}$ has an arrowhead at X on the edge between X and Y, then

$$X \notin \mathrm{An}(Y, \mathcal{G});$$

(ii) if $\mathcal{P}$ contains $X \to Y$, then

$$X \in \mathrm{An}(Y, \mathcal{G});$$

(iii) if

$$X \notin \mathrm{An}(Y, \mathcal{G}') \quad \textit{for every } \mathcal{G}' \in [\mathcal{G}]_\sigma,$$

then $\mathcal{P}$ has an arrowhead at X;

(iv) if

$$X \in \mathrm{An}(Y, \mathcal{G}') \quad \textit{for every } \mathcal{G}' \in [\mathcal{G}]_\sigma,$$

then $\mathcal{P}$ has a tail at X.

Lemma 19. *Let $\mathcal{G}$ be a DMG over $\mathbf{O}$ and let $\mathcal{H}$ be its corresponding σ-MAG. For any adjacent vertices $X, Y \in \mathbf{O}$,*

$$\mathcal{H} \textit{ has an arrowhead at } X \textit{ on } X \leftarrow\!\!* Y \iff X \notin \mathrm{An}(Y, \mathcal{G}),$$
$$\mathcal{H} \textit{ has a tail at } X \textit{ on } X \twoheadrightarrow Y \iff X \in \mathrm{An}(Y, \mathcal{G}).$$

Proof of Lemma 19. By Definition 5 of Yao & Mooij (2025), the result follows directly from the endpoint semantics of a σ-MAG. □

Lemma 20. *Let $\mathcal{G}$ be a DMG over $\mathbf{O}$, and let $\mathcal{H}$ be its corresponding σ-MAG. For any adjacent vertices $X, Y \in \mathbf{O}$ in $\mathcal{H}$,*

$$\textit{every } \mathcal{H}' \in [\mathcal{H}]_m \textit{ has an arrowhead at } X \textit{ on } X \leftarrow\!\!* Y$$
$$\iff X \notin \mathrm{An}(Y, \mathcal{G}') \quad \textit{for every } \mathcal{G}' \in [\mathcal{G}]_\sigma,$$

and

$$\textit{every } \mathcal{H}' \in [\mathcal{H}]_m \textit{ has a tail at } X \textit{ on } X \twoheadrightarrow Y$$
$$\iff X \in \mathrm{An}(Y, \mathcal{G}') \quad \textit{for every } \mathcal{G}' \in [\mathcal{G}]_\sigma.$$

Proof of Lemma 20. We prove the first equivalence; the second follows analogously.

Suppose first that every member of $[\mathcal{H}]_m$ has an arrowhead at X. Furthermore, suppose that there exists $\mathcal{G}' \in [\mathcal{G}]_\sigma$ such that $X \in \mathrm{An}(Y, \mathcal{G}')$. By Lemma 17, its corresponding σ-MAG $\mathcal{H}'$ belongs to $[\mathcal{H}]_m$. By Lemma 19, $\mathcal{H}'$ has a tail at X, i.e., $X \twoheadrightarrow Y$ in $\mathcal{H}'$. This contradicts the assumption that every member of $[\mathcal{H}]_m$ has an arrowhead at X. Hence,

$$X \notin \mathrm{An}(Y, \mathcal{G}') \quad \text{for every } \mathcal{G}' \in [\mathcal{G}]_\sigma.$$

Conversely, suppose that

$$X \notin \mathrm{An}(Y, \mathcal{G}') \quad \text{for every } \mathcal{G}' \in [\mathcal{G}]_\sigma.$$

Furthermore, suppose that there exists $\mathcal{H}' \in [\mathcal{H}]_m$ that has a tail at X. By Lemma 17, its underlying DMG $\mathcal{G}'$ belongs to $[\mathcal{G}]_\sigma$. By Lemma 19, X is an ancestor of Y in $\mathcal{G}'$. This contradicts the assumption that $X \notin \mathrm{An}(Y, \mathcal{G}')$ for every member of $[\mathcal{G}]_\sigma$.

Replacing arrowheads and non-ancestry by tails and ancestry gives the second equivalence. □

Theorem 5 (Soundness and completeness of σ-FCI for σ-MAGs)**.** *Let $\mathcal{G}$ be a DMG over $\mathbf{O}$, let $\mathcal{H}$ be its corresponding σ-MAG, and let*

$$\mathcal{P} = \mathcal{P}_{\mathrm{FCI}}\big(\mathrm{IM}_\sigma(\mathcal{G})\big).$$

Then $\mathcal{P}$ is the σ-PAG representing $[\mathcal{H}]_m$. Consequently, σ-FCI is sound and complete with respect to the m-Markov equivalence class of σ-MAGs.

Proof of Theorem 5. By the soundness of FCI and Definition 5 of Yao & Mooij (2025), X and Y are adjacent in $\mathcal{P}$ if and only if they are adjacent in $\mathcal{H}$. Since m-Markov equivalent σ-MAGs have the same adjacencies, $\mathcal{P}$ has exactly the adjacencies shared by all members of $[\mathcal{H}]_m$.

For endpoint marks, by Lemma 20, an arrowhead at X is invariant over $[\mathcal{H}]_m$ if and only if $X \notin \mathrm{An}(Y, \mathcal{G}')$ for every $\mathcal{G}' \in [\mathcal{G}]_\sigma$. By the soundness and arrowhead completeness of FCI in Lemma 18, this holds if and only if $\mathcal{P}$ has an arrowhead at X.

Similarly, by Lemma 20, a tail at X is invariant over $[\mathcal{H}]_m$ if and only if $X \in \mathrm{An}(Y, \mathcal{G}')$ for every $\mathcal{G}' \in [\mathcal{G}]_\sigma$. By the soundness and tail completeness of FCI in Lemma 18, this holds if and only if $\mathcal{P}$ has a tail at X.

Therefore, $\mathcal{P}$ contains exactly the adjacencies and endpoint marks that are invariant over $[\mathcal{H}]_m$, while every non-invariant endpoint remains a circle. By Definition 5, $\mathcal{P}$ is the σ-PAG representing $[\mathcal{H}]_m$. □

Although σ-MAGs may contain undirected edges, as do standard MAGs in the presence of selection bias, obtaining the complete σ-PAG does not require the additional orientation rules introduced specifically for selection bias. Apart from using R0 to orient the initial collider structures, rules R1–R4 together with R8a, R9, and R10 are sufficient in our setting. This naturally raises the question of why the additional rules for undirected edges are unnecessary in this setting. The following lemma provides the key insight: the tails on an undirected edge in a σ-MAG are not invariant over its m-Markov equivalence class.

Lemma 21 (Non-invariance of tails on undirected edges)**.** *Let $\mathcal{G}$ be a DMG over $\mathbf{O}$, and let $\mathcal{H}$ be its corresponding σ-MAG. Suppose that $X - Y$ is an edge in $\mathcal{H}$. Then there exist $\mathcal{H}_X, \mathcal{H}_Y \in [\mathcal{H}]_m$ such that $Y \to X$ is in $\mathcal{H}_X$ and $X \to Y$ is in $\mathcal{H}_Y$. Consequently, neither tail on an undirected edge is invariant over $[\mathcal{H}]_m$.*

Proof of Lemma 21. Since $X - Y$ is present in $\mathcal{H}$, both endpoints of the edge are tails. By Lemma 19, $X \in \mathrm{An}(Y, \mathcal{G})$ and $Y \in \mathrm{An}(X, \mathcal{G})$. Hence, X and Y belong to the same strongly connected component of $\mathcal{G}$.

Consider an acyclification $\mathcal{G}_X$ of $\mathcal{G}$. Within the strongly connected component containing X and Y, choose a fully connected DAG whose topological ordering places Y before X. Such an acyclification exists because a strongly connected component can be replaced by an arbitrary fully connected DAG (Mooij & Claassen, 2020). Therefore, $Y \to X$ is in $\mathcal{G}_X$. Since $\mathcal{G}_X$ is acyclic, we have $Y \in \mathrm{An}(X, \mathcal{G}_X)$ and $X \notin \mathrm{An}(Y, \mathcal{G}_X)$.

Let $\mathcal{H}_X$ be the corresponding σ-MAG of $\mathcal{G}_X$. By Lemma 19, the endpoint at Y is a tail and the endpoint at X is an arrowhead. Hence, $Y \to X$ is in $\mathcal{H}_X$. Moreover, since $\mathcal{G}_X$ is an acyclification of $\mathcal{G}$, Lemma 22 gives $\mathcal{G}_X \in [\mathcal{G}]_\sigma$. By Lemma 17, it follows that $\mathcal{H}_X \in [\mathcal{H}]_m$.

Analogously, choose an acyclification $\mathcal{G}_Y$ in which the topological ordering of the same strongly connected component places X before Y. Then $X \to Y$ is in $\mathcal{G}_Y$. Let $\mathcal{H}_Y$ be its corresponding σ-MAG. By Lemma 19, $X \to Y$ is in $\mathcal{H}_Y$. Since $\mathcal{G}_Y \in [\mathcal{G}]_\sigma$, Lemma 17 further gives $\mathcal{H}_Y \in [\mathcal{H}]_m$.

Thus, for either endpoint of the undirected edge $X - Y$, there exists a member of $[\mathcal{H}]_m$ in which that endpoint is an arrowhead. Hence, neither tail is invariant over $[\mathcal{H}]_m$. □

Lemma 22 (Proposition 2 in Mooij & Claassen (2020))**.** *Let $\mathcal{G}$ be a DMG over $\mathbf{O}$, and let $\mathcal{G}'$ be any acyclification of $\mathcal{G}$. Then*

$$\mathrm{IM}_\sigma(\mathcal{G}) = \mathrm{IM}_\sigma(\mathcal{G}') = \mathrm{IM}_d(\mathcal{G}').$$

In particular, $\mathcal{G}'$ is σ-Markov equivalent to $\mathcal{G}$, i.e., $\mathcal{G}' \in [\mathcal{G}]_\sigma$.

E Experimental Details and Additional Results

E.1 Baseline Implementations and Experimental Setup

For all compared methods, we use the default hyperparameter settings provided by their official or publicly available implementations, unless otherwise stated. For conditional-independence-test-based methods, the same significance level $\alpha = 0.01$ is used throughout all experiments to ensure comparability across methods. The implementations used for the compared methods are summarized in Table 3. All experiments are conducted on a machine equipped with an Apple M5 Pro CPU and 24 GB RAM.

E.2 Evaluation Metrics

Following the evaluation protocol in (Li et al., 2026; Xie et al., 2024), let $\mathcal{P}$ and $\widehat{\mathcal{P}}$ denote the ground-truth σ-PAG and the estimated graph, respectively, represented using the same PAG matrix

Table 3: Implementation sources of the compared methods.

Method	Code Source
σ-FCI	https://github.com/cmu-phil/causal-learn
L-MARVEL	https://github.com/ban-epfl/rcd
MB-by-MB	https://github.com/kuiy/pyCausalFS
CMB	https://github.com/kuiy/pyCausalFS
LatentLCD	https://github.com/z-dragonl/Causal-Learner

encoding. Since we focus on target-specific causal discovery, structural accuracy is evaluated only on the entries associated with the target variable T. Specifically, define

$$\Omega_T = \{(T, V) : V \in \mathbf{O} \setminus \{T\}\} \cup \{(V, T) : V \in \mathbf{O} \setminus \{T\}\}.$$

Thus, Ω_T contains the target row and target column, excluding the diagonal entry (T, T). A nonzero entry represents an endpoint mark, whereas a zero entry indicates the absence of an edge.

Endpoint-Mark Precision, Recall, and F1. We evaluate endpoint marks at the entry level. A predicted mark is counted as correct only if it is nonzero and exactly matches the corresponding ground-truth mark. Let

$$\mathrm{TP}_{\mathrm{mark}} = \sum_{(i,j)\in\Omega_T} \mathbb{I}\Big[\widehat{\mathcal{P}}(i,j) \neq 0 \ \wedge \ \widehat{\mathcal{P}}(i,j) = \mathcal{P}(i,j)\Big].$$

The mark-level Precision and Recall are then defined as

$$\mathrm{Precision} = \frac{\mathrm{TP}_{\mathrm{mark}}}{\sum_{(i,j)\in\Omega_T} \mathbb{I}\Big[\widehat{\mathcal{P}}(i,j) \neq 0\Big]}, \qquad \mathrm{Recall} = \frac{\mathrm{TP}_{\mathrm{mark}}}{\sum_{(i,j)\in\Omega_T} \mathbb{I}[\mathcal{P}(i,j) \neq 0]}.$$

Their harmonic mean gives

$$\mathrm{F1} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}.$$

If the denominator of any metric is zero, the corresponding score is set to zero.

Local Structural Hamming Distance. We further use a target-specific version of Structural Hamming Distance (SHD), termed *Local-SHD*, defined as

$$\text{Local-SHD} = \sum_{(i,j)\in\Omega_T} \mathbb{I}\Big[\widehat{\mathcal{P}}(i,j) \neq \mathcal{P}(i,j)\Big].$$

Local-SHD therefore counts the number of mismatched PAG entries associated with the target, capturing errors due to missing or false adjacencies as well as incorrect endpoint marks. Lower Local-SHD indicates better target-specific structural recovery.

Computational Efficiency. Computational efficiency is evaluated using runtime and the total number of conditional independence (CI) tests performed by each method. Lower values indicate higher computational efficiency.

E.3 Additional Results on Random Structures

Table 4: Results for $d = 50$ with sample sizes $n \in \{1000, 3000, 5000\}$.

	Structural Accuracy				Efficiency	
Method	Precision ↑	Recall ↑	F1 ↑	Local SHD ↓	Runtime ↓	CI tests ($\times 10^3$) ↓
Sample size $n = 1000$						
σ-FCI	0.80 ± 0.31	0.78 ± 0.30	0.78 ± 0.30	1.00 ± 1.45	0.14 ± 0.03	3.54 ± 0.601
L-MARVEL	0.82 ± 0.32	0.82 ± 0.32	0.82 ± 0.32	0.70 ± 1.26	0.05 ± 0.04	1.77 ± 0.691
MB-by-MB	0.35 ± 0.28	0.53 ± 0.34	0.40 ± 0.29	4.80 ± 3.11	0.34 ± 0.21	3.19 ± 2.34
CMB	0.56 ± 0.40	0.57 ± 0.40	0.57 ± 0.40	1.90 ± 1.68	0.98 ± 1.39	10.70 ± 16.20
LatentLCD	0.78 ± 0.27	0.81 ± 0.27	0.80 ± 0.27	1.05 ± 1.43	0.38 ± 0.29	2.45 ± 1.32
LoCL (ours)	$\mathbf{0.85 \pm 0.27}$	$\mathbf{0.85 \pm 0.27}$	$\mathbf{0.85 \pm 0.27}$	$\mathbf{0.60 \pm 1.10}$	$\mathbf{0.03 \pm 0.10}$	$\mathbf{0.17 \pm 0.15}$
Sample size $n = 3000$						
σ-FCI	0.84 ± 0.27	0.81 ± 0.28	0.82 ± 0.27	0.75 ± 1.12	0.17 ± 0.06	5.22 ± 1.69
L-MARVEL	0.80 ± 0.37	0.80 ± 0.37	0.80 ± 0.37	0.80 ± 1.47	0.11 ± 0.08	2.31 ± 1.02
MB-by-MB	0.32 ± 0.27	0.47 ± 0.32	0.36 ± 0.25	5.60 ± 3.35	0.94 ± 0.58	5.73 ± 3.69
CMB	0.57 ± 0.35	0.56 ± 0.35	0.56 ± 0.34	1.95 ± 1.57	2.11 ± 2.97	14.40 ± 22.10
LatentLCD	0.73 ± 0.32	0.74 ± 0.31	0.73 ± 0.31	1.35 ± 1.57	0.46 ± 0.33	4.00 ± 2.12
LoCL (ours)	$\mathbf{0.86 \pm 0.29}$	$\mathbf{0.86 \pm 0.29}$	$\mathbf{0.86 \pm 0.29}$	$\mathbf{0.55 \pm 1.15}$	$\mathbf{0.03 \pm 0.09}$	$\mathbf{0.29 \pm 0.46}$
Sample size $n = 5000$						
σ-FCI	0.86 ± 0.25	0.86 ± 0.25	0.86 ± 0.25	0.55 ± 1.00	0.22 ± 0.07	6.77 ± 1.95
L-MARVEL	0.82 ± 0.34	0.82 ± 0.34	0.82 ± 0.34	0.80 ± 1.64	0.20 ± 0.19	3.05 ± 2.01
MB-by-MB	0.30 ± 0.22	0.56 ± 0.36	0.38 ± 0.25	6.55 ± 4.10	2.16 ± 1.67	9.71 ± 8.01
CMB	0.53 ± 0.40	0.56 ± 0.43	0.54 ± 0.41	2.05 ± 1.79	2.95 ± 5.49	16.80 ± 34.60
LatentLCD	0.70 ± 0.32	0.71 ± 0.30	0.70 ± 0.31	1.45 ± 1.70	0.51 ± 0.27	5.95 ± 3.82
LoCL (ours)	$\mathbf{0.90 \pm 0.19}$	$\mathbf{0.90 \pm 0.19}$	$\mathbf{0.90 \pm 0.19}$	$\mathbf{0.40 \pm 0.75}$	$\mathbf{0.03 \pm 0.09}$	$\mathbf{0.36 \pm 0.49}$

Note: ↑ indicates that a higher value is better, while ↓ indicates that a lower value is better. The best result in each group is highlighted in bold.

Table 5: Results for $d = 100$ with sample sizes $n \in \{1000, 3000, 5000\}$.

	Structural Accuracy				Efficiency	
Method	Precision ↑	Recall ↑	F1 ↑	Local SHD ↓	Runtime ↓	CI tests ($\times 10^3$) ↓
			Sample size $n = 1000$			
σ-FCI	$\mathbf{0.82 \pm 0.27}$	0.80 ± 0.26	0.81 ± 0.26	1.10 ± 1.86	0.48 ± 0.08	13.90 ± 2.69
L-MARVEL	0.74 ± 0.39	0.74 ± 0.39	0.74 ± 0.39	1.15 ± 1.81	0.17 ± 0.12	6.54 ± 1.82
MB-by-MB	0.25 ± 0.28	0.39 ± 0.35	0.29 ± 0.29	7.45 ± 4.83	0.98 ± 0.73	6.86 ± 6.38
CMB	0.46 ± 0.40	0.47 ± 0.40	0.47 ± 0.40	2.40 ± 1.88	6.90 ± 9.62	64.30 ± 96.00
LatentLCD	0.79 ± 0.21	0.80 ± 0.19	0.80 ± 0.20	0.90 ± 1.02	0.71 ± 0.49	9.28 ± 6.75
LoCL (ours)	$\mathbf{0.82 \pm 0.34}$	$\mathbf{0.82 \pm 0.34}$	$\mathbf{0.82 \pm 0.34}$	$\mathbf{0.80 \pm 1.64}$	$\mathbf{0.03 \pm 0.10}$	$\mathbf{0.12 \pm 0.16}$
			Sample size $n = 3000$			
σ-FCI	0.86 ± 0.25	0.84 ± 0.26	0.85 ± 0.25	0.65 ± 1.04	0.75 ± 0.22	20.30 ± 4.90
L-MARVEL	0.76 ± 0.43	0.76 ± 0.43	0.76 ± 0.43	1.05 ± 1.93	0.23 ± 0.17	7.09 ± 2.16
MB-by-MB	0.20 ± 0.26	0.39 ± 0.36	0.25 ± 0.28	8.35 ± 4.52	3.78 ± 3.08	18.50 ± 16.60
CMB	0.54 ± 0.38	0.55 ± 0.38	0.55 ± 0.38	2.00 ± 1.78	10.53 ± 11.79	59.00 ± 73.50
LatentLCD	0.68 ± 0.32	0.72 ± 0.29	0.70 ± 0.31	1.60 ± 1.76	0.89 ± 0.61	17.40 ± 12.80
LoCL (ours)	$\mathbf{0.90 \pm 0.31}$	$\mathbf{0.90 \pm 0.31}$	$\mathbf{0.90 \pm 0.31}$	$\mathbf{0.60 \pm 1.85}$	$\mathbf{0.03 \pm 0.09}$	$\mathbf{0.13 \pm 0.21}$
			Sample size $n = 5000$			
σ-FCI	0.83 ± 0.30	0.84 ± 0.28	0.83 ± 0.29	0.95 ± 1.50	0.80 ± 0.36	25.20 ± 11.50
L-MARVEL	0.89 ± 0.29	0.86 ± 0.30	0.87 ± 0.29	0.55 ± 1.19	0.45 ± 0.37	8.35 ± 3.58
MB-by-MB	0.19 ± 0.26	0.39 ± 0.35	0.24 ± 0.27	8.95 ± 4.54	8.89 ± 6.95	26.10 ± 22.80
CMB	0.44 ± 0.35	0.47 ± 0.37	0.45 ± 0.36	2.80 ± 1.91	22.27 ± 22.38	71.90 ± 82.80
LatentLCD	0.71 ± 0.35	0.74 ± 0.31	0.72 ± 0.32	1.65 ± 1.90	1.54 ± 0.84	33.80 ± 20.20
LoCL (ours)	$\mathbf{0.95 \pm 0.15}$	$\mathbf{0.93 \pm 0.18}$	$\mathbf{0.93 \pm 0.17}$	$\mathbf{0.30 \pm 0.73}$	$\mathbf{0.04 \pm 0.10}$	$\mathbf{0.27 \pm 0.39}$

Note: ↑ indicates that a higher value is better, while ↓ indicates that a lower value is better. The best result in each group is highlighted in bold.

Table 6: Results for $d = 150$ with sample sizes $n \in \{1000, 3000, 5000\}$.

	Structural Accuracy				Efficiency	
Method	Precision ↑	Recall ↑	F1 ↑	Local SHD ↓	Runtime ↓	CI tests ($\times 10^3$) ↓
			Sample size $n = 1000$			
σ-FCI	0.81 ± 0.32	0.81 ± 0.32	0.81 ± 0.32	0.75 ± 1.29	0.93 ± 0.18	27.20 ± 6.86
L-MARVEL	0.86 ± 0.31	0.86 ± 0.31	0.86 ± 0.31	0.55 ± 1.23	0.23 ± 0.13	12.80 ± 2.03
MB-by-MB	0.14 ± 0.17	0.30 ± 0.34	0.18 ± 0.20	8.00 ± 4.70	1.46 ± 0.96	9.35 ± 8.03
CMB	0.46 ± 0.42	0.46 ± 0.42	0.46 ± 0.42	2.35 ± 2.01	3.60 ± 4.79	29.30 ± 43.60
LatentLCD	0.66 ± 0.32	0.70 ± 0.30	0.68 ± 0.31	1.70 ± 1.81	1.69 ± 0.82	29.10 ± 18.00
LoCL (ours)	$\mathbf{0.88 \pm 0.28}$	$\mathbf{0.88 \pm 0.28}$	$\mathbf{0.88 \pm 0.28}$	$\mathbf{0.50 \pm 1.10}$	$\mathbf{0.04 \pm 0.10}$	$\mathbf{0.24 \pm 0.40}$
			Sample size $n = 3000$			
σ-FCI	0.88 ± 0.25	0.89 ± 0.24	0.88 ± 0.24	0.55 ± 1.10	1.21 ± 0.20	35.20 ± 6.65
L-MARVEL	0.90 ± 0.22	0.90 ± 0.22	0.90 ± 0.22	0.40 ± 0.88	0.55 ± 0.47	16.10 ± 5.53
MB-by-MB	0.21 ± 0.27	0.40 ± 0.38	0.26 ± 0.30	8.60 ± 6.56	6.60 ± 6.86	26.30 ± 33.00
CMB	0.45 ± 0.42	0.46 ± 0.42	0.46 ± 0.42	2.45 ± 2.04	23.17 ± 29.26	102.00 ± 156.00
LatentLCD	0.67 ± 0.28	0.71 ± 0.28	0.68 ± 0.28	1.55 ± 1.43	2.04 ± 1.45	43.20 ± 35.30
LoCL (ours)	$\mathbf{0.95 \pm 0.17}$	$\mathbf{0.95 \pm 0.17}$	$\mathbf{0.95 \pm 0.17}$	$\mathbf{0.20 \pm 0.70}$	$\mathbf{0.03 \pm 0.10}$	$\mathbf{0.15 \pm 0.19}$
			Sample size $n = 5000$			
σ-FCI	0.84 ± 0.24	0.85 ± 0.22	0.84 ± 0.23	0.70 ± 1.13	1.41 ± 0.36	40.00 ± 10.80
L-MARVEL	0.94 ± 0.18	0.94 ± 0.18	0.94 ± 0.18	0.25 ± 0.72	0.64 ± 0.63	16.20 ± 6.42
MB-by-MB	0.15 ± 0.16	0.38 ± 0.39	0.21 ± 0.21	8.80 ± 4.30	19.88 ± 19.15	47.50 ± 53.40
CMB	0.40 ± 0.42	0.41 ± 0.42	0.40 ± 0.42	2.85 ± 2.16	62.44 ± 67.22	153.00 ± 178.00
LatentLCD	0.67 ± 0.34	0.71 ± 0.28	0.68 ± 0.31	1.85 ± 2.03	2.20 ± 1.60	57.90 ± 50.90
LoCL (ours)	$\mathbf{0.96 \pm 0.09}$	$\mathbf{0.96 \pm 0.09}$	$\mathbf{0.96 \pm 0.09}$	$\mathbf{0.15 \pm 0.37}$	$\mathbf{0.04 \pm 0.10}$	$\mathbf{0.50 \pm 0.89}$

Note: ↑ indicates that a higher value is better, while ↓ indicates that a lower value is better. The best result in each group is highlighted in bold.

F AI USE STATEMENT

We used ChatGPT to refine writing only. The prompt was: "I am preparing a paper for submission to an international conference and would like your help to check for any grammatical issues and refine the wording or sentence structure where necessary to ensure conciseness and precision." Edits were applied paragraph-by-paragraph, and all outputs were verified and revised by the authors; no scientific content, analyses, or references were generated by the tool.