Adaptive Online Experimental Design for Causal Discovery

Muhammad Qasim Elahi [*1] **Lai Wei** [*2] **Murat Kocaoglu** [1] **Mahsa Ghasemi** [1]

Abstract

Causal discovery aims to uncover cause-and-effect relationships encoded in causal graphs by leveraging observational, interventional data, or their combination. The majority of existing causal discovery methods are developed assuming infinite interventional data. We focus on interventional data efficiency and formalize causal discovery from the perspective of online learning, inspired by pure exploration in bandit problems. A graph separating system, consisting of interventions that cut every edge of the graph at least once, is sufficient for learning causal graphs when infinite interventional data is available, even in the worst case. We propose a track-and-stop causal discovery algorithm that adaptively selects interventions from the graph separating system via allocation matching and learns the causal graph based on sampling history. Given any desired confidence value, the algorithm determines a termination condition and runs until it is met. We analyze the algorithm to establish a problem-dependent upper bound on the expected number of required interventional samples. Our proposed algorithm outperforms existing methods in simulations across various randomly generated causal graphs. It achieves higher accuracy, measured by the structural hamming distance (SHD) between the learned causal graph and the ground truth, with significantly fewer samples.

1. Introduction

Causal discovery is a fundamental problem encountered across various scientific and engineering disciplines (Pearl, 2009; Spirtes et al., 2000; Peters et al., 2017). Observational data is generally inadequate for establishing causal relationships and interventional data, obtained by deliberately perturbing the system, becomes necessary. Consequently, contemporary approaches propose leveraging both observational and interventional data for causal discovery (Hauser & Bühlmann, 2014; Greenewald et al., 2019). A well-established model for depicting causal relationships is the directed acyclic graph (DAG). A directed edge between two variables indicates a direct causal effect, while a directed path indicates an indirect causal effect (Spirtes et al., 2000).

A causal graph is typically identifiable only up to its Markov equivalence class (MEC) (Verma & Pearl, 2022) using observational data. Markov equivalence class is a set of DAGs that encode the same set of conditional independencies. There is a growing focus on developing algorithms for the design of interventions, specifically aimed at learning causal graphs (Hu et al., 2014; Shanmugam et al., 2015; Ghassami et al., 2017). These algorithms rely on the availability of an infinite amount of interventional data, whose collection in real-world settings is often more challenging and expensive than gathering observational data. In numerous medical contexts, abundant observational clinical data is readily available (Subramani & Cooper, 1999), whereas conducting randomized controlled trials can be costly or sometimes present ethical challenges. In this work, we consider a scenario where access is limited to only a finite number of interventional samples. Similar to (Hu et al., 2014; Shanmugam et al., 2015; Ghassami et al., 2017), we assume causal sufficiency, meaning that all variables are observed, and no latent or hidden variables are involved.

The PC algorithm (Spirtes et al., 2000) utilizes conditional independence tests in combination with Meek orientation rules (Meek, 1995) to learn the causal structure with all identifiable causal relations from the data. The graph separating system, which is a set of interventions that cuts every edge of the graph at least once is sufficient for learning the full causal graph when infinite interventional data is available. (Shanmugam et al., 2015; Kocaoglu et al., 2017). Bayesian causal discovery is a valuable tool for efficiently learning causal models from limited interventional data, but it encounters challenges when it comes to computing probabilities over the combinatorial space of DAGs (Heckerman et al., 1997; Annadani et al., 2023; Toth et al., 2022). Deal-

[*]Equal contribution [1]School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA [2]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan, USA. Correspondence to: Muhammad Qasim Elahi <elahi0@purdue.edu>, Lai Wei <weilaitim@gmail.com>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

Reference	Adaptive/Non-adaptive	Graph structure constraints	Interventional sample efficiency
Hauser & Bühlmann, 2012	Non-adaptive	None	✗
Shanmugam et al., 2015	Non-adaptive	None	✗
Kocaoglu et al., 2017	Non-adaptive	None	✗
Greenewald et al., 2019	Adaptive	Trees only	✓
Squires et al., 2020	Adaptive	None	✗
Choo & Shiragur, 2023	Adaptive	None	✗
Track & Stop Causal Discovery (Ours)	Adaptive	None	✓

Table 1: A comparison of existing causal discovery techniques with our proposed algorithm

ing with the search complexity for DAGs without relying on specific parametric assumptions remains a challenge.

A comparison between our proposed algorithm and existing methods is presented in Table 1. Causal discovery algorithms can be broadly classified into two categories: adaptive and non-adaptive. In the offline or non-adaptive setting, interventions are predetermined before algorithm execution. The majority of existing offline discovery algorithms require access to infinite interventional samples (Hauser & Bühlmann, 2012; Shanmugam et al., 2015; Kocaoglu et al., 2017). Existing online discovery algorithms apply interventions sequentially, with adaptively chosen targets at each step, still necessitating access to interventional distributions, i.e., an infinite number of interventional samples (Squires et al., 2020; Choo & Shiragur, 2023). Although the algorithm by Greenewald et al. works with finite interventional data, it is applicable only when the underlying causal structure is a tree. Our proposed tracking and stopping algorithm does not require any graphical assumptions and provides a sample-efficient alternative for general graphs.

We approach causal discovery from an online learning standpoint, emphasizing knowledge acquisition and incremental decision-making. Inspired by the pure exploration problem in multi-armed bandit (Kaufmann et al., 2016; Degenne et al., 2019), we view the possible interventions as the action space. We propose a discovery algorithm that adaptively selects interventions from the graph separating system via an allocation matching approach similar to one employed in Wei et al., 2024. Our objective is to uncover the true DAG with a predefined level of confidence while minimizing the number of interventional samples required. The main contributions of our work are listed below:

- We study the causal discovery problem with fixed confidence and proposed a track-and-stop causal discovery algorithm that can adaptively select informative interventions according to the sampling history.
- We analyze the algorithm to show it can learn the true DAG with any given confidence level and provide an upper bound on the expected number of required interventional samples.
- We conduct a series of experiments using random DAGs and the SACHS Bayesian network from bnlibrary (Scutari, 2009) to compare our algorithm with other baselines. The results show that our algorithm outperforms the baselines, requiring fewer samples.

2. Problem Formulation

A causal graph $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ is a DAG with the vertex set $\mathbf{V}$ corresponding to a set of random variables. If there is a directed edge $(X, Y) \in \mathbf{E}$ from variable X to variable Y, denoted as $X \to Y$, it means that X is a direct cause or an immediate parent of Y. The parent set of a variable Y is denoted by $\mathsf{Pa}(Y)$. The induced graph $\mathcal{D}_{\mathbf{X}}$ has a vertex set $\mathbf{X}$, and the edge set contains all edges with both endpoints in $\mathbf{X}$. The cut at a set of vertices $\mathbf{X}$, denoted by $E[\mathbf{X}, \mathbf{V} \setminus \mathbf{X}]$, is the set of edges between the nodes in $\mathbf{X}$ and $\mathbf{V} \setminus \mathbf{X}$. Based on the Markov assumption, the joint distribution can be factorized as $P(\mathbf{v}) = \prod_{i=1}^{n} P(v_i|\mathsf{pa}(X_i))$. A causal graph implies specific conditional independence (CI) relationships among variables through d-separation statements. A collection of DAGs is considered Markov equivalent when they exhibit the same set of CI relations. In any learning environment, it is necessary to make some form of faithfulness assumption in order to deduce graphical characteristics from the constraints imposed by the distribution (Yang et al., 2018; Zhang, 2008; Jaber et al., 2020).

Definition 1 (Faithfulness (Zhang & Spirtes, 2012))**.** *In the population distribution, no conditional independence relations exist other than those implied by the d-separation statements in the true causal DAG.*

Observational faithfulness assumption implies that observed independencies in the population arise from its underlying structure rather than coincidence and is widely used in causal discovery algorithms (Scheines, 1997; Hauser & Bühlmann, 2012). The observational data can be used to learn the skeleton of the underlying DAG with some additional edge orientations. In order to completely orient the causal graph, we need access to interventional samples. An intervention on a subset of variables $\mathbf{S} \subseteq \mathbf{V}$, denoted by the do-operator $do(\mathbf{S} = \mathbf{s})$, involves setting each $S_j \in \mathbf{S}$

to s_j. Let $\mathcal{D}_{\overline{\mathbf{S}}}$ denote the corresponding post interventional causal graph with incoming edges to nodes in $\mathbf{S}$ removed. Using the truncated factorization formula over $\mathcal{D}_{\overline{\mathbf{S}}}$, if $\mathbf{v}$ is consistent with the realization $\mathbf{s}$, we have:

$$P_{\mathbf{s}}(\mathbf{v}) := P(\mathbf{v} \mid do(\mathbf{S} = \mathbf{s})) = \prod_{V_i \notin \mathbf{s}} P(v_i | \mathsf{pa}(V_i)) \quad (1)$$

For a DAG $\mathcal{D}$, we denote the interventional and observational distributions as $P_{\mathbf{s}}^{\mathcal{D}}(\mathbf{v})$ and $P^{\mathcal{D}}(\mathbf{v})$ respectively. In many scenarios, abundant observational data allow for an accurate approximation of the ground truth observational distribution. Therefore, we make the following assumption:

Assumption 1. *We assume that each variable $V \in \mathbf{V}$ is discrete and that the observational distribution is available and faithful to the true causal graph.*

Causal Discovery with Fixed Confidence: Under assumption 1, the causal DAG can be learned up to the MEC with the PC algorithm (Spirtes et al., 2000). To orient remaining edges, we need interventional data. We consider a fixed confidence setting, where the learner is given a confidence level $\delta \in (0, 1)$ and is required to output the true DAG with probability at least $1 - \delta$. This problem setup is inspired by the pure exploration problem in multi-armed bandits (Kaufmann et al., 2016). It requires the learner to adaptively select informative interventions to reveal the underlying causal structure. With a set of interventional targets $\mathbf{S}$, let the action space be $\mathcal{I} = \bigcup_{\mathbf{S} \in \mathcal{S}} \omega(\mathbf{S})$, where each $\omega(\mathbf{S})$ includes a finite number of interventions $\mathbf{S}$ or its finite number of realizations. The learner sequentially selects intervention $\mathbf{s}_t \in \mathcal{I}$ and observes a sample from the interventional distribution $\mathbf{v}_t \sim P_{\mathbf{s}_t}(\mathbf{v})$. A policy π is a sequence $\{\pi_t\}_{t \in \mathbb{N}}$, where each π_t determines the probability distribution of taking intervention $\mathbf{s}_t \in \mathcal{I}$ given intervention and observation history $\pi_t(\mathbf{s}_t \mid \mathbf{s}_1, \mathbf{v}_1, \ldots, \mathbf{s}_{t-1}, \mathbf{v}_{t-1})$.

In a fixed confidence level setting, the number of required interventional samples to output a DAG with a confidence level is unknown beforehand. For a given $\delta \in (0, 1)$, the learner is required to select a stopping time τ_δ adapted to filtration $\{\mathcal{F}_t\}_{t \in \mathbb{N}_{>0}}$ where $\mathcal{F}_t = \sigma(\mathbf{s}_1, \mathbf{v}_1, \ldots, \mathbf{s}_{t-1}, \mathbf{v}_{t-1})$. At τ_δ, the learner selects a causal graph based on select rule $\psi(\mathbf{s}_1, \mathbf{v}_1, \ldots, \mathbf{s}_{\tau_\delta - 1}, \mathbf{v}_{\tau_\delta - 1})$. The stopping time τ_δ represents the time when the learner halts and reaches confidence level δ about a selected causal graph ψ. Putting the policy, stopping time, and selection rule together, the triple (π, τ_δ, ψ) is called a causal discovery algorithm. The objective is to design an algorithm that takes as few interventional samples as possible.

3. Preliminaries

In this section, we introduce definitions and some fundamental concepts about partially directed graphs, which can be used to encode the Markov equivalence class (MEC), i.e., represent an equivalence class or set of DAGs. We employ them for algorithm design and analysis in our work.

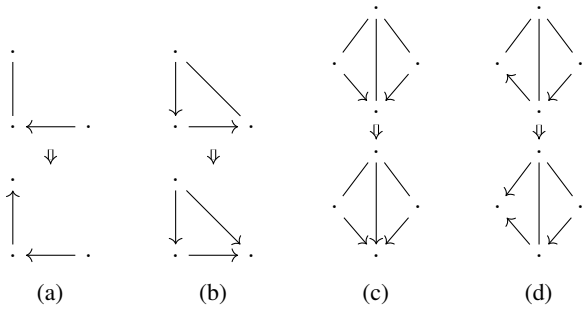


Figure 1: Forbidden induced subgraphs and orientation rules for partially directed graphs (Meek, 1995).

Partially Directed Graphs (PDAGs): A partially directed acyclic graph (PDAG) is a partially directed graph that is free from directed cycles (Perkovic, 2020). Markov equivalent DAGs can be represented by a **completed partially directed acyclic graph (CPDAG)**, denoted by $\mathcal{C}$, which has the same set of adjacencies and unshielded colliders as all the DAGs in the MEC. A triple (X, Y, Z) is called an unshielded collider if there exists a v-structure $X \to Y \leftarrow Z$ with X and Z being non-adjacent.

Definition 2. *(MPDAG) A PDAG is classified as a maximally oriented PDAG (MPDAG) if and only if it does not contain any of the forbidden induced subgraphs in the first row of the Figure 1.*

In other words, CPDAG with some additional edges oriented by the combination of side information and propagation using Meek rules is classified as a maximally oriented PDAG (MPDAG). The set of all DAGs represented by the CPDAG $\mathcal{C}$ is denoted by $[\mathcal{C}]$, and similarly, the set of all DAGs represented by the MPDAG $\mathcal{M}$ is denoted by $[\mathcal{M}]$. Both the CPDAGs and MPDAGs always take the form of a chain graph with chordal chain components. In chordal graphs, every cycle of four or more vertices always contains an additional edge, called a chord (Andersson et al., 1997).

Partial Causal Ordering (PCO) in PDAGs : A path between vertices X and Y is termed a causal path when all edges in the path are directed toward Y. A path of the form $P :=< V_1 = X, V_2, ..., V_n = Y >$ is categorized as a possibly causal path when it does not contain any edge in the form of $V_i \leftarrow V_j$, where $i < j$. A proper path from $\mathbf{X}$ to $\mathbf{Y}$ is one where only the first node belongs to $\mathbf{X}$ while the remaining nodes do not. If there is a causal path from vertex x to vertex y, it implies that x is an ancestor of y, i.e., $x \in \mathsf{An}(y)$. Likewise, if there is a possibly causal path from vertex x to vertex y, it implies that x is a possible ancestor of y, i.e., $x \in \mathsf{PoAn}(y)$. The $\mathsf{An}(\mathbf{X}, \mathcal{M})$ and $\mathsf{PoAn}(\mathbf{X}, \mathcal{M})$

for a set of nodes $\mathbf{X}$ in $\mathcal{M}$ is the union of over all vertices in $\mathbf{X}$. We adhere to the convention that each node is considered a descendant, ancestor, and possible ancestor of itself.

Definition 3 (Partial Causal Ordering)**.** *A total ordering of a subset of vertices $\mathbf{X} \subseteq \mathbf{V}$ is a causal ordering of $\mathbf{X}$ in a DAG $\mathcal{D}(\mathbf{V}, \mathbf{E})$ if $\forall\ X_i, X_j \in \mathbf{X}$ such that $X_i < X_j$ there exists an edge $X_i \to X_j \in \mathbf{E}$. In the context of an MPDAG, where unoriented edges are present, we can define the Partial Causal Ordering (PCO) of a subset $\mathbf{X} \subseteq \mathbf{V}$ in $\mathcal{M}(\mathbf{V}, \mathbf{E})$ as a total ordering of pairwise disjoint subsets $(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_m)$ such that $\bigcup_{i=1}^{m} \mathbf{X}_i = \mathbf{X}$. The PCO must fulfill the following requirement: if $\mathbf{X}_i < \mathbf{X}_j$ and there is an edge between some $X_i \in \mathbf{X}_i$ and $X_j \in \mathbf{X}_j$ in $\mathcal{M}$, then edge $X_i \to X_j$ is present in $\mathcal{M}$.*

4. Algorithm Initialization

In our problem setup, we assume access to the observational distribution and the corresponding CPDAG $\mathcal{G}$. We proceed by constructing a graph-separating system for $\mathcal{G}$ and enumerating possible causal effects.

4.1. Constructing Graph Separating System

Definition 4 (Graph Separating System)**.** *Given a graph $G = (\mathbf{V}, \mathbf{E})$, a set of different subsets of the vertex set V, $\mathcal{S} = \{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_m\}$ is a graph separating system when, for every edge $\{a, b\} \in E$, there exists a set $S_i \in \mathcal{S}$ such that either $a \in S_i$ and $b \notin S_i$ or $a \notin S_i$ and $b \in S_i$.*

In a setting where infinite interventional data is available, the interventional distributions from targets in the graph separating system for unoriented edges in CPDAG $\mathcal{C}$ are necessary and sufficient to learn the true DAG (Kocaoglu et al., 2017; Shanmugam et al., 2015). Graph coloring which can be used a method used to generate a separating system asisgns adjacent vertices are assigned distinct colors, is computationally challenging for general graphs. However, for perfect graphs like chordal graphs, efficient polynomial-time algorithms can color the graph using the minimum number of colors (Král', 2004). For a set of n variables, a separating system of the form $\mathcal{S} = \{S_1, S_2, ..., S_m\}$, such that $|S_a| \leq k, ; \forall a \in [m]$, is called a (n, k)-separating system (Katona, 1966; Wegener, 1979). In our work, we use (n, k) separating systems to ensure that the size of every intervention set is bounded by k, and enumeration of causal effects is feasible. The detailed procedure to construct (n, k) separating system is outlined in the supplementary material.

4.2. Enumerating Causal Effects

While causal effects are generally not identifiable from CPDAGs, we can still enumerate interventional distributions. We assign all possible orientations to the edge cut at every $\mathbf{S} \in \mathcal{S}$, and subsequently apply Meek's rules to generate a set of MPDAGs. From the set of resulting MPDAGs, we can then enumerate all the candidate interventional distributions using identification formula from (Perkovic, 2020).

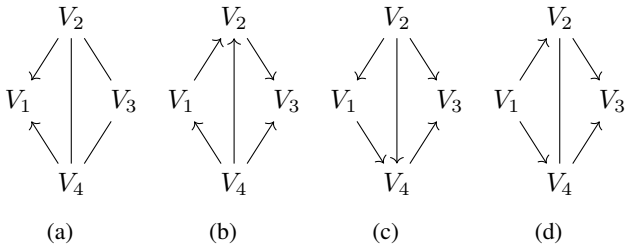


Figure 2: MPDAGs obtained by assigning orientations to edges $E[V_1, \mathbf{V} \setminus V_1]$ in corresponding skeleton i.e. CPDAG

Lemma 1 (Causal Identification Formula for MPDAG (Perkovic, 2020))**.** *Consider an MPDAG $\mathcal{M}(\mathbf{V}, \mathbf{E})$ and two disjoint sets of variables $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$. The interventional distribution $P_{\mathbf{x}}(\mathbf{y})$ is identifiable from any observational distribution consistent with $\mathcal{M}$ if there exists no possibly proper causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{M}$ that starts with an undirected edge and is given below:*

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{b}} \prod_{i=1}^{m} P(\mathbf{b}_i | \mathsf{Pa}(\mathbf{b}_i, \mathcal{M})). \tag{2}$$

The assignment for $\mathsf{Pa}(\mathbf{b}_i, \mathcal{M})$ must be in consistence with $do(\mathbf{X})$. Also $(\mathbf{B}_1, \ldots, \mathbf{B}_m)$ is a partial causal ordering of $\mathsf{An}(\mathbf{Y}, \mathcal{M}_{\mathbf{V}\setminus\mathbf{X}})$ in $\mathcal{M}$ and $\mathbf{b} = \mathsf{An}(\mathbf{Y}, \mathcal{M}_{\mathbf{V}\setminus\mathbf{X}}) \setminus \mathbf{Y}$.

The algorithm for finding the partial causal ordering (PCO) and enumerating all possible causal effects in an MPDAG, leveraging Lemma 1, is provided in the supplementary material. In cases where one or multiple possibly proper causal paths exist from $\mathbf{X}$ to $\mathbf{Y}$ and start with an undirected edge, the interventional distribution $P_{\mathbf{x}}(\mathbf{y})$ cannot be uniquely determined. However, we can enumerate all possible values for $P_{\mathbf{x}}(\mathbf{y})$ for the set of DAGs represented by the MPDAG $([\mathcal{M}])$. Suppose that $|\mathbf{X}| = k$ and the maximum degree of $\mathcal{M}$ is d. This implies that there can be a maximum of kd edges adjacent to the vertices in set $\mathbf{X}$. To enumerate all candidate values of $P_{\mathbf{x}}(\mathbf{y})$ for every DAG in the set $[\mathcal{M}]$, we assign orientations to all the unoriented edges in $E[\mathbf{X}, V \setminus \mathbf{X}]$ and propagate using Meek rules. We denote an orientation of cutting edges $(E[\mathbf{X}, \mathbf{V} \setminus \mathbf{X}])$ as $\mathsf{C}(\mathbf{X})$. This process results in a maximum of 2^{kd} partially directed graphs, each one being a valid MPDAG. It's worth noting that, since the first edge of all paths from $\mathbf{X}$ to $\mathbf{Y}$ is oriented, the condition for the identifiability of $P_{\mathbf{x}}(\mathbf{y})$ is satisfied in all of the newly generated MPDAGs. With slight abuse of notation we denote the interventional distribution in the MPDAG $\mathcal{M}$ with cut configuration $\mathsf{C}(\mathbf{X})$ by $P_{\mathbf{x}}^{\mathsf{C}(\mathbf{X})}$.

Consider a CPDAG $\mathcal{G}$ on the vertex set $\mathbf{V} = \{V_1, V_2, V_3, V_4\}$, which has the same skeleton as the com-

plete undirected graph on $\mathbf{V}$ with the edge $V_1 - V_3$ removed. A valid separating set system for $\mathcal{G}$ is $\{\{V_1\}, \{V_1, V_2\}\}$. Figure 2 shows the 4 possible MPDAGs by assigning different orientations to the cut at $\mathbf{X} = \{V_1\}$. By applying the causal identification formula (Lemma 1) to all the 4 MPDAGs, we can identify all possible interventional distributions $P_{\mathbf{x}}(\mathbf{v})$ for MPDAGs in Fig. 2 given below.

$$P_{v_1}(v_2, v_3, v_4) = \begin{cases} P(v_2, v_3, v_4) & (a) \\ P(v_4)P(v_3|v_2, v_4)P(v_2|v_1, v_4) & (b) \\ P(v_4|v_1, v_2)P(v_3|v_2, v_4)P(v_2) & (c) \\ P(v_3|v_2, v_4)P(v_2, v_4|v_1) & (d) \end{cases}$$

For example, for the MPDAG in Figure 2(a), using the algorithm to find PCO, we obtain $\mathsf{PCO}(\mathbf{V} \setminus \mathbf{X}, \mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}) = \{V_2, V_3, V_4\}$, and $P_{\mathbf{x}}(\mathbf{v}) = P(v_2, v_3, v_4)$. We repeat this process for all the MPDAGs in Figure 2 to enumerate all the possible candidate interventional distributions in the above equation. We show in Lemma 2 that all candidate interventional distributions are different from one another. This implies that we can orient the cutting edges by comparing the candidate interventional distributions with the empirical interventional distribution. In order to ensure that the enumeration step is feasible, we use (n, k) separating system, which implies that for any target set $\mathbf{S}$ we will have at most 2^{kd} possible interventional distributions, where d is the maximum degree in the graph. We can then orient the entire DAG by repeating this procedure for all the intervention targets in the separating system.

We define the collection of interventional distributions $P_{\mathbf{S}}^{\mathcal{D}} = \{P_{\mathbf{s}}^{\mathcal{D}}\}_{\forall \mathbf{s} \in \mathrm{Dom}(\mathbf{S})}$, where $\mathrm{Dom}(\mathbf{S})$ refers to the domain of $\mathbf{S}$. We show that we have a unique $P_{\mathbf{S}}^{\mathcal{D}}$ for every possible cutting edge configuration $\mathsf{C}(\mathbf{S})$ in $\mathcal{D}$ (Lemma 2). The Lemma 2 implies that there exists a one-to-one mapping between the candidate interventional distributions and the cutting edge orientation $\mathsf{C}(\mathbf{S})$. The proof of Lemma 2 relies on (Hauser & Bühlmann, 2012, Th. 10), which requires revisiting some concepts and definitions from the paper Hauser & Bühlmann, 2012.

Lemma 2. *Assume that the faithfulness assumption in Definition 1 holds and $\mathcal{D}^*$ is the true DAG. For any DAG $\mathcal{D} \neq \mathcal{D}^*$, if $P_{\mathbf{S}}^{\mathcal{D}} = P_{\mathbf{S}}^{\mathcal{D}^*}$ for some $\mathbf{S} \subseteq \mathbf{V}$, they must share the same cutting edge orientation $\mathsf{C}(\mathbf{S})$.*

Definition 5. *Let $\mathcal{D}$ be a DAG on the vertex set $\mathbf{V}$, and let $\mathcal{S}$ be a family of targets. Then we define $\mathcal{MK}_{\mathcal{S}}(\mathcal{D})$ as follows:*

$$\mathcal{MK}_{\mathcal{S}}(\mathcal{D}) = \{(P_{\mathbf{S}}^{\mathcal{D}})_{\mathbf{S}\in\mathcal{S}} \mid \textit{condition (1) and (2) is true.}\}$$

(1) Markov property: $P_{\mathbf{S}}^{\mathcal{D}} \in \mathcal{MK}(\mathcal{D}_{\overline{\mathbf{S}}})$ for all $\mathbf{S} \in \mathcal{S}$.

(2) Local invariance property: for any pair of intervention targets $\mathbf{S_1}, \mathbf{S_2} \in \mathcal{S}$, for any non-intervened node $U \notin \mathbf{S}_1 \cup \mathbf{S}_2$, $P_{\mathbf{S}_1}^{\mathcal{D}}(U|\mathsf{Pa}_{\mathcal{D}}(U)) = P_{\mathbf{S}_2}^{\mathcal{D}}(U|\mathsf{Pa}_{\mathcal{D}}(U))$.

The $\mathcal{MK}_{\mathcal{S}}(\mathcal{D})$ is space of interventional distribution tuples $(P_{\mathbf{S}}^{\mathcal{D}})_{\mathbf{S}\in\mathcal{S}}$, where each $P_{\mathbf{S}}^{\mathcal{D}}$ is Markov relative to the post-interventional DAG $\mathcal{D}_{\overline{\mathbf{S}}}$, as indicated by $P_{\mathbf{S}}^{\mathcal{D}} \in \mathcal{MK}(\mathcal{D}_{\overline{\mathbf{S}}})$. This suggests that the expression for each $P_{\mathbf{s}}^{\mathcal{D}}$ can be formulated using truncated factorization over $\mathcal{D}_{\overline{\mathbf{S}}}$ in equation (1). Additionally, for any non-intervened variable U, the conditional distribution given its parents remains invariant across different interventions. A family of targets $\mathcal{S}$ is considered conservative if, for any $V \in \mathbf{V}$, there exists at least one $\mathbf{S} \in \mathcal{S}$ such that $V \notin \mathbf{S}$. This implies that any $\mathcal{S}$ containing the empty set, i.e., observational distribution being available, is indeed conservative. Two DAGs D and D^* are $\mathcal{S}$-Markov equivalent denoted by $D \sim_{\mathcal{S}} \mathcal{D}^*$ if $\mathcal{MK}_{\mathcal{S}}(\mathcal{D}) = \mathcal{MK}_{\mathcal{S}}(\mathcal{D}^*)$.

Lemma 3 ((Hauser & Bühlmann, 2012), Th. 10)**.** *Let $\mathcal{D}$ and $\mathcal{D}^*$ be two DAGs on $\mathbf{V}$, and $\mathcal{S}$ be a conservative family of targets. Then, the following statements are equivalent:*

1. *$\mathcal{D} \sim_{\mathcal{S}} \mathcal{D}^*$.*
2. *$\mathcal{D}$ and $\mathcal{D}^*$ have the same skeleton and the same v-structures, and $\mathcal{D}_{\overline{\mathbf{S}}}$ and $\mathcal{D}^*_{\overline{\mathbf{S}}}$ have the same skeleton for all $\mathbf{S} \in \mathcal{S}$.*

Proof of Lemma 2. From the definition of interventional markov equivalence, for any two Markov Equivalent DAGs, $\mathcal{D}$ and $\mathcal{D}^*$, if they are not $\mathcal{S}$-Markov Equivalent, i.e., $\mathcal{D} \not\sim_{\mathcal{S}} \mathcal{D}^*$, this implies $\mathcal{MK}_{\mathcal{S}}(\mathcal{D}) \neq \mathcal{MK}_{\mathcal{S}}(\mathcal{D}^*)$, which in turn implies there exists $S \in \mathcal{S}$ such that $P_{\mathbf{S}}^{\mathcal{D}}(\mathbf{v}) \neq P_{\mathbf{S}}^{\mathcal{D}^*}(\mathbf{v})$. Also, note that for any set of nodes $\mathbf{S}$, the DAGs with incoming edges to S removed $\mathcal{D}_{\overline{\mathbf{S}}}$ and $\mathcal{D}^*_{\overline{\mathbf{S}}}$ share the same skeleton if and only if they have the same cutting edge orientations at $\mathbf{S}$, i.e., $\mathsf{C}(\mathbf{S})$. Now, considering $\mathcal{S} = \{\emptyset, \mathbf{S}\}$ and using Lemma 3, we have an equivalence relationship between statements 1 and 2, i.e., 1 $\iff$ 2. This equivalence implies that for any two Markov Equivalent DAGs, if they have different cutting-edge configurations $\mathsf{C}(\mathbf{S})$, statement 2 does not hold, which, in turn, implies that statement 1 does not hold. Consequently, $\mathcal{D} \not\sim_{\mathcal{S}} \mathcal{D}^*$, suggesting that the joint interventional distribution will differ across the two DAGs, i.e., $P_{\mathbf{S}}^{\mathcal{D}}(\mathbf{v}) \neq P_{\mathbf{S}}^{\mathcal{D}^*}(\mathbf{v})$. The converse of the previous statement, which is that if two Markov equivalent DAGs have the same interventional distribution with some target $\mathbf{S}$, i.e., $P_{\mathbf{S}}^{\mathcal{D}}(\mathbf{v}) = P_{\mathbf{S}}^{\mathcal{D}^*}(\mathbf{v})$, they must have the same cutting edge configuration at the target $\mathbf{S}$, is also true. □

5. Online Algorithm Design and Analysis

We design a data-efficient causal discovery algorithm. After initialization, the CPDAG, a graph separating system, and all possible causal effects are available. We proceed to propose a track-and-stop causal discovery algorithm that adaptively selects informative interventions. We analyze it to show it can discover the true DAG with any given confidence level $1 - \delta$ for any $\delta \in (0, 1)$. In casual discovery, reaching a confidence level $1 - \delta$ itself is not a challenging task since

the learner can take arbitrarily many interventional samples. The overarching objective is to minimize the number of interventions required to reach the accuracy level τ_δ. Since the stopping time τ_δ is random, in fact, $\mathbb{E}[\tau_\delta]$ is minimized. A sound algorithm needs to be instance-dependent, which means it is capable of detecting any DAG $\mathcal{D}^* \in [\mathcal{C}]$ if it is the ground truth. Also in line with the definition of stopping times, for a poorly designed algorithm, it is possible that $\tau_\delta = \infty$, which means the learner can never make a decision. Bringing both aspects together, a sound causal discovery algorithm is formally defined as follows.

Definition 6 (Soundness of Algorithm)**.** *For a given confidence level $\delta \in (0,1)$, a causal discovery algorithm (π, τ_δ, ψ) is sound if for any $\mathcal{D}^* \in [\mathcal{C}]$, it satisfies*

$$\mathbb{P}(\tau_\delta < \infty, \psi = \mathcal{D}^*) \geq 1 - \delta.$$

The following theorem gives a lower bound on $\mathbb{E}[\tau_\delta]$ for all sound algorithms to discover the true DAG, which serves as the ultimate target we follow in algorithm design. It has a similar form to the sampling complexity of the bandit problem, whose objective is to identify the optimal arm. The proof follows a similar procedure as (Kaufmann et al., 2016), and we defer its proof to the appendix.

Theorem 1. *For the causal discovery problem, suppose the MEC represented by CPDAG $\mathcal{C}$ and observational distributions are available. Assume that (π, τ_δ, ψ) is sound for $\mathcal{D}^*$ at confidence level $\delta \in (0,1)$. It holds that $\mathbb{E}[\tau_\delta] \geq \log(4/\delta)/c(\mathcal{D}^*)$, where*

$$c(\mathcal{D}^*) = \sup_{\boldsymbol{\alpha} \in \Delta(\mathcal{I})} \min_{\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}^*} \sum_{\mathbf{s} \in \mathcal{I}} \alpha_{\mathbf{s}} \mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}\right), \quad (3)$$

and $\Delta(\mathcal{I}) := \left\{\boldsymbol{\alpha} \in \mathbb{R}_{\geq 0}^{|\mathcal{I}|} \mid \sum_{\mathbf{s} \in \mathcal{I}} \alpha_{\mathbf{s}} = 1\right\}$.

The lower bound can be interpreted as follows. By mixing up interventions in $\mathcal{I}$ with oracle allocation $\boldsymbol{\alpha}$, the average information distance generated from $\mathcal{D}^*$ to $\mathcal{D}$ is $\sum_{\mathbf{s} \in \mathcal{I}} \alpha_{\mathbf{s}} \mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}})$. To identify the true DAG with probability at least $1-\delta$, Theorem 1 suggest at least $\log(4/\delta)$ information distance is required to be generated from $\mathcal{D}^*$ to any other $\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}^*$, which explains the minimization term in (3). The optimal parameter α^* that solves (3) suggests an optimal allocation of interventions in $\mathcal{I}$. The role of the optimal allocation α^* is to make the lower bound tight. However, computing optimal allocation requires true interventional distribution $P_{\mathbf{s}}$ for each $\mathbf{s} \in \mathcal{I}$. The allocation matching principle essentially replaces the true interventional distributions with an estimated one to compute α and select samples to match it. The key idea will be elaborated in the upcoming algorithm design section.

5.1. The Exact Algorithm

In this section, we propose the track-and-stop causal discovery whose pseudo-code is shown in Algorithm 1. It is asymptotically optimal as it achieves the $O(\log(1/\delta)/c(\mathcal{D}^*))$ lower bound in Theorem 1 on the expected number of required interventions. However, it is computationally intense. In the next section, we propose its practical implementation that reduces computational complexity at the cost of acceptable reduced efficiency.

Algorithm 1: Track-and-stop Causal Discovery

Input : CPDAG $\mathcal{C}$, δ, $\mathcal{I}$ and $(P_{\mathbf{s}})_{\mathbf{s} \in \mathcal{I}}$
Output : causal discovery result

select each intervention $\mathbf{s} \in \mathcal{I}$ once
while $f_t(d_t) \leq \delta$ *or* $d_t < |\mathcal{I}|\,(|\omega(V)| - 1)$ **do**
  compute $\boldsymbol{\alpha}_t$ via (8) (or (10))
1 **if** $\min_{\mathbf{s} \in \mathcal{I}} N_t(\mathbf{s}) < \sqrt{t}$ **then**
    % forced exploration
    select $\mathbf{s}_t = \arg\min_{\mathbf{s} \in \mathcal{I}} N_t(\mathbf{s})$
2 **else**
    % allocation matching
    select $\mathbf{s}_t = \arg\max_{\mathbf{s} \in \mathcal{I}} \sum_{i=1}^{t} \alpha_{\mathbf{s},i}/N_i(\mathbf{s})$
3 observe $\mathbf{v}_t$ and update
  $N_t(\mathbf{s}_t) \leftarrow N_t(\mathbf{s}_t) + 1$, $N_t(\mathbf{s}_t, \mathbf{v}_t) \leftarrow N_t(\mathbf{s}_t, \mathbf{v}_t) + 1$
return $\mathcal{D}_t^*$ in (4) (or $(\mathsf{C}_t^*(\mathbf{S}))_{\mathbf{S} \in \mathcal{S}}$ in (9))

Tracking and Termination Condition: Let $N_t(\mathbf{s})$ be the number of intervention $do(\mathbf{S} = \mathbf{s})$ taken till t, and let $N_t(\mathbf{s}, \mathbf{v})$ be the number of times $\mathbf{v}$ is observed by taking intervention $do(\mathbf{S} = \mathbf{s})$. The most probable DAG can be computed as

$$\mathcal{D}_t^* \in \arg\max_{\mathcal{D} \in [\mathcal{C}]} \sum_{\mathbf{s} \in \mathcal{I}} N_t(\mathbf{s}, \mathbf{v}) \log P_{\mathbf{s}}^{\mathcal{D}}(\mathbf{v}), \quad (4)$$

where $P_{\mathbf{s}}^{\mathcal{D}}$ can be computed based on the configuration of cutting edges of $\mathbf{S}$ in $\mathcal{D}$ according to Lemma 1. Let $\bar{P}_{\mathbf{s},t}(\mathbf{v}) = N_t(\mathbf{v}, \mathbf{s})/N_t(\mathbf{s})$ be the empirical distribution conditioned on taking intervention $do(\mathbf{S} = \mathbf{s})$. To evaluate if $\mathcal{D}_t^*$ has reached the confidence level $1 - \delta$, we compute

$$d_t = \min_{\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}_t^*} \sum_{\mathbf{s} \in \mathcal{I}} N_t(\mathbf{s}) \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathcal{D}}), \quad (5)$$

which is the cumulative information distance between the empirical distribution and interventional distribution from the second most probable DAG. The algorithm terminates if $f_t(d_t) < \delta$, where

$$f_t(x) = \left(\frac{x\lceil x \ln t + 1\rceil 2e}{|\mathcal{I}|\,(|\omega(V)| - 1)}\right)^{|\mathcal{I}|(|\omega(V)|-1)} e^{1-x}, \quad (6)$$

and returns $\mathcal{D}_t^*$. The function $f_t(x)$ is selected according to a concentration bound for Categorical distributions (Van Parys & Golrezaei, 2020), and it guarantees the probability of $\mathcal{D}_t^* \neq \mathcal{D}^*$ to be lower than δ.

Intervention Selection Rule: Inspired by Theorem 1, we intend to design an efficient causal discovery strategy such

that $N_t(\mathbf{s}) \approx \alpha_{\mathbf{s}} t$ for each $\mathbf{s} \in \mathcal{I}$. Since ground truth $(P^*_{\mathbf{s}})_{\mathbf{s}\in\mathcal{I}}$ is unavailable, at each time t, we use $(\bar{P}_{\mathbf{s},t})_{\mathbf{s}\in\mathcal{I}}$ instead to solve for α_t to approximate the oracle allocation $\boldsymbol{\alpha}$. To make this approach work, we need to ensure every intervention is taken a sufficient amount of times so that each $\bar{P}_{\mathbf{s},t}$ converges to the $P^*_{\mathbf{s}}$ in a fast enough rate. Accordingly, if $\min_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s}) \le \sqrt{t}$, the forced exploration step selects the least selected intervention so that it guarantees that each intervention is selected at least $\Omega(\sqrt{t})$ times.

To solve for the sequence $\{\boldsymbol{\alpha}_t\}_{t=1}^T$, we substitute $\bar{P}_{\mathbf{s},t}$ and $\mathcal{D}^*_t$ into (3) and take an online optimization procedure to

$$\underset{\forall t:\boldsymbol{\alpha}_t\in\Delta(\mathcal{I})}{\text{maximize}} \sum_{t=1}^{T} \min_{\mathcal{D}\in[\mathcal{C}]\setminus\mathcal{D}^*_t} \sum_{\mathbf{s}\in\mathcal{I}} \alpha_{\mathbf{s},t}\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P^{\mathcal{D}}_{\mathbf{s}}). \quad (7)$$

Let $\mathcal{D}'_t \in \arg\min_{\mathcal{D}\in[\mathcal{M}]\setminus\mathcal{D}^*_t} \sum_{\mathbf{s}\in\mathcal{I}} \alpha_{\mathbf{s},t}\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P^{\mathcal{D}}_{\mathbf{s}})$ and $\boldsymbol{r}_t \in \mathbb{R}^{|\mathcal{I}|}$ be a vector with entries $r_{\mathbf{s},t} = \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P^{\mathcal{D}'_t}_{\mathbf{s}})$. Note that $\boldsymbol{\alpha}_t \in \Delta(\mathcal{I})$ for all t. We follow the AdaHedge algorithm (De Rooij et al., 2014) to set

$$\alpha_{\mathbf{s},1} = \frac{1}{|\mathcal{I}|},\ \alpha_{\mathbf{s},t+1} = \frac{\alpha_{\mathbf{s},t}e^{\eta_t r_{\mathbf{s},t}}}{\sum_{\mathbf{s}\in\mathcal{I}} \alpha_{\mathbf{s},t}e^{\eta_t r_{\mathbf{s},t}}}, \forall \mathbf{s}\in\mathcal{I}, \quad (8)$$

where η_t is a decreasing learning rate with update rule

$$\eta_{t+1} = \frac{\ln K}{\Delta_t},\ \Delta_t = \sum_{i=1}^{t} \frac{1}{\eta_t} \ln\langle \boldsymbol{\alpha}_t, e^{\eta_t \boldsymbol{r}_t}\rangle - \sum_{\mathbf{s}\in\mathcal{I}} \alpha_{\mathbf{s},t} r_{\mathbf{s},t}.$$

To make $N_t(\mathbf{s})$ track $\sum_{i=1}^t \alpha_{\mathbf{s}}$, the allocation matching step selects $\arg\max_{\mathbf{s}\in\mathcal{I}} \sum_{i=1}^t \alpha_{\mathbf{s},i}/N_t(\mathbf{s})$.

Remark 1. *The proposed algorithm is computationally intensive since in equations (4) (5) and (7), it needs to enumerate DAGs in $[\mathcal{C}]$ which can be exponentially many. The worst-case computational complexity can be $\Omega(2^n)$, where n is the number of unoriented edges in $\mathcal{C}$.*

5.2. Practical Algorithm Implementation

In a practical implementation of track-and-stop causal discovery, we treat learning the configuration of edge cut $\mathsf{C}^*(\mathbf{S})$ for each node set $\mathbf{S} \in \mathcal{S}$ as an individual task, and apply a local learning strategy. The global strategy assigns allocation according to feedback from local learning results.

Local strategy: With Lemma 2, the intervention $\mathbf{S} \in \mathcal{S}$ is sufficient to learn the edge cut corresponding to $\mathbf{S}$. At time t, we compute a local allocation rule $\boldsymbol{\xi}^{\mathbf{S}}_t \in \Delta(\omega(\mathbf{S}))$ to learn the edge cut of $\mathbf{S}$. Let the most probable configuration of edge cut be computed as

$$\mathsf{C}^*_t(\mathbf{S}) = \arg\max_{\mathsf{C}(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} N_t(\mathbf{s},\mathbf{v}) \log P^{\mathsf{C}(\mathbf{S})}_{\mathbf{s}}(\mathbf{v}). \quad (9)$$

Similar to (7), we solve $\xi_{\mathbf{S},t}$ via online optimization

$$\underset{\forall t:\boldsymbol{\xi}^{\mathbf{S}}_t\in\Delta(\omega(\mathbf{S}))}{\text{maximize}} \sum_{t=1}^{T} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*_t(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},t}\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P^{\mathsf{C}(\mathbf{S})}_{\mathbf{s}}).$$

The update rule for $\boldsymbol{\xi}^{\mathbf{S}}_t \in \Delta(\omega(\mathbf{S}))$ is similar to (8). Let $\mathsf{C}'_t(\mathbf{S}) \in \arg\min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*_t(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},t}\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P^{\mathsf{C}(\mathbf{S})}_{\mathbf{s}})$ and define vector $\boldsymbol{r}^{\mathbf{S}}_t \in \mathbb{R}^{|\omega(\mathbf{S})|}$ with each entry to be $r^{\mathbf{S}}_{\mathbf{s},t} = \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P^{\mathsf{C}'_t(\mathbf{S})}_{\mathbf{s}})$. Then we set

$$\xi^{\mathbf{S}}_{\mathbf{s},1} = \frac{1}{|\omega(\mathbf{S})|},\ \xi^{\mathbf{S}}_{\mathbf{s},t+1} = \frac{\xi^{\mathbf{S}}_{\mathbf{s},t}e^{\eta_t r_{\mathbf{s},t}}}{\sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},t}e^{\eta_t r_{\mathbf{s},t}}}, \forall \mathbf{s}\in\omega(\mathbf{S}),$$

where $\eta_{t+1} = \ln K/\Delta_t$ and

$$\Delta_t = \sum_{i=1}^{t} \frac{1}{\eta_t} \ln\langle \boldsymbol{\xi}^{\mathbf{S}}_t, e^{\eta_t \boldsymbol{r}_t(\mathbf{S})}\rangle - \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},t} r^{\mathbf{S}}_{\mathbf{s},t}.$$

Global Strategy: To allocate interventions on different node sets in $\mathcal{S}$, we design a global allocation strategy $\boldsymbol{\gamma}_t \in \Delta(\mathcal{S})$ at each step. Taking feedback from $|\mathcal{S}|$ local strategies, let $c_t(\mathbf{S}) = \frac{1}{t}\sum_{i=1}^t \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},t} r^{\mathbf{S}}_{\mathbf{s},t}$. The value $1/c_t(\mathbf{S})$ corresponds to the estimated difficulty of learning the edge cut of $\mathbf{S}$. Accordingly, set $\gamma_{\mathbf{S},t} = \frac{1/c_t(\mathbf{S})}{\sum_{\mathbf{S}\in\mathcal{S}} 1/c_t(\mathbf{S})}$ and let

$$\alpha_{\mathbf{s}} = \gamma_{\mathbf{S},t}\xi^{\mathbf{S}}_{\mathbf{s},t}\ \forall \mathbf{S} \in \mathcal{I}. \quad (10)$$

Tracking and Termination: The algorithm keeps track of $(\mathsf{C}^*_t(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$ as the candidate causal discovery result. To evaluate if the confidence level δ is reached about $(\mathsf{C}^*_t(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$, for each $\mathbf{S} \in \mathcal{S}$, let

$$Z_t(\mathbf{S}) = \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*_t(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} N_t(\mathbf{s})\mathrm{KL}(P^{\mathsf{C}(\mathbf{S})}_{\mathbf{s},t} \parallel P^{\mathsf{C}^*_t(\mathbf{S})}_{\mathbf{s}}),$$

which is the minimal additional information distance by changing the edge cut of $\mathbf{S}$. Then, we set d_t to be

$$d_t = \min_{\mathbf{S}\in\mathcal{S}} Z_t(\mathbf{S}) + \sum_{\mathbf{S}\in\mathcal{S}} \sum_{\mathbf{s}\in\omega(\mathbf{S})} N_t(\mathbf{s})\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P^{\mathsf{C}^*_t(\mathbf{S})}_{\mathbf{s}}).$$

If $f_t(d_t) < \delta$, the algorithm stops and returns $(\mathsf{C}^*_t(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$. Other aspects of the algorithm remains unchanged.

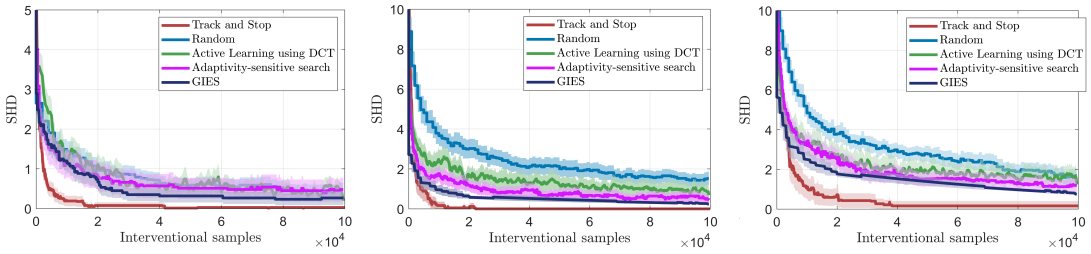
Remark 2. *Instead of enumerating all DAGs in $[\mathcal{C}]$, the practical implementation enumerates configurations of cutting edges for each $\mathbf{S} \in \mathcal{S}$. It is possible to output $(\mathsf{C}_t(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$ with contradictory edge orientations or violation of the DAG criteria. But the overall probability of $(\mathsf{C}^*_{\tau_\delta}(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$ not matching with the true DAG $\mathcal{D}^*$ is bounded by δ. If $(\mathsf{C}^*_{\tau_\delta}(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$ is not a DAG, it is suggested to reduce δ and continue the causal discovery experiment.*

5.3. An Asymptotic Analysis of Algorithm

Let $\mathcal{A}_{\mathsf{I}}$ and $\mathcal{A}_{\mathsf{P}}$ denote the exact track-and-stop causal discovery algorithm and its practical implementation, respectively. To characterize the performance for $\mathcal{A}_{\mathsf{P}}$, we define

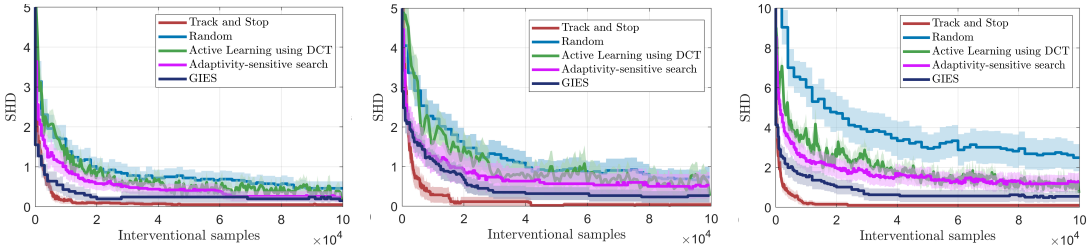
$$\underline{c}(\mathcal{D}^*) := \sup_{\boldsymbol{\alpha}\in\Delta(\mathcal{I})} \min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \alpha_{\mathbf{s}}\mathrm{KL}(P^{\mathcal{D}^*}_{\mathbf{s}} \parallel P^{\mathsf{C}(\mathbf{S})}_{\mathbf{s}}),$$

which is a lower bound for $c(\mathcal{D}^*)$ in (3).



(a) $N = 5, \rho = 1$ (b) $N = 6, \rho = 1$ (c) $N = 7, \rho = 1$

Figure 3: SHD vs interventional samples for complete random graphs with varying graph orders.



(a) $N = 10, \rho = 0.1$ (b) $N = 10, \rho = 0.15$ (c) $N = 10, \rho = 0.2$

Figure 4: SHD vs interventional samples for random graphs with varying graph density.

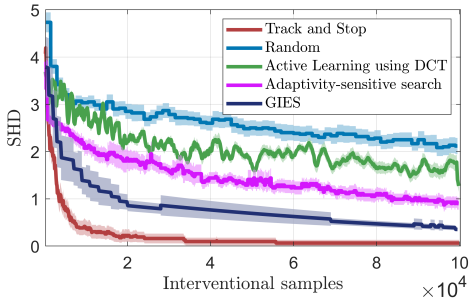


Figure 5: SHD vs No. of samples for SACHS dataset.

Theorem 2. *For the causal discovery problem, suppose the MEC represented by CPDAG $\mathcal{C}$ and observational distributions are available. If the faithfulness assumption in Definition 1 holds, for both $\mathcal{A}_{\mathsf{I}}$ and $\mathcal{A}_{\mathsf{P}}$,*

- $\mathbb{P}(\psi \neq \mathcal{D}^*) \leq \delta$ *and* $\mathbb{P}(\tau_\delta = \infty) = 0$.
- *The expected number of required interventions*

$$\lim_{\delta \to 0} \frac{\log(1/\delta)}{\mathbb{E}[\tau_\delta]} = \begin{cases} c(\mathcal{D}^*), & \mathcal{A}_{\mathsf{I}}, \\ \underline{c}(\mathcal{D}^*), & \mathcal{A}_{\mathsf{P}}, \end{cases}$$

where $c(\mathcal{D}^*) \geq \underline{c}(\mathcal{D}^*)$.

Theorem 2 establishes a problem-dependent upper bound on the expected number of required interventional samples to learn the true causal graph with a given confidence level. The output of ψ can be either $D_{\tau_\delta^*}$ or $(\mathsf{C}_{\tau_\delta}(\mathbf{S}))_{\mathbf{S} \in \mathcal{S}}$, and $\psi \neq \mathcal{D}^*$ in general means that the output does not match $\mathcal{D}^*$. The theorem shows that both $\mathcal{A}_{\mathsf{I}}$ and $\mathcal{A}_{\mathsf{P}}$ are sound, and $\mathcal{A}_{\mathsf{I}}$ archives a asymptotic performance matching with the lower bound in Theorem 1.

6. Experiments

We compare the proposed track-and-stop causal discovery algorithm with four other baselines. The first baseline involves random interventions on targets within the graph-separating system. At each time step, only one sample is collected, and statistical independence tests are used to learn the cuts at the targets based on the available samples from each intervention target within the graph-separating system. The second baseline employs Active Structure Learning of Causal DAGs via Directed Clique Trees (DCTs) (Squires et al., 2020). The third one is the adaptive sensitivity search algorithm proposed in (Choo & Shiragur, 2023). The fourth baseline is Greedy Interventional Equivalence Search (GIES), which is a generalization of Greedy Equivalence Search (GES) and is used for regularized maximum likelihood estimation in an interventional setting (Hauser & Bühlmann, 2012).

We randomly sample connected moral Directed Acyclic Graphs (DAGs) using a modified Erdős-Rényi sampling approach. A DAG whose CPDAG (Completed Partially Directed Acyclic Graph) has a single chain component is called a moral DAG. We consider moral DAGs because once we can orient moral DAGs, we can easily generalize

to general DAGs. We start by generating a random ordering σ over the vertices. Subsequently, for the n^{th} node, we sample its in-degree as $X_n = \max(1, \text{Bin}(n-1, \rho))$ and select its parents by uniformly sampling from the nodes that precede it in the ordering. In the final step, we chordalize the graph by applying the elimination algorithm (Koller & Friedman, 2009), using an elimination ordering that is the reverse of σ. This procedure is similar to the one used by (Squires et al., 2020). Finally, we randomly sample the conditional probability tables (CPTs) for all the nodes consistent with the sampled DAG and run our proposed track-and-stop discovery algorithm along with the baseline algorithms.

Figures 3 and 4 plot the Structural Hamming Distance (SHD) between the true and learned DAGs in relation to the number of interventional samples. SHD measures the number of edge additions, deletions, and reversals needed to transform one DAG into another. The shaded region represents a range of two standard deviations above and below the mean SHD. For active learning using DCT and the Adaptive-Sensitive Search algorithms, they rely on perfect interventions, i.e., an infinite number of intervention samples. However, we evaluate their performance using a limited number of intervention samples for statistical independence tests. To conduct these tests, we utilize the Chi-Square independence test available in the Causal Discovery Toolbox (Kalainathan et al., 2020).

The results in Figures 3 and 4 demonstrate that the track-and-stop algorithm outperforms other causal discovery algorithms. This is evident in the notably faster decrease in the SHD compared to the baseline algorithms. In Figure 3, we illustrate the performance of causal discovery algorithms on complete graphs with 5, 6, and 7 vertices, highlighting the superior performance of the track-and-stop algorithm compared to other baseline methods. The number of samples required by the other algorithms to achieve a low SHD increases significantly faster with the number of nodes compared to our proposed algorithm. A comparison of the plots in Figure 4(a), 4(b), and 4(c) reveals that as the DAGs become denser, the number of samples required by our algorithm to learn the DAG does not increase significantly. In contrast, the baseline algorithms are impacted by the increases in density of the graph. This is because the number of samples required by the baseline algorithms to achieve the same level of SHD increases significantly with the increase in the graph's density

We also evaluate the performance of causal discovery algorithms using the SACHS Bayesian network from the Discrete Bayesian Networks Repository in the bnlearn library (Scutari, 2009). The SACHS dataset measures the expression levels of various proteins and phospholipids in human cells (Sachs et al., 2005). The corresponding Bayesian network in the bnlearn library comprises 13 nodes and 17 edges. As depicted in Figure 5, the track-and-stop algorithm clearly outperforms the other baseline methods, resulting in significantly lower Structural Hamming Distance (SHD) for the same number of samples. The simulation results from the set of synthetic and semi-synthetic experiments establish the superior performance of the proposed track-and-stop causal algorithm compared to the baseline methods in scenarios when the interventional samples are limited. The code to reproduce our experimental results and for running the baseline algorithms and our track-and-stop discovery algorithm is available at `https://github.com/CausalML-Lab/Track-and-Stop-Discovery`.

7. Conclusion

Causal discovery aims to reconstruct the causal structure explaining the mechanism of the underlying data-generating process through observation and experimentation. It is crucial in many fields, such as economics, medicine, and social sciences, as it helps identify the underlying causes of various phenomena. Inspired by pure exploration problems in bandits, we propose a track-and-stop causal discovery algorithm that intervenes adaptively and employs a decision rule to return the most probable causal graph at any stage. We establish a problem-dependent upper bound on the expected number of interventional samples required by the algorithm. We also conduct a series of experiments on synthetic and semi-synthetic data to compare our proposed track-and-stop algorithm with existing baseline causal discovery algorithms. Our proposed algorithm outperforms baseline algorithms by requiring considerably fewer interventional samples to learn the true causal graph.

Acknowledgements

Murat Kocaoglu acknowledges the support of NSF CAREER 2239375, Amazon Research Award, and Adobe Research. Most of the work was completed when Lai Wei was a postdoctoral researcher at Purdue University.

Impact Statement

This paper presents work with the goal of advancing the field of Causality and Machine Learning. In our opinion, there are some potential societal consequences of our work, including ethical considerations related to interventions and the possibility of biased or incomplete understanding of causal relationships, which can lead to misguided decision-making or policy recommendations in real-world situations. Therefore, special attention and care are necessary for critical applications before drawing conclusions using causal discovery algorithms.

References

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

Annadani, Y., Pawlowski, N., Jennings, J., Bauer, S., Zhang, C., and Gong, W. Bayesdag: Gradient-based posterior sampling for causal discovery. *arXiv preprint arXiv:2307.13917*, 2023.

Boyd, S. P. and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004.

Choo, D. and Shiragur, K. Adaptivity complexity for causal graph discovery. *arXiv preprint arXiv:2306.05781*, 2023.

Combes, R. and Proutiere, A. Unimodal bandits: Regret lower bounds and optimal algorithms. In *International Conference on Machine Learning*, pp. 521–529. PMLR, 2014.

De Rooij, S., Van Erven, T., Grünwald, P. D., and Koolen, W. M. Follow the leader if you can, hedge if you must. *The Journal of Machine Learning Research*, 15(1):1281–1316, 2014.

Degenne, R., Koolen, W. M., and Ménard, P. Non-asymptotic pure exploration by solving games. *Advances in Neural Information Processing Systems*, 32, 2019.

Ghassami, A., Salehkaleybar, S., and Kiyavash, N. Optimal experiment design for causal discovery from fixed number of experiments. *arXiv preprint arXiv:1702.08567*, 2017.

Greenewald, K., Katz, D., Shanmugam, K., Magliacane, S., Kocaoglu, M., Boix Adsera, E., and Bresler, G. Sample efficient active learning of causal trees. *Advances in Neural Information Processing Systems*, 32, 2019.

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

Hauser, A. and Bühlmann, P. Two optimal strategies for active learning of causal models from interventional data. *International Journal of Approximate Reasoning*, 55(4): 926–939, 2014.

Heckerman, D., Meek, C., and Cooper, G. A bayesian approach to causal discovery. Technical report, Technical report msr-tr-97-05, Microsoft Research, 1997.

Hu, H., Li, Z., and Vetta, A. R. Randomized experimental design for causal graph discovery. *Advances in neural information processing systems*, 27, 2014.

Jaber, A., Kocaoglu, M., Shanmugam, K., and Bareinboim, E. Causal discovery from soft interventions with unknown targets: Characterization and learning. *Advances in neural information processing systems*, 33:9551–9561, 2020.

Kalainathan, D., Goudet, O., and Dutta, R. Causal discovery toolbox: Uncovering causal relationships in python. *The Journal of Machine Learning Research*, 21(1):1406–1410, 2020.

Katona, G. On separating systems of a finite set. *Journal of Combinatorial Theory*, 1(2):174–194, 1966.

Kaufmann, E., Cappé, O., and Garivier, A. On the complexity of best arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17:1–42, 2016.

Kocaoglu, M., Dimakis, A., and Vishwanath, S. Cost-optimal learning of causal graphs. In *International Conference on Machine Learning*, pp. 1875–1884. PMLR, 2017.

Koller, D. and Friedman, N. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

Král', D. Coloring powers of chordal graphs. *SIAM Journal on Discrete Mathematics*, 18(3):451–461, 2004.

Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020.

Meek, C. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pp. 403–410, 1995.

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

Perkovic, E. Identifying causal effects in maximally oriented partially directed acyclic graphs. In *Conference on Uncertainty in Artificial Intelligence*, pp. 530–539. PMLR, 2020.

Peters, J., Janzing, D., and Schölkopf, B. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

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

Scheines, R. An introduction to causal inference. 1997.

Scutari, M. Learning bayesian networks with the bnlearn r package. *arXiv preprint arXiv:0908.3817*, 2009.

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

Siegmund, D. *Sequential analysis: tests and confidence intervals*. Springer Science & Business Media, 1985.

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

Squires, C., Magliacane, S., Greenewald, K., Katz, D., Kocaoglu, M., and Shanmugam, K. Active structure learning of causal dags via directed clique trees. *Advances in Neural Information Processing Systems*, 33:21500–21511, 2020.

Subramani, M. and Cooper, G. Causal discovery from medical textual data, 1999.

Toth, C., Lorch, L., Knoll, C., Krause, A., Pernkopf, F., Peharz, R., and Von Kügelgen, J. Active bayesian causal inference. *Advances in Neural Information Processing Systems*, 35:16261–16275, 2022.

Van Parys, B. P. and Golrezaei, N. Optimal learning for structured bandits. *arXiv preprint arXiv:2007.07302*, 2020.

Verma, T. S. and Pearl, J. Equivalence and synthesis of causal models. In *Probabilistic and causal inference: The works of Judea Pearl*, pp. 221–236. 2022.

Wegener, I. On separating systems whose elements are sets of at most k elements. *Discrete Mathematics*, 28(2): 219–222, 1979.

Wei, L., Elahi, M. Q., Ghasemi, M., and Kocaoglu, M. Approximate allocation matching for structural causal bandits with unobserved confounders. *Advances in Neural Information Processing Systems*, 36, 2024.

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

Zhang, J. 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.

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

A. Supplementary Material

A.1. Procedure to construct (n, k) seperating system

Lemma 4 ((Shanmugam et al., 2015)). *There exists a labeling procedure that gives distinct labels of length ℓ for all elements in $[n]$ using letters from the integer alphabet $\{0, 1 \dots a\}$, where $\ell = \lceil \log_a n \rceil$. Furthermore, in every position, any integer letter is used at most $\lceil n/a \rceil$ times.*

The string labeling method in Lemma 4 from (Shanmugam et al., 2015) is described below:

Labelling Procedure: Let $a > 1$ be a positive integer. Let x be the integer such that $a^x < n \leq a^{x+1}$. $x + 1 = \lceil \log_a n \rceil$. Every element $j \in [1 : n]$ is given a label $L(j)$ which is a string of integers of length $x + 1$ drawn from the alphabet $\{0, 1, 2 \dots a\}$ of size $a + 1$. Let $n = p_d a^d + r_d$ and $n = p_{d-1} a^{d-1} + r_{d-1}$ for any integers $p_d, p_{d-1}, r_d, r_{d-1}$, where $r_d < a^d$ and $r_{d-1} < a^{d-1}$. Now, we describe the sequence of the d-th digit across the string labels of all elements from 1 to n:

1. Repeat the integer 0 a total of a^{d-1} times, and then repeat the subsequent integer, 1, also a^{d-1} times [1] from $\{0, 1 \dots a - 1\}$ till $p_d a^d$.

2. Following this, repeat the integer 0 a number of times equal to $\lceil r_d/a \rceil$, and then repeat the integer 1 $\lceil r_d/a \rceil$ times, continuing this pattern until we reach the nth position. It is evident that the nth integer in the sequence will not exceed $a - 1$.

3. Each integer that appears beyond the position $a^{d-1} p_{d-1}$ is incremented by 1.

Once we have a set of n string labels, we can easily construct a (n, k) separating system using Lemma 5, stated as follows:

Lemma 5 ((Shanmugam et al., 2015)). *Consider an alphabet $\mathcal{A} = [0 : \lceil \frac{n}{k} \rceil]$ of size $\lceil \frac{n}{k} \rceil + 1$ where $k < n/2$. Label every element of an n element set using a distinct string of letters from $\mathcal{A}$ of length $\ell = \lceil \log_{\lceil \frac{n}{k} \rceil} n \rceil$ using the labeling procedure in Lemma 4 with $a = \lceil \frac{n}{k} \rceil$. For every $1 \leq a \leq \ell$ and $1 \leq b \leq \lceil \frac{n}{k} \rceil$,we choose the subset $I_{a,b}$ of vertices whose string's a-th letter is b. The set of all such subsets $\mathcal{S} = \{s_{a,b}\}$ is a k-separating system on n elements and $|\mathcal{S}| \leq (\lceil \frac{n}{k} \rceil) \lceil \log_{\lceil \frac{n}{k} \rceil} n \rceil$.*

A.2. Meek Rules

The following algorithm can be used to apply Meek orientation rules to PDAGs.

Algorithm 2: Apply Meek Rules to a Skeleton

Function `ApplyMeekRules(`$\mathcal{M}$`)`**:**
Orient as many undirected edges as possible by repeated application of the following three rules:
(R1) Orient $b - c$ into $b \to c$ whenever there is an arrow $a \to b$ such that a and c are nonadjacent.
(R2) Orient $a - b$ into $a \to b$ whenever there is a chain $a \to c \to b$.
(R3) Orient $a - b$ into $a \to b$ whenever there are two chains $a - k \to b$ and $a - l \to b$ such that k and l are nonadjacent.
(R4) Orient $a - b$ into $a \to b$ whenever there is an edge $a - k$ and chain $k \to l \to b$ such that k and b are nonadjacent.
return A valid MPDAG: $\mathcal{M}$
End Function

A.3. Algorithms to find Partial Causal Ordering (PCO) and Enumerate all possible causal effects in the MPDAG

Definition 7. *(Bucket (Perkovic, 2020)) Consider an MPDAG $\mathcal{M}(\mathbf{V}, \mathbf{E})$ and set of vertices $\mathbf{S} \in \mathbf{V}$. The maximal undirected connected subset of S in $\mathcal{M}$ is defined as a bucket in S.*

Definition 7 permits the presence of directed edges connecting nodes within the same bucket. This definition allows for a unique decomposition, known as the bucket decomposition, to be applied to any set of vertices in the MPDAG.

[1] Circular means that after $a - 1$ is completed, we start with 0 again.

Algorithm 3: Partial Causal Ordering (Perkovic, 2020)

Function PCO $(\mathcal{M}(\mathbf{V},\mathbf{E}),\mathbf{S})$ **:**
  $\mathbf{CC}$ = Bucket decomposition of $\mathbf{V}$ in $\mathcal{M}$
  $\mathbf{B}$ = an empty list
  while $\mathbf{CC} \neq \emptyset$ **do**
    Let $\mathbf{c} \in \mathbf{CC}$ //First element in set CC
    $\overline{\mathbf{c}} = \mathbf{CC} \setminus \mathbf{c}$
    if *all edges in* $E(\mathbf{c},\overline{\mathbf{c}})$ *have a head in* $\mathbf{c}$ **then**
      $\mathbf{CC} = \overline{\mathbf{c}}$
      $\overline{B} = \mathbf{S} \cap \mathbf{c}$
      if $\overline{B} \neq \emptyset$ **then**
        Add $\overline{B}$ to the beginning of $\mathbf{B}$
      end
    end
  end
return $\mathbf{B}$ (An ordered list of Bucket Decomposition of $\mathbf{S}$)
End Function

Algorithm 4: Identify Causal Effect in an MPDAG

Input : MPDAG $\mathcal{M}(\mathbf{V},\mathbf{E})$, $\mathbf{X},\mathbf{Y} \subseteq \mathbf{V}$
Output : The interventional distribution $P(\mathbf{y}|do(\mathbf{x}))$ in MPDAG
Function `IdentifyCausalEffect` $(\mathcal{M}(\mathbf{V},\mathbf{E}),\mathbf{X},\mathbf{Y})$ **:**
  $(\mathbf{B}_1,\mathbf{B}_2,....,\mathbf{B}_m)$ =PCO$(\mathsf{An}(\mathbf{Y},\mathcal{M}_{\mathbf{V}\setminus\mathbf{X}}),\mathcal{M})$
  $\mathbf{b} = \mathsf{An}(\mathbf{Y},\mathcal{M}_{\mathbf{V}\setminus\mathbf{X}}) \setminus \mathbf{Y}$
  $P(\mathbf{y}|do(\mathbf{x})) = \sum_{\mathbf{b}} \prod_{i=1}^{m} P(\mathbf{b}_i|\mathsf{Pa}(\mathbf{b}_i,\mathcal{M}))$
return $P(\mathbf{y}|do(\mathbf{x}))$
End Function

Algorithm 5: Enumerate Causal Effect in an MPDAG

Input : MPDAG $\mathcal{M}(\mathbf{V},\mathbf{E})$, $\mathbf{X},\mathbf{Y} \subseteq \mathbf{V}$
Output : All possible interventional distribution $P(\mathbf{y}|do(\mathbf{x}))$ in MPDAG
Function `EnumerateCausalEffect` $(\mathcal{M}(\boldsymbol{V},\boldsymbol{E}),\boldsymbol{X},\boldsymbol{Y})$ **:**
  List = an empty list
  $\mathbf{E}$ = Unoriented edges in cut at $\mathbf{X}$
  if $\boldsymbol{e} = \emptyset$ **then**
    $P(\mathbf{y}|do(\mathbf{x}))$ = IdentifyCausalEffect($\mathcal{M}$, $\mathbf{X}$, $\mathbf{Y}$)
    Add $P(\mathbf{y}|do(\mathbf{x}))$ to the **List**
  else
    for *All possible orientations of edges in* $\boldsymbol{E}$ **do**
      Orient the corresponding edges $\mathbf{E}$ in $\mathcal{M}$ to get $\hat{\mathcal{M}}$
      $\bar{\mathcal{M}}$ = ApplyMeekRules($\hat{\mathcal{M}}$)
      $P(\mathbf{y}|do(\mathbf{x}))$ = IdentifyCausalEffect($\bar{\mathcal{M}}$, $\mathbf{X}$, $\mathbf{Y}$)
      Add $P(\mathbf{y}|do(\mathbf{x}))$ to the **List**
    end
  end
return List (A List of all candidate values of $P(\mathbf{y}|do(\mathbf{x}))$ for all DAGs in $[\mathcal{M}]$)
End Function

Lemma 6. *(Bucket Decomposition (Perkovic, 2020)) Consider an MPDAG* $\mathcal{M}(\mathbf{V},\mathbf{E})$ *and set of vertices* $\mathbf{S} \in \mathbf{V}$*. There exists a unique partition of* $\mathbf{S}$ *into pairwise disjoint subsets* $(\mathbf{B}_1,\mathbf{B}_2,....,\mathbf{B}_m)$ *such that* $\bigcup_{i=1}^{m} \mathbf{B}_i = \mathbf{S}$ *and* $\mathbf{B}_i$ *is a bucket of* $\mathbf{S}$ $\forall i \in [m]$.

The Algorithm 3 returns an ordered list of bucket decomposition of $\mathbf{S}$ in $\mathcal{M}$. Also ordered list of buckets output by Algorithm 3 is a partial causal ordering of $\mathbf{S}$ in $\mathcal{M}$.

Algorithm 5 provides a systematic procedure for enumerating all possible values for $P(\mathbf{y}|\mathbf{x})$ in a given MPDAG.

A.4. Proof of Lower Bound in Theorem 1

The lower bound is derived following the same strategy in (Lattimore & Szepesvári, 2020) by applying divergence decomposition and Bretagnolle–Huber inequality. For completeness, we reproduce both proofs in this section. Readers familiar with these results can skip them.

Recall that a policy π is composed of a sequence $\{\pi_t\}_{t\in\mathbb{N}_{>0}}$, where at each time $t \in \{1, \dots, T\}$, π_t determines the probability distribution of taking intervention $\mathbf{s}_t \in \mathcal{I}$ given intervention and observation history $\pi_t(\mathbf{s}_t \mid \mathbf{s}_1, \mathbf{v}_1, \dots, \mathbf{s}_{t-1}, \mathbf{v}_{t-1})$. So the intervention and observation sequence $\{\mathbf{s}_t, \mathbf{v}_t\}_{t\in\mathbb{N}_{>0}}$ is a production of the interactions between the interventional distribution tuple $(P_{\mathbf{s}})_{\mathbf{s}\in\mathcal{I}}$ and policy π. For any $T \in \mathbb{N}_{>0}$, we define a probability measure $\mathbb{P}$ on the sequence of outcomes induced by $(P_{\mathbf{s}})_{\mathbf{s}\in\mathcal{I}}$ and π such that

$$\mathbb{P}(\mathbf{s}_1, \mathbf{v}_1, \dots, \mathbf{s}_T, \mathbf{v}_T) = \prod_{t=1}^{T} \pi_t(\mathbf{s}_t \mid \mathbf{s}_1, \mathbf{v}_1, \dots, \mathbf{s}_{t-1}, \mathbf{v}_{t-1}) P_{\mathbf{s}_t}(\mathbf{v}_t). \tag{11}$$

The following decomposition is a standard result in Bandit literature (Lattimore & Szepesvári, 2020, Ch. 15).

Lemma 7 (Divergence Decomposition)**.** *In the causal discovery problem, assume $\mathcal{D}^*$ is the true DAG. for any fixed policy π, let $\mathbb{P}$ and $\mathbb{P}'$ be the probability measures corresponding to applying interventions on $\mathcal{D}^*$ and $\mathcal{D}'$, respectively. Let $\mathcal{F} = \{\mathcal{F}_t\}_{t\in\mathbb{N}_{>0}}$ be a filtration, where $\mathcal{F}_t = \sigma(\mathbf{s}_1, \mathbf{v}_1, \dots, \mathbf{s}_{t-1}, \mathbf{v}_{t-1})$, and let τ be a $\mathcal{F}$-measurable stopping time. Then for any event E that is $\mathcal{F}_\tau$ measurable,*

$$\mathrm{KL}(\mathbb{P}(E) \parallel \mathbb{P}'(E)) = \sum_{\mathbf{s}\in\mathcal{I}} \mathbb{E}[N_\tau(\mathbf{s})] \mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}),$$

where the expectation is computed with probability measure $\mathbb{P}$.

Proof. For a given sequence $\{\mathbf{s}_t, \mathbf{v}_t\}_{t\in\mathbb{N}_{>0}}$, let τ be the stopping time. Since policy π and stopping time τ are fixed, it follows from (11) that

$$\mathbb{P}(\mathbf{s}_1, \mathbf{v}_1, \dots, \mathbf{s}_\tau, \mathbf{v}_\tau) = \prod_{t=1}^{\tau} \pi_t(\mathbf{s}_t \mid \mathbf{s}_1, \mathbf{v}_1, \dots, \mathbf{s}_{t-1}, \mathbf{v}_{t-1}) P_{\mathbf{s}_t}(\mathbf{v}_t).$$

Accordingly, we define random variable

$$L_\tau := \log \frac{\mathbb{P}(\mathbf{s}_1, \mathbf{v}_1, \dots, \mathbf{s}_\tau, \mathbf{v}_\tau)}{\mathbb{P}'(\mathbf{s}_1, \mathbf{v}_t, \dots, \mathbf{s}_\tau, \mathbf{v}_\tau)} = \sum_{t=1}^{\tau} \log \frac{P_{\mathbf{s}_t}^{\mathcal{D}^*}(\mathbf{v}_t)}{P_{\mathbf{s}_t}^{\mathcal{D}'}(\mathbf{v}_t)}, \tag{12}$$

in which π_t is reduced. Equation (12) shows that the distinction between $\mathbb{P}$ and $\mathbb{P}'$ is exclusively due to the separations of $P_{\mathbf{s}}$ and $P'_{\mathbf{s}}$ for each $\mathbf{s} \in \mathcal{I}$. Let $\{\mathbf{v}_{\mathbf{s},i}\}_{i\in\mathbb{N}_{>0}}$ be the sequence of observations by applying intervention $do(\mathbf{S} = \mathbf{s})$. Then we have

$$L_\tau = \sum_{\mathbf{s}\in\mathcal{I}} \sum_{i=1}^{N_\tau(\mathbf{s})} \log \frac{P_{\mathbf{s}}(\mathbf{v}_{\mathbf{s},i})}{P'_{\mathbf{s}}(\mathbf{v}_{\mathbf{s},i})} \text{ and } \mathbb{E}\left[\log \frac{P_{\mathbf{s}}(\mathbf{v}_{\mathbf{s},i})}{P'_{\mathbf{s}}(\mathbf{v}_{\mathbf{s},i})}\right] = \mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}). \tag{13}$$

Since event E that is $\mathcal{F}_\tau$ measurable, we apply log sum inequality to get

$$\mathrm{KL}(\mathbb{P}(E) \parallel \mathbb{P}'(E)) \le \mathbb{E}\left[\log \frac{\mathbb{P}(\mathbf{s}_1, \mathbf{v}_1, \dots, \mathbf{s}_\tau, \mathbf{v}_\tau)}{\mathbb{P}'(\mathbf{s}_1, \mathbf{v}_t, \dots, \mathbf{s}_\tau, \mathbf{v}_\tau)}\right] = \mathbb{E}[L_\tau].$$

With (13), we apply Wald's Lemma (e.g. (Siegmund, 1985)) to get

$$\mathrm{KL}(\mathbb{P}(E) \parallel \mathbb{P}'(E)) \le \mathbb{E}\left[\sum_{\mathbf{s}\in\mathcal{I}} \sum_{i=1}^{N_\tau(\mathbf{s})} \log \frac{P_{\mathbf{s}}(\mathbf{v}_{\mathbf{s},i})}{P'_{\mathbf{s}}(\mathbf{v}_{\mathbf{s},i})}\right] = \sum_{\mathbf{s}\in\mathcal{I}} \mathbb{E}[N_\tau(\mathbf{s})] \mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}),$$

which concludes the proof. □

The other tool to prove the regret lower bound is the Bretagnolle–Huber Inequality.

Lemma 8 (Bretagnolle–Huber Inequality (Lattimore & Szepesvári, 2020), Th 14.2)**.** *Let P and Q be two probability measures on a measurable space $(\omega, \mathcal{F})$, and let $E \in \mathcal{F}$ be an arbitrary event. Then*

$$P(E) + Q(E^{\complement}) \geq \frac{1}{2}\exp\left(-\mathrm{KL}(P \parallel Q)\right),$$

where $E^{\complement} = \omega \setminus E$ is complement of E.

Proof of Theorem 1. If $\mathbb{E}[\tau_\delta] = \infty$, the result is trivial. Assume that $\mathbb{E}[\tau_\delta] < \infty$, which also indicates $\mathbb{P}(\tau_\delta = \infty) = 0$. Recall $\mathbb{P}$ and $\mathbb{P}'$ are the probability measures corresponding to applying interventions on $\mathcal{D}^*$ and $\mathcal{D}'$ respectively. Define event $E = \{\tau_\delta \leq \infty, \psi \neq \mathcal{D}'\}$. For a sound casual discovery algorithm, we have

$$2\delta \geq \mathbb{P}\big(\tau_\delta \leq \infty, \psi \neq \mathcal{D}^*\big) + \mathbb{P}'\big(\tau_\delta \leq \infty, \psi \neq \mathcal{D}'\big) \tag{14}$$

$$\geq \mathbb{P}\big(E^{\complement}\big) + \mathbb{P}'(E). \tag{15}$$

We apply Bretagnolle–Huber inequality to get

$$2\delta \geq \frac{1}{2}\exp\left(-\mathrm{KL}(\mathbb{P}(E) \parallel \mathbb{P}'(E))\right). \tag{16}$$

With Lemma 7, we substitute $\sum_{\mathbf{s}\in\mathcal{I}} \mathbb{E}[N_T(\mathbf{s})]\mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}'})$ into $\mathrm{KL}(\mathbb{P}(E) \parallel \mathbb{P}'(E))$ and rearrange (16) to get

$$\log\frac{4}{\delta} \leq \sum_{\mathbf{s}\in\mathcal{I}} \mathbb{E}[N_{\tau_\delta}(\mathbf{s})]\mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}'}) \leq \mathbb{E}[\tau_\delta] \sum_{\mathbf{s}\in\mathcal{I}} \frac{\mathbb{E}[N_{\tau_\delta}(\mathbf{s})]}{\mathbb{E}[\tau_\delta]}\mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}'}) \leq \mathbb{E}[\tau_\delta]c(\mathcal{D}^*). \tag{17}$$

Since (17) holds for any $\mathcal{D}' \in [\mathcal{C}] \setminus \mathcal{D}^*$, we have

$$\begin{aligned}\log\frac{4}{\delta} &\leq \mathbb{E}[\tau_\delta] \min_{\mathcal{D}\in[\mathcal{C}]\setminus\mathcal{D}^*} \sum_{\mathbf{s}\in\mathcal{I}} \frac{\mathbb{E}[N_{\tau_\delta}(\mathbf{s})]}{\mathbb{E}[\tau_\delta]}\mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}) \\ &\leq \mathbb{E}[\tau_\delta] \max_{\boldsymbol{\alpha}\in\Delta(\mathcal{I})} \min_{\mathcal{D}\in[\mathcal{C}]\setminus\mathcal{D}^*} \sum_{\mathbf{s}\in\mathcal{I}} \alpha_{\mathbf{s}}\mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}\right) = \mathbb{E}[\tau_\delta]c(\mathcal{D}^*),\end{aligned}$$

where the last inequality is due to $\sum_{\mathbf{s}\in\mathcal{I}} \mathbb{E}[N_{\tau_\delta}(\mathbf{s})]/\mathbb{E}[\tau_\delta] = 1$. We conclude the proof. □

A.5. Supporting Lemmas for Theorem 2

A.5.1. Supporting Lemmas on Online Maxmin Optimization

In (3), we define a variable $c(\mathcal{D}^*) = \sup_{\boldsymbol{\alpha}\in\Delta(\mathcal{I})} \min_{\mathcal{D}\in[\mathcal{C}]\setminus\mathcal{D}^*} \sum_{\mathbf{s}\in\mathcal{I}} \alpha_{\mathbf{s}}\mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}\right)$. We define another variable for the local discovery result

$$c_{\mathbf{S}}(\mathcal{D}^*) = \sup_{\boldsymbol{\xi}^{\mathbf{S}}\in\Delta(\omega(\mathbf{S}))} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi_{\mathbf{s}}^{\mathbf{S}}\mathrm{KL}(P_{\mathbf{s}}^{\mathsf{C}^*(\mathbf{S})} \parallel P_{\mathbf{s},t}^{\mathsf{C}(\mathbf{S})}). \tag{18}$$

Let $\mathsf{CF}(\mathbf{S})$ denote all the possible configurations of cutting edges attached to the node set $\mathbf{S}$. The following theorem shows these values from the maxmin optimization equal to their minmax counterparts.

Lemma 9. *The following two inequality holds:*

$$c(\mathcal{D}^*) = \inf_{\boldsymbol{w}\in\Delta([\mathcal{C}]\setminus\mathcal{D}^*)} \max_{\mathbf{s}\in\mathcal{I}} \sum_{\mathcal{D}\in[\mathcal{C}]\setminus\mathcal{D}^*} w_{\mathcal{D}}\mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}\right),$$

$$c_{\mathbf{S}}(\mathcal{D}^*) = \inf_{\boldsymbol{\zeta}^{\mathbf{S}}\in\Delta(\mathsf{CF}(\mathbf{S})\setminus\mathsf{C}^*(\mathbf{S}))} \max_{\mathbf{s}\in\omega(\mathbf{S})} \sum_{\mathsf{C}(\mathbf{S})\in\mathsf{CF}(\mathbf{S})\setminus\mathsf{C}^*(\mathbf{S})} \zeta_{\mathsf{C}(\mathbf{S})}^{\mathbf{S}}\mathrm{KL}(P_{\mathbf{s}}^{\mathsf{C}^*(\mathbf{S})} \parallel P_{\mathbf{s},t}^{\mathsf{C}(\mathbf{S})}), \forall \mathbf{S} \in \mathcal{S}.$$

Besides,

$$\underline{c}(\mathcal{D}^*) = \sup_{\boldsymbol{\alpha}\in\Delta(\mathcal{I})} \min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \alpha_{\mathbf{s}}\mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) = \gamma_{\mathbf{S}}^* c_{\mathbf{S}}(\mathcal{D}^*), \text{ where } \gamma_{\mathbf{S}}^* = \frac{1/c_{\mathbf{S}}(\mathcal{D}^*)}{\sum_{\mathbf{S}\in\mathcal{S}} 1/c_{\mathbf{S}}(\mathcal{D}^*)}.$$

Sketch of Proof. These two max-min optimization problems correspond to designing a mixed strategy for matrix games. To elaborate, the reward matrix $R \in \mathbb{R}^{(|[\mathcal{C}]|-1)\times|\mathcal{I}|}$ has entries represented as $\mathrm{KL}\left(P^{\mathcal{D}^*}\mathbf{s} \parallel P^{\mathcal{D}}\mathbf{s}\right)$, and solving for $c(\mathcal{D})$ is equivalent to the following optimization problem:

$$\begin{aligned} &\text{maximize} \min_{i=\{1,\ldots,|[\mathcal{C}]|-1\}} (R\boldsymbol{\alpha})_i \\ &\text{subject to} \quad \alpha \succeq 1, \mathbf{1}^T\boldsymbol{\alpha} = 1. \end{aligned}$$

For such a problem, it is shown in (Boyd & Vandenberghe, 2004, CH 5.2.5) that strong duality holds. Similar argument can be made on $c_{\mathbf{S}}(\mathcal{D}^*)$. Detailed proofs are omitted.

To prove the last equality, let $\gamma_{\mathbf{S}} = \sum_{\mathbf{s}\in\omega(\mathbf{S})} \alpha_{\mathbf{s}}$ and let $\alpha_{\mathbf{s}} = \gamma_{\mathbf{S}}\xi_{\mathbf{s}}^{\mathbf{S}}$. We also have $\sum_{\mathbf{S}\in\mathcal{S}} \gamma_{\mathbf{S}} = 1$ and $\sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi_{\mathbf{s}}^{\mathbf{S}} = 1$. It follows that,

$$\begin{aligned} \underline{c}(\mathcal{D}^*) &= \sup_{\boldsymbol{\alpha}\in\Delta(\mathcal{I})} \min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \alpha_{\mathbf{s}} \mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) \\ &= \sup_{\{\gamma_{\mathbf{S}}\}_{\mathbf{S}\in\mathcal{S}}} \sup_{\{\xi_{\mathbf{s}}^{\mathbf{S}}\}_{\mathbf{s}\in\omega(\mathbf{S})}} \min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \gamma_{\mathbf{S}}\xi_{\mathbf{s}}^{\mathbf{S}} \mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) \\ &= \sup_{\{\gamma_{\mathbf{S}}\}_{\mathbf{S}\in\mathcal{S}}} \gamma_{\mathbf{S}} \min_{\mathbf{S}\in\mathcal{S}} \sup_{\{\xi_{\mathbf{s}}^{\mathbf{S}}\}_{\mathbf{s}\in\omega(\mathbf{S})}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi_{\mathbf{s}}^{\mathbf{S}} \mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) \\ &= \sup_{\{\gamma_{\mathbf{S}}\}_{\mathbf{S}\in\mathcal{S}}} \gamma_{\mathbf{S}} \min_{\mathbf{S}\in\mathcal{S}} c_{\mathbf{S}}(\mathcal{D}^*). \end{aligned}$$

Besides, the solution for above problem satisfies $\gamma_{\mathbf{S}} \propto 1/c_{\mathbf{S}}(\mathcal{D}^*)$. We conclude the proof. □

A.5.2. Supporting Lemma for AdaHedge Algorithm

The AdaHedge deals with such a sequential decision-making problem. At each $t = 1, 2, \ldots$, the learner needs to decide a weight vector $\boldsymbol{\alpha}_t = (\alpha_{1,t}, \ldots, \alpha_{K,t})$ over K "experts". Nature then reveals a K-dimensional vector containing the rewards of the experts $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{K,t}) \in \mathbb{R}^K$. The actual received reward is the dot product $h_t = \boldsymbol{\alpha}_t \cdot \boldsymbol{r}_t$, which can be interpreted as the expected loss with a mixed strategy. The learn's task it to maximize the cumulative reward $H_T = \sum_{t=1}^T h_t$ or equivalently minimize the regret defined as

$$R_T = \max_{k\in\{1,\ldots,K\}} \sum_{t=1}^T r_{k,t} - H_T.$$

The performance guarantee of AdaHedge is as follows.

Lemma 10 ((De Rooij et al., 2014)). *If for any $t \in \mathbb{N}_{>0}$, $r_{k,t} \in [0, D]$ for all $k \in \{1, \ldots, K\}$, let R_T^{AH} be the regret for AdaHedge for horizon T. It satisfies that*

$$R_T^{\mathrm{AH}} \leq \sqrt{DT \ln K} + D\left(\frac{4}{3}\ln K + 2\right).$$

The following lemma is also used in the proof of Theorem 2.

Lemma 11. *If for any $t \in \mathbb{N}_{>0}$, $r_{k,t} \in [0, D]$ for all $k \in \{1, \ldots, K\}$, for any $T \geq \tau > 0$,*

$$\max_{\mathbf{S}\in\mathcal{I}} \sum_{t=\tau+1}^T r_{\mathbf{s},t} - \sum_{t=\tau+1}^T h_t \geq R_T - \tau D.$$

Proof. We apply the fact that $max_{k\in\{1,\ldots,K\}} \sum_{t=1}^T r_{k,t} \leq \tau D + \max_{k\in\{1,\ldots,K\}} \sum_{t=\tau+1}^T r_{\mathbf{s},t}$.

$$\begin{aligned} R_T &= \max_{k\in\{1,\ldots,K\}} \sum_{t=1}^T r_{k,t} - \sum_{t=1}^T h_t \leq \tau D + \max_{k\in\{1,\ldots,K\}} \sum_{t=\tau+1}^T r_{\mathbf{s},t} - \sum_{t=1}^T h_t \\ &\leq \tau D + \max_{\mathbf{S}\in\mathcal{I}} \sum_{t=\tau+1}^T r_{\mathbf{s},t} - \sum_{t=\tau+1}^T h_t, \end{aligned}$$

which concludes the proof. □

In the exact version of track-and-stop causal discovery algorithm $\mathcal{A}_{\mathsf{I}}$, the AdaHege is run with $|\mathcal{I}|$ dimensional reward vector $(r_{\mathbf{s},t})_{\mathbf{s}\in\mathcal{I}}$ with entries $r_{\mathbf{s},t} = \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathcal{D}'_t})$, where

$$\mathcal{D}'_t \in \operatorname*{arg\,min}_{\mathcal{D}\in[\mathcal{M}]\setminus\mathcal{D}_t^*} \sum_{\mathbf{s}\in\mathcal{I}} \alpha_{\mathbf{s},t}\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathcal{D}}).$$

In the practical algorithm $\mathcal{A}_{\mathsf{P}}$, for each $\mathbf{S}\in\mathcal{S}$, the AdaHege is run to compute $\boldsymbol{\xi}_t^{\mathbf{S}}$. The feedback is $\omega(\mathbf{S})$ dimensional vector $(r_{\mathbf{s},t}^{\mathbf{S}})_{\mathbf{s}\in\omega(\mathbf{S})}$ with entries $r_{\mathbf{s},t}^{\mathbf{S}} = \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathsf{C}'_t(\mathbf{S})})$, where

$$\mathsf{C}'_t(\mathbf{S}) \in \operatorname*{arg\,min}_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}_t^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi_{\mathbf{s},t}^{\mathbf{S}}\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}).$$

Also in our setup $D = \max_{\mathcal{D}\in[\mathcal{C}]} \sup_{P_{\mathbf{s}}} \mathrm{KL}(P_{\mathbf{s}} \parallel P_{\mathbf{s}}^{\mathcal{D}})$, where KL-divergence follows the convention that $0\log 0 = 0$ $\log 0/0 = 0$ and $x\log x/0 = +\infty$ for $x > 0$.

A.5.3. Supporting Lemmas on Allocation Matching

Lemma 12. *For the track-and-stop causal discovery algorithm, for any $t \geq |\mathcal{I}|$ and any $\mathbf{s}\in\mathcal{I}$*

$$\sum_{i=1}^{t} \alpha_{\mathbf{s},i} - (|\mathcal{I}|-1)(\sqrt{t}+2) \leq N_t(\mathbf{s}) \leq \max\Big\{1 + \sum_{i=1}^{t} \alpha_{\mathbf{s},i}, \sqrt{t}+1\Big\}.$$

Proof. We first show that for any $t \geq \mathcal{I}$, the following is true.

$$N_t(\mathbf{s}) \leq \max\Big\{1 + \sum_{i=1}^{t} \alpha_{\mathbf{s},i}, \sqrt{t}+1\Big\}. \tag{19}$$

We prove this claim by induction. At time $t' = \mathcal{I}$, $N_{t'}(\mathbf{s}) = 1$ for all $\mathbf{s}\in\mathcal{I}$, so that (19) is true. Suppose $N_{t'}(\mathbf{s}) \leq \max\Big\{1 + \sum_{i=1}^{t'} \alpha_{\mathbf{s},i}, \sqrt{t}+1\Big\}$ is true. If $do(\mathbf{s})$ is not selected at $t'+1$, we have

$$N_{t'+1}(\mathbf{s}) = N_{t'}(\mathbf{s}) \leq \max\Big\{1 + \sum_{i=1}^{t'+1} \alpha_{\mathbf{s},i}, \sqrt{t'+1}+1\Big\}. \tag{20}$$

If $do(\mathbf{s})$ is selected at $t'+1$ by force exploration, we have

$$N_{t'+1}(\mathbf{s}) = N_{t'}(\mathbf{s}) + 1 < \sqrt{t+1} + 1. \tag{21}$$

If $do(\mathbf{s})$ is selected at $t'+1$ by allocation matching, since $\sum_{\mathbf{s}\in\mathcal{I}}\sum_{i=1}^{t} \alpha_{\mathbf{s},i} = t$ and $\sum_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s}) = t$, we have

$$\min_{\mathbf{s}\in\mathcal{I}} \frac{N_t(\mathbf{s})}{\sum_{i=1}^{t}\alpha_{\mathbf{s},i}} \leq 1.$$

Accordingly,

$$\frac{N_{t+1}(\mathbf{s}_t)}{\sum_{i=1}^{t+1}\alpha_{\mathbf{s},i}} = \frac{N_t(\mathbf{s}_t)}{\sum_{i=1}^{t+1}\alpha_{\mathbf{s},i}} + \frac{1}{\sum_{i=1}^{t+1}\alpha_{\mathbf{s},i}} \leq 1 + \frac{1}{\sum_{i=1}^{t}\alpha_{\mathbf{s},i}}. \tag{22}$$

Combining (20) (21) (22), we show (19) is true. Also notice that for all $\mathbf{s}\in\mathcal{I}$,

$$N_t(\mathbf{s}) \leq \max\Big\{1 + \sum_{i=1}^{t} \alpha_{\mathbf{s},i}, \sqrt{t}+1\Big\} \leq \sqrt{t}+2 + \sum_{i=1}^{t} \alpha_{\mathbf{s},i}.$$

It follows from that $\sum_{\mathbf{s}\in\mathcal{I}}\sum_{i=1}^{t} \alpha_{\mathbf{s},i} = t$ and $\sum_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s}) = t$,

$$N_t(\mathbf{s}) \geq \sum_{i=1}^{t} \alpha_{\mathbf{s},i} - (|\mathcal{I}|-1)(\sqrt{t}+2).$$

We conclude the proof. □

A.5.4. Supporting Lemmas on Concentration Inequality of Empirical Mean

The following Lemma 13 proposed in (Combes & Proutiere, 2014) extends Hoeffding's inequality to provide an upper bound on the deviation of the empirical mean sampled at a stopping time. In our problem, each time the intervention is selected is a stopping time.

Lemma 13 (Extension of Hoeffding's Inequality (Combes & Proutiere, 2014), Lemma 4.3)**.** *Let $\{Z_t\}_{t\in\mathbb{N}_{>0}}$ be a sequence of independent random variables with values in $[0,1]$. Let $\mathcal{F}_t$ be the σ-algebra such that $\sigma(Z_1,\dots,Z_t)\subset\mathcal{F}_t$ and the filtration $\mathcal{F}=\{\mathcal{F}_t\}_{t\in\mathbb{N}_{>0}}$. Consider $s\in\mathbb{N}$, and $T\in\mathbb{N}_{>0}$. We define $S_t=\sum_{j=1}^t\epsilon_j(Z_j-\mathbb{E}[Z_j])$, where $\epsilon_j\in\{0,1\}$ is a $\mathcal{F}_{j-1}$-measurable random variable. Further define $N_t=\sum_{j=1}^t\epsilon_j$. Define $\phi\in\{1,\dots,T+1\}$ a $\mathcal{F}$-stopping time such that either $N_\phi\geq s$ or $\phi=T+1$. Then we have that*

$$P[S_\phi\geq N_\phi\delta]\leq\exp(-2s\delta^2).$$

As a consequence,

$$P[|S_\phi|\geq N_\phi\delta]\leq 2\exp(-2s\delta^2).$$

In Corollary 1, we extend Lemma 13 to bound the L_1 deviation of the empirical distribution.

Corollary 1 (L_1 deviation of the empirical distribution)**.** *Let $\mathcal{A}$ denote finite set $\{1,\dots,a\}$. For two probability distribution Q and Q' on $\mathcal{A}$, let $\|Q'-Q\|_1=\sum_{k=1}^a|Q'(k)-Q(k)|$. Let $X_t\in\mathcal{A}$ be a sequence of independent random variables with common distribution Q. Let $\mathcal{F}_t$ be the σ-algebra such that $\sigma(X_1,\dots,X_t)\subset\mathcal{F}_t$ and the filtration $\mathcal{F}=\{\mathcal{F}_t\}_{t\in\mathbb{N}_{>0}}$. Let $\epsilon_t\in\{0,1\}$ be a $\mathcal{F}_{t-1}$-measurable random variable. We define*

$$N_t=\sum_{j=1}^t\epsilon_j,S_t(i)=\sum_{j=1}^t\epsilon_j\mathbf{1}\left\{X_j=i\right\},\text{ and }\bar{Q}_t(i)=\frac{S_t(i)}{N_t},\forall i\in\mathcal{A}.$$

For $s\in\mathbb{N}$, and $T\in\mathbb{N}_{>0}$, let $\phi\in\{1,\dots,T+1\}$ be a $\mathcal{F}$-stopping time such that either $N_\phi\geq s$ or $\phi=T+1$. Then we have

$$P\left(\left\|\bar{Q}_\phi-Q\right\|_1\geq\delta\right)\leq(2^a-2)\exp\left(\frac{-s\delta^2}{2}\right).$$

Proof. It is known that for any distribution Q' on $\mathcal{A}$,

$$\|Q'-Q\|_1=2\max_{A\subseteq\mathcal{A}}(Q'(A)-Q(A)).$$

Then we apply a union bound to get

$$\begin{aligned}P\left(\left\|\bar{Q}_\phi-Q\right\|_1\geq\delta\right)&\leq\sum_{A\subseteq\mathcal{A}}P\left(\bar{Q}_\phi(A)-Q(A)\geq\frac{\delta}{2}\right)\\&\leq\sum_{A\subseteq\mathcal{A}:A\neq\mathcal{A}\text{ or }\emptyset}P\left(\bar{Q}_\phi(A)-Q(A)\geq\frac{\delta}{2}\right)\\&\leq(2^a-2)\exp\left(\frac{-s\delta^2}{2}\right),\end{aligned}$$

which concludes the proof. □

Corollary 2. *For the causal discovery problem with the track-and-stop algorithm and any $\epsilon>0$, define the random time*

$$\tau_p(\epsilon)=\max\left\{t\in\mathbb{N}_{>0}\;\middle|\;\exists\mathbf{s}\in\mathcal{I}:\left\|\bar{P}_{\mathbf{s},t}-P_{\mathbf{s}}^{\mathcal{D}^*}\right\|_1>\epsilon\right\}.$$

Then there exists a constant $c(\epsilon)>0$ such that $\mathbb{E}[\tau_p(\epsilon)]\leq c(\epsilon)$.

Proof. The forced exploration step guarantees that each intervention is selected at least $\Omega(\sqrt{t})$ times at time t. To show that, we first note that the following two facts are true:

- $\min_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s})$ is non-decreasing over t.
- If $\min_{\mathbf{s}\in\mathcal{I}} N_{t_i}(\mathbf{s}) < \sqrt{i}$, then $\min_{\mathbf{s}\in\mathcal{I}} N_{t_{i+|\mathcal{I}|}}(\mathbf{s}) \geq \min N_{t_i}(\mathbf{s}) + 1$.

Since $N_t(\mathbf{s})$ for each $\mathbf{s} \in \mathcal{I}$ is non-decreasing over t, the first statement is true. The second statement is true since otherwise, after at least $|\mathcal{I}|$ forced exploration steps, $\min_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s})$ does not increase. With these two facts, we are ready to show for any $\alpha \in (0,1)$ and $t \geq \alpha |\mathcal{I}|^2/(1-\alpha)^2$, $\min_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s}) \geq \sqrt{\alpha t}$. The proof is provided by contradiction. Suppose there exists time step i such that

$$\min_{\mathbf{s}\in\mathcal{I}} N_i(\mathbf{s}) < \sqrt{\alpha i}.$$

According to the first fact, we have for any $j \geq \alpha i$

$$\min_{\mathbf{s}\in\mathcal{I}} N_j(\mathbf{s}) \leq \min_{\mathbf{s}\in\mathcal{I}} N_i(\mathbf{s}) < \sqrt{\alpha i}.$$

Then we apply the second fact. For any $i \geq \alpha |\mathcal{I}|^2/(1-\alpha)^2$, we have

$$\min_{\mathbf{s}\in\mathcal{I}} N_i(\mathbf{s}) \geq \frac{i-j}{|\mathcal{I}|} \geq \frac{(1-\alpha)i}{|\mathcal{I}|} \geq \sqrt{\alpha i},$$

which creates a contradiction.

To show $\mathbb{E}[\tau_p(\epsilon)] \leq c$, we first notice that

$$\mathbb{P}(\tau_p(\epsilon) \geq x) = \mathbb{P}(\exists t \geq x : \exists \mathbf{s} \in \mathcal{I} : \left\|\bar{P}_{\mathbf{s},t} - P_{\mathbf{s}}\right\|_1 \geq \epsilon) \leq \sum_{\mathbf{s}\in\mathcal{I}} \sum_{t\geq x} \mathbb{P}(\left\|\bar{P}_{\mathbf{s},t} - P_{\mathbf{s}}\right\|_1 \geq \epsilon),$$

where the last inequality is from the union bound. Accordingly, for any $x \geq \alpha |\mathcal{I}|^2/(1-\alpha)^2$, we apply Corollary 1 to get

$$\begin{aligned}
\mathbb{P}(\tau_p(\epsilon) \geq x) &\leq (2^{|\omega(\mathbf{V})|} - 2)\,|\mathcal{I}| \sum_{t\geq x} \exp\Big(\frac{-\sqrt{\alpha t}\epsilon^2}{2}\Big) \\
&\leq (2^{|\omega(\mathbf{V})|} - 2)\,|\mathcal{I}| \int_{x-1}^{+\infty} \exp\Big(\frac{-\sqrt{\alpha t}\epsilon^2}{2}\Big) dx \\
&= (2^{|\omega(\mathbf{V})|} - 2)\,|\mathcal{I}| \frac{8}{\alpha\epsilon^4} \exp\Big(\frac{-\sqrt{\alpha(x-1)}\epsilon^2}{2}\Big)\Big(\frac{\sqrt{\alpha(x-1)}\epsilon^2}{2} + 1\Big)
\end{aligned}$$

Let $\beta = \alpha |\mathcal{I}|^2/(1-\alpha)^2$. It follows that

$$\begin{aligned}
\mathbb{E}[\tau_p(\epsilon)] &\leq \beta + 1 + \int_{\beta+1}^{+\infty} \mathbb{P}(\tau_p(\epsilon) \geq x) dx \\
&\leq \beta + 1 + (2^{|\omega(\mathbf{V})|} - 2)\,|\mathcal{I}| \frac{64}{\alpha^2\epsilon^8} \exp\Big(\frac{-\sqrt{\alpha\beta}\epsilon^2}{2}\Big)\Big(\frac{\alpha\beta\epsilon^4}{4} + \frac{3\sqrt{\alpha\beta}\epsilon^2}{2} + 3\Big) := g(\epsilon, \alpha).
\end{aligned}$$

Taking $c(\epsilon) = \inf_{\alpha\in(0,1)} g(\epsilon, \alpha)$, we conclude the proof. □

A.6. Proof of Theorem 2

We decompose Theorem 2 into Lemmas 15, 16 and 18 and prove them in separate sections.

A.6.1. Accuracy of the Track-and-stop Causal Discovery Algorithm

In this section, we prove that for any $\delta \in (0,1)$, the confidence level $1-\delta$ can be reached by the track-and-stop causal discovery algorithm (exact and practical version). The following concentration inequality is crucial in the proof. For an active learning setup with feedback drawn from Categorical distributions, a concentration bound on the empirical distribution is presented in Lemma 6 of (Van Parys & Golrezaei, 2020). In the causal discovery problem, the actions space is $\mathcal{I}$, and the discrete support of feedback is $\omega(\mathbf{V})$. At each time t, for each intervention $\mathbf{s} \in \mathcal{I}$, recall $\bar{P}_{\mathbf{s},t}$ is the empirical interventional distribution of $\mathbf{V}$ and $N_t(\mathbf{s})$ is the number of times the intervention $do(\mathbf{S} = \mathbf{s})$ is taken till t. For each intervention $\mathbf{s} \in \mathcal{I}$, the true interventional distribution is $P_{\mathbf{s}}^{\mathcal{D}^*}$.

Lemma 14 (Concentration Inequality for Information Distance (Van Parys & Golrezaei, 2020)). *Let $x \geq |\mathcal{I}|\left(|\omega(V)|-1\right)$. Then for any $t>0$,*

$$\mathbb{P}\left[\sum_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s})\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathcal{D}^*}) \geq x\right] \leq \left(\frac{x\lceil x\ln t+1\rceil 2e}{|\mathcal{I}|\left(|\omega(V)|-1\right)}\right)^{|\mathcal{I}|(|\omega(V)|-1)} \exp(1-x).$$

Lemma 15. *For the causal discovery problem with the MEC represented by CPDAG $\mathcal{C}$ and observational distributions being available, if the faithfulness assumption in Definition 1 holds, for both $\mathcal{A}_{\mathsf{I}}$ and $\mathcal{A}_{\mathsf{P}}$, $\mathbb{P}(\psi \neq \mathcal{D}^*) \leq \delta$.*

Proof. The track-and-stop causal discovery algorithm keeps track of the most probable DAG

$$\mathcal{D}_t^* = \arg\max_{\mathcal{D}\in[\mathcal{C}]} \sum_{\mathbf{s}\in\omega(\mathbf{S})} N_t(\mathbf{s},\mathbf{v}) \log P^{\mathcal{D}}(\mathbf{v}).$$

For $\mathcal{A}_{\mathsf{I}}$, at stopping time τ_δ, by the design if $\mathcal{D}_{\tau_\delta}^* \neq \mathcal{D}^*$, we have

$$d_{\tau_\delta} = \min_{\mathcal{D}\in[\mathcal{C}]\setminus\mathcal{D}_{\tau_\delta}^*} \sum_{\mathbf{s}\in\mathcal{I}} N_{\tau_\delta}(\mathbf{s})\mathrm{KL}(\bar{P}_{\tau_\delta,\mathbf{s}} \parallel P_{\mathbf{s}}^{\mathcal{D}}) \leq \sum_{\mathbf{s}\in\mathcal{I}} N_{\tau_\delta}(\mathbf{s})\mathrm{KL}(\bar{P}_{\tau_\delta,\mathbf{s}} \parallel P_{\mathbf{s}}^{\mathcal{D}^*}).$$

Then we apply Lemma 14 to get

$$\mathbb{P}\left[\psi \neq \mathcal{D}^*\right] \leq \mathbb{P}\left[d_{\tau_\delta} \leq \sum_{\mathbf{s}\in\mathcal{I}} N_{\tau_\delta}(\mathbf{s})\mathrm{KL}(\bar{P}_{\tau_\delta,\mathbf{s}} \parallel P_{\mathbf{s}}^{\mathcal{D}^*})\right] \leq f_{\tau_\delta}(d_\tau) \leq \delta, \tag{23}$$

where the last inequality is due to the termination condition of the algorithm.

With $\mathcal{A}_{\mathsf{P}}$, instead of searching $\mathcal{D}^*$ in $[\mathcal{C}]$, we search $(\mathsf{C}^*(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$ in the space $(\mathsf{CF}(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$. Recall $Z_t(\mathbf{S}) = \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}_t^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} N_t(\mathbf{s})\mathrm{KL}(P_{\mathbf{s},t}^{\mathsf{C}(\mathbf{S})} \parallel P_{\mathbf{s}}^{\mathsf{C}_t^*(\mathbf{S})})$. As a matter of fact,

$$\begin{aligned} d_t &= \min_{\mathbf{S}\in\mathcal{S}} Z_t(\mathbf{S}) + \sum_{\mathbf{S}\in\mathcal{S}}\sum_{\mathbf{s}\in\omega(\mathbf{S})} N_t(\mathbf{s})\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathsf{C}_t^*(\mathbf{S})}) \\ &= \min_{(\mathsf{C}_{\tau_\delta}^*(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}\neq(\mathsf{C}^*(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}} \sum_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s})\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathsf{C}_t^*(\mathbf{S})}). \end{aligned} \tag{24}$$

If $(\mathsf{C}_{\tau_\delta}^*(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}} \neq (\mathsf{C}^*(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$, following a similar reasoning, it can be seen (23) still holds. We conclude the proof. □

A.6.2. Asymptotic performance of Exact Algorithm

Lemma 16. *For the causal discovery problem, suppose the MEC represented by CPDAG $\mathcal{C}$ and observational distributions are available. If the faithfulness assumption in Definition 1 holds, for the exact algorithm $\mathcal{A}_{\mathsf{I}}$, we have $\mathbb{P}(\tau_\delta=\infty)=0$ and*

$$\lim_{\delta\to 0} \frac{\log(1/\delta)}{\mathbb{E}[\tau_\delta]} = c(\mathcal{D}^*).$$

Proof. Let an arbitrary intervention distribution tuple be $\mathcal{P}=(P_{\mathbf{s}})_{\mathbf{s}\in\mathcal{I}}$. By the continuity of KL-divergence, there exists a small enough constant $c>0$ such that if $\left\|P_{\mathbf{s}} - P_{\mathbf{s}}^{\mathcal{D}^*}\right\|_1 \leq c$ holds for all $\mathbf{s}\in\mathcal{I}$, for any $\mathcal{D}\in[\mathcal{C}]\setminus[\mathcal{D}^*]$, it satisfies that

$$\forall \mathbf{s}\in\mathcal{I}: \mathrm{KL}(P_{\mathbf{s}} \parallel P_{\mathbf{s}}^{\mathcal{D}^*}) \leq \mathrm{KL}(P_{\mathbf{s}} \parallel P_{\mathbf{s}}^{\mathcal{D}}) \text{ and } \exists \mathbf{s}\in\mathcal{I}: \mathrm{KL}(P_{\mathbf{s}} \parallel P_{\mathbf{s}}^{\mathcal{D}^*}) < \mathrm{KL}(P_{\mathbf{s}} \parallel P_{\mathbf{s}}^{\mathcal{D}}). \tag{25}$$

Recall that at time t, the track-and-stop causal discovery algorithm tracks the most probable DAG

$$\mathcal{D}_t^* \in \arg\max_{\mathcal{D}\in[\mathcal{C}]} \sum_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s},\mathbf{v}) \log P_{\mathbf{s}}^{\mathcal{D}}(\mathbf{v}).$$

As a matter of fact,

$$\mathcal{D}_t^* \in \arg\min_{\mathcal{D}\in[\mathcal{C}]} \sum_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s})\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathcal{D}}).$$

Thus, if $\forall \mathbf{s} \in \mathcal{I} : \left\|\bar{P}_{\mathbf{s},t} - P_{\mathbf{s}}^{\mathcal{D}^*}\right\|_1 \leq c$, according to conditions in (25), $\mathcal{D}_t^* = \mathcal{D}^*$ can be uniquely determined.

For $\epsilon \in (0, c]$, define time $\tau_p(\epsilon) = \max\{t \in \mathbb{N}_{>0} \mid \exists \mathbf{s} : \left\|\bar{P}_{\mathbf{s},t} - P_{\mathbf{s}}\right\|_1 \geq \epsilon\}$. Therefore, for any $T \geq \tau_p(\epsilon)$, $\mathcal{D}_T^* = \mathcal{D}^*$. As a result,

$$d_T = \min_{\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}_t^*} \sum_{\mathbf{s} \in \mathcal{I}} N_T(\mathbf{s}) \mathrm{KL}(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathcal{D}}) = \min_{\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}^*} \sum_{\mathbf{s} \in \mathcal{I}} N_T(\mathbf{s}) \mathrm{KL}\left(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathcal{D}}\right). \tag{26}$$

It follows from Lemma 12 that

$$\begin{aligned}
(26) &\geq \min_{\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}^*} \sum_{\mathbf{s} \in \mathcal{I}} \sum_{t=1}^{T} \alpha_{\mathbf{s},t} \mathrm{KL}\left(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathcal{D}}\right) - |\mathcal{I}|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2)D \\
&\geq \min_{\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}^*} \sum_{t=\tau_p(\epsilon)+1}^{T} \sum_{\mathbf{s} \in \mathcal{I}} \alpha_{\mathbf{s},t} \mathrm{KL}(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathcal{D}}) - |\mathcal{I}|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2)D \\
&\geq \min_{\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}^*} \sum_{t=\tau_p(\epsilon)+1}^{T} \sum_{\mathbf{s} \in \mathcal{I}} \alpha_{\mathbf{s},t} \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathcal{D}}) - 2[T - \tau_p(\epsilon)]u(\epsilon) - |\mathcal{I}|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2)D \\
&\geq \sum_{t=\tau_p(\epsilon)+1}^{T} \min_{\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}^*} \sum_{\mathbf{s} \in \mathcal{I}} \alpha_{\mathbf{s},t} \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathcal{D}}) - 2[T - \tau_p(\epsilon)]u(\epsilon) - |\mathcal{I}|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2)D,
\end{aligned} \tag{27}$$

where

$$u(\epsilon) = \sup_{(P_{\mathbf{s}})_{\mathbf{s} \in \mathcal{I}}} \left\{ \max_{\mathcal{D} \in [\mathcal{C}]} \left| \mathrm{KL}(P_{\mathbf{s}} \parallel P_{\mathbf{s}}^{\mathcal{D}}) - \mathrm{KL}(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}) \right| : \left\| P_{\mathbf{s}} - P_{\mathbf{s}}^{\mathcal{D}^*} \right\|_1 \leq \epsilon, \forall \mathbf{s} \in \mathcal{I} \right\}.$$

Recall that we define $\mathcal{D}_t' \in \arg\min_{\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}_t^*} \sum_{\mathbf{s} \in \mathcal{I}} \alpha_{\mathbf{s},t} \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathcal{D}})$. With Lemma 11, we have

$$\begin{aligned}
(27) &\geq \sum_{t=\tau_p(\epsilon)+1}^{T} \sum_{\mathbf{s} \in \mathcal{I}} \alpha_{\mathbf{s},t} \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathcal{D}_t'}) - 2[T - \tau_p(\epsilon)]u(\epsilon) - |\mathcal{I}|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2)D \\
&\geq \max_{\mathbf{s} \in \mathcal{I}} \sum_{t=\tau_p(\epsilon)+1}^{T} \mathrm{KL}\left(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathcal{D}_t'}\right) - R_T^{\mathrm{AH}} - \tau_p(\epsilon)D - 2[T - \tau_p(\epsilon)]u(\epsilon) - |\mathcal{I}|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2)D \\
&\geq \max_{\mathbf{s} \in \mathcal{I}} \sum_{t=\tau_p(\epsilon)+1}^{T} \mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}_t'}\right) - R_T^{\mathrm{AH}} - \tau_p(\epsilon)D - 3[T - \tau_p(\epsilon)]u(\epsilon) - |\mathcal{I}|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2)D \\
&= \max_{\mathbf{s} \in \mathcal{I}} \sum_{\mathcal{D} \in [\mathcal{C}]} N_{\tau_p(\epsilon):T}(\mathcal{D}) \mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}\right) - R_T^{\mathrm{AH}} - \tau_p(\epsilon)D - 3[T - \tau_p(\epsilon)]u(\epsilon) - |\mathcal{I}|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2)D,
\end{aligned} \tag{28}$$

where $N_{\tau_p(\epsilon):T}(\mathcal{D}) = \sum_{t=\tau_p(\epsilon)+1}^{T} \mathbf{1}\left\{\mathcal{D}_t' = \mathcal{D}\right\}$. With Lemma 9, we have

$$\begin{aligned}
\max_{\mathbf{s} \in \mathcal{I}} \sum_{\mathcal{D} \in [\mathcal{C}]} N_{\tau_p(\epsilon):T}(\mathcal{D}) \mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}\right) &\geq [T - \tau_p(\epsilon)] \inf_{\boldsymbol{w} \in \Delta([\mathcal{C}] \setminus \mathcal{D}^*)} \max_{\mathbf{s} \in \mathcal{I}} \sum_{\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}^*} w_{\mathcal{D}} \mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathcal{D}}\right) \\
&= [T - \tau_p(\epsilon)] c(\mathcal{D}^*).
\end{aligned}$$

Plugging the above result into (28), we get

$$(28) \geq [T - \tau_p(\epsilon)][c(\mathcal{D}^*) - 3u(\epsilon)] - R_T^{\mathrm{AH}} - \tau_p(\epsilon)D - |\mathcal{I}|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2)D := \underline{d}_t. \tag{29}$$

Define time

$$\bar{\tau}_\delta = \max_{\tau} \left\{ \tau \in \mathbb{N}_{>0} : d_\tau \geq |\mathcal{I}|\left(|\omega(V)| - 1\right), f_\tau(\underline{d}_\tau) \leq \delta \right\}.$$

According to the termination condition of the causal discovery algorithm, the algorithm terminates at $\tau_\delta \leq \bar{\tau}_\delta$. Since Corollary 2 shows that $\mathbb{E}[\tau_p(\epsilon)]$ is bounded by a constant, which means $\mathbb{P}(\mathbb{E}[\tau_p(\epsilon)] = \infty) = 0$, we have

$$\mathbb{P}(\tau_\delta = \infty) \leq \mathbb{P}(\bar{\tau}_\delta = \infty) = 0.$$

With Lemma 10, notice that (29) $= T[c(\mathcal{D}^*) - 3u(\epsilon)] + o(T)$ and $f_t(x)$ is dominated by $\exp(-x)$. We have for any $\epsilon \in (0, c]$

$$\lim_{\delta\to 0} \frac{\log(1/\delta)}{\mathbb{E}[\tau_\delta]} \geq \frac{\log(1/\delta)}{\mathbb{E}[\bar{\tau}_\delta]} = c(\mathcal{D}^*) - 3u(\epsilon),$$

The continuity of KL-divergence ensures that $\lim_{\epsilon\to 0} u(\epsilon) = 0$. Then we have that

$$\lim_{\delta\to 0} \frac{\log(1/\delta)}{\mathbb{E}[\tau_\delta]} \geq c(\mathcal{D}^*).$$

Combining with the lower bound result in Theorem 1, we conclude the proof. □

A.6.3. Asymptotic performance of Practical Algorithm

Lemma 17. *For the causal discovery problem with $\omega(\mathbf{S}) \subseteq \mathcal{I}$ contains all interventions on the node set $\mathbf{S}$, we have $c(\mathcal{D}^*) \geq \underline{c}(\mathcal{D}^*)$.*

Proof. Since the set of interventions $\mathcal{I}$ can be partitioned into interventions on different node sets, we have $\mathcal{I} = \cup_{\mathbf{S}\in\mathcal{S}} \omega(\mathbf{S})$ and $\omega(\mathbf{S}) \cap \omega(\mathbf{S}') = \emptyset$ for $\mathbf{S} \neq \mathbf{S}'$. Accordingly, for every $\mathcal{D} \in [\mathcal{C}] \setminus \mathcal{D}^*$, there exists at least one edge cut that has a different configuration compared with $\mathcal{D}^*$

$$\begin{aligned}\min_{\mathcal{D}\in[\mathcal{C}]\setminus\mathcal{D}^*} \sum_{\mathbf{s}\in\mathcal{I}} \alpha_{\mathbf{s}} \mathrm{KL}\left(P^*_{\mathbf{s}} \parallel P^{\mathcal{D}}_{\mathbf{s}}\right) &= \min_{\mathcal{D}\in[\mathcal{C}]\setminus\mathcal{D}^*} \sum_{\mathbf{S}\in\mathcal{S}} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \alpha_{\mathbf{s}} \mathrm{KL}\left(P^*_{\mathbf{s}} \parallel P^{\mathcal{D}}_{\mathbf{s}}\right) \\ &\geq \min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \alpha_{\mathbf{s}} \mathrm{KL}\left(P^*_{\mathbf{s}} \parallel P^{\mathsf{C}(\mathbf{S})}_{\mathbf{s}}\right),\end{aligned}$$

for any $\boldsymbol{\alpha} \in \Delta(\mathcal{I})$. Thus we get

$$\begin{aligned}c(\mathcal{D}^*) &= \sup_{\boldsymbol{\alpha}\in\Delta(\mathcal{I})} \min_{\mathcal{D}\in[\mathcal{C}]\setminus\mathcal{D}^*} \sum_{\mathbf{s}\in\mathcal{I}} \alpha_{\mathbf{s}} \mathrm{KL}\left(P^*_{\mathbf{s}} \parallel P^{\mathcal{D}}_{\mathbf{s}}\right) \\ &\leq \sup_{\boldsymbol{\alpha}\in\Delta(\mathcal{I})} \min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \alpha_{\mathbf{s}} \mathrm{KL}(P^{\mathcal{D}^*}_{\mathbf{s}} \parallel P^{\mathsf{C}(\mathbf{S})}_{\mathbf{s}}) = \underline{c}(\mathcal{D}^*).\end{aligned}$$

We reached the result. □

Lemma 18. *For the causal discovery problem, suppose the MEC represented by CPDAG $\mathcal{C}$ and observational distributions are available. If the faithfulness assumption in Definition 1 holds, for the practical algorithm $\mathcal{A}_{\mathsf{P}}$, we have $\mathbb{P}(\tau_\delta = \infty) = 0$ and*

$$\lim_{\delta\to 0} \frac{\log(1/\delta)}{\mathbb{E}[\tau_\delta]} = \underline{c}(\mathcal{D}^*).$$

Sketch of Proof. With the practical algorithm $\mathcal{A}_{\mathsf{P}}$, instead of searching $\mathcal{D}^*$ in $[\mathcal{C}]$, we search $(\mathsf{C}^*(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$ in the space $(\mathsf{CF}(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}$. Recall

$$\mathsf{C}'_t(\mathbf{S}) \in \operatorname*{arg\,min}_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*_t(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},t} \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P^{\mathsf{C}(\mathbf{S})}_{\mathbf{s}}),$$

and $\tau_p(\epsilon) = \max\{t \in \mathbb{N}_{>0} \mid \exists \mathbf{s} : \left\|\bar{P}_{\mathbf{s},t} - P_{\mathbf{s}}\right\|_1 \geq \epsilon\}$. For each $\mathbf{S} \in \mathcal{S}$, we have

$$\begin{aligned} Tc_T(\mathbf{S}) =& \sum_{t=1}^{T} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},i} \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathsf{C}'_t(\mathbf{S})}) \\ \geq& \sum_{t=\tau_p(\epsilon)+1}^{T} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},i} \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathsf{C}'_t(\mathbf{S})}) - \tau_p(\epsilon) D \\ \geq& \max_{\mathbf{s}\in\omega(\mathbf{S})} \sum_{t=\tau_p(\epsilon)+1}^{T} \mathrm{KL}\left(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathsf{C}'_t(\mathbf{S})}\right) - R_T^{\mathrm{AH}} - \tau_p(\epsilon) D \\ \geq& \max_{\mathbf{s}\in\omega(\mathbf{S})} \sum_{t=\tau_p(\epsilon)+1}^{T} \mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathsf{C}'_t(\mathbf{S})}\right) - R_T^{\mathrm{AH}} - \tau_p(\epsilon) D - [T - \tau_p(\epsilon)] u(\epsilon), \end{aligned} \tag{30}$$

where we apply Lemma 10 in the second inequality. Let $N_{\tau_p(\epsilon):T}(\mathsf{C}(\mathbf{S})) = \sum_{t=\tau_p(\epsilon)+1}^{T} \mathbf{1}\{\mathsf{C}'_t(\mathbf{S}) = \mathsf{C}(\mathbf{S})\}$. We apply the second inequality in Lemma 9 to get

$$\begin{aligned} &\max_{\mathbf{s}\in\omega(\mathbf{S})} \sum_{t=\tau_p(\epsilon)+1}^{T} \mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathsf{C}'_t(\mathbf{S})}\right) \\ =& \max_{\mathbf{s}\in\omega(\mathbf{S})} \sum_{\mathsf{C}(\mathbf{S})\in\mathsf{CF}(\mathbf{S})} N_{\tau_p(\epsilon):T}(\mathsf{C}(\mathbf{S})) \mathrm{KL}\left(P_{\mathbf{s}}^{\mathcal{D}^*} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}\right) \\ \geq& [T - \tau_p(\epsilon)] \inf_{\boldsymbol{\zeta}^{\mathbf{S}} \in \Delta(\mathsf{CF}(\mathbf{S})\setminus\mathsf{C}^*(\mathbf{S}))} \max_{\mathbf{s}\in\omega(\mathbf{S})} \sum_{\mathsf{C}(\mathbf{S})\in\mathsf{CF}(\mathbf{S})\setminus\mathsf{C}^*(\mathbf{S})} \zeta^{\mathbf{S}}_{\mathsf{C}(\mathbf{S})} \mathrm{KL}(P_{\mathbf{s}}^{\mathsf{C}^*(\mathbf{S})} \parallel P_{\mathbf{s},t}^{\mathsf{C}(\mathbf{S})}) \\ =& [T - \tau_p(\epsilon)][c_{\mathbf{S}}(\mathcal{D}^*) - u(\epsilon)], \end{aligned}$$

where $c_{\mathbf{S}}(\mathcal{D}^*)$ is defined in (18). Plugging the above result into (30), we get

$$Tc_T(\mathbf{S}) \geq [T - \tau_p(\epsilon)][c_{\mathbf{S}}(\mathcal{D}^*) - u(\epsilon)] - R_T^{\mathrm{AH}} - \tau_p(\epsilon) D. \tag{31}$$

Since $R_T^{\mathrm{AH}} \leq \sqrt{DT \ln K} + D\left(\frac{4}{3}\ln K + 2\right)$, (31) indicates $c_t(\mathbf{S})/t \to c_{\mathbf{S}}(\mathcal{D}^*) - u(\epsilon)$ as $t \to \infty$. Furthermore, since $\gamma_{\mathbf{S},t} \propto 1/c_t(\mathbf{S})$, we can define a stopping time

$$\tau_{p,\gamma}(\epsilon) := \max\left\{t \geq \tau_p(\epsilon) \;\middle|\; \sum_{\mathbf{S}\in\mathcal{S}} |\gamma_{\mathbf{S},t} - \gamma^*_{\mathbf{S}}| \geq \epsilon\right\}. \tag{32}$$

With $\mathbb{E}[\tau_P(\epsilon)] \leq c(\epsilon)$ according to Corollary 2, we have $\mathbb{E}[\tau_{P,\gamma}(\epsilon)] \leq c'(\epsilon)$ for some $c'(\epsilon) \leq \infty$.

Similar to the proof of Lemma 16, by the continuity of KL-divergence, there exists a small enough constant $c > 0$ such that if $\forall \mathbf{s} \in \mathcal{I} : \left\|\bar{P}_{\mathbf{s},t} - P_{\mathbf{s}}^{\mathcal{D}^*}\right\|_1 \leq c$ holds for all $\mathbf{s} \in \mathcal{I}$, each $\mathsf{C}^*_t(\mathbf{S}) = \mathsf{C}^*(\mathbf{S})$ for all $\mathbf{S} \in \mathcal{S}$ can be uniquely determined. Therefore, for any $\epsilon \in (0, c]$, , $\forall \mathbf{S} \in \mathcal{S} : \mathsf{C}^*_T(\mathbf{S}) = \mathsf{C}^*(\mathbf{S})$, if $T \geq \tau_p(\epsilon)$. It follows from (24) that for $T \geq \tau_p(\epsilon)$,

$$\begin{aligned} d_T =& \min_{(\mathsf{C}(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}} \neq (\mathsf{C}^*_t(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}} \sum_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s}) \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) \\ =& \min_{(\mathsf{C}(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}} \neq (\mathsf{C}^*(\mathbf{S}))_{\mathbf{S}\in\mathcal{S}}} \sum_{\mathbf{s}\in\mathcal{I}} N_t(\mathbf{s}) \mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) \\ \geq& \min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{\mathbf{s}\in\omega(\mathbf{S})} N_T(\mathbf{s}) \mathrm{KL}\left(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}\right) \\ \geq& \min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \left[\sum_{\mathbf{s}\in\omega(\mathbf{S})} \sum_{t=1}^{T} \alpha_{\mathbf{s},t} \mathrm{KL}\left(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S}}\right) - |\omega(\mathbf{S})|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2) D\right] \\ \geq& \min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{t=\tau_{p,\gamma}(\epsilon)+1}^{T} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \alpha_{\mathbf{s},t} \mathrm{KL}(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) - |\mathcal{I}|\left(|\mathcal{I}| - 1\right)(\sqrt{t} + 2) D \end{aligned} \tag{33}$$

where we apply Lemma 12 in the second inequality. Since $\alpha_{\mathbf{s}} = \gamma_{\mathbf{S},t}\xi^{\mathbf{S}}_{\mathbf{s},t}$

$$\begin{aligned}
&\min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{t=\tau_{p,\gamma}(\epsilon)+1}^{T} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \alpha_{\mathbf{s},t}\mathrm{KL}(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) \\
=&\min_{\mathbf{S}\in\mathcal{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{t=\tau_{p,\gamma}(\epsilon)+1}^{T} \gamma_{\mathbf{S},t} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},t}\mathrm{KL}(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) \\
\geq&\min_{\mathbf{S}\in\mathcal{S}} \gamma^*_{\mathbf{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{t=\tau_{p,\gamma}(\epsilon)+1}^{T} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},t}\mathrm{KL}(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) - [T-\tau_{p,\gamma}(\epsilon)]\epsilon D \\
\geq&\min_{\mathbf{S}\in\mathcal{S}} \gamma^*_{\mathbf{S}} \min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{t=\tau_{p,\gamma}(\epsilon)+1}^{T} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},t}\mathrm{KL}(\bar{P}_{\mathbf{s},t} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) - 2[T-\tau_{p,\gamma}(\epsilon)]u(\epsilon) - [T-\tau_{p,\gamma}(\epsilon)]\epsilon D \qquad (34)
\end{aligned}$$

where the first inequality is due to definition (32). With (31), we have

$$\begin{aligned}
\min_{\mathsf{C}(\mathbf{S})\neq\mathsf{C}^*(\mathbf{S})} \sum_{t=\tau_{p,\gamma}(\epsilon)+1}^{T} \sum_{\mathbf{s}\in\omega(\mathbf{S})} \xi^{\mathbf{S}}_{\mathbf{s},t}\mathrm{KL}(\bar{P}_{\mathbf{s},T} \parallel P_{\mathbf{s}}^{\mathsf{C}(\mathbf{S})}) &\geq Tc_T(\mathbf{S}) - \tau_{p,\gamma}(\epsilon)D \\
&\geq [T-\tau_{p,\gamma}(\epsilon)][c_{\mathbf{S}}(\mathcal{D}^*) - u(\epsilon)] - R_T^{\mathrm{AH}} - \tau_p(\epsilon)D - \tau_{p,\gamma}(\epsilon)D \qquad (35)
\end{aligned}$$

Putting together (33), (34) and (35), we apply the third equality in Lemma 9 to get

$$d_T \geq [T-\tau_{p,\gamma}(\epsilon)][\underline{c}(\mathcal{D}^*) - 3u(\epsilon) - \epsilon D] - R_T^{\mathrm{AH}} - \tau_{p,\gamma}(\epsilon)D - \tau_{p,\gamma}(\epsilon)D - |\mathcal{I}|\,(|\mathcal{I}|-1)(\sqrt{t}+2)D := \underline{d}_t.$$

The remaining proof is similar to that of Lemma 16. Define time

$$\bar{\tau}_\delta = \max_{\tau}\Big\{\tau \in \mathbb{N}_{>0} : d_\tau \geq |\mathcal{I}|\,(|\omega(V)|-1), f_\tau(\underline{d}_\tau) \leq \delta\Big\}.$$

According to the termination condition of the causal discovery algorithm, the algorithm terminates at $\tau_\delta \leq \bar{\tau}_\delta$. Since Corollary 2 shows that $\mathbb{E}[\tau_p(\epsilon)]$ is bounded, so is $\mathbb{E}[\tau_{P,\gamma}(\epsilon)]$. Accordingly, $\mathbb{P}(\mathbb{E}[\tau_p(\epsilon)] = \infty) = 0$, and we have

$$\mathbb{P}(\tau_\delta = \infty) \leq \mathbb{P}(\bar{\tau}_\delta = \infty) = 0.$$

With Lemma 10, notice that (29) $= T[\underline{c}(\mathcal{D}^*) - 3u(\epsilon) - \epsilon D] + o(T)$ and $f_t(x)$ is dominated by $\exp(-x)$. For any $\epsilon \in (0, c]$, it satisfies that

$$\lim_{\delta\to 0} \frac{\log(1/\delta)}{\mathbb{E}[\tau_\delta]} \geq \frac{\log(1/\delta)}{\mathbb{E}[\bar{\tau}_\delta]} = \underline{c}(\mathcal{D}^*) - 3u(\epsilon) - \epsilon D,$$

The continuity of KL-divergence ensures that $\lim_{\epsilon\to 0} u(\epsilon) = 0$. Then we have that

$$\lim_{\delta\to 0} \frac{\log(1/\delta)}{\mathbb{E}[\tau_\delta]} \geq \underline{c}(\mathcal{D}^*),$$

which concludes the proof.

□