FOCUS: FIXED-CONFIDENCE ONLINE CAUSAL LEARNING USING SEQUENTIAL ADAPTIVE INTERVENTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study fully online fixed-confidence causal discovery without any historical observational data. Starting from zero samples, the learner sequentially selects interventions to recover both the causal DAG and its edge weights under a linear-Gaussian structural equation model. We establish an instance-dependent lower bound for any (ϵ, δ)-correct algorithm and propose FOCUS, which adaptively allocates interventions through an online max–min game. A key contribution is a computable concentration inequality for the accumulated KL divergence involving causal parameters shared across interventions. We prove that FOCUS is (ϵ, δ)-correct and that its expected stopping time matches the lower bound in its $\Theta(\log(1/\delta))$ dependence up to an instance-dependent constant. Experiments demonstrate improved structure and edge-weight recovery and confirm the predicted stopping-time trend. Our codes are available on `https://anonymous.4open.science/r/FOCUS_code-76E5`

1 INTRODUCTION

Causal discovery seeks to uncover the directed relationships that govern a system and is fundamental to scientific reasoning and decision making (Pearl, 2009; Spirtes et al., 2000; Peters et al., 2017). Observational data, however, generally identify a causal directed acyclic graph (DAG) only up to its Markov equivalence class, leaving many edge directions unresolved (Spirtes et al., 2000; Peters et al., 2017). Interventions can resolve this ambiguity by actively perturbing the system, but are often substantially more costly than passive observations (Hauser & Bühlmann, 2014; Shanmugam et al., 2015). In clinical trials, for example, premature stopping for apparent benefit can yield exaggerated treatment-effect estimates (Montori et al., 2005), whereas unnecessary continuation consumes resources and may expose additional patients to inferior treatments (Pallmann et al., 2018). A sample-efficient learner therefore needs both an adaptive intervention strategy and a statistically valid stopping rule: **which intervention should be performed next, and when is the evidence sufficient to return a reliable causal model?**

Existing intervention-design methods do not fully address both questions. Classical methods typically study graph identification under population-level observational and interventional information, effectively treating distributions or edge orientations as known without sampling error (Shanmugam et al., 2015; Kocaoglu et al., 2017; Squires et al., 2020; Choo & Shiragur, 2023). Recent finite-data methods adaptively select interventions, but generally operate under a prescribed budget rather than a target confidence level (Scherrer et al., 2021; Tigas et al., 2022; Olko et al., 2023). The closest fixed-confidence approach is Elahi et al. (2024), but it assumes an available faithful observational distribution, restricts variables to finite discrete state spaces, and targets only DAG recovery. This leaves open fully online fixed-confidence learning for continuous systems, where no historical observational data are available and both causal structure and effect magnitudes must be learned. The distinction matters because the graph may be recovered well before its edge weights are estimated accurately.

We address this problem for linear-Gaussian structural equation models (SEM). The learner starts from zero samples and, at every round, chooses either passive observation or a single-node intervention, observes one new sample, and updates its estimate of the DAG, edge weights, and noise

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

variances. Given an accuracy level ϵ and confidence level δ, the objective is to stop as early as possible while returning the true DAG and an edge-weight matrix whose maximum entrywise error is smaller than ϵ, with probability at least $1 - \delta$. To this end, we introduce **F**ixed-confidence **O**nline **C**ausal learning method **U**sing **S**equential adaptive interventions (FOCUS), which is an online algorithm that does not need any historical data. We first characterize the instance-dependent information complexity of the problem and show that the continuous intervention intervals can be reduced without loss to their endpoints. FOCUS then combines joint maximum-likelihood estimation with an online max–min game to learn an informative intervention allocation, which is converted into actual interventions through forced exploration and cumulative tracking. A data-dependent stopping rule determines when the accumulated evidence is sufficient. We prove that FOCUS is (ϵ, δ)-correct (see Definition 3.2) and that its expected stopping time has the optimal $\Theta(\log(1/\delta))$ dependence up to an instance-dependent constant.

Contributions. **1) Fully online formulation and lower bound.** We formulate fixed-confidence joint recovery of causal structure and edge weights in linear-Gaussian SEMs without historical observational data, and derive an instance-dependent information-theoretic lower bound for every (ϵ, δ)-correct algorithm. **2) Adaptive intervention design.** We prove that restricting each continuous intervention interval to its two endpoints preserves the information complexity, and propose FOCUS, which learns and tracks an informative intervention allocation through an online max–min game. **3) Fixed-confidence guarantees.** We establish that FOCUS is (ϵ, δ)-correct and derive an instance-dependent upper bound on its expected stopping time, showing that it is order-optimal in the confidence level. A technical ingredient is a computable concentration inequality for the accumulated KL divergence under adaptive interventions. **4) Implementation and experiments.** We provide finite implementations of the optimization procedures used by FOCUS. Experiments across graphs of different sizes and densities demonstrate improved structure and edge-weight recovery and confirm the predicted dependence of the stopping time on $\log(1/\delta)$.

2 Related Work

Non-adaptive causal discovery. Causal structures can be learned from observational data using constraint-based, score-based, or continuous optimization methods (Spirtes et al., 2000; Chickering, 2002; Zheng et al., 2018). Under the standard Markov and faithfulness assumptions, however, observational data generally identify only a Markov equivalence class. In particular, a linear-Gaussian SEM with unrestricted noise variances is not identifiable from its observational distribution alone (Spirtes et al., 2000; Peters et al., 2017); exact identification requires additional restrictions, such as non-Gaussian noise or equal error variances (Shimizu et al., 2006; Peters & Bühlmann, 2014). Interventional data can distinguish observationally equivalent DAGs. Classical non-adaptive designs select all interventions in advance and study worst-case requirements, bounded intervention sizes, intervention costs, or fixed budgets (Eberhardt et al., 2005; Hu et al., 2014; Shanmugam et al., 2015; Kocaoglu et al., 2017; Ghassami et al., 2018; Lindgren et al., 2018; Sussex et al., 2021). Much of this literature assumes that the observational essential graph is known and that each intervention reveals its orientation information without sampling error, effectively requiring population-level observational and interventional information. Methods based on a fixed finite interventional dataset (Brouillard et al., 2020) also do not adapt subsequent interventions to the accumulated evidence.

Adaptive causal discovery. Adaptive methods select interventions using the outcomes of previous experiments. Existing work develops strategies based on equivalence classes, trees, clique decompositions, and adaptive graph search (He & Geng, 2008; Hauser & Bühlmann, 2014; Greenewald et al., 2019; Squires et al., 2020; Choo et al., 2022; Choo & Shiragur, 2023). Most graph-theoretic methods rely on an accurately known observational equivalence class and population-level interventional information. Recent finite-sample methods include AIT and GIT (Scherrer et al., 2021; Olko et al., 2023), as well as Bayesian approaches such as ABCD, CBED, and ABCI, which select interventions using posterior utilities or expected information gain (Agrawal et al., 2019; Tigas et al., 2022; Toth et al., 2022). These methods generally operate under a prescribed budget. Although some provide guarantees for the experimental-design objective, they do not establish frequentist fixed-confidence stopping guarantees for joint graph and edge-weight recovery. The most closely related work is Elahi et al. (2024), which proposes a track-and-stop algorithm using finitely many interventional samples. However, it assumes access to infinitely many observational samples, con-

siders only finite-state discrete variables, and targets only DAG structure recovery. In contrast, our method starts without historical observational data and jointly learns the graph, edge weights, and noise variances in a continuous linear-Gaussian SEM.

3 Problem Formulation

Notation. For a positive integer p, let $[p] := \{1, \dots, p\}$. For a matrix $\boldsymbol{M}$, we write $\|\boldsymbol{M}\|_{\max} := \max_{j,k} |M_{jk}|$. We use $\mathrm{KL}(P\|Q)$ for the Kullback–Leibler (KL) divergence and $\mathrm{kl}(x, y) := x \log(x/y) + (1-x)\log((1-x)/(1-y))$ for the binary relative entropy. For a measurable space $\mathcal{X}$, $\mathcal{P}(\mathcal{X})$ denotes the set of probability measures on $\mathcal{X}$.

3.1 Linear Gaussian Causal Model and Interventions

We consider a linear Gaussian SEM over p observed variables,

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{e}, \qquad \boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{1}$$

parameterized by $\theta = (D, \boldsymbol{A}, \boldsymbol{\Sigma}) \in \Theta$. Here, D is a directed acyclic graph (DAG) on $[p]$, $A_{jk} \neq 0$ if and only if $k \to j$ is an edge in D, and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \dots, \sigma_p^2)$ is the diagonal noise covariance. Thus, A_{jk} represents the direct causal effect of node k on node j.

For every node j, the learner may perform a perfect single-node intervention $\mathrm{do}(X_j = u)$. Such an intervention replaces the structural equation of node j by $X_j = u$, thereby removing all incoming edges to j while leaving the remaining structural equations unchanged. The intervention value can be chosen from a known interval $[U_{j,1}, U_{j,2}]$. The available action space is therefore

$$\mathcal{A}_\infty := \{0\} \cup \big\{(j, u) : j \in [p],\ u \in [U_{j,1}, U_{j,2}]\big\}, \tag{2}$$

where $a = 0$ denotes passive observation and $a = (j, u)$ denotes $\mathrm{do}(X_j = u)$. At this stage, $\mathcal{A}_\infty$ is generally uncountable; Section 4 will show that restricting each interval to its two endpoints entails no loss of information.

Let P_a^θ denote the non-degenerate Gaussian distribution observed under intervention a, with $a = 0$ corresponding to passive observation. For $a = (j, u)$, let $R = [p] \setminus \{j\}$. Since the intervention removes all incoming edges to node j and fixes $X_j = u$,

$$\boldsymbol{X}_R = (\boldsymbol{I} - \boldsymbol{A}_{R,R})^{-1} \boldsymbol{A}_{R,j} u + (\boldsymbol{I} - \boldsymbol{A}_{R,R})^{-1} \boldsymbol{e}_R. \tag{3}$$

The distributions under passive observation and single-node intervention are therefore summarized as

$$P_a^\theta = \begin{cases} \mathcal{N}\big(\boldsymbol{0},\ (\boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{\Sigma} (\boldsymbol{I} - \boldsymbol{A})^{-\mathsf{T}}\big), & a = 0, \\ \mathcal{N}\big((\boldsymbol{I} - \boldsymbol{A}_{R,R})^{-1} \boldsymbol{A}_{R,j} u,\ (\boldsymbol{I} - \boldsymbol{A}_{R,R})^{-1} \boldsymbol{\Sigma}_{R,R} (\boldsymbol{I} - \boldsymbol{A}_{R,R})^{-\mathsf{T}}\big), & a = (j, u). \end{cases} \tag{4}$$

In particular, the interventional mean is linear in u, whereas the covariance is independent of u. This structure will allow us to reduce the continuous intervention space to finitely many interventions.

Assumption 3.1 (Bounded parameter class)**.** *The number of nodes p is fixed, and every candidate graph is a DAG. There exist known constants $\beta_{\min}, A_{\max}, \sigma_{\min}, \sigma_{\max}, C_{\max} > 0$ such that*

$$\beta_{\min} \le |A_{jk}| \le A_{\max} \quad \textit{for every nonzero } A_{jk}, \quad \sigma_{\min}^2 \le \sigma_j^2 \le \sigma_{\max}^2, \quad \max_{j, u \in [U_{j,1}, U_{j,2}]} |u| \le C_{\max}.$$

Since there are finitely many DAGs on $[p]$ and the parameter constraints in Assumption 3.1 are closed and bounded, the resulting parameter space Θ is compact. The edge-strength lower bound separates distinct graph supports, while the variance and magnitude bounds prevent degenerate distributions and provide uniform control of the action-wise KL divergences.

3.2 Fixed-Confidence Online Causal Learning

Let $\theta^\star = (D^\star, \boldsymbol{A}^\star, \boldsymbol{\Sigma}^\star) \in \Theta$ denote the unknown data-generating parameter. At round t, the learner selects an intervention action $a_t \in \mathcal{A}_\infty$ based on the past and obtains an independent observation from $P_{a_t}^{\theta^\star}$. Let

$$\mathcal{F}_t := \sigma(a_1, \boldsymbol{X}_1, \dots, a_t, \boldsymbol{X}_t)$$

be the resulting filtration, where $\boldsymbol{X}_t$ denotes the non-degenerate random coordinates observed under a_t. A fixed-confidence causal learning algorithm is a triple

$$\mathfrak{A} = (\pi, \tau_\delta, \widehat{\theta}).$$

Here, $\pi = (\pi_t)_{t\geq 1}$ is an adaptive intervention policy, where π_t selects the next intervention action a_t based on the history available in $\mathcal{F}_{t-1}$; τ_δ is an $(\mathcal{F}_t)$-stopping time at which the algorithm terminates, indicating that its current estimate has reached the prescribed confidence level; and $\widehat{\theta} = (\widehat{\theta}_t)_{t\geq 1}$ is an adapted estimator sequence, where $\widehat{\theta}_t = (\widehat{D}_t, \widehat{\boldsymbol{A}}_t, \widehat{\boldsymbol{\Sigma}}_t)$. The terminal output is $\widehat{\theta}_{\tau_\delta}$.

Definition 3.2 ((ϵ,δ)-correctness)**.** Fix an accuracy level $0 < \epsilon \leq \beta_{\min}/2$ and a confidence level $\delta \in (0,1)$. An algorithm $\mathfrak{A}$ is (ϵ,δ)-*correct* over Θ if, for every $\theta^\star \in \Theta$,

$$\mathbb{P}_{\theta^\star}\Big(\tau_\delta < \infty,\ \widehat{D}_{\tau_\delta} = D^\star,\ \|\widehat{\boldsymbol{A}}_{\tau_\delta} - \boldsymbol{A}^\star\|_{\max} < \epsilon\Big) \geq 1-\delta. \tag{5}$$

Although $\boldsymbol{\Sigma}^\star$ is not included in the reported accuracy criterion, it must be estimated because both the likelihood and the KL-based sampling and stopping rules depend on it. Our objective is to construct an (ϵ,δ)-correct algorithm with the smallest possible expected stopping time.

3.3 Fundamental Information-Theoretic Limit

To characterize the intrinsic difficulty of the problem, write $\theta = (D^\theta, \boldsymbol{A}^\theta, \boldsymbol{\Sigma}^\theta)$ and $\lambda = (D^\lambda, \boldsymbol{A}^\lambda, \boldsymbol{\Sigma}^\lambda)$. For $r > 0$, define the alternative set

$$\mathrm{Alt}_r(\theta) := \left\{\lambda \in \Theta : D^\lambda \neq D^\theta \text{ or } \|\boldsymbol{A}^\lambda - \boldsymbol{A}^\theta\|_{\max} \geq r\right\}. \tag{6}$$

Thus, $\mathrm{Alt}_r(\theta)$ contains all parameters that are either structurally different from θ or at least r away from it in edge weights.

Let $\mathcal{P}(\mathcal{A}_\infty)$ denote the set of probability measures over the intervention space $\mathcal{A}_\infty$. We define the instance-dependent information complexity as

$$C_r(\theta) := \sup_{w\in\mathcal{P}(\mathcal{A}_\infty)} \inf_{\lambda\in\mathrm{Alt}_r(\theta)} \int_{\mathcal{A}_\infty} \mathrm{KL}(P_a^\theta \| P_a^\lambda)\, w(\mathrm{d}a). \tag{7}$$

It is the largest information rate that an intervention design can guarantee against its least distinguishable alternative.

Theorem 3.3 (Information-theoretic lower bound)**.** *Let $\delta \in (0, 1/2)$. Every algorithm that is (ϵ,δ)-correct uniformly over Θ satisfies*

$$\mathbb{E}_{\theta^\star}[\tau_\delta] \geq \frac{\mathrm{kl}(1-\delta,\delta)}{C_{2\epsilon}(\theta^\star)}. \tag{8}$$

Consequently,

$$\liminf_{\delta\downarrow 0} \frac{\mathbb{E}_{\theta^\star}[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{C_{2\epsilon}(\theta^\star)}. \tag{9}$$

The radius 2ϵ arises because two parameters separated by at least 2ϵ cannot share an ϵ-accurate estimate. Thus, $C_{2\epsilon}(\theta^\star)$ quantifies how quickly the best possible intervention strategy can distinguish the true model from every incompatible alternative. Theorem 3.3 provides the benchmark that guides our algorithm design: we seek an (ϵ,δ)-correct procedure whose expected stopping time approaches this fundamental limit as closely as possible.

Our goal. We aim to develop an (ϵ,δ)-correct algorithm whose upper bound on $\mathbb{E}_{\theta^\star}[\tau_\delta]$ is as close as possible to the information-theoretic lower bound in Theorem 3.3. We do not assume access to any historical or pre-collected observational data; all observational and interventional samples are acquired sequentially by the algorithm.

4 FOCUS: Adaptive Fixed-Confidence Causal Learning

The lower bound in Theorem 3.3 suggests allocating interventions to maximize the information collected against the least distinguishable alternative. This optimal allocation depends on the unknown parameter $\theta^\star$ and is therefore unavailable to the learner. We now construct FOCUS, which sequentially estimates the causal model, learns an informative intervention allocation, tracks this allocation through actual interventions, and stops once sufficient evidence has accumulated.

4.1 Reducing the Intervention Space to Endpoints

The intervention space $\mathcal{A}_\infty$ in Eq. (2) is generally uncountable. The following result shows that interior intervention values provide no advantage in the information-design problem.

Proposition 4.1 (Endpoint reduction). *For every $\theta \in \Theta$ and $r > 0$, the information complexity $C_r(\theta)$ in Eq. (7) admits an optimal intervention design supported on passive observation and the two endpoints of each intervention interval.*

For any fixed intervention target and any pair of parameters, the Gaussian KL divergence is a convex quadratic function of the intervention value. Therefore, probability mass assigned to an interior value can be redistributed to the two endpoints without decreasing the information against any alternative. Consequently, an optimal intervention design can always be chosen to use only the endpoints.

We henceforth restrict attention to the finite intervention set

$$\mathcal{A} := \{0\} \cup \big\{(j, U_{j,s}) : j \in [p],\ s \in \{1,2\}\big\}, \qquad K := |\mathcal{A}| = 2p + 1. \tag{10}$$

Accordingly, the information complexity in Eq. (7) has the equivalent finite-dimensional form

$$C_r(\theta) = \sup_{\boldsymbol{\alpha} \in \Delta(\mathcal{A})} \inf_{\lambda \in \mathrm{Alt}_r(\theta)} \sum_{a \in \mathcal{A}} \alpha_a \, \mathrm{KL}(P_a^\theta \| P_a^\lambda), \tag{11}$$

where $\Delta(\mathcal{A}) := \{\boldsymbol{\alpha} \in \mathbb{R}_+^K : \sum_{a \in \mathcal{A}} \alpha_a = 1\}$. Hence, the lower bound in Theorem 3.3 remains unchanged after replacing the continuous intervention space by $\mathcal{A}$.

Proposition 4.1 is essential because it converts intervention selection over an uncountable space into an exactly equivalent problem involving only $2p + 1$ interventions. This reduction preserves the information complexity and enables the finite-dimensional AdaHedge and tracking procedures developed below. Throughout the remainder of the paper, all intervention designs, intervention counts, and sums over interventions are defined with respect to $\mathcal{A}$ rather than $\mathcal{A}_\infty$.

4.2 Joint Estimation and Alternative Best Response

Given the observations collected by time t, define the adaptive log-likelihood and joint maximum-likelihood estimator (MLE) by

$$\ell_t(\theta) := \sum_{s=1}^{t} \log p_{a_s}^\theta(\boldsymbol{X}_s), \qquad \widehat{\theta}_t \in \arg\max_{\theta \in \Theta} \ell_t(\theta), \tag{12}$$

where p_a^θ is the density of P_a^θ. The estimator jointly learns the graph, edge weights, and noise variances from all observational and interventional samples: $\widehat{\theta}_t = (\widehat{D}_t, \widehat{\boldsymbol{A}}_t, \widehat{\boldsymbol{\Sigma}}_t)$.

If $\theta^\star$ were known, an optimizer $\boldsymbol{\alpha}^\star$ of the max–min problem defining $C_\epsilon(\theta^\star)$ in Eq. (11) would specify the most informative intervention allocation. Since both $\theta^\star$ and $\boldsymbol{\alpha}^\star$ are unknown, we replace $\theta^\star$ with the current estimate $\widehat{\theta}_t$ and learn the intervention allocation sequentially. Given the current allocation $\boldsymbol{\alpha}_t \in \Delta(\mathcal{A})$, an alternative-model best response is

$$\lambda_t \in \arg\min_{\lambda \in \mathrm{Alt}_\epsilon(\widehat{\theta}_t)} \sum_{a \in \mathcal{A}} \alpha_{t,a} \, \mathrm{KL}(P_a^{\widehat{\theta}_t} \| P_a^\lambda). \tag{13}$$

Here, $\alpha_{t,a}$ is the current estimated weight assigned to intervention a, rather than its realized sampling proportion. As detailed in Section 4.3, these weights are updated online by AdaHedge using the accumulated KL gains and are subsequently converted into actual interventions through cumulative tracking.

The resulting gain from intervention a is

$$r_{t,a} := \mathrm{KL}(P_a^{\widehat{\theta}_t} \| P_a^{\lambda_t}), \qquad a \in \mathcal{A}. \tag{14}$$

Thus, $r_{t,a}$ measures how informative intervention a is for distinguishing the current estimate from its least distinguishable ϵ-alternative.

4.3 Learning and Tracking the Intervention Allocation

Directly solving the max–min problem in Eq. (11) at every round can be computationally expensive. Following the game-based approach to fixed-confidence pure exploration, we instead let an online learner update the intervention allocation, while the alternative player returns the best response in Eq. (13) (Garivier & Kaufmann, 2016; Degenne et al., 2019; Elahi et al., 2024).

We use the gain-form AdaHedge algorithm (de Rooij et al., 2014). Initialize $\alpha_{1,a} = 1/K$, $\Delta_0 = 0$, and update

$$\alpha_{t+1,a} = \frac{\alpha_{t,a} \exp(\eta_t r_{t,a})}{\sum_{b \in \mathcal{A}} \alpha_{t,b} \exp(\eta_t r_{t,b})}, \tag{15}$$

$$m_t = \frac{1}{\eta_t} \log \left(\sum_{a \in \mathcal{A}} \alpha_{t,a} \exp(\eta_t r_{t,a}) \right), \qquad h_t = \sum_{a \in \mathcal{A}} \alpha_{t,a} r_{t,a}, \tag{16}$$

$$\Delta_t = \Delta_{t-1} + m_t - h_t, \qquad \eta_{t+1} = \frac{\log K}{\Delta_t}. \tag{17}$$

The initialization $\eta_1 = +\infty$ and zero-gap cases are interpreted by continuity. AdaHedge adapts its learning rate to the observed gains and approaches the best fixed intervention allocation in hindsight without requiring prior knowledge of the gain scale.

The allocation $\boldsymbol{\alpha}_t$ is fractional, whereas only one intervention can be implemented at each round. Let $N_t(a) := \sum_{s=1}^{t} \mathbf{1}\{a_s = a\}$ be the number of times intervention a has been implemented. Cumulative tracking aims to maintain $N_t(a) \approx \sum_{s=1}^{t} \alpha_{s,a}$, rather than matching $N_t(a)$ to the single-round weight $\alpha_{t,a}$. Combined with forced exploration, the next intervention is selected by

$$a_{t+1} \in \begin{cases} \arg\min_{a \in \mathcal{A}} N_t(a), & \text{if } \min_{a \in \mathcal{A}} N_t(a) < \sqrt{t}, \\ \arg\max_{a \in \mathcal{A}} \dfrac{\sum_{s=1}^{t} \alpha_{s,a}}{N_t(a)}, & \text{otherwise.} \end{cases} \tag{18}$$

Forced exploration ensures that every intervention is implemented infinitely often, while cumulative tracking translates the learned fractional allocation into actual interventions.

4.4 Stopping Rule and the Complete Algorithm

To evaluate whether the current estimate is sufficiently separated from all incorrect alternatives, define the accumulated information statistic

$$d_t := \inf_{\lambda \in \mathrm{Alt}_\epsilon(\widehat{\theta}_t)} \sum_{a \in \mathcal{A}} N_t(a) \, \mathrm{KL}(P_a^{\widehat{\theta}_t} \| P_a^{\lambda}). \tag{19}$$

The statistic is small when some ϵ-alternative remains difficult to distinguish and increases as evidence accumulates against every such alternative.

The stopping rule is built on the following concentration result for the accumulated KL divergence between the joint MLE and the true causal model.

Lemma 4.2 (KL concentration)**.** *Under Assumption 3.1, for every deterministic $t \geq 1$ and $x \geq 0$,*

$$\mathbb{P}_{\theta^\star}\Big(\sum_{a \in \mathcal{A}} N_t(a) \, \mathrm{KL}(P_a^{\widehat{\theta}_t} \| P_a^{\theta^\star}) \geq x\Big) \leq f_t(x), \tag{20}$$

where the tail function is

$$f_t(x) := \min\left\{1, \ \left[e^q \mathcal{N}(\zeta_t(x)) + 1\right] e^{-\kappa_B x}\right\}. \tag{21}$$

In particular,

$$f_t(x) = \mathrm{poly}\big(t, x, \log(t+1)\big) e^{-\kappa_B x}.$$

The quantities q, κ_B, $\mathcal{N}(\cdot)$, and $\zeta_t(\cdot)$ are explicit and can be computed from p and the five known constants in Assumption 3.1. Their specific expressions are provided in Appendix C.6.

Algorithm 1 FOCUS: Fixed-Confidence Online Causal Learning

Require: Accuracy ϵ, confidence δ, intervention intervals, and parameter class Θ
Construct $\mathcal{A}$ according to Eq. (10)
Initialize $N_0(a) \leftarrow 0$ and $\alpha_{1,a} \leftarrow 1/K$ for every $a \in \mathcal{A}$
Initialize the remaining AdaHedge variables with $\Delta_0 \leftarrow 0$ and $\eta_1 \leftarrow +\infty$
for $t = 1, 2, \ldots$ **do**
    Select intervention a_t according to Eq. (18)
    Implement a_t and observe $\boldsymbol{X}_t \sim P_{a_t}^{\theta^\star}$
    for all $a \in \mathcal{A}$ **do**
        $N_t(a) \leftarrow N_{t-1}(a) + \mathbf{1}\{a_t = a\}$
    end for
    Compute the joint MLE $\widehat{\theta}_t$ according to Eq. (12)
    Compute d_t according to Eq. (19)
    if $\frac{\pi^2 t^2}{6} f_t(d_t) < \delta$ **then**
        return $\widehat{\theta}_t$
    end if
    Compute λ_t according to Eq. (13)
    Compute $\{r_{t,a}\}_{a \in \mathcal{A}}$ according to Eq. (14)
    Update $\boldsymbol{\alpha}_{t+1}$ and the AdaHedge variables according to Eqs. (15)–(17)
end for

Lemma 4.2 is a key technical contribution of this work. Existing fixed-confidence methods typically establish stopping validity through likelihood-ratio or generalized likelihood-ratio concentration (Garivier & Kaufmann, 2016; Degenne et al., 2019). In contrast, our result directly controls the accumulated KL divergence involving a joint estimator of the shared causal parameters. Establishing this KL concentration requires a new argument tailored to online causal estimation.

Using the tail function in Eq. (21), FOCUS stops whenever

$$\boxed{\frac{\pi^2 t^2}{6} f_t(d_t) < \delta} \tag{22}$$

and returns $\widehat{\theta}_t$. The complete procedure is summarized in Algorithm 1.

Remark 4.3 (Oracle-based formulation)**.** The theoretical version of FOCUS assumes access to two optimization oracles: a joint-MLE oracle for Eq. (12) and an alternative-model best-response oracle for Eqs. (13) and (19). Such oracle-based formulations are standard in related fixed-confidence pure-exploration and online causal-learning problems (Degenne et al., 2019; Elahi et al., 2024). Appendix F provides concrete finite procedures that can replace both oracle calls.

5 Theoretical Guarantees

We now establish the fixed-confidence correctness and sample-complexity guarantees of FOCUS.

Theorem 5.1 (Fixed-confidence correctness)**.** *Suppose Assumption 3.1 holds and the exact joint-MLE and alternative-model best-response procedures are used. Then, for every $\theta^\star \in \Theta$, $0 < \epsilon \leq \beta_{\min}/2$, and $\delta \in (0, 1)$, Algorithm 1 satisfies*

$$\mathbb{P}_{\theta^\star}\Big(\tau_\delta < \infty,\ \widehat{D}_{\tau_\delta} = D^\star,\ \|\widehat{\boldsymbol{A}}_{\tau_\delta} - \boldsymbol{A}^\star\|_{\max} < \epsilon\Big) \geq 1 - \delta. \tag{23}$$

Therefore, FOCUS *is (ϵ, δ)-correct in the sense of Definition 3.2.*

Theorem 5.1 guarantees that FOCUS simultaneously identifies the true causal graph and estimates every edge weight within the prescribed accuracy. This guarantee holds despite the adaptive choice of interventions and requires no historical observational data: all information used for estimation and stopping is collected online. Although $\boldsymbol{\Sigma}^\star$ is not part of the reported accuracy criterion, it is jointly estimated to evaluate the likelihoods, KL divergences, and stopping statistic.

Theorem 5.2 (Expected stopping-time upper bound)**.** *Suppose Assumption 3.1 holds and the exact joint-MLE and alternative-model best-response procedures are used. Then Algorithm 1 terminates almost surely,*

$$\mathbb{P}_{\theta^\star}(\tau_\delta < \infty) = 1, \tag{24}$$

and its expected stopping time satisfies

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta^\star}[\tau_\delta]}{\log(1/\delta)} \leq \frac{1}{\kappa_B C_\epsilon(\theta^\star)}. \tag{25}$$

Theorem 5.2 shows that the expected number of online samples grows at most logarithmically with the inverse confidence level. The information complexity $C_\epsilon(\theta^\star)$ captures the intrinsic distinguishability of the true model: a larger value means that informative interventions separate $\theta^\star$ from its closest alternatives more quickly. The factor κ_B is introduced by the direct KL concentration inequality in Lemma 4.2. Together, AdaHedge, forced exploration, and cumulative tracking ensure that the stopping statistic accumulates information at a linear rate, leading to the upper bound in Eq. (25).

Remark 5.3 (Comparison with the lower bound)**.** Combining Theorem 3.3 with Theorem 5.2 yields

$$\frac{1}{C_{2\epsilon}(\theta^\star)} \leq \liminf_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta^\star}[\tau_\delta]}{\log(1/\delta)} \leq \limsup_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta^\star}[\tau_\delta]}{\log(1/\delta)} \leq \frac{1}{\kappa_B C_\epsilon(\theta^\star)}. \tag{26}$$

Thus, FOCUS is (ϵ, δ)-correct and order-optimal in the confidence level: for fixed $\theta^\star$ and ϵ, both its expected stopping time and the information-theoretic lower bound scale as $\Theta(\log(1/\delta))$. The remaining gap is the instance-dependent multiplicative factor

$$\frac{C_{2\epsilon}(\theta^\star)}{\kappa_B C_\epsilon(\theta^\star)},$$

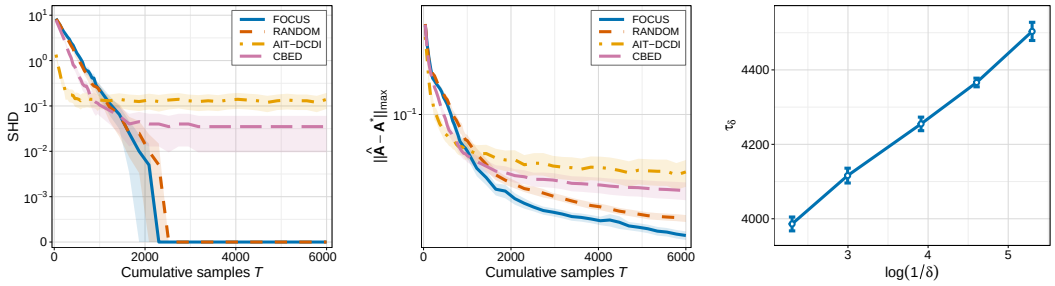
which is independent of δ.

6 Experiments

We evaluate whether FOCUS can efficiently recover both the causal structure and the associated edge weights through sequentially selected interventions. We further examine whether its empirical stopping time exhibits the logarithmic dependence on the confidence level predicted by Theorem 5.2.

Baselines. We compare FOCUS with the following three methods. RANDOM uses the same joint estimator and intervention set as FOCUS, but selects interventions uniformly at random, thereby isolating the benefit of the proposed adaptive allocation strategy. AIT-DCDI combines active intervention targeting (AIT) (Scherrer et al., 2021) with a linear-Gaussian implementation of differentiable causal discovery from interventional data (DCDI) (Brouillard et al., 2020). Finally, CBED is a causal Bayesian experimental-design method that selects informative intervention targets and values using expected information gain (Tigas et al., 2022). We adapt all methods to the same collection of observational and single-node perfect interventions and give them the same online sample budget. For a fair comparison of parameter estimation, the edge weights of each estimated graph are fitted using all samples collected by the corresponding method.

Simulation settings. For each Monte Carlo repetition, we first sample a random topological ordering of p nodes and generate an Erdős–Rényi DAG by independently including each admissible directed edge with probability ρ. Conditional on the resulting graph, the nonzero edge weights are independently sampled from $\mathrm{Unif}([-A_{\max}, -\beta_{\min}] \cup [\beta_{\min}, A_{\max}])$, and the diagonal entries of $\mathbf{\Sigma}^\star$ are independently sampled from $\mathrm{Unif}([\sigma^2_{\min}, \sigma^2_{\max}])$. Throughout the experiments, we set $\beta_{\min} = 0.15$, $A_{\max} = 1.5$, $\sigma^2_{\min} = 0.8$, and $\sigma^2_{\max} = 1.2$. Samples are then drawn from the linear-Gaussian SEM $\boldsymbol{X} = \boldsymbol{A}^\star \boldsymbol{X} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}^\star)$. At each online round, an algorithm selects either the observational action or one of the two endpoint interventions on a single node, as defined in Eq. (10), and receives one new sample. The intervention interval is $[-2, 2]$ for every node, so the two endpoint interventions set the selected variable to either -2 or 2. For FOCUS, we use the target estimation accuracy $\epsilon = 0.07 \leq \beta_{min}/2$. Within each repetition, all four methods are evaluated on the same underlying DAG, coefficients, noise variances, and sample budget.



(a) Structure error (b) Edge-weight error (c) Stopping-time scaling

Figure 1: Performance for $p = 6$ and $\rho = 0.3$. Panels (a)–(c) show SHD, maximum edge-weight estimation error, and the stopping time of FOCUS versus $\log(1/\delta)$, respectively. Shaded bands and error bars denote 95% confidence intervals.

We consider five graph settings: $(p, \rho) \in \{(5, 0.3), (6, 0.3), (7, 0.3), (6, 0.4), (6, 0.5)\}$.The representative setting $(p, \rho) = (6, 0.3)$ is reported in the main text. For each setting, we conduct 1000 independent Monte Carlo repetitions, regenerating the DAG and its parameters in every repetition. The curves report the sample means, and the shaded regions show the corresponding 95% confidence intervals. The remaining experimental results are provided in Appendix G due to space constraints.

Evaluation metrics. Each experiment contains three panels. The first reports the structural Hamming distance $\mathrm{SHD}(\widehat{D}_n, D^\star)$ against the number n of collected samples, where a smaller value indicates more accurate causal-structure recovery. The second reports the maximum edge-weight error $\|\widehat{\boldsymbol{A}}_n - \boldsymbol{A}^\star\|_{\max} := \max_{i,j} \left|\widehat{A}_{n,ij} - A^\star_{ij}\right|$, which evaluates parameter recovery in addition to graph identification. Both metrics are reported for all four methods under the same sample budget. The third panel concerns the fixed-confidence behavior of FOCUS and plots its empirical expected stopping time τ_δ against $\log(1/\delta)$ for confidence levels $\delta \in \{0.1, 0.05, 0.02, 0.01, 0.005\}$.

Results. Figure 1 presents the representative setting with $p = 6$ and $\rho = 0.3$. As shown in Figures 1a and 1b, although FOCUS improves slightly more slowly during the initial rounds, it soon overtakes the competing methods and achieves a clear advantage in both structural and edge-weight recovery. Its improvement over RANDOM highlights the benefit of learning a nonuniform intervention allocation rather than distributing samples uniformly across actions, while its advantage over AIT-DCDI and CBED shows the value of designing interventions for joint graph and edge-weight recovery. Comparing these two figures also shows that SHD often reaches zero before $\|\widehat{\boldsymbol{A}}_n - \boldsymbol{A}^\star\|_{\max}$ stabilizes, indicating that graph recovery alone may stop before the edge weights are accurately estimated and motivating the joint target of FOCUS. Finally, Figure 1c shows an approximately linear relationship between τ_δ and $\log(1/\delta)$, consistent with Theorem 5.2. Together, the three panels indicate that, when FOCUS terminates and returns its estimate, it already substantially outperforms the other methods under comparable sample budgets.

7 Conclusion

We studied fully online fixed-confidence recovery of causal structure and edge weights without any historical observational data. Starting from zero samples, FOCUS adaptively selects interventions and determines when sufficient evidence has been collected. It jointly updates the causal model and intervention allocation using only sequentially collected data. We established an instance-dependent lower bound, proved that FOCUS is (ϵ, δ)-correct, and showed that it achieves the optimal $\Theta(\log(1/\delta))$ dependence up to an instance-dependent constant. Experiments demonstrate improved structural and parameter recovery and confirm the predicted stopping-time trend. They also show that edge-weight estimation remains meaningful after the SHD reaches zero. Future work includes extending the framework to nonlinear models and latent confounding, and more general intervention mechanisms under practical experimental constraints.

AI USE STATEMENT

In this work, we used generative AI tools to assist with writing refinement and language polishing, as well as literature retrieval and discovery for identifying related research. We have not used generative AI tools for generating research ideas, developing the theoretical framework, designing the proposed algorithm, conducting experiments, analyzing experimental results, generating code, or producing mathematical proofs. Other uses of generative AI tools beyond the above-mentioned purposes are not applicable to this work. We have carefully reviewed all AI-assisted content and verified its correctness, consistency, and originality through manual inspection. In particular, all theoretical results, including the information-theoretic lower bound, concentration analysis, and correctness guarantees, as well as the algorithmic design and experimental implementation, were independently developed and validated by the authors. We take full responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI tools.

ETHICS STATEMENT

This work does not involve human subjects, personal data, or experiments on real-world participants. All experiments are conducted on synthetic datasets generated under controlled simulation settings. The proposed method is developed for fundamental research in causal learning and experimental design. We are not aware of any ethical concerns, privacy issues, or potential misuse arising from this work.

REPRODUCIBILITY STATEMENT

We have made efforts to ensure the reproducibility of this work. The source code for the proposed FOCUS algorithm and experimental implementation is provided in the appendix and abstract through an anonymous repository. The main paper describes the problem formulation, model assumptions, intervention setting, and algorithmic framework. Complete proofs of all theoretical results, including the information-theoretic lower bound, concentration inequality, and fixed-confidence guarantees, are provided in the appendix. Detailed experimental configurations, baseline implementations, and additional results are also included in the appendix. Together, these materials provide sufficient information to reproduce the theoretical analysis and empirical findings presented in this work.

REFERENCES

Raj Agrawal, Chandler Squires, Karren Yang, Karthikeyan Shanmugam, and Caroline Uhler. ABCD-Strategy: Budgeted experimental design for targeted causal structure discovery. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 3400–3409. PMLR, 2019.

Philippe Brouillard, Sébastien Lachapelle, Alexandre Lacoste, Simon Lacoste-Julien, and Alexandre Drouin. Differentiable causal discovery from interventional data. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, 2002.

Davin Choo and Kirankumar Shiragur. Adaptivity complexity for causal graph discovery. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pp. 391–402, 2023.

Davin Choo, Kirankumar Shiragur, and Arnab Bhattacharyya. Verification and search algorithms for causal DAGs. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Steven de Rooij, Tim van Erven, Peter D. Grünwald, and Wouter M. Koolen. Follow the leader if you can, hedge if you must. *Journal of Machine Learning Research*, 15(37):1281–1316, 2014.

Rémy Degenne, Wouter M. Koolen, and Pierre Ménard. Non-asymptotic pure exploration by solving games. In *Advances in Neural Information Processing Systems*, volume 32, pp. 14492–14501, 2019.

Frederick Eberhardt, Clark Glymour, and Richard Scheines. On the number of experiments sufficient and in the worst case necessary to identify all causal relations among n variables. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, pp. 178–184, 2005.

Muhammad Qasim Elahi, Lai Wei, Murat Kocaoglu, and Mahsa Ghasemi. Adaptive online experimental design for causal discovery. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12385–12408. PMLR, 2024.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 998–1027. PMLR, 2016.

AmirEmad Ghassami, Saber Salehkaleybar, Negar Kiyavash, and Elias Bareinboim. Budgeted experiment design for causal structure learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1724–1733, 2018.

Kristjan Greenewald, Dmitriy Katz, Karthikeyan Shanmugam, Sara Magliacane, Murat Kocaoglu, Enric Boix Adsera, and Guy Bresler. Sample efficient active learning of causal trees. In *Advances in Neural Information Processing Systems*, volume 32, pp. 14279–14289, 2019.

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

Yang-Bo He and Zhi Geng. Active learning of causal networks with intervention experiments and optimal designs. *Journal of Machine Learning Research*, 9:2523–2547, 2008.

Huining Hu, Zhentao Li, and Adrian R. Vetta. Randomized experimental design for causal graph discovery. In *Advances in Neural Information Processing Systems*, volume 27, pp. 2339–2347, 2014.

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

Murat Kocaoglu, Alexandros G. Dimakis, and Sriram Vishwanath. Cost-optimal learning of causal graphs. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1875–1884, 2017.

Béatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000. doi: 10.1214/aos/1015957395.

Erik M. Lindgren, Murat Kocaoglu, Alexandros G. Dimakis, and Sriram Vishwanath. Experimental design for cost-aware learning of causal graphs. In *Advances in Neural Information Processing Systems*, volume 31, pp. 5279–5289, 2018.

Victor M. Montori, P. J. Devereaux, Neill K. J. Adhikari, Karen E. A. Burns, Christoph H. Eggert, Matthias Briel, Christina Lacchetti, Teresa W. Leung, Elizabeth Darling, Dianne M. Bryant, Heiner C. Bucher, Holger J. Schünemann, Maureen O. Meade, Deborah J. Cook, Patricia J. Erwin, Amit Sood, Richa Sood, Benjamin Lo, Carly A. Thompson, Qi Zhou, Edward Mills, and Gordon H. Guyatt. Randomized trials stopped early for benefit: A systematic review. *JAMA*, 294 (17):2203–2209, 2005. doi: 10.1001/jama.294.17.2203.

Mateusz Olko, Michał Zajac, Aleksandra Nowak, Nino Scherrer, Yashas Annadani, Stefan Bauer, Łukasz Kuciński, and Piotr Miłoś. Trust your ∇: Gradient-based intervention targeting for causal discovery. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Philip Pallmann, Alun W. Bedding, Babak Choodari-Oskooei, Munyaradzi Dimairo, Laura Flight, Lisa V. Hampson, Jane Holmes, Adrian P. Mander, Lang'o Odondi, Matthew R. Sydes, Sofía S. Villar, James M. S. Wason, Christopher J. Weir, Graham M. Wheeler, Christina Yap, and Thomas Jaki. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Medicine*, 16(1):29, 2018. doi: 10.1186/s12916-018-1017-7.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2 edition, 2009.

Jonas Peters and Peter Bühlmann. Identifiability of gaussian structural equation models with equal error variances. *The Annals of Statistics*, 42(6):2191–2213, 2014.

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

Nino Scherrer, Olexa Bilaniuk, Yashas Annadani, Anirudh Goyal, Patrick Schwab, Bernhard Schölkopf, Michael C. Mozer, Yoshua Bengio, Stefan Bauer, and Nan Rosemary Ke. Learning neural causal models with active interventions. *arXiv preprint arXiv:2109.02429*, 2021.

Karthikeyan Shanmugam, Murat Kocaoglu, Alexandros G. Dimakis, and Sriram Vishwanath. Learning causal graphs with small interventions. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–2030, 2006.

Peter Spirtes, Clark N. Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, 2 edition, 2000.

Chandler Squires, Sara Magliacane, Kristjan Greenewald, Dmitriy Katz, Murat Kocaoglu, and Karthikeyan Shanmugam. Active structure learning of causal DAGs via directed clique trees. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21500–21511, 2020.

Scott Sussex, Caroline Uhler, and Andreas Krause. Near-optimal multi-perturbation experimental design for causal structure learning. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Panagiotis Tigas, Yashas Annadani, Andrew Jesson, Bernhard Schölkopf, Yarin Gal, and Stefan Bauer. Interventions, where and how? experimental design for causal models at scale. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24130–24143, 2022.

Christian Toth, Lars Lorch, Christian Knoll, Andreas Krause, Franz Pernkopf, Robert Peharz, and Julius von Kügelgen. Active bayesian causal inference. In *Advances in Neural Information Processing Systems*, volume 35, pp. 16261–16275, 2022.

Xun Zheng, Bryon Aragam, Pradeep K. Ravikumar, and Eric P. Xing. DAGs with NO TEARS: Continuous optimization for structure learning. In *Advances in Neural Information Processing Systems*, volume 31, pp. 9472–9483, 2018.

APPENDIX CONTENTS

A Proof of Proposition 4.1 **13**

B Proof of Theorem 3.3 **14**

C Proof of Lemma 4.2 **16**

D Proof of Theorem 5.1 **22**

E Proof of Theorem 5.2 **23**

F Finite Implementations of the Optimization Oracles **28**

G Additional Experimental Results **34**

Additional notation. For a vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|_2$ denotes its Euclidean norm. For a matrix $\boldsymbol{M}$, $\|\boldsymbol{M}\|_F$ and $\|\boldsymbol{M}\|_{op}$ denote its Frobenius and operator norms, respectively. For a DAG D, $\mathrm{pa}_D(j)$ denotes the parent set of node j.

A PROOF OF PROPOSITION 4.1

Proof. Fix $\theta, \lambda \in \Theta$ and an intervention target $j \in [p]$. Let $R = [p] \setminus \{j\}$ and define

$$\boldsymbol{g}_j^\theta := (\boldsymbol{I} - \boldsymbol{A}_{R,R}^\theta)^{-1}\boldsymbol{A}_{R,j}^\theta, \qquad \boldsymbol{V}_j^\theta := (\boldsymbol{I} - \boldsymbol{A}_{R,R}^\theta)^{-1}\boldsymbol{\Sigma}_{R,R}^\theta(\boldsymbol{I} - \boldsymbol{A}_{R,R}^\theta)^{-\mathsf{T}}.$$

Because $\boldsymbol{A}^\theta$ corresponds to a DAG, $\boldsymbol{I} - \boldsymbol{A}_{R,R}^\theta$ is invertible. Under the intervention $\mathrm{do}(X_j = u)$, the non-degenerate coordinates therefore follow

$$P_{j,u}^\theta = \mathcal{N}(\boldsymbol{g}_j^\theta u, \boldsymbol{V}_j^\theta).$$

The covariance does not depend on u, while the mean is linear in u.

For two Gaussian distributions, the KL divergence gives

$$\mathrm{KL}(P_{j,u}^\theta \| P_{j,u}^\lambda) = c_j(\theta, \lambda) + \frac{u^2}{2}(\boldsymbol{g}_j^\lambda - \boldsymbol{g}_j^\theta)^\mathsf{T}(\boldsymbol{V}_j^\lambda)^{-1}(\boldsymbol{g}_j^\lambda - \boldsymbol{g}_j^\theta), \tag{27}$$

where

$$c_j(\theta, \lambda) := \frac{1}{2}\left[\mathrm{tr}\big\{(\boldsymbol{V}_j^\lambda)^{-1}\boldsymbol{V}_j^\theta\big\} - (p-1) + \log\frac{\det(\boldsymbol{V}_j^\lambda)}{\det(\boldsymbol{V}_j^\theta)}\right]$$

is independent of u. The coefficient of u^2 in Eq. (27) is nonnegative because $\boldsymbol{V}_j^\lambda$ is positive definite. Hence

$$u \longmapsto \mathrm{KL}(P_{j,u}^\theta \| P_{j,u}^\lambda)$$

is convex on $[U_{j,1}, U_{j,2}]$ for every pair $\theta, \lambda \in \Theta$.

Suppose first that $U_{j,1} < U_{j,2}$. Every $u \in [U_{j,1}, U_{j,2}]$ can be written as

$$u = \rho_j(u)U_{j,1} + \big(1 - \rho_j(u)\big)U_{j,2}, \qquad \rho_j(u) := \frac{U_{j,2} - u}{U_{j,2} - U_{j,1}}.$$

By convexity,

$$\begin{aligned}\mathrm{KL}(P_{j,u}^\theta \| P_{j,u}^\lambda) \le{}& \rho_j(u)\,\mathrm{KL}(P_{j,U_{j,1}}^\theta \| P_{j,U_{j,1}}^\lambda) \\ &+ \big(1 - \rho_j(u)\big)\,\mathrm{KL}(P_{j,U_{j,2}}^\theta \| P_{j,U_{j,2}}^\lambda).\end{aligned} \tag{28}$$

If $U_{j,1} = U_{j,2}$, there is only one admissible intervention value and the conclusion is immediate.

Now take an arbitrary probability measure $w \in \mathcal{P}(\mathcal{A}_\infty)$. We construct a new measure $\widetilde{w}$ supported on $\mathcal{A}$ by retaining the mass of the passive action and transferring the mass of every interior intervention to its two endpoints. Specifically, let w_j be the restriction of w to interventions on node j, identified with a measure on $[U_{j,1}, U_{j,2}]$, and set

$$\begin{aligned}\widetilde{w}(\{0\}) &:= w(\{0\}),\\ \widetilde{w}(\{(j,U_{j,1})\}) &:= \int_{U_{j,1}}^{U_{j,2}} \rho_j(u)\, w_j(\mathrm{d}u),\\ \widetilde{w}(\{(j,U_{j,2})\}) &:= \int_{U_{j,1}}^{U_{j,2}} \big(1-\rho_j(u)\big)\, w_j(\mathrm{d}u).\end{aligned}$$

The transferred masses are nonnegative and preserve the total mass assigned to interventions on each node. Thus, $\widetilde{w}$ is a probability measure supported on the finite set $\mathcal{A}$.

Integrating Eq. (28) with respect to w_j and summing over all intervention targets shows that, for every $\lambda \in \Theta$,

$$\int_{\mathcal{A}_\infty} \mathrm{KL}(P_a^\theta \| P_a^\lambda)\, w(\mathrm{d}a) \le \sum_{a\in\mathcal{A}} \widetilde{w}(a)\, \mathrm{KL}(P_a^\theta \| P_a^\lambda). \tag{29}$$

Importantly, the construction of $\widetilde{w}$ depends only on w and the intervention intervals, not on the alternative parameter λ. Therefore, taking the infimum over $\lambda \in \mathrm{Alt}_r(\theta)$ preserves the inequality:

$$\inf_{\lambda\in\mathrm{Alt}_r(\theta)} \int_{\mathcal{A}_\infty} \mathrm{KL}(P_a^\theta \| P_a^\lambda)\, w(\mathrm{d}a) \le \inf_{\lambda\in\mathrm{Alt}_r(\theta)} \sum_{a\in\mathcal{A}} \widetilde{w}(a)\, \mathrm{KL}(P_a^\theta \| P_a^\lambda).$$

Hence every design over the continuous action space is weakly dominated by an endpoint-supported design. Since $\mathcal{A} \subseteq \mathcal{A}_\infty$, the reverse inequality between the corresponding optimal values is immediate, and thus

$$C_r(\theta) = \sup_{\boldsymbol{\alpha}\in\Delta(\mathcal{A})} \inf_{\lambda\in\mathrm{Alt}_r(\theta)} \sum_{a\in\mathcal{A}} \alpha_a\, \mathrm{KL}(P_a^\theta \| P_a^\lambda).$$

Finally, $\Delta(\mathcal{A})$ is compact, $\mathrm{Alt}_r(\theta)$ is compact, and the action-wise KL divergences are continuous in their parameters. Therefore, the displayed max–min objective attains its supremum on $\Delta(\mathcal{A})$. An optimal design supported only on the passive action and the two endpoint interventions for each node consequently exists. □

B Proof of Theorem 3.3

We first state the change-of-measure inequality used in the proof. Its finite-action version is Lemma 1 of Kaufmann et al. (2016). The following formulation extends it to a general measurable action space by replacing the expected action counts with an expected occupation measure.

Lemma B.1 (Change of measure for adaptive interventions)**.** *Fix two parameters $\theta, \lambda \in \Theta$ and consider any adaptive intervention policy. Let τ be an $(\mathcal{F}_t)$-stopping time such that $\mathbb{E}_\theta[\tau] < \infty$, and let $E \in \mathcal{F}_\tau$. Define the expected occupation measure under θ by*

$$\Gamma_\tau^\theta(B) := \mathbb{E}_\theta\left[\sum_{t=1}^{\tau} \mathbf{1}\{a_t \in B\}\right], \qquad B \subseteq \mathcal{A}_\infty.$$

Then

$$\int_{\mathcal{A}_\infty} \mathrm{KL}(P_a^\theta \| P_a^\lambda)\, \Gamma_\tau^\theta(\mathrm{d}a) \ge \mathrm{kl}(\mathbb{P}_\theta(E), \mathbb{P}_\lambda(E))\,. \tag{30}$$

For a finite action space, the left-hand side reduces to

$$\sum_{a\in\mathcal{A}} \mathbb{E}_\theta[N_\tau(a)]\, \mathrm{KL}(P_a^\theta \| P_a^\lambda).$$

Proof. Let p_a^θ and p_a^λ denote the densities of the observations under action a. The log-likelihood ratio of the history up to time t is

$$L_t(\theta, \lambda) := \sum_{s=1}^{t} \log \frac{p_{a_s}^\theta(\boldsymbol{X}_s)}{p_{a_s}^\lambda(\boldsymbol{X}_s)}.$$

Although the actions are selected adaptively, the intervention policy is the same under θ and λ. Conditional on the past, the probability with which the policy selects a_s therefore appears in both likelihoods and cancels from their ratio. Consequently, $L_\tau(\theta, \lambda)$ is the log-likelihood ratio between the stopped histories under θ and λ.

Taking conditional expectations under θ gives

$$\begin{aligned}
\mathbb{E}_\theta[L_\tau(\theta, \lambda)] &= \mathbb{E}_\theta\left[\sum_{s=1}^{\tau} \mathbb{E}_\theta\left[\log \frac{p_{a_s}^\theta(\boldsymbol{X}_s)}{p_{a_s}^\lambda(\boldsymbol{X}_s)} \middle| \mathcal{F}_{s-1}, a_s\right]\right] \\
&= \mathbb{E}_\theta\left[\sum_{s=1}^{\tau} \mathrm{KL}(P_{a_s}^\theta \| P_{a_s}^\lambda)\right] \\
&= \int_{\mathcal{A}_\infty} \mathrm{KL}(P_a^\theta \| P_a^\lambda)\, \Gamma_\tau^\theta(\mathrm{d}a). \qquad (31)
\end{aligned}$$

The interchange is valid because $\mathbb{E}_\theta[\tau] < \infty$ and the action-wise KL divergences are uniformly bounded over the compact parameter and intervention spaces.

The KL divergence between the distributions of the complete stopped histories equals $\mathbb{E}_\theta[L_\tau(\theta, \lambda)]$. Mapping a stopped history to the binary variable $\mathbf{1}\{E\}$ and applying the data-processing inequality for KL divergence yields

$$\mathbb{E}_\theta[L_\tau(\theta, \lambda)] \geq \mathrm{kl}(\mathbb{P}_\theta(E), \mathbb{P}_\lambda(E)).$$

Combining this inequality with Eq. (31) proves the result. □

We now apply Lemma B.1 to the fixed-confidence causal learning problem.

Proof of Theorem 3.3. If $\mathbb{E}_{\theta^\star}[\tau_\delta] = \infty$, the claimed lower bound holds trivially. We may therefore assume $\mathbb{E}_{\theta^\star}[\tau_\delta] < \infty$, which also implies that $\tau_\delta < \infty$ almost surely under $\theta^\star$.

Define the event that the algorithm returns an ϵ-accurate estimate of $\theta^\star$ by

$$E_{\theta^\star} := \left\{\tau_\delta < \infty,\ \widehat{D}_{\tau_\delta} = D^\star,\ \|\widehat{\boldsymbol{A}}_{\tau_\delta} - \boldsymbol{A}^\star\|_{\max} < \epsilon\right\}. \qquad (32)$$

Since the algorithm is (ϵ, δ)-correct uniformly over Θ,

$$\mathbb{P}_{\theta^\star}(E_{\theta^\star}) \geq 1 - \delta.$$

Fix an arbitrary $\lambda = (D^\lambda, \boldsymbol{A}^\lambda, \boldsymbol{\Sigma}^\lambda) \in \mathrm{Alt}_{2\epsilon}(\theta^\star)$. Under λ, define its corresponding success event

$$E_\lambda := \left\{\tau_\delta < \infty,\ \widehat{D}_{\tau_\delta} = D^\lambda,\ \|\widehat{\boldsymbol{A}}_{\tau_\delta} - \boldsymbol{A}^\lambda\|_{\max} < \epsilon\right\}.$$

Uniform correctness also gives $\mathbb{P}_\lambda(E_\lambda) \geq 1 - \delta$.

The events $E_{\theta^\star}$ and E_λ are disjoint. Indeed, if $D^\lambda \neq D^\star$, they require the algorithm to return two different graphs. Otherwise, the definition of $\mathrm{Alt}_{2\epsilon}(\theta^\star)$ gives $\|\boldsymbol{A}^\lambda - \boldsymbol{A}^\star\|_{\max} \geq 2\epsilon$. An estimate belonging to both events would satisfy

$$\|\boldsymbol{A}^\lambda - \boldsymbol{A}^\star\|_{\max} \leq \|\boldsymbol{A}^\lambda - \widehat{\boldsymbol{A}}_{\tau_\delta}\|_{\max} + \|\widehat{\boldsymbol{A}}_{\tau_\delta} - \boldsymbol{A}^\star\|_{\max} < 2\epsilon,$$

which is a contradiction. Therefore,

$$\mathbb{P}_\lambda(E_{\theta^\star}) \leq 1 - \mathbb{P}_\lambda(E_\lambda) \leq \delta.$$

This disjointness argument explains the appearance of the radius 2ϵ in the lower bound.

Applying Lemma B.1 with $\theta = \theta^\star$, stopping time τ_δ, and event $E_{\theta^\star}$ gives, for every $\lambda \in \mathrm{Alt}_{2\epsilon}(\theta^\star)$,

$$\int_{\mathcal{A}_\infty} \mathrm{KL}(P_a^{\theta^\star} \| P_a^\lambda)\, \Gamma_{\tau_\delta}^{\theta^\star}(\mathrm{d}a) \geq \mathrm{kl}(1-\delta, \delta). \tag{33}$$

Here we used that binary relative entropy is increasing in its first argument and decreasing in its second argument whenever the first argument exceeds the second, together with $\delta < 1/2$.

Normalize the expected occupation measure by defining

$$\overline{w}_\delta(B) := \frac{\Gamma_{\tau_\delta}^{\theta^\star}(B)}{\mathbb{E}_{\theta^\star}[\tau_\delta]}.$$

Because exactly one action is selected at every round, $\Gamma_{\tau_\delta}^{\theta^\star}(\mathcal{A}_\infty) = \mathbb{E}_{\theta^\star}[\tau_\delta]$, so $\overline{w}_\delta \in \mathcal{P}(\mathcal{A}_\infty)$. Dividing Eq. (33) by $\mathbb{E}_{\theta^\star}[\tau_\delta]$ and taking the infimum over $\lambda \in \mathrm{Alt}_{2\epsilon}(\theta^\star)$ yields

$$\begin{aligned} \mathrm{kl}(1-\delta, \delta) &\leq \mathbb{E}_{\theta^\star}[\tau_\delta] \inf_{\lambda \in \mathrm{Alt}_{2\epsilon}(\theta^\star)} \int_{\mathcal{A}_\infty} \mathrm{KL}(P_a^{\theta^\star} \| P_a^\lambda)\, \overline{w}_\delta(\mathrm{d}a) \\ &\leq \mathbb{E}_{\theta^\star}[\tau_\delta]\, C_{2\epsilon}(\theta^\star), \end{aligned} \tag{34}$$

where the final inequality follows from the definition of $C_{2\epsilon}(\theta^\star)$ as the supremum over all intervention designs. Rearranging gives

$$\mathbb{E}_{\theta^\star}[\tau_\delta] \geq \frac{\mathrm{kl}(1-\delta, \delta)}{C_{2\epsilon}(\theta^\star)}.$$

Finally,

$$\mathrm{kl}(1-\delta, \delta) = (1-2\delta) \log \frac{1-\delta}{\delta},$$

and hence

$$\lim_{\delta \downarrow 0} \frac{\mathrm{kl}(1-\delta, \delta)}{\log(1/\delta)} = 1.$$

Dividing the finite-confidence lower bound by $\log(1/\delta)$ and taking the limit inferior proves

$$\liminf_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta^\star}[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{C_{2\epsilon}(\theta^\star)}.$$

□

C Proof of Lemma 4.2

We establish several auxiliary results before proving Lemma 4.2. For each $a \in \mathcal{A}$, let $J(a)$ denote the random coordinates observed under action a: $J(0) = [p]$, while $J(a) = [p] \setminus \{j\}$ if a intervenes on node j. Write $d_a := |J(a)|$ and

$$P_a^\theta = \mathcal{N}(\boldsymbol{\mu}_a(\theta), \boldsymbol{V}_a(\theta)) \quad \text{on } \mathbb{R}^{d_a}.$$

C.1 Uniform bounds for the interventional Gaussian family

Define the computable constants

$$G_p := \sum_{k=0}^{p-1} (p A_{\max})^k, \qquad B_\mu := G_p \sqrt{p}\, A_{\max} C_{\max}, \tag{35}$$

and

$$v_- := \frac{\sigma_{\min}^2}{(1 + p A_{\max})^2}, \qquad v_+ := G_p^2 \sigma_{\max}^2. \tag{36}$$

Lemma C.1 (Uniform Gaussian bounds)**.** *Under Assumption 3.1, for every $\theta \in \Theta$ and $a \in \mathcal{A}$,*

$$\|\boldsymbol{\mu}_a(\theta)\|_2 \leq B_\mu, \qquad v_- \boldsymbol{I}_{d_a} \preceq \boldsymbol{V}_a(\theta) \preceq v_+ \boldsymbol{I}_{d_a}. \tag{37}$$

Consequently, the action-wise KL divergences are uniformly bounded:

$$\sup_{a\in\mathcal{A}}\sup_{\theta,\lambda\in\Theta}\mathrm{KL}(P_a^\theta\|P_a^\lambda)\le D_{\max},$$

where one valid computable choice is

$$D_{\max}:=\frac{1}{2}\left[p\left(\frac{v_+}{v_-}-1+\log\frac{v_+}{v_-}\right)+\frac{4B_\mu^2}{v_-}\right]. \tag{38}$$

Proof. Because $\boldsymbol{A}$ represents a DAG, it becomes strictly triangular after a simultaneous permutation of its rows and columns. Hence $\boldsymbol{A}^p=0$ and

$$(\boldsymbol{I}-\boldsymbol{A})^{-1}=\sum_{k=0}^{p-1}\boldsymbol{A}^k.$$

Since $\|\boldsymbol{A}\|_{\mathrm{op}}\le\|\boldsymbol{A}\|_{\mathrm{F}}\le pA_{\max}$, we obtain $\|(\boldsymbol{I}-\boldsymbol{A})^{-1}\|_{\mathrm{op}}\le G_p$. The same argument applies to every principal submatrix $\boldsymbol{A}_{R,R}$.

For an intervention $a=(j,U_{j,s})$, Eq. (4) gives

$$\boldsymbol{\mu}_a(\theta)=(\boldsymbol{I}-\boldsymbol{A}_{R,R})^{-1}\boldsymbol{A}_{R,j}U_{j,s}.$$

Using $\|\boldsymbol{A}_{R,j}\|_2\le\sqrt{p}A_{\max}$ and $|U_{j,s}|\le C_{\max}$ proves the mean bound. The observational mean is zero and therefore satisfies the same bound.

For the upper covariance bound, use $\|\boldsymbol{\Sigma}_{R,R}\|_{\mathrm{op}}\le\sigma_{\max}^2$ to obtain

$$\lambda_{\max}(\boldsymbol{V}_a(\theta))\le\|(\boldsymbol{I}-\boldsymbol{A}_{R,R})^{-1}\|_{\mathrm{op}}^2\sigma_{\max}^2\le v_+.$$

For the lower bound,

$$\lambda_{\min}(\boldsymbol{V}_a(\theta))\ge\sigma_{\min}^2 s_{\min}\big((\boldsymbol{I}-\boldsymbol{A}_{R,R})^{-1}\big)^2=\frac{\sigma_{\min}^2}{\|\boldsymbol{I}-\boldsymbol{A}_{R,R}\|_{\mathrm{op}}^2}\ge v_-.$$

The observational covariance is handled identically.

Finally, for $P=\mathcal{N}(\boldsymbol{\mu},\boldsymbol{V})$ and $Q=\mathcal{N}(\boldsymbol{\nu},\boldsymbol{W})$ in dimension $d\le p$,

$$\mathrm{KL}(P\|Q)=\frac{1}{2}\left[\mathrm{tr}(\boldsymbol{W}^{-1}\boldsymbol{V})-d+\log\frac{\det\boldsymbol{W}}{\det\boldsymbol{V}}+(\boldsymbol{\nu}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{W}^{-1}(\boldsymbol{\nu}-\boldsymbol{\mu})\right].$$

The bounds in Eq. (37) imply $\mathrm{tr}(\boldsymbol{W}^{-1}\boldsymbol{V})\le d\,v_+/v_-$, $\log(\det\boldsymbol{W}/\det\boldsymbol{V})\le d\log(v_+/v_-)$, and

$$(\boldsymbol{\nu}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{W}^{-1}(\boldsymbol{\nu}-\boldsymbol{\mu})\le\frac{4B_\mu^2}{v_-}.$$

This proves Eq. (38). □

C.2 A UNIFORM BHATTACHARYYA–KL COMPARISON

For distributions P and Q with densities p and q, define their Bhattacharyya distance by

$$\mathsf{B}(P,Q):=-\log\int\sqrt{pq}.$$

For Gaussian distributions $P=\mathcal{N}(\boldsymbol{\mu}_0,\boldsymbol{V}_0)$ and $Q=\mathcal{N}(\boldsymbol{\mu}_1,\boldsymbol{V}_1)$, direct Gaussian integration gives

$$\mathsf{B}(P,Q)=\frac{1}{8}\boldsymbol{\Delta}^{\mathsf{T}}\overline{\boldsymbol{V}}^{-1}\boldsymbol{\Delta}+\frac{1}{2}\log\frac{\det\overline{\boldsymbol{V}}}{\sqrt{\det\boldsymbol{V}_0\det\boldsymbol{V}_1}}, \tag{39}$$

where $\boldsymbol{\Delta}=\boldsymbol{\mu}_1-\boldsymbol{\mu}_0$ and $\overline{\boldsymbol{V}}=(\boldsymbol{V}_0+\boldsymbol{V}_1)/2$.

Let $\kappa_V:=v_+/v_-$ and define

$$g(r):=\frac{1}{2}(r-1-\log r),\qquad h(r):=\frac{1}{2}\log\frac{1+r}{2\sqrt{r}}.$$

At $r = 1$, define the continuous extension $h(1)/g(1) := 1/4$. Set

$$\kappa_\mu := \frac{v_-}{4v_+}, \qquad \kappa_\Sigma := \min_{r\in[\kappa_V^{-1},\kappa_V]} \frac{h(r)}{g(r)}, \qquad \kappa_B := \min\{\kappa_\mu, \kappa_\Sigma\}. \tag{40}$$

The minimization defining κ_Σ is one-dimensional and deterministic. Moreover, continuity and strict positivity away from $r = 1$ imply $\kappa_B > 0$.

Lemma C.2 (Bhattacharyya distance dominates reverse KL)**.** *For every $a \in \mathcal{A}$ and $\theta, \lambda \in \Theta$,*

$$\mathsf{B}(P_a^\theta, P_a^\lambda) \geq \kappa_B \,\mathrm{KL}(P_a^\lambda \| P_a^\theta). \tag{41}$$

Proof. We compare the mean and covariance parts separately. From $\overline{\boldsymbol{V}} \preceq v_+ \boldsymbol{I}$,

$$\mathsf{B}_\mu(P, Q) \geq \frac{\|\boldsymbol{\Delta}\|_2^2}{8v_+}.$$

Meanwhile,

$$\mathrm{KL}_\mu(Q\|P) = \frac{1}{2}\boldsymbol{\Delta}^\mathsf{T}\boldsymbol{V}_0^{-1}\boldsymbol{\Delta} \leq \frac{\|\boldsymbol{\Delta}\|_2^2}{2v_-}.$$

Thus, $\mathsf{B}_\mu(P, Q) \geq \kappa_\mu \,\mathrm{KL}_\mu(Q\|P)$.

For the covariance terms, let $r_1, \ldots, r_d$ be the eigenvalues of $\boldsymbol{V}_0^{-1/2}\boldsymbol{V}_1\boldsymbol{V}_0^{-1/2}$. The uniform covariance bounds imply $r_i \in [\kappa_V^{-1}, \kappa_V]$. Diagonalizing the two covariance expressions gives

$$\mathrm{KL}_\Sigma(Q\|P) = \sum_{i=1}^d g(r_i), \qquad \mathsf{B}_\Sigma(P, Q) = \sum_{i=1}^d h(r_i).$$

By the definition of κ_Σ, $\mathsf{B}_\Sigma(P, Q) \geq \kappa_\Sigma \,\mathrm{KL}_\Sigma(Q\|P)$. Adding the mean and covariance parts and taking their smaller constant proves the lemma. □

C.3 A FIXED-PARAMETER LIKELIHOOD-RATIO BOUND

For a fixed candidate parameter $\theta \in \Theta$, define

$$L_t(\theta) := \ell_t(\theta) - \ell_t(\theta^\star), \tag{42}$$

$$D_t(\theta, \theta^\star) := \sum_{a\in\mathcal{A}} N_t(a) \,\mathrm{KL}(P_a^\theta \| P_a^{\theta^\star}). \tag{43}$$

Lemma C.3 (Fixed-parameter deviation)**.** *For every deterministic $t \geq 1$, fixed $\theta \in \Theta$, and $b, y \geq 0$,*

$$\mathbb{P}_{\theta^\star}(L_t(\theta) \geq -b,\ D_t(\theta, \theta^\star) \geq y) \leq \exp\left(\frac{b}{2} - \kappa_B y\right). \tag{44}$$

Proof. Define

$$M_t^\theta := \exp\left\{\frac{1}{2}L_t(\theta) + \sum_{s=1}^t \mathsf{B}(P_{a_s}^{\theta^\star}, P_{a_s}^\theta)\right\}. \tag{45}$$

Because a_s is selected using only $\mathcal{F}_{s-1}$, conditional expectation under $\theta^\star$ gives

$$\mathbb{E}_{\theta^\star}\left[\exp\left\{\frac{1}{2}\log\frac{p_{a_s}^\theta(\boldsymbol{X}_s)}{p_{a_s}^{\theta^\star}(\boldsymbol{X}_s)} + \mathsf{B}(P_{a_s}^{\theta^\star}, P_{a_s}^\theta)\right\}\middle|\, \mathcal{F}_{s-1}\right]$$
$$= \exp\left\{\mathsf{B}(P_{a_s}^{\theta^\star}, P_{a_s}^\theta)\right\} \int \sqrt{p_{a_s}^{\theta^\star}(x) p_{a_s}^\theta(x)}\,\mathrm{d}x = 1.$$

Hence $(M_t^\theta)_{t\geq 0}$ is a nonnegative mean-one martingale.

On the event in Eq. (44), Lemma C.2 implies

$$M_t^\theta \geq \exp\{-b/2 + \kappa_B y\}.$$

Markov's inequality and $\mathbb{E}_{\theta^\star}[M_t^\theta] = 1$ then give the desired bound. □

C.4 Covering and Lipschitz bounds

Define a metric on the induced collection of action distributions by

$$d_{\mathsf{G}}(\theta,\lambda) := \max_{a\in\mathcal{A}} \{\|\boldsymbol{\mu}_a(\theta)-\boldsymbol{\mu}_a(\lambda)\|_2 + \|\boldsymbol{V}_a(\theta)-\boldsymbol{V}_a(\lambda)\|_{\mathrm{F}}\}. \tag{46}$$

Let $s_a := d_a(d_a+1)/2$ and

$$q := \sum_{a\in\mathcal{A}}(d_a+s_a) = p + \frac{p(p+1)}{2} + 2p\left[(p-1)+\frac{p(p-1)}{2}\right]. \tag{47}$$

A standard volumetric argument shows that a Euclidean ball of radius R in $\mathbb{R}^d$ has an ξ-net of cardinality at most $(1+2R/\xi)^d$: take a maximal ξ-separated subset and compare the volumes of the disjoint balls of radius $\xi/2$ around its elements with the enclosing ball of radius $R+\xi/2$.

Applying this argument to every mean and covariance component, and retaining one feasible parameter from each nonempty product cell, gives a ζ-net of the image of Θ under d_{G} with cardinality at most

$$\mathcal{N}(\zeta) := \prod_{a\in\mathcal{A}}\left(1+\frac{8B_\mu}{\zeta}\right)^{d_a}\left(1+\frac{8\sqrt{d_a}\,v_+}{\zeta}\right)^{s_a}, \qquad 0<\zeta\le 1. \tag{48}$$

The constant 8 accounts for selecting feasible representatives from the nonempty covering cells.

We also require a tail bound for the observed Gaussian vectors.

Lemma C.4 (Uniform Gaussian norm tail)**.** *For $z>0$, define*

$$R(z) := B_\mu + \sqrt{v_+}\big(\sqrt{p}+\sqrt{2z}\big). \tag{49}$$

Then, for every deterministic $t\ge 1$,

$$\mathbb{P}_{\theta^\star}\left(\max_{1\le s\le t}\|\boldsymbol{X}_s\|_2 > R(z)\right) \le te^{-z}. \tag{50}$$

Proof. Lemma 1 of Laurent & Massart (2000) states that if $Z\sim\chi^2_d$, then, for every $z>0$,

$$\mathbb{P}\Big(Z\ge d+2\sqrt{dz}+2z\Big)\le e^{-z}.$$

In particular, for $\boldsymbol{Z}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_d)$,

$$\mathbb{P}\Big(\|\boldsymbol{Z}\|_2 > \sqrt{d}+\sqrt{2z}\Big)\le e^{-z}.$$

Conditionally on $\mathcal{F}_{s-1}$ and a_s, write $\boldsymbol{X}_s = \boldsymbol{\mu}_{a_s}(\theta^\star)+\boldsymbol{V}_{a_s}(\theta^\star)^{1/2}\boldsymbol{Z}_s$. By Lemma C.1,

$$\|\boldsymbol{X}_s\|_2 \le B_\mu + \sqrt{v_+}\|\boldsymbol{Z}_s\|_2.$$

Since $d_{a_s}\le p$, the conditional probability of $\|\boldsymbol{X}_s\|_2 > R(z)$ is at most e^{-z}. A union bound over $s=1,\dots,t$ proves the claim. □

On the event $\max_{s\le t}\|\boldsymbol{X}_s\|_2\le R$, differentiation of the Gaussian log density gives the per-sample Lipschitz constant

$$L_\ell(R) := \frac{R+B_\mu}{v_-} + \frac{1}{2}\left[\frac{\sqrt{p}}{v_-}+\frac{(R+B_\mu)^2}{v_-^2}\right]. \tag{51}$$

Indeed, for a Gaussian density with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{V}$,

$$\nabla_{\boldsymbol{\mu}}\log p_{\boldsymbol{\mu},\boldsymbol{V}}(x) = \boldsymbol{V}^{-1}(x-\boldsymbol{\mu}),$$

while

$$\nabla_{\boldsymbol{V}}\log p_{\boldsymbol{\mu},\boldsymbol{V}}(x) = -\frac{1}{2}\boldsymbol{V}^{-1}+\frac{1}{2}\boldsymbol{V}^{-1}(x-\boldsymbol{\mu})(x-\boldsymbol{\mu})^{\top}\boldsymbol{V}^{-1}.$$

The uniform mean and covariance bounds therefore imply

$$\|\nabla_{\boldsymbol{\mu}}\log p_{\boldsymbol{\mu},\boldsymbol{V}}(x)\|_2 \le \frac{R+B_\mu}{v_-},$$

and

$$\|\nabla_{\boldsymbol{V}} \log p_{\boldsymbol{\mu},\boldsymbol{V}}(x)\|_{\mathrm{F}} \le \frac{1}{2}\left[\frac{\sqrt{p}}{v_-} + \frac{(R+B_\mu)^2}{v_-^2}\right].$$

The mean-value theorem consequently gives, whenever $d_{\mathsf{G}}(\theta,\lambda) \le \zeta$,

$$|L_t(\theta) - L_t(\lambda)| \le t L_\ell(R)\zeta. \tag{52}$$

Similarly, differentiating $\mathrm{KL}(P_a^\theta \| P_a^{\theta^\star})$ with respect to the mean and covariance of its first argument gives the uniform constant

$$L_{\mathrm{K}} := \frac{2B_\mu + \sqrt{p}}{v_-}. \tag{53}$$

Hence

$$|D_t(\theta,\theta^\star) - D_t(\lambda,\theta^\star)| \le t L_{\mathrm{K}}\zeta \qquad \text{whenever } d_{\mathsf{G}}(\theta,\lambda) \le \zeta. \tag{54}$$

C.5 Completion of the concentration proof

For $t \ge 1$ and $x \ge 0$, define

$$z_t(x) := \kappa_B x + 2\log(t+1), \tag{55}$$

$$H_t(x) := \frac{1}{2} L_\ell(R(z_t(x))) + \kappa_B L_{\mathrm{K}}, \tag{56}$$

$$\zeta_t(x) := \min\left\{1,\ \frac{q}{tH_t(x)}\right\}. \tag{57}$$

All quantities in these definitions depend only on t, x, p and the known constants in Assumption 3.1.

Proof of Lemma 4.2. For brevity, write

$$K_t^\star := \sum_{a\in\mathcal{A}} N_t(a)\,\mathrm{KL}(P_a^{\widehat{\theta}_t} \| P_a^{\theta^\star}) = D_t(\widehat{\theta}_t, \theta^\star).$$

Since $\widehat{\theta}_t$ is a global MLE and $\theta^\star \in \Theta$,

$$L_t(\widehat{\theta}_t) = \ell_t(\widehat{\theta}_t) - \ell_t(\theta^\star) \ge 0.$$

Therefore,

$$\{K_t^\star \ge x\} \subseteq \{\exists \theta \in \Theta : L_t(\theta) \ge 0,\ D_t(\theta,\theta^\star) \ge x\}. \tag{58}$$

Fix $z > 0$ and $\zeta \in (0,1]$, and let $\{\theta_1,\dots,\theta_M\}$ be a ζ-net under d_{G}, with $M \le \mathcal{N}(\zeta)$. On the truncation event $\max_{s\le t}\|\boldsymbol{X}_s\|_2 \le R(z)$, any witness θ in Eq. (58) has a net representative θ_i satisfying

$$L_t(\theta_i) \ge -tL_\ell(R(z))\zeta, \qquad D_t(\theta_i,\theta^\star) \ge x - tL_{\mathrm{K}}\zeta.$$

Applying Lemma C.3 to every net point and then using a union bound yields

$$\mathbb{P}_{\theta^\star}(K_t^\star \ge x) \le \mathcal{N}(\zeta)\exp\left\{-\kappa_B x + t\zeta\left[\frac{1}{2}L_\ell(R(z)) + \kappa_B L_{\mathrm{K}}\right]\right\} + te^{-z}. \tag{59}$$

If $x - tL_{\mathrm{K}}\zeta < 0$, the first term on the right-hand side is at least one, so the same inequality remains trivially valid.

We now choose $z = z_t(x)$ and $\zeta = \zeta_t(x)$. By construction,

$$te^{-z_t(x)} = \frac{t}{(t+1)^2}e^{-\kappa_B x} \le e^{-\kappa_B x},$$

and $t\zeta_t(x)H_t(x) \le q$. Substituting these choices into Eq. (59) gives

$$\mathbb{P}_{\theta^\star}(K_t^\star \ge x) \le \left[e^q \mathcal{N}(\zeta_t(x)) + 1\right] e^{-\kappa_B x}.$$

Since every probability is at most one, we conclude that

$$\mathbb{P}_{\theta^\star}(K_t^\star \ge x) \le f_t(x),$$

where

$$f_t(x) := \min\left\{1,\, [e^q \mathcal{N}(\zeta_t(x)) + 1]\, e^{-\kappa_B x}\right\}.$$

Finally, Eq. (48) implies $\mathcal{N}(\zeta) \le C_{\mathrm{cov}} \zeta^{-q}$ for an explicit constant C_{cov}. Moreover, $H_t(x) = O(1 + x + \log(t+1))$, and hence $\zeta_t(x)^{-q}$ is polynomial in t, x, and $\log(t+1)$. Therefore,

$$f_t(x) = \mathrm{poly}\big(t, x, \log(t+1)\big) e^{-\kappa_B x},$$

up to truncation at one, as claimed. □

C.6 Explicit computation of the concentration function

We now give an explicit construction of every quantity appearing in $f_t(x)$. All the constants below are deterministic and depend only on p and the known bounds in Assumption 3.1; in particular, they do not depend on $\theta^\star$ or on the observations.

First, define

$$G_p := \sum_{k=0}^{p-1} (p A_{\max})^k, \qquad B_\mu := G_p \sqrt{p}\, A_{\max} C_{\max}. \tag{60}$$

Because every candidate graph is acyclic, after a suitable permutation its weighted adjacency matrix is nilpotent. Hence $(\boldsymbol{I} - \boldsymbol{A})^{-1} = \sum_{k=0}^{p-1} \boldsymbol{A}^k$, and G_p uniformly bounds the operator norm of this inverse. Consequently, the mean vector under every action satisfies

$$\|\boldsymbol{\mu}_a(\theta)\|_2 \le B_\mu, \qquad a \in \mathcal{A}, \quad \theta \in \Theta.$$

Uniform lower and upper bounds on all action-wise covariance matrices are given by

$$v_- := \frac{\sigma_{\min}^2}{(1 + p A_{\max})^2}, \qquad v_+ := G_p^2 \sigma_{\max}^2. \tag{61}$$

Thus, for every $a \in \mathcal{A}$ and $\theta \in \Theta$,

$$v_- \boldsymbol{I} \preceq \boldsymbol{V}_a(\theta) \preceq v_+ \boldsymbol{I}.$$

Let

$$\kappa_V := \frac{v_+}{v_-}.$$

To define the exponential rate in the concentration inequality, introduce

$$g(r) := \frac{1}{2}\big(r - 1 - \log r\big), \qquad h(r) := \frac{1}{2} \log\left(\frac{1+r}{2\sqrt{r}}\right).$$

At $r = 1$, the ratio $h(r)/g(r)$ is defined by continuity and equals $1/4$. Set

$$\kappa_\mu := \frac{v_-}{4 v_+}, \qquad \kappa_\Sigma := \min_{r \in [\kappa_V^{-1}, \kappa_V]} \frac{h(r)}{g(r)}, \qquad \kappa_B := \min\{\kappa_\mu, \kappa_\Sigma\}. \tag{62}$$

The interval in Eq. (62) is compact and the extended ratio is continuous and strictly positive. Therefore, $\kappa_B > 0$ and can be computed numerically from v_- and v_+.

It remains to specify the covering term. Let d_a denote the dimension of the nondegenerate random coordinates under action a. Thus,

$$d_a = \begin{cases} p, & a = 0, \\ p - 1, & a = (j, u), \end{cases} \qquad s_a := \frac{d_a(d_a+1)}{2}.$$

Since there is one observational action and two endpoint interventions for each node, we may take the covering dimension to be

$$q := \sum_{a \in \mathcal{A}} (d_a + s_a) = p + \frac{p(p+1)}{2} + 2p\left(p - 1 + \frac{p(p-1)}{2}\right). \tag{63}$$

A valid explicit upper bound on the required covering number is

$$\mathcal{N}(\zeta) := \prod_{a\in\mathcal{A}} \left(1 + \frac{8B_\mu}{\zeta}\right)^{d_a} \left(1 + \frac{8\sqrt{d_a}\, v_+}{\zeta}\right)^{s_a}, \qquad \zeta > 0. \tag{64}$$

For $z \geq 0$, define the uniform sample-radius bound

$$R(z) := B_\mu + \sqrt{v_+}\big(\sqrt{p} + \sqrt{2z}\big). \tag{65}$$

The two Lipschitz constants used in the covering argument can then be chosen as

$$L_\ell(R) := \frac{R + B_\mu}{v_-} + \frac{1}{2}\left\{\frac{\sqrt{p}}{v_-} + \frac{(R+B_\mu)^2}{v_-^2}\right\}, \tag{66}$$

$$L_{\mathrm{KL}} := \frac{2B_\mu + \sqrt{p}}{v_-}. \tag{67}$$

For each $t \geq 1$ and $x \geq 0$, let

$$z_t(x) := \kappa_B x + 2\log(t+1), \tag{68}$$

$$H_t(x) := \frac{1}{2} L_\ell(R(z_t(x))) + \kappa_B L_{\mathrm{KL}}, \tag{69}$$

$$\zeta_t(x) := \min\left\{1,\ \frac{q}{tH_t(x)}\right\}. \tag{70}$$

Substituting Eqs. (62), (64), and (70) into Eq. (21) gives the fully computable tail function

$$f_t(x) = \min\left\{1,\ \left[e^q \prod_{a\in\mathcal{A}} \left(1 + \frac{8B_\mu}{\zeta_t(x)}\right)^{d_a} \left(1 + \frac{8\sqrt{d_a}\, v_+}{\zeta_t(x)}\right)^{s_a} + 1\right] e^{-\kappa_B x}\right\}. \tag{71}$$

Therefore, $f_t(x)$ can be evaluated using only p, $A_{\max}$, $\sigma_{\min}$, $\sigma_{\max}$, and $C_{\max}$. The remaining constant $\beta_{\min}$ in Assumption 3.1 is needed to separate different graph supports, but it is not required for this uniform Gaussian concentration bound. Hence every quantity in f_t is known before the online experiment begins.

D Proof of Theorem 5.1

Proof. Define

$$K_t^\star := \sum_{a\in\mathcal{A}} N_t(a)\, \mathrm{KL}\Big(P_a^{\widehat{\theta}_t} \Big\| P_a^{\theta^\star}\Big).$$

Suppose that FOCUS stops at time t but returns an incorrect estimate, namely,

$$\widehat{D}_t \neq D^\star \quad \text{or} \quad \|\widehat{\boldsymbol{A}}_t - \boldsymbol{A}^\star\|_{\max} \geq \epsilon.$$

By the definition of $\mathrm{Alt}_\epsilon(\widehat{\theta}_t)$, this implies $\theta^\star \in \mathrm{Alt}_\epsilon(\widehat{\theta}_t)$. Therefore, Eq. (19) gives

$$d_t \leq \sum_{a\in\mathcal{A}} N_t(a)\, \mathrm{KL}\Big(P_a^{\widehat{\theta}_t} \Big\| P_a^{\theta^\star}\Big) = K_t^\star. \tag{72}$$

Moreover, stopping at time t implies

$$f_t(d_t) < \frac{6\delta}{\pi^2 t^2}.$$

Combining this inequality with Eq. (72) and the fixed-time concentration bound in Lemma 4.2, the nestedness of the upper-tail events $\{K_t^\star \geq x\}$ yields

$$\mathbb{P}_{\theta^\star}\Big(\tau_\delta = t,\ \widehat{D}_t \neq D^\star \text{ or } \|\widehat{\boldsymbol{A}}_t - \boldsymbol{A}^\star\|_{\max} \geq \epsilon\Big) \leq \frac{6\delta}{\pi^2 t^2}. \tag{73}$$

This argument applies directly to the stopping condition $f_t(d_t) < 6\delta/(\pi^2 t^2)$.

Taking a union bound over all possible stopping times gives

$$\mathbb{P}_{\theta^\star}\Big(\tau_\delta < \infty,\, \widehat{D}_{\tau_\delta} \neq D^\star \text{ or } \|\widehat{\boldsymbol{A}}_{\tau_\delta} - \boldsymbol{A}^\star\|_{\max} \geq \epsilon\Big) \quad \leq \sum_{t=1}^{\infty} \frac{6\delta}{\pi^2 t^2} = \delta.$$

Finally, Theorem 5.2 guarantees that $\mathbb{P}_{\theta^\star}(\tau_\delta < \infty) = 1$. Hence,

$$\mathbb{P}_{\theta^\star}\Big(\tau_\delta < \infty,\, \widehat{D}_{\tau_\delta} = D^\star,\, \|\widehat{\boldsymbol{A}}_{\tau_\delta} - \boldsymbol{A}^\star\|_{\max} < \epsilon\Big) \geq 1 - \delta.$$

Thus, FOCUS is (ϵ, δ)-correct. □

E Proof of Theorem 5.2

This section proves the almost-sure termination and expected stopping-time upper bound of FOCUS. We first collect the tracking and online-learning properties used in the proof, then establish a finite expected localization time for the joint MLE, and finally show that the stopping statistic grows linearly.

For notational convenience, write

$$k_a(\theta, \lambda) := \mathrm{KL}(P_a^\theta \| P_a^\lambda), \qquad a \in \mathcal{A}.$$

Compactness of Θ, continuity of the action-wise Gaussian distributions, and the uniform covariance lower bound imply that

$$B_{\mathrm{KL}} := \max_{a \in \mathcal{A}} \sup_{\theta, \lambda \in \Theta} k_a(\theta, \lambda) < \infty. \tag{74}$$

Thus, the gain vectors in Eq. (14) satisfy $0 \leq r_{t,a} \leq B_{\mathrm{KL}}$ uniformly over t and a.

As usual in the analysis of stopping algorithms, we consider the infinite continuation of the sampling rule: if the stopping condition is met, the algorithm may be regarded as continuing to generate actions, estimates, and gain vectors according to the same update rules. This continuation does not change the actual stopping time.

E.1 Tracking and AdaHedge guarantees

We first state the cumulative-tracking property used below.

Lemma E.1 (Forced exploration and cumulative tracking)**.** *Under the intervention-selection rule in Eq. (18), for every $t \geq K$ and $a \in \mathcal{A}$,*

$$\sum_{s=1}^{t} \alpha_{s,a} - (K-1)(\sqrt{t}+2) \leq N_t(a) \leq \max\left\{1 + \sum_{s=1}^{t} \alpha_{s,a}, \sqrt{t}+1\right\}. \tag{75}$$

Moreover, there exist deterministic constants $c_{\mathrm{fe}} > 0$ and $t_{\mathrm{fe}} < \infty$, depending only on K, such that

$$\min_{a \in \mathcal{A}} N_t(a) \geq c_{\mathrm{fe}} \sqrt{t}, \qquad t \geq t_{\mathrm{fe}}. \tag{76}$$

The bound in Eq. (75) is precisely the allocation-matching result of Lemma 12 in Elahi et al. (2024), specialized to the action set $\mathcal{A}$. The lower bound in Eq. (76) is a direct consequence of the forced-exploration branch: whenever the least-sampled action falls below the threshold $\sqrt{t}$, one of the least-sampled actions is selected. Since there are only K actions and the threshold increases by $o(1)$ per round, the minimum count can lag behind $\sqrt{t}$ by at most an additive constant depending on K. Reducing the multiplicative constant if necessary gives Eq. (76).

We next record the regret guarantee for the gain-form AdaHedge procedure.

Lemma E.2 (AdaHedge regret)**.** *Suppose $0 \leq r_{t,a} \leq B_{\mathrm{KL}}$ for every t and a. Define*

$$\mathcal{R}_T^{\mathrm{AH}} := \max_{a \in \mathcal{A}} \sum_{t=1}^{T} r_{t,a} - \sum_{t=1}^{T} \sum_{a \in \mathcal{A}} \alpha_{t,a} r_{t,a}. \tag{77}$$

Then

$$\mathcal{R}_T^{\mathrm{AH}} \le B_{\mathrm{KL}}\sqrt{T\log K} + B_{\mathrm{KL}}\left(\frac{4}{3}\log K + 2\right). \tag{78}$$

In particular, $\mathcal{R}_T^{\mathrm{AH}} = o(T)$*. Furthermore, for every integer* $0 \le T_0 < T$*,*

$$\sum_{t=T_0+1}^{T}\sum_{a\in\mathcal{A}} \alpha_{t,a} r_{t,a} \ge \max_{a\in\mathcal{A}} \sum_{t=T_0+1}^{T} r_{t,a} - \mathcal{R}_T^{\mathrm{AH}} - T_0 B_{\mathrm{KL}}. \tag{79}$$

Proof. The first inequality follows from Corollary 9 of de Rooij et al. (2014), applied to the loss vectors $-r_t$ and rescaled from unit range to range B_{KL}. Indeed, the per-round range is at most B_{KL}, so the range-dependent bound in that corollary gives Eq. (78).

For the truncated bound, the definition of $\mathcal{R}_T^{\mathrm{AH}}$ gives

$$\sum_{t=1}^{T}\sum_{a\in\mathcal{A}} \alpha_{t,a} r_{t,a} \ge \max_{a\in\mathcal{A}} \sum_{t=1}^{T} r_{t,a} - \mathcal{R}_T^{\mathrm{AH}}.$$

The first T_0 mixed gains are nonnegative, while for each fixed action their sum over these rounds is at most $T_0 B_{\mathrm{KL}}$. Removing the first T_0 rounds therefore yields Eq. (79). □

We will also use the mixed-strategy form of the information game.

Lemma E.3 (Mixed-alternative representation)**.** *For every* $r > 0$*,*

$$C_r(\theta^\star) = \inf_{\nu\in\mathcal{P}(\mathrm{Alt}_r(\theta^\star))} \max_{a\in\mathcal{A}} \int_{\mathrm{Alt}_r(\theta^\star)} k_a(\theta^\star,\lambda)\,\nu(\mathrm{d}\lambda), \tag{80}$$

where $\mathcal{P}(\mathrm{Alt}_r(\theta^\star))$ *denotes the set of probability measures on* $\mathrm{Alt}_r(\theta^\star)$*.*

This is the continuous-alternative analogue of the standard max–min duality used in fixed-confidence pure exploration (Degenne et al., 2019); see also Lemma 9 of Elahi et al. (2024). It follows by allowing the alternative player to randomize and applying minimax duality to the compact allocation simplex and the convex set of probability measures over the compact alternative set.

E.2 Finite expected localization time

We measure the distance between two causal parameters by

$$\mathsf{d}_\Theta(\theta,\lambda) := \max\left\{\mathbf{1}\{D^\theta \ne D^\lambda\}, \|\boldsymbol{A}^\theta - \boldsymbol{A}^\lambda\|_{\max}, \|\boldsymbol{\Sigma}^\theta - \boldsymbol{\Sigma}^\lambda\|_{\max}\right\}. \tag{81}$$

For a fixed localization radius $\xi > 0$, define the last-exit time

$$\tau_{\mathrm{loc}}(\xi) := \inf\left\{n \ge 1 : \mathsf{d}_\Theta(\widehat{\theta}_t, \theta^\star) < \xi \text{ for every } t \ge n\right\}. \tag{82}$$

This random time is introduced only for the analysis and is not observed by the algorithm. Notice that ξ is unrelated to the AdaHedge learning rate η_t.

Define the parameter-separation constant

$$\Delta_\Theta(\xi;\theta^\star) := \inf_{\substack{\theta\in\Theta:\\ \mathsf{d}_\Theta(\theta,\theta^\star)\ge\xi}} \sum_{a\in\mathcal{A}} k_a(\theta,\theta^\star). \tag{83}$$

Compactness of Θ, continuity of the Gaussian KL divergence, and identifiability of the finite interventional family imply

$$\Delta_\Theta(\xi;\theta^\star) > 0 \qquad \text{for every fixed } \xi > 0. \tag{84}$$

Lemma E.4 (Finite expected localization time)**.** *For every fixed* $\xi > 0$*,*

$$\mathbb{E}_{\theta^\star}[\tau_{\mathrm{loc}}(\xi)] < \infty. \tag{85}$$

Proof. Recall the accumulated divergence between the joint MLE and the true parameter,

$$K_t^\star = \sum_{a\in\mathcal{A}} N_t(a) k_a(\widehat{\theta}_t, \theta^\star).$$

For $t \geq t_{\mathrm{fe}}$, if $\mathsf{d}_\Theta(\widehat{\theta}_t, \theta^\star) \geq \xi$, then Eqs. (76) and (83) give

$$\begin{aligned} K_t^\star &\geq \min_{a\in\mathcal{A}} N_t(a) \sum_{a\in\mathcal{A}} k_a(\widehat{\theta}_t, \theta^\star) \\ &\geq c_{\mathrm{fe}}\sqrt{t}\,\Delta_\Theta(\xi;\theta^\star). \end{aligned} \tag{86}$$

Therefore, Lemma 4.2 implies

$$\mathbb{P}_{\theta^\star}\Big(\mathsf{d}_\Theta(\widehat{\theta}_t, \theta^\star) \geq \xi\Big) \leq f_t\Big(c_{\mathrm{fe}}\sqrt{t}\,\Delta_\Theta(\xi;\theta^\star)\Big). \tag{87}$$

The event $\{\tau_{\mathrm{loc}}(\xi) > n\}$ means that the estimator leaves the ξ-neighborhood at least once after time n. Hence

$$\mathbb{P}_{\theta^\star}(\tau_{\mathrm{loc}}(\xi) > n) \leq \sum_{t=n}^{\infty} \mathbb{P}_{\theta^\star}\Big(\mathsf{d}_\Theta(\widehat{\theta}_t, \theta^\star) \geq \xi\Big).$$

Using the tail-sum formula and Eq. (87), we obtain

$$\begin{aligned} \mathbb{E}_{\theta^\star}[\tau_{\mathrm{loc}}(\xi)] &= \sum_{n=0}^{\infty} \mathbb{P}_{\theta^\star}(\tau_{\mathrm{loc}}(\xi) > n) \\ &\leq t_{\mathrm{fe}} + \sum_{t=t_{\mathrm{fe}}}^{\infty} (t+1) f_t\Big(c_{\mathrm{fe}}\sqrt{t}\,\Delta_\Theta(\xi;\theta^\star)\Big). \end{aligned} \tag{88}$$

By the explicit expression derived in Appendix C.6, the summand in Eq. (88) is bounded by a polynomial in t times $\exp(-c\sqrt{t})$ for some $c > 0$. The series is therefore finite. □

In particular, $\tau_{\mathrm{loc}}(\xi) < \infty$ almost surely.

E.3 Continuity error terms

We introduce two deterministic continuity moduli. The first controls the effect of replacing the true first argument of a KL divergence by a nearby estimated parameter:

$$u(\xi) := \max_{a\in\mathcal{A}} \sup_{\lambda\in\Theta} \sup_{\substack{\theta\in\Theta:\\ \mathsf{d}_\Theta(\theta,\theta^\star)\leq\xi}} |k_a(\theta,\lambda) - k_a(\theta^\star,\lambda)|. \tag{89}$$

To define the second modulus, for $\boldsymbol{\alpha} \in \Delta(\mathcal{A})$ let

$$b_\epsilon(\boldsymbol{\alpha},\theta) \in \operatorname*{arg\,min}_{\lambda\in \mathrm{Alt}_\epsilon(\theta)} \sum_{a\in\mathcal{A}} \alpha_a k_a(\theta,\lambda) \tag{90}$$

denote the best response selected by the algorithm. We use the same deterministic tie-breaking rule throughout. Define the auxiliary response

$$b^\star_{\epsilon-\xi}(\boldsymbol{\alpha},\theta) \in \operatorname*{arg\,min}_{\lambda\in \mathrm{Alt}_{\epsilon-\xi}(\theta^\star)} \sum_{a\in\mathcal{A}} \alpha_a k_a(\theta,\lambda). \tag{91}$$

The associated best-response modulus is

$$\begin{aligned} \widetilde{u}_\epsilon(\xi) := \sup\Big\{ &\big|k_a(\vartheta, b_\epsilon(\boldsymbol{\alpha},\theta)) - k_a\big(\vartheta, b^\star_{\epsilon-\xi}(\boldsymbol{\alpha},\theta)\big)\big| : \\ &a\in\mathcal{A},\ \boldsymbol{\alpha}\in\Delta(\mathcal{A}),\ \mathsf{d}_\Theta(\theta,\theta^\star)\leq\xi,\ \mathsf{d}_\Theta(\vartheta,\theta^\star)\leq\xi\Big\}. \end{aligned} \tag{92}$$

Since the parameter class and allocation simplex are compact, the Gaussian KL divergence is uniformly continuous, and the same deterministic tie-breaking rule is used for both best responses, we have

$$u(\xi) \longrightarrow 0, \qquad \widetilde{u}_\epsilon(\xi) \longrightarrow 0, \qquad C_{\epsilon-\xi}(\theta^\star) \longrightarrow C_\epsilon(\theta^\star) \quad \text{as } \xi \downarrow 0. \tag{93}$$

E.4 Linear growth of the stopping statistic

Lemma E.5 (Pathwise linear information growth). *Fix $0 < \xi < \min\{\epsilon, 1\}$ and let $T_0 := \tau_{\mathrm{loc}}(\xi)$. For every $T > T_0$,*

$$d_T \geq (T - T_0)\left[C_{\epsilon-\xi}(\theta^\star) - 3u(\xi) - 2\widetilde{u}_\epsilon(\xi)\right] - \mathcal{R}_T^{\mathrm{AH}} - T_0 B_{\mathrm{KL}} - K(K-1)(\sqrt{T}+2)B_{\mathrm{KL}}. \tag{94}$$

Proof. By the definition of T_0, for every $t \geq T_0$,

$$\mathsf{d}_\Theta(\widehat{\theta}_t, \theta^\star) < \xi. \tag{95}$$

In particular, since $\xi < 1$, all these estimates have the correct graph: $\widehat{D}_t = D^\star$.

For every $t > T_0$, write

$$\lambda_t := b_\epsilon(\boldsymbol{\alpha}_t, \widehat{\theta}_t)$$

for the actual alternative selected by the algorithm, and define the auxiliary response

$$\lambda_t^\circ := b^\star_{\epsilon-\xi}(\boldsymbol{\alpha}_t, \widehat{\theta}_t).$$

By construction, $\lambda_t^\circ \in \mathrm{Alt}_{\epsilon-\xi}(\theta^\star)$.

We begin with the stopping statistic

$$d_T = \inf_{\lambda \in \mathrm{Alt}_\epsilon(\widehat{\theta}_T)} \sum_{a \in \mathcal{A}} N_T(a) k_a(\widehat{\theta}_T, \lambda).$$

Using the lower tracking bound in Eq. (75), nonnegativity of KL divergence, and the uniform upper bound B_{KL}, we obtain

$$d_T \geq \inf_{\lambda \in \mathrm{Alt}_\epsilon(\widehat{\theta}_T)} \sum_{t=T_0+1}^{T} \sum_{a \in \mathcal{A}} \alpha_{t,a} k_a(\widehat{\theta}_T, \lambda) - K(K-1)(\sqrt{T}+2)B_{\mathrm{KL}}. \tag{96}$$

For nonnegative functions $q_t(\lambda)$, $\inf_\lambda \sum_t q_t(\lambda) \geq \sum_t \inf_\lambda q_t(\lambda)$. Therefore, the first term in Eq. (96) is at least

$$\sum_{t=T_0+1}^{T} \inf_{\lambda \in \mathrm{Alt}_\epsilon(\widehat{\theta}_T)} \sum_{a \in \mathcal{A}} \alpha_{t,a} k_a(\widehat{\theta}_T, \lambda).$$

Because $\mathsf{d}_\Theta(\widehat{\theta}_T, \theta^\star) < \xi$, the triangle inequality gives

$$\mathrm{Alt}_\epsilon(\widehat{\theta}_T) \subseteq \mathrm{Alt}_{\epsilon-\xi}(\theta^\star). \tag{97}$$

Indeed, a graph alternative to $\widehat{\theta}_T$ is also a graph alternative to $\theta^\star$, while for a same-graph alternative,

$$\|\boldsymbol{A}^\lambda - \boldsymbol{A}^\star\|_{\max} \geq \|\boldsymbol{A}^\lambda - \widehat{\boldsymbol{A}}_T\|_{\max} - \|\widehat{\boldsymbol{A}}_T - \boldsymbol{A}^\star\|_{\max} \geq \epsilon - \xi.$$

Both $\widehat{\theta}_T$ and $\widehat{\theta}_t$ lie in the ξ-neighborhood of $\theta^\star$. Applying Eq. (89) twice gives

$$\left|k_a(\widehat{\theta}_T, \lambda) - k_a(\widehat{\theta}_t, \lambda)\right| \leq 2u(\xi).$$

Combining this inequality with Eq. (97), for every $t > T_0$,

$$\inf_{\lambda \in \mathrm{Alt}_\epsilon(\widehat{\theta}_T)} \sum_{a \in \mathcal{A}} \alpha_{t,a} k_a(\widehat{\theta}_T, \lambda) \geq \sum_{a \in \mathcal{A}} \alpha_{t,a} k_a(\widehat{\theta}_t, \lambda_t^\circ) - 2u(\xi). \tag{98}$$

By the definition of $\widetilde{u}_\epsilon(\xi)$,

$$k_a(\widehat{\theta}_t, \lambda_t^\circ) \geq k_a(\widehat{\theta}_t, \lambda_t) - \widetilde{u}_\epsilon(\xi).$$

Substituting this into Eq. (98) and summing over t gives

$$d_T \geq \sum_{t=T_0+1}^{T} \sum_{a\in\mathcal{A}} \alpha_{t,a} k_a(\widehat{\theta}_t, \lambda_t) - (T-T_0)\big[2u(\xi) + \widetilde{u}_\epsilon(\xi)\big] - K(K-1)(\sqrt{T}+2)B_{\mathrm{KL}}. \tag{99}$$

Since $r_{t,a} = k_a(\widehat{\theta}_t, \lambda_t)$, the truncated AdaHedge bound in Eq. (79) implies

$$d_T \geq \max_{a\in\mathcal{A}} \sum_{t=T_0+1}^{T} k_a(\widehat{\theta}_t, \lambda_t) - \mathcal{R}_T^{\mathrm{AH}} - T_0 B_{\mathrm{KL}} - (T-T_0)\big[2u(\xi) + \widetilde{u}_\epsilon(\xi)\big] - K(K-1)(\sqrt{T}+2)B_{\mathrm{KL}}. \tag{100}$$

We next replace the first KL argument by $\theta^\star$. From the definition of $u(\xi)$,

$$k_a(\widehat{\theta}_t, \lambda_t) \geq k_a(\theta^\star, \lambda_t) - u(\xi).$$

The definition of $\widetilde{u}_\epsilon(\xi)$ also gives

$$k_a(\theta^\star, \lambda_t) \geq k_a(\theta^\star, \lambda_t^\circ) - \widetilde{u}_\epsilon(\xi).$$

Consequently,

$$\max_{a\in\mathcal{A}} \sum_{t=T_0+1}^{T} k_a(\widehat{\theta}_t, \lambda_t) \geq \max_{a\in\mathcal{A}} \sum_{t=T_0+1}^{T} k_a(\theta^\star, \lambda_t^\circ) - (T-T_0)\big[u(\xi) + \widetilde{u}_\epsilon(\xi)\big]. \tag{101}$$

Let ν_T be the empirical probability measure that assigns mass $1/(T-T_0)$ to each λ_t°, $t = T_0+1, \ldots, T$. Since every λ_t° belongs to $\mathrm{Alt}_{\epsilon-\xi}(\theta^\star)$, Lemma E.3 yields

$$\max_{a\in\mathcal{A}} \sum_{t=T_0+1}^{T} k_a(\theta^\star, \lambda_t^\circ) = (T-T_0) \max_{a\in\mathcal{A}} \int k_a(\theta^\star, \lambda)\, \nu_T(\mathrm{d}\lambda) \geq (T-T_0) C_{\epsilon-\xi}(\theta^\star). \tag{102}$$

Combining Eqs. (100), (101), and (102) proves Eq. (94). □

E.5 Completion of the proof

Proof of Theorem 5.2. Fix $0 < \xi < \min\{\epsilon, 1\}$ and define

$$c_\xi := C_{\epsilon-\xi}(\theta^\star) - 3u(\xi) - 2\widetilde{u}_\epsilon(\xi). \tag{103}$$

By Eq. (93), ξ can be chosen sufficiently small so that $c_\xi > 0$.

Let $T_0 := \tau_{\mathrm{loc}}(\xi)$. The pathwise information-growth bound in Eq. (94) can be written as

$$d_T \geq c_\xi T - (c_\xi + B_{\mathrm{KL}})T_0 - q_T, \tag{104}$$

where

$$q_T := \mathcal{R}_T^{\mathrm{AH}} + K(K-1)(\sqrt{T}+2)B_{\mathrm{KL}}.$$

Lemma E.2 implies that $q_T = o(T)$ deterministically.

Fix any $\gamma \in (0, c_\xi)$. There exists a deterministic constant $T_{\xi,\gamma}^{(0)} < \infty$ such that $q_T \leq \gamma T/2$ for every $T \geq T_{\xi,\gamma}^{(0)}$. Therefore, whenever

$$T \geq \max\left\{ T_{\xi,\gamma}^{(0)}, \frac{2(c_\xi + B_{\mathrm{KL}})}{\gamma} \tau_{\mathrm{loc}}(\xi) \right\},$$

Eq. (104) yields

$$d_T \geq (c_\xi - \gamma)T. \tag{105}$$

From the explicit form of f_T in Eq. (21), there exist deterministic constants $C_f, m_f < \infty$, depending only on the known parameter bounds, such that

$$f_T(x) \le C_f(T + x + 2)^{m_f} e^{-\kappa_B x}.$$

Since the right-hand side is decreasing in x for all sufficiently large $T + x$, Eq. (105) gives

$$\frac{\pi^2 T^2}{6} f_T(d_T) \le C_{\xi,\gamma} T^{m_f+2} e^{-\kappa_B(c_\xi - \gamma)T} \tag{106}$$

for a deterministic constant $C_{\xi,\gamma} < \infty$ and all sufficiently large T.

Let $\overline{T}_\delta(\xi, \gamma)$ be a deterministic time after which the right-hand side of Eq. (106) is smaller than δ. Taking logarithms shows that

$$\kappa_B(c_\xi - \gamma)\overline{T}_\delta(\xi, \gamma) = \log(1/\delta) + O(\log\log(1/\delta)),$$

and hence

$$\limsup_{\delta \downarrow 0} \frac{\overline{T}_\delta(\xi, \gamma)}{\log(1/\delta)} \le \frac{1}{\kappa_B(c_\xi - \gamma)}. \tag{107}$$

Consequently, for some deterministic constant $C^{(0)}_{\xi,\gamma} < \infty$,

$$\tau_\delta \le \overline{T}_\delta(\xi, \gamma) + \frac{2(c_\xi + B_{\mathrm{KL}})}{\gamma} \tau_{\mathrm{loc}}(\xi) + C^{(0)}_{\xi,\gamma}. \tag{108}$$

Since $\mathbb{E}_{\theta^\star}[\tau_{\mathrm{loc}}(\xi)] < \infty$ by Lemma E.4, this bound first implies $\mathbb{P}_{\theta^\star}(\tau_\delta < \infty) = 1$. Taking expectations in Eq. (108), dividing by $\log(1/\delta)$, and using Eq. (107) gives

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta^\star}[\tau_\delta]}{\log(1/\delta)} \le \frac{1}{\kappa_B(c_\xi - \gamma)}.$$

Finally, letting $\gamma \downarrow 0$ and then $\xi \downarrow 0$, together with Eq. (93), yields

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta^\star}[\tau_\delta]}{\log(1/\delta)} \le \frac{1}{\kappa_B C_\epsilon(\theta^\star)}.$$

This proves the theorem. □

F Finite Implementations of the Optimization Oracles

The theoretical description of FOCUS uses a joint-MLE oracle and an alternative-model best-response oracle. This section provides finite implementations of both procedures. The resulting algorithms require only finite enumeration and convex quadratic programming. They are exponential in p, as expected for unrestricted DAG search, but remove the need for abstract optimization oracles.

For an action $a \in \mathcal{A}$, let $\iota(a)$ denote its intervention target, with $\iota(0) = \varnothing$. Because $\boldsymbol{X}_s$ in the main text contains only the nondegenerate random coordinates, let $\widetilde{\boldsymbol{X}}_s \in \mathbb{R}^p$ denote the corresponding full vector obtained by reinserting the known intervention value when $a_s \neq 0$. Thus, $\widetilde{X}_{s,j}$ equals the assigned intervention value when $j = \iota(a_s)$ and equals the observed random coordinate otherwise.

F.1 Finite joint-MLE implementation

Under a candidate DAG D, the density of a sample collected under action a_s factorizes over the nodes that were not intervened upon. Ignoring constants independent of θ, its negative log-likelihood is

$$-\ell_t(\theta) = \frac{1}{2} \sum_{j=1}^{p} \sum_{\substack{s \le t: \\ \iota(a_s) \neq j}} \left[\log \sigma_j^2 + \frac{\left(\widetilde{X}_{s,j} - \boldsymbol{A}_{j,\mathrm{pa}_D(j)} \widetilde{\boldsymbol{X}}_{s,\mathrm{pa}_D(j)} \right)^2}{\sigma_j^2} \right]. \tag{109}$$

Algorithm 2 Finite implementation of the joint-MLE oracle

Require: History $\{(a_s, \boldsymbol{X}_s)\}_{s=1}^t$ and parameter bounds $\beta_{\min}$, $A_{\max}$, $\sigma_{\min}$, $\sigma_{\max}$
Ensure: A global joint MLE $\widehat{\theta}_t = (\widehat{D}_t, \widehat{\boldsymbol{A}}_t, \widehat{\boldsymbol{\Sigma}}_t)$

Reinsert each known intervention value to construct $\{\widetilde{\boldsymbol{X}}_s\}_{s=1}^t$
$\mathcal{L}_{\text{best}} \leftarrow +\infty$
$\widehat{\theta}_t \leftarrow$ null
for all permutations $\pi = (\pi_1, \dots, \pi_p)$ of $[p]$ **do**
  Initialize an empty candidate graph D_π
  $\mathcal{L}_\pi \leftarrow 0$
  for $m = 1, \dots, p$ **do**
    $j \leftarrow \pi_m$
    $\mathrm{Pred}_\pi(j) \leftarrow \{\pi_1, \dots, \pi_{m-1}\}$
    $\mathcal{L}_j^{\text{best}} \leftarrow +\infty$
    for all $S \subseteq \mathrm{Pred}_\pi(j)$ **do**
      if $S = \varnothing$ **then**
        Set $\boldsymbol{b}$ to the empty vector and compute $Q_{j,t}(\boldsymbol{b}; \varnothing)$
        Compute v_j and $\mathcal{L}_{j,t}$ from Eqs. (111)–(112)
        Store $S = \varnothing$, $\boldsymbol{b}$, v_j, and the local score as the best local solution if $\mathcal{L}_{j,t} < \mathcal{L}_j^{\text{best}}$
      else
        for all sign vectors $\boldsymbol{\nu} \in \{-1, +1\}^{|S|}$ **do**
          Solve the convex quadratic program

$$\boldsymbol{b}^\star \in \operatorname*{arg\,min}_{\boldsymbol{b} \in \mathbb{R}^{|S|}} Q_{j,t}(\boldsymbol{b}; S) \quad \text{s.t.} \quad \beta_{\min} \le \nu_k b_k \le A_{\max}, \;\; k \in S$$

          if the quadratic program is feasible **then**
            Compute $v_j^\star = v_{j,t}^\star(\boldsymbol{b}^\star; S)$ and $\mathcal{L}_{j,t}(\boldsymbol{b}^\star; S)$
            Store S, $\boldsymbol{b}^\star$, $v_j^\star$, and the local score as the best local solution if its score is smaller than $\mathcal{L}_j^{\text{best}}$
          end if
        end for
      end if
    end for
    Set $\mathrm{pa}_{D_\pi}(j)$, $\boldsymbol{A}_{j,\mathrm{pa}_{D_\pi}(j)}$, and σ_j^2 to the best local solution
    $\mathcal{L}_\pi \leftarrow \mathcal{L}_\pi + \mathcal{L}_j^{\text{best}}$
  end for
  if $\mathcal{L}_\pi < \mathcal{L}_{\text{best}}$ **then**
    Store $\widehat{\theta}_t \leftarrow (D_\pi, \boldsymbol{A}_\pi, \boldsymbol{\Sigma}_\pi)$
    $\mathcal{L}_{\text{best}} \leftarrow \mathcal{L}_\pi$
  end if
end for
return $\widehat{\theta}_t$

The structural equation of an intervened node is excluded because it is replaced by the intervention assignment. Nevertheless, that fixed coordinate remains available as a regressor in the structural equations of the other nodes.

For a node j, define

$$\mathcal{I}_{j,t} := \{s \in [t] : \iota(a_s) \neq j\}, \qquad n_{j,t} := |\mathcal{I}_{j,t}|.$$

Given a proposed parent set $S \subseteq [p] \setminus \{j\}$ and coefficient vector $\boldsymbol{b} \in \mathbb{R}^{|S|}$, let

$$Q_{j,t}(\boldsymbol{b}; S) := \sum_{s \in \mathcal{I}_{j,t}} \left(\widetilde{X}_{s,j} - \boldsymbol{b}^\mathsf{T} \widetilde{\boldsymbol{X}}_{s,S} \right)^2. \tag{110}$$

For fixed $\boldsymbol{b}$ and S, the optimal noise variance is

$$v_{j,t}^\star(\boldsymbol{b}; S) := \Pi_{[\sigma_{\min}^2, \sigma_{\max}^2]} \left(\frac{Q_{j,t}(\boldsymbol{b}; S)}{n_{j,t}} \right), \tag{111}$$

where $\Pi_{[l,u]}(x) := \min\{u, \max\{l, x\}\}$. If $n_{j,t} = 0$, the local likelihood is constant and we set $v^\star_{j,t}(\boldsymbol{b}; S) = \sigma^2_{\min}$ by convention.

Substituting Eq. (111) into the likelihood gives the local score

$$\mathcal{L}_{j,t}(\boldsymbol{b}; S) := \frac{1}{2}\left[n_{j,t} \log v^\star_{j,t}(\boldsymbol{b}; S) + \frac{Q_{j,t}(\boldsymbol{b}; S)}{v^\star_{j,t}(\boldsymbol{b}; S)}\right]. \tag{112}$$

The optimized score is nondecreasing in $Q_{j,t}(\boldsymbol{b}; S)$. Therefore, for a fixed support and coefficient-sign pattern, it suffices to minimize the convex quadratic function in Eq. (110).

Why Algorithm 2 is exact. For a fixed topological order, every parent of node j must occur among its predecessors. Enumerating all predecessor subsets therefore enumerates every DAG compatible with that order. Enumerating the coefficient signs converts the nonconvex constraint $\beta_{\min} \le |A_{jk}| \le A_{\max}$ into finitely many linear box constraints. Conditional on the order, parent set, and signs, the remaining coefficient problem is convex.

Moreover, Eq. (109) decomposes over nodes. Hence the best parent set and parameters for each node can be selected independently once the order is fixed. Finally, every DAG admits at least one topological order, so enumeration over all permutations includes every parameter in Θ. Algorithm 2 consequently returns a global maximizer of Eq. (12). Duplicate appearances of the same DAG under different topological orders do not affect correctness.

F.2 Local representation of the alternative objective

The second finite algorithm computes both the allocation best response in Eq. (13) and the stopping statistic in Eq. (19). Its generic inputs are:

$$\overline{\theta} = (\overline{D}, \overline{\boldsymbol{A}}, \overline{\boldsymbol{\Sigma}}) \in \Theta, \qquad \boldsymbol{w} = (w_a)_{a\in\mathcal{A}} \in \mathbb{R}^K_+, \qquad r > 0.$$

It solves

$$\inf_{\lambda \in \mathrm{Alt}_r(\overline{\theta})} \sum_{a\in\mathcal{A}} w_a \,\mathrm{KL}(P^{\overline{\theta}}_a \| P^{\lambda}_a). \tag{113}$$

Taking $\boldsymbol{w} = \boldsymbol{\alpha}_t$ and $r = \epsilon$ gives the best response λ_t. Taking $w_a = N_t(a)$ and $r = \epsilon$ gives d_t and its minimizing alternative.

For each action a, compute analytically under the reference parameter $\overline{\theta}$ the full-vector moments

$$\overline{\boldsymbol{\mu}}^{(a)} := \mathbb{E}_{\overline{\theta},a}[\boldsymbol{X}], \qquad \overline{\boldsymbol{H}}^{(a)} := \mathbb{E}_{\overline{\theta},a}[\boldsymbol{X}\boldsymbol{X}^\mathsf{T}] = \overline{\boldsymbol{V}}^{(a)} + \overline{\boldsymbol{\mu}}^{(a)}\overline{\boldsymbol{\mu}}^{(a)\mathsf{T}}. \tag{114}$$

For an intervention, the fixed coordinate is reinserted into the mean and second-moment matrix and has zero conditional variance. These quantities are available explicitly from Eq. (4).

For a proposed parent set S of node j, define

$$W_j := \sum_{\substack{a\in\mathcal{A}:\\ \iota(a)\neq j}} w_a, \tag{115}$$

$$\boldsymbol{G}_j(S) := \sum_{\substack{a\in\mathcal{A}:\\ \iota(a)\neq j}} w_a \overline{\boldsymbol{H}}^{(a)}_{S,S}, \tag{116}$$

$$\boldsymbol{h}_j(S) := \sum_{\substack{a\in\mathcal{A}:\\ \iota(a)\neq j}} w_a \overline{\boldsymbol{H}}^{(a)}_{S,j}, \tag{117}$$

$$g_j := \sum_{\substack{a\in\mathcal{A}:\\ \iota(a)\neq j}} w_a \overline{H}^{(a)}_{j,j}. \tag{118}$$

The weighted expected squared residual associated with coefficient vector $\boldsymbol{b} \in \mathbb{R}^{|S|}$ is

$$R_j(\boldsymbol{b}; S) := g_j - 2\boldsymbol{b}^\mathsf{T}\boldsymbol{h}_j(S) + \boldsymbol{b}^\mathsf{T}\boldsymbol{G}_j(S)\boldsymbol{b}. \tag{119}$$

Since $\boldsymbol{G}_j(S)$ is positive semidefinite, this is a convex quadratic function of $\boldsymbol{b}$.

For a candidate conditional variance v, the local contribution to the objective in Eq. (113) is

$$\Phi_j(\boldsymbol{b}, v; S) := \frac{W_j}{2} \log \frac{v}{\overline{\sigma}_j^2} + \frac{R_j(\boldsymbol{b}; S)}{2v} - \frac{W_j}{2}. \tag{120}$$

For a candidate parameter $\lambda = (D^\lambda, \boldsymbol{A}^\lambda, \boldsymbol{\Sigma}^\lambda)$, let $S_j = \mathrm{pa}_{D^\lambda}(j)$, $\boldsymbol{b}_j = \boldsymbol{A}^\lambda_{j,S_j}$, and $v_j = (\sigma_j^\lambda)^2$. The weighted KL objective decomposes as

$$\sum_{a \in \mathcal{A}} w_a \, \mathrm{KL}(P_a^{\overline{\theta}} \| P_a^\lambda) = \sum_{j=1}^{p} \Phi_j(\boldsymbol{b}_j, v_j; S_j). \tag{121}$$

When $W_j > 0$, its minimizing variance is

$$v_j^\star(\boldsymbol{b}; S) = \Pi_{[\sigma^2_{\min}, \sigma^2_{\max}]} \left(\frac{R_j(\boldsymbol{b}; S)}{W_j} \right). \tag{122}$$

The optimized value of Eq. (120) is nondecreasing in $R_j(\boldsymbol{b}; S)$. Thus, for each support and sign pattern, the coefficient vector is found by minimizing the convex quadratic function $R_j(\boldsymbol{b}; S)$ and the variance is subsequently obtained from Eq. (122).

If $W_j = 0$, the available weighted actions contain no likelihood contribution from the structural equation of node j. We then set its local score to zero and select any feasible coefficients and variance using a fixed deterministic tie-breaking rule.

For later use, define

$$\mathrm{LocalFit}(j, S, \mathcal{C}) \tag{123}$$

as the following finite subroutine: enumerate all sign vectors $\boldsymbol{\nu} \in \{-1, +1\}^{|S|}$, minimize $R_j(\boldsymbol{b}; S)$ subject to

$$\beta_{\min} \le \nu_k b_k \le A_{\max}, \qquad k \in S,$$

together with any additional linear constraints in $\mathcal{C}$, discard infeasible sign patterns, compute the optimal variance from Eq. (122), and return the feasible parameters having the smallest value of Eq. (120). For $S = \varnothing$, the coefficient vector is empty and $R_j(\boldsymbol{b}; S) = g_j$.

Whenever the local problem is feasible, we write

$$\big(\boldsymbol{b}_j^\star(S, \mathcal{C}), v_j^\star(S, \mathcal{C}), \Phi_j^\star(S, \mathcal{C})\big) := \mathrm{LocalFit}(j, S, \mathcal{C})$$

for its returned coefficients, variance, and optimized local score. When $\mathcal{C} = \varnothing$, we abbreviate $\Phi_j^\star(S, \varnothing)$ as $\Phi_j^\star(S)$.

Algorithm 3 Finite implementation of the alternative-model oracle

Require: Reference parameter $\overline{\theta} = (\overline{D}, \overline{\boldsymbol{A}}, \overline{\boldsymbol{\Sigma}})$, weights $\boldsymbol{w} = (w_a)_{a \in \mathcal{A}}$, and radius r
Ensure: Global optimizer $(\lambda^\star, \mathcal{V}^\star)$ of Eq. (113)

Compute $\{\overline{\boldsymbol{\mu}}^{(a)}, \overline{\boldsymbol{H}}^{(a)}\}_{a \in \mathcal{A}}$ using Eq. (114)
Compute the local quantities in Eqs. (115)–(119)
$(\lambda_A, \mathcal{V}_A) \leftarrow$ SAMEGRAPHALTERNATIVE$(\overline{\theta}, \boldsymbol{w}, r)$ (Algorithm 4)
$(\lambda_B, \mathcal{V}_B) \leftarrow$ DIFFERENTGRAPHALTERNATIVE$(\overline{\theta}, \boldsymbol{w})$ (Algorithm 5)
if $\mathcal{V}_A \le \mathcal{V}_B$ **then**
    $(\lambda^\star, \mathcal{V}^\star) \leftarrow (\lambda_A, \mathcal{V}_A)$
else
    $(\lambda^\star, \mathcal{V}^\star) \leftarrow (\lambda_B, \mathcal{V}_B)$
end if
return $(\lambda^\star, \mathcal{V}^\star)$

Algorithm 4 SAMEGRAPHALTERNATIVE: same-graph branch

Require: Reference parameter $\overline{\theta}$, weights $\boldsymbol{w}$, and radius r
Ensure: Best same-graph alternative $(\lambda_A, \mathcal{V}_A)$

$\mathcal{V}_A \leftarrow +\infty$ and $\lambda_A \leftarrow$ null
for $j = 1, \ldots, p$ **do**
  $\overline{S}_j \leftarrow \mathrm{pa}_{\overline{D}}(j)$
  for all $k \in \overline{S}_j$ and $d \in \{-1, +1\}$ **do**
    Define $\mathcal{C}_{j,k,d} := \{d(b_k - \overline{A}_{jk}) \geq r\}$
    Call $\mathrm{LocalFit}(j, \overline{S}_j, \mathcal{C}_{j,k,d})$
    if the local problem is feasible **then**
      Let $(\boldsymbol{b}^\star, v^\star, \Phi_j^\star(\overline{S}_j, \mathcal{C}_{j,k,d}))$ be the returned local fit
      Construct λ using the returned row and variance for node j
      Keep all other coefficient rows and noise variances equal to their reference values
      Let $\mathcal{V}$ be the returned local score
      if $\mathcal{V} < \mathcal{V}_A$ **then**
        $(\lambda_A, \mathcal{V}_A) \leftarrow (\lambda, \mathcal{V})$
      end if
    end if
  end for
end for
return $(\lambda_A, \mathcal{V}_A)$

F.3 FINITE ALTERNATIVE-MODEL IMPLEMENTATION

The alternative set in Eq. (6) is the union of two branches:

$$D^\lambda = \overline{D}, \quad \|\boldsymbol{A}^\lambda - \overline{\boldsymbol{A}}\|_{\max} \geq r,$$

and

$$D^\lambda \neq \overline{D}.$$

Algorithms 4 and 5 solve the two branches separately, and Algorithm 3 returns the better solution.

Explanation of Algorithm 4. When $D^\lambda = \overline{D}$, membership in $\mathrm{Alt}_r(\overline{\theta})$ requires at least one existing coefficient to differ from its reference value by at least r. This condition can be written as the finite union

$$b_k - \overline{A}_{jk} \geq r \qquad \text{or} \qquad \overline{A}_{jk} - b_k \geq r$$

over all existing edges $k \to j$. Each alternative is a linear constraint.

Only one structural row needs to be modified. Indeed, once node j and edge $k \to j$ satisfy the separation constraint, changing any other row is not needed for feasibility. Keeping those rows equal to the reference parameter gives zero local KL contribution, whereas optimizing them cannot reduce the total KL below their reference value. The constrained problem for the selected row is therefore sufficient to obtain the global optimum of the same-graph branch.

The coefficient optimizer may change several entries in row j, even though only one entry is required to deviate by at least r. This is intentional: allowing all coefficients in that row to adjust can reduce the KL objective while preserving the required separation.

Explanation of Algorithm 5. For a fixed topological order, every candidate parent set must be a subset of the node's predecessors. The local KL objective then decomposes across nodes, but the global constraint $D^\lambda \neq \overline{D}$ couples the local choices. Algorithm 5 handles this constraint using a two-state dynamic program:

$$q = 0 : \text{all parent sets chosen so far agree with } \overline{D},$$

and

$$q = 1 : \text{at least one chosen parent set differs from } \overline{D}.$$

Only state $q = 1$ is accepted at the end. This prevents the reference graph from being returned by the different-graph branch. Enumerating all topological orders ensures that every candidate DAG is considered.

Algorithm 5 DIFFERENTGRAPHALTERNATIVE: different-graph branch

Require: Reference parameter $\overline{\theta}$, weights $\boldsymbol{w}$
Ensure: Best different-graph alternative $(\lambda_B, \mathcal{V}_B)$

$\mathcal{V}_B \leftarrow +\infty$ and $\lambda_B \leftarrow$ null
for all permutations $\pi = (\pi_1, \dots, \pi_p)$ of $[p]$ **do**
  $F_0(0) \leftarrow 0$ and $F_0(1) \leftarrow +\infty$
  Initialize backtracking records
  for $m = 1, \dots, p$ **do**
    $j \leftarrow \pi_m$ and $\mathrm{Pred}_\pi(j) \leftarrow \{\pi_1, \dots, \pi_{m-1}\}$
    $F_m(0), F_m(1) \leftarrow +\infty$
    for all $q \in \{0, 1\}$ such that $F_{m-1}(q) < +\infty$ **do**
      for all $S \subseteq \mathrm{Pred}_\pi(j)$ **do**
        Call $\mathrm{LocalFit}(j, S, \varnothing)$
        if the local problem is feasible **then**
          Let $(\boldsymbol{b}^\star, v^\star, \Phi_j^\star(S))$ be the returned local fit
          $q' \leftarrow \max\{q, \mathbf{1}\{S \neq \mathrm{pa}_{\overline{D}}(j)\}\}$
          $\mathcal{V} \leftarrow F_{m-1}(q) + \Phi_j^\star(S)$
          if $\mathcal{V} < F_m(q')$ **then**
            $F_m(q') \leftarrow \mathcal{V}$
            Store $(S, \boldsymbol{b}^\star, v^\star)$ and the predecessor state for (m, q')
          end if
        end if
      end for
    end for
  end for
  if $F_p(1) < \mathcal{V}_B$ **then**
    Backtrack state 1 to reconstruct λ
    $(\lambda_B, \mathcal{V}_B) \leftarrow (\lambda, F_p(1))$
  end if
end for
return $(\lambda_B, \mathcal{V}_B)$

Using Algorithm 3 in FOCUS. At round t, the allocation best response is obtained by calling Algorithm 3 with

$$\overline{\theta} = \widehat{\theta}_t, \qquad w_a = \alpha_{t,a}, \qquad r = \epsilon.$$

The returned parameter is λ_t, and the action-wise gains are then computed as

$$r_{t,a} = \mathrm{KL}(P_a^{\widehat{\theta}_t} \| P_a^{\lambda_t}).$$

The stopping statistic is obtained from a second call with

$$\overline{\theta} = \widehat{\theta}_t, \qquad w_a = N_t(a), \qquad r = \epsilon.$$

In this case, the returned objective value is exactly

$$\mathcal{V}^\star = \inf_{\lambda \in \mathrm{Alt}_\epsilon(\widehat{\theta}_t)} \sum_{a \in \mathcal{A}} N_t(a)\, \mathrm{KL}(P_a^{\widehat{\theta}_t} \| P_a^{\lambda}) = d_t.$$

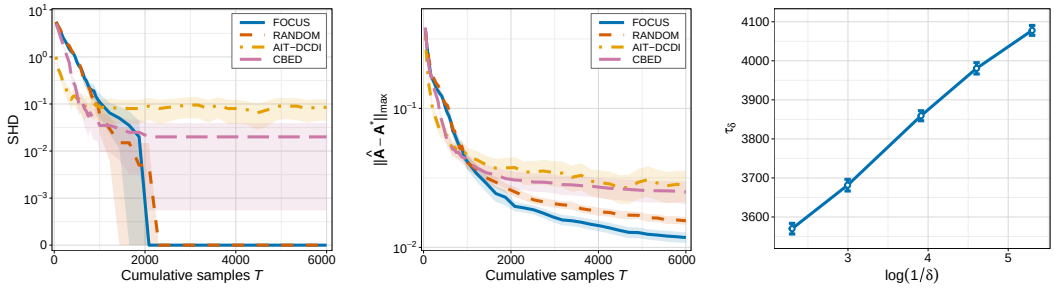
Thus, the same finite procedure implements both alternative-model optimization calls required by Algorithm 1.

Computational cost. Both algorithms are finite but exponential in the number of nodes. The joint-MLE procedure enumerates topological orders, parent sets, and coefficient-sign patterns. The alternative-model procedure additionally maintains the two graph-difference states. Every continuous subproblem is a convex quadratic program with linear constraints and can therefore be solved to global optimality.

These algorithms are intended to demonstrate that the oracle formulation does not hide an ill-defined or noncomputable operation. For the small graphs considered in our experiments, the finite procedures can be implemented directly. For substantially larger graphs, restricted parent-set sizes, order-based dynamic programming, or approximate continuous structure search may be used to reduce the computational cost, although such replacements no longer automatically inherit the exact-oracle guarantees of Theorems 5.1 and 5.2.

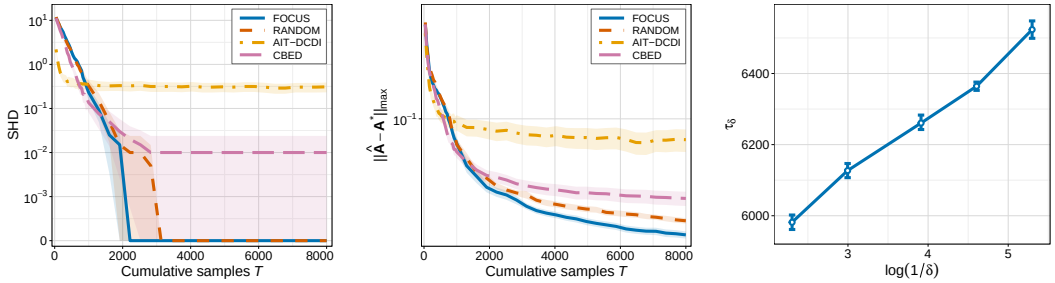
G Additional Experimental Results

Following the experimental protocol described in Section 6, we consider five settings: $(p, \rho) \in \{(5, 0.3), (6, 0.3), (7, 0.3), (6, 0.4), (6, 0.5)\}$. The representative setting $(p, \rho) = (6, 0.3)$ is reported in the main text, while the remaining four settings are shown in Figures 2–5. The additional results are consistent with the main-text experiment and lead to the same conclusions.



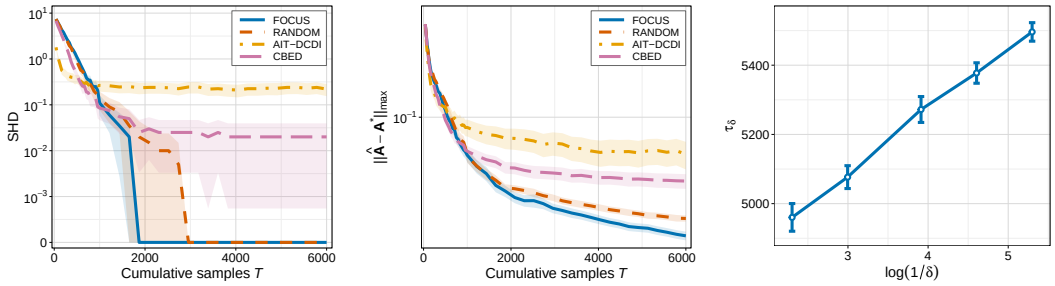
(a) Structure error (b) Edge-weight error (c) Stopping-time scaling

Figure 2: Performance for $p = 5$ and $\rho = 0.3$. Panels (a)–(c) show SHD, maximum edge-weight estimation error, and the stopping time of FOCUS versus $\log(1/\delta)$, respectively. Shaded bands and error bars denote 95% confidence intervals.



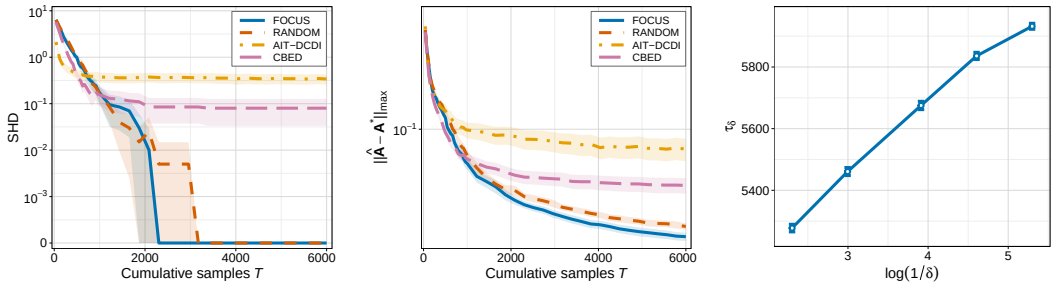
(a) Structure error (b) Edge-weight error (c) Stopping-time scaling

Figure 3: Performance for $p = 7$ and $\rho = 0.3$. Panels (a)–(c) show SHD, maximum edge-weight estimation error, and the stopping time of FOCUS versus $\log(1/\delta)$, respectively. Shaded bands and error bars denote 95% confidence intervals.



(a) Structure error (b) Edge-weight error (c) Stopping-time scaling

Figure 4: Performance for $p = 6$ and $\rho = 0.4$. Panels (a)–(c) show SHD, maximum edge-weight estimation error, and the stopping time of FOCUS versus $\log(1/\delta)$, respectively. Shaded bands and error bars denote 95% confidence intervals.



(a) Structure error (b) Edge-weight error (c) Stopping-time scaling

Figure 5: Performance for $p = 6$ and $\rho = 0.5$. Panels (a)–(c) show SHD, maximum edge-weight estimation error, and the stopping time of FOCUS versus $\log(1/\delta)$, respectively. Shaded bands and error bars denote 95% confidence intervals.