PAG-CBO: CAUSAL BAYESIAN OPTIMIZATION UNDER PARTIAL IDENTIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Causal Bayesian optimization uses causal structure and observational data to reduce costly experimentation. With hidden confounding and uncertain structure, however, intervention effects may remain only partially identified. We introduce PAG-CBO for systems whose structural uncertainty is represented by a partial ancestral graph (PAG). Our method combines structural pruning with bounds on intervention-response means, accounts for observational sampling uncertainty, and updates these bounds using interventional data. It eliminates actions that cannot be near-optimal and stops when the remaining uncertainty permits an ϵ-optimality certificate. Under a correct PAG and simultaneous confidence-band coverage, we prove certificate validity and localize a standard kernel-bandit simple-regret bound to the region surviving observational pruning. In discrete benchmarks, informative positive-width causal bounds can certify a recommendation before any intervention; when these bounds do not settle the decision, sequential experiments reduce the remaining uncertainty. An empirical calibration study records no false recommendations among 4,835 stopping claims under correctly specified PAGs. When informative causal bounds are unavailable, PAG-CBO uses the known outcome range and relies on experimental learning.

1 INTRODUCTION

Experiments in scientific and operational systems are often costly, while observational data are abundant but potentially confounded. Bayesian optimization (BO) addresses expensive objectives through probabilistic surrogates and sequential experimentation (Garnett, 2023; Srinivas et al., 2010). Causal Bayesian optimization (CBO) additionally uses causal structure and observational data to guide intervention selection (Aglietti et al., 2020). Extensions cover dynamic, model-based, constrained, functional, adversarial, and multi-objective settings (Aglietti et al., 2021; Sussex et al., 2023; Aglietti et al., 2023; Gultchin et al., 2023; Sussex et al., 2024; Bhatija et al., 2025). These methods typically assume enough structural knowledge to transfer observational information to intervention responses; others maintain uncertainty over candidate graphs (Branchini et al., 2023).

With latent confounding, observational discovery may recover only a *partial ancestral graph* (PAG), representing a Markov equivalence class of *maximal ancestral graphs* (MAGs) (Richardson & Spirtes, 2002; Zhang, 2008). Structural causal bandits use possibly-optimal minimal intervention sets (POMISs) to prune intervention targets (Lee & Bareinboim, 2018), including under incomplete causal knowledge (Malek et al., 2023; Elahi et al., 2024; Konobeev et al., 2025). Most relevant here, Park et al. (2025) characterize these targets under MAG/PAG equivalence. Other work avoids full graph recovery during optimization (Shahverdikondori et al., 2026) or extends CBO to uncertain structure (Branchini et al., 2023; Mukherjee et al., 2024; Durand et al., 2025). After structural pruning, a separate question remains: how much uncertainty about intervention responses must be resolved to make a good decision? The PAG identification algorithm IDP recovers interventional distributions when they are identifiable (Jaber et al., 2019); subsequent work also addresses conditional queries (Jaber et al., 2022). Partial-identification (PI) methods bound effects that remain ambiguous (Bellot, 2024). Bounds can aid sequential allocation (Zhang & Bareinboim, 2017; Sharma et al., 2020), but their use in CBO requires a link between PAG-compatible responses, finite-sample uncertainty, and decision certification. PAG-CBO develops this link. Appendix A discusses the closest approaches. For a target set $\mathbb{S}$ and intervention value $\boldsymbol{x}$, CBO minimizes the mean response $f_{\mathbb{S}}(\boldsymbol{x}) = \mathbb{E}[Y \mid \mathrm{do}(X_{\mathbb{S}} = \boldsymbol{x})]$. Identifying the full interventional marginal of Y identifies this mean, but a useful optimization decision may require less: intervals for the competing means can already

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.

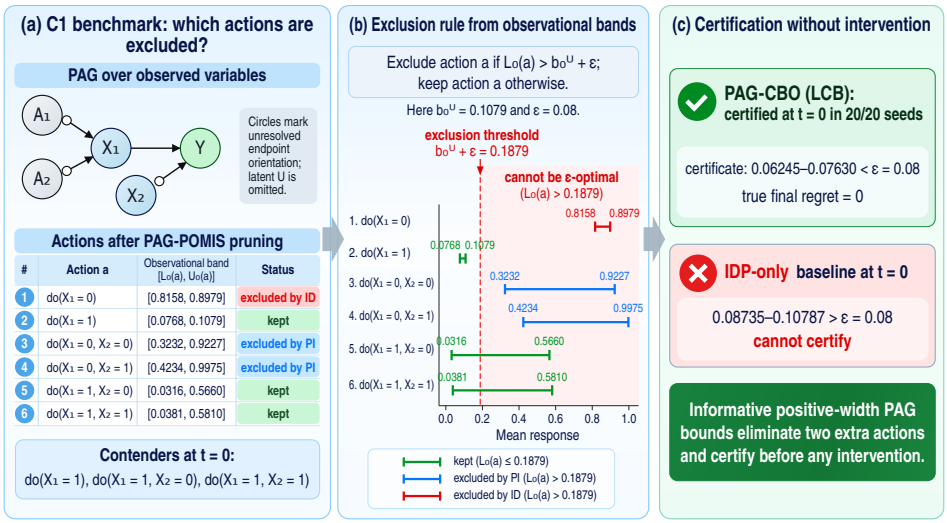

Figure 1: Motivating example: informative PAG PI bounds can certify an ϵ-optimal recommendation without intervention.

establish that one action is within ϵ of the optimum. We therefore consider the set of response means induced by all causal models compatible with the PAG and observational distribution. Its endpoint difference, the *optimization identification width*, measures the ambiguity that remains about the objective. Section 3.1 defines this set formally. Figure 1 illustrates why this distinction matters. In benchmark C1, structural pruning leaves six actions. Confidence bands for point-identified responses exclude one action, and informative partial-identification bands exclude two more under the same ϵ-contender rule. Although some responses retain positive identification width, the observational bands certify an ϵ-optimal recommendation before any intervention in all 20 runs. The IDP-only ablation, which uses the point-identification procedure of Jaber et al. (2019) and assigns the outcome range to other actions, cannot certify at this stage. Section 4 reports the comparison. Our contributions are threefold. **(i)** We formulate PAG-based causal optimization through identified sets of response means, distinguish distributional from mean identification, and characterize their equivalence for binary outcomes. **(ii)** We combine valid causal bounds with observational and interventional uncertainty to obtain safe contender pruning and a finite-sample ϵ-optimality certificate. We also localize a standard kernel-bandit regret bound to the region surviving observational pruning. **(iii)** We evaluate when partial-identification bounds resolve the decision, when interventions remain necessary, and when the method must fall back on the known outcome range. Calibration and misspecification studies examine both the validity and the limits of certification.

2 Problem Setup

Let $\mathbf{V} = (X_1, \ldots, X_d, Y)$ denote the observed variables of an underlying structural causal model (SCM) $M^\star$ that may contain latent variables, and let $p^\star(\mathbf{V})$ denote the population joint distribution in the observational regime. The available causal knowledge is a partial ancestral graph (PAG) $\mathcal{G}$, representing a Markov equivalence class of maximal ancestral graphs (MAGs).

Assumption 1 (Correct PAG and positivity)**.** *The MAG induced by $M^\star$ over $\mathbf{V}$ belongs to the equivalence class represented by $\mathcal{G}$, and there is no selection bias. The observational distribution satisfies the positivity conditions required by each identifying functional and bound construction used (Jaber et al., 2019; Bellot, 2024). For discrete variables, this means positive probability on the stated joint support; for continuous variables, identifying functionals must also be well defined at the admissible intervention values.*

All theoretical guarantees are conditional on Assumption 1. We do not model uncertainty in the causal-discovery procedure itself; Section 4 includes misspecified-PAG stress tests.

Assumption 2 (Bounded outcome)**.** *The outcome has a known finite range, rescaled without loss of generality so that $Y \in [0, 1]$ under the observational regime and every admissible intervention.*

The known range supplies a universal response envelope and conditionally sub-Gaussian intervention noise. Gaussian outcomes are not required. For unbounded raw measurements, a bounded utility must be specified as the objective; clipping is not an affine rescaling and can change the optimal action. Appendix D.1 states the transformations used in the experiments.

Interventions and objective. Let $\mathbb{X} \subseteq \{X_1, \ldots, X_d\}$ denote the manipulable observed variables. Let $\mathcal{T}_0 \subseteq 2^{\mathbb{X}}$ be the family of admissible intervention target sets. An action is $a = (\mathbb{S}, \boldsymbol{x})$, where $\mathbb{S} \in \mathcal{T}_0$ and $\boldsymbol{x} \in \mathbb{D}_{\mathbb{S}} \subseteq \mathbb{R}^{|\mathbb{S}|}$ is an admissible intervention value. Its mean response is

$$f_{\mathbb{S}}(\boldsymbol{x}) = \mathbb{E}_{M^\star}[Y \mid \mathrm{do}(X_{\mathbb{S}} = \boldsymbol{x})], \qquad f(a) := f_{\mathbb{S}}(\boldsymbol{x}). \tag{1}$$

The action space before structural pruning is $\mathbb{A}_0 = \bigcup_{\mathbb{S}\in\mathcal{T}_0} \{\mathbb{S}\} \times \mathbb{D}_{\mathbb{S}}$. If the no-intervention action is admissible, $\varnothing \in \mathcal{T}_0$ has the singleton set $\mathbb{D}_\varnothing$ containing the empty tuple. We minimize the response and seek $a^\star \in \arg\min_{a\in\mathbb{A}_0} f(a)$. For a recommendation $\widehat{a}$, its simple regret is $r(\widehat{a}) = f(\widehat{a}) - f(a^\star)$. Given a user-specified $\epsilon > 0$, $\widehat{a}$ is ϵ-optimal when $r(\widehat{a}) \leq \epsilon$. We seek such a certificate with as few interventional outcomes as possible. We assume that the minima used below are attained; numerical approximations require the qualifications in Section 3.3.

Observational and interventional data. Before optimization, we observe $\mathbb{D}_{\mathrm{obs}} = \{\mathbf{V}_i\}_{i=1}^{n_O}$ with $\mathbf{V}_i \overset{\text{i.i.d.}}{\sim} p^\star(\mathbf{V})$. At round t, the algorithm selects $a_t = (\mathbb{S}_t, \boldsymbol{x}_t)$ and collects a batch of $m_t \geq 1$ outcomes that, conditional on the past and a_t, are independent and satisfy $Y_{t,j} \sim p^\star(Y \mid \mathrm{do}(X_{\mathbb{S}_t} = \boldsymbol{x}_t))$, $j = 1, \ldots, m_t$. Single-observation rounds are the special case $m_t = 1$.

Structural reduction. A *possibly-optimal minimal intervention set* (POMIS) is a minimal intervention target set that can be optimal for at least one causal model consistent with the available structure (Lee & Bareinboim, 2018). For MAG/PAG equivalence classes, Park et al. (2025) provide a sound-and-complete graphical characterization. Since our optimization problem is stated as minimization, we apply the corresponding maximization result to the reward $1 - Y$. POMIS pruning concerns intervention target sets; for each surviving target we retain its full admissible value domain $\mathbb{D}_{\mathbb{S}}$. Let $\mathcal{T} = \mathrm{POMIS}(\mathcal{G}, Y)$ and define the reduced action space $\mathbb{A} = \bigcup_{\mathbb{S}\in\mathcal{T}} \{\mathbb{S}\} \times \mathbb{D}_{\mathbb{S}}$. We require this reduction to preserve the minimum over $\mathbb{A}_0$:

$$\min_{a\in\mathbb{A}} f(a) = \min_{a\in\mathbb{A}_0} f(a). \tag{2}$$

PAG-POMIS soundness supplies this property for the outcome-only intervention setting covered by its characterization. Arbitrary restrictions on intervention values, feasibility constraints, or action-dependent penalties require a separate justification; if preservation is not established, let $\mathcal{T}$ contain all admissible targets and retain $\mathbb{A} = \mathbb{A}_0$. All guarantees below require equation 2. We choose $a^\star \in \mathbb{A}$ as a global optimizer. Structural pruning reduces the targets to consider, while response identification determines what can be learned about their values.

3 PAG-CBO: Identification, Optimization, and Certification

PAG-CBO first bounds the responses of the surviving actions, then combines those bounds with experimental confidence intervals. The resulting intervals support both intervention selection and a stopping certificate.

3.1 Optimization-response identification under the PAG

For an observational distribution $p(\mathbf{V})$ and PAG $\mathcal{G}$, let $\mathbb{M}(\mathcal{G}, p)$ denote the set of SCMs, possibly containing latent variables, whose induced MAG over $\mathbf{V}$ belongs to the class represented by $\mathcal{G}$, whose observed marginal is $p(\mathbf{V})$, and whose outcome obeys Assumption 2. Here the induced MAG represents the conditional independences among observed variables entailed by the causal graph; it need not coincide with the usual acyclic directed mixed graph (ADMG) latent projection. We consider only p for which this model class is nonempty. For an action $a = (\mathbb{S}, \boldsymbol{x}) \in \mathbb{A}$, define the *optimization-response identified set*

$$\mathbb{Q}_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}) = \{\mathbb{E}_M[Y \mid \mathrm{do}(X_{\mathbb{S}} = \boldsymbol{x})] : M \in \mathbb{M}(\mathcal{G}, p)\}. \tag{3}$$

At the population truth, we write $\mathbb{Q}_{\mathbb{S}}(\boldsymbol{x}) := \mathbb{Q}_{\mathbb{S}}(\boldsymbol{x}; p^\star, \mathcal{G})$ whenever the dependence on $(p^\star, \mathcal{G})$ is clear. Define $\underline{f}_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}) = \inf \mathbb{Q}_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G})$ and $\overline{f}_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}) = \sup \mathbb{Q}_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G})$. The corresponding *optimization identification width* is

$$w_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}) = \overline{f}_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}) - \underline{f}_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}). \tag{4}$$

We suppress $(p^\star, \mathcal{G})$ from these quantities at the population truth. Under Assumption 1, $M^\star \in \mathbb{M}(\mathcal{G}, p^\star)$, so $\mathbb{Q}_{\mathbb{S}}(\boldsymbol{x})$ is nonempty and contains the true response $f_{\mathbb{S}}(\boldsymbol{x})$. Under Assumption 2, its envelope $[\underline{f}_{\mathbb{S}}(\boldsymbol{x}), \overline{f}_{\mathbb{S}}(\boldsymbol{x})]$ lies in $[0, 1]$. We do not require $\mathbb{Q}_{\mathbb{S}}(\boldsymbol{x})$ itself to be an interval.

A response is point identified at $\boldsymbol{x}$ when $w_{\mathbb{S}}(\boldsymbol{x}) = 0$ and remains partially identified when $w_{\mathbb{S}}(\boldsymbol{x}) > 0$. Thus the width quantifies ambiguity in the causal quantity actually optimized rather than ambiguity in the entire causal model or interventional distribution. Positive width need not imply unresolved optimization: non-singleton response sets may still be sufficiently separated across actions to certify an ϵ-optimal recommendation.

What IDP implies for the mean. The soundness and completeness of IDP concern interventional distributions (Jaber et al., 2019). PAG-CBO needs only their means, so success implies:

Theorem 1 (IDP success implies zero optimization identification width)**.** *Under Assumption 1, if IDP identifies the query $p(Y \mid \mathrm{do}(X_{\mathbb{S}} = \boldsymbol{x}))$ as a functional of $p^\star$, then $\mathbb{Q}_{\mathbb{S}}(\boldsymbol{x})$ is a singleton and $w_{\mathbb{S}}(\boldsymbol{x}) = 0$.*

Remark 1 (Graphical nonidentification versus ambiguity at a fixed distribution)**.** *IDP failure rules out a single identifying functional valid over the PAG model class. It supplies observationally equivalent witness models at some compatible distribution, not necessarily at the particular $p^\star$. Moreover, distinct interventional distributions can have the same mean. Thus failure alone does not establish $w_{\mathbb{S}}(\boldsymbol{x}; p^\star, \mathcal{G}) > 0$ at a specified intervention value.*

For binary outcomes, distributional and mean ambiguity coincide at each fixed observational distribution and action.

Corollary 1 (Binary-outcome equivalence)**.** *Suppose every model in $\mathbb{M}(\mathcal{G}, p)$ has $Y \in \{0, 1\}$ under the intervention. Then $w_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}) = 0$ if and only if all models in $\mathbb{M}(\mathcal{G}, p)$ agree on $p_M(Y \mid \mathrm{do}(X_{\mathbb{S}} = \boldsymbol{x}))$. Consequently, nonidentification of the query family in this binary-outcome class implies that there exist a compatible positive observational distribution p and an intervention value $\boldsymbol{x}$ in that family for which $w_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}) > 0$. This conclusion applies to an IDP failure when its completeness result covers the stated model class and query family; it does not assert positive width at every $p^\star$ or within an arbitrarily restricted intervention domain.*

Valid population response bounds. The sharp endpoints of $\mathbb{Q}_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G})$ need not be available in closed form. PAG-CBO therefore permits any valid outer bound

$$\mathbb{B}^O_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}) = [\ell^O_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}), u^O_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G})] \supseteq [\underline{f}_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}), \overline{f}_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G})]. \tag{5}$$

A bound construction is *valid* on a stated model class when equation 5 holds for every admissible $(p, \mathcal{G})$ to which the construction is applied. The optimization theory requires validity, not sharpness.

When IDP succeeds, integrating Y against the identified interventional distribution yields a mean functional $\Phi_{\mathbb{S}}(p; \boldsymbol{x})$, and the outer interval collapses to $\ell^O_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}) = u^O_{\mathbb{S}}(\boldsymbol{x}; p, \mathcal{G}) = \Phi_{\mathbb{S}}(p; \boldsymbol{x})$. When point identification is unavailable, PAG-level partial-identification methods may provide informative outer bounds where their assumptions apply. For discrete benchmarks, we use the analytical PAG bounds of Bellot (2024) where applicable. Appendix D.3 compares the resulting finite-sample bands with exact linear-programming (LP) reference intervals for a specified canonical SCM class. That comparison does not establish sharpness over the full PAG model class.

If no informative valid bound is available, the known outcome range $[0, 1]$ provides a conservative fallback. PAG-CBO first seeks a point-identified response. If point identification fails, it uses a valid PI bound when available; otherwise, it retains the known outcome range. The population width characterizes causal ambiguity; the algorithm uses valid outer bands, which may be wider. In particular, use of the $[0, 1]$ fallback does not imply that the sharp identified set fills that range.

3.2 FINITE-SAMPLE OBSERVATIONAL AND INTERVENTIONAL BANDS

The population quantities above depend on $p^\star(\mathbf{V})$, whereas PAG-CBO observes only $\mathbb{D}_{\text{obs}}$. Both identified mean functionals and partial-identification endpoints therefore carry sampling uncertainty.

Assumption 3 (Simultaneous observational outer-band coverage)**.** *From $(\mathbb{D}_{\text{obs}}, \mathcal{G})$ we construct functions $L^O_{n_O}(a)$ and $U^O_{n_O}(a)$ such that, for $\delta_O \in (0,1)$,*

$$\mathcal{E}_O := \left\{ \forall a = (\mathbb{S}, \boldsymbol{x}) \in \mathbb{A} :\ L^O_{n_O}(a) \le \underline{f}_{\mathbb{S}}(\boldsymbol{x}) \le \overline{f}_{\mathbb{S}}(\boldsymbol{x}) \le U^O_{n_O}(a) \right\}, \qquad \Pr(\mathcal{E}_O) \ge 1 - \delta_O. \tag{6}$$

For point-identified responses, these may be simultaneous confidence bands for $\Phi_{\mathbb{S}}(p^\star; \cdot)$. For partially identified responses, they must also account for uncertainty in the estimated lower and upper endpoints. For a finite action space, simultaneous coverage can be obtained, for example, by a union bound over action-wise confidence intervals; continuous action spaces require uniform confidence bands over the relevant intervention domains. Inference for partially identified quantities can be non-regular near boundary cases (Imbens & Manski, 2004; Stoye, 2009a; Chernozhukov et al., 2007), so the required coverage must be justified for the particular construction used. The conservative choice $L^O_{n_O}(a) = 0$ and $U^O_{n_O}(a) = 1$ always qualifies under Assumption 2.

Interventions provide a second source of response bounds. Their intersection with the observational bands can be narrower, although successive intersections need not be nested over time.

Assumption 4 (Reproducing-kernel Hilbert space (RKHS) regularity)**.** *The kernels and norm bounds are fixed before collecting intervention outcomes. For each retained target $\mathbb{S}$, $f_{\mathbb{S}} \in \mathcal{H}_{\kappa_{\mathbb{S}}}$ with $\|f_{\mathbb{S}}\|_{\kappa_{\mathbb{S}}} \le B_{\mathbb{S}}$ and $\kappa_{\mathbb{S}}(\boldsymbol{x}, \boldsymbol{x}) \le 1$ for every $\boldsymbol{x} \in \mathbb{D}_{\mathbb{S}}$.*

Define the block kernel on $\mathbb{A}$ by $\kappa((\mathbb{S}, \boldsymbol{x}), (\mathbb{S}', \boldsymbol{x}')) = \mathbf{1}\{\mathbb{S} = \mathbb{S}'\}\kappa_{\mathbb{S}}(\boldsymbol{x}, \boldsymbol{x}')$. Then f belongs to the corresponding direct-sum RKHS with $B := \left(\sum_{\mathbb{S} \in \mathcal{T}} B_{\mathbb{S}}^2\right)^{1/2}$ as a valid norm bound.

Index individual intervention outcomes by $z_1, \dots, z_n$ and their queried actions by $\bar{a}_1, \dots, \bar{a}_n$, including repetitions within batches. With $K_n = (\kappa(\bar{a}_i, \bar{a}_j))_{i,j}$ and $k_n(a) = (\kappa(\bar{a}_i, a))_i$, define the zero-mean kernel ridge-regression quantities

$$\begin{aligned} \mu_n(a) &= k_n(a)^\top (K_n + \lambda I)^{-1} z_{1:n}, \\ \sigma_n^2(a) &= \kappa(a, a) - k_n(a)^\top (K_n + \lambda I)^{-1} k_n(a). \end{aligned} \tag{7}$$

We fix $\lambda > 0$, and set $\mu_0 = 0$ and $\sigma_0^2(a) = \kappa(a, a)$. These are Gaussian process (GP) posterior formulas under a working Gaussian likelihood; the guarantees use RKHS concentration and bounded noise instead. Define the realized information gain

$$\widehat{\gamma}_n = \frac{1}{2} \log\det(I + \lambda^{-1} K_n), \qquad \widehat{\gamma}_0 = 0. \tag{8}$$

By Assumption 2, the centered intervention noise is conditionally R-sub-Gaussian with $R = 1/2$. For $\delta_I \in (0,1)$, use the data-dependent anytime confidence parameter

$$\beta_{n+1} = \left(B + \frac{R}{\sqrt{\lambda}} \sqrt{2 \left(\widehat{\gamma}_n + 1 + \ln \frac{1}{\delta_I} \right)} \right)^2, \qquad R = \frac{1}{2}. \tag{9}$$

The regularized self-normalized argument underlying Chowdhury & Gopalan (2017) yields the following event; Appendix B.2 derives it for the realized information gain and fixed λ used here:

$$\mathcal{E}_I = \left\{ \forall n \ge 0,\ \forall a \in \mathbb{A} : |f(a) - \mu_n(a)| \le \sqrt{\beta_{n+1}}\, \sigma_n(a) \right\}, \qquad \Pr(\mathcal{E}_I) \ge 1 - \delta_I. \tag{10}$$

Let $N_t = \sum_{s=1}^t m_s$ denote the number of intervention outcomes available after round t. For $N_t > 0$, intersect the observational and intervention-derived intervals:

$$\begin{aligned} L_t(a) &= \max\left\{ L^O_{n_O}(a), \mu_{N_t}(a) - \sqrt{\beta_{N_t+1}} \sigma_{N_t}(a) \right\}, \\ U_t(a) &= \min\left\{ U^O_{n_O}(a), \mu_{N_t}(a) + \sqrt{\beta_{N_t+1}} \sigma_{N_t}(a) \right\}. \end{aligned} \tag{11}$$

At $t = 0$, we deliberately take the intervention-derived interval as uninformative and set $L_0 = L^O_{n_O}$ and $U_0 = U^O_{n_O}$. Thus a certificate at $t = 0$ is based entirely on the PAG and observational data. We allocate the confidence budget so that $\delta_O + \delta_I \leq \delta$. On the resulting joint coverage event, equation 11 is nonempty for every action. If an empty intersection occurs, PAG-CBO records the event, reverts that action to its observational interval $[L^O_{n_O}(a), U^O_{n_O}(a)]$, which remains valid on $\mathcal{E}_O$, and suppresses certification for that round. Empty intersections are therefore treated as incompatibility diagnostics rather than as a test of PAG correctness; Appendix D.8 reports their behavior under stress tests.

3.3 Decision state and PAG-CBO algorithm

For minimization, define

$$b^L_t = \inf_{a \in \mathbb{A}} L_t(a), \qquad b^U_t = \inf_{a \in \mathbb{A}} U_t(a). \tag{12}$$

The ϵ-contender set is

$$\mathcal{C}^\epsilon_t = \left\{ a \in \mathbb{A} : L_t(a) \leq b^U_t + \epsilon \right\}. \tag{13}$$

An action outside $\mathcal{C}^\epsilon_t$ is already too poor, even under its most favorable currently admissible response, to be ϵ-optimal. Let $\widehat{a}_t \in \arg\min_{a \in \mathbb{A}} U_t(a)$ and define the decision-certificate gap

$$\widetilde{R}_t = U_t(\widehat{a}_t) - b^L_t = b^U_t - b^L_t. \tag{14}$$

PAG-CBO certifies and stops when $\widetilde{R}_t \leq \epsilon$ and no empty-intersection diagnostic has blocked certification in that round. These definitions use the full reduced action space $\mathbb{A}$. For numerical optimization on a finite grid $\mathbb{A}_h \subseteq \mathbb{A}$, replacing $\mathbb{A}$ by $\mathbb{A}_h$ defines $\widehat{a}_{t,h}$ and $\widetilde{R}_{t,h}$. The resulting certificate is relative to the grid optimum. Relative to the original optimum, it gives

$$f(\widehat{a}_{t,h}) - f(a^\star) \leq \widetilde{R}_{t,h} + \tau_h, \qquad \tau_h = \min_{a \in \mathbb{A}_h} f(a) - f(a^\star). \tag{15}$$

A full-domain ϵ-certificate therefore requires a valid bound $\overline{\tau}_h \geq \tau_h$ and the stopping test $\widetilde{R}_{t,h} + \overline{\tau}_h \leq \epsilon$. Without such a bound, certification is limited to the grid. Approximate numerical minimization likewise requires certified optimization-error bounds. For finite benchmarks whose grid is the full action space, $\tau_h = 0$. When certification has not occurred, the default policy uses the optimistic lower-confidence-bound (LCB) rule

$$a_{t+1} \in \arg\min_{a \in \mathbb{A}} L_t(a). \tag{16}$$

On the joint coverage event, $L_t(a) \leq U_t(a)$ for every a, so each global minimizer of L_t belongs to $\mathcal{C}^0_t$. Restricting equation 16 to $\mathcal{C}^0_t$ therefore produces the same optimistic action on that event when ties are resolved consistently. The contender set describes which actions can be excluded from the current search. Since the bands are not necessarily nested, permanent removal would require retaining past valid bounds; here contenders are recomputed each round. The complete PAG-CBO procedure is summarized in Algorithm 1. Here $\mathsf{ObsBand}(\mathcal{G}, \mathbb{D}_{\text{obs}}, \mathbb{A}, \delta_O)$ denotes a prescribed observational-band construction that uses, in order, IDP point identification, a valid PI bound, or the outcome-range fallback, with simultaneous coverage across all targets and values as required by Assumption 3. The level δ_O is a global error budget, not a separate budget for each target.

Algorithm 1 checks certification both before the first experiment and after the final batch, including when the intervention budget is zero. A grid-based procedure uses $\mathbb{A}_h$ throughout and reports grid certification unless the correction in equation 15 is supplied. The certificate applies for arbitrary batch sizes. The regret analysis below specializes to $m_t = 1$, matching the standard sequential kernel-bandit setting. The experiments also consider an optional model-based decision-value-of-information acquisition; its likelihood assumptions and acquisition rule are given in Appendix C. It does not enter any guarantee below.

3.4 Theoretical guarantees

Proposition 1 (Safety of contender pruning)**.** *Under equation 2, on any event for which $L_t(a) \leq f(a) \leq U_t(a)$ for every $a \in \mathbb{A}$, every ϵ-optimal action in $\mathbb{A}$ belongs to $\mathcal{C}^\epsilon_t$.*

This guarantee concerns response-based pruning within $\mathbb{A}$; structural pruning need preserve only one optimum, not every near-optimal action in $\mathbb{A}_0$.

Algorithm 1 PAG-CBO on a fixed action domain

Inputs: PAG $\mathcal{G}$, outcome Y, admissible targets and domains, observations $\mathbb{D}_{\text{obs}}$, tolerance ϵ, global confidence budgets δ_O, δ_I, outcome budget $\mathcal{B}_I \in \mathbb{N}_0$, batch sizes, fixed kernels, regularization, RKHS norm bounds, and observational-band construction ObsBand.

1. Construct $\mathbb{A}$ by sound structural pruning, or retain $\mathbb{A}_0$ if equation 2 is not established.
2. Construct $L^O_{n_O}, U^O_{n_O}$ jointly over $\mathbb{A}$ using the global budget δ_O. Set $t = 0$, $N_0 = 0$, $L_0 = L^O_{n_O}$, and $U_0 = U^O_{n_O}$.
3. At the current round, record any action with $L_t(a) > U_t(a)$ and restore its observational interval. If any such action exists, block certification for this round.
4. Compute $\mathcal{C}^\epsilon_t$, $\widehat{a}_t$, and $\widetilde{R}_t$. If $\widetilde{R}_t \le \epsilon$ and certification is not blocked, return the recommendation, gap, and status CERTIFIED.
5. If $N_t = \mathcal{B}_I$, return the recommendation, gap, contender set, and status UNCERTIFIED.
6. Select $a_{t+1} \in \arg\min_{a\in\mathbb{A}} L_t(a)$ and collect $m_{t+1} \leftarrow \min\{m_{t+1}, \mathcal{B}_I - N_t\}$ independent outcomes there.
7. Set $N_{t+1} = N_t + m_{t+1}$; update kernel regression and the confidence coefficient; compute the combined bands in equation 11; increase t by one and return to step 3.

Theorem 2 (Finite-sample ϵ-decision certificate)**.** *Under Assumptions 1, 2, 3, and 4, and equation 2, choose $\delta_O + \delta_I \le \delta$ and the anytime intervention confidence sequence in equation 10. Then, with probability at least $1-\delta$, simultaneously at every round,*

$$f(\widehat{a}_t) - f(a^\star) \le \widetilde{R}_t. \tag{17}$$

Consequently, $\widetilde{R}_t \le \epsilon$ certifies that $\widehat{a}_t$ is ϵ-optimal.

Corollary 2 (Zero-intervention certification)**.** *Under Assumptions 1–2 and 3, and equation 2, with probability at least $1-\delta_O$, any recommendation certified by $\widetilde{R}_0 \le \epsilon$ is ϵ-optimal. No intervention outcomes or RKHS assumption are needed for this statement.*

Corollary 2 formalizes the distinction between causal and optimization ambiguity. Individual intervention responses may remain partially identified while their remaining uncertainty is already too small to change the decision by more than ϵ.

Regret localization. To isolate the role of observational PAG information in intervention complexity, let $\mathbb{A}_h \subseteq \mathbb{A}$ be the finite candidate set used by the theoretical optimizer and let $a^\star_h \in \arg\min_{a\in\mathbb{A}_h} f(a)$. Define

$$\tau_h = f(a^\star_h) - f(a^\star) \ge 0. \tag{18}$$

For a finite action space with $\mathbb{A}_h = \mathbb{A}$, $\tau_h = 0$. For a finite nonempty domain $\mathcal{D}$, define its maximum information gain by

$$\gamma_T(\mathcal{D};\kappa) = \max_{(a_1,\dots,a_T)\in\mathcal{D}^T} \frac{1}{2}\log\det\left(I + \lambda^{-1}K_T\right), \tag{19}$$

where repeated query locations are permitted, $\gamma_0 = 0$, and the same λ is used as above. All information-gain domains used below are finite, so the maximum exists. Define the observationally surviving grid

$$\mathcal{C}_{O,h} = \left\{ a \in \mathbb{A}_h : L^O_{n_O}(a) \le \inf_{a'\in\mathbb{A}_h} U^O_{n_O}(a') \right\}. \tag{20}$$

For the sequential rule, the tth action is selected from the confidence state after $t-1$ intervention outcomes. Define the computable confidence radius $q_t = \sqrt{\beta_t}\sigma_{t-1}(a_t)$, choose $\widehat{t} \in \arg\min_{1\le t\le T} q_t$, and set $a^{\text{rec}}_T := a_{\widehat{t}}$. On the joint coverage event, all queried actions lie in $\mathcal{C}_{O,h}$. Consequently, $\widehat{\gamma}_{T-1} \le \gamma_{T-1}(\mathcal{C}_{O,h};\kappa)$. Define the corresponding localized upper bound

$$\overline{\beta}_{T,h} = \left(B + \frac{R}{\sqrt{\lambda}}\sqrt{2\left(\gamma_{T-1}(\mathcal{C}_{O,h};\kappa) + 1 + \ln\frac{1}{\delta_I}\right)} \right)^2. \tag{21}$$

Theorem 3 (Simple regret localized to the PAG-surviving region)**.** *Consider $T \geq 1$ sequential queries ($m_t = 1$) selected by equation 16 over $\mathbb{A}_h$, before stopping or along a continuation of that rule. Under Assumptions 1, 2, 3, and 4, and equation 2, use the confidence sequence in equation 9. Then, with probability at least $1 - \delta_O - \delta_I$,*

$$f(a_T^{\mathrm{rec}}) - f(a^\star) \leq C_\lambda \sqrt{\frac{\overline{\beta}_{T,h}\, \gamma_T(\mathcal{C}_{O,h}; \kappa)}{T}} + \tau_h, \qquad C_\lambda = 2\sqrt{\frac{2}{\ln(1+\lambda^{-1})}}. \tag{22}$$

The recommendation a_T^{rec} is computable from the query history; it need not coincide with the certificate recommendation $\widehat{a}_T \in \arg\min_a U_T(a)$. PAG-CBO returns $\widehat{a}_T$ together with its certificate gap $\widetilde{R}_T$; Theorem 3 additionally identifies a queried action with the stated high-probability simple-regret guarantee. Since $\mathcal{C}_{O,h} \subseteq \mathbb{A}_h$, both $\gamma_T(\mathcal{C}_{O,h}; \kappa)$ and the information-gain term entering $\overline{\beta}_{T,h}$ are no larger than their full-domain counterparts. For fixed kernels, norm bounds, regularization, and confidence levels, observational pruning therefore cannot increase this regret upper bound. It can reduce the bound when the removed actions contribute to maximum information gain. If $\widetilde{R}_0 \leq \epsilon$, Corollary 2 gives the stronger conclusion of certification with no intervention. Theorem 3 is a localization of standard kernel-bandit analysis, not a new GP-UCB rate. A sharper instance-dependent analysis that exploits the full width profile $U_{n_O}^O(a) - L_{n_O}^O(a)$ across scales is left for future work.

4 Experiments

We evaluate four questions: *(i)* do the observational response bands contain the canonical SCM reference intervals; *(ii)* can positive-width partial-identification information change the optimization decision before intervention; *(iii)* are the finite-sample observational bands and decision certificates empirically calibrated; and *(iv)* does observational pruning produce the localization predicted by Theorem 3? Additional benchmark and protocol details, baseline results, sensitivity studies, and robustness audits are in Appendix D.

Setup and metrics. The E0–E6 suite includes continuous and discrete SCMs with point identification, informative partial identification, and weak observational bounds. It combines adaptations of the CBO and Causal Entropy Optimization benchmarks (Aglietti et al., 2020; Branchini et al., 2023), a discrete example from Bellot (2024), and synthetic controls. The two-treatment instances C1 and C2 share a graph but have different response probabilities. Appendix D.1 gives the sources, complete structural equations, intervention domains, and graph assumptions. Primary PAG-CBO optimization results use 20 seeds; the optional DVI acquisition uses the prespecified five-seed subset. The separate calibration study uses E0, C1, C2, and E6 with $n_O \in \{300, 1000, 3000\}$, $\delta_O = \delta_I \in \{0.02, 0.05, 0.10\}$, and 200 fresh observational samples per configuration, for 36 cells. The decision tolerances are fixed in advance: $\epsilon = 0.08$ for C1/C2 and $\epsilon = 0.05$ for E0/E6. For continuous benchmarks, the stopping rule is evaluated on the candidate grid. Its formal guarantee is relative to that grid unless the correction in equation 15 is supplied; an empirical check against a known SCM optimum does not replace this correction.

We report simple regret relative to the benchmark reference optimum, the certificate gap (using $\widetilde{R}_{t,h}$ on grids), the number of intervention outcomes N_t, the number of intervention batches t, contender size, observational-band coverage, false-certificate frequency, and the greedy information-gain ratio $g_T^{\mathrm{gr}}(\mathcal{C}_{O,h}; \kappa)/g_T^{\mathrm{gr}}(\mathbb{A}_h; \kappa)$, where $\mathcal{C}_{O,h}$ is the $\epsilon = 0$ observationally surviving grid used by Theorem 3. Here g_T^{gr} is the log determinant attained by greedy query selection, a numerical proxy for the maximum γ_T (Appendix D.6). Comparisons include the IDP-only ablation, POMIS+GP-UCB and full-space GP-UCB (GP upper-confidence-bound methods expressed as lower confidence bounds for minimization), and, on discrete problems, the ZB17-style bound-UCB (Zhang & Bareinboim, 2017) and UCB-SI-style (Sharma et al., 2020) baselines. Complete curves and additional oracle, arbitrary-MAG, shared-surrogate, cost-aware, and DVI comparisons appear in the appendix.

Partial identification can resolve the decision before intervention. Figures 1 and 2(a) isolate the contribution of partial identification in C1. Its PAG-POMIS grid has six actions: one is excluded by point identification and two additional actions are excluded only by PI bounds, leaving half the grid as initial contenders. PAG-CBO certifies at $t = 0$ in all 20 optimization seeds, whereas the IDP-only ablation—which replaces positive-width PI intervals by the $[0, 1]$ fallback—does not certify at

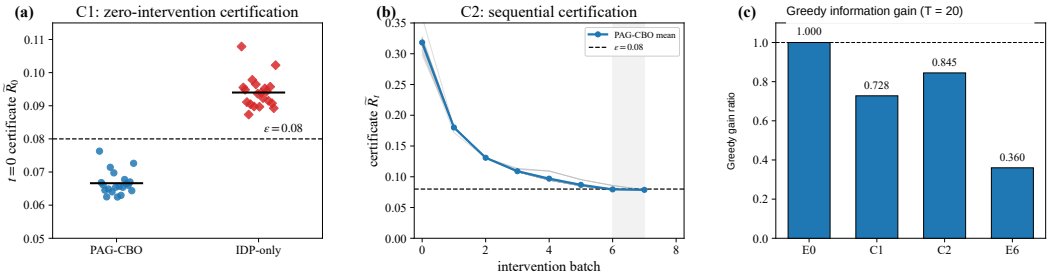

Figure 2: **Main experimental results.** **(a)** C1 certificate gap at $t = 0$ across seeds for PAG-CBO and the IDP-only ablation; the dashed line is $\epsilon = 0.08$. **(b)** C2 certificate gap versus 250-outcome intervention batch, showing certification after six or seven batches. **(c)** Greedy information-gain localization at $T = 20$ using the $\epsilon = 0$ surviving set: $g_T^{\mathrm{gr}}(\mathcal{C}_{O,h})/g_T^{\mathrm{gr}}(\mathbb{A}_h)$ equals 1.000 (E0), 0.728 (C1), 0.845 (C2), and 0.360 (E6). These attained greedy values are not exact maximum-information-gain values. The theorem uses the exact maxima over the same domains at horizons T and $T-1$.

$t = 0$. Thus point identification alone is insufficient for the immediate certificate; informative non-singleton PAG bounds provide the missing decision information. Bound-aware finite-arm methods supplied with the same valid intervals can also exploit this information, so the result isolates the value of the PAG response bounds, rather than claiming a novel warm-start principle. Figure 2(b) shows C2, where overlapping initial bands prevent immediate certification. With the fixed batch size of 250 outcomes per intervention, all 20 PAG-CBO runs certify after six or seven batches (1500–1750 outcomes), and all finish with zero regret relative to the benchmark reference optimum.

Calibration and certificate validity. Figure 12(a) reports observational-band coverage, and Figure 12(b) reports certificate claims and errors. The coverage targets are point-identified responses, exact canonical-SCM LP intervals for the C1/C2 joint-treatment actions, and $[0, 1]$ for fallback actions. These LP targets are sharp within the stated canonical class; equality with the full PAG identified set is not assumed. Across 36 cells with 200 replicates each, the prespecified Wilson criterion finds no cell statistically incompatible with its nominal $1 - \delta_O$ coverage target. This cellwise diagnostic does not prove nominal coverage. Pooling stopping claims at the benchmark-specific tolerances yields 4,835 claims and no false recommendations. The observational bands use approximate inference, and the experimental intervention radius differs from the sufficient radius in Theorem 2. These results therefore provide empirical calibration evidence, not a verification of every premise of that theorem; Appendix D.2 states the differences precisely.

Localization, fallback behavior, and limitations. Figure 2(c) illustrates a reduction in the greedy information-gain proxy, consistent with the localization mechanism in Theorem 3: informative observational bounds reduce the decision-relevant domain, whereas E0, where most nonidentified continuous responses use the valid $[0, 1]$ fallback, obtains essentially no reduction. The continuous E0–E4 benchmarks similarly retain nearly the full contender domain when no validated informative continuous PAG mean bound is available. The numerical regret diagnostics exceed the $[0, 1]$ response range in several cases and do not evaluate all constants and confidence conditions of Theorem 3. They therefore assess localization rather than the numerical tightness of its bound. Likewise, DVI remains a model-based heuristic: on the five C2 DVI seeds it requires 15–37 of the same 250-outcome batches, compared with 6–7 for the default LCB rule. Finally, deliberately misspecified PAGs can invalidate certification: the endpoint-perturbation study contains two C2 certificate violations, while the correct/oracle-PAG cases contain none (Appendix D).

Conclusion and Future Work

PAG-CBO uses bounds on intervention-response means to connect causal identification with sequential optimization. Under a correct PAG and valid simultaneous bands, it safely prunes actions and certifies ϵ-optimal recommendations. Informative PI bounds can resolve a decision before intervention; otherwise, they can restrict the region that must be explored. The experiments illustrate these mechanisms and their limitations: weak bounds offer little pruning, approximate bands require calibration, and a misspecified PAG can invalidate certification. Future work will develop informative continuous-action bounds, account for uncertainty in learned PAGs, and improve acquisition efficiency while retaining valid certification.

AI USE STATEMENT

We used generative AI tools to assist with generating synthetic datasets, implementing methods, and developing and writing proofs of mathematical claims. We also used these tools to improve the clarity, organization, and consistency of the manuscript and to assist with preparing the conceptual illustration. The authors reviewed and revised the AI-assisted material, checked the mathematical arguments against the stated assumptions, and examined the synthetic-data generation procedures for consistency with the benchmark definitions. We take responsibility for the final content of this work, including all text, mathematical claims, experimental results, and artifacts produced with generative AI assistance.

ETHICS STATEMENT

This work studies causal Bayesian optimization using computational experiments with structural causal models. The reported experiments do not involve interventions on human participants; the healthcare and epidemiology benchmarks serve as simulation studies rather than clinical evaluations. A potential risk in applying PAG-CBO is treating an optimization certificate as an unconditional guarantee of safety. Its guarantees depend on the stated causal assumptions, valid confidence bands, and the action domain over which optimization is performed. As the misspecification experiments demonstrate, an incorrect PAG can invalidate certification. Moreover, ϵ-optimality for the chosen objective does not establish fairness, clinical benefit, or the absence of harmful side effects. Applications affecting people therefore require domain-specific validation, appropriate oversight, and explicit consideration of outcomes and constraints beyond the optimization objective.

REPRODUCIBILITY STATEMENT

Sections 2 and 3 specify the problem, assumptions, response bounds, optimization procedure, and certification rule. Appendix B provides proofs of the theoretical results. Section 4 and Appendix D.2 describe the benchmark suite, experimental protocol, seed counts, and evaluation metrics. Appendices D.3, D.4, and D.8 report causal-bound audits, confidence-band calibration, and PAG-misspecification tests, respectively. Additional comparisons and sensitivity analyses appear in Appendix D. The source code is available anonymously at `https://anonymous.4open.science/r/pagcbo-794D`.

REFERENCES

Virginia Aglietti, Xiaoyu Lu, Andrei Paleyes, and Javier González. Causal bayesian optimization. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 3155–3164. PMLR, 2020.

Virginia Aglietti, Neil Dhir, Javier González, and Theodoros Damoulas. Dynamic causal bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 34, pp. 10549–10560, 2021.

Virginia Aglietti, Alan Malek, Ira Ktena, and Silvia Chiappa. Constrained causal bayesian optimization. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 304–321. PMLR, 2023.

Andreas Andersson and Nicholas R. Bates. In situ measurements used for coral and reef-scale calcification structural equation modeling including environmental and chemical measurements, and coral calcification rates in bermuda from 2010 to 2012 (BEACON project). Biological and Chemical Oceanography Data Management Office, dataset 720788, 2018. URL `https://www.bco-dmo.org/dataset/720788`. Dataset dated March 2, 2018.

Alexis Bellot. Towards bounding causal effects under Markov equivalence. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pp. 308–332. PMLR, 2024.

Shriya Bhatija, Paul-David Zuercher, Jakob Thumm, and Thomas Bohné. Multi-objective causal bayesian optimization. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 4172–4195. PMLR, 2025.

Nicola Branchini, Virginia Aglietti, Neil Dhir, and Theodoros Damoulas. Causal entropy optimization. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 8586–8605. PMLR, 2023.

Victor Chernozhukov, Han Hong, and Elie Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284, 2007. doi: 10.1111/j.1468-0262.2007.00794.x.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70 of *Proceedings of Machine Learning Research*, pp. 844–853. PMLR, 2017.

Travis A. Courtney, Mario Lebrato, Nicholas R. Bates, Andrew Collins, Samantha J. de Putron, Rebecca Garley, Rodney Johnson, Juan-Carlos Molinero, Timothy J. Noyes, Christopher L. Sabine, and Andreas J. Andersson. Environmental controls on modern scleractinian coral and reef-scale calcification. *Science Advances*, 3(11):e1701356, 2017. doi: 10.1126/sciadv.1701356.

Jean Durand, Yashas Annadani, Stefan Bauer, and Sonali Parbhoo. Causal bayesian optimization with unknown graphs. *arXiv preprint arXiv:2503.19554*, 2025.

Muhammad Qasim Elahi, Mahsa Ghasemi, and Murat Kocaoglu. Partial structure discovery is sufficient for no-regret learning in causal bandits. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Ana Ferro, Francisco Pina, Milton Severo, Pedro Dias, Francisco Botelho, and Nuno Lunet. Use of statins and serum levels of prostate specific antigen. *Acta Urológica Portuguesa*, 32(2):71–77, 2015. doi: 10.1016/j.acup.2015.02.002.

Peter I. Frazier, Warren B. Powell, and Savas Dayanik. A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439, 2008. doi: 10.1137/070693424.

Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.

Limor Gultchin, Virginia Aglietti, Alexis Bellot, and Silvia Chiappa. Functional causal bayesian optimization. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pp. 756–765. PMLR, 2023.

Guido W. Imbens and Charles F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004. doi: 10.1111/j.1468-0262.2004.00555.x.

Amin Jaber, Jiji Zhang, and Elias Bareinboim. Causal identification under markov equivalence. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 978–987, 2018.

Amin Jaber, Jiji Zhang, and Elias Bareinboim. Causal identification under Markov equivalence: Completeness results. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2981–2989. PMLR, 2019.

Amin Jaber, Adele Ribeiro, Jiji Zhang, and Elias Bareinboim. Causal identification under Markov equivalence: Calculus, algorithm, and completeness. In *Advances in Neural Information Processing Systems*, volume 35, pp. 3679–3690, 2022.

Mohammad Ali Javidian. Transferring information across interventions in causal bayesian optimization. *arXiv preprint arXiv:2606.01457*, 2026.

Markus Kalisch, Martin Mächler, Diego Colombo, Marloes H. Maathuis, and Peter Bühlmann. Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software*, 47(11):1–26, 2012. doi: 10.18637/jss.v047.i11.

Mikhail Konobeev, Jalal Etesami, and Negar Kiyavash. Causal bandits without graph learning. In *Proceedings of the Fourth Conference on Causal Learning and Reasoning*, volume 275 of *Proceedings of Machine Learning Research*, pp. 31–63. PMLR, 2025.

Sanghack Lee and Elias Bareinboim. Structural causal bandits: Where to intervene? In *Advances in Neural Information Processing Systems*, volume 31, pp. 2573–2583, 2018.

Alan Malek, Virginia Aglietti, and Silvia Chiappa. Additive causal bandits with unknown graph. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 23574–23589. PMLR, 2023.

Charles F. Manski. Nonparametric bounds on treatment effects. *American Economic Review*, 80(2): 319–323, 1990.

Charles F. Manski. Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice. *Journal of Econometrics*, 95(2): 415–442, 2000.

Sumantrak Mukherjee, Mengyan Zhang, Seth Flaxman, and Sebastian Josef Vollmer. Graph agnostic causal bayesian optimisation. *arXiv preprint arXiv:2411.03028*, 2024.

Min Woo Park, Andy Arditi, Elias Bareinboim, and Sanghack Lee. Structural causal bandits under markov equivalence. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

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

Mohammad Shahverdikondori, Jalal Etesami, and Negar Kiyavash. Graph learning is suboptimal in causal bandits. In *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 3088–3096. PMLR, 2026.

Nihal Sharma, Soumya Basu, Karthikeyan Shanmugam, and Sanjay Shakkottai. Warm starting bandits with side information from confounded data. *arXiv preprint arXiv:2002.08405*, 2020.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning*, 2010.

Jörg Stoye. More on confidence intervals for partially identified parameters. *Econometrica*, 77(4): 1299–1315, 2009a. doi: 10.3982/ECTA7347.

Jörg Stoye. Minimax regret treatment choice with finite samples. *Journal of Econometrics*, 151(1): 70–81, 2009b.

Scott Sussex, Anastasiia Makarova, and Andreas Krause. Model-based causal bayesian optimization. In *International Conference on Learning Representations*, 2023.

Scott Sussex, Pier Giuseppe Sessa, Anastasiia Makarova, and Andreas Krause. Adversarial causal bayesian optimization. In *International Conference on Learning Representations*, 2024.

Tian-Zuo Wang, Tian Qin, and Zhi-Hua Zhou. Estimating possible causal effects with latent variables via adjustment. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 36308–36335. PMLR, 2023.

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

Junzhe Zhang and Elias Bareinboim. Transfer learning in multi-armed bandits: A causal approach. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 1340–1346, 2017. doi: 10.24963/ijcai.2017/186.