Causal Bayesian Optimization: Foundations, Methods, and Applications

Anonymous authors
Paper under double-blind review

Abstract

Causal Bayesian Optimization (CBO) integrates causal inference with Bayesian optimization to enable sample-efficient intervention selection in systems governed by causal structure. This survey provides a comprehensive and systematic review of the CBO landscape, organizing the growing literature through a unified BO-loop perspective that reveals how causal assumptions shape four core components: intervention search spaces, surrogate construction, acquisition design, and decision policies. We classify methods along recurring design axes, i.e., graph knowledge, intervention representation, uncertainty source, and budget allocation, and establish formal connections between CBO and adjacent fields, including causal bandits, Bayesian experimental design, safe optimization, and policy search. To address the lack of standardized evaluation in the field, we introduce a reproducibility-oriented benchmark that covers hard- and soft-intervention settings, implements both the standard GAP metric and a new trajectory-aware Path-Aware GAP (PA-GAP) metric, and evaluates seven CBO methods alongside a non-causal BO baseline under a common scoring protocol. Our empirical study across thirteen datasets, three budget levels, and two metrics reveals that no single method dominates uniformly: rankings depend critically on the dataset, budget, and metric, and strong non-causal baselines remain competitive in several settings. We conclude by identifying six open challenges, including robustness to hidden confounding, scalable unknown-graph optimization, mixed intervention types, realistic cost models, tighter theoretical guarantees, and integration with modern representation learning, that must be addressed for CBO to transition from proof-of-concept demonstrations to reliable real-world deployment.

1 Introduction

Many scientific and engineering problems require the selection of interventions in systems whose variables are causally coupled. A clinician chooses a drug dosage that propagates through metabolic pathways to affect a clinical endpoint; an engineer tunes process parameters that cascade through a manufacturing pipeline; a policymaker adjusts economic levers whose effects depend on the causal structure of the economy. In each case, the key challenge is the same: the system's response to an intervention is not governed by statistical association but by a causal mechanism, and purely correlational optimization can recommend actions that appear promising in observational data, yet fail or even cause harm when deployed (Pearl, 2009; Peters et al., 2017).

Bayesian Optimization (BO) offers a basic framework for sample-efficient optimization when evaluations are expensive, noisy, or constrained (Jones et al., 1998; Shahriari et al., 2016). By iteratively fitting a probabilistic surrogate and selecting queries via an acquisition function that balances exploitation against exploration, BO can identify near-optimal solutions with far fewer evaluations than grid or random search. However, standard BO treats the objective as a black box: it is agnostic to the causal relationships among the variables it manipulates. Causal Bayesian Optimization (CBO) closes this gap by embedding structural causal knowledge into the BO loop, enabling the optimizer to reason about *which* variables to intervene on, *how* observational data relate to interventional effects, and *where* the experimental budget is best spent (Aglietti et al., 2020).

Since its introduction, CBO has rapidly expanded into a diverse family of methods. These developments can be organized around a small number of recurring design choices: (i) what is assumed known about the causal graph, (ii) whether the surrogate models the intervention-to-outcome mapping directly or instead models individual causal mechanisms and propagates uncertainty through the graph, (iii) what form interventions take (hard clamps, soft mechanism modifications or context-dependent policies), and (iv) how the method allocates its budget across competing sources of uncertainty. Representative extensions include dynamic environments (Aglietti et al., 2021), safety constraints (Aglietti et al., 2023), mechanism-level surrogates with regret guaranties (Sussex et al., 2023), functional and policy interventions (Gultchin et al., 2023), contextual scope selection (Arsenyan et al., 2026), adversarial non-stationarity (Sussex et al., 2024), high-dimensional scaling (Wu et al., 2024), unknown-graph optimization via information-theoretic exploration (Branchini et al., 2023) or optimistic model selection (Mukherjee et al., 2024), and noise-robust acquisition with learned priors (Li et al., 2023).

Despite this progress, the field lacks two critical ingredients for maturation. First, a **unified conceptual framework** that reveals the shared structure across CBO variants and connects them to adjacent fields, such as causal bandits, Bayesian experimental design, active causal discovery, and safe reinforcement learning, is missing. Individual papers typically position their contributions relative to the original CBO formulation, but rarely articulate how different notions of uncertainty (effect, mechanism, and graph uncertainty) lead to fundamentally different acquisition principles, or how contextual and functional intervention methods are instances of a common policy-search abstraction. Second, **standardized evaluation infrastructure** is absent. Existing papers use different datasets, incompatible codebases, inconsistent intervention conventions, and different metrics, making cross-paper comparisons unreliable. A method that excels when the graph is known and interventions are low-dimensional may behave very differently under soft interventions, unknown graphs, or context-dependent policies, yet these distinctions are obscured when each paper evaluates in isolation.

This survey addresses both gaps. We provide a systematic, technically detailed review of the CBO landscape organized through a unified BO-loop perspective, and we introduce a reproducibility-oriented benchmark with standardized evaluation. In terms of concreteness, our contributions are as follows.

- **A unified design-space perspective.** We decompose CBO methods into four interacting design axes, including intervention representation, surrogate architecture, decision rule, and dominant uncertainty source, and show how each published method instantiates specific choices along these axes (Section 3). This reveals which components are interchangeable and which are tightly coupled to specific identifiability or structural assumptions.
- **Formal connections to adjacent fields.** We establish explicit links between CBO and causal bandits, Bayesian experimental design, active causal discovery, safe optimization, and policy search, clarifying what CBO inherits from these fields and where it introduces genuinely new challenges (Section 3.7).
- **A reproducibility-oriented benchmark.** We release a benchmark covering eight hard-intervention and five soft-intervention scenarios, with standardized GAP and a new trajectory-aware PA-GAP metric, and a common best-so-far trajectory interface that enables fair re-scoring of outputs from heterogeneous implementations (Section 4).
- **A unified empirical comparison.** We evaluate seven CBO methods and a non-causal BO baseline across thirteen datasets, three budget levels, and two metrics, revealing that no method dominates uniformly, that rankings depend critically on dataset–budget-metric interactions, and that strong non-causal baselines remain competitive in several settings (Section 4.4–4.5).
- **Structured open problems.** We identify six concrete research directions (robustness to causal assumptions, scalability, richer intervention models, realistic evaluation, theoretical foundations, and integration with representation learning) that must be addressed for CBO to achieve reliable real-world deployment (Section 5).

Paper outline. Section 1 places this survey relative to existing reviews. Section 2 establishes the theoretical foundations in causal inference and Bayesian optimization. Section 3 presents the methodological

frameworks organized by design axes. Section 4 describes the benchmark, metrics, datasets, and empirical results. Section 5 discusses open problems, and Section 6 concludes.

Related Surveys and Scope

Several surveys cover topics adjacent to CBO, but none provides the unified treatment of causal optimization methods and standardized empirical evaluation that we pursue here.

Bayesian optimization surveys. Comprehensive reviews of BO cover surrogate modeling, acquisition functions, and scalability (Shahriari et al., 2016; Frazier, 2018; Garnett, 2023), but treat the objective as a black box and do not address causal structure, intervention design, or the observation-intervention trade-off central to CBO.

Causal inference and causal discovery. Textbooks and surveys on causal inference (Pearl, 2009; Peters et al., 2017; Hernán & Robins, 2020; Glymour et al., 2019) provide the framework of the structural causal model that underlies CBO, but focus on the identification and estimation of causal effects rather than on sequential optimization of interventions under budget constraints. Causal discovery surveys (Spirtes et al., 2000; Chickering, 2002; Glymour et al., 2019) address the learning of the graph from data, which is a component of unknown-graph CBO but not its main goal.

Causal bandits. The causal bandit literature (Lattimore et al., 2016; Lu et al., 2021; Nair et al., 2021) shares CBO's goal of selecting interventions using causal structure, but operates under different modeling assumptions: bandits typically assume discrete action sets with parametric reward models, whereas CBO handles continuous intervention domains with nonparametric (GP-based) surrogates. We discuss the relationship between CBO and causal bandits in detail in Section 3.7.

Bayesian experimental design. The Bayesian optimal experimental design (Chaloner & Verdinelli, 1995; Foster et al., 2021) selects experiments to maximize the information about the model parameters, which overlaps with the structure-learning component of the CBO's in the context of unknown-graphs. However, experimental design typically does not include an optimization objective beyond parameter estimation.

Scope of this survey. We focus on methods that explicitly combine (i) a structural causal model or causal graph with (ii) a sequential Bayesian optimization loop to select interventions. We exclude pure causal effect estimation without an optimization loop, causal bandits without GP-based surrogates (except when contrasting with CBO), and reinforcement learning methods that learn policies through environment interaction without an explicit causal model. We include both known- and unknown-graph settings, hard and soft interventions, and single- and multi-objective formulations.

Survey protocol and inclusion criteria. To make the construction of the review corpus explicit, we followed a targeted search-and-screening protocol. We searched OpenReview, PMLR, arXiv, Google Scholar, and the proceedings of major machine learning and causality venues using keyword combinations such as "causal Bayesian optimization," "causal global optimization," "Bayesian optimization with interventions," "causal optimization," and "optimization under causal structure." Candidate papers were first screened by title and abstract, and then by their methodological formulation. A work was included if its central contribution couples an explicit causal model, such as an SCM, causal graph, or distribution over graphs, with a sequential Bayesian-optimization-style procedure for intervention selection. We further used backward and forward citation tracing from the original CBO paper and subsequent representative extensions to identify closely related work. Papers whose primary objective is only causal effect estimation, graph recovery, bandit learning, experimental design, or reinforcement learning were excluded from the main corpus unless they directly clarify the boundary between CBO and adjacent fields. This protocol yields a focused corpus of methods in which causal assumptions actively modify the Bayesian optimization loop, rather than a broad survey of causal decision-making methods.

Table 1: Summary of notation used throughout this survey.

Symbol	Meaning
$M = \langle U, V, F, P(U) \rangle$	Structural Causal Model
$V = \{V_1, \ldots, V_n\}$	Endogenous (observed) variables
U	Exogenous (unmodeled) variables
Pa_i	Parents of V_i in the causal graph $\mathcal{G}$
$\mathcal{G}$	Causal directed acyclic graph (DAG)
$do(X{=}\mathbf{x})$	Hard intervention setting X to value $\mathbf{x}$
X, C, Y	Manipulable variables, context variables, target outcome
$X_s \subseteq X$	Intervention scope (subset of manipulable variables)
$\mathbf{x}_s \in \mathcal{D}(X_s)$	Intervention value in the feasible domain
$f(X_s, \mathbf{x}_s)$	Interventional objective: $\mathbb{E}[Y \mid do(X_s{=}\mathbf{x}_s), C]$
$\mathcal{GP}(m, k)$	Gaussian process with mean m and kernel k
$\mu_t(\cdot),\ \sigma_t^2(\cdot)$	GP posterior mean and variance after t observations
$\alpha(\cdot)$	Acquisition function
T	Total number of interventional trials (budget)
y^*	Reference optimum (for metric computation)
R_t	Best-so-far improvement ratio at trial t

2 Theoretical Foundations

This section establishes the conceptual and mathematical foundations underlying Causal Bayesian Optimization. We first review structural causal models, interventions, and identifiability (Section 2.1). We then summarize classical Bayesian optimization with an emphasis on components that CBO modifies (Section 2.2). Finally, we formulate the CBO problem and highlight the key ways in which causality reshapes the BO loop (Section 2.3). Table 1 consolidates the notation used throughout the paper.

2.1 Causal Inference

Most CBO methods adopt the framework of Structural Causal Models (SCMs) (Pearl, 2009), which provide a compact, modular representation of how variables are generated and how they respond to interventions.

Definition 1 (Structural Causal Model (SCM))**.** *An SCM is a tuple $M = \langle U, V, F, P(U) \rangle$ where U is a set of exogenous variables, $V = \{V_1, \ldots, V_n\}$ is a set of endogenous variables, and $F = \{f_1, \ldots, f_n\}$ is a collection of structural assignments*

$$V_i := f_i(\mathrm{Pa}_i, U_i), \qquad i = 1, \ldots, n, \tag{1}$$

where $\mathrm{Pa}_i \subseteq V \setminus \{V_i\}$ denotes the direct causes (parents) of V_i in the causal graph and $U_i \subseteq U$ is the exogenous noise that affects V_i. The distribution $P(U)$ completes the model and, together with F, induces a unique joint distribution on V.

Causal graphs, assumptions, and the causal Markov property. An SCM induces a causal graph $\mathcal{G}$, commonly represented as a directed acyclic graph (DAG), with a node for each endogenous variable and an edge $V_j \to V_i$ whenever $V_j \in \mathrm{Pa}_i$. The graph encodes a qualitative causal structure in which variables are direct causes of others but does not, by itself, specify functional forms or effect magnitudes. That quantitative information resides in the structural equations F.

Two standard assumptions link the graph to the distribution. The *causal Markov condition* states that each variable is conditionally independent of its non-descendants given its parents (Pearl, 2009; Spirtes et al., 2000). Under the additional assumption that exogenous variables are mutually independent (*causal sufficiency*), the observational distribution factorizes according to the graph.

$$P(V_1, \ldots, V_n) \;=\; \prod_{i=1}^{n} P(V_i \mid \mathrm{Pa}_i). \tag{2}$$

The *faithfulness* assumption further requires that all conditional independencies in P are entailed by the Markov condition applied to $\mathcal{G}$, ruling out cancelations that could hide edges. Together, these assumptions enable structure learning from observational data and underpin the identifiability results that CBO exploits.

Hard interventions. In Pearl's framework, a *hard* (or *perfect*) intervention $do(X{=}\mathbf{x})$ replaces the structural equation for X with the constant $\mathbf{x}$, severing all incoming edges to X while leaving other mechanisms unchanged. This produces a modified SCM $M_{do(X=\mathbf{x})}$ and an interventional distribution that admits a *truncated factorization*:

$$P(V_1, \ldots, V_n \mid do(X{=}\mathbf{x})) \;=\; \prod_{V_i \notin X} P(V_i \mid \mathrm{Pa}_i) \,\Big|_{X=\mathbf{x}}. \tag{3}$$

Hard interventions are the default in the original CBO formulation and in most Hard intervention benchmarks.

Soft interventions. A *soft* (or *imperfect*) intervention modifies the structural mechanism of a variable without completely severing its incoming edges. Formally, a soft intervention on V_i replaces f_i with a new mechanism $\tilde{f}_i(\mathrm{Pa}_i, U_i; \theta)$ parameterized by an intervention parameter θ. This is important in several CBO variants, notably MCBO (Sussex et al., 2023) and fCBO (Gultchin et al., 2023), that model mechanism modifications rather than variable clamping. Soft interventions yield a richer action space but require modeling how the intervention parameter interacts with the existing mechanism.

Counterfactuals. SCMs also define counterfactual quantities that describe outcomes under hypothetical actions in a specific latent world (indexed by $u \in U$), typically written as $Y_{\mathbf{x}}(u)$ (Pearl, 2009). Although most CBO methods optimize interventional expectations, counterfactual reasoning becomes relevant when personalizing interventions or reasoning about alternative actions given historical data.

Identifiability and the *do*-calculus. In practice, the full SCM is rarely known, so CBO relies on *identifiability*: whether an interventional quantity such as $\mathbb{E}[Y \mid do(X = \mathbf{x})]$ can be expressed purely in terms of the observational distribution given assumptions about the graph. Pearl's *do*-calculus provides three general inference rules for transforming interventional expressions into observational quantities when identification is possible (Pearl, 2009; 2012). A common special case is the *backdoor adjustment*: if a set Z satisfies the backdoor criterion relative to (X, Y), then

$$\mathbb{E}[Y \mid do(X{=}\mathbf{x})] \;=\; \sum_{z} \mathbb{E}[Y \mid X{=}\mathbf{x}, Z{=}z]\, P(Z{=}z). \tag{4}$$

When identification holds but closed-form evaluation is difficult, one can approximate the required expectations via Monte Carlo integration of the identified estimand.

Identifiability as an algorithmic assumption in CBO. In CBO, identifiability is not merely a causal inference condition; it directly shapes the optimizer's behavior. When an intervention effect is identifiable from observational data, the observational sample can be used to construct informative GP priors, initialize surrogates, or eliminate dominated intervention scopes *before* any costly intervention is performed. When identifiability fails, for example, due to hidden confounding, the same procedure can produce overconfident and biased priors that mislead the acquisition function. This asymmetry means that empirical comparisons should carefully distinguish methods that assume correct identifiable priors from those that rely primarily on interventional data or explicitly model graph and identification uncertainty. Throughout this survey, we follow the common CBO assumption of causal sufficiency and perfect interventions unless otherwise stated.

2.2 Classical Bayesian Optimization

Bayesian Optimization (BO) is a sequential strategy for optimizing an expensive black-box objective $f : \mathcal{X} \to \mathbb{R}$ on a bounded domain $\mathcal{X} \subset \mathbb{R}^d$ (Kushner, 1964; Zhilinskas, 1975; Mockus et al., 1978; Mockus, 1989; Jones et al., 1998; Shahriari et al., 2016). At each iteration t, BO selects a query $\mathbf{x}_t$, observes $y_t = f(\mathbf{x}_t) + \epsilon_t$ with noise ϵ_t, and updates a probabilistic model to guide future queries. BO is characterized by two interacting components.

Probabilistic surrogate. The most common surrogate is a Gaussian Process (GP) (Rasmussen & Williams, 2006), specified by a prior mean function $m_0(\cdot)$ and a kernel (covariance function) $k(\cdot, \cdot)$. Given t observations $\mathcal{D}_t = \{(\mathbf{x}_i, y_i)\}_{i=1}^t$, the GP posterior of any candidate input $\mathbf{x}$ is available in closed form:

$$\mu_t(\mathbf{x}) = m_0(\mathbf{x}) + \mathbf{k}_t(\mathbf{x})^\top (\mathbf{K}_t + \sigma_\epsilon^2 \mathbf{I})^{-1} (\mathbf{y}_t - \mathbf{m}_0), \tag{5}$$

$$\sigma_t^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}_t(\mathbf{x})^\top (\mathbf{K}_t + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{k}_t(\mathbf{x}), \tag{6}$$

where $\mathbf{k}_t(\mathbf{x})$ is the vector of covariances between $\mathbf{x}$ and the observed inputs, $\mathbf{K}_t$ is the kernel matrix over the observed inputs, and σ_ϵ^2 is the variance of the observation noise. The posterior mean μ_t provides a point estimate, while σ_t^2 quantifies the predictive uncertainty, enabling the acquisition function to distinguish well-explored from poorly-understood regions.

Acquisition functions. The acquisition function $\alpha(\mathbf{x}; \mathcal{D}_t)$ scores candidate inputs by balancing exploitation (questioning where μ_t is favorable) against exploration (questioning where σ_t is large). Two widely used choices are Expected Improvement (EI) and the Upper Confidence Bound (UCB):

$$\alpha_{\mathrm{EI}}(\mathbf{x}) = \mathbb{E}\big[\max(f(\mathbf{x}) - f^+, 0) \mid \mathcal{D}_t\big], \tag{7}$$

$$\alpha_{\mathrm{UCB}}(\mathbf{x}) = \mu_t(\mathbf{x}) + \beta_t^{1/2}\, \sigma_t(\mathbf{x}), \tag{8}$$

where f^+ is the best observed value and $\beta_t > 0$ is a schedule parameter that controls the exploration–exploitation trade-off (Jones et al., 1998; Srinivas et al., 2010). The BO loop iterates: maximize α to select $\mathbf{x}_{t+1}$, evaluate f, and update the GP posterior.

Regret and sample efficiency. Under regularity conditions in the kernel and the appropriate choice of β_t, GP-UCB achieves a sublinear cumulative regret $\sum_{t=1}^T [f(\mathbf{x}^*) - f(\mathbf{x}_t)] = \mathcal{O}(\sqrt{T\gamma_T})$, where γ_T is the maximum information gain of the kernel (Srinivas et al., 2010; Chowdhury & Gopalan, 2017). This theoretical grounding is relevant to CBO because MCBO (Sussex et al., 2023) and ACBO (Sussex et al., 2024) extend GP-UCB-style regret analysis to causal settings, obtaining bounds that can improve over the non-causal case when the causal graph induces a factored reward structure.

Beyond standard GPs. Although GPs remain the dominant surrogate in CBO, the broader literature on BO also uses random forests, Bayesian neural networks, and deep kernel learning (Shahriari et al., 2016; Garnett, 2023). These alternatives are relevant when mechanism functions are high-dimensional or exhibit non-stationary structure, though their integration with causal priors remains largely unexplored.

2.3 Causal Bayesian Optimization

Causal Bayesian Optimization (CBO) extends standard BO to settings where the decision variable includes not only the intervention level but also the intervention scope, which variables to intervene on, and where observational data can inform the optimizer through causal identification. This coupling of combinatorial scope selection with continuous value optimization, guided by a causal graph, is the defining feature that distinguishes CBO from black-box BO.

CBO introduces three innovations relative to standard BO. First, *graph-based scope reduction* uses the causal graph to prune intervention scopes that are structurally redundant or observationally dominated, concentrating the budget on scopes that the graph predicts can meaningfully move the outcome. Second, *observationally informed priors* convert abundant observational data into prior beliefs about interventional effects when identifiability holds, giving the surrogate a reasonable global shape before costly interventions begin. Third, CBO explicitly models an *observation–intervention trade-off*, allowing the algorithm to decide whether additional observational or interventional data is more valuable at each step.

Section 3.1 formalizes the CBO problem, defines the key concepts of minimal and possibly-optimal intervention sets, presents the surrogate and acquisition design, and identifies the limitations that motivate the extensions reviewed in the remainder of Section 3.

3 Methodological Frameworks

Causal Bayesian Optimization (CBO) is best viewed as a design space rather than a single algorithm. All CBO methods instantiate a common loop: choose an intervention (what to manipulate and how), predict its causal effect using a probabilistic model, select the next data-collection action via an acquisition rule, and update beliefs using newly obtained observational and/or interventional data. What differentiates methods is where the causal structure is injected into this loop and which uncertainty sources are modeled.

A unifying lens: four design axes. In the literature, most variants can be understood through the following axes (Figure 1):

1. **Representation of the intervention.** Hard interventions vs. soft interventions; static actions vs. functional policies; context-dependent actions; or adversarial interventions.
2. **Surrogate modeling choice.** Modeling the intervention-to-outcome effect directly versus modeling mechanisms on the causal graph and propagating uncertainty to the outcome.
3. **Decision rule and budget allocation.** EI-/UCB-style BO, information-gain search, bandit-style allocation across scopes, or online-learning strategies in non-stationary settings.
4. **What uncertainty dominates.** Effect uncertainty, structure uncertainty, and robustness uncertainty.

We use these axes to highlight what each method adds, when it helps, and what it costs in terms of computational or statistical resources.

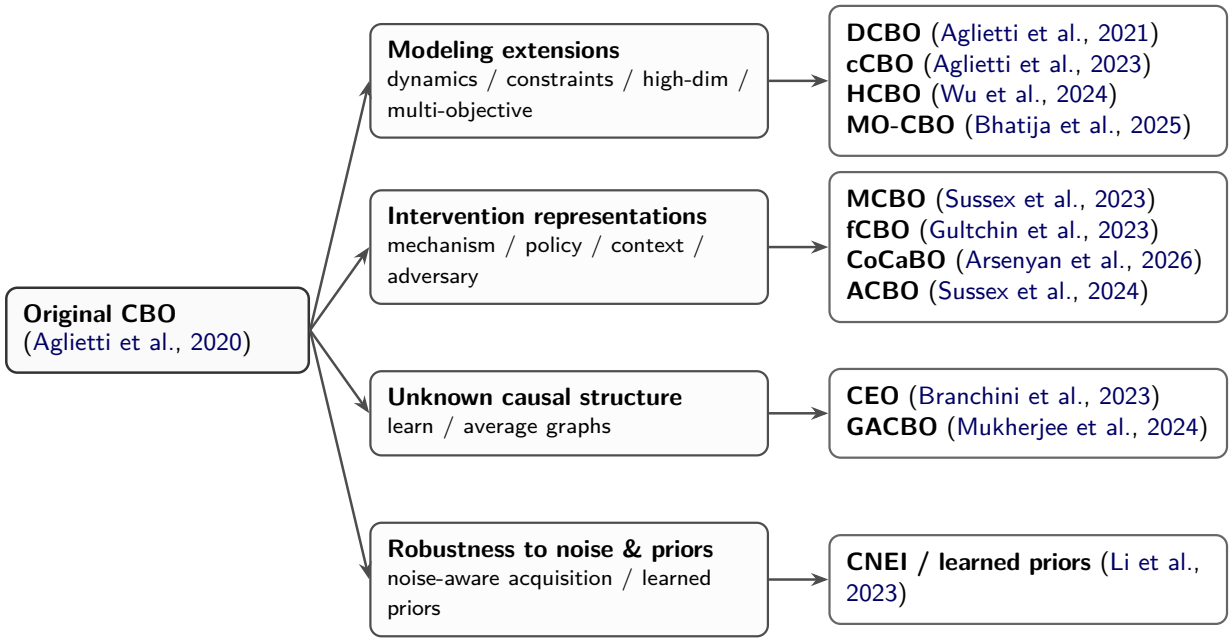

Figure 1: Causal Bayesian Optimization development map organized by the primary bottleneck addressed.

How we summarize methods. For each approach, we emphasize: (i) assumptions (graph knowledge, intervention type, observed variables); (ii) the main algorithmic lever (search-space reduction, surrogate design, acquisition/decision policy); and (iii) failure modes and scaling constraints. This avoids re-stating paper narratives verbatim and instead clarifies how the methods relate at the level of design choices in the CBO loop.

3.1 Causal Bayesian Optimization: Core Framework

The original CBO framework (Aglietti et al., 2020) formalizes causal optimization as a search over interventions in a known causal graph. Let X denote manipulable variables, C non-manipulable context variables, and Y the target outcome. An action consists of an intervention scope $X_s \subseteq X$ and an assignment $\mathbf{x}_s$ in the feasible intervention domain $\mathcal{D}(X_s)$.

Definition 2 (CBO objective and action space)**.** *A CBO action is a pair $(X_s, \mathbf{x}_s)$ where $X_s \subseteq X$ is an intervention scope and $\mathbf{x}_s \in \mathcal{D}(X_s)$ is an assignment in the feasible intervention domain. The interventional objective is*

$$f(X_s, \mathbf{x}_s) \;=\; \mathbb{E}[\, Y \mid do(X_s = \mathbf{x}_s),\, C\,]\,. \tag{9}$$

The global causal optimization problem is

$$(X_s^\star, \mathbf{x}_s^\star) \in \underset{X_s \subseteq X,\, \mathbf{x}_s \in \mathcal{D}(X_s)}{\arg\min} f(X_s, \mathbf{x}_s), \tag{10}$$

with $\arg\min$ *replaced by* $\arg\max$ *for maximization tasks.*

The key point is that the optimizer is free to choose the scope of the intervention, not only the level. This yields a mixed discrete–continuous search problem and motivates two core ideas in CBO: graph-based pruning of candidate scopes and observationally informed priors over interventional effects.

Why causality changes the BO problem. A naive approach would flatten $(X_s, \mathbf{x}_s)$ into a single input vector and run the standard BO on the interventional objective. This ignores the structural information encoded in the causal graph. Causal knowledge enables two advantages that standard BO cannot exploit: (i) *scope reduction*: identifying which variables need to be manipulated at all, thereby shrinking the combinatorial search; and (ii) *informative priors*: converting plentiful observational data into prior beliefs about interventional effects, which is possible only when identifiability holds (Section 2.1).

Scope reduction: minimal intervention sets. A central insight of CBO is that the causal graph can prune intervention scopes that are structurally redundant. Two key notions formalize this:

Definition 3 (Minimal Intervention Set (MIS))**.** *An intervention scope $X_s \subseteq X$ is a* minimal intervention set *for the target Y if (i) intervention in X_s can affect the distribution of Y, and (ii) no proper subset of X_s achieves the same interventional coverage of causal pathways to Y.*

Definition 4 (Possibly-Optimal MIS (POMIS))**.** *A MIS X_s is* possibly optimal *if observational evidence does not allow its optimal value to be dominated by another MIS. That is, the best achievable interventional objective under X_s is not provably worse than under another scope, given the available data.*

By restricting the search to POMIS, the CBO avoids wasting budget on scopes that are causally irrelevant (no path to Y) or observationally dominated (another scope provably achieves a better objective). This reduction is not merely computational; it changes the statistical problem by concentrating evaluations on scopes that the graph predicts can meaningfully move the outcome.

Surrogate modeling: effect-level GPs with causal priors. The CBO models the interventional objective for each candidate scope with an effect-level GP. The key innovation is the prior construction: when the interventional effect is identifiable, observational data are used to build an initial estimate of the interventional mean and its uncertainty:

$$f_s(\mathbf{x}_s) := \mathbb{E}[Y \mid do(X_s = \mathbf{x}_s), C], \tag{11}$$

$$f_s(\cdot) \sim \mathcal{GP}(m_s(\cdot),\, k_s(\cdot,\cdot))\,, \tag{12}$$

$$m_s(\mathbf{x}_s) \approx \widehat{\mathbb{E}}[Y \mid do(X_s = \mathbf{x}_s), C]\,, \tag{13}$$

$$k_s(\mathbf{x}_s, \mathbf{x}_s') = k_{\mathrm{RBF}}(\mathbf{x}_s, \mathbf{x}_s') + \widehat{\sigma}_s(\mathbf{x}_s)\,\widehat{\sigma}_s(\mathbf{x}_s'), \tag{14}$$

where $k_{\mathrm{RBF}}(\mathbf{x}, \mathbf{x}') = \exp\big(-\|\mathbf{x} - \mathbf{x}'\|^2/(2\ell^2)\big)$ and $\widehat{\sigma}_s(\mathbf{x}_s)$ is the uncertainty in the observationally-derived effect estimate. The observational data thus serve as prior information, not as additional labels, giving the surrogate a reasonable global shape before costly interventions begin.

Algorithm 1 Generic Causal Bayesian Optimization Loop

Require: Causal graph $\mathcal{G}$ (or prior over graphs), observational data $\mathcal{D}_{\text{obs}}$, budget T
Ensure: Best intervention $(X_s^\star, \mathbf{x}_s^\star)$
1: **Scope reduction** [causal]: Use $\mathcal{G}$ to identify candidate scopes (MIS/POMIS)
2: **Prior construction** [causal]: Use $\mathcal{D}_{\text{obs}}$ and identifiability to set GP prior means $\{m_s\}$
3: **for** $t = 1, \ldots, T$ **do**
4: **Surrogate update**: Update GP posteriors $\{\mu_t^{(s)}, \sigma_t^{(s)}\}$ for each scope s
5: **Acquisition** [causal]: Select $(X_s, \mathbf{x}_s) = \arg\max_{s,\mathbf{x}} \alpha_s(\mathbf{x}; \mathcal{D}_t)$ across scopes
6: **Evaluate**: Observe $y_t = f(X_s, \mathbf{x}_s) + \epsilon_t$ via intervention
7: **Update**: Add $(X_s, \mathbf{x}_s, y_t)$ to interventional dataset $\mathcal{D}_t$
8: **end for**
9: **return** $(X_s^\star, \mathbf{x}_s^\star)$ with best observed objective

Acquisition and the observation–intervention trade-off. CBO uses a cost-aware expected improvement to choose between pairs (scope, value):

$$\text{CEI}_s(\mathbf{x}_s) = \frac{\mathbb{E}\big[(y^\star - f_s(\mathbf{x}_s))_+\big]}{\text{Co}(X_s, \mathbf{x}_s)}, \tag{15}$$

where $y^\star$ is the best value observed so far and $\text{Co}(\cdot)$ is an intervention cost. Beyond the standard exploration–exploitation trade-off, CBO introduces an *observation–intervention trade-off*: the algorithm can decide whether to allocate budget to additional observational samples (cheap but potentially biased for interventional decisions) or to new interventions (expensive but causally informative). This framing is foundational: many later variants can be interpreted as changing *what is uncertain* and therefore changing *which data is worth acquiring*.

The generic CBO loop. The Algorithm 1 summarizes the generic CBO procedure, highlighting where causal knowledge enters. Steps 1–2 exploit the graph *before* the optimization loop; Steps 4–5 use a causal structure *within* each iteration. This template applies, with variations, to all methods reviewed in the remainder of this section.

Limitations that motivate extensions. The modularity of CBO's also exposes its bottlenecks: (i) reliance on a correctly known graph and identifiable observational priors; (ii) limited robustness to noise and prior misspecification; and (iii) scaling challenges due to exponentially many candidate scopes and separate surrogates per scope. The remainder of this section can be read as a sequence of responses to these bottlenecks, organized by the design axis, each method primarily addressing.

3.2 Modeling Extensions

The original CBO formulation assumes a static causal system, a single objective, and an intervention space that remains tractable after graph-based pruning. Modeling extensions relax these assumptions by changing what the optimizer must represent: time variation that renders the objective non-stationary, feasibility that restricts attention to a safe set, dimensionality that enlarges the scope of the space, or multiple competing goals that shift the target from a single optimum to a Pareto set. A useful way to read this branch of the literature is that it keeps the same BO loop based on a surrogate, an acquisition rule, and iterative updates, but introduces an additional structure that preserves sample efficiency by decomposing the problem into smaller subproblems and sharing information across them.

Time-evolving causal systems (DCBO). Dynamic Causal Bayesian Optimization (DCBO) addresses environments where the best intervention is non-stationary, typically because mechanisms drift or the system state evolves (Aglietti et al., 2021). The main contribution is not a new acquisition rule, but a transfer mechanism. Under invariance assumptions such as a fixed graph and additive-noise structure, interventional knowledge from earlier time steps can be mapped into an informative prior for the current step. This treats

time as a source of reusable evidence so that exploration can be amortized rather than restarted. The trade-off is the sensitivity of the assumption. If the graph or key mechanisms shift in ways not captured by the transfer recursion, the prior can become overconfident and bias the search, so DCBO is strongest when temporal variation is structured and not adversarial.

Constraints and safe causal optimization (cCBO). Constrained Causal Bayesian Optimization (cCBO) targets settings where interventions must satisfy safety or feasibility constraints (Aglietti et al., 2023). The key insight from the modeling is that constraints change the definition of a useful intervention scope. A scope that can move the objective but cannot maintain feasibility is effectively irrelevant; cCBO therefore extends graph-based pruning to constraint-relevant pathways through constrained minimal intervention sets, and it couples objective and constraints through multi-output surrogates so evidence about feasibility shapes where the objective surrogate should be trusted. In practice, this often acts as a regularizer that prevents spending the budget in regions that look promising under the objective surrogate but are unlikely to be deployable. The main limitation is scaling with the number of constraints and candidate scopes, although structural pruning mitigates this in graphs with sparse causal connectivity.

High-dimensional intervention spaces (HCBO). High-Dimensional Causal Bayesian Optimization (HCBO) addresses the combinatorial bottleneck of scope selection in large graphs (Wu et al., 2024). Its central move is to replace exact enumeration of minimal or possibly-optimal scopes with an efficient coverage-based approximation. Rather than asking which scopes are minimal, HCBO constructs a small candidate family of scopes that jointly cover causal influence on the outcome well enough to support optimization. This reframes high-dimensional CBO as a hierarchical problem: select a promising scope family, then run continuous BO within the chosen scopes. HCBO also introduces normalized cross-scope scoring so that surrogates trained on different scopes can be compared in a meaningful way despite scale and data imbalance. The cost of scalability is that any coverage proxy can miss narrow, high-impact pathways, so HCBO is most reliable when the system exhibits causal sparsity or a low intrinsic causal dimension.

Multiple outcomes and Pareto structure (MO-CBO). Multi-Objective Causal Bayesian Optimization (MO-CBO) generalizes CBO to settings with multiple targets, where the goal is to approximate a Pareto front rather than a single optimum (Bhatija et al., 2025). The key observation is that causal structure is especially valuable when objectives compete. Intervening on everything can obscure trade-offs by coupling variables that need not be coupled. MO-CBO uses the graph to reduce the search to a minimal collection of local multi-objective subproblems and then allocates evaluation budget across them using a cross-scope criterion based on relative hypervolume improvement. Within each local problem, standard multi-objective BO components can be used, for example, DGEMO (Konakovic Lukovic et al., 2020). The main challenge remaining is combinatorial. The number of local subproblems can still grow quickly with dense graphs or many objectives, motivating stronger cross-scope sharing and joint surrogates that exploit causal factorization across objectives.

Summary. Across DCBO, cCBO, HCBO, and MO-CBO, a recurring pattern is that causal graphs function as a decomposition tool. They identify which parts of the intervention space matter, which tasks, such as objectives, constraints, or time steps, can share statistical strength, and which comparisons must be made across heterogeneous subproblems. These extensions are therefore less about inventing new acquisition rules and more about engineering representations so that acquisition decisions remain meaningful under time variation, feasibility restrictions, high dimensionality, or multi-objective trade-offs.

3.3 Intervention Strategy Developments

Beyond modeling extensions, another major line of work changes what an intervention looks like and how decisions are made during optimization. The common motivation is that treating each intervention as a single black-box query can waste information. These methods modify the BO loop so it can reuse causal structure inside the system, express richer actions such as policies, and stay meaningful under contextual variation or non-stationarity.

Mechanism-level modeling with guarantees (MCBO). Model-based Causal Bayesian Optimization (MCBO) models the causal system at the level of structural mechanisms rather than only the intervention-to-outcome effect (Sussex et al., 2023). It places Gaussian process priors on individual structural equations and propagates uncertainty through the known causal graph to obtain uncertainty over the target outcome. This changes data reuse. Observations of intermediate variables update upstream mechanisms and therefore tighten uncertainty about downstream outcomes, which can reduce the number of costly interventions needed to identify good actions. MCBO then uses an optimism-based decision rule in the spirit of UCB and provides regret guaranties, connecting CBO to theoretical results from Gaussian process bandits. The main limitations are practical. Mechanism posteriors must be maintained for multiple equations, uncertainty propagation increases computational cost, and acquisition optimization can be challenging. MCBO also assumes a known graph and a mechanism model class that is expressive enough for the system at hand.

Policies as interventions (fCBO). Functional Causal Bayesian Optimization (fCBO) expands the action space from selecting a value to selecting a policy that maps the observed context to an action (Gultchin et al., 2023). This is motivated by domains such as personalized treatment, where a single fixed intervention is rarely optimal for all individuals. fCBO makes policy search tractable by defining a kernel over policies using RKHS tools and running Bayesian optimization in this function space with a functional expected improvement rule. The key contribution is that it unifies causal optimization and policy learning within a single sequential procedure, where the surrogate models policy-to-outcome effects and the acquisition balances exploration across qualitatively different decision rules. The main challenge is representation. Because the policy space is large, empirical performance depends strongly on the chosen parameterization and on whether the kernel captures a meaningful similarity between policies.

Context-dependent scope selection (CoCaBO). Contextual Causal Bayesian Optimization (CoCaBO) targets settings where the exogenous context affects outcomes and where the best intervention depends on the realized context (Arsenyan et al., 2026). The additional difficulty is that one must decide not only which action to take, but also which contextual variables should be used to condition decisions. CoCaBO addresses this by separating the problem into two levels. At the outer level, it allocates the budget across candidate mixed policy scopes using a bandit-style UCB rule. At the inner level, it optimizes interventions within the selected scope using a mixed-variate Bayesian optimization routine such as HEBO (Cowen-Rivers et al., 2022). This architecture makes explicit that context introduces a discrete model selection problem on top of continuous optimization. The main limitations are derived from the hierarchy. The outer allocation must spend trials to distinguish scopes, and improvements learned within one scope do not necessarily transfer to others.

Non-stationarity and adversarial environments (ACBO). Adversarial Causal Bayesian Optimization (ACBO) addresses cases where the system response changes over time or is influenced by an adversary that can also intervene in the system (Sussex et al., 2024). This shifts the objective from identifying a single best intervention to achieving low regret over a series of rounds. ACBO combines multiplicative weight update from adversarial online learning (Littlestone & Warmuth, 1994; Freund & Schapire, 1997) with causal modeling to construct calibrated optimistic reward estimates via a causal UCB oracle. A central benefit of using the causal graph is that it can reduce regret dependence on the nominal action dimension when rewards factor through causal compositions, improving over non-causal analogs such as GP-MW (Sessa et al., 2019). The main challenges are computational and statistical. The oracle can be expensive to optimize, and in non-stationary regimes, the method relies on uncertainty sets that must track drift accurately to avoid misleading optimism.

Summary. These methods broaden the CBO in two directions. They expand the intervention language, from values to mechanisms and policies, and refine uncertainty modeling, from effect uncertainty to mechanism uncertainty, scope uncertainty, and adversarial uncertainty. The resulting algorithms are often more reliable in complex environments, but they typically require stronger modeling choices and heavier computation than effect-level CBO.

3.4 Unknown causal structure

A major practical barrier for Causal Bayesian Optimization is that many applications do not provide a trusted causal graph in advance. In this regime, the optimization loop must allocate a limited intervention budget across two coupled goals: improving the outcome and reducing the ambiguity about which causal explanations remain consistent with the data. This introduces an additional uncertainty source, structure uncertainty, on top of uncertainty about causal mechanisms and observation noise. Existing approaches mainly differ in how they represent uncertainty over candidate graphs and in whether they explicitly target the decision-relevant structure, that is, the parts of the graph that materially change which intervention should be selected.

Information-seeking joint learning and optimization (CEO). Causal Entropy Optimization, CEO, couples structure learning and optimization using an explicitly information-seeking decision rule (Branchini et al., 2023). The method maintains a posterior distribution over the candidate graphs and predicts the interventional outcomes by averaging across this posterior, producing a surrogate that reflects both the mechanism uncertainty and the graph uncertainty. The next intervention is then chosen to maximize the expected information gain about the identity of the best intervention, rather than to maximize the predicted reward alone. This focus has an important practical consequence. CEO is not optimized to recover a globally accurate graph. It spends budget on experiments that resolve the structural ambiguities that matter to decide how to act so that it can focus on a decision-relevant subgraph for the target Y.

The main limitation is computational overhead. Approximating a posterior over graphs typically requires sampling in a large discrete space, and estimating information gain must be repeated across many candidate interventions. As a result, CEO is most practical when the graph is small to medium or when strong structural priors substantially narrow the set of plausible graphs.

Optimism over plausible models (GACBO). Graph-Agnostic Causal Bayesian Optimization, GACBO, follows a more algorithmic route that emphasizes scalability (Mukherjee et al., 2024). Rather than choosing interventions by entropy reduction, it maintains a confidence set of causal models that remain statistically plausible given the data and selects interventions that could be optimal under at least one model in this set. This implements optimism in the face of uncertainty and replaces posterior averaging with best-case evaluation over models consistent with the observations. The benefit is that GACBO avoids explicit entropy computations and aligns closely with optimistic bandit and Bayesian optimization principles, which can make it easier to implement and scale.

The main risk is over-optimism early in the learning process. When the confidence set is wide, the method may conduct trials on interventions that look good only under structural hypotheses that have not yet been ruled out. Performance, therefore, depends on the calibration of the confidence set and on how quickly interventional evidence eliminates incorrect graphs and mechanisms. Recent work extends this direction by exploring alternative uncertainty decompositions, including approaches that emphasize learning exogenous distributions and broader formulations of unknown-graph problems (Ren et al., 2026; Durand et al., 2025).

Summary. Unknown-graph CBO turns intervention design into an adaptive experimental design problem in which structure learning is not a preprocessing step but part of the optimization loop. The CEO and GACBO represent two complementary ways to couple these goals. The CEO prioritizes information gain about the best intervention under a Bayesian posterior, while GACBO prioritizes optimistic improvement under a frequentist-style set of plausible models. The choice between them largely reflects an information-versus-optimism trade-off and the computational budget available for representing and updating uncertainty over graphs.

3.5 Robustness to noise and priors

Even with a known graph, practical CBO pipelines face two recurring robustness issues. First, interventional outcomes can be noisy, heteroscedastic, or affected by unmodeled disturbances, which can destabilize improvement-based acquisitions. Second, observationally derived priors may be misspecified, either because

Table 2: Classification of major CBO methods by key dimensions. Each row lists a representative method with its causal assumption (graph knowledge), surrogate model type, optimization strategy, and distinctive intervention features.

Method	Causal graph	Surrogate model	Acquisition / strategy	Intervention features
CBO (2020)	Known DAG	Effect-level GP for Y with causal prior	Causal EI; cost-aware; observe vs. intervene	Scope + value, hard interventions
DCBO (2021)	Known DAG	Dynamic effect-level GP	Time-aware causal EI	Intervention sequences over time
MCBO (2023)	Known DAG	Mechanism-level GPs, full SCM	UCB / optimism, regret bounds	Hard/soft interventions; mechanism modeling
cCBO (2023)	Known DAG	Multi-output GP, objective and constraints	Constrained EI, feasibility-weighted	Safety and feasibility constraints
fCBO (2023)	Known DAG	GP over policies, RKHS function space	Functional EI	Functional and policy interventions
CoCaBO (2026)	Known DAG	BO within scope plus bandit over scopes	Two-layer, MAB-UCB then BO within scope	Contextual mixed interventions
ACBO (2024)	Known DAG	Mechanism-level modeling, full SCM	MW-style online learning plus causal UCB oracle	Adversarial and non-stationary environments
MO-CBO (2025)	Known DAG	GP per objective, local MOBO subproblems	Relative hypervolume improvement	Multi-objective Pareto optimization
HCBO (2024)	Known DAG	Multiple effect-level GPs, selected scopes	Normalized UCB-style scope scoring	High-dimensional scope selection
CEO (2023)	Unknown DAG	Bayesian model average over graphs, mixture	Entropy and information gain	Learns graph structure during optimization
GACBO (2024)	Unknown DAG	Confidence set over graphs and mechanisms	Optimism / UCB over plausible models	Structure discovery via optimistic exploration
CNEI (2023)	Known DAG	GP for Y with learned supervised prior	Counter-noise EI	Noise-robust acquisition; learned priors

identifiability assumptions fail, or because the estimand is hard to estimate accurately from finite data, or because observational correlations leak confounding into a prior mean.

Counter-noise Expected Improvement (CNEI) (Li et al., 2023) targets both issues through a noise-aware acquisition and a learned prior construction. CNEI modifies EI-style selection to account for the noise level so that the acquisition does not reward an apparent improvement dominated by measurement variance. In parallel, it estimates a surrogate prior mean by fitting predictive models to observational data and using the resulting predictor as a warm-start mean for the GP. This can improve the early-stage search by giving the surrogate a reasonable global shape before sufficient interventional data accumulate.

A key caveat is that observationally learned priors are not inherently causal. They can encode spurious associations when confounding is present, so their role should be weak guidance with appropriately inflated uncertainty rather than a substitute for interventional evidence. This motivates a broader open direction: robust CBO procedures that explicitly control how strongly observational information can influence acquisition decisions under noise, misspecification, and possible hidden confounding.

Table 3: Datasets used in empirical evaluations across CBO variants (as reported in the corresponding papers). We add a category column to make the dataset's family structure explicit and use light grid lines to improve readability. The dataset-usage table is restricted to single-objective CBO evaluations.

Category	Dataset	CBO	cCBO	fCBO	MCBO	HCBO	CEO	GACBO	CNEI	ACBO	CoCaBO
Synthetic (hard)	ToyGraph	✓	✓	✓	✓		✓	✓			
	Synthetic	✓							✓		
	Synthetic-2		✓								
	Chain-hard			✓							
Synthetic (soft)	Ackley				✓			✓		✓	
	Rosenbrock				✓			✓		✓	
	Dropwave				✓			✓		✓	
	Alpine2				✓			✓		✓	
	Chain-soft			✓							
Real / fitted SCMs (hard)	Ecology	✓				✓					
	Healthcare	✓	✓	✓	✓	✓	✓				✓
	Protein-reconstructed		✓								
	Epidemiology						✓	✓			

Summary. Robustness to noise and prior misspecification is a cross-cutting concern for all CBO methods, not only for CNEI. Any pipeline that constructs observational priors or learns surrogate hyperparameters from finite data faces the risk that estimation error or unmodeled confounding will mislead the acquisition function. The core open question is how to design CBO procedures that adaptively control the influence of observational information, treating it as useful guidance when identification is strong and discounting it when evidence of misspecification accumulates.

3.6 Comparison tables and unifying observations

Tables 2 and 3 summarize the field from two complementary angles. Table 2 organizes methods by key design dimensions: assumptions about the causal graph, the surrogate modeling target (effect-level, mechanism-level, or graph-uncertainty) and the acquisition or decision strategy. Table 3 maps empirical coverage and reveals substantial fragmentation: beyond a small core of shared benchmarks (ToyGraph, Healthcare), many datasets appear in only one or two papers, which complicates cross-paper performance comparison.

Together, these tables highlight two orthogonal axes of variation. The *modeling axis* ranges from effect-level surrogates (CBO, DCBO, cCBO, HCBO, CNEI) through mechanism-level models with uncertainty propagation (MCBO, ACBO) to explicit graph-uncertainty methods (CEO, GACBO). The *action axis* ranges from scope-and-value selection (CBO, DCBO, HCBO) through policy- or context-dependent decisions (fCBO, CoCaBO) to adversarial interaction (ACBO). These axes interact: mechanism-level models naturally support soft interventions but require maintaining multiple GPs, whereas effect-level models are lighter but cannot reuse intermediate observations. Recognizing this interaction is important for practitioners choosing a method and for researchers identifying gaps in the design space.

3.7 Connections to Adjacent Fields

CBO does not exist in isolation. Several established research areas share overlapping goals, tools, or assumptions. Understanding these connections clarifies what the CBO inherits, where it innovates, and what cross-pollination opportunities remain.

Causal bandits. The causal bandit literature (Lattimore et al., 2016; Lu et al., 2021; Nair et al., 2021) also selects interventions using causal structure, but typically assumes discrete action sets, parametric reward models, and focuses on cumulative regret in stationary environments. CBO extends this setting in three directions: (i) continuous intervention domains requiring nonparametric surrogates, (ii) mixed discrete–continuous action spaces (scope value + and (iii) more robust uncertainty modeling through GP. Conversely, causal bandits offer tighter finite-sample regret bounds and principled treatments of partial graph knowledge

that CBO methods have only begun to adopt (e.g., MCBO's regret analysis draws on GP-bandit theory). A promising direction is to unify the two frameworks by developing CBO methods with bandit-style theoretical guaranties for continuous domains.

Bayesian experimental design. The Bayesian optimal experimental design (BOED) (Chaloner & Verdinelli, 1995; Foster et al., 2021) selects experiments to maximize information about the model parameters. CEO (Branchini et al., 2023) can be viewed as a BOED method applied to intervention selection: choose experiments to maximize the information gained about the identity of the best intervention. The key distinction is that standard BOED targets parameter estimation without an optimization objective, whereas CEO jointly optimizes information and reward. This connection suggests that advances in amortized BOED (Foster et al., 2021) could accelerate information-seeking CBO methods.

Active causal discovery. Active structure learning (Spirtes et al., 2000; Chickering, 2002; Eberhardt & Scheines, 2007) selects interventions to learn the causal graph itself, while CBO selects interventions to optimize an outcome. Unknown-graph CBO methods (CEO, GACBO) bridge these goals by learning graph structure as a means to better optimization. The distinction is that active discovery aims for global graph accuracy, whereas CBO needs only *decision-relevant* structure, the parts of the graph that change which intervention is optimal. This focus on decision-relevant learning is a truly novel aspect of unknown-graph CBO.

Safe and constrained optimization. Safe Bayesian optimization (Sui et al., 2015; Berkenkamp et al., 2023) maintains feasibility during optimization, which is closely related to cCBO's (Aglietti et al., 2023) constrained causal optimization. The causal contribution is that safety constraints can be analyzed through the graph: a scope that cannot affect a constraint variable need not model constraint feasibility, reducing the multi-output modeling burden. Future work could combine safe BO's theoretical safety guaranties with CBO's causal constraint analysis.

Policy search and reinforcement learning. Functional CBO (fCBO) (Gultchin et al., 2023) and contextual CBO (CoCaBO) (Arsenyan et al., 2026) can be viewed as policy search methods operating within an explicit causal model. Unlike model-free reinforcement learning, these methods exploit the known graph to decompose the policy-to-outcome mapping and to construct informative priors from observational data. The main limitation relative to RL is the assumption of a single-step (or few-step) decision problem; extending CBO to sequential multi-stage intervention policies remains an open challenge.

4 Experiments

Although this paper is primarily a survey, we include a benchmark study to support two of the survey's main claims. First, existing CBO papers are often evaluated in heterogeneous codebases and on only partially overlapping datasets, making cross-paper comparison difficult. Second, commonly reported efficiency metrics can be sensitive to how optimization trajectories are summarized, especially when improvements occur late in the evaluation budget. Our benchmark is therefore designed to separate algorithm execution from evaluation: methods can run in their native implementations, but their outputs are exported into a common best-so-far trajectory format and re-scored with identical metric code.

4.1 CBO benchmark and evaluation protocol

We release a CBO benchmark at `https://anonymous.4open.science/status/CausalBO_Benchmark`. The benchmark includes eight curated hard-intervention scenarios and a separate set of soft-intervention benchmarks. It provides standardized implementations of GAP and PA-GAP. Because existing CBO methods often rely on incompatible software environments, the benchmark does not force all algorithms into a single execution stack. Instead, each method is run in its native or vendored implementation, and its trajectory is exported and re-scored using the same evaluation pipeline. This design enables reproducible comparisons across heterogeneous implementations and makes it easy to add new methods, datasets, or metrics.

Execution protocol. For each data set, method, and trial limit, we run 20 random seeds. A seed fixes the observational sample, the initial interventional design, and the stochastic components of the optimizer. All methods are evaluated under the same trial limits, intervention domains, optimization direction, and scoring code. An interventional evaluation counts as one trial. Observational data used to construct priors, learn graphs, or initialize surrogates are fixed before the optimization loop and are not counted as interventional trials unless otherwise stated. Each method is run with its released default hyperparameters unless explicitly stated in Appendix A.2; we do not tune hyperparameters separately for each data set.

Trajectory export and scoring. The output of each method is converted to a common trajectory format with two required fields: trial number and best-so-far objective value in the natural optimization direction of the task. For minimization tasks, this is the minimum value observed up to the trial t; for maximization tasks, it is the maximum value observed up to the trial t. The trajectory includes an initial value followed by interventional trials. Thus, if the budget is B interventions, the exported trajectory contains $B + 1$ entries: one initial best value and B post-intervention best-so-far values. GAP and PA-GAP are then computed from these exported trajectories using the same evaluation script for all methods.

Reference optimum. Both GAP and PA-GAP require a reference optimum y^*. In our benchmark, y^* is computed offline with oracle access to the benchmark SCM, function generator, or precomputed interventional data. This value is used only for scoring and is never provided to any optimizer. For low-dimensional or discrete intervention domains, we enumerate or densely evaluate admissible intervention scopes and values. For continuous domains, we use high-budget offline search with multi-start optimization or stored oracle intervention sets. The exact offline budgets, grids, restarts, and dataset-specific procedures are reported in the Appendix A.3.

Failure handling. When a method proposes an infeasible or out-of-domain intervention, we follow the behavior of the released implementation. If the implementation projects the proposals back to the admissible domain, the projected intervention is evaluated. Otherwise, the trial is marked as a failed proposal, and the trajectory records the current best-so-far value. Runs with numerical errors, missing outputs, or non-finite objective values are logged and are not silently removed. Runtime is not used as a primary metric because implementations differ substantially in language, hardware support, and dependency stacks.

Cost model. Unless otherwise stated, each intervention has a unit cost. This choice isolates the efficiency of the sample but does not evaluate the full cost-aware behavior of the CBO acquisition functions. In real applications, the cost of the intervention can depend on the target variable, the number of variables manipulated, the magnitude of the intervention, safety constraints, or whether the action is observational or experimental. We therefore report the current results as unit-cost benchmarks and treat realistic intervention-cost modeling as a separate evaluation dimension.

4.2 Performance metrics

We evaluated each run using two numerical efficiency metrics, GAP and PA-GAP, and one visualization metric based on the best-so-far objective trajectory. All scalar metrics are computed from the same exported trajectories described above.

GAP. GAP was introduced in DCBO to quantify the efficiency of optimization by combining the final improvement with the speed with which the best value is discovered (Aglietti et al., 2021). Let R_t denote the best improvement ratio so far in the trial t, computed in the natural optimization direction of the task. For maximization tasks,

$$R_t = \min \left\{ \frac{y(\mathbf{x}_t^*) - y(\mathbf{x}_{\text{init}})}{y^* - y(\mathbf{x}_{\text{init}})}, 1 \right\}, \tag{16}$$

and for minimization tasks,

$$R_t = \min \left\{ \frac{y(\mathbf{x}_{\text{init}}) - y(\mathbf{x}_t^*)}{y(\mathbf{x}_{\text{init}}) - y^*}, 1 \right\}. \tag{17}$$

Here, $\mathbf{x}_t^*$ denotes the best point so far in the trial t. The clipping prevents numerical artifacts when an observed value exceeds the offline reference optimum. Let R_T be the final improvement ratio and let t^* be the first trial in which R_T is achieved. If no improvement occurs, we set $t^* = T$. GAP is then

$$\mathrm{GAP} = \left[\underbrace{R_T}_{\text{Final improvement}} + \underbrace{\frac{T - t^*}{T}}_{\text{Discovery efficiency}} \right] \Bigg/ \underbrace{\left(1 + \frac{T-1}{T}\right)}_{\text{Normalization}}. \tag{18}$$

Under this convention, GAP lies in $[0, 1]$, with larger values indicating larger final improvement and earlier discovery.

Path-Aware GAP (PA-GAP). GAP depends only on the final best value and the first trial at which it is reached. This can under-value runs that improve late or make steady progress over the full budget. To better reflect the full optimization trajectory, we use Path-Aware GAP:

$$\text{PA-GAP} = \frac{1}{T} \sum_{t=1}^{T} \left(\underbrace{R_t}_{\text{Improvement ratio}} \cdot \underbrace{\frac{T - (t-1)}{T}}_{\text{Efficiency weight}} \right). \tag{19}$$

PA-GAP multiplies the best-so-far improvement by a time-dependent efficiency weight, so later improvements are down-weighted but not discarded, and averages these weighted improvements over the full trajectory. Since $R_t \in [0, 1]$, raw PA-GAP lies in

$$\left[0, \frac{T+1}{2T}\right].$$

The upper bound is attained when the reference optimum is found in the first trial and remains the best value so far thereafter. Thus, PA-GAP is a raw trajectory-weighted efficiency score rather than a unit-normalized metric. Larger values indicate faster and stronger trajectory-level improvement, and raw PA-GAP values should be compared directly only under the same trial budget. Appendix A.1.2 gives a counterexample where GAP ranks a worse final trajectory above a better late-improving one, while PA-GAP resolves the ordering.

Trajectory visualization. In addition to scalar metrics, we plot the best-so-far trajectories over trials, reported as mean ± standard deviation over seeds. We do not emphasize average-reward trajectories because they are sensitive to the acquisition rule. For example, UCB-style methods may intentionally sample high-uncertainty points, reducing the running mean while improving long-term search (Srinivas et al., 2010; Brochu et al., 2010).

4.3 Datasets and task conventions

Table 3 summarizes the datasets used in the CBO literature. Most evaluations rely on SCM-generated data, where observational samples are drawn from the unintervened SCM and interventional samples are generated by applying hard or soft interventions to the SCM mechanisms. The optimization direction is data-dependent. For hard-intervention SCM benchmarks, we follow the standard CGO/CBO convention and treat the target as a minimization objective unless explicitly stated otherwise. ToyGraph, Synthetic, Synthetic-2, Chain-hard, Healthcare, Protein-reconstructed, and Epidemiology are minimization tasks, while Ecology is a maximization task. For the soft-intervention benchmarks for the function-network, Ackley, Rosenbrock, Dropwave, and Alpine2 are treated as reward-maximization tasks following the convention used in the BO, MCBO, and ACBO function-network. Chain-soft is a minimization task because the objective is to reduce the target variable Y.

4.3.1 Hard-intervention SCM benchmarks

Hard-intervention benchmarks evaluate optimization when intervened variables are clamped to fixed values. They include small synthetic SCMs, a reconstructed protein-signaling SCM, and real-world-inspired SCMs

from ecology, healthcare, and epidemiology. The benchmark also provides optional high-dimensional generators, but the main cross-method comparison focuses on the eight curated scenarios for which multiple implementations can be evaluated under the same protocol.

ToyGraph. ToyGraph was introduced in the original CBO work (Aglietti et al., 2020). It is a minimal three-node chain-structured SCM with variables $X \to Z \to Y$, where X and Z are manipulable and Y is the target. The task is to minimize Y. The benchmark uses a simple nonlinear mechanism in which Z depends exponentially on X and Y is generated from a nonlinear transformation of Z. Its small size makes it a useful sanity-check benchmark for hard-intervention methods. Detailed equations and visualization are provided in the Appendix A.5.1.

Synthetic. Synthetic was also introduced in the original CBO work (Aglietti et al., 2020). It contains seven observed variables, A, B, C, D, E, F, Y and two latent variables, U_1 and U_2. The variables B, D, and E are manipulable, while Y is the target. The task is to minimize Y. The SCM combines nonlinear dependencies, latent variables, and multiple intervention options, making it a richer testbed than ToyGraph for structured causal optimization. Detailed equations and visualization are provided in the Appendix A.5.2.

Synthetic-2. Synthetic-2 was introduced into the cCBO line of work (Aglietti et al., 2023). It is a lightweight three-node SCM with variables $X \to Z \to Y$, where X and Z are manipulable and Y is the target. The task is to minimize Y. It follows a nonlinear chain structure similar to ToyGraph but uses explicitly Gaussian exogenous noise and is generated directly from the specified SCM. Detailed equations and visualization are provided in the Appendix A.5.3.

Chain-hard. Chain-hard is the hard-intervention version of the Chain benchmark introduced in fCBO (Gultchin et al., 2023). It consists of four observed variables X, W, Z, Y, where W and Z are manipulable, X is a non-manipulable context variable and Y is the target. The task is to minimize Y. In the setting of hard-interventions, interventions in W and Z replace the corresponding structural equations with fixed constants in $[-1, 1]$. Because the result depends on the interaction between Z and X, this benchmark is more expressive than the simple chain benchmarks while remaining low-dimensional. Detailed equations and visualization are provided in the Appendix A.5.4.

Ecology. The Ecology benchmark was introduced in the CBO (Aglietti et al., 2020) and is based on a Bermuda reef ecosystem model (Courtney et al., 2017). The task is to maximize net ecosystem calcification under environmental interventions. The SCM includes variables such as chlorophyll on the surface of the sea a, salinity, total alkalinity, dissolved inorganic carbon, seawater pCO_2, bottom temperature, bottom light, availability of nutrients, seawater pH, saturation state of aragonite, and the target variable NEC. We follow the CBO setup and use the available SCM and the data released in the public CBO repository. Detailed equations and visualization are provided in the Appendix A.6.1.

Protein-reconstructed. The Protein benchmark is based on the protein-signaling data of Sachs et al. (2005) and the graph used in cCBO (Aglietti et al., 2023). The target is the extracellular signal-regulated kinases 1 and 2, Erk, and the task is to minimize Erk by perturbing Mek, PKC, PKA. and Akt while respecting biological constraints on PKC and PKA. Because the fitted SCM used in the original cCBO Protein experiments is not released, our benchmark is a reconstruction rather than an exact reproduction. We retain the connected subgraph used in cCBO, consisting of Raf, Erk, P38, Jnk, Akt, Mek, PKA, and PKC, and fit structural mechanisms using 852 observational samples from Sachs et al. (2005). We label this data set as Protein-reconstructed in the benchmark files to make it clear that it tests transparent SCM reconstruction rather than bitwise replication of cCBO. Appendix A.6.2 reports the retained variables, parent sets, preprocessing, functional model class, fitting procedure, constraints, intervention domains, equations, and visualization.

Healthcare. Healthcare was introduced in CBO (Aglietti et al., 2020). It is based on a causal graph over age, BMI, cancer status, statin use, aspirin use, and PSA level (Thompson & Yung, 2019; Ferro et al., 2015). The objective is to minimize PSA through interventions with statin and aspirin. Although the original

data template specifies several variables as binary, existing CBO implementations commonly relax aspirin and statin to continuous variables in $(0, 1)$ for BO-style optimization. We follow this convention to remain comparable to the prior work. Detailed equations and visualization are provided in the Appendix A.6.3.

Epidemiology. Epidemiology was introduced in CEO (Branchini et al., 2023). The task is to minimize the viral load of HIV by selecting doses for two possible treatments, denoted by T and R (Havercroft & Didelez, 2012). In the CEO benchmark, Y is the target viral load, B is a baseline or background non-manipulable variable, and L is an intermediate non-manipulable health variable. Since CEO uses a modified HIV SCM, more specific clinical interpretations of B and L are not provided in the original article. We follow CEO and use the given SCM to generate both observational and interventional datasets. Detailed equations and visualization are provided in the Appendix A.6.4.

4.3.2 Soft-intervention function-network benchmarks

Soft-intervention benchmarks evaluate settings where interventions modify structural mechanisms rather than simply clamping variables to fixed values. Ackley, Rosenbrock, Dropwave and Alpine2 are classical global optimization benchmarks (Jamil & Yang, 2013) reformulated as BO tasks of the function-network and adopted in MCBO and ACBO (Sussex et al., 2023; 2024). In these tasks, the output node is treated as a reward to be maximized, even when this corresponds to a sign-flipped version of a standard minimization benchmark. Chain-soft is instead a minimization task because the objective is to reduce the target variable Y.

Ackley. Ackley is a continuous, differentiable, non-separable, scalable, and multimodal benchmark. In its function-network form, the task is to maximize the output node. The network decomposes the objective into two intermediate components: one for the average squared magnitude of the inputs and one for the average cosine term. These components are combined at the output node to recover the Ackley objective in reward-maximization form. We use the domain $[-2, 2]^D$ with $D = 6$. Detailed equations and visualization are provided in the Appendix A.7.1.

Rosenbrock. Rosenbrock is a continuous, differentiable, non-separable, scalable, and unimodal benchmark known for its narrow curved valley. In the function-network formulation, the task is to maximize the output node. The network is chain-structured: the first node computes the first Rosenbrock term, and each subsequent node recursively adds the contribution of an adjacent variable pair to the accumulated output. We use the domain $[-2, 2]^D$ and consider $D \in \{3, 5, 7\}$. Detailed equations and visualization are provided in the Appendix A.7.2.

Dropwave. Dropwave is a highly multimodal two-dimensional benchmark. In function-network form, the task is to maximize the output node. An intermediate node first computes the radial quantity $\sqrt{x_1^2 + x_2^2}$, and then the output node applies the oscillatory Dropwave transformation. We use the standard domain $[-5.12, 5.12]^2$. Detailed equations and visualization are provided in the Appendix A.7.3.

Alpine2. Alpine2 is a continuous, differentiable, separable, scalable, and multimodal benchmark. In function-network form, the task is to maximize the output node. It is represented as a chain graph in which each node combines its own decision variable with the previous node through the multiplicative structure of Alpine2. We use the domain $[0, 10]^K$ with $K = 6$. Detailed equations and visualization are provided in the Appendix A.7.4.

Chain-soft. Chain-soft extends Chain-hard by allowing the mechanism of Z to be modified through a context-dependent policy, rather than restricting interventions to fixed clamped values (Gultchin et al., 2023). The task is to minimize Y. In this setting, Z can depend on the upstream context X through a policy $\pi_0(X)$, while W remains subject to standard hard interventions on $[-1, 1]$. The target depends on the interaction between Z and X, so the optimal intervention in Z generally varies with X. This makes Chain-soft a richer benchmark for context-adaptive intervention strategies. Detailed equations, policy specification, and visualization are provided in the Appendix A.7.5.

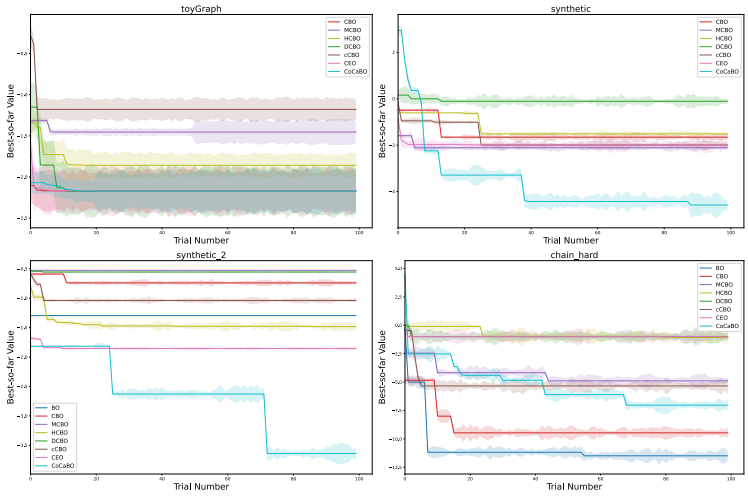

Figure 2: Best-so-far trajectories on four synthetic hard-intervention datasets: ToyGraph, Synthetic, Synthetic-2, and Chain-hard. Solid lines show the mean across 20 seeds and shaded regions show the standard deviation. Lower values indicate better objective values for these minimization tasks.

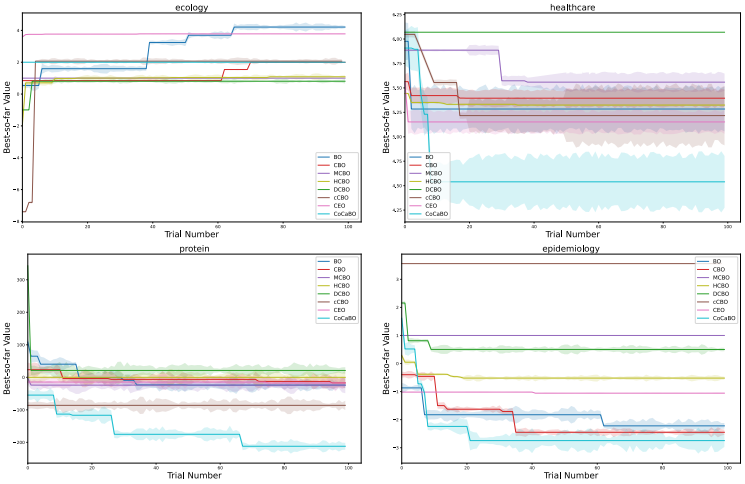

Figure 3: Best-so-far trajectories on four real or fitted-SCM hard-intervention datasets: Ecology, Protein-reconstructed, Healthcare, and Epidemiology. Solid lines show the mean across 20 seeds and shaded regions show the standard deviation. Ecology is a maximization task, while Protein-reconstructed, Healthcare, and Epidemiology are minimization tasks.

4.4 Hard-intervention benchmark

Compared methods. We compare a non-causal baseline, BO, with seven CBO-family methods: CBO, cCBO, DCBO, MCBO, HCBO, CoCaBO, and CEO (Aglietti et al., 2020; 2023; 2021; Sussex et al., 2023; Wu et al., 2024; Arsenyan et al., 2026; Branchini et al., 2023). This comparison is designed to evaluate whether the use of causal structure improves the search for the intervention relative to a standard black-box BO baseline under a shared scoring protocol. We exclude fCBO, CNEI, and GACBO because we did not obtain compatible public implementations of the benchmark protocol. We exclude ACBO and MO-CBO from this hard-intervention benchmark because their released implementations or task settings are not aligned with the single-objective hard-intervention protocol considered here. CEO is included as an unknown-graph method and therefore spends part of its budget resolving structure uncertainty, whereas known-graph methods receive the benchmark graph.

Table 4: Hard-intervention performance across different trial limits and datasets. Each value is the average ± standard deviation across 20 random seeds and different initializations of the observational and interventional data. Higher is better for both GAP and PA-GAP. Best mean values in each row are in **bold**.

Trial limit	Metric	Dataset	BO	CBO	cCBO	DCBO	MCBO	HCBO	CoCaBO	CEO
100	GAP	ToyGraph	0.346±.204	0.810±.198	0.730±.102	**0.949±.213**	0.547±.142	0.727±.136	0.625±.241	0.654±.178
		Synthetic	0.401±.172	0.658±.104	0.691±.123	0.482±.032	0.652±.132	0.559±.106	**0.757±.219**	0.565±.172
		Synthetic-2	0.000±.000	0.544±.092	0.549±.078	0.485±.012	0.000±.000	0.174±.071	**0.562±.202**	0.378±.172
		Chain-hard	0.721±.142	0.688±.061	**0.734±.114**	0.542±.121	0.502±.205	0.135±.108	0.205±.081	0.502±.152
		Ecology	**0.799±.184**	0.511±.078	0.697±.121	0.637±.115	0.000±.000	0.334±.132	0.000±.000	0.336±.142
		Protein-reconstructed	0.751±.132	0.449±.194	0.000±.000	0.000±.000	**0.878±.284**	0.412±.142	0.664±.231	0.275±.092
		Healthcare	0.310±.089	0.577±.108	0.842±.132	0.487±.004	0.510±.104	0.335±.032	**0.964±.195**	0.696±.294
		Epidemiology	0.634±.218	0.844±.128	0.000±.000	0.375±.125	0.000±.000	0.557±.109	**0.898±.262**	0.308±.092
	PA-GAP	ToyGraph	0.224±.139	0.069±.039	0.235±.135	**0.337±.193**	0.071±.045	0.285±.162	0.087±.053	0.295±.167
		Synthetic	0.198±.155	0.172±.124	0.184±.119	0.036±.020	0.162±.103	0.107±.091	**0.411±.013**	0.358±.200
		Synthetic-2	0.000±.000	0.055±.037	0.065±.037	0.002±.001	0.000±.000	0.099±.054	**0.129±.094**	0.035±.021
		Chain-hard	**0.495±.009**	0.280±.189	0.244±.144	0.049±.028	0.111±.065	0.026±.022	0.038±.017	0.002±.001
		Ecology	**0.303±.173**	0.117±.026	0.204±.122	0.144±.086	0.000±.000	0.210±.115	0.000±.000	0.016±.008
		Protein-reconstructed	**0.407±.195**	0.254±.162	0.000±.000	0.000±.000	0.045±.026	0.000±.000	0.240±.132	0.014±.018
		Healthcare	0.125±.064	0.175±.088	0.170±.099	0.000±.000	0.100±.093	0.010±.005	**0.329±.201**	0.200±.114
		Epidemiology	0.299±.169	0.346±.092	0.000±.000	0.175±.098	0.000±.000	0.138±.075	**0.392±.212**	0.005±.005
50	GAP	ToyGraph	0.296±.192	0.676±.201	0.723±.098	**0.896±.183**	0.513±.106	0.643±.132	0.629±.216	0.690±.162
		Synthetic	0.531±.082	0.597±.108	0.565±.103	0.422±.012	0.630±.099	0.432±.087	**0.793±.251**	0.372±.152
		Synthetic-2	0.000±.000	0.459±.093	**0.531±.061**	0.467±.032	0.000±.000	0.160±.091	0.438±.103	0.213±.192
		Chain-hard	**0.919±.081**	0.462±.041	0.712±.102	0.535±.109	0.215±.120	0.302±.071	0.106±.072	0.505±.142
		Ecology	**0.798±.165**	0.363±.065	0.680±.102	0.624±.094	0.000±.000	0.246±.071	0.000±.000	0.151±.082
		Protein-reconstructed	0.656±.121	0.553±.172	0.000±.000	0.000±.000	**0.752±.184**	0.320±.078	0.731±.182	0.377±.113
		Healthcare	0.366±.086	0.598±.087	0.758±.121	0.474±.002	0.315±.087	0.155±.010	**0.927±.162**	0.694±.204
		Epidemiology	0.750±.194	0.644±.109	0.000±.000	0.596±.092	0.000±.000	0.461±.082	**0.793±.182**	0.096±.003
	PA-GAP	ToyGraph	0.177±.129	0.068±.038	0.233±.134	0.316±.180	0.063±.044	0.262±.147	**0.357±.225**	0.296±.164
		Synthetic	0.130±.095	0.131±.113	0.191±.074	0.033±.018	0.150±.099	0.050±.060	0.349±.181	**0.357±.141**
		Synthetic-2	0.000±.000	0.043±.035	0.063±.035	0.002±.001	0.000±.000	**0.091±.049**	0.051±.063	0.033±.020
		Chain-hard	**0.500±.009**	0.178±.135	0.233±.140	0.049±.028	0.078±.060	0.012±.015	0.034±.016	0.002±.001
		Ecology	**0.306±.173**	0.085±.058	0.194±.120	0.140±.085	0.000±.000	0.208±.111	0.000±.000	0.015±.008
		Protein-reconstructed	**0.339±.140**	0.192±.150	0.000±.000	0.000±.000	0.045±.026	0.000±.000	0.157±.098	0.003±.003
		Healthcare	0.120±.064	0.131±.082	0.263±.170	0.000±.000	0.034±.049	0.009±.005	**0.295±.190**	0.202±.114
		Epidemiology	0.247±.175	0.222±.088	0.000±.000	0.169±.093	0.000±.000	0.127±.065	**0.354±.182**	0.000±.001
20	GAP	ToyGraph	0.383±.098	0.510±.198	0.702±.103	**0.729±.172**	0.406±.089	0.372±.098	0.539±.194	0.508±.124
		Synthetic	0.239±.062	0.392±.089	0.367±.098	0.231±.011	0.558±.072	0.166±.010	**0.756±.231**	0.390±.123
		Synthetic-2	0.000±.000	0.285±.076	**0.474±.081**	0.408±.009	0.000±.000	0.186±.078	0.000±.000	0.282±.102
		Chain-hard	**0.819±.063**	0.136±.014	0.641±.098	0.510±.082	0.360±.103	0.000±.000	0.065±.051	0.513±.125
		Ecology	**0.795±.176**	0.208±.015	0.625±.091	0.585±.071	0.000±.000	0.271±.054	0.000±.000	0.340±.192
		Protein-reconstructed	0.503±.098	0.508±.152	0.000±.000	0.000±.000	0.540±.162	0.028±.073	**0.648±.183**	0.168±.071
		Healthcare	0.604±.069	0.392±.092	0.490±.083	0.432±.001	0.000±.000	0.172±.003	**0.810±.076**	0.689±.162
		Epidemiology	0.646±.135	0.452±.098	0.000±.000	0.457±.078	0.000±.000	0.154±.065	**0.791±.126**	0.000±.000
	PA-GAP	ToyGraph	0.088±.096	**0.392±.228**	0.227±.131	0.260±.151	0.040±.037	0.200±.177	0.190±.130	0.300±.160
		Synthetic	0.026±.035	0.034±.053	0.147±.077	0.024±.013	0.116±.092	0.000±.001	0.334±.157	**0.361±.190**
		Synthetic-2	0.000±.000	0.016±.021	0.057±.030	0.002±.001	0.000±.000	**0.074±.037**	0.000±.000	0.027±.018
		Chain-hard	**0.357±.158**	0.032±.030	0.200±.128	0.047±.028	0.034±.041	0.000±.000	0.031±.016	0.002±.001
		Ecology	**0.316±.173**	0.030±.034	0.166±.112	0.126±.084	0.000±.000	0.205±.106	0.000±.000	0.015±.007
		Protein-reconstructed	0.250±.107	0.070±.091	0.000±.000	**0.290±.159**	0.047±.026	0.000±.000	0.182±.200	0.001±.001
		Healthcare	0.121±.064	0.034±.081	0.143±.030	0.000±.000	0.000±.000	0.008±.004	**0.268±.199**	0.208±.114
		Epidemiology	0.158±.148	0.078±.093	0.000±.000	0.152±.083	0.000±.000	0.102±.050	**0.331±.151**	0.000±.000

Selected datasets. We evaluated eight hard-intervention datasets: ToyGraph, Synthetic, Synthetic-2, Chain-hard, Ecology, Protein-reconstructed, Healthcare, and Epidemiology. These datasets cover small synthetic SCMs, real or fitted-SCM benchmarks, and settings that are unevenly shared across prior CBO papers. Protein-reconstructed is included as a transparent reconstruction of the cCBO Protein benchmark, rather than as an exact reproduction of the unreleased fitted SCM used in the original cCBO experiments. Ecology is treated as a maximization task; the remaining hard-intervention datasets are treated as minimization tasks.

Results summary. The hard-intervention benchmark shows that no method dominates uniformly across datasets, budgets, and metrics. CoCaBO has the strongest overall rank, followed by cCBO, CBO, and BO, but this ranking is not explained by GAP alone. GAP emphasizes the final best value and the time at which that final value is first found, so methods that quickly reach a strong final solution are rewarded. In this regard, CoCaBO performs especially well in Synthetics, Healthcare, and Epidemiology, DCBO performs well in ToyGraph, and MCBO performs well in Protein-reconstructed. In contrast, PA-GAP rewards the entire best-so-far trajectory, so methods that make early or steady progress can rank well even when they are not the best final performer. This explains why CBO has a strong aggregate rank despite not being the dominant GAP winner, and why BO remains competitive: on datasets such as Chain-hard, Ecology, and Healthcare, it often finds useful interventions early enough to receive favorable trajectory-level scores. These results show

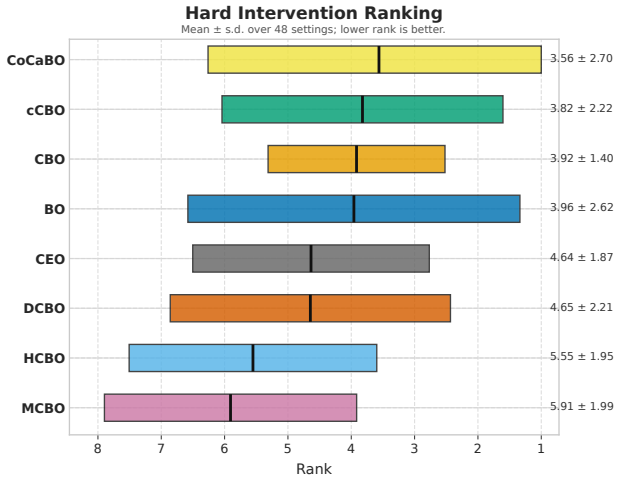

Figure 4: Average rank across 48 hard-intervention settings, covering eight datasets, three trial limits, and two metrics. Ranks are computed within each dataset-budget-metric setting before averaging. Lower rank is better; error bars show standard deviation across settings.

that causal structure can improve optimization, but its benefit depends on how the method's assumptions match the dataset and on whether performance is measured by final discovery or by the full optimization path.

Trajectory-level behavior. Figures 2 and 3 complement the scalar table by showing how the methods reach their final scores. In synthetic datasets, many methods improve through a few large drops followed by long plateaus, suggesting that much of the gain comes from identifying a useful intervention scope rather than from gradual local refinement. This also explains why GAP and PA-GAP can disagree. A method that reaches a moderate value early can receive a high GAP if that value is its final best, while PA-GAP gives more credit to trajectories that continue moving toward the reference optimum. In Synthetic, CoCaBO shows strong path-level progress, consistent with its high scores under both metrics. On ToyGraph, DCBO is strong under GAP, while PA-GAP shows that early progress with CBO and HCBO-style can also matter. In Chain-hard, BO is highly competitive, indicating that a flexible black-box optimizer can perform well when the effective intervention landscape is simple or when the causal prior does not provide a decisive advantage.

The real or fitted-SCM datasets show even stronger metric dependence. Some methods remain nearly flat on particular datasets, which explains the zero or near-zero metric values in Table 4. Ecology and Healthcare also show that strong performance can arise through different paths: some methods improve early and then plateau, while others discover a better intervention only after many trials. Reporting both GAP and PA-GAP is therefore important. GAP captures the final improvement and discovery time, while PA-GAP captures whether the optimizer makes meaningful progress throughout the budget.

Aggregate ranking across settings. Figure 4 summarizes the benchmark from a rank-based perspective. CoCaBO has the best average rank, followed by cCBO, CBO, and BO. The ordering highlights two points. First, causal methods can provide an advantage when their structural assumptions match the benchmark, as seen for CoCaBO and cCBO. Second, the advantage is not automatic: BO remains close to the CBO-family methods because PA-GAP rewards reliable early progress, and BO often performs well on datasets with simple or smooth effective intervention landscapes. The ranking therefore reinforces the main benchmark message: CBO methods should be compared using scalar tables, trajectory plots, and rank summaries, because a single final-value metric can hide important differences in how methods make progress.

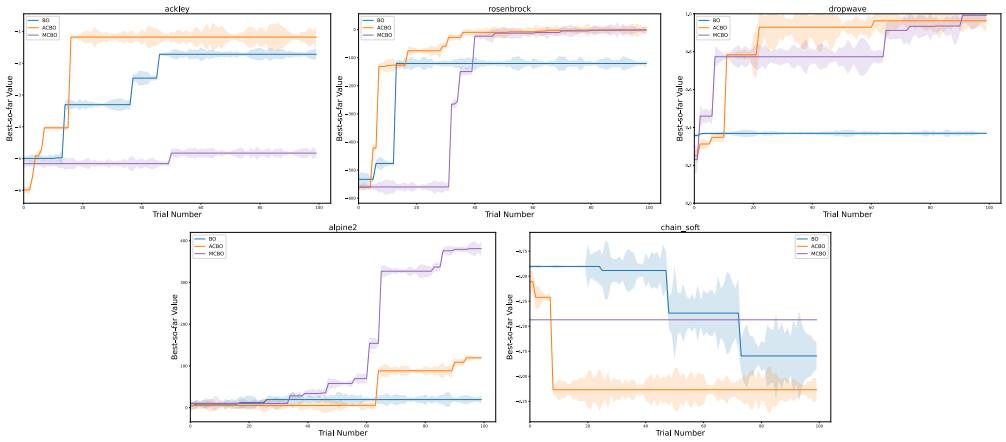

Figure 5: Best-so-far trajectories on five soft-intervention datasets: Ackley, Rosenbrock, Dropwave, Alpine2, and Chain-soft. Solid lines show the mean over 20 random seeds and shaded regions denote the standard deviation across runs.

4.5 Soft-intervention benchmark

For soft interventions, we compare BO, MCBO, and ACBO in Ackley, Rosenbrock, Dropwave, Alpine2 and Chain-soft (Sussex et al., 2023; 2024). BO serves as a non-causal baseline, while MCBO and ACBO represent mechanism-level and adversarial CBO-style approaches. fCBO also supports policy interventions, but it is excluded from the comparison because we did not obtain a compatible reference implementation (Gultchin et al., 2023). Ackley, Rosenbrock, Dropwave, and Alpine2 are treated as reward-maximization tasks; Chain-soft is a minimization task.

Results summary. The soft-intervention benchmark shows that ACBO has the strongest overall rank, followed by BO and then MCBO, but the table also shows that this ordering is not due to uniform dominance. ACBO is especially strong on Ackley and Chain-soft, where it obtains the best GAP across budgets and often achieves strong PA-GAP values, indicating fast and sustained progress. BO remains a highly competitive baseline on Rosenbrock and Alpine2. In these landscapes, standard black-box optimization often reaches strong final values and also receives favorable PA-GAP scores when it improves early and then maintains its best-so-far value. MCBO is most competitive on Dropwave, where it achieves the best GAP at 50 and 20 trials and the best PA-GAP at 50 and 20 trials. This pattern highlights the role of metric choice: GAP rewards the final best value and its discovery time, while PA-GAP rewards the whole best-so-far trajectory. As a result, methods such as BO can rank well when they make reliable early progress, even if they are not always the best final performer. Overall, the soft-intervention results reinforce the main conclusion from the hard-intervention benchmark: causal modeling can improve optimization, but its advantage depends on how well the method's assumptions match the intervention type, objective landscape, and evaluation metric.

Trajectory-level behavior. Figure 5 shows that the three methods behave differently in objective landscapes. ACBO improves quickly on Ackley and Chain-soft, which explains its strong scalar scores on these datasets. BO remains highly competitive in Rosenbrock and Alpine2, where its best-so-far trajectory often improves early enough to receive strong PA-GAP values as well as strong GAP values. Dropwave shows a complementary pattern: MCBO and ACBO both improve substantially, but MCBO is stronger under tighter budgets after the corrected PA-GAP value for ACBO at 50 trials. These trajectory patterns explain why the aggregate rank is not determined only by the number of final GAP wins. PA-GAP gives additional credit to methods that improve early and maintain progress, which is why BO remains close to ACBO overall despite being a non-causal baseline.

Aggregate ranking across settings. Figure 6 summarizes performance in all soft-intervention settings. ACBO has the best average rank, followed by BO and then MCBO. This ranking supports the interpretation

Table 5: Soft-intervention performance across different trial limits and datasets. Each value is the average ± standard deviation across 20 random seeds and initializations of the observational and interventional data. Higher is better; best mean values per row are in **bold**.

Trial limit	Metric	Dataset	BO	MCBO	ACBO
100	GAP	Ackley	0.328±.098	0.310±.013	**0.922±.367**
		Rosenbrock	**0.825±.245**	0.507±.184	0.623±.210
		Dropwave	0.024±.024	0.558±.292	**0.690±.419**
		Alpine2	**0.536±.207**	0.522±.231	0.178±.061
		Chain-soft	0.634±.124	0.512±.097	**0.693±.234**
	PA-GAP	Ackley	0.033±.020	0.012±.014	**0.403±.027**
		Rosenbrock	0.308±.207	0.217±.020	**0.412±.033**
		Dropwave	0.008±.004	0.355±.184	**0.379±.024**
		Alpine2	**0.120±.063**	0.076±.085	0.015±.024
		Chain-soft	**0.290±.051**	0.000±.003	0.205±.126
50	GAP	Ackley	0.090±.024	0.072±.007	**0.843±.318**
		Rosenbrock	**0.761±.189**	0.515±.219	0.631±.234
		Dropwave	0.235±.107	**0.803±.381**	0.318±.180
		Alpine2	**0.403±.153**	0.085±.032	0.175±.041
		Chain-soft	0.515±.104	0.000±.000	**0.655±.148**
	PA-GAP	Ackley	0.018±.014	0.000±.001	**0.320±.186**
		Rosenbrock	0.239±.182	0.053±.078	**0.344±.208**
		Dropwave	0.008±.004	**0.318±.180**	0.171±.107
		Alpine2	**0.085±.035**	0.003±.005	0.000±.000
		Chain-soft	0.135±.169	0.000±.000	**0.181±.118**
20	GAP	Ackley	0.166±.057	0.000±.000	**0.593±.219**
		Rosenbrock	**0.559±.157**	0.000±.000	0.499±.210
		Dropwave	0.252±.105	**0.701±.301**	0.293±.184
		Alpine2	**0.142±.058**	0.000±.000	0.000±.000
		Chain-soft	0.000±.000	0.000±.000	**0.534±.109**
	PA-GAP	Ackley	0.003±.006	0.000±.000	**0.151±.076**
		Rosenbrock	0.082±.082	0.000±.000	**0.216±.168**
		Dropwave	0.007±.004	**0.253±.134**	0.130±.094
		Alpine2	**0.049±.025**	0.000±.000	0.000±.000
		Chain-soft	0.000±.000	0.000±.000	**0.116±.086**

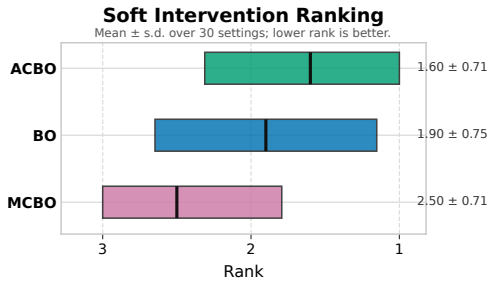

Figure 6: Average rank across 30 soft-intervention settings, covering five datasets, three trial limits, and two metrics. Ranks are computed within each dataset-budget-metric setting before averaging. Lower rank is better; error bars show standard deviation across settings.

that ACBO is the most reliable of the three methods in this benchmark, but it also shows that the advantage of causal or adversarial modeling is not automatic. BO obtains a strong rank because it is competitive in Rosenbrock and Alpine2 and because PA-GAP rewards its early trajectory quality in several settings. MCBO is more data set dependent: it is strong on Dropwave, where the mechanism-level structure appears useful, but weaker on Ackley, Alpine2, and Chain-soft. As in the hard-intervention benchmark, the combination of scalar tables, trajectory plots, and rank summaries gives a more balanced comparison than a single final-value metric.

4.6 Cross-cutting analysis

The hard and soft-intervention benchmarks together span 78 dataset–budget-metric settings. Rather than summarizing each setting individually, we distill the main patterns into four cross-sectional observations.

When does causal structure help? Causal methods provide the clearest advantage when the graph enables effective scope reduction or when observational priors are well-calibrated. In the hard-intervention setting, CoCaBO and cCBO benefit from structured scope selection on datasets with multiple intervenable variables and clear causal pathways (Synthetic, Healthcare, Epidemiology). In the soft-intervention setting, ACBO benefits from mechanism-level modeling on datasets with complex intermediate structure (Ackley, Chain-soft). Conversely, causal structure provides little or no advantage on datasets where the effective intervention landscape is simple (Chain-hard) or where causal priors are poorly calibrated (Ecology under some methods). BO remains competitive precisely in these settings, suggesting that the value of causal modeling is conditional on the alignment between the method's structural assumptions and the problem's actual causal complexity.

Budget sensitivity. Rankings are moderately sensitive to budget. At tight budgets ($T = 20$), methods that rely on observational priors or scope reduction (CBO, cCBO, DCBO) tend to perform relatively better because their prior information is most valuable when interventional data are scarce. At larger budgets ($T = 100$), methods with richer exploration mechanisms (CoCaBO, ACBO) can overtake prior-driven methods by accumulating sufficient interventional evidence to overcome initial advantages from prior information.

Metric sensitivity. GAP and PA-GAP can produce different rankings for the same dataset and budget. GAP rewards the final best value and its discovery time, so methods that make a single large improvement early are favored. PA-GAP rewards steady trajectory-level progress, so methods that gradually improve throughout the budget can rank well even without the best final value. This discrepancy is most pronounced when comparing methods that plateau early (e.g., DCBO on Synthetic-2) versus methods that improve late (e.g., HCBO on Synthetic-2). We recommend that future CBO evaluations report both metrics, as well as trajectory plots, to avoid misleading conclusions from any single scalar summary.

Threats to validity. Several factors limit the generalizability of our benchmark findings. First, the benchmark uses unit-cost interventions, which does not capture realistic cost heterogeneity. Second, all hard-intervention datasets assume known graphs and causal sufficiency; results may differ under graph misspecification or hidden confounding. Third, we use released default hyperparameters for all methods; per-dataset tuning could change relative rankings. Fourth, the benchmark covers only single-objective settings (except for MO-CBO, which is excluded from the cross-method comparison). Finally, the set of compared methods is limited by the availability of compatible public implementations; fCBO, CNEI, GACBO and ACBO (for hard interventions) are not included in the hard intervention comparison.

5 Open Problems and Future Directions

The benchmark results and methodological analysis presented in this survey reveal that CBO is a maturing but incomplete field. No single method dominates across settings, and the value of causal structure depends critically on the alignment between a method's assumptions and the problem's actual causal complexity. In the following, we organize the most important open challenges into six concrete research directions.

5.1 Robustness to causal assumptions

Most CBO methods assume a correctly known graph, causal sufficiency (no hidden confounders), and identifiable interventional effects. In practice, each of these assumptions can fail.

Hidden confounding. When unobserved confounders are present, observational priors constructed via backdoor adjustment or the g-formula can be systematically biased, leading the acquisition function to favor interventions that appear promising under confounded associations but fail under true causal effects. Developing CBO methods that explicitly model sensitivity to hidden confounding, for example, by propagating bounds on the causal effect under varying degrees of unmeasured confounding (Pearl, 2009), is an important open direction.

Graph misspecification. Even when a graph is provided, edges may be missing, spurious, or incorrectly oriented. The impact of such errors on CBO performance has been largely unexplored. Future work should characterize how graph misspecification degrades scope reduction and prior quality and develop methods that are robust to bounded graph errors.

Prior misspecification. CNEI (Li et al., 2023) addresses noise robustness, but the broader problem of controlling how strongly observational information influences acquisition decisions under potential misspecification remains open. Methods that adaptively downweight observational priors as interventional data accumulate, analogous to Bayesian robustness techniques in other domains, would be valuable.

5.2 Scalability

Current CBO methods are mainly demonstrated on graphs with tens of variables and a handful of intervention scopes. Scaling to realistic systems requires advances on multiple fronts.

High-dimensional graphs. HCBO (Wu et al., 2024) addresses the selection of the scope in larger graphs, but the combinatorial explosion of candidate scopes remains a fundamental bottleneck. Future methods should take advantage of causal sparsity, hierarchical decomposition, or learned embeddings of the intervention scopes to avoid exhaustive enumeration.

Unknown-graph scalability. CEO (Branchini et al., 2023) and GACBO (Mukherjee et al., 2024) demonstrate unknown-graph CBO on small graphs, but maintaining a posterior or confidence set over graphs becomes intractable as the number of variables grows. Scalable alternatives might include local structure learning focused on the decision-relevant subgraph, amortized inference over graph posteriors, or graph neural network-based surrogates that learn causal structure implicitly.

Computational cost. Mechanism-level methods (MCBO, ACBO) require maintaining and updating multiple GPs and propagating uncertainty through the graph at each iteration. The per-iteration cost scales with the number of mechanism models and the depth of the graph. Sparse GP approximations, variational inference, or neural process surrogates could reduce this computational burden.

5.3 Richer intervention models

The current CBO literature mainly addresses either pure hard interventions or pure soft interventions. Real-world intervention settings are often more complex.

Mixed intervention types. Many practical systems involve a mix of hard interventions (e.g. turning a device on or off), soft interventions (e.g., adjusting a controller's parameters) and stochastic interventions (e.g., randomizing a treatment assignment probability). No existing CBO method handles arbitrary mixtures of these types within a single optimization loop.

Multi-agent and sequential interventions. ACBO (Sussex et al., 2024) considers adversarial environments, but the more general setting of multi-agent intervention, where multiple decision-makers intervene in the same system, possibly with conflicting objectives, remains unexplored. Similarly, extending CBO to sequential multi-stage intervention policies, where the intervention at each stage depends on the outcomes of previous stages, would bridge CBO and causal reinforcement learning.

Intervention cost heterogeneity. Our benchmark uses unit-cost interventions to isolate sample efficiency, but real interventions differ substantially in cost. The cost may depend on the target variable, the magnitude of the intervention, the safety requirements, or the logistics of implementation. Future CBO benchmarks should incorporate explicit cost models and future methods should optimize cost-adjusted efficiency rather than pure sample efficiency.

5.4 Realistic evaluation

Our benchmark study reveals that the evaluation design significantly affects conclusions about the quality of the method. Several aspects of the CBO evaluation require further development.

Standardized benchmarks. The fragmentation of the data set documented in Table 3 makes the comparison between papers unreliable. The community would benefit from a shared and versioned benchmark suite with standardized SCMs, intervention domains, observational datasets, and evaluation code, analogous to established benchmarks in reinforcement learning or supervised learning.

Evaluation under distribution shift. Most benchmarks assume that the test-time SCM matches the training-time SCM. In practice, the system can change between the observational data collection phase and the intervention deployment phase. Evaluating CBO under a controlled distribution shift would test the robustness of observational priors and structure assumptions.

Trajectory-aware metrics. Our analysis shows that GAP and PA-GAP can produce different rankings, highlighting the need for trajectory-aware evaluation. Future work should develop principled metrics that capture the full optimization trajectory, potentially incorporating cost, risk, and constraint satisfaction alongside objective improvement.

5.5 Theoretical foundations

The theoretical understanding of CBO lags behind its empirical development. Several foundational questions remain open.

Finite-sample regret bounds. MCBO (Sussex et al., 2023) provides regret bounds under mechanism-level GP models, and ACBO (Sussex et al., 2024) extends these to adversarial settings. However, regret bounds for effect-level CBO with observational priors, for unknown-graph CBO, and for constrained CBO are largely missing. Developing such bounds would clarify when causal structure provably improves over black-box BO and under what conditions.

Identifiability-aware acquisition. Current methods either assume full identifiability or ignore identifiability entirely. An intermediate approach would design acquisition functions that explicitly account for the degree of identifiability: using observational information aggressively when identification is strong and falling back to interventional exploration when it is weak. This would connect CBO to the literature on partial identification and sensitivity analysis in causal inference.

Sample complexity of scope reduction. The POMIS-based scope reduction in CBO eliminates dominated scopes using observational evidence, but the sample complexity of this elimination, i.e., how many observational samples are needed to reliably identify the correct POMIS, has not been analyzed. Such an analysis would clarify when scope reduction is reliable and when it may prematurely discard useful scopes.

5.6 Integration with modern representation learning

Recent advances in foundation models and causal representation learning suggest new directions for CBO.

Causal representation learning. Methods for learning causal variables and structure from high-dimensional observations (Schölkopf et al., 2021) could enable CBO in settings where the causal graph operates over latent variables rather than directly observed quantities. This would extend CBO to image- or text-based intervention settings.

Large language models for causal reasoning. Large language models (LLMs) have shown preliminary ability to reason about causal relationships from text descriptions. An intriguing direction is to use LLMs to elicit or refine causal graphs from domain experts, providing CBO with better structural priors without requiring formal causal discovery from data.

Neural surrogates. Replacing GP-based surrogates with neural network-based models (e.g., neural processes, transformers) could improve scalability and flexibility, particularly for high-dimensional mechanism functions or non-stationary environments. The challenge is maintaining the well-calibrated uncertainty quantification that makes GP-based CBO effective.

6 Conclusion

This survey has provided a comprehensive review of Causal Bayesian Optimization, organizing the growing literature through a unified design-space perspective that decomposes methods into four interacting axes: intervention representation, surrogate architecture, decision rule, and dominant uncertainty source. By analyzing each major CBO variant through this lens, we have shown that the field's diversity reflects systematic responses to specific bottlenecks of the original CBO framework (temporal dynamics, safety constraints, high dimensionality, unknown graphs, soft interventions, context dependence, and adversarial non-stationarity) rather than ad hoc extensions.

Our benchmark study reinforces a central finding: the value of causal structure in optimization is conditional. Causal methods provide clear advantages when their structural assumptions align with the problem, when scope reduction eliminates irrelevant interventions, when observational priors are well-calibrated, or when mechanism-level modeling captures the system's compositional structure. However, when these conditions are not met, strong non-causal BO baselines remain competitive, and method rankings depend substantially on the choice of dataset, budget, and evaluation metric. The introduction of PA-GAP as a trajectory-aware complement to GAP highlights that even the definition of "good performance" in CBO is not straightforward: final-value metrics and trajectory-level metrics can favor different methods.

Looking ahead, the most pressing challenges are not purely algorithmic. Robustness to violated causal assumptions, scalable structure learning integrated with optimization, realistic intervention cost models, and standardized evaluation infrastructure are all prerequisites for CBO to move from controlled benchmarks to real-world deployment. We hope that the unified perspective, benchmark, and structured open problems presented in this survey will help researchers navigate the CBO landscape and focus their efforts on the directions most likely to have a practical impact.

References

Virginia Aglietti, Xiaoyu Lu, Andrei Paleyes, and Javier González. Causal Bayesian optimization. In Silvia Chiappa and Roberto Calandra (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 3155–3164. PMLR, 26–28 Aug 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. Curran Associates, Inc., 2021.

Virginia Aglietti, Alan Malek, Ira Ktena, and Silvia Chiappa. Constrained causal Bayesian optimization. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 304–321. PMLR, 23–29 Jul 2023.

Vahan Arsenyan, Antoine Grosnit, Haitham Bou Ammar, and Arnak S. Dalalyan. Contextual causal Bayesian optimisation. In *International Conference on Learning Representations*. OpenReview.net, 2026. ICLR 2026 Poster.

Felix Berkenkamp, Angela P. Schoellig, and Andreas Krause. No-regret Bayesian optimization with unknown hyperparameters. *Journal of Machine Learning Research*, 24(124):1–50, 2023.

Shriya Bhatija, Paul-David Zuercher, Jakob Thumm, and Thomas Bohné. Multi-objective causal Bayesian optimization. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan

Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 4172–4195. PMLR, 13–19 Jul 2025.

Nicola Branchini, Virginia Aglietti, Neil Dhir, and Theodoros Damoulas. Causal entropy optimization. In Francisco Ruiz, Jennifer Dy, and Jan-Willem van de Meent (eds.), *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 8586–8605. PMLR, 25–27 Apr 2023.

Eric Brochu, Vlad M. Cora, and Nando de Freitas. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*, 2010.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10 (3):273–304, 1995.

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

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proceedings of the 34th International Conference on Machine Learning*, 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, Rod 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.

Alexander I. Cowen-Rivers, Wenlong Lyu, Rasul Tutunov, Zhi Wang, Antoine Grosnit, Ryan Rhys Griffiths, Alexandre Max Maraval, Hao Jianye, Jun Wang, Jan Peters, and Haitham Bou-Ammar. Hebo: Pushing the limits of sample-efficient hyper-parameter optimisation. *Journal of Artificial Intelligence Research*, 74: 1269–1349, 2022. doi: 10.1613/jair.1.13643.

Jean Durand, Yashas Annadani, Stefan Bauer, and Sonali Parbhoo. Causal Bayesian optimization with unknown graphs, 2025.

Frederick Eberhardt and Richard Scheines. Interventions and causal inference. *Philosophy of Science*, 74(5): 981–995, 2007. doi: 10.1086/525638.

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.

Adam Foster, Desi R. Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential Bayesian experimental design. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3384–3395. PMLR, 2021.

Peter I. Frazier. A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018. doi: 10.48550/arXiv.1807.02811.

Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997. doi: 10.1006/jcss.1997.1504.

Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023. doi: 10.1017/9781108348973.

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10:524, 2019. doi: 10.3389/fgene.2019.00524.

Limor Gultchin, Virginia Aglietti, Alexis Bellot, and Silvia Chiappa. Functional causal Bayesian optimization. In Robin J. Evans and Ilya Shpitser (eds.), *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pp. 756–765. PMLR, 31 Jul–04 Aug 2023.

William G. Havercroft and Vanessa Didelez. Simulating from marginal structural models with time-dependent confounding. *Statistics in Medicine*, 31(30):4190–4206, 2012. doi: 10.1002/sim.5472.

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

Momin Jamil and Xin-She Yang. A literature survey of benchmark functions for global optimisation problems. *International Journal of Mathematical Modelling and Numerical Optimisation*, 4(2):150–194, 2013.

Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998. doi: 10.1023/A:1008306431147.

Mina Konakovic Lukovic, Yunsheng Tian, and Wojciech Matusik. Diversity-guided multi-objective Bayesian optimization with batch evaluations. In *Advances in Neural Information Processing Systems*, volume 33, pp. 17708–17720. Curran Associates, Inc., 2020.

Harold J. Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, 86(1):97–106, 1964. doi: 10.1115/1.3653121.

Finnian Lattimore, Tor Lattimore, and Mark D. Reid. Causal bandits: Learning good interventions via causal inference. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Zheng Li, Yongxuan Li, Zongqiang Lian, and Rui Zheng. Improved causal Bayesian optimization algorithm with counter-noise acquisition function and supervised prior estimation. *Journal of Physics: Conference Series*, 2547(1):012017, 2023. doi: 10.1088/1742-6596/2547/1/012017.

Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994. doi: 10.1006/inco.1994.1009.

Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Causal bandits with unknown graph structure. In *Advances in Neural Information Processing Systems*, volume 34. Curran Associates, Inc., 2021.

Jonas Mockus. *Bayesian Approach to Global Optimization*. Springer, 1989. doi: 10.1007/978-94-009-0909-0.

Jonas Mockus, Vytautas Tiesis, and Antanas Zilinskas. The application of Bayesian methods for seeking the extremum. In *Towards Global Optimization 2*, pp. 117–129. Elsevier, 1978.

Sumantrak Mukherjee, Mengyan Zhang, Seth Flaxman, and Sebastian Josef Vollmer. Graph agnostic causal Bayesian optimisation, 2024.

Vineet Nair, Vishakha Patil, and Gaurav Sinha. Budgeted and non-budgeted causal bandits. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 2017–2025. PMLR, 2021.

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

Judea Pearl. The do-calculus revisited. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, 2012.

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

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Shaogang Ren, Zihao Wang, Yuzhou Chen, and Xiaoning Qian. Causal Bayesian optimization via exogenous distribution learning, 2026.

Karen Sachs, Omar Perez, Dana Pe'er, Douglas A. Lauffenburger, and Garry P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005. doi: 10.1126/science.1105809.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5): 612–634, 2021. doi: 10.1109/JPROC.2021.3058954.

Pier Giuseppe Sessa, Ilija Bogunovic, Maryam Kamgarpour, and Andreas Krause. No-regret learning in unknown games with correlated payoffs. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016. doi: 10.1109/JPROC.2015.2494218.

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

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.

Yanan Sui, Alkis Gotovos, Joel W. Burdick, and Andreas Krause. Safe exploration for optimization with Gaussian processes. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 997–1005. PMLR, 2015.

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

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

Clay Thompson and Yiu-Fai Yung. Causal graph analysis with the CAUSALGRAPH procedure. In *Proceedings of SAS Global Forum*, 2019. Paper SAS2998-2019.

Yupeng Wu, Weiye Wang, Yangwenhui Zhang, Mingjia Li, Yuanhao Liu, Hong Qian, and Aimin Zhou. High-dimensional causal Bayesian optimization. In *ECAI 2024 – 27th European Conference on Artificial Intelligence*, volume 392 of *Frontiers in Artificial Intelligence and Applications*, pp. 2990–2997. IOS Press, 2024. doi: 10.3233/FAIA240839.

A. G. Zhilinskas. Single-step bayesian search method for an extremum of functions of a single variable. *Cybernetics and Systems Analysis*, 11:160–166, 1975. doi: 10.1007/BF01069961.

A Technical Appendix

This appendix provides the complete supplementary material needed to reproduce, interpret, and extend the benchmark study presented in Section 4. Each sub-appendix is self-contained but cross-referenced to the relevant parts of the main text, and together they specify every detail required for independent replication.

- **Appendix A.1: Metrics analysis.** We formally analyze the value ranges of GAP and PA-GAP, state key properties of PA-GAP (trajectory monotonicity, budget dependence), and provide both a synthetic counterexample and an empirical case study from the Synthetic-2 benchmark showing how PA-GAP resolves a ranking bias present in GAP.

- **Appendix A.2: Implementation protocol.** We document all benchmark-controlled settings, wrapper design choices, software environment details, and trajectory normalization procedures used to ensure fair cross-method comparison.
- **Appendix A.3: Reference optima.** We list the reference optimum y^* for each data set, the offline computation protocol used to obtain it, and a discussion of how reference quality affects the reliability of the metric.
- **Appendix A.4: Summary of the consolidated dataset** Table 9 provides a single-table overview of all 13 benchmark datasets, including structural properties, intervention types, domains, and task directions.
- **Appendix A.5: Synthetic dataset SCMs.** Complete structural equations, noise distributions, intervention domains, and causal graph diagrams for ToyGraph, Synthetic, Synthetic-2 and Chain-hard.
- **Appendix A.6: Real dataset SCMs.** Complete structural equations, fitting procedures, and causal graphs for Ecology, Protein-reconstructed, Healthcare, and Epidemiology.
- **Appendix A.7: Soft-intervention datasets.** Function-network specifications and equations for Ackley, Rosenbrock, Dropwave, Alpine2, and the Chain-soft policy-intervention benchmark.

A.1 Metrics analysis

DCBO introduced the GAP metric as a scalar that combines how much an optimizer improves on its initial value with how early it reaches its best value (Aglietti et al., 2021). For a minimization problem, let $y(\mathbf{x}) = \mathbb{E}[Y \mid do(X = \mathbf{x})]$, let y^* denote the reference optimum (see Appendix A.3), let $\mathbf{x}_{\text{init}}$ be the initial best point, and let $\mathbf{x}_t^*$ be the best point found up to trial t. Let T be the total number of trials, and let t^* be the first trial at which the final best value $y(\mathbf{x}_T^*)$ is reached. GAP is

$$\text{GAP} = \left[\underbrace{\frac{y(\mathbf{x}_T^*) - y(\mathbf{x}_{\text{init}})}{y^* - y(\mathbf{x}_{\text{init}})}}_{\text{Improvement term}} + \underbrace{\frac{T - t^*}{T}}_{\text{Efficiency term}} \right] \Bigg/ \underbrace{\left(1 + \frac{T-1}{T}\right)}_{\text{Normalization}}. \tag{20}$$

A.1.1 GAP value range

When the global optimum is found in the first trial ($t^* = 1$), the efficiency term is equal to $\frac{T-1}{T}$, and normalization makes the maximum GAP equal to 1. However, DCBO sets $t^* = 0$ when no improvement is observed. In that case, the improvement term is 0, but the efficiency term becomes 1, which produces a nonzero GAP even without improvement. Specifically, GAP is then lower-bounded by

$$\frac{1}{1 + \frac{T-1}{T}} = \frac{T}{2T - 1}. \tag{21}$$

To correct for this, we set $t^* = T$ when no improvement occurs. Then both the improvement term and the efficiency term are 0, so the GAP ranges over $[0, 1]$ as intended.

A.1.2 Why PA-GAP can rank trajectories more consistently

GAP depends only on the final best value and the trial in which that final best value is first reached. This can bias the rankings when two optimization runs share the same early progress but differ in later improvements. Figure 7 illustrates this issue with a simple two-model example. Both models start from the same initial value $y(\mathbf{x}_{\text{init}}) = 10$ and reach an improved value $y(\mathbf{x}_t^*) = 5$ at the trial $t = 10$. Model 1 does not make further progress, while Model 2 later reaches the global optimum $y^* = 1$ at trial $t = 30$. With $T = 40$, GAP assigns Model 1 a higher score than Model 2, although Model 2 achieves the better final solution. This occurs because GAP adds an improvement term and an efficiency term and then evaluates only the final best value and its discovery time, rather than the full optimization path.

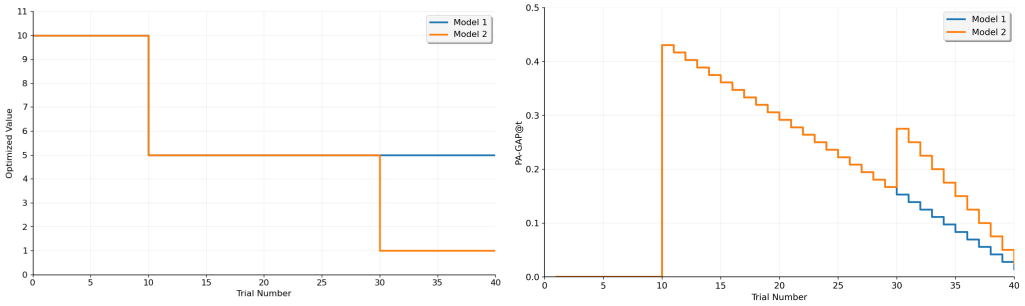

Figure 7: Demo comparison illustrating the ranking bias of GAP. The left panel shows best-so-far trajectories, while the right panel shows the corresponding per-trial PA-GAP (PA-GAP@t) contributions. The two models share the same early improvement, but Model 2 later reaches the global optimum. GAP can favor the earlier but worse final trajectory, while PA-GAP assigns credit to later improvement by averaging weighted best-so-far progress across the full path.

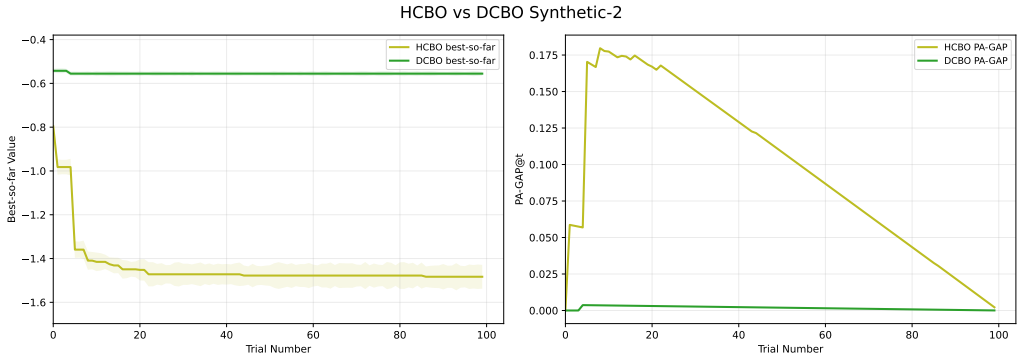

Figure 8: HCBO and DCBO on Synthetic-2. The left panel shows best-so-far trajectories, while the right panel shows the corresponding per-trial PA-GAP (PA-GAP@t) contributions. DCBO reaches its own best-so-far value early, which gives it a higher GAP score, but HCBO reaches better objective values over the trajectory and receives the higher PA-GAP score.

To better reflect the full optimization process, PA-GAP computes a weighted improvement value at each trial and then averages these values over the trajectory. For minimization tasks, the best-so-far improvement ratio in the trial t is

$$R_t = \min\left\{\frac{y(\mathbf{x}_{\text{init}}) - y(\mathbf{x}_t^*)}{y(\mathbf{x}_{\text{init}}) - y^*}, 1\right\}, \tag{22}$$

and for maximization tasks, the numerator and denominator are reversed:

$$R_t = \min\left\{\frac{y(\mathbf{x}_t^*) - y(\mathbf{x}_{\text{init}})}{y^* - y(\mathbf{x}_{\text{init}})}, 1\right\}. \tag{23}$$

The clipping prevents numerical artifacts when an observed best value exceeds the offline reference optimum. The PA-GAP score used in our benchmark is then

$$\text{PA-GAP} = \frac{1}{T}\sum_{t=1}^{T}\left(\underbrace{R_t}_{\text{Improvement ratio}} \cdot \underbrace{\frac{T-(t-1)}{T}}_{\text{Efficiency weight}}\right). \tag{24}$$

Here T denotes the number of interventional trials, excluding the initial value in the best-so far trajectory.

Table 6: Formal comparison of GAP and PA-GAP evaluation properties.

Property	GAP	PA-GAP
What is evaluated	Final best value + discovery time	Full best-so-far trajectory
Value range	$[0, 1]$ (after correction)	$[0, \frac{T+1}{2T}]$
Normalization	Unit-normalized	Raw; budget-dependent upper bound
Sensitivity to late improvement	Low: only the time of the *final* best matters	High: all improvements contribute
Sensitivity to budget T	Moderate: efficiency term scales with T	Strong: values not comparable across different T
Plateau handling	Rewards early plateau at good value	Penalizes early plateau if further improvement is possible
Computational cost	$\mathcal{O}(T)$	$\mathcal{O}(T)$

This definition is not normalized to have a maximum value 1. Since $R_t \in [0, 1]$ and the efficiency weight decreases linearly from 1 at the first trial to $1/T$ at the last trial, the raw PA-GAP range is

$$0 \leq \text{PA-GAP} \leq \frac{1}{T} \sum_{t=1}^{T} \frac{T-(t-1)}{T} = \frac{T+1}{2T}. \tag{25}$$

The upper bound is attained when the global optimum is found in the first trial and remains the best value so far for the rest of the budget. Thus, PA-GAP should be interpreted as a raw trajectory-weighted efficiency score, not as a unit-normalized score. Larger values indicate faster and stronger best-so-far progress, but raw PA-GAP values should be compared directly only within the same trial budget.

Figure 7 shows why this trajectory-aware averaging can produce a more intuitive ranking than GAP: Model 2 receives additional credit for its late improvement instead of being judged only by the delayed discovery time of its final best value. Figure 8 shows the same issue in the Synthetic-2 benchmark by comparing HCBO and DCBO. DCBO quickly reaches a stable best-so-far value, so GAP rewards its early discovery and assigns it a higher score ($0.485 \pm .012$) than HCBO ($0.174 \pm .071$). However, this early DCBO value is much worse than the values eventually reached by HCBO. PA-GAP reverses the ranking: DCBO receives only $0.002 \pm .001$, while HCBO receives $0.099 \pm .054$. This better reflects the fact that HCBO makes more meaningful progress toward the reference optimum over the optimization path, even if its improvement is not captured well by GAP's final-discovery-time summary.

Summary and formal properties. GAP is simple and interpretable, but can produce counterintuitive rankings when methods differ primarily in their late-stage behavior or trajectory shape. PA-GAP complements GAP by rewarding sustained progress throughout the budget. Table 6 contrasts the two metrics along key evaluation dimensions.

We state two key properties of PA-GAP that justify its use as a trajectory-aware complement to GAP.

Lemma 1 (Trajectory monotonicity of PA-GAP)**.** *Let $\{R_t\}_{t=1}^T$ and $\{\tilde{R}_t\}_{t=1}^T$ be two sequences of the best improvement ratios so far satisfying $R_t \geq \tilde{R}_t$ for all $t \in \{1, \ldots, T\}$. Then* PA-GAP$(\{R_t\}) \geq$ PA-GAP$(\{\tilde{R}_t\})$*, with equality if and only if $R_t = \tilde{R}_t$ for all t.*

Proof. Since the efficiency weights $w_t = \frac{T-(t-1)}{T} > 0$ for all t, the PA-GAP functional $\frac{1}{T}\sum_{t=1}^{T} R_t \cdot w_t$ is a positively weighted average of the improvement ratios. The conclusion follows from the strict positivity of the weights. □

This property ensures that if one trajectory Pareto-dominates another at every trial, PA-GAP reflects this dominance. GAP does not satisfy this property in general: a trajectory with uniformly higher R_t can receive a lower GAP score if its final best is reached later.

Table 7: Implementation settings used in the benchmark. Method-specific defaults are preserved unless a wrapper override is needed for shared evaluation.

Method	Benchmark-controlled settings	Method-specific notes
MCBO	Trial budget and seed fixed by wrapper; noise scale set to 0; $\beta = 10$.	Uses GP networks with MC acquisition. Acquisition optimization follows the released implementation defaults, including restart and raw-sample rules.
ACBO	Trial budget and seed fixed by wrapper; noise scale set to 0; $\beta = 10$; default discrete action cardinality 4.	Used only for soft-intervention function-network tasks. Runs CBO-MW with multiplicative-weights updates.
cCBO	Trial budget fixed by wrapper; dataset configs determine observational samples, intervention sets, and constraints.	Uses the single-task cCBO setting in our benchmark. Protein uses an internal offset so that the exported trajectory contains the intended number of evaluated trials.
DCBO	Trial budget and seed fixed by wrapper; dataset YAML files specify the SCM and intervention domain.	Benchmark uses a single-time setting to make DCBO comparable with static hard-intervention tasks.
CEO	Trial budget and seed fixed by wrapper; small candidate graph and anchor-point settings follow the benchmark runner.	Uses the benchmark-native CEO runner by default. The original vendored CEO implementation can be enabled separately.
CoCaBO	Trial budget, seed, and objective direction fixed by wrapper.	Uses UCB over active policy scopes and HEBO-style inner optimization. Benchmark configs do not include additional contextual variables unless specified.
HCBO	Seed and trial settings fixed by wrapper when supported.	Used for compatible hard-intervention settings. Acquisition and initialization constants follow the released HCBO-style implementation.

Lemma 2 (Budget dependence of PA-GAP)**.** *The maximum attainable PA-GAP value is* $\frac{T+1}{2T}$*, which is strictly decreasing in* T *and approaches* $\frac{1}{2}$ *as* $T \to \infty$*. Consequently, raw PA-GAP values are comparable only within the same budget* T*.*

Proof. The maximum is achieved when $R_t = 1$ for all t, giving PA-GAP $= \frac{1}{T}\sum_{t=1}^{T} \frac{T-(t-1)}{T} = \frac{T+1}{2T}$. Since $\frac{d}{dT}\left(\frac{T+1}{2T}\right) = -\frac{1}{2T^2} < 0$, the upper bound is strictly decreasing in T. □

Practical recommendation. We recommend reporting both metrics alongside trajectory plots to obtain a balanced view of optimizer performance. GAP is appropriate for identifying methods that quickly find a strong final solution, while PA-GAP is appropriate for evaluating whether a method makes sustained, meaningful progress throughout the available budget.

A.2 Implementation protocol

Reproducibility philosophy. All methods run with their default hyperparameters released, unless noted in Table 7. We deliberately avoid per-dataset hyperparameter tuning to ensure that the benchmark reflects each method's out-of-the-box behavior. This choice prioritizes reproducibility and fairness over per-method optimization, but it means that individual methods may underperform relative to what a practitioner could achieve with careful tuning. The trade-off is intentional: our goal is to compare algorithmic design choices, not implementation engineering.

Software environment. All experiments are conducted in Python 3.10 using PyTorch 2.0 and GPyTorch 1.11 as the primary GP backend for methods that depend on Gaussian process surrogates. Each CBO method is run in its own isolated virtual environment to prevent dependency conflicts. The benchmark evaluation pipeline–including trajectory export, GAP/PA-GAP computation, and figure generation–is

implemented in a shared codebase that is independent of any individual method's software stack. This separation ensures that scoring differences reflect algorithmic behavior rather than library version effects. The exact package versions for each method are documented in the benchmark repository.

Computational resources. All experiments were conducted on a single-GPU CUDA workstation equipped with one NVIDIA A100 GPU with 40 GiB of VRAM, Intel Xeon Platinum 8481C CPUs, and 754 GiB of system memory. Most methods complete 100 trials on a single dataset within 5–30 minutes. The main exceptions are CEO, which incurs additional overhead from graph posterior sampling, and MCBO, which propagates uncertainty through multiple mechanism-level GPs. GPU acceleration is used when supported by the corresponding implementation; otherwise, experiments are executed on the CPU.

Wrapper design. Wrapper-level overrides are limited to shared experimental controls: trial budget, random seed, dataset mapping, objective direction (minimization or maximization), and trajectory export format. The Method-internal choices—including the selection of the GP kernel, the acquisition optimization strategy, the restart schedules, the scope enumeration, and the internal noise models—are left at their defaults released. Each wrapper converts the method's internal output into the common trajectory format described below. Wrappers do not modify the optimization loop itself; they only control initialization and output extraction.

Trajectory normalization. To make results comparable across heterogeneous codebases, every method's output is converted to a common trajectory format with two fields: trial number and best-so-far objective value. For minimization tasks, this value is the running minimum observed up to the trial t; for maximization tasks, the running maximum. The trajectory includes an initial best-of-so-far value (before any interventional trial), followed by one entry per interventional trial. Thus, a budget of B interventions produces a trajectory of $B + 1$ entries. GAP and PA-GAP are then computed from these exported trajectories using a single evaluation script shared by all methods, ensuring that scoring differences reflect algorithmic behavior rather than implementation artifacts. This design also makes it easy to add new methods or metrics without modifying existing implementations.

A.3 Reference optima

The reference optimum y^* used by GAP and PA-GAP is computed offline from the SCM benchmark, function generator or oracle interventional data, as detailed in Table 8. It is used *only* for scoring and is never provided to any optimizer during a run. At evaluation time, all methods on a given data set are compared against the same fixed reference value using the same task direction.

Reliability categories. The reference optima fall into two reliability classes. For datasets with analytically known optima (Dropwave, Alpine2, Ackley, Rosenbrock), the reference is exact and introduces no scoring uncertainty. For SCM-based datasets, the reference is obtained by dense grid search or large-sample uniform random search over the admissible intervention domain. In these cases, y^* represents the *best known achievable outcome*, not necessarily the true global optimum. This distinction is important for datasets with complex, potentially multimodal objective landscapes: if the reference underestimates the true optimum, both GAP and PA-GAP will overestimate the improvement ratios R_t, potentially making all methods appear closer to optimal than they actually are. In the opposite direction, an overestimated reference (which cannot occur by construction in our protocol, since we use oracle SCM access) would inflate the denominator and deflate all scores. To mitigate this risk, we use high-budget offline search (10,000 samples for Ecology and Epidemiology, 200×200 grids for Protein-reconstructed, and 1,000-point grids for Synthetic-2) and cross-validate against known interventional datasets where available.

Impact on metric comparisons. Because both GAP and PA-GAP normalize the improvement relative to $y^* - y(\mathbf{x}_{\text{init}})$, the quality of y^* affects the absolute metric values, but does not affect *the relative rankings* between methods in the same dataset, provided that all methods are scored against the same reference. Cross data set comparisons of absolute GAP or PA-GAP values should therefore be made with caution, as differences in reference quality across datasets can introduce systematic biases.

Table 8: Reference optima used for GAP and PA-GAP scoring. Analytic optima are exact; grid/random-search references represent the best known achievable value.

Dataset	Task	y^*	Offline computation protocol
Dropwave	Max	1.0000	Analytic optimum from the known function generator.
Alpine2	Max	400.0000	Analytic optimum from the known function generator.
Ackley	Max	0.0000	Analytic optimum of the negated Ackley objective.
Rosenbrock	Max	0.0000	Analytic optimum of the negated Rosenbrock objective.
Chain-soft	Min	-3.3936	Reference value from the Chain-soft function generator.
Protein-recon.	Min	-240.0234	Reconstructed linear SEM; 200×200 grid over PKC and PKA; minimum Erk.
Synthetic-2	Min	-3.9690	Best offline oracle outcome; 1000-point grids on X and Z.
Synthetic	Min	-4.4474	Best value across BO and CBO oracle interventional datasets.
ToyGraph	Min	-2.5718	Best value across BO and CBO oracle interventional datasets.
Chain-hard	Min	-15.8060	Best value across BO and CBO oracle interventional datasets.
Ecology	Max	9.2652	10,000-sample uniform search over the manipulative-variable domain.
Epidemiology	Min	-3.3906	10,000-sample uniform search over the manipulative-variable domain.
Healthcare	Min	4.0401	Reference value from the distributed real-data benchmark artifact.

Table 9: Consolidated summary of all thirteen benchmark datasets. For each dataset, we list the category, number of observed nodes, manipulable variables, intervention type, admissible domain per variable, optimization direction, and source reference. Latent variables (if any) are noted in parentheses.

Dataset	Category	#Nodes	Manipulable vars	Intervention type	Domain	Task	Source
ToyGraph	Synthetic (hard)	3	X, Z	Hard	$X \in [-5, 5]$; $Z \in [-5, 20]$	Min Y	Aglietti et al. (2020)
Synthetic	Synthetic (hard)	7 (+2 latent)	B, D, E	Hard	$B \in [-5, 4]$; $D \in [-5, 5]$; $E \in [-6, 3]$	Min Y	Aglietti et al. (2020)
Synthetic-2	Synthetic (hard)	3	X, Z	Hard	$X \in [-3, 2]$; $Z \in [-1, 1]$	Min Y	Aglietti et al. (2023)
Chain-hard	Synthetic (hard)	4	W, Z	Hard	$[-1, 1]$ each	Min Y	Gultchin et al. (2023)
Ecology	Real/fitted (hard)	11	N, O, C, T, D	Hard	$N \in [-2, 5]$; $O \in [2, 4]$; $C \in [0, 1]$; $T \in [2200, 2500]$; $D \in [1950, 2100]$	Max Y	Aglietti et al. (2020)
Protein-recon.	Real/fitted (hard)	8	PKC, PKA, Mek, Akt	Hard	PKC$\in [0.5, 106.5]$; PKA$\in [1.45, 4491.5]$; Mek$\in [0.5, 389.5]$; Akt$\in [1.2, 3555.5]$	Min Erk	Aglietti et al. (2023)
Healthcare	Real/fitted (hard)	6	Aspirin, Statin	Hard	$[0, 1]$ each	Min PSA	Aglietti et al. (2020)
Epidemiology	Real/fitted (hard)	5	T, R	Hard	$T \in [0.48, 801.79]$; $R \in [-0.99, 1.0]$	Min Y	Branchini et al. (2023)
Ackley	Synthetic (soft)	9	$a_0, \ldots, a_5$	Soft (function network)	$[-2, 2]$ each	Max Y	Sussex et al. (2023)
Rosenbrock	Synthetic (soft)	$D+D-1$	$a_0, \ldots, a_{D-1}$	Soft (function network)	$[-2, 2]$ each	Max Y	Sussex et al. (2023)
Dropwave	Synthetic (soft)	4	a_0, a_1	Soft (function network)	$[-5.12, 5.12]$ each	Max Y	Sussex et al. (2023)
Alpine2	Synthetic (soft)	12	$a_0, \ldots, a_5$	Soft (function network)	$[0, 10]$ each	Max Y	Sussex et al. (2023)
Chain-soft	Synthetic (soft)	4	W (hard), Z (policy)	Mixed hard/soft	$W \in [-1, 1]$; $Z = \pi(X)$ with coeffs. $\in [-0.27, 0.27]$	Min Y	Gultchin et al. (2023)

A.4 Consolidated dataset summary

Table 9 provides a unified overview of all 13 benchmark datasets, consolidating key structural and experimental properties in a single reference. This table complements the SCM specifications per-dataset in Appendices A.5–A.7 and the usage-coverage table (Table 3) in the main text. The column "#Nodes" counts endogenous (observed) variables only; latent variables are listed separately where applicable. The column "Intervention domain" specifies the admissible range per manipulable variable as used in the benchmark.

A.5 Synthetic dataset SCMs

This section provides the full structural equations, noise distributions, intervention domains, and causal graph diagrams for the four synthetic hard-intervention datasets used in our benchmark. For each data set, we follow a standardized format: (i) a brief description including the source reference, the number of nodes, and the direction of the task; (ii) the exogenous variables and their distributions; (iii) the endogenous structural equations; (iv) the noise model with explicit distributional parameters; (v) the intervention domain; and (vi) a figure of the causal graph with a consistent color scheme (green = manipulable, red = target, light-gray = non-manipulable, dark-gray = latent).

A.5.1 ToyGraph

A minimal three-node chain-structured SCM introduced by Aglietti et al. (2020), consisting of variables X, Z, and Y. Both X and Z are manipulable (intervenable); Y is the target. The task is to minimize Y. Figure 9 shows the causal graph.

Exogenous variable.

$$X \sim P_X \quad \text{(input variable, exogenous)}$$

Endogenous variables.

$$Z = e^{-X} + \varepsilon_Z,$$
$$Y = \cos Z - e^{-Z/20} + \varepsilon_Y.$$

Noise model. All noise terms $\varepsilon_Z, \varepsilon_Y$ are mutually independent. In the original CBO implementation, these are drawn from $\mathcal{N}(0, \sigma^2)$ with a small variance σ^2.

Intervention domain. Hard interventions on X and Z use dataset-specific domains:

$$X \in [-5, 5], \qquad Z \in [-5, 20].$$

The admissible scopes are $\{X\}$, $\{Z\}$, and $\{X, Z\}$.

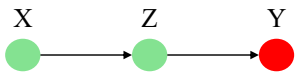

Figure 9: ToyGraph SCM. Green nodes represent manipulable variables; the red node is the target variable Y.

A.5.2 Synthetic

A seven-node SCM with two latent variables, introduced by Aglietti et al. (2020). The observed nodes are A, B, C, D, E, F, Y, with latent confounders U_1 and U_2. Among the observed nodes, B, D, and E are manipulable; A, C, and F are not manipulable; and Y is the target. The task is to minimize Y. Figure 10 shows the causal graph.

Latent variables.

$$U_1 = \varepsilon_{YA} \sim \mathcal{N}(0, 1),$$
$$U_2 = \varepsilon_{YB} \sim \mathcal{N}(0, 1).$$

Exogenous variable.

$$F = \varepsilon_F \sim \mathcal{N}(0, 1).$$

Endogenous variables.

$$\begin{aligned}
A &= F^2 + U_1 + \varepsilon_A, \\
B &= U_2 + \varepsilon_B, \\
C &= e^{-B} + \varepsilon_C, \\
D &= \frac{e^{-C}}{10} + \varepsilon_D, \\
E &= \cos A + \frac{C}{10} + \varepsilon_E, \\
Y &= \cos D + \sin E + U_1 + U_2\, \varepsilon_Y.
\end{aligned}$$

Noise model. Each $\varepsilon_\bullet$ is an independent noise term. In the original CBO implementation, the endogenous noise terms are drawn from $\mathcal{N}(0, \sigma^2)$ with small variance σ^2.

Intervention domain. Hard interventions on B, D, and E use dataset-specific domains:

$$B \in [-5, 4], \qquad D \in [-5, 5], \qquad E \in [-6, 3].$$

The admissible scopes include all non-empty subsets of $\{B, D, E\}$ considered by the benchmark intervention-set construction.

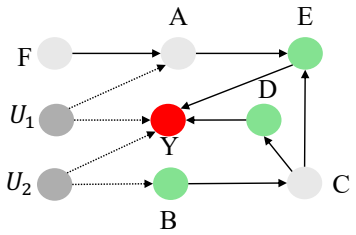

Figure 10: Synthetic SCM. Light-gray nodes are non-manipulable variables; dark-gray nodes are latent confounders; green nodes are manipulable variables; the red node is the target variable Y.

A.5.3 Synthetic-2

A three-node SCM introduced in the cCBO line of work (Aglietti et al., 2023), consisting of variables X, Z, and Y. Both X and Z are manipulable; Y is the target. The task is to minimize Y. The structure is similar to ToyGraph but uses explicitly specified Gaussian noise. Figure 11 shows the causal graph.

Exogenous variable.

$$X = \varepsilon_X, \qquad \varepsilon_X \sim \mathcal{N}(0, 1).$$

Endogenous variables.

$$\begin{aligned}
Z &= e^{-X} + \varepsilon_Z, \\
Y &= \cos Z - e^{-Z/20} + \varepsilon_Y.
\end{aligned}$$

Noise model. $\varepsilon_X \sim \mathcal{N}(0,1)$, $\varepsilon_Z \sim \mathcal{N}(0,1)$, and $\varepsilon_Y \sim \mathcal{N}(0,1)$. All noise terms are mutually independent.

Intervention domain. Hard interventions on X and Z are each drawn from $[-3,3]$. The admissible scopes are $\{X\}$, $\{Z\}$, and $\{X,Z\}$.

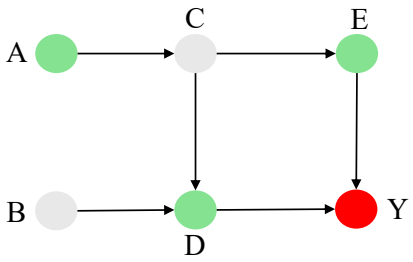

Figure 11: Synthetic-2 SCM. Green nodes represent manipulable variables; the red node is the target variable Y.

A.5.4 Chain-hard

A four-node SCM introduced by Gultchin et al. (2023) as the hard-intervention version of the Chain benchmark. The observed variables are X, W, Z, and Y. Among them, W and Z are manipulable; X is a non-manipulable context variable; and Y is the target. The task is to minimize Y. Figure 12 illustrates the causal graph.

Structural equations.

$$\begin{aligned} X &= U_X, \\ W &= U_W, \\ Z &= -0.5X + U_Z, \\ Y &= -W - 3ZX + U_Y. \end{aligned}$$

Noise model. $U_X, U_W, U_Z, U_Y \sim \mathcal{N}(0,1)$, mutually independent.

Intervention domain. Hard interventions in W and Z are each restricted to $[-1,1]$. A hard intervention replaces the corresponding structural equation with a fixed constant. The admissible scopes are $\{W\}$, $\{Z\}$, and $\{W,Z\}$.

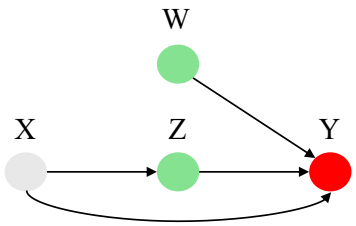

Figure 12: Chain-hard SCM. Light-gray nodes are non-manipulable variables; green nodes are manipulable variables; the red node is the target variable Y.

A.6 Real dataset SCMs

This section provides the full structural equations, noise distributions, fitting details, and intervention domains for the four real or fitted-SCM hard-intervention datasets. Unlike the synthetic datasets above, these

SCMs are either fitted to real observational data or derived from domain-specific causal models. We follow the same standardized format as in the Appendix A.5. For protein reconstruction, we additionally document the fitting procedure to ensure transparent reproducibility.

A.6.1 Ecology

An 11-node SCM based on the Bermuda reef calcification model of Courtney et al. (2017), introduced in CBO (Aglietti et al., 2020). The manipulable variables are Nut, Chlα, TA, DIC, and Ω_A. The non-manipulable variables are Tem, Sal, P_{co_2}, Light, and pHsw. The target is NEC (net ecosystem calcification). The task is to *maximize* NEC. Figure 13 shows the causal graph.

Exogenous variables.

$$\begin{aligned}
\text{Tem} &= U_{\text{Tem}} \sim \mathcal{N}(24.184130,\ 3.220405^2),\\
\text{Sal} &= U_{\text{Sal}} \sim \mathcal{N}(36.591624,\ 0.149197^2),\\
\text{Nut} &= U_{\text{Nut}} \sim \mathcal{N}(0.492065,\ 1.592408^2),\\
\text{TA} &= U_{\text{TA}} \sim \mathcal{N}(2357.893696,\ 27.609355^2).
\end{aligned}$$

Endogenous variables.

$$\begin{aligned}
P_{co_2} &= 18.798174 + 15.797384\,\text{Tem} + U_{P_{co_2}},\\
\text{Chl}\alpha &= 0.373420 - 0.002400\,\text{Nut} + U_{\text{Chl}\alpha},\\
\text{Light} &= 6665.081996 - 10737.462582\,\text{Chl}\alpha + U_{\text{Light}},\\
\text{pHsw} &= 8.427706 - 0.000966\,P_{co_2} + U_{\text{pHsw}},\\
\text{DIC} &= 2131.672107 - 0.216560\,P_{co_2} + U_{\text{DIC}},\\
\Omega_A &= 3.245248 + 0.094332\,\text{Tem} + 0.006754\,\text{Sal} - 0.005737\,P_{co_2} + U_{\Omega_A},\\
\text{NEC} &= 211.422555 - 0.000030\,\text{Light} + 0.016680\,\text{Nut} - 25.719277\,\text{pHsw} - 0.500403\,\Omega_A + U_{\text{NEC}}.
\end{aligned}$$

Noise model. $U_{P_{co_2}} \sim \mathcal{N}(0, 28.896639^2)$, $U_{\text{Chl}\alpha} \sim \mathcal{N}(0, 0.039295^2)$, $U_{\text{Light}} \sim \mathcal{N}(0, 1546.913034^2)$, $U_{\text{pHsw}} \sim \mathcal{N}(0, 0.005258^2)$, $U_{\text{DIC}} \sim \mathcal{N}(0, 18.412100^2)$, $U_{\Omega_A} \sim \mathcal{N}(0, 0.036958^2)$, and $U_{\text{NEC}} \sim \mathcal{N}(0, 1.073340^2)$. All noise terms are mutually independent.

Intervention domain. Hard interventions in Nut, Chlα, TA, DIC, and Ω_A are applied in domain-specific ranges corresponding to the natural variability reported in Courtney et al. (2017). The benchmark uses the intervention domains specified in the public CBO repository.

A.6.2 Protein-reconstructed

An eight-node SCM reconstructed from protein-signaling data from Sachs et al. (2005), following the DAG used in cCBO (Aglietti et al., 2023). The manipulable variables are PKC, PKA, Mek, and Akt. The non-manipulable variables are Raf, P38, and Jnk. The target is Erk (extracellular signal-regulated kinase). The task is to minimize Erk. Figure 14 shows the causal graph.

Fitting procedure. This is a transparent reconstruction, not an exact reproduction of the unreleased fitted SCM from the original cCBO experiments. We retain the connected subgraph used in cCBO and fit *linear* structural mechanisms using ordinary least squares on 852 observational samples from Sachs et al. (2005). The parent sets are determined by the cCBO DAG. The linear model class was chosen for reproducibility and interpretability; nonlinear alternatives (e.g. kernel regression) were not explored.

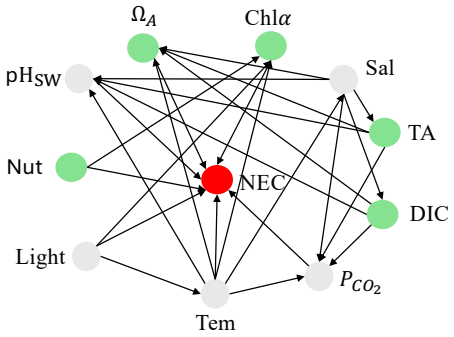

Figure 13: Ecology SCM. Light-gray nodes are non-manipulable variables; green nodes are manipulable variables; the red node is the target variable NEC.

Exogenous variable.

$$\mathrm{PKC} = U_{\mathrm{PKC}} \sim \mathrm{Uniform}(1,\, 106).$$

Endogenous variables.

$$\begin{aligned}
\mathrm{PKA} &= 554.390731 + 0.841153\,\mathrm{PKC} + U_{\mathrm{PKA}},\\
\mathrm{Raf} &= 62.199046 - 0.177745\,\mathrm{PKC} - 0.000379\,\mathrm{PKA} + U_{\mathrm{Raf}},\\
\mathrm{Mek} &= -1.090275 + 0.039662\,\mathrm{PKC} - 0.000652\,\mathrm{PKA} + 0.520845\,\mathrm{Raf} + U_{\mathrm{Mek}},\\
\mathrm{P38} &= 15.144328 + 1.234783\,\mathrm{PKC} + 0.000591\,\mathrm{PKA} + U_{\mathrm{P38}},\\
\mathrm{Jnk} &= 52.953603 - 0.764801\,\mathrm{PKC} - 0.005306\,\mathrm{PKA} + U_{\mathrm{Jnk}},\\
\mathrm{Akt} &= -31.110747 + 0.128905\,\mathrm{PKA} + U_{\mathrm{Akt}},\\
\mathrm{Erk} &= -23.248743 + 0.081707\,\mathrm{PKA} - 0.029886\,\mathrm{Mek} + U_{\mathrm{Erk}}.
\end{aligned}$$

Noise model. $U_{\mathrm{PKA}} \sim \mathcal{N}(0,\, 427.437996^2)$, $U_{\mathrm{Raf}} \sim \mathcal{N}(0,\, 41.768782^2)$, $U_{\mathrm{Mek}} \sim \mathcal{N}(0,\, 16.695865^2)$, $U_{\mathrm{P38}} \sim \mathcal{N}(0,\, 13.111164^2)$, $U_{\mathrm{Jnk}} \sim \mathcal{N}(0,\, 42.074715^2)$, $U_{\mathrm{Akt}} \sim \mathcal{N}(0,\, 113.958504^2)$, and $U_{\mathrm{Erk}} \sim \mathcal{N}(0,\, 82.760198^2)$. All noise terms are mutually independent.

Intervention domain and constraints. Hard interventions in PKC and PKA are drawn from the range $[0.5, 106.5]$ and $[1.45, 4491.5]$, respectively. The interventions in Mek and Akt use the domains observed in the Sachs data, which are $[0.5, 389.5]$ and $[1.2, 3555.5]$. Biological plausibility constraints on PKC and PKA are enforced by clamping the proposed values to their observed ranges.

A.6.3 Healthcare

A six-node SCM derived from a real-world clinical setting (Thompson & Yung, 2019), involving Age, BMI, Aspirin, Statin, Cancer, and PSA. The manipulable variables are Aspirin and Statin. Non-manipulable variables are Age, BMI, and Cancer. The target is PSA (prostate-specific antigen). The task is to minimize PSA. Figure 15 shows the causal graph.

Exogenous variable.

$$\mathrm{Age} \sim \mathrm{Uniform}(55,\, 75).$$

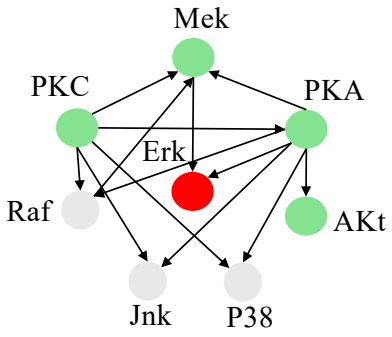

Figure 14: Protein-signaling SCM (reconstructed). Light-gray nodes are non-manipulable variables; green nodes are manipulable variables; the red node is the target variable Erk.

Endogenous variables.

$$
\begin{aligned}
\text{BMI} &= 27.0 - 0.01\,\text{Age} + \varepsilon_{\text{BMI}},\\
\text{Aspirin} &= \sigma\big(-8.0 + 0.10\,\text{Age} + 0.03\,\text{BMI}\big) + \varepsilon_{\text{Aspirin}},\\
\text{Statin} &= \sigma\big(-13.0 + 0.10\,\text{Age} + 0.20\,\text{BMI}\big) + \varepsilon_{\text{Statin}},\\
\text{Cancer} &= \sigma\big(2.2 - 0.05\,\text{Age} + 0.01\,\text{BMI} - 0.04\,\text{Statin} + 0.02\,\text{Aspirin}\big) + \varepsilon_{\text{Cancer}},\\
\text{PSA} &= 6.8 + 0.04\,\text{Age} - 0.15\,\text{BMI} - 0.60\,\text{Statin} + 0.55\,\text{Aspirin} + 1.00\,\text{Cancer} + \varepsilon_{\text{PSA}},
\end{aligned}
$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ denotes the sigmoid function.

Noise model. $\varepsilon_{\text{BMI}} \sim \mathcal{N}(0,\, 0.7^2)$, $\varepsilon_{\text{PSA}} \sim \mathcal{N}(0,\, 0.4^2)$. Each remaining $\varepsilon_\bullet$ is an independent noise term with small variance.

Intervention domain. Although the original data template specifies Aspirin and Statin as binary variables, existing CBO implementations commonly relax them to continuous values in $(0, 1)$ for BO-style optimization. We follow this convention to remain comparable to the prior work. The admissible scopes are $\{\text{Aspirin}\}$, $\{\text{Statin}\}$, and $\{\text{Aspirin}, \text{Statin}\}$.

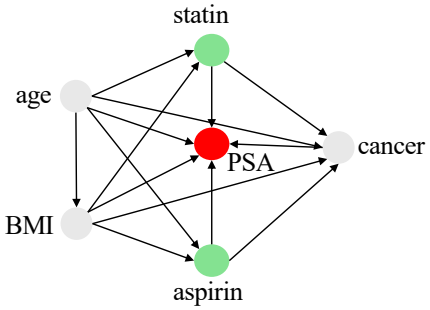

Figure 15: Healthcare SCM. Light-gray nodes are non-manipulable variables; green nodes are manipulable variables; the red node is the target variable PSA.

A.6.4 Epidemiology

A five-node SCM introduced in CEO (Branchini et al., 2023), modeling an HIV treatment scenario based on Havercroft & Didelez (2012). The observed variables are B, T, L, R, and Y. The manipulable variables

are T (treatment dose) and R (second treatment). The non-manipulable variables are B (baseline) and L (intermediate health indicator). The target is Y (viral load). The task is to minimize Y. Figure 16 shows the causal graph.

Exogenous variables.

$$B \sim \text{Uniform}(-1,\, 1),$$
$$T \sim \text{Uniform}(4,\, 8).$$

Endogenous variables.

$$L = \exp(0.5\,T + U), \qquad U \sim \mathcal{N}(0,1),$$
$$R = 4 + L\,T,$$
$$Y = 0.5 + \cos(4T) + \sin(-L + 2R) + B + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0,1).$$

Noise model. $U \sim \mathcal{N}(0,1)$ and $\varepsilon \sim \mathcal{N}(0,1)$. All noise terms are mutually independent.

Intervention domain. Hard interventions in T are drawn from $[4, 8]$ and in R from $[4, 8]$. The admissible scopes are $\{T\}$, $\{R\}$, and $\{T, R\}$. Since CEO uses a modified HIV SCM, more specific clinical interpretations of B and L are not provided in the original article.

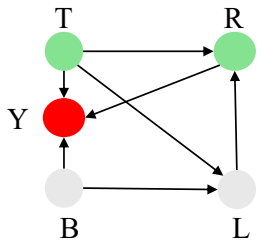

Figure 16: Epidemiology SCM. Light-gray nodes are non-manipulable variables; green nodes are manipulable variables; the red node is the target variable Y (viral load).

A.7 Soft-intervention datasets

This section details the five soft-intervention datasets used in the benchmark. These datasets fall into two categories:

1. **Function network benchmarks** (Ackley, Rosenbrock, Dropwave, Alpine2): Classical global optimization benchmarks reformulated as deterministic function networks, where each action variable propagates through intermediate nodes to the target. Interventions correspond to setting action variables, and the network structure defines soft causal relationships. These are *maximization* tasks.
2. **SCM-based policy benchmark** (Chain-soft): A full structural causal model in which a soft intervention replaces the structural equation of one variable with a context-dependent policy. This is a *minimization* task.

For each data set, we specify the network architecture, all structural equations, the action domains, and the target computation. Figures use a consistent color scheme: blue = intermediate variables, orange squares = action (intervention) variables, red = target.

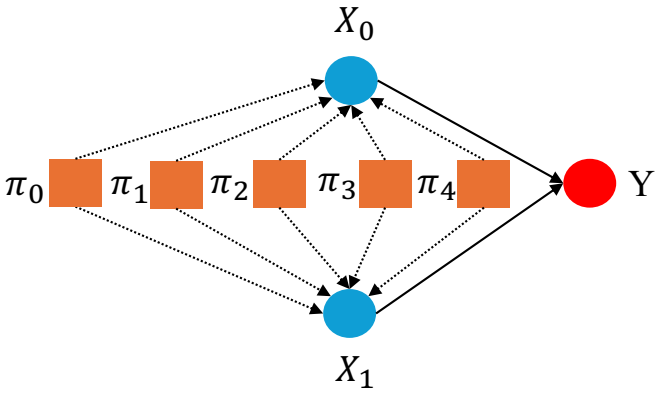

Figure 17: Ackley function network ($D = 6$). Blue nodes indicate intermediate variables; orange squares represent action (intervention) variables; the red node is the target variable Y.

A.7.1 Ackley

A multimodal, non-separable function-network benchmark with six action variables $a_0, \ldots, a_5$, two intermediate nodes X_0 and X_1, and target node Y. The task is to *maximize* Y. Following the standard configuration in the BO function-network (Sussex et al., 2023; 2024), we use the domain $a_i \in [-2, 2]$ for $i = 0, \ldots, 5$, corresponding to $D = 6$. Figure 17 shows the function network.

Action variables. $a_i \in [-2, 2], \quad i = 0, \ldots, 5.$

Intermediate variables.

$$X_0 = \frac{1}{D} \sum_{i=0}^{5} a_i^2,$$

$$X_1 = \frac{1}{D} \sum_{i=0}^{5} \cos(2\pi a_i),$$

where $D = 6$.

Target variable.

$$Y = 20 \exp\left(-0.2\sqrt{X_0}\right) + \exp(X_1) - 20 - e.$$

Interpretation. X_0 aggregates the average squared magnitude of the action variables, while X_1 aggregates the average cosine term. The target Y combines these two intermediate quantities to recover the Ackley objective in the form of maximum reward. The observational (unintervened) setting corresponds to $a_i = 0$ for all i.

A.7.2 Rosenbrock

A unimodal, non-separable function-network benchmark with a narrow curved valley. The network has action variables $a_0, \ldots, a_{D-1}$, intermediate nodes $X_0, \ldots, X_{D-2}$, and target node Y. The task is to *maximize* Y. We use the domain $a_i \in [-2, 2]$ and consider $D \in \{3, 5, 7\}$. Figure 18 shows the chain-structured function network.

Action variables. $a_i \in [-2, 2], \quad i = 0, \ldots, D-1.$

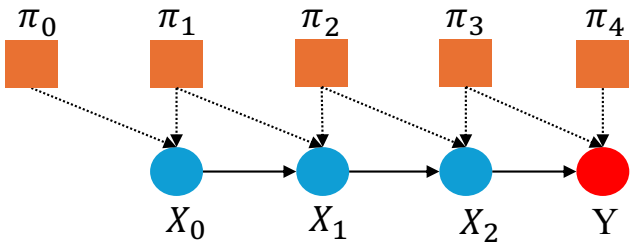

Figure 18: Rosenbrock function network (shown for $D = 3$). Blue nodes indicate intermediate variables; orange squares represent action (intervention) variables; the red node is the target variable Y.

Intermediate variables. The first intermediate node computes the first Rosenbrock term:

$$X_0 = -100\big(a_1 - a_0^2\big)^2 - (1 - a_0)^2.$$

Each subsequent node recursively accumulates the next pairwise contribution:

$$X_k = -100\big(a_{k+1} - a_k^2\big)^2 - (1 - a_k)^2 + X_{k-1}, \qquad k = 1, \ldots, D - 2.$$

Target variable.

$$Y = X_{D-2}.$$

Interpretation. The Rosenbrock function network forms a causal chain in which the first node computes the initial Rosenbrock component, and each later node adds one additional pairwise contribution to the accumulated objective. The target Y recovers the Rosenbrock objective in the form of maximum reward. The observational setting corresponds to $a_i = 0$ for all i.

A.7.3 Dropwave

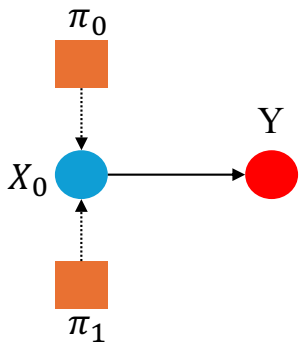

Figure 19: Dropwave function network. Blue nodes indicate intermediate variables; orange squares represent action (intervention) variables; the red node is the target variable Y.

A highly multimodal two-dimensional function-network benchmark with two action variables a_0 and a_1, an intermediate node X_0 and a target node Y. The task is to *maximize* Y. Following the BO setup of the standard function-network (Sussex et al., 2023; 2024), we use the domain $a_0, a_1 \in [-5.12, 5.12]$. Figure 19 shows the function network.

Action variables. $a_0, a_1 \in [-5.12, 5.12]$.

Intermediate variable.

$$X_0 = \sqrt{a_0^2 + a_1^2} + \epsilon_0.$$

Target variable.

$$Y = \frac{1 + \cos(12X_0)}{2 + 0.5X_0^2} + \epsilon_Y.$$

Interpretation. The intermediate node X_0 computes the radial distance from the origin induced by the two action variables. The target Y applies the oscillatory Dropwave transformation to this radial value. Following MCBO and related function-network BO settings, this benchmark can also be evaluated in a noisy setting by choosing non-zero ϵ_0 and ϵ_Y.

A.7.4 Alpine2

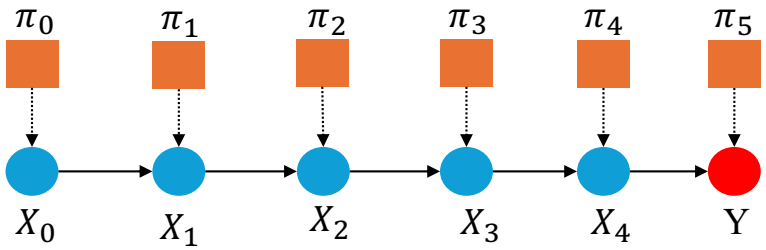

Figure 20: Alpine2 function network ($K = 6$). Blue nodes indicate intermediate variables; orange squares represent action (intervention) variables; the red node is the target variable Y.

A multimodal, separable chain-structured function-network benchmark with action variables $a_0, \ldots, a_5$, intermediate nodes $X_0, \ldots, X_4$, and target node Y. The task is to *maximize* Y. We use the setting $K = 6$ with the action domain $a_i \in [0, 10]$ for $i = 0, \ldots, 5$. Figure 20 shows the function network.

Action variables. $a_i \in [0, 10], \quad i = 0, \ldots, 5.$

Intermediate variables. The first intermediate node is defined as

$$X_0 = -\sqrt{a_0}\sin(a_0).$$

Each subsequent node combines its own action variable with the output of the previous node through the Alpine2 multiplicative structure:

$$X_k = \sqrt{a_k}\sin(a_k)\, X_{k-1}, \qquad k = 1, \ldots, 4.$$

Target variable.

$$Y = \sqrt{a_5}\sin(a_5)\, X_4.$$

Interpretation. Alpine2 forms a causal chain in which each stage combines its own decision variable with the output of the previous stage through a multiplicative structure. The analytical global maximum is $\prod_{i=0}^{5}\sqrt{a_i^*}\sin(a_i^*) = 2.808^6 \approx 400$ at $a_i^* \approx 7.917$ for all i. This subsection corresponds to the $K = 6$ configuration used in our experiments.

A.7.5 Chain-soft

The soft-intervention Chain benchmark uses the same base SCM as Chain-hard (Appendix A.5.4), with observed nodes X, W, Z, Y. Here, X is a non-manipulable context variable, W is hard-intervenable, Z is *soft*-intervenable via a context-dependent policy, and Y is the target. The task is to *minimize* Y. Figure 21 illustrates the SCM with the intervention policy.

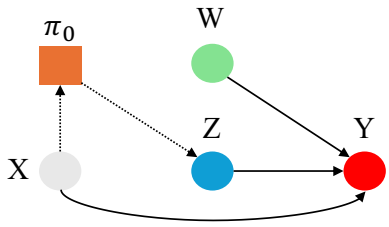

Figure 21: Chain SCM under soft intervention. Light-gray nodes are non-manipulable variables; green nodes indicate hard-intervenable variables; blue nodes indicate intermediate variables; the orange rectangle denotes the intervention policy π_0; the red node is the target variable Y.

Exogenous variables.

$$X = U_X \sim \mathcal{N}(0,1),$$
$$W = U_W \sim \mathcal{N}(0,1).$$

Endogenous variables (unintervened).

$$Z = -0.5X + U_Z,$$
$$Y = -W - 3ZX + U_Y,$$

where $U_Z \sim \mathcal{N}(0,1)$ and $U_Y \sim \mathcal{N}(0,1)$, all mutually independent.

Intervention policy. Under soft intervention, the structural mechanism of Z is replaced by a context-dependent policy π_0:

$$Z := \pi_0(X),$$

so that the intervened SCM becomes:

$$X = U_X,$$
$$W = U_W \quad \text{or} \quad W := w,\ w \in [-1,1],$$
$$Z = \pi_0(X),$$
$$Y = -W - 3ZX + U_Y.$$

Intervention domain and policy parameterization. The admissible scopes include a purely functional intervention in Z (through the policy π_0) and a mixed intervention combining the functional intervention in Z with a hard intervention in W over $[-1,1]$. Following Gultchin et al. (2023), each functional intervention is represented using samples $\texttt{GridSize} = 10$ and $N_\alpha = N_\beta = 10$ for the context variable, with coefficients α_i and β_j uniformly sampled from $[-0.27, 0.27]$. This parameterization keeps the induced intervention values comparable to the hard-intervention range $[-1,1]$.

Interpretation. Because the target $Y = -W - 3ZX + U_Y$ depends on the interaction between Z and the context X, the optimal intervention in Z generally varies with X. This makes Chain-soft a key benchmark for evaluating context-adaptive intervention strategies. In the figure, the dashed arrows $X \dashrightarrow \pi_0 \dashrightarrow Z$ indicate that the policy takes X as input and determines the intervened value of Z.