Active Learning for Optimal Intervention Design in Unknown Linear Causal Models

Runshi Yang
University of Toronto
ryang@cs.toronto.edu

Walter Virany
University of Toronto
wvirany@cs.toronto.edu

Bo Gong
University of Toronto
bogong@cs.utoronto.ca

Abstract

Designing effective interventions in causal systems is a central challenge in domains where experiments are costly and the intervention space is large. We study this problem in the setting of linear structural causal models (SCMs), where the goal is to select soft interventions that steer the post-intervention mean of the system toward a desired target. Prior work—most notably CIV and CIV-OW [1]—frames this task as an active learning problem and proposes an acquisition function that reduces uncertainty about the optimal intervention. However, these methods assume that the causal graph is known, which is often unrealistic.

In this project, we focus on extending CIV-based active learning to settings where *the causal graph is unknown*, and the learner must reason jointly over parameter and structural uncertainty. We develop a Bayesian formulation that maintains a posterior over both SCM parameters and DAG structures, approximated using samples from the interventional Bayesian Gaussian equivalent (iBGe) score. We then integrate this structural uncertainty directly into the CIV objective, yielding a DAG-averaged acquisition rule.

We compare an oracle baseline with known graph structure, a misspecified fixed graph baseline, and our Bayesian approach that marginalizes over graph uncertainty using weighted DAG samples. We study how incorporating structural uncertainty impacts sample efficiency and intervention performance, and find that by incorporating uncertainty about the graph structure, we can improve upon the setting where the causal graph is misspecified.

All code can be found at `https://github.com/wvirany/actlearn_optint`.

1 Introduction

Understanding how to design effective interventions in complex systems is a central goal of causal inference. In many scientific domains—such as biology, economics, and the social sciences—interventions are costly, and each experimental round provides only limited information about the underlying causal mechanisms. This creates a need for *sample-efficient, goal-directed experimental design*, where the experimenter strategically selects interventions that accelerate progress toward a desired objective.

In this work, we study optimal intervention design in the setting of *linear structural causal models* (SCMs). The system consists of variables $\mathbf{x} \in \mathbb{R}^p$ governed by a directed acyclic graph (DAG), and a *soft intervention* introduces an additive perturbation vector $\mathbf{a}$ into the structural equations. The goal is to choose an intervention $\mathbf{a}^\star$ whose resulting post-intervention mean $\mathbb{E}[\mathbf{x} \mid \mathrm{do}(\mathbf{a}^\star)]$ closely matches a specified target mean $\boldsymbol{\mu}^\star$. This formulation captures a broad class of practical tasks, including steering gene expression profiles toward therapeutically relevant states or driving engineered systems to desirable equilibria.

A promising recent approach to this problem is the *causal integrated variance* (CIV) framework introduced by Zhang et al. [1]. CIV views optimal intervention design as an *active learning* problem: at each round,

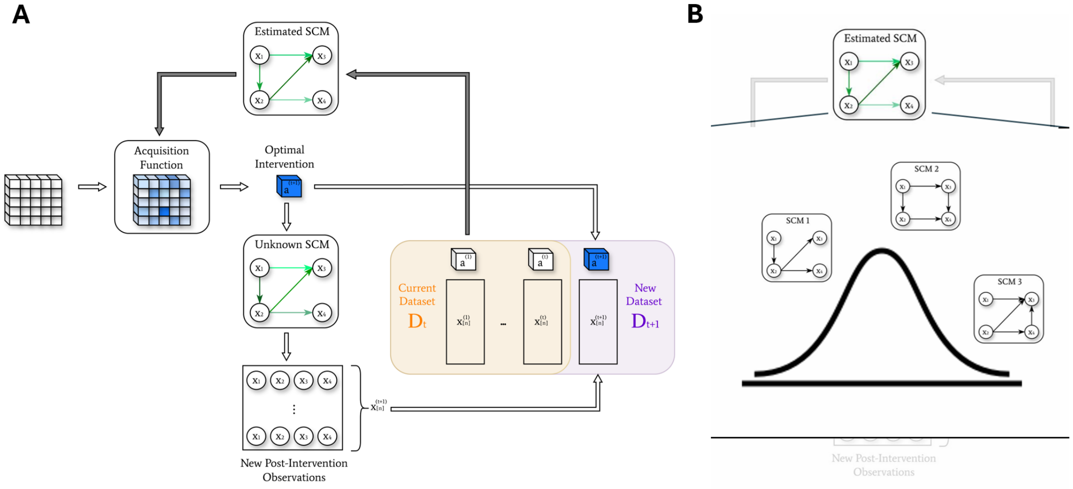

Figure 1: Schematic diagram of our work. (A) In the causal integrated variance (CIV) framework, an optimal intervention is selected by an acquisition function with strong empirical performance and principled Bayesian motivation. However,this approach assumes a known causal graph structure. (B) We develop an extension of CIV-based active learning to the setting where the causal graph is unknown, by maintaining a posterior distribution over both DAGs and SCM parameters.

one selects the intervention that most effectively reduces uncertainty about the optimality gap between a candidate intervention and $\mathbf{a}^\star$. Their derivation yields a closed-form acquisition function with strong empirical performance and principled Bayesian motivation. However, a key limitation of CIV and CIV-OW is the assumption that the causal graph structure is known *a priori*. In realistic settings, especially in high-dimensional biological systems, the graph must be inferred from data, and its uncertainty can significantly affect downstream decisions (**Figure** 1A).

Our project addresses this gap. We develop an extension of CIV-based active learning to the setting where the causal graph is *unknown*, and both structural and parametric uncertainty must be taken into account. Leveraging the interventional Bayesian Gaussian equivalent (iBGe) score, we maintain a posterior distribution over DAGs and SCM parameters, and we construct an acquisition rule that marginalizes the CIV objective over sampled graph structures. This leads to a *DAG-averaged* acquisition function that incorporates uncertainty not only in model parameters but also in the underlying causal structure (**Figure** 1B).

Due to the lack of available processed single-cell perturbation data, our experiments focus on synthetic causal graphs, where ground truth allows controlled comparisons across approximation regimes. We evaluate three settings: (i) a known ground-truth SCM, (ii) a misspecified single graph, and (iii) our Bayesian treatment that averages over posterior DAG samples. These experiments highlight the situations in which structural uncertainty meaningfully affects intervention selection and demonstrate the potential benefits of incorporating Bayesian structure learning into active causal experimental design.

Overall, our work positions DAG-averaged CIV as a practical and theoretically motivated approach for optimal intervention design when the causal structure is uncertain—a scenario that arises frequently in modern data-driven science.

2 Related Work

Our project builds on several strands of research spanning causal inference, active learning, Bayesian structure learning, and optimal experimental design. We summarize the most relevant contributions below and highlight how our work differs from and extends previous approaches.

Active learning for optimal interventions. The most directly related work is Zhang et al. [1], who introduced the *causal integrated variance* (CIV) and *output-weighted CIV* (CIV-OW) acquisition functions.

Their framework treats optimal intervention design as an active learning problem over a linear SCM with a *known* causal graph, deriving a closed-form Bayesian measure of uncertainty in the optimality gap. CIV-OW reweights the intervention space to emphasize directions that matter most for the target. Their experiments demonstrate strong performance on both synthetic SCMs and single-cell perturbation data. Our work follows their formulation of the optimality gap but removes the assumption of a known graph, extending CIV to incorporate uncertainty over DAG structures.

Graph learning from observational and interventional data. A significant body of work studies how to infer causal DAGs from a mixture of observational and interventional datasets. Hauser and Bühlmann [2] formalized joint likelihood models for Gaussian interventions and characterized interventional Markov equivalence classes. He and Geng [3] examined *active structure learning*, selecting interventions that optimally reduce graph uncertainty. More recent approaches, such as Scherrer et al. [4], integrate active interventions with differentiable structure learning for nonlinear causal models. These works focus primarily on recovering the true graph, whereas our objective is *not* structural accuracy itself: we study how graph uncertainty affects downstream intervention design.

Bayesian structure learning and iBGe. Bayesian DAG learning traditionally relies on decomposable likelihood scores such as the Bayesian Gaussian equivalent (BGe) score, combined with MCMC schemes over DAG space. Kuipers and Moffa [5] proposed *partition MCMC*, which samples DAGs via a partition-based state space for improved mixing. More recently, Kuipers and Moffa introduced the *interventional Bayesian Gaussian equivalent* (iBGe) score [6], which generalizes BGe to incorporate interventional data, including settings where soft intervention targets may be unknown. Our work leverages iBGe-based posterior sampling as a modular way to obtain graph samples for marginalizing the CIV objective over DAG uncertainty. Rather than using expensive MCMC sampling, we adopt a bootstrap-based approximation following Agrawal et al. [7], where candidate DAGs are weighted by their iBGe scores to form an importance-weighted posterior approximation.

Optimal experimental design. Classical optimal design criteria, such as A-, D-, and G-optimality, are widely used to select informative experiments in statistical models [8, 9]. Bayesian optimal design extends these ideas by selecting experiments that maximize expected utilities (e.g., reduction in posterior entropy). While these classical methods do not directly address causal intervention targets, they motivate CIV's variance-based criterion. CIV can be interpreted as a decision-focused analogue of Bayesian optimal design, where utility is tied not to parameter estimation alone but specifically to reducing uncertainty about the *optimality gap*. Our project continues this tradition while expanding the uncertainty set to include DAG structure.

Positioning our contribution. Overall, our work lies at the intersection of the above lines of research. Relative to Zhang et al. [1], we contribute a version of CIV that accounts for DAG uncertainty. Compared to Bayesian DAG learning methods [5, 6], we do not aim to recover the true graph but instead use graph samples to improve intervention selection. And unlike structure-learning interventions [3, 4], our objective is decision-oriented: we seek interventions that steer the system toward a predefined target mean. By integrating structure uncertainty directly into the acquisition function, our method provides a principled and practically motivated extension of CIV for settings where the causal graph is not known in advance.

3 Formal Description

We consider a system of variables $\mathbf{x} \in \mathbb{R}^p$ governed by a linear structural causal model (SCM). The causal relationships are encoded by a directed acyclic graph (DAG) $\mathcal{G}$ with weighted adjacency matrix $\mathbf{B}$. Each variable x_i is generated according to a linear structural equation with additive Gaussian noise:

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \Sigma). \tag{1}$$

Assuming the matrix $(\mathbf{I} - \mathbf{B})$ is invertible, the observational distribution is

$$\mathbf{x} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\epsilon}, \qquad \mathbb{E}[\mathbf{x}] = \mathbf{0}, \quad \mathrm{Cov}(\mathbf{x}) = (\mathbf{I} - \mathbf{B})^{-1}\Sigma(\mathbf{I} - \mathbf{B})^{-T}. \tag{2}$$

3.1 Soft Interventions and Post-Intervention Mean

A *soft intervention* introduces an additive shift $\mathbf{a} \in \mathbb{R}^p$ into the structural equations:

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{a} + \boldsymbol{\epsilon}. \tag{3}$$

The corresponding post-intervention distribution is

$$\mathbf{x} = (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{a} + \boldsymbol{\epsilon}), \tag{4}$$

with mean

$$\mu_{\mathbf{a}} := \mathbb{E}[\mathbf{x} \mid \mathrm{do}(\mathbf{a})] = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{a}. \tag{5}$$

Given a target mean $\boldsymbol{\mu}^\star$, the goal is to find the optimal intervention

$$\mathbf{a}^\star = (\mathbf{I} - \mathbf{B})\, \boldsymbol{\mu}^\star. \tag{6}$$

3.2 Optimality Gap

Following Zhang et al. [1], we define the *optimality gap* as

$$g(\mathbf{a}) := \|\mu_{\mathbf{a}} - \boldsymbol{\mu}^\star\|_2 = \left\|(\mathbf{I} - \mathbf{B})^{-1}\mathbf{a} - \boldsymbol{\mu}^\star\right\|_2. \tag{7}$$

In a Bayesian treatment of the SCM parameters, uncertainty in $\mathbf{B}$ induces uncertainty in $g(\mathbf{a})$, giving rise to the posterior variance

$$\sigma^2_{g(\mathbf{a})|\mathcal{D}} := \mathrm{Var}(g(\mathbf{a}) \mid \mathcal{D}). \tag{8}$$

3.3 Classical CIV Acquisition Function

The CIV acquisition function selects the next intervention $\mathbf{a}'$ by integrating the posterior predictive variance of the optimality gap over a reference measure ν:

$$h_{\mathrm{CIV}}(\mathbf{a}', \mathcal{D}) := \int_{\mathbf{a}} \sigma^2_{g(\mathbf{a})|\mathcal{D}(\mathbf{a}')}\, d\nu(\mathbf{a}), \tag{9}$$

where $\mathcal{D}(\mathbf{a}')$ denotes the augmented dataset after intervening with $\mathbf{a}'$.

Critically, this formulation assumes that the causal graph $\mathcal{G}$ is known.

3.4 Bayesian Model with Unknown DAG

We relax this assumption by treating $\mathcal{G}$ as a latent variable. We place a joint Bayesian model over $(\mathbf{B}, \Sigma, \mathcal{G})$:

$$p(\mathbf{B}, \Sigma, \mathcal{G} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \mathbf{B}, \Sigma, \mathcal{G})\; p(\mathbf{B}, \Sigma \mid \mathcal{G})\; p(\mathcal{G}), \tag{10}$$

where $p(\mathcal{D} \mid \mathbf{B}, \Sigma, \mathcal{G})$ incorporates both observational and interventional data.

Following Kuipers and Moffa [6], we use the interventional Bayesian Gaussian equivalent (iBGe) score to compute

$$p(\mathcal{D} \mid \mathcal{G}). \tag{11}$$

We approximate the posterior over DAGs by sampling

$$\mathcal{G}_1, \ldots, \mathcal{G}_K \sim p(\mathcal{G} \mid \mathcal{D}) \tag{12}$$

using the DAG bootstrap method discussed in Agrawal et al. [7], and define normalized importance weights

$$\omega_k := \frac{p(\mathcal{D} \mid \mathcal{G}_k)\, p(\mathcal{G}_k)}{\sum_{j=1}^K p(\mathcal{D} \mid \mathcal{G}_j)\, p(\mathcal{G}_j)}. \tag{13}$$

3.5 CIV with unknown graphs (Our Contribution)

We marginalize the optimality-gap variance over sampled DAGs:

$$\sigma^2_{g(\mathbf{a})|\mathcal{D}} \approx \sum_{k=1}^K \omega_k\ \mathrm{Var}(g(\mathbf{a}) \mid \mathcal{G}_k, \mathcal{D}). \tag{14}$$

This yields our DAG-averaged version of CIV for unknown graphs (CIV-UG):

$$h_{\text{CIV-UG}}(\mathbf{a}', \mathcal{D}) := \sum_{k=1}^K \omega_k\ h_{\mathrm{CIV}}(\mathbf{a}', \mathcal{D} \mid \mathcal{G}_k). \tag{15}$$

When $K = 1$, this formulation reduces to computing CIV with a single learned DAG structure. For K > 1, the weighted sum integrates over structural uncertainty, with weights ω_k determined by how well each candidate DAG $\mathcal{G}_k$ scores the observed data according to the iBGe score.

3.6 Algorithm: DAG-Averaged CIV Active Learning

Algorithm 1 DAG-Averaged CIV Active Learning

```
**Input:** Prior $p(\mathbf{B}, \Sigma, \mathcal{G})$, target mean $\boldsymbol{\mu}^\star$, intervention space $\mathcal{A}$, initial dataset $\mathcal{D}_0$.
**for** $t = 0, 1, 2, \dots, T-1$ **do**
    **Sample DAGs:** Approximate $p(\mathcal{G} \mid \mathcal{D}_t)$ to obtain candidate graphs $\mathcal{G}_1, \dots, \mathcal{G}_K$:
        Generate $K$ bootstrap resamples of $\mathcal{D}_t$.
        Run structure learning (e.g., GES) on each resample to obtain $\mathcal{G}_k$.
        Score each $\mathcal{G}_k$ on the full dataset $\mathcal{D}_t$ using iBGe.
    **Compute importance weights:**
        $\omega_k = \frac{p(\mathcal{D}_t \mid \mathcal{G}_k) \cdot p(\mathcal{G}_k)}{\sum_{j=1}^{K} p(\mathcal{D}_t \mid \mathcal{G}_j) \cdot p(\mathcal{G}_j)}.$
    **for** each candidate intervention $\mathbf{a} \in \mathcal{A}$ **do**
        Compute $h_{\text{CIV}}(\mathbf{a}, \mathcal{D}_t \mid \mathcal{G}_k)$ for each $k = 1, \dots, K$.
        Compute DAG-averaged acquisition score:
            $h_{\text{DAG-CIV}}(\mathbf{a}, \mathcal{D}_t) = \sum_{k=1}^{K} \omega_k \, h_{\text{CIV}}(\mathbf{a}, \mathcal{D}_t \mid \mathcal{G}_k).$
    **end for**
    **Select next intervention:**
        $\mathbf{a}_{t+1} = \arg\min_{\mathbf{a} \in \mathcal{A}} \; h_{\text{DAG-CIV}}(\mathbf{a}, \mathcal{D}_t).$
    Perform intervention $\mathbf{a}_{t+1}$ and collect $n$ samples $\mathbf{x}_{[n]}$.
    Update dataset $\mathcal{D}_{t+1} = \mathcal{D}_t \cup (\mathbf{x}_{[n]}, \mathbf{a}_{t+1})$.
**end for**
**Output:** Estimated optimal intervention $\mathbf{a}^\star$.
```

3.7 Comparison to Standard CIV

Key difference:

Standard CIV assumes a *single known causal graph.* Our method replaces that assumption with a Bayesian average over candidate graphs.

Benefits:

- Robust to graph misspecification.
- Incorporates structural uncertainty into decision-making.
- Reduces overconfidence when data weakly identifies the DAG.

When they coincide:

If the posterior over $\mathcal{G}$ is sharply peaked around a single DAG, DAG-CIV and classical CIV behave identically.

4 Experiments and Demonstration

In this section, we evaluate our extension of the CIV framework to cases with unknown causal structure. All experiments are performed on synthetic causal models, and we describe our data generation process, baseline models, and experimental results.

4.1 Synthetic Data Generation

Following the experimental setup from Zhang et al. [1], we construct linear–Gaussian SCMs with p variables and a ground-truth DAG $\mathcal{G}^\star$. DAGs are generated randomly using the Erdős–Rényi model with edge probability $q = 0.2$, and edge weights are sampled uniformly from $[-1, -0.25] \cup [0.25, 1]$. Given $\mathcal{G}^\star$ and weight matrix $\mathbf{B}^\star$, we draw observational samples from

$$\mathbf{x} = (\mathbf{I} - \mathbf{B}^\star)^{-1}\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \Sigma^\star),$$

where $\Sigma^\star = 0.4 \cdot \mathbf{I}$. Soft interventions with shift $\mathbf{a}$ are simulated using the modified structural equations

$$\mathbf{x} = (\mathbf{I} - \mathbf{B}^\star)^{-1}(\mathbf{a} + \boldsymbol{\epsilon}).$$

The optimal intervention with respect to the oracle is given by $\mathbf{a}^\star = (\mathbf{I} - \mathbf{B}^\star)\mu^\star$ for a randomly sampled target mean $\mu^\star$, normalized such that $\|\mathbf{a}^\star\| = 1$. For each experiment, we generate $n_{\text{obs}} = 100$ initial observational samples.

4.2 Baselines

We compare our proposed CIV-UG method against three baselines:

1. **Oracle (Ground-truth CIV):** Classical CIV computed using the true DAG $\mathcal{G}^\star$. The parameters $\mathbf{B}$ are learned via Bayesian linear regression from interventional data. This represents the best-case performance when the causal structure is known.
2. **Misspecified:** CIV computed on a fixed misspecified DAG $\tilde{\mathcal{G}}$, constructed by randomly adding and removing edges from $\mathcal{G}^\star$ until a target Structural Hamming Distance (SHD) is reached. This baseline demonstrates the scenario when the true graph is unknown, and uncertainty is not taken into account.
3. **Random:** At each step, $\mathbf{a}^{(t+1)}$ is sampled uniformly from the unit ball and used as the intervention. This provides a lower bound demonstrating the utility of active learning.
4. **CIV-UG (ours):** CIV extended to the unknown graph setting. At each active learning step, we approximate the posterior distribution $P(\mathcal{G}|\mathcal{D})$ over DAG structures using the DAG bootstrap method and compute a weighted average of the CIV acquisition function over candidate graphs, as described in **Algorithm** 1. The importance weights ω_k are computed using the iBGe score [6], and we select the intervention that minimizes the DAG-averaged acquisition score (equation 15).

The iBGe hyperparameter a_m is set to 0.1 following [6]. Each method performs $T = 50$ active learning steps with $n = 10$ samples per intervention, following a warm-up period of $W = 3$ random interventions.

4.3 Results

Experiment 1: Graph-uncertainty improves upon misspecified baseline. We evaluate all four methods on synthetic DAGs with $p \in \{10, 20\}$ nodes, averaged over 15 independent trials. **Figure** 2 shows convergence curves plotting the distance to the optimal intervention $\|\mathbf{a}^{(t)} - \mathbf{a}^\star\|$ over time.

The Oracle baseline achieves the best performance at the end of the active learning procedure, reaching a final distance of approximately 0.2 by step 50. The misspecified baseline fails to converge to the optimal intervention, with final error approximately **2–3 times larger** than the oracle, demonstrating the severe cost of committing to an incorrect graph structure. The random baseline shows no improvement over time.

Interestingly, CIV-UG appears to perform well in early stages of the procedure under both settings, initially outperforming the oracle. Despite averaging over 15 random trials, the results appear to have high variance. This suggests that the improvement in performance between CIV-UG and the misspecified setting is modest at best. Of course, this depends on the degree to which the model is misspecified—when the causal graph is mostly or partially known, the Bayesian framework may not provide significant improvements. Future work could systematically explore the relationship between degree of misspecification (e.g., measured by SHD) and the performance gain achieved by CIV-UG relative to fixed misspecified baselines.

Experiment 2: $K = 1$ provides sufficient approximation. To assess whether averaging over multiple DAG samples improves performance beyond using a single learned DAG, we fix $p = 10$ and vary the number of bootstrap samples $K \in \{1, 5, 10, 20\}$. For each value of K, we bootstrap the combined dataset,

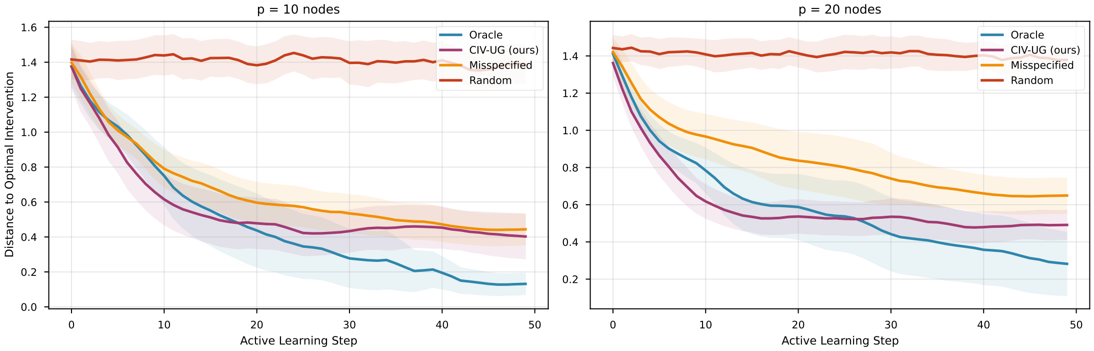

Figure 2: Distance to optimal intervention over $T = 50$ active learning steps for Oracle, Misspecified, Random, and CIV-UG methods on $p =$ (left) 10 and (right) 20 node graphs. Shaded regions indicated $\pm 1/2$ standard deviation across 15 random trials.

run GES on each sample to obtain K candidate DAGs, compute importance weights based on iBGe scores, and evaluate the DAG-averaged acquisition function as a weighted sum over all K DAGs according to Equation (15). When $K = 1$, this reduces to using a single learned DAG structure.

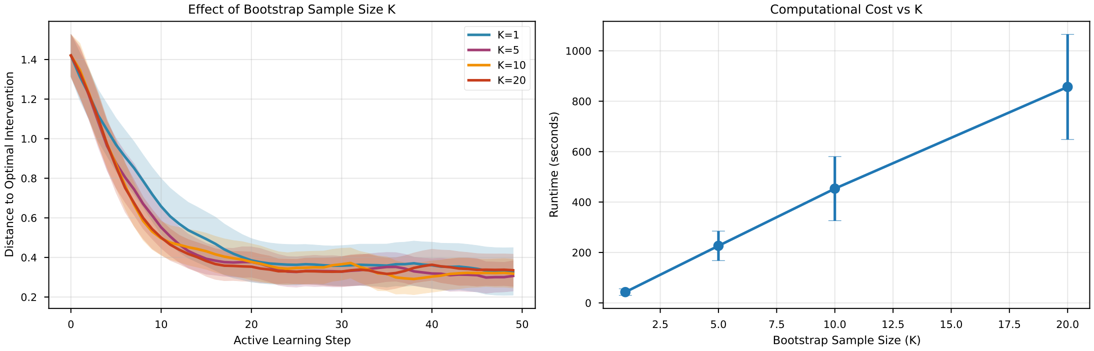

Figure 3: Effect of bootstrap sample size K on performance and computational cost for $p = 10$. **Left:** Convergence curves for $K \in \{1, 5, 10, 20\}$. Shaded regions indicate $\pm 1/2$ standard deviation across 15 random trials. **Right:** Computational cost in terms of wall-clock time for $K \in \{1, 5, 10, 20\}$. Error bars indicate ± 1 standard deviation across 15 trials.

Figure 3 (left) shows convergence curves for different values of K, and **Figure** 3 (right) shows the computational cost in terms of average wall-clock time of each active learning step as a function of K. We observe that all values of K achieve nearly identical performance, with convergence curves overlapping throughout the 50 active learning steps. This suggests that $K = 1$ **already provides a good approximation** to the posterior-averaged acquisition function for this task. While increasing K to 5, 10, or 20 provides no measurable performance benefit, it does increase computational cost approximately linearly with K (**Figure** 3, right). The average runtime per active learning step increases from $\sim$50 seconds for $K = 1$ to $\sim$900 seconds for $K = 20$.

This finding validates our choice of $K = 1$ as a computationally efficient approximation that retains the essential benefits of Bayesian structure learning. The lack of improvement with larger K may be due to the graph structure being relatively easy to recover by GES for small synthetic problems. Thus, it would be interesting to explore whether this holds on larger datasets.

5 Limitations

While our extension of CIV to settings with unknown causal structure provides a more realistic formulation of optimal intervention design, several important limitations remain. We discuss these below, together with opportunities for future work.

Posterior sampling over DAGs is computationally expensive. Our approach relies on sampling graphs $\mathcal{G}_1, \ldots, \mathcal{G}_K$ from the iBGe posterior. Even with partition MCMC, DAG sampling remains super-exponential in the worst case and scales poorly as the number of variables p increases. For moderate p (e.g., $p \leq 20$), sampling is feasible, but larger graphs may be intractable. This limitation is fundamental to Bayesian structure learning and suggests several possible directions: continuous relaxations of DAGs, differentiable structure learning, or amortized inference networks that approximate $p(\mathcal{G} \mid \mathcal{D})$ without expensive MCMC.

Linear-Gaussian assumptions limit expressiveness. We adopt a linear-Gaussian SCM for analytical tractability and to leverage closed-form CIV updates. However, real systems (e.g., biological networks) often exhibit nonlinear and heterogeneous causal mechanisms. Under such nonlinearities, the inversion $(\mathbf{I}-\mathbf{B})^{-1}$ may fail to capture true intervention effects. Extending CIV and DAG-averaged CIV to nonlinear SCMs—perhaps via local linearizations, Gaussian process mechanisms, or neural causal models—remains an open direction that could significantly broaden applicability.

The CIV objective may be misaligned with downstream goals. CIV quantifies uncertainty in the optimality gap, but this objective may not always align with all intervention design tasks. For example, in cases where stability or risk sensitivity matters, minimizing variance alone may lead to myopic decisions. Incorporating risk-adjusted utilities (e.g., CVaR or variance penalties), or adapting CIV to multi-objective intervention design, is a promising avenue for future work.

Reliance on synthetic data limits external validity. Experiments using biological datasets (single cell perturb-CITE-seq data) were planned but unfortunately not carried out due to the unavailability of processed perturb-CITE-seq data, despite several data requests. As a result, our empirical evaluation focuses exclusively on synthetic graphs. Although synthetic settings allow controlled comparisons, they may not reflect the complexities of real experimental noise, latent confounding, or imperfect intervention targeting. Future work should validate DAG-averaged CIV on real perturbation datasets, once suitable processed formats are available or once an alternative tractable dataset is identified.

Approximation errors accumulate across levels of uncertainty. Our method introduces two layers of approximation: Bayesian linear regression over edge weights *and* posterior approximation over graph structures. Errors at either stage can compound, affecting the acquisition score. Understanding how these approximations propagate and deriving bounds on the error of DAG-averaged CIV are open problems that would strengthen the theoretical foundations of our approach.

Future directions. Promising extensions include amortized DAG posteriors (similar to variational causal discovery), scalable nonparametric approximations of $g(\mathbf{a})$, tighter theoretical guarantees for structure-averaged utilities, and integration with intervention-cost models. Addressing these limitations would enable DAG-averaged CIV to operate in more realistic, high-dimensional environments and could support applications such as gene regulatory perturbation design, adaptive experimentation in policy, and robotics.

6 Conclusions

We investigated optimal intervention design in linear structural causal models under uncertainty about both parameters and causal graph structure. Building on the causal integrated variance (CIV) framework of Zhang et al. [1], we proposed a DAG-averaged acquisition function that integrates uncertainty over candidate causal graphs using samples from the iBGe posterior. This extension removes the unrealistic assumption that the causal graph is known a priori and provides a more robust formulation of active intervention selection.

Our experiments on synthetic datasets demonstrate that marginalizing over DAG uncertainty can meaningfully improve sample efficiency and reduce sensitivity to graph misspecification. In settings where the posterior over graphs is diffuse, DAG-averaged CIV avoids overconfident or brittle decisions that arise when committing to a single misspecified graph. When the posterior concentrates on a single DAG, our method naturally reduces to classical CIV, recovering its favorable properties.

Overall, our results highlight the importance of explicitly modeling structural uncertainty in causal experimental design. Although computational challenges remain, DAG-averaged CIV represents a principled and practical extension of previous approaches, and opens the door to more robust active learning strategies in domains where the causal structure is only partially known or difficult to recover.

7 Team Contributions

The authors contributed collaboratively. All authors contributed to the study proposal and design. RY contributed to the *Abstract*, *Introduction*, *Related Work* and *Formal Description* sections. WV conducted the experiments on synthetic datasets, plotted the results, and contributed to the *Experiments and Demonstration* and *Discussion* sections. BG contributed to the *Limitations* and *Conclusions* sections, and proofread the report.

References

[1] Jiaqi Zhang, Louis Cammarata, Chandler Squires, Themistoklis P. Sapsis, and Caroline Uhler. Active learning for optimal intervention design in causal models. *Nature Machine Intelligence*, 5(10): 1066–1075, 2023.

[2] Alain Hauser and Peter Bühlmann. Jointly interventional and observational data: Estimation of interventional markov equivalence classes of directed acyclic graphs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):291–318, 2015.

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

[4] Nino Scherrer, Olexa Bilaniuk, Yashas Annadani, Anirudh Goyal, Patrick Schwab, Bernhard Schölkopf, Michael C. Mozer, Yoshua Bengio, Stefan Bauer, and Nan Rosemary Ke. Learning neural causal models with active interventions. In *International Conference on Learning Representations (ICLR)*, 2022.

[5] Jack Kuipers and Giusi Moffa. Partition MCMC for inference on acyclic digraphs. *Journal of the American Statistical Association*, 112(517):282–299, 2017.

[6] Jack Kuipers and Giusi Moffa. The interventional bayesian gaussian equivalent score for bayesian causal inference with unknown soft interventions. In *Proceedings of the Fourth Conference on Causal Learning and Reasoning*, volume 275 of *Proceedings of Machine Learning Research*, pages 772–791. PMLR, 2025.

[7] Raj Agrawal, Chandler Squires, Karren Yang, Karthik Shanmugam, and Caroline Uhler. ABCD-Strategy: Budgeted Experimental Design for Targeted Causal Structure Discovery, February 2019.

[8] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statist Sci*, 10(3): 273 – 304, 1995.

[9] Friedrich Pukelsheim. *Optimal Design of Experiments*. SIAM, Philadelphia, 2nd edition, 2006.