Differentiable Multi-Target Causal Bayesian Experimental Design

Panagiotis Tigas [1][*]   **Yashas Annadani** [2][*]   **Desi R. Ivanova** [4]   **Andrew Jesson** [1]   **Yarin Gal** [1]

Adam Foster [3][†]   **Stefan Bauer** [2,5][†]

[1]OATML, University of Oxford   [2]KTH Stockholm   [3]Microsoft Research
[4]Department of Statistics, University of Oxford   [5]CIFAR Azrieli Global Scholar

Abstract

We introduce a gradient-based approach for the problem of Bayesian optimal experimental design to learn causal models in a batch setting — a critical component for causal discovery from finite data where interventions can be costly or risky. Existing methods rely on greedy approximations to construct a batch of experiments while using black-box methods to optimize over a *single target-state* pair to intervene with. In this work, we completely dispose of the black-box optimization techniques and greedy heuristics and instead propose a conceptually simple end-to-end gradient-based optimization procedure to acquire a set of optimal intervention target-state pairs. Such a procedure enables parameterization of the design space to efficiently optimize over a batch of *multi-target-state* interventions, a setting which has hitherto not been explored due to its complexity. We demonstrate that our proposed method outperforms baselines and existing acquisition strategies in both single-target and multi-target settings across a number of synthetic datasets.

1 Introduction

Imagine a scientist entering a wet lab to conduct experiments in order to discover the underlying causal relations within the system of interest. The scientist first comes up with some hypotheses, based on prior knowledge and past observations. Then, based on the formed hypotheses, an experimentation protocol to disambiguate between the competing hypotheses is devised. Additionally, because of the financial and ethical costs and risks involved in such experimentation (e.g. CRISPR or economic interventions), it is in the scientist's interest to minimize the number of batches required.

This process is known as experimental design, and assuming that the question of interest concerns discovering the causal structure of the system of interest, the process is known as experimental design for causal discovery. A Bayesian framework for this process has been proposed in prior work [55, 40, 8, 2, 53, 54, 56] which typically consists of updating an approximate posterior with past experimental data and using this updated posterior to compute experiments that are maximally informative, as evaluated by expected information gain—the objective of interest in *Bayesian Optimal Experimental Design* (BOED) [35, 7].
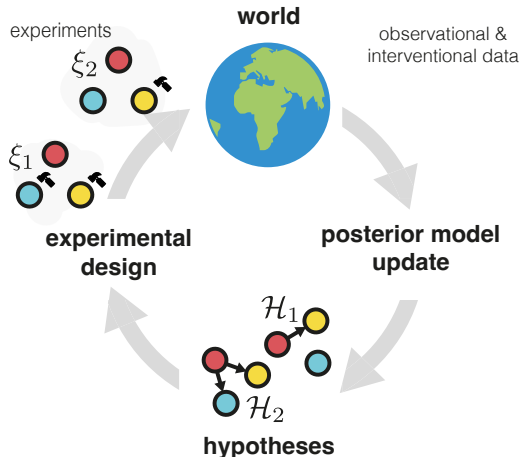
Figure 1: Causal Bayesian Experimental Design optimizes experiments that help disambiguate between competing causal hypotheses.

---

[*]Equal contribution. [†] Equal advising. Correspondence to: `ptigas@robots.ox.ac.uk`, `yashas.annadani@gmail.com`.

The problem of Bayesian Optimal Experimental Design for Causal Discovery (BOECD) is hard; computing the Bayesian posterior over Structural Causal Models (SCM)—a common framework for capturing causal relationships—is intractable. More importantly, the estimation and optimization of batch BOED objectives are computationally challenging, which has resulted in heuristics like the greedy batch strategy [2, 53] and soft-top-k batch strategy [54]. Additionally, in causal discovery, one is interested not only in identifying the variable (target) to intervene *on* but also the state to set the intervention *to*, resulting in a design space which is a product space of discrete targets and continuous states, making experimental design even more challenging. Tigas et al. [54] proposed to use Bayesian Optimization (BO) to optimize over the continuous state-space of the interventions and a soft-top-k heuristic to select a batch.

In this work, we propose a method for estimating and optimizing the BOED objective in a differentiable end-to-end manner, alleviating the inefficiencies introduced by the heuristics of the batch selection but also the black-box optimization over the intervening states. Specifically, we use estimators of mutual information based on nested estimation considered in Ryan [48], Myung et al. [41], Huan & Marzouk [23], Foster et al. [13] and extended it to the problem of causal discovery where the optimization is over both discrete nodes and continuous states. We cast the problem of batch experiment selection as a policy optimization where the policy uses either the Gumbel-Softmax or relaxed Bernoulli distribution [25, 37] for single target and multi-target settings respectively. When combined with the straight-through gradient estimator [5] to optimize over the targets and gradient ascent over corresponding states, we can explore the space of optimal designs efficiently with powerful optimizers [26]. Our proposed method requires very few assumptions about the causal model and can explore wide range of design settings as compared to prior work (see Table 2), thus opening up possibilities of experimental design for causal discovery in a broader range of applications.

2 BACKGROUND

2.1 CAUSALITY

Notation. Let $\mathbf{V} = \{1, \ldots, d\}$ be the vertex set of any Directed Acyclic Graph (DAG) $\mathbf{g} = (\mathbf{V}, E)$ and $\mathbf{X_V} = \{X_1, \ldots, X_d\} \subseteq \mathcal{X}$ be the random variables of interest indexed by $\mathbf{V}$.

Structural Causal Model. To deal with questions related to modelling causal relations between variables of interest, we employ the framework of Structural Causal Models (SCM) [46]. In many fields of empirical sciences like network inference in single cell genomics [19], inferring protein-signalling networks [49] and medicine [51], SCMs provide a framework to model the effects of interventions [44]– experiments which perturb the state of a variable to a desired state, thereby allowing to study the mechanisms which affect the downstream variables (for example, CRISPR gene knockouts [38] in genomics). Under this framework, each variable X_i has an associated *structural equation*, and is assigned a value which is a deterministic function of its direct causes $X_{\mathrm{pa}(i)}$ as well as an exogenous noise variable ϵ_i with a distribution P_{ϵ_i}:

$$X_i := f_i(X_{\mathrm{pa}(i)}, \epsilon_i) \ \ \forall i \in \mathbf{V}$$

f_i's are mechanisms that relate how the direct causes affect the variable X_i. If the structural assignments are assumed to be acyclic, these equations induce a DAG $\mathbf{g} = (\mathbf{V}, E)$ whose vertices correspond to the variables and edges indicate direct causes. An intervention on any variable X_i corresponds to changing the structural equation of that variable to the desired state (value), $X_i := s_i$, where $s_i \in \mathcal{X}_i$. It is denoted by the do-operator [44] as $\mathrm{do}(X_i = s_i)$.

In this work, we assume that the SCM is causally sufficient, i.e. all the variables are measurable, and the noise variables are mutually independent. Though the mechanisms f can be nonparametric in the general case, we assume that there exists a parametric approximation to these mechanisms with parameters $\gamma \in \Gamma$. In the case of linear SCMs, γ corresponds to the weights of the edges in E. We further require that the functions f are differentiable with respect to their parameters. Many classes of SCMs fall under this category, including the most commonly studied one – the Gaussian additive

noise models (ANM)[1]:

$$X_i := f_i(X_{\mathrm{pa}(i)}; \boldsymbol{\gamma}_i) + \epsilon_i, \ \ \epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$$

Bayesian Causal Discovery. If the SCM for a given set of variables $\mathbf{X_V}$ is unknown, it has to be estimated from a combination of observational data (data obtained in an unperturbed state of a system) and experimental data under an intervention. This problem is called causal induction or causal discovery [52]. This amounts to learning the parameters of the unknown SCM given by DAG $\mathbf{g}$, parameters of mechanisms, $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_d]$, and variances, $\sigma^2 = [\sigma_1^2, \ldots, \sigma_d^2]$. For notational brevity, henceforth we denote $\boldsymbol{\phi} = (\boldsymbol{\gamma}, \sigma^2)$ and all the parameters of interest with $\boldsymbol{\theta} = (\mathbf{g}, \boldsymbol{\phi})$. In Bayesian causal discovery [22], the parameters of SCM are treated as random variables whose beliefs are updated according to the Bayes rule. A Bayesian method for causal discovery is preferable to model epistemic uncertainty about the model due to finite data as well as characterize equivalence classes of SCM like Markov Equivalence Class (MEC) in the case of non-identifiability [45]. Interventions improve identifiability, but they have to be planned carefully. After acquiring interventional data, Bayesian methods update the posterior distribution to reduce the uncertainty of the SCM.

2.2 BAYESIAN OPTIMAL EXPERIMENTAL DESIGN

Bayesian Optimal Experimental Design (BOED) [35, 7] is an information theoretic approach to the problem of selecting the optimal experiment to estimate any parameter θ. For BOED, the *utility* of the experiment ξ is the mutual information (MI) between the observation $\mathbf{y}$ and $\boldsymbol{\theta}$:

$$\begin{aligned} \mathrm{U_{BOED}}(\xi) &\triangleq \mathcal{I}(\mathbf{Y}; \boldsymbol{\Theta} \mid \xi) \\ &= \mathop{\mathbb{E}}_{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta},\xi)} [\log p(\mathbf{y} \mid \boldsymbol{\theta}, \xi) - \log p(\mathbf{y} \mid \xi)] \end{aligned}$$

This objective is also known as the *Expected Information Gain* (EIG). The goal of BOED is to select the experiment that maximizes this objective $\xi^* = \arg\max_\xi \mathrm{U_{BOED}}(\xi)$. Unfortunately, evaluating and optimizing this objective is challenging because of the nested expectations [47] and several estimators have been introduced [13, 28], which can be combined with various optimization methods to select the designs [14, 24, 15, 6].

A common setting, called *static*, *fixed* or *batch* design, is to optimize B designs $\{\xi_1, \ldots, \xi_B\}$ at the same time. The designs are then executed, and the experimental outcomes are collected for a Bayesian update of the model parameters.

2.3 CAUSAL BAYESIAN EXPERIMENTAL DESIGN

Causal Bayesian Experimental Design is concerned with designing the most informative experiments to identify the true SCM so that the number of experiments required is as few as possible. An experiment in causal discovery corresponds to picking the intervention targets $I \in \mathcal{P}(\mathbf{V})$ and the corresponding states $S^I \in \bigcup_{i \in I} \mathcal{X}_i$ to set those targets to. A key component of such methods is computing a posterior over the parameters of the SCM. However, computing the posterior is a difficult task since the number of DAGs grows exponentially in the number of variables. Nevertheless, a plethora of methods exist [16, 3, 36, 9] which can be used with our approach.

Having access to such posterior models, one can estimate the EIG objective. One difficulty that still remains though is that optimizing the EIG objective over the experiments is a mixed discrete and continuous optimization problem, for which previous work has proposed to find the optimal state per node via the use of black-box methods like *Bayesian Optimization* (BO) [54]. Additionally, for the construction of the batch of experimental designs, a greedy approximation is used to incrementally select experiments, a method that is $1 - \frac{1}{e}$-approximate to the optimal solution [32].

3 DIFFERENTIABLE BAYESIAN CAUSAL EXPERIMENTAL DESIGN

Let $\boldsymbol{\Theta}$ be a random variable that models the uncertainty in the parameters of the true SCM, of which $\boldsymbol{\theta} := (\mathbf{g}, \boldsymbol{\phi})$ is a realization. An experiment to identify an intervention is denoted by $\xi :=$

---

[1]Note that differentiability of f is the only assumption we require with respect to an SCM. We do not require that the noise is additive. For clarity of exposition, we restrict our focus to an ANM as they are the most commonly studied class of SCMs.

$\{(I, S^I)\} \coloneqq \mathrm{do}(\mathbf{X}_I = S^I)$, where $I \in \mathcal{P}(\mathbf{V})$ is a set of target indices in the multi-target setting, and S^I are the corresponding states of those targets under intervention. The outcome of the experiment is denoted by $\mathbf{y} \sim P\left(X_1 = x_1, \ldots, X_d = x_d \mid \mathrm{do}\left(\mathbf{X}_I = S^I\right)\right) = p(\mathbf{y} \mid \xi)$. Here, $\mathbf{y}$ is an instance of the random variable $\mathbf{Y} \subseteq \mathcal{X}$ distributed according to the interventional distribution[2]. Due to causal sufficiency, the likelihood of data for a given $\boldsymbol{\theta}$ satisfies the causal Markov condition:

$$p(\mathbf{y} \mid \boldsymbol{\theta}, \xi) = \prod_{j \in \mathbf{V} \setminus I} p\left(x_j \mid \phi_j, \mathbf{x}_{\mathrm{pa}_\mathbf{g}(j)}, \mathrm{do}\left(\mathbf{X}_I = S^I\right)\right) \tag{1}$$

Along with a prior $p(\boldsymbol{\theta})$, the above equation defines a generative model of the data.

Design setting. As in prior work [54, 53], we are interested in the setting of batch design where we design B experiments at once before collecting experimental data. In other words, we seek a multiset of intervention targets and corresponding states which are jointly maximally informative about the parameters. We denote this multiset as $\xi_{1:B} \coloneqq (I_{1:B}, S^I_{1:B})$. After executing a batch of experiments and collecting experimental outcomes, an experimenter might wish to design a new batch of experiments based on collected data (as summarized by the posterior distribution). Let h_t denote experimental history $(\xi^1, \mathbf{y}^1), \ldots, (\xi^t, \mathbf{y}^t)$ after t batches of acquisition. The BOED objective for this batch setting at any point t is given by the joint mutual information:

$$\mathcal{I}(\mathbf{Y}^t_{1:B}; \boldsymbol{\Theta} \mid \xi^t_{1:B}, h_{t-1}) = \underset{\substack{p(\boldsymbol{\theta}|h_{t-1}) \\ p(\mathbf{y}^t_{1:B}|\boldsymbol{\theta}, \xi^t_{1:B})}}{\mathbb{E}} \left[\log \frac{p(\mathbf{y}^t_{1:B} \mid \boldsymbol{\theta}, \xi^t_{1:B})}{p(\mathbf{y}^t_{1:B} \mid \xi^t_{1:B}, h_{t-1})}\right] \tag{2}$$

where $\mathbf{Y}^t_{1:B}$ are the random variables corresponding to experimental outcomes for iteration t, $\mathbf{y}^t_{1:B}$ are the instances of these random variables and $\xi^t_{1:B}$ is the corresponding multiset of experimental designs. We drop the superscript t from these variables for simplicity of exposition. Ideally, we wish to maximize the above objective by obtaining the gradients $\nabla_{\xi_{1:B}} I$ and performing gradient ascent. However, the above objective is doubly intractable [47] and approximations are required. This usually leads to a two-stage procedure where the above objective is first estimated with respect to an inference network and then maximized with respect to designs [13], which can be typically inefficient [14].

3.1 ESTIMATORS OF THE JOINT MUTUAL INFORMATION

NESTED MONTE CARLO

Following [23, 14, 15], we consider an estimator that allows for approximating the EIG objective while *simultaneously* optimizing for the experiment ξ that maximizes the objective via gradient-based methods. This estimator, called Nested Monte Carlo (NMC), is based on contrastive estimation of the experimental likelihood and has been extensively used in Bayesian experimental design [48, 41]. More precisely, assuming some past observational and interventional data $h_{t-1} = \{(\xi^1, \mathbf{y}^1), \ldots, (\xi^{t-1}, \mathbf{y}^{t-1})\}$, for every parameter sample from the posterior distribution $\boldsymbol{\theta}_0 \sim p(\boldsymbol{\theta} \mid h_{t-1})$, a set of contrastive samples $\boldsymbol{\theta}_{1:L} \sim p(\boldsymbol{\theta} \mid h_{t-1})$ are considered to obtain a unified objective:

$$\mathcal{U}^t_{\mathrm{NMC}}(\xi_{1:B}) = \underset{\substack{p(\boldsymbol{\theta}_{0:L}|h_{t-1}) \\ p(\mathbf{y}_{1:B}|\boldsymbol{\theta}_0, \xi_{1:B}))}}{\mathbb{E}} \left[\log \frac{p(\mathbf{y}_{1:B} \mid \xi_{1:B}, \boldsymbol{\theta}_0)}{\frac{1}{L} \sum_{\ell=1}^L p(\mathbf{y}_{1:B} \mid \xi_{1:B}, \boldsymbol{\theta}_\ell))}\right] \tag{3}$$

This estimator converges to the true mutual information as $L \to \infty$ [47]. If the design space is continuous, the optimal *batch* of experiment $\xi^*_{1:B}$ can be found by *directly* maximizing the NMC objective ($\xi^*_{1:B} \leftarrow \arg\max_{\xi_{1:B}} \mathcal{U}^t_{\mathrm{NMC}}(\xi_{1:B})$) with gradient-based techniques [23].

The above objective requires estimating the posterior distribution $p(\boldsymbol{\theta} \mid h_{t-1})$ after every acquisition. For causal models, while it is generally hard to estimate this posterior due to DAG space of causal structures being discrete and super-exponential in the number of variables [55], many approaches exist in the literature [2, 36, 9]. These approximate posteriors can be nevertheless used for estimating the NMC objective.

---

[2]Note that when $I = \varnothing$, it corresponds to an observational/ non-experimental setting. In this case, $\mathbf{Y} = \mathbf{X_V}$.

IMPORTANCE WEIGHTED NESTED MONTE CARLO

To establish an alternative path to estimating the mutual information, we begin by utilizing an observation from Foster et al. [13] that it is possible to draw the contrastive samples from a distribution other than $p(\boldsymbol{\theta} \mid h_{t-1})$ and obtain an asymptotically exact estimator, up to a constant C that does not depend on $\xi_{1:B}^t$. Drawing samples from the *original* prior $p(\boldsymbol{\theta})$ gives the estimator

$$\mathcal{I}(\mathbf{Y}_{1:B}^t; \boldsymbol{\Theta} \mid \xi_{1:B}^t, h_{t-1}) - C = \lim_{L \to \infty} \mathop{\mathbb{E}}_{\substack{p(\boldsymbol{\theta}_0|h_{t-1})p(\boldsymbol{\theta}_{1:L}) \\ p(\mathbf{y}_{1:B}|\boldsymbol{\theta}_0, \xi_{1:B})}} \left[\log \frac{p(\mathbf{y}_{1:B}|\xi_{1:B}, \boldsymbol{\theta}_0)}{\frac{1}{L}\sum_{\ell=1}^{L} p(\mathbf{y}_{1:B}|\xi_{1:B}, \boldsymbol{\theta}_\ell)p(h_{t-1}|\boldsymbol{\theta}_\ell)} \right]$$

The remaining wrinkle is that we must sample $\boldsymbol{\theta}_0$ from $p(\boldsymbol{\theta}_0|h_{t-1})$. We propose the conceptually simplest approach of applying self-normalized importance sampling (SNIS) to the outer expectation. The resulting objective, based on efficiently re-using samples in a leave-one-out manner, can optimize designs by just sampling parameters from the prior, without having to estimate the posterior:

$$\mathcal{U}_{\text{IWNMC}}^t(\xi_{1:B}) = \mathbb{E}\left[\sum_{m=1}^{L} \omega_m \log \frac{p(\mathbf{y}_{m,1:B}|\boldsymbol{\theta}_m, \xi_{1:B})}{\frac{1}{L-1}\sum_{\ell \neq m} p(\mathbf{y}_{m,1:B}|\boldsymbol{\theta}_\ell, \xi_{1:B})p(h_{t-1}|\boldsymbol{\theta}_\ell)} \right] \tag{4}$$

where $\boldsymbol{\theta}_{1:L} \sim p(\boldsymbol{\theta}_{1:L})$ are sampled from the original prior, $\mathbf{y}_{m,1:B} \sim p(\mathbf{y}_{1:B}|\boldsymbol{\theta}_m, \xi_{1:B})$ are all the experimental outcomes in the batch for parameter $\boldsymbol{\theta}_m$ and $\omega_m \propto p(h_{t-1}|\boldsymbol{\theta}_m)$ are self-normalized weights. A full derivation is given in Section A.

As IWNMC does not require any posterior estimation but instead relies entirely on the prior, it completely sidesteps the causal discovery process for designing experiments. This is a paradigm change from the NMC estimator which requires causal discovery through the estimation of the posterior. However, we note that using IWNMC with just the prior (Eq. 4) as opposed to NMC (Eq. 3) comes with trade-offs. IWNMC typically requires a large L to get a good estimate of the EIG. In high dimensions, this can be computationally infeasible. Having a small L on the other hand might result in a failure case if the effective sample size of importance samples becomes 1. We can alleviate this issue if there is some prior information available which could be leveraged to design better proposal distributions. This might consist of knowledge of certain causal mechanisms of the system under study or access to some initial observational data. In such a case, a proposal distribution which encodes this information (for example with support on graphs which are in the Markov Equivalence Class (MEC) of the observational distribution) can be used instead of the prior. If no prior information is available or a good approximate inference technique is at our disposal, NMC is preferable in high dimensions. Surprisingly, we get good results on variables of size up to 5 with IWNMC from just the prior and up to 40 variables from a proposal distribution which has support on the MEC of observational distribution (see Sec 4.3).

3.2 OPTIMIZING OVER TARGETS AND STATES (DIFFCBED)

While the NMC estimator provides a unified objective to directly optimize over the designs $\xi_{1:B}$, it requires that the design space is continuous so that the gradients $\frac{\partial \mathcal{U}_{\text{NMC}}}{\partial I_{1:B}}$ and $\frac{\partial \mathcal{U}_{\text{NMC}}}{\partial S_{1:B}^I}$ can be computed. However, in the case of designing experiments for causal models, the challenge still remains that optimizing over intervention targets I with gradient-based techniques is not possible because it is a discrete choice.

In order to address this problem, we introduce a *design policy* π_ϕ with learnable parameters ϕ that parameterize a joint distribution over possible intervention targets and corresponding states. Instead of seeking the gradients $\frac{\partial \mathcal{U}_{\text{NMC}}}{\partial I_{1:B}}$ and $\frac{\partial \mathcal{U}_{\text{NMC}}}{\partial S_{1:B}^I}$, the goal now instead is to estimate $\frac{\mathcal{U}_{\text{NMC}}}{\partial \phi}$ so that policy can be updated to be close to optimal. Such a characterization of the design space allows us to use continuous relaxations of discrete distributions [37, 25] to obtain samples of designs and estimate NMC gradients.

Let $\mathbf{I}$ and $\mathbf{S}$ be the random variables which model all possible intervention target combinations and states for a batch design respectively. While there are many possibilities of instantiating the policy in practice, we consider the simplest case where $\pi_\phi(\mathbf{I}, \mathbf{S}) \triangleq \pi_{\phi_n}(\mathbf{I})\pi_{\phi_m}(\mathbf{S})$. As the state space is continuous[3], π_{ϕ_m} can be either deterministic (a delta Dirac with $\phi_m \in \mathbb{R}^{B \times d}$) or Gaussian with

---

[3]If the state space is discrete, optimizing π_{ϕ_m} would be similar to π_{ϕ_n} which involves reparameterized gradients.

$\phi_m \in \mathbb{R}^{2 \times B \times d}$ parameterizing its mean and log variance. In this work, we found it sufficient to use a deterministic policy over the state space. For the interventional targets, $\phi_n \in \mathbb{R}^{B \times d}$ parameterizes the logits of different relaxed versions of discrete distributions depending on the setting, which we describe below. The diffCBED algorithm is outlined in Algorithm 1.

SINGLE TARGET ($q = 1$)

In this setting, the intervention targets are one-hot vectors, as demonstrated in Figure 2. To sample one-hot vectors in a differentiable manner, we parametrize π_{ϕ_n} as a Gumbel-Softmax distribution [37, 25] over intervention targets, which is a continuous relaxation of the categorical distribution (in *one-hot* form). Additionally, we use the straight-through (ST) gradient estimator [5].
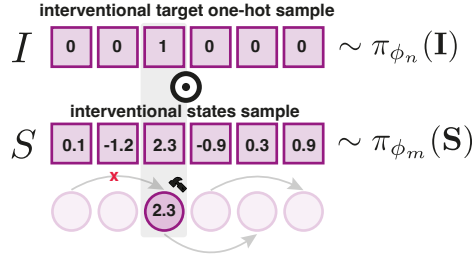
Figure 2: A design sample is obtained by first sampling $I_{1:B} \sim \pi_{\phi_n}(\mathbf{I})$, $S_{1:B} \sim \pi_{\phi_m}(\mathbf{S})$ and then setting states to be $S_{1:B}^I = S_{1:B} \odot I_{1:B}$. To obtain hard samples of I, we use the straight-through estimator [5]. Illustration for $B = 1$.

UNCONSTRAINED MULTI-TARGET ($q \leq d$)

If instead of a continuous relaxation of the categorical distribution, we parametrise the policy π_{ϕ_n} as a continuous relaxation of the Bernoulli distribution (Binary Concrete) [37], we can now sample multi-target experiments. Notice that since each interventional target sample will have at most d non-zero entries, this policy is suitable for multi-target experiments with an unconstrained number of interventions per experiment.

CONSTRAINED MULTI-TARGET ($q = k$)

Finally, when considering a setting where the number of targets per intervention is exactly k. However, this is a significantly more challenging case, since the policy needs to select a subset of k from d nodes. By using a continuous relaxation of subset sampling, as introduced in Xie & Ermon [58], combined with straight-through gradient estimator, we can efficiently optimize the policy to select a subset of nodes to intervene on.

Algorithm 1: Differentiable CBED

Input : $\mathcal{E}$ SCM Environment, N Initial observational samples, B Batch Size
1   $\mathcal{D}_{\text{obs}} \leftarrow \mathcal{E}.\text{sample}(N)$, $\mathcal{D}_{\text{int}} \leftarrow \varnothing$
2   Train $q(\mathbf{\Theta} \mid \mathcal{D}_{\text{obs}}) \approx p(\mathbf{\Theta} \mid \mathcal{D}_{\text{int}})$ using appropriate algorithm.
3   **for** *batch* $t = 1 \ldots T$ *Batches* **do**
4   Initialize design policy parameters $\phi = \{\phi_n, \phi_m\}$: trainable logits ϕ_n for the targets; trainable parameters ϕ_m for the states.
5   **for** *update step* $c = 1 \ldots C$ **do**
         ▷ Sample Interventional Targets and States
6          $\{\xi_{1:B}^{(o)}\}_{o=1}^O \sim \pi_\phi(\mathbf{I}, \mathbf{S})$
         ▷ Gradient ascent with straight-through gradient estimator
7   Update $\phi \rightarrow \phi + \alpha \frac{\partial}{\partial \phi} \frac{1}{O} \sum_{o=1}^O \left[\mathcal{U}_{\text{NMC}}^t(\xi_{1:B}^{(o)}) \right]$
     ▷ Intervene with learned policy
8      $\xi_{1:B} \sim \pi_\phi$
9      $\mathcal{D}_{\text{int}} \leftarrow \mathcal{D}_{\text{int}} \cup \mathcal{E}.\text{intervene}(\xi_{1:B})$
10   Update the posterior $q(\mathbf{\Theta} \mid \mathcal{D}_{\text{obs}} \cup \mathcal{D}_{\text{int}})$.

---

4   EXPERIMENTS

We evaluate the performance of our method on synthetic graphs and a range of baselines. We aim to investigate the following aspects empirically: (1) To what extent can we design good experiments

without performing intermediate causal discovery/ posterior estimation with IWNMC estimator from the prior? (2) Ability to design good experiments with a proposal distribution with IWNMC (3) the performance of our policy-based design in combination with the differentiable NMC estimator in single-target and multi-target settings, as compared to suitable baselines (Section E).

4.1 RESULTS

4.2 EVALUATION OF THE IWNMC ESTIMATOR
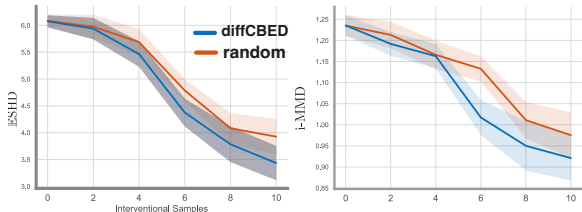
Figure 3: We test the designs acquired with IWNMC estimator with just the prior as opposed to the random policy (with random target and state acquisition) on variables of 5 dimensions. Plots correspond to unconstrained multi-target setting with $B = 2$ (shaded area represents 95% confidence intervals - 60 seeds).

In this section, we consider optimizing the designs with respect to the IWNMC estimator entirely from the prior, introduced in 3.1, sidestepping the causal discovery procedure. As noted before, estimating posteriors of causal models is hard, so it is important to understand to what extent IWNMC can be considered a suitable candidate for designing good experiments in the absence of a posterior. For this setting, we sample from the prior distribution over graphs by first sampling an ordering of nodes at random and then sampling edges with probability $p = 0.25$ which adhere to this topological order to give a DAG. We sample the mechanism parameters and noise variances of ANM at random from a Gaussian distribution with mean 0 and variance 1. Figure 3 demonstrates results for 5 variable unconstrained multi-target setting with batch size 2. For evaluation, we train DAG Bootstrap [16] with GIES [21] on the data acquired from each policy. We can see that we can recover the ground truth SCM faster than a random strategy. This is a surprising, but positive result given that our policy was trained entirely from samples from the prior. We also tested this approach for 10 nodes (results in Appendix G). While this resulted in better performance of the policy as opposed to random in terms of downstream metrics, we observed effective sample size reach 1 indicating that for 10 dimensions or higher, indicating that we might need a better proposal distribution or a posterior estimate.

4.3 EVALUATION IN HIGHER DIMENSIONS

EVALUATION OF IWNMC WITH PROPOSAL DISTRIBUTION

In this experiment, we consider 40 variables, constrained $(q = 5)$ multi-target and batch size $B = 2$. Further, we use the same setup as Sussex et al. [53] to make a fair comparison as well as to construct a proposal distribution. To construct a proposal distribution, we use 800 observational samples to train DAG Bootstrap [16, 2] and augment our posterior samples with samples of dags from the Markov Equivalence Class of the true graph, to make sure that there is support over the graphs from the MEC of the true graph (see Sussex et al. [53] for details). We then acquire a single batch of experiments from IWNMC estimator for our approach. For the baseline, we acquire a single batch of experiments from the estimator defined in [53].

For random and Sussex et al. [53] baseline, we set the interventional state to 5, as explained in [53]. Notice that in contrast to the baselines, our approach doesn't fix the state to 5 but optimizes over the state-space to perform the intervention with. In Table 1 we summarize our results. As we can see, our method outperforms random and SSGb [53] by a great margin, indicating that with a good proposal distribution, IWNMC can still be a promising candidate in higher dimensions.

RESULTS WITH NMC ESTIMATOR

For the following results, we use DAG-Bootstrap [2] with 20 components, an approximate posterior method based on GIES causal discovery method [21]. As GIES is not a differentiable method, once we compute the posterior via the DAG-Bootstrap algorithm, we transfer the weights of the posterior

Table 1: Results of multi-target experiments on graphs of size 40 (30 seeds $\pm$ s.e.). Similarly to [53], we are using posterior samples trained on observational data and re-weighting them with likelihoods.

Method	$\mathbb{E}$SHD $\downarrow$	F1 $\uparrow$	iMMD $\downarrow$
Random	43.78 ± 46.67	0.91 ± 0.08	0.16 ± 0.07
SSGb	15.59 ± 29.66	0.97 ± 0.05	0.10 ± 0.06
diffCBED	0.44 ± 0.21	0.99 ± 0.00	0.07 ± 0.01

samples (the bootstraps) into JAX [1] tensors to allow for the gradients to be computed with respect to the experiments.

Single-target synthetic graphs: In this experiment, we test against synthetic graphs of 50 nodes and batch size 5, where the graph is sampled from the class of Erdos-Renyi (a common benchmark in the literature [54, 56, 50]). In Figure 8 **(A,B,C)** we summarize the results. We observe that our method performs significantly better than the baselines.

20 nodes, unconstrained ($q \leq 20$), **batch size** $B = 2$**:** In this experiment, we want to evaluate the performance of our method as compared with the baselines, on sparse graphs over several acquisitions. Figure 8 **(D,E,F)** summarizes the results of this setting. We observe strong empirical performance as compared to all the baselines. Additional results are given in Section I.1.
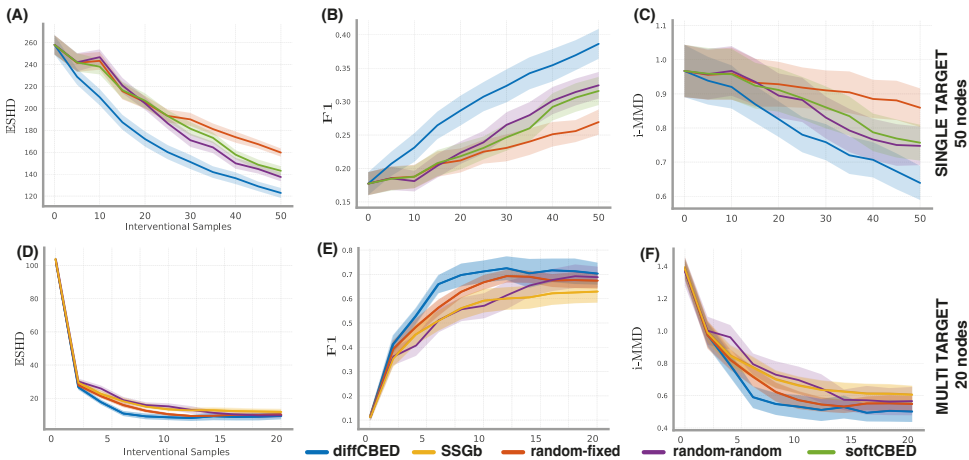
Figure 4: **(A,B,C)** Single target-state design setting results for Erdős–Rényi [12] graphs with $d = 50$ variables. **(D,E,F)** Multi target-state design setting results for Erdős–Rényi [12] graphs with $d = 20$ variables. Each experiment was run with 30 random seeds (shaded area represents 95% CIs)

5 DISCUSSION

Limitations: A primary limitation of our method is that it needs to estimate a posterior after every acquisition. While the proposed IWNMC estimator presents an interesting alternative, the designs are still non-adaptive. As demonstrated by [15, 24], a promising and exciting direction is to train a policy to be adaptive and propose new experiments in *real-time*.

Conclusion: We presented a gradient-based method for differentiable Bayesian Optimal Experimental for causal discovery. Our method allows not only for single-target but also various multi-target (constrained and unconstrained) batch acquisition of experiments. While prior work in Causal Bayesian Experimental Design relies on greedy approximations for the selection of a batch [2, 54] or black-box methods [56, 54] for optimizing over interventional states, our method utilizes gradient-based optimization procedures to simultaneously optimize for various design choices. Evaluation on different benchmarks suggests that our method is competitive with baselines.

REFERENCES

[1] Jax: Autograd and XLA. https://github.com/google/jax. 8

[2] Raj Agrawal, Chandler Squires, Karren Yang, Karthikeyan Shanmugam, and Caroline Uhler. Abcd-strategy: Budgeted experimental design for targeted causal structure discovery. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 3400–3409. PMLR, 2019. 1, 2, 4, 7, 8, 14, 16, 17

[3] Yashas Annadani, Jonas Rothfuss, Alexandre Lacoste, Nino Scherrer, Anirudh Goyal, Yoshua Bengio, and Stefan Bauer. Variational causal networks: Approximate bayesian inference over causal structures. *arXiv preprint arXiv:2106.07635*, 2021. 3, 14

[4] Vladimir Batagelj and Ulrik Brandes. Efficient generation of large random networks. *Physical Review E*, 71(3):036113, 2005. 17

[5] Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013. 2, 6

[6] Tom Blau, Edwin V Bonilla, Iadine Chades, and Amir Dezfouli. Optimizing sequential experimental design with deep reinforcement learning. In *International Conference on Machine Learning*, pp. 2107–2128. PMLR, 2022. 3, 14

[7] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pp. 273–304, 1995. 1, 3

[8] Hyunghoon Cho, Bonnie Berger, and Jian Peng. Reconstructing causal biological networks through active learning. *PloS one*, 11(3):e0150611, 2016. 1, 14, 17

[9] Chris Cundy, Aditya Grover, and Stefano Ermon. Bcd nets: Scalable variational approaches for bayesian causal discovery. *Advances in Neural Information Processing Systems*, 34, 2021. 3, 4, 14

[10] Tristan Deleu, António Góis, Chris Emezue, Mansi Rankawat, Simon Lacoste-Julien, Stefan Bauer, and Yoshua Bengio. Bayesian structure learning with generative flow networks. *arXiv preprint arXiv:2202.13903*, 2022. 14

[11] Frederick Eberhardt, Clark Glymour, and Richard Scheines. On the number of experiments sufficient and in the worst case necessary to identify all causal relations among n variables. *arXiv preprint arXiv:1207.1389*, 2012. 14

[12] P Erdős and A Rényi. On random graphs i. publicationes mathematicae (debrecen). 1959. 8, 16, 17

[13] Adam Foster, Martin Jankowiak, Eli Bingham, Paul Horsfall, Yee Whye Teh, Tom Rainforth, and Noah Goodman. Variational bayesian optimal experimental design. *arXiv preprint arXiv:1903.05480*, 2019. 2, 3, 4, 5, 13, 14

[14] Adam Foster, Martin Jankowiak, Matthew O'Meara, Yee Whye Teh, and Tom Rainforth. A unified stochastic gradient approach to designing bayesian-optimal experiments. In *International Conference on Artificial Intelligence and Statistics*, pp. 2959–2969. PMLR, 2020. 3, 4, 14

[15] Adam Foster, Desi R Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential bayesian experimental design. *arXiv preprint arXiv:2103.02438*, 2021. 3, 4, 8, 14

[16] Nir Friedman, Moises Goldszmidt, and Abraham Wyner. Data analysis with bayesian networks: A bootstrap approach. *arXiv preprint arXiv:1301.6695*, 2013. 3, 7, 14

[17] Juan L Gamella and Christina Heinze-Deml. Active invariant causal prediction: Experiment selection through stability. *arXiv preprint arXiv:2006.05690*, 2020. 14

[18] AmirEmad Ghassami, Saber Salehkaleybar, Negar Kiyavash, and Elias Bareinboim. Budgeted experiment design for causal structure learning. In *International Conference on Machine Learning*, pp. 1724–1733. PMLR, 2018. 14

[19] Alex Greenfield, Aviv Madar, Harry Ostrer, and Richard Bonneau. Dream4: Combining genetic and dynamic information to identify biological networks and dynamical models. *PloS one*, 5 (10):e13397, 2010. 2

[20] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012. 15

[21] Alain Hauser and Peter Bühlmann. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. *The Journal of Machine Learning Research*, 13(1):2409–2464, 2012. 7

[22] David Heckerman, Christopher Meek, and Gregory Cooper. A bayesian approach to causal discovery. Technical report, Technical report msr-tr-97-05, Microsoft Research, 1997. 3

[23] Xun Huan and Youssef Marzouk. Gradient-based stochastic optimization methods in bayesian experimental design. *International Journal for Uncertainty Quantification*, 4(6), 2014. 2, 4, 14

[24] Desi R Ivanova, Adam Foster, Steven Kleinegesse, Michael U Gutmann, and Thomas Rainforth. Implicit deep adaptive design: policy-based experimental design without likelihoods. *Advances in Neural Information Processing Systems*, 34:25785–25798, 2021. 3, 8, 14

[25] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016. 2, 5, 6

[26] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 2

[27] Steven Kleinegesse and Michael Gutmann. Bayesian experimental design for implicit models by mutual information neural estimation. In *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pp. 5316–5326. PMLR, 2020. 14

[28] Steven Kleinegesse and Michael U Gutmann. Efficient bayesian experimental design for implicit models. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 476–485. PMLR, 2019. 3

[29] Steven Kleinegesse and Michael U. Gutmann. Gradient-based bayesian experimental design for implicit models using mutual information lower bounds. *arXiv preprint arXiv:2105.04379*, 2021. 14

[30] Murat Kocaoglu, Alex Dimakis, and Sriram Vishwanath. Cost-optimal learning of causal graphs. In *International Conference on Machine Learning*, pp. 1875–1884. PMLR, 2017. 14

[31] Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Experimental design for learning causal graphs with latent variables. *Advances in Neural Information Processing Systems*, 30, 2017. 14

[32] Andreas Krause and Daniel Golovin. Submodular function maximization. *Tractability*, 3: 71–104, 2014. 3

[33] Vincent Lim, Ellen Novoseller, Jeffrey Ichnowski, Huang Huang, and Ken Goldberg. Policy-based bayesian experimental design for non-differentiable implicit models. *arXiv preprint arXiv:2203.04272*, 2022. 14

[34] Erik Lindgren, Murat Kocaoglu, Alexandros G Dimakis, and Sriram Vishwanath. Experimental design for cost-aware learning of causal graphs. *Advances in Neural Information Processing Systems*, 31, 2018. 14

[35] Dennis V Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pp. 986–1005, 1956. 1, 3

[36] Lars Lorch, Jonas Rothfuss, Bernhard Schölkopf, and Andreas Krause. Dibs: Differentiable bayesian structure learning. *arXiv preprint arXiv:2105.11839*, 2021. 3, 4, 14

[37] Chris J Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712*, 2016. 2, 5, 6

[38] Nicolai Meinshausen, Alain Hauser, Joris M Mooij, Jonas Peters, Philip Versteeg, and Peter Bühlmann. Methods for causal inference from gene perturbation experiments and validation. *Proceedings of the National Academy of Sciences*, 113(27):7361–7368, 2016. 2

[39] Ehsan Mokhtarian, Saber Salehkaleybar, AmirEmad Ghassami, and Negar Kiyavash. A unified experiment design approach for cyclic and acyclic causal models. *arXiv preprint arXiv:2205.10083*, 2022. 14

[40] Kevin P Murphy. Active learning of causal bayes net structure. 2001. 1, 14, 17

[41] Jay I Myung, Daniel R Cavagnaro, and Mark A Pitt. A tutorial on adaptive design optimization. *Journal of mathematical psychology*, 57(3-4):53–67, 2013. 2, 4

[42] Mizu Nishikawa-Toomey, Tristan Deleu, Jithendaraa Subramanian, Yoshua Bengio, and Laurent Charlin. Bayesian learning of causal structure and mechanisms with gflownets and variational bayes. *arXiv preprint arXiv:2211.02763*, 2022. 14

[43] Mateusz Olko, Michał Zając, Aleksandra Nowak, Nino Scherrer, Yashas Annadani, Stefan Bauer, Łukasz Kuciński, and Piotr Miłoś. Trust your ∇: Gradient-based intervention targeting for causal discovery. *arXiv preprint arXiv:2211.13715*, 2022. 14

[44] Judea Pearl. *Causality*. Cambridge university press, 2009. 2

[45] Jonas Peters, Joris Mooij, Dominik Janzing, and Bernhard Schölkopf. Identifiability of causal graphs using functional models. *arXiv preprint arXiv:1202.3757*, 2012. 3

[46] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017. 2

[47] Tom Rainforth, Rob Cornish, Hongseok Yang, Andrew Warrington, and Frank Wood. On nesting monte carlo estimators. In *International Conference on Machine Learning*, pp. 4267–4276. PMLR, 2018. 3, 4

[48] Kenneth J Ryan. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *Journal of Computational and Graphical Statistics*, 12 (3):585–603, 2003. 2, 4

[49] 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. 2

[50] Nino Scherrer, Olexa Bilaniuk, Yashas Annadani, Anirudh Goyal, Patrick Schwab, Bernhard Schölkopf, Michael C Mozer, Yoshua Bengio, Stefan Bauer, and Nan Rosemary Ke. Learning neural causal models with active interventions. *arXiv preprint arXiv:2109.02429*, 2021. 8, 14

[51] Xinpeng Shen, Sisi Ma, Prashanthi Vemuri, and Gyorgy Simon. Challenges and opportunities with causal discovery algorithms: application to alzheimer's pathophysiology. *Scientific reports*, 10(1):2975, 2020. 2

[52] Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2000. 3

[53] Scott Sussex, Caroline Uhler, and Andreas Krause. Near-optimal multi-perturbation experimental design for causal structure learning. *Advances in Neural Information Processing Systems*, 34:777–788, 2021. 1, 2, 4, 7, 8, 14, 16, 17

[54] Panagiotis Tigas, Yashas Annadani, Andrew Jesson, Bernhard Schölkopf, Yarin Gal, and Stefan Bauer. Interventions, where and how? experimental design for causal models at scale. *Advances in Neural Information Processing Systems*, 35, 2022. 1, 2, 3, 4, 8, 14, 16, 17

[55] Simon Tong and Daphne Koller. Active learning for structure in bayesian networks. In *International joint conference on artificial intelligence*, volume 17, pp. 863–869. Citeseer, 2001. 1, 4, 14, 17

[56] Christian Toth, Lars Lorch, Christian Knoll, Andreas Krause, Franz Pernkopf, Robert Peharz, and Julius von Kügelgen. Active bayesian causal inference. *arXiv preprint arXiv:2206.02063*, 2022. 1, 8, 14, 16, 17

[57] Julius von Kügelgen, Paul K Rubenstein, Bernhard Schölkopf, and Adrian Weller. Optimal experimental design via bayesian optimization: active causal structure learning for gaussian process networks. *arXiv preprint arXiv:1910.03962*, 2019. 14

[58] Sang Michael Xie and Stefano Ermon. Reparameterizable subset sampling via continuous relaxations. *arXiv preprint arXiv:1901.10517*, 2019. 6

APPENDIX A DERIVATION OF IMPORTANCE WEIGHTED NESTED MONTE CARLO ESTIMATOR

In this section, we derive the $\mathcal{U}_{\text{IWNMC}}$ (Eq. 4) estimator. We derive the estimator for a single design with an experiment denoted by ξ, parameters $\boldsymbol{\theta}$ and experimental outcome random variable $\mathbf{Y}$ and its instance $\mathbf{y}$. Since it is a static design, all the steps of the derivation hold if we replace ξ with $\xi_{1:B}$, $\mathbf{Y}$ with $\mathbf{Y_{1:B}}$ and $\mathbf{y}$ with $\mathbf{y_{1:B}}$. We begin from the variational NMC (VNMC) estimator, introduced by Foster et al. [13]

$$\mathcal{I}(\mathbf{Y};\boldsymbol{\Theta} \mid \xi) \leq \mathcal{U}_{\text{VNMC}}(\xi) = \underset{\substack{p(\boldsymbol{\theta}_0|h_{t-1}) \\ p(\mathbf{y}|\boldsymbol{\theta}_0,\xi) \\ q(\boldsymbol{\theta}_{1:L}|h_{t-1},\mathbf{y})}}{\mathbb{E}} \left[\log \frac{p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_0)}{\frac{1}{L}\sum_{\ell=1}^{L} \frac{p(\mathbf{y}|\xi,\boldsymbol{\theta}_\ell)p(\boldsymbol{\theta}_\ell|h_{t-1})}{q(\boldsymbol{\theta}_{1:L}|h_{t-1},\mathbf{y})}} \right]. \tag{5}$$

This can be rewritten as

$$\mathcal{U}_{\text{VNMC}}(\xi) = \underset{\substack{p(\boldsymbol{\theta}_0|h_{t-1}) \\ p(\mathbf{y}|\boldsymbol{\theta}_0,\xi) \\ q(\boldsymbol{\theta}_{1:L}|h_{t-1},\mathbf{y})}}{\mathbb{E}} \left[\log \frac{p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_0)}{\frac{1}{L}\sum_{\ell=1}^{L} \frac{p(\mathbf{y}|\xi,\boldsymbol{\theta}_\ell)p(\boldsymbol{\theta}_\ell)p(h_{t-1}|\boldsymbol{\theta}_\ell)}{q(\boldsymbol{\theta}_{1:L}|h_{t-1},\mathbf{y})}} + \log p(h_{t-1}) \right]. \tag{6}$$

and Foster et al. [13] observed that $\log p(h_{t-1})$ is a constant that does not depend on ξ and so can be safely neglected when optimizing over designs. If we take the original prior $p(\boldsymbol{\theta}_\ell)$ as our proposal distribution q, then we arrive at

$$\mathcal{U}_{\text{VNMC-prior}}(\xi) = \underset{\substack{p(\boldsymbol{\theta}_0|h_{t-1})p(\boldsymbol{\theta}_{1:L}) \\ p(\mathbf{y}|\boldsymbol{\theta}_0,\xi)}}{\mathbb{E}} \left[\log \frac{p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_0)}{\frac{1}{L}\sum_{\ell=1}^{L} p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_\ell)p(h_{t-1} \mid \boldsymbol{\theta}_\ell)} \right] + C \tag{7}$$

where $C = \log p(h_{t-1})$. This allows us to sample contrastive samples from any distribution, but does not account for $\boldsymbol{\theta}_0$. If we were to sample $\boldsymbol{\theta}_0$ from $p(\boldsymbol{\theta}_0)$, we can correct using an importance weight

$$\mathcal{U}_{\text{VNMC-prior}}(\xi) = \underset{\substack{p(\boldsymbol{\theta}_{0:L}) \\ p(\mathbf{y}|\boldsymbol{\theta}_0,\xi)}}{\mathbb{E}} \left[\frac{p(\boldsymbol{\theta}_0 \mid h_{t-1})}{p(\boldsymbol{\theta}_0)} \log \frac{p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_0)}{\frac{1}{L}\sum_{\ell=1}^{L} p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_\ell)p(h_{t-1} \mid \boldsymbol{\theta}_\ell)} \right] + C, \tag{8}$$

but unfortunately, this relies on knowing the density of the posterior or using the fact that $p(\boldsymbol{\theta}_0 \mid h_{t-1})/p(\boldsymbol{\theta}_0) = p(h_{t-1} \mid \boldsymbol{\theta}_0)/p(h_{t-1})$, knowing the marginal likelihood of the data h_{t-1}. Neither of these is usually tractable. Instead, we can use a self-normalized importance sampling approach, which amounts to estimating $p(h_{t-1})$ by a sum over $\theta_{0:L}$, giving the *approximation* IWNMC:

$$\mathcal{U}_{\text{IWNMC}}(\xi) = \underset{\substack{p(\boldsymbol{\theta}_{0:L}) \\ p(\mathbf{y}|\boldsymbol{\theta}_0,\xi)}}{\mathbb{E}} \left[\frac{p(h_{t-1} \mid \boldsymbol{\theta}_0)}{\frac{1}{L+1}\sum_{k=0}^{L} p(h_{t-1} \mid \boldsymbol{\theta}_k)} \log \frac{p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_0)}{\frac{1}{L}\sum_{\ell=1}^{L} p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_\ell)p(h_{t-1} \mid \boldsymbol{\theta}_\ell)} \right] + C. \tag{9}$$

The form that is given in equation 4 is obtained by first relabelling the $\boldsymbol{\theta}$ samples to start from 1

$$\mathcal{U}_{\text{IWNMC}}(\xi) = \underset{\substack{p(\boldsymbol{\theta}_{1:L}) \\ p(\mathbf{y}|\boldsymbol{\theta}_1,\xi)}}{\mathbb{E}} \left[\frac{p(h_{t-1} \mid \boldsymbol{\theta}_1)}{\frac{1}{L}\sum_{k=1}^{L} p(h_{t-1} \mid \boldsymbol{\theta}_k)} \log \frac{p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_1)}{\frac{1}{L-1}\sum_{\ell=2}^{L} p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_\ell)p(h_{t-1} \mid \boldsymbol{\theta}_\ell)} \right] + C, \tag{10}$$

noting that the role of $\boldsymbol{\theta}_1$ is arbitrary and can be replaced by any $m \in \{1, \ldots, L\}$

$$\mathcal{U}_{\text{IWNMC}}(\xi) = \underset{\substack{p(\boldsymbol{\theta}_{1:L}) \\ p(\mathbf{y}|\boldsymbol{\theta}_m,\xi)}}{\mathbb{E}} \left[\frac{p(h_{t-1} \mid \boldsymbol{\theta}_m)}{\frac{1}{L}\sum_{k=1}^{L} p(h_{t-1} \mid \boldsymbol{\theta}_k)} \log \frac{p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_m)}{\frac{1}{L-1}\sum_{\ell\neq m}^{L} p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_\ell)p(h_{t-1} \mid \boldsymbol{\theta}_\ell)} \right] + C, \tag{11}$$

and finally taking the mean over m, noting that this does not change the expected value due to linearity

$$\mathcal{U}_{\text{IWNMC}}(\xi) = \underset{\substack{p(\boldsymbol{\theta}_{1:L}) \\ p(\mathbf{y}|\boldsymbol{\theta}_m,\xi)}}{\mathbb{E}} \left[\sum_{m=1}^{L} \frac{p(h_{t-1} \mid \boldsymbol{\theta}_m)}{\sum_{k=1}^{L} p(h_{t-1} \mid \boldsymbol{\theta}_k)} \log \frac{p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_m)}{\frac{1}{L-1}\sum_{\ell\neq m}^{L} p(\mathbf{y} \mid \xi, \boldsymbol{\theta}_\ell)p(h_{t-1} \mid \boldsymbol{\theta}_\ell)} \right] + C. \tag{12}$$

We finally drop the constant C as it is independent of ξ and take

$$\omega_m = \frac{p(h_{t-1} \mid \boldsymbol{\theta}_m)}{\sum_{k=1}^{L} p(h_{t-1} \mid \boldsymbol{\theta}_k)}. \tag{13}$$

APPENDIX B RELATED WORK

Differentiable Bayesian Optimal Experimental Design. Huan & Marzouk [23], Foster et al. [13; 14], Kleinegesse & Gutmann [27; 29] developed a unified framework for estimating Expected Information Gain and optimizing the designs with gradient-based methods. In Ivanova et al. [24], Foster et al. [15], the authors introduced a policy-based method for performing adaptive experimentation. More recently, work like Blau et al. [6], Lim et al. [33] used Reinforcement Learning to train policies for adaptive experimental design.

Experimental Design for Causal Discovery. One of the earliest works of experimental design for causal discovery in a BOED setting was proposed by [40] and [55] in the case of discrete variables for single target acquisition. Since then, a number of works have attempted to address this problem for continuous variables in both the BOED framework [2, 57, 56, 8] and other frameworks [30, 17, 11, 34, 39, 18, 43, 50]. In contrast to the setting studied in this paper, of particular note, are the approaches for experimental design for causal discovery in a non-BOED setting in the presence of cycles [39] and latent variables [31]. Closer to our BOED setting are the approaches of [54] and [53]. Specifically, in [54], the authors introduce a method for selecting single target-state pair with stochastic batch acquisition while [53] introduce a method for selecting a batch of multi-target experiments with a greedy strategy, based on a gradient-based approximation to mutual information, without selecting the intervention state. Our presented method in contrast can acquire a batch of multi-target-state pairs.

Bayesian Causal Discovery. While our work is not directly concerned with Bayesian causal discovery as such, our design procedure method benefits from using the approximate posteriors for causal models to estimate mutual information and design interventions [16, 3, 36, 9, 10, 42].

APPENDIX C   EXPECTED INFORMATION GAIN FOR 6 NODES AND BATCH SIZE 2
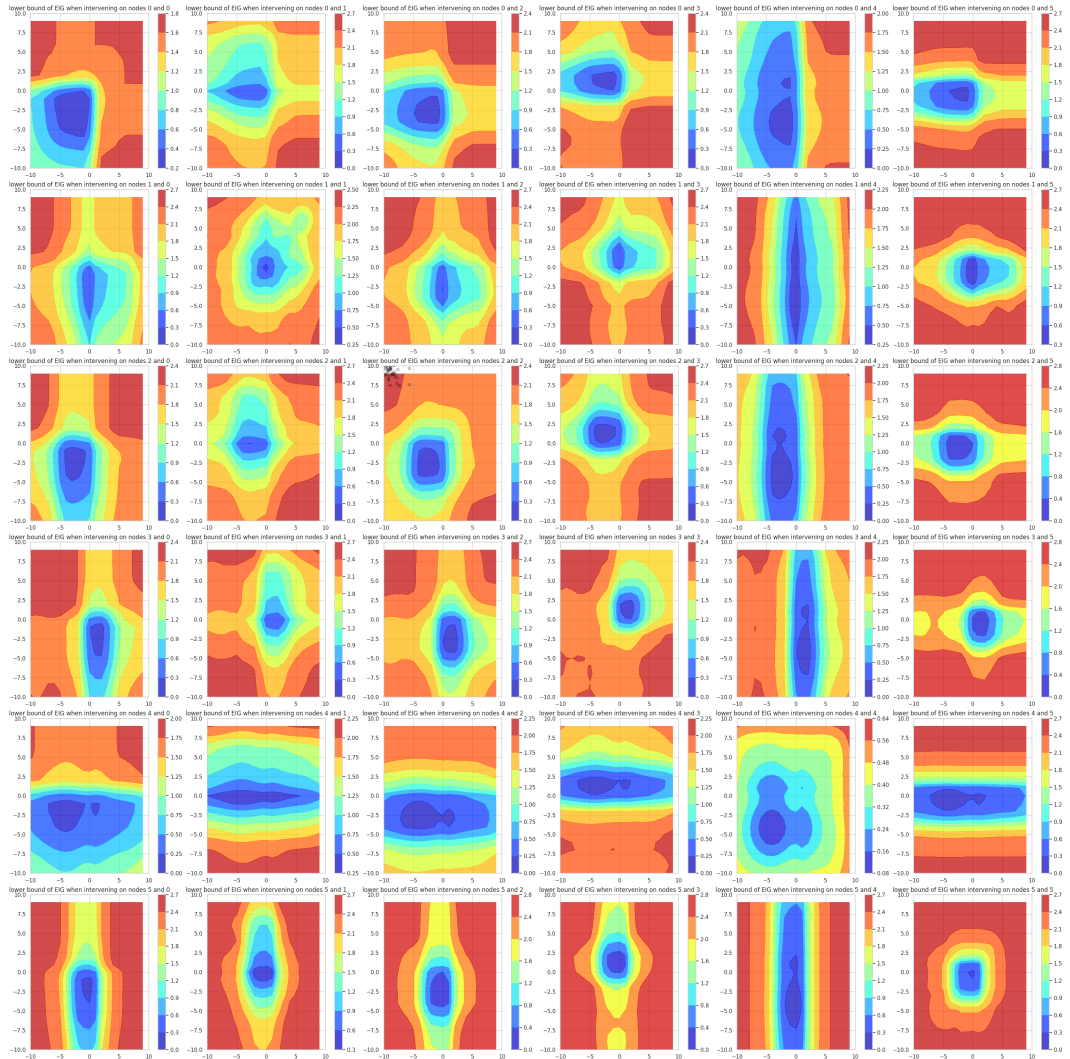
Figure 5: Here we visualize the Expected Information Gain of batch size two, on two nodes over different interventional values of the range $[-10, 10]$.

APPENDIX D   METRICS

.

$\mathbb{E}$-**SHD**: Defined as the *expected structural hamming distance* between samples from the posterior model over graphs and the true graph $\mathbb{E}\text{-SHD} := \mathbb{E}_{\mathbf{g} \sim p(\mathcal{G}|\mathcal{D})} \big[\text{SHD}(\mathbf{g}, \check{\mathbf{g}}) \big]$

Expected edges F1: The expected F1 score of the binary classification task of predicting the presence/ absence of all edges. The expectation is taken over multiple posterior samples.

i-MMD: Interventional MMD is defined as MMD distance [20] between the true interventional distribution and the interventional distribution induced by θ and $\mathbf{g}$ (posterior sample). We take an expectation over different posterior samples, interventional targets and interventional states. For the kernel choice, we use the median heuristic as described in [20].

APPENDIX E   BASELINES

Before we evaluate the IWNMC estimator with a proposal distribution more informative than the prior and the NMC estimator with a posterior estimate of SCM, we present the baselines with which we can also compare the overall performance of our designs.

SINGLE TARGET

Random-Fixed: Uniform random selection of node, fixing the state to a state to 0 (as introduced in [2, 54]). **Random-Random**: Uniform selection of node, uniform selection of state (introduced in [56]). **SoftCBED**: A stochastic approximation of greedy batch selection as introduced in [54].
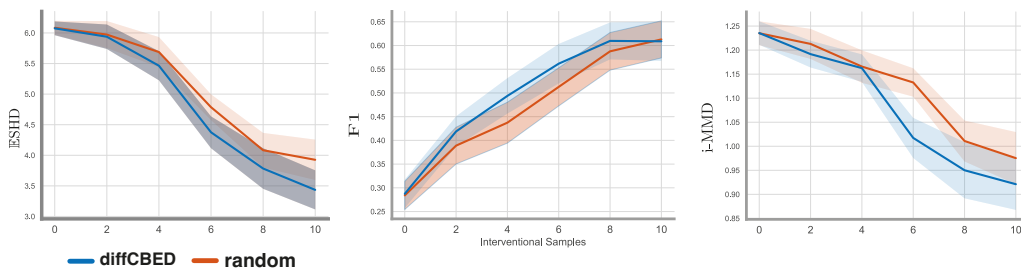
MULTI-TARGET

Random-Random: Multitarget version of Uniform selection of node, uniform selection of state (introduced in [56]). **Random-Fixed**: Multitarget version of Uniform selection of node, fixed state to 5 [53], as suggested by the authors. **SSGb**: Finite sample baseline from [53] with fixed state equal 5. We emphasize that in contrast to our method, the baselines cannot select states, but they either assume a fixed predefined value or select a value at random.

APPENDIX F   DAG BOOTSTRAP

The DAG bootstrap bootstraps observations and interventions to infer a different causal structure per bootstrap. We used GIES as the causal inference algorithm because of the adaptation of GES on interventional data as well. In our experiments, we used the pcalg R implementation `https://github.com/cran/pcalg/blob/master/R/gies.R` to discover 100 graphs. Each graph can be seen as a posterior sample from $p(\mathbf{G} \mid h_{t-1})$. For each of the sampled graphs G_i we compute the appropriate θ_{MLE} under linear Gaussian assumption for the conditional distributions.

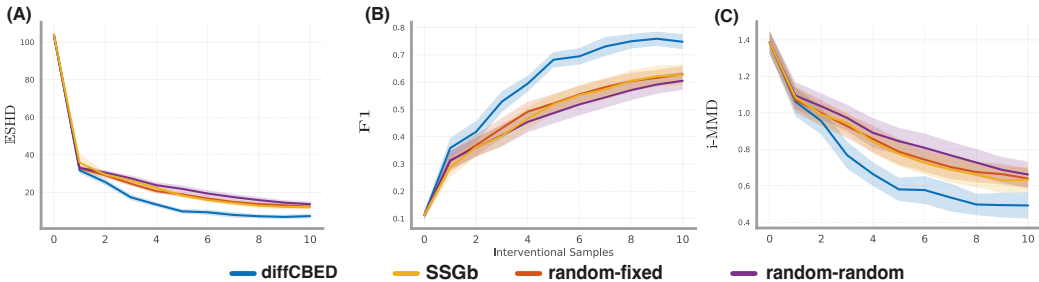## APPENDIX G   IMPORTANCE WEIGHTED NESTED MONTE CARLO FULL RESULTS

Figure 6: Multi target-state design setting results for Erdős–Rényi [12] graphs with $d = 5$ variables. Each experiment was run with 60 random seeds (shaded area represents 95% CIs)
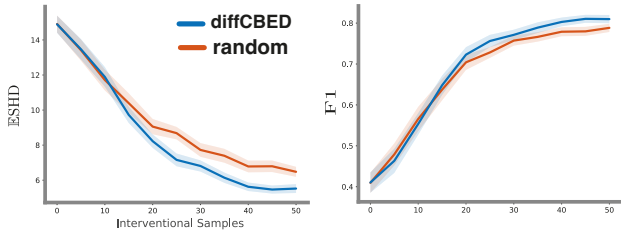
Figure 7: Single target-state design setting results for Erdős–Rényi [12] graphs with $d = 10$ variables. Each experiment was run with 60 random seeds (shaded area represents 95% CIs)

APPENDIX H   20 NODES, UNCONSTRAINED $(q \leq 20)$, BATCH SIZE $B = 1$:

Figure 8: Multi target-state design setting results for Erdős–Rényi [12] graphs with $d = 50$ variables. Each experiment was run with 30 random seeds (the shaded area represents 95% CIs). We observe that for batch size 1, the difference between the methods becomes more significant.

APPENDIX I   DATASETS AND EXPERIMENT DETAILS

I.1   SYNTHETIC GRAPHS EXPERIMENTS

In the synthetic data experiments, we focus on Erdős-Rényi graph model. We used `networkx`[4] and method `fast_gnp_random_graph` [4] to generate graphs based on the Erdős-Rényi model. We set the expected number of edges per vertex to 1.

APPENDIX J   TABLE SUMMARIZING PRIOR WORK

Table 2: Comparison of different BOED for Causal Discovery methods based on their design space assumptions.

	Design Space Assumptions				
	Node Acquisition (Single Target)	Value Acquisition (Single Target)	Node Acquisition (Multi-target)	Value Acquisition (Multi-target)	Batch Acquisition
Murphy [40]	✓				
Tong & Koller [55]	✓				
Cho et al. [8]	✓				
Agrawal et al. [2]	✓				✓
Toth et al. [56]	✓	✓			
Tigas et al. [54]	✓	✓			✓
Sussex et al. [53]	✓		✓		✓
Ours	✓	✓	✓	✓	✓

---

[4]https://networkx.org/documentation/networkx-1.10/reference/generated/
networkx.generators.random_graphs.fast_gnp_random_graph.html

APPENDIX K   OPTIMIZER SETTINGS

Table 3: Table indicating the hyperparameters and optimizer settings for different experimental results.

	Optimization settings			
	Single Target NMC	Multi-Target NMC	Multi-Target IWNMC with prior	Multi-Target IWNMC with proposal
L	30	30	1000	60
Number of outer DAGs N_o	30	30	1000	60
Batch Size	5	2	2	2
Relaxation temperature	$5 \to .5$	$5 \to .5$	0.1	$5 \to .5$
Optimizer	Adam	Adam	Adam	Adam
Learning rate of optimizer	0.1	0.1	0.01	0.1
Number of starting samples (observational)	60	60	2	800
Number of batches	10	10	5	1
Number of DAG Bootstraps	30	30	-	60
Number of training steps per batch	100	100	100	100