Combinatorial Causal Bandits

Shi Feng,[1] **Wei Chen**[2]

[1]Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China
[2]Microsoft Research, Beijing, China
shifeng-thu@outlook.com, weic@microsoft.com

Abstract

In combinatorial causal bandits (CCB), the learning agent chooses at most K variables in each round to intervene, collects feedback from the observed variables, with the goal of minimizing expected regret on the target variable Y. We study under the context of binary generalized linear models (BGLMs) with a succinct parametric representation of the causal models. We present the algorithm BGLM-OFU for Markovian BGLMs (i.e. no hidden variables) based on the maximum likelihood estimation method, and show that it achieves $O(\sqrt{T}\log T)$ regret, where T is the time horizon. For the special case of linear models with hidden variables, we apply causal inference techniques such as the do-calculus to convert the original model into a Markovian model, and then show that our BGLM-OFU algorithm and another algorithm based on the linear regression both solve such linear models with hidden variables. Our novelty includes (a) considering the combinatorial intervention action space and the general causal models including ones with hidden variables, (b) integrating and adapting techniques from diverse studies such as generalized linear bandits and online influence maximization, and (c) avoiding unrealistic assumptions (such as knowing the joint distribution of the parents of Y under all interventions) and regret factors exponential to causal graph size in prior studies.

1 Introduction

Causal bandit problem is first introduced in (Lattimore, Lattimore, and Reid 2016). It consists of a causal graph $G = (\boldsymbol{X} \cup \{Y\}, E)$ indicating the causal relationship among the observed variables, where the structure of the graph is known but the underlying probability distributions governing the causal model are unknown. In each round, the learning agent selects one or a few variables in $\boldsymbol{X}$ to intervene, gains the reward as the output of Y, and observes the values of all variables in $\boldsymbol{X} \cup \{Y\}$. The goal is to either maximize the cumulative reward over T rounds, or find the intervention closest to the optimal one after T rounds. The former setting, which is the one our study focuses on, is typically translated to minimizing cumulative regret, which is defined as the difference in reward between always playing the optimal intervention and playing interventions according to a learning algorithm. Causal bandits can be applied in many settings that include causal relationships, such as medical drug testing, policy making, scientific experimental process, performance tuning, etc. Causal bandit is a form of multi-armed bandit (cf. (Lattimore and Szepesvári 2020)), with the main difference being that causal bandits may use the causal relationship and more observed feedback to achieve a better regret.

All the causal bandit studies so far (Lattimore, Lattimore, and Reid 2016; Sen et al. 2017; Nair, Patil, and Sinha 2021; Lu et al. 2020; Maiti, Nair, and Sinha 2021) focus on the case where the number of possible interventions is small. However, in many scenarios we need to intervene on a set of variables and the number of possible choices of sets is large. For example, in tuning performance of a large system, one often needs to tune a large number of parameters simultaneously to achieve the best system performance, and in drug testing, a combination of several drugs with a number of possible dosages needs to be tested for the best result. Intervening on a set of variables raises new challenges to the learning problem, since the number of possible interventions is exponentially large to the size of the causal graph. In this paper, we address this challenge and propose the new framework of combinatorial causal bandits (CCB) and its solutions. In each round of CCB, the learning agent selects a set of at most K observed variables in $\boldsymbol{X}$ to intervene instead of one variable. Other aspects including the feedback and reward remain the same.

We use the binary generalized linear models (BGLMs) to give the causal model a succinct parametric representation where all variables are binary. Using the maximum likelihood estimation (MLE) method, we design an online learning algorithm BGLM-OFU for the causal models without hidden variables (called Markovian models), and the algorithm achieves $O(\sqrt{T}\log T)$ regret, where T is the time horizon. The algorithm is based on the earlier study on generalized linear bandits (Li, Lu, and Zhou 2017), but our BGLM model is more general and thus requires new techniques (such as a new initialization phase) to solve the problem. Our regret analysis also integrates results from online influence maximization (Li et al. 2020; Zhang et al. 2022) in order to obtain the final regret bound.

Furthermore, for the binary linear models (BLMs), we show how to transform a BLM with hidden variables into one without hidden variables by utilizing the tools in causal inference such as do-calculus, and thus we can handle causal model even with hidden variables in linear models. Then, for BLMs, we show (a) the regret bound when applying the BGLM-OFU algorithm to the linear model, and (b) a new algorithm and its regret bound based on the linear regression method. We show a tradeoff between the MLE-based BGLM-OFU algorithm and the linear-regression-based algorithm on BLMs: the latter removes the assumption needed by the former but has an extra factor in the regret bound. We demonstrate effectiveness of our algorithms by experimental evaluations in Appendix H. Besides BLMs, we give similar results for linear models with continuous variables. Due to space limits, this part is put in Appendix F.

In summary, our contribution includes (a) proposing the new combinatorial causal bandit framework, (b) considering general causal models including ones with hidden variables, (c) integrating and adapting techniques from diverse studies such as generalized linear bandits and online influence maximization, and (d) achieving competitive online learning results without unrealistic assumptions and regret factors exponential to the causal graph size as appeared in prior studies. The intrinsic connection between causal bandits and online influence maximization revealed by this study may further benefit researches in both directions.

2 Related Works

Causal Bandits. The causal bandit problem is first defined in (Lattimore, Lattimore, and Reid 2016). The authors introduce algorithms for a specific parallel model and more general causal models to minimize the simple regret, defined as the gap between the optimal reward and the reward of the action obtain after T rounds. A similar algorithm to minimize simple regret has also been proposed in (Nair, Patil, and Sinha 2021) for a graph with no backdoor. To optimally trade-off observations and interventions, they have also discussed the budgeted version of causal bandit when interventions are more expensive than observations. Sen et al. (2017) use the importance sampling method to propose an algorithm to optimize simple regret when only soft intervention on a single node is allowed. These studies all focus on simple regret for the pure exploration setting, while our work focuses on cumulative regret. There are a few studies on the cumulative regret of causal bandits (Lu et al. 2020; Nair, Patil, and Sinha 2021; Maiti, Nair, and Sinha 2021). However, the algorithms in (Lu et al. 2020) and (Nair, Patil, and Sinha 2021) are both under an unrealistic assumption that for every intervention, the joint distribution of the parents of the reward node Y is known. The algorithm in (Maiti, Nair, and Sinha 2021) does not use this assumption, but it only works on Markovian models without hidden variables. The regrets in (Lu et al. 2020; Maiti, Nair, and Sinha 2021) also have a factor exponential to the causal graph size. Yabe et al. (2018) designs an algorithm by estimating each $P(X|\boldsymbol{Pa}(X))$ for $X \in \boldsymbol{X} \cup \{Y\}$ that focuses on simple regret but works for combinatorial settings. However, it requires the round number $T \geq \sum_{X \in \boldsymbol{X}} 2^{|\boldsymbol{Pa}(X)|}$, which is still unrealistic. We avoid this by working on the BGLM model with a linear number of parameters, and it results in completely different techniques. Lee and Bareinboim (2018, 2019, 2020) also consider the combinatorial settings, but they focus on empirical studies, while we provide theoretical regret guarantees. Furthermore, studying causal bandit problem without the full casual structure is an important future direction. One such study, (Lu, Meisami, and Tewari 2021), exists but it is only for the atomic setting and has a strong assumption that $|\boldsymbol{Pa}(Y)| = 1$.

Combinatorial Multi-Armed Bandits. CCB can be viewed as a type of combinatorial multi-armed bandits (CMAB) (Chen, Wang, and Yuan 2013; Chen et al. 2016), but the feedback model is not the semi-bandit model studied in (Chen, Wang, and Yuan 2013; Chen et al. 2016). In particular, in CCB individual random variables cannot be treated as base arms, because each intervention action changes the distribution of the remaining variables, violating the i.i.d assumption for base arm random variables across different rounds. Interestingly, it has a closer connection to recent studies on online influence maximization (OIM) with node-level feedback (Li et al. 2020; Zhang et al. 2022). These studies consider influence cascades in a social network following either the independent cascade (IC) or the linear threshold (LT) model. In each round, one selects a set of at most K seed nodes, observes the cascade process as which nodes are activated in which steps, and obtains the reward as the total number of nodes activated. The goal is to learn the cascade model and minimize the cumulative regret. Influence propagation is intrinsically a causal relationship, and thus OIM has some intrinsic connection with CCB, and our study does borrow techniques from (Li et al. 2020; Zhang et al. 2022). However, there are a few key differences between OIM and CCB, such that we need adaptation and integration of OIM techniques into our setting: (a) OIM does not consider hidden variables, while we do consider hidden variables for causal graphs; (b) OIM allows the observation of node activations at every time step, but in CCB we only observe the final result of the variables; and (c) current studies in (Li et al. 2020; Zhang et al. 2022) only consider IC and LT models, while we consider the more general BGLM model, which includes IC model (see (Zhang et al. 2022) for transformation from IC model to BGLM) and LT model as two special cases.

Linear and Generalized Linear Bandits. Our work is also based on some previous online learning studies on linear and generalized linear bandits. Abbasi-Yadkori, Pál, and Szepesvári (2011) propose an improved theoretical regret bound for the linear stochastic multi-armed bandit problem. Some of their proof techniques are used in our proofs. Li, Lu, and Zhou (2017) propose an online algorithm and its analysis based on MLE for generalized linear contextual bandits, and our MLE method is adapted from this study. However, our setting is much more complicated, in that (a) we have a combinatorial action space, and (b) we have a network of generalized linear relationships while they only consider one generalized linear relationship. As the result, our algorithm and analysis are also more sophisticated.

3 Model and Preliminaries

Following the convention in causal inference literature (e.g. (Pearl 2009)), we use capital letters $(U, X, Y \ldots)$ to represent variables, and their corresponding lower-case letters to represent their values. We use boldface letters such as $\boldsymbol{X}$ and $\boldsymbol{x}$ to represent a set or a vector of variables or values.

A *causal graph* $G = (\boldsymbol{U} \cup \boldsymbol{X} \cup \{Y\}, E)$ is a directed acyclic graph where $\boldsymbol{U} \cup \boldsymbol{X} \cup \{Y\}$ are sets of nodes with $\boldsymbol{U}$ being the set of unobserved or hidden nodes, $\boldsymbol{X} \cup \{Y\}$ being the set of observed nodes, Y is a special target node with no outgoing edges, and E is the set of directed edges connecting nodes in $\boldsymbol{U} \cup \boldsymbol{X} \cup \{Y\}$. For simplicity, in this paper we consider all variables in $\boldsymbol{U} \cup \boldsymbol{X} \cup \{Y\}$ are $(0,1)$-binary random variables. For a node X in the graph G, we call its in-neighbor nodes in G the *parents* of X, denoted as $\boldsymbol{Pa}(X)$, and the values taken by these parent random variables are denoted $\boldsymbol{pa}(X)$.

Following the causal Bayesian model, the causal influence from the parents of X to X is modeled as the probability distribution $P(X|\boldsymbol{Pa}(X))$ for every possible value combination of $\boldsymbol{Pa}(X)$. Then, for each X, the full non-parametric characterization of $P(X|\boldsymbol{Pa}(X))$ requires $2^{|\boldsymbol{Pa}(X)|}$ values, which would be difficult for learning. Therefore, we will describe shortly the succinct parametric representation of $P(X|\boldsymbol{Pa}(X))$ as a generalized linear model to be used in this paper.

The causal graph G is *Markovian* if there are no hidden variables in G and every observed variable X has certain randomness not caused by other observed parents. To model this effect of the Markovian model, in this paper we dedicate random variable X_1 to be a special variable that always takes the value 1 and is a parent of all other observed random variables. We study the Markovian causal model first, and in Section 5 we will consider causal models with more general hidden variable forms.

An *intervention* in the causal model G is to force a subset of observed variables $\boldsymbol{S} \subseteq \boldsymbol{X}$ to take some predetermined values $\boldsymbol{s}$, to see its effect on the target variable Y. The intervention on $\boldsymbol{S}$ to Y is denoted as $\mathbb{E}[Y|do(\boldsymbol{S}=\boldsymbol{s})]$. In this paper, we study the selection of $\boldsymbol{S}$ to maximize the expected value of Y, and our parametric model would have the monotonicity property such that setting a variable X to 1 is always better than setting it to 0 in maximizing $\mathbb{E}[Y]$, so our intervention would always be setting $\boldsymbol{S}$ to all 1's, for which we simply denote as $\mathbb{E}[Y|do(\boldsymbol{S})]$.

In this paper, we study the online learning problem of *combinatorial causal bandit*, as defined below. A learning agent runs an algorithm π for T rounds. Initially, the agent knows the observed part of the causal graph G induced by observed variables $\boldsymbol{X} \cup \{Y\}$, but does not know the probability distributions $P(X|\boldsymbol{Pa}(X))$'s. In each round $t = 1, 2, \ldots, T$, the agent selects at most K observed variables in $\boldsymbol{X}$ for intervention, obtains the observed Y value as the reward, and observes all variable values in $\boldsymbol{X} \cup \{Y\}$ as the feedback. The agent needs to utilize the feedback from the past rounds to adjust its action in the current round, with the goal of maximizing the cumulative reward from all T rounds.

The performance of the agent is measured by the *regret* of the algorithm π, which is defined as the difference between the expected cumulative reward of the best action $\boldsymbol{S}^*$ and the cumulative reward of algorithm π, where $\boldsymbol{S}^* \in \operatorname{argmax}_{\boldsymbol{S} \subseteq \boldsymbol{X}, |\boldsymbol{S}|=K} \mathbb{E}[Y|do(\boldsymbol{S})]$, as given below:

$$R^{\pi}(T) = \mathbb{E}\left[\sum_{t=1}^{T}(\mathbb{E}[Y|do(\boldsymbol{S}^*)] - \mathbb{E}[Y|do(\boldsymbol{S}_t^{\pi})])\right], \quad (1)$$

where $\boldsymbol{S}_t^{\pi}$ is the intervention set selected by algorithm π in round t, and the expectation is taking from the randomness of the causal model as well as the possible randomness of the algorithm π. In some cases online learning algorithm uses a standard offline oracle that takes the causal graph G and the distributions $P(X|\boldsymbol{Pa}(X))$'s as inputs and outputs a set of nodes $\boldsymbol{S}$ that achieves an α-approximation with probability β with $\alpha, \beta \in (0,1]$. In this case, one could consider the (α,β)-approximation regret, as in (Chen et al. 2016):

$$R^{\pi}_{\alpha,\beta}(T) = \mathbb{E}\left[\sum_{t=1}^{T}(\alpha\beta\mathbb{E}[Y|do(\boldsymbol{S}^*)] - \mathbb{E}[Y|do(\boldsymbol{S}_t^{\pi})])\right]. \quad (2)$$

As pointed out earlier, the non-parametric form of distributions $P(X|\boldsymbol{Pa}(X))$'s needs an exponential number of quantities and is difficult to learn. In this paper, we adopt the general linear model as the parametric model, which is widely used in causal inference literature (Hernán and Robins 2010; Garcia-Huidobro and Michael Oakes 2017; Han, Yu, and Friedberg 2017; Arnold et al. 2020; Vansteelandt and Dukes 2020). Since we consider binary random variables, we refer to such models as binary general linear models (BGLMs). In BGLM, the functional relationship between a node X and its parents $\boldsymbol{Pa}(X)$ in G is $P(X=1|\boldsymbol{Pa}(X)=\boldsymbol{pa}(X)) = f_X(\boldsymbol{\theta}^*_X \cdot \boldsymbol{pa}(X)) + \varepsilon_X$, where $\boldsymbol{\theta}^*_X$ is the unknown weight vector in $[0,1]^{|\boldsymbol{Pa}(X)|}$, ε_X is a zero-mean sub-Gaussian noise that ensures that the probability does not exceed 1, $\boldsymbol{pa}(X)$ here is the vector form of the values of parents of X, and f_X is a scalar and monotonically non-decreasing function. It is worth noticing that our BGLM here is a binary version of traditional generalized linear models (Aitkin et al. 2009; Hilbe 2011; Sakate and Kashid 2014; Hastie and Pregibon 2017). Instead of letting $X = f_X(\boldsymbol{\theta}^*_X \cdot \boldsymbol{pa}(X)) + \varepsilon_X$ directly, we take $f_X(\boldsymbol{\theta}^*_X \cdot \boldsymbol{pa}(X)) + \varepsilon_X$ as the probability for X to be 1. The vector of all the weights in a BGLM is denoted by $\boldsymbol{\theta}^*$ and the feasible domain for the weights is denoted by Θ. We use the notation $\theta^*_{Z,X}$ to denote the parameter in $\boldsymbol{\theta}^*$ that corresponds to the edge (Z,X), or equivalently the entry in vector $\boldsymbol{\theta}^*_X$ that corresponds to node Z. We also use notation $\boldsymbol{\varepsilon}$ to represent all noise random variables $(\varepsilon_X)_{X \in \boldsymbol{X} \cup Y}$.

A special case of BGLM is the linear model where function f_X is the identity function for all X's, then $P(X=1|\boldsymbol{Pa}(X)=\boldsymbol{pa}(X)) = \boldsymbol{\theta}^*_X \cdot \boldsymbol{pa}(X) + \varepsilon_X$. We refer to this model as BLM. Moreover, when we remove the noise variable ε_X, BLM coincides with the *linear threshold (LT)* model for influence cascades (Kempe, Kleinberg, and Tardos 2003) in a DAG. In the LT model, each node X has a random threshold λ_X uniformly drawn from $[0,1]$, and each edge (Z,X) has a weight $w_{Z,X} \in [0,1]$, such that node X is activated (equivalent to X being set to 1) when the cumulative weight of its active in-neighbors is at least λ_X. It is easy to see that when we set $\theta^*_{Z,X} = w_{Z,X}$, the activation condition is translated to the probability of $X = 1$ being exactly $\boldsymbol{\theta}^*_X \cdot \boldsymbol{pa}(X)$. It is not surprising that a linear causal model is equivalent to an influence cascade model, since the influence relationship is intrinsically a causal relationship.

4 Algorithm for BGLM CCB

In this section, we present an algorithm that solves the online CCB problem for the Markovian BGLM. The algorithm requires three assumptions.

Assumption 1. *For every $X \in \boldsymbol{X} \cup \{Y\}$, f_X is twice differentiable. Its first and second order derivatives are upper-bounded by $L^{(1)}_{f_X} > 0$ and $L^{(2)}_{f_X} > 0$.*

Let $\kappa = \inf_{X \in \boldsymbol{X} \cup \{Y\}, \boldsymbol{v} \in [0,1]^{|\boldsymbol{Pa}(X)|}, ||\boldsymbol{\theta} - \boldsymbol{\theta}^*_X|| \leq 1} \dot{f}_X(\boldsymbol{v} \cdot \boldsymbol{\theta})$.

Assumption 2. *We have $\kappa > 0$.*

Assumption 3. *There exists a constant $\zeta > 0$ such that for any $X \in \boldsymbol{X} \cup \{Y\}$ and $X' \in \boldsymbol{Pa}(X)$, for any value vector $\boldsymbol{v} \in \{0,1\}^{|\boldsymbol{Pa}(X) \setminus \{X', X_1\}|}$, the following inequalities hold:*

$$\Pr_{\boldsymbol{\varepsilon}, \boldsymbol{X}, Y} \{X' = 1 | \boldsymbol{Pa}(X) \setminus \{X', X_1\} = \boldsymbol{v}\} \geq \zeta, \qquad (3)$$

$$\Pr_{\boldsymbol{\varepsilon}, \boldsymbol{X}, Y} \{X' = 0 | \boldsymbol{Pa}(X) \setminus \{X', X_1\} = \boldsymbol{v}\} \geq \zeta. \qquad (4)$$

The first two assumptions for our BGLM are also adopted in a previous work on GLM (Li, Lu, and Zhou 2017), which ensure that the change of $P(X = 1|\boldsymbol{pa}(X))$ is not abrupt. It is worth noting that Assumption 2 only needs the lower bound of the first derivative in the neighborhood of θ^*_X, which is weaker than Assumption 1 in (Filippi et al. 2010). Finally, Assumption 3 makes sure that each parent node of X still has a constant probability of taking either 0 or 1 even when all other parents of X already fix their values. This means that each parent has some independence and is not fully determined by other parents. In Appendix B we give some further justification of this assumption.

We first introduce some notations. Let n, m be the number of nodes and edges in G respectively. Let $D = \max_{X \in \boldsymbol{X} \cup \{Y\}} |\boldsymbol{Pa}(X)|$, $L^{(1)}_{\max} = \max_{X \in \boldsymbol{X} \cup \{Y\}} L^{(1)}_{f_X}$, and c be the constant in Lecué and Mendelson's inequality (Nie 2022) (see Lemma 10 in Appendix C.4). Let $(\boldsymbol{X}_t, Y_t)$ be the propagation result in the t^{th} round of intervention. It contains $\boldsymbol{V}_{t,X}$ and X^t for each node $X \in \boldsymbol{X} \cup \{Y\}$ where X^t is the propagating result of X and $\boldsymbol{V}_{t,X}$ is the propagating result of parents of X. Additionally, our estimation of the weight vectors is denoted by $\hat{\boldsymbol{\theta}}$.

Algorithm 1: BGLM-OFU for BGLM CCB Problem

```
1: Input: Graph $G = (\boldsymbol{X} \cup \{Y\}, E)$, intervention budget $K \in \mathbb{N}$, parameter $L^{(1)}_{f_X}, L^{(2)}_{f_X}, \kappa, \zeta$ in Assumption 1 , 2 and 3.
2: Initialize $M_{0,X} \leftarrow \mathbf{0} \in \mathbb{R}^{|\boldsymbol{Pa}(X)| \times |\boldsymbol{Pa}(X)|}$ for all $X \in \boldsymbol{X} \cup \{Y\}$, $\delta \leftarrow \frac{1}{3n\sqrt{T}}$, $R \leftarrow \lceil \frac{512D(L^{(2)}_{f_X})^2}{\kappa^4}(D^2 + \ln\frac{1}{\delta}) \rceil$, $T_0 \leftarrow \max\left\{\frac{c}{\zeta^2}\ln\frac{1}{\delta}, \frac{(8n^2-16n+2)R}{\zeta}\right\}$ and $\rho \leftarrow \frac{3}{\kappa}\sqrt{\log(1/\delta)}$.
3: /* Initialization Phase: */
4: Do no intervention on BGLM $G$ for $T_0$ rounds and observe feedback $(\boldsymbol{X}_t, Y_t), 1 \leq t \leq T_0$.
5: /* Iterative Phase: */
6: for $t = T_0 + 1, T_0 + 2, \cdots, T$ do
7:     $\{\hat{\boldsymbol{\theta}}_{t-1,X}, M_{t-1,X}\}_{X \in \boldsymbol{X} \cup \{Y\}} =$ BGLM-Estimate($(\boldsymbol{X}_1, Y_1), \cdots, (\boldsymbol{X}_{t-1}, Y_{t-1})$) (see Algorithm 2).
8:     Compute the confidence ellipsoid $\mathcal{C}_{t,X} = \{\boldsymbol{\theta}'_X \in [0,1]^{|\boldsymbol{Pa}(X)|} : \left\|\boldsymbol{\theta}'_X - \hat{\boldsymbol{\theta}}_{t-1,X}\right\|_{M_{t-1,X}} \leq \rho\}$ for any node $X \in \boldsymbol{X} \cup \{Y\}$.
9:     $(\boldsymbol{S}_t, \tilde{\boldsymbol{\theta}}_t) = \operatorname{argmax}_{\boldsymbol{S} \subseteq \boldsymbol{X}, |\boldsymbol{S}| \leq K, \boldsymbol{\theta}'_{t,X} \in \mathcal{C}_{t,X}} \mathbb{E}[Y|do(\boldsymbol{S})]$.
10:    Intervene all the nodes in $\boldsymbol{S}_t$ to 1 and observe the feedback $(\boldsymbol{X}_t, Y_t)$.
11: end for
```

We now propose the algorithm BGLM-OFU in Algorithm 1, where OFU stands for optimism in the face of uncertainty. The algorithm contains two phases. The first phase is the initialization phase with only pure observations without doing any intervention to ensure that our maximum likelihood estimation of $\boldsymbol{\theta}^*$ is accurate enough. Based on Lecué and Mendelson's inequality (Nie 2022), it is designed to ensure Eq. (5) in Lemma 1 holds. In practice, one alternative implementation of the initialization phase is doing no intervention until Eq. (5) holds for every $X \in \boldsymbol{X} \cup \{Y\}$. The required number of rounds is usually much less than T_0. Then in the second iterative phase, we use maximum likelihood estimator (MLE) method to estimate $\boldsymbol{\theta}^*$ and can therefore create a confidence region that contains the real parameters $\boldsymbol{\theta}^*$ with high probability around it to balance the exploration and exploitation. More specifically, in each iteration of intervention selections, we find an optimal intervention set together with a set of parameters $\tilde{\boldsymbol{\theta}}$ from the region around our unbiased estimation $\hat{\boldsymbol{\theta}}$ and take the found intervention set in this iteration. Intuitively, this method to select intervention sets follows the OFU spirit: the argmax operator in line 9 of Algorithm 1 selects the best (optimistic) solution in a confidence region (to address uncertainty). The empirical mean $\hat{\boldsymbol{\theta}}$ calculated corresponds to exploration while the confidence region surrounding it is for exploration.

The regret analysis of Algorithm 1 requires two technical components to support the main analysis. The first component indicates that when the observations are sufficient, we can get a good estimation $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}^*$, while the second component shows that a small change in the parameters should not lead to a big change in the reward $\mathbb{E}[Y|do(\boldsymbol{S})]$.

The first component is based on the result of maximum-likelihood estimation. In the studies of (Filippi et al. 2010) and (Li, Lu, and Zhou 2017), a standard log-likelihood function used in the updating process should be $L^{std}_{t,X}(\boldsymbol{\theta}_X) = \sum_{i=1}^t [X^i \ln f_X(\boldsymbol{V}^\intercal_{i,X}\boldsymbol{\theta}_X) + (1 - X^i)\ln(1 - f_X(\boldsymbol{V}^\intercal_{i,X}\boldsymbol{\theta}_X))]$. However, the analysis in their work needs the gradient of the log-likelihood function to have the form $\sum_{i=1}^t \left[X^i - f_X(\boldsymbol{V}^\intercal_{i,X}\boldsymbol{\theta}_X)\right]\boldsymbol{V}_{i,X}$, which is not true here. Therefore, using the same idea in (Zhang et al. 2022), we use

Algorithm 2: BGLM-Estimate

```
1: **Input:** All observations $((\boldsymbol{X}_1, Y_1), \cdots, (\boldsymbol{X}_t, Y_t))$ until round $t$.
2: **Output:** $\{\hat{\boldsymbol{\theta}}_{t,X}, M_{t,X}\}_{X \in \boldsymbol{X} \cup \{Y\}}$
3: For each $X \in \boldsymbol{X} \cup \{Y\}$, $i \in [t]$, construct data pair $(\boldsymbol{V}_{i,X}, X^i)$ with $\boldsymbol{V}_{i,X}$ the parent value vector of $X$ in round $i$, and $X^i$ the value of $X$ in round $i$ if $X \notin S_i$.
4: **for** $X \in \boldsymbol{X} \cup \{Y\}$ **do**
5:     Calculate the maximum-likelihood estimator $\hat{\boldsymbol{\theta}}_{t,X}$ by solving the equation $\sum_{i=1}^{t}(X^i - f_X(\boldsymbol{V}_{i,X}^{\intercal}\boldsymbol{\theta}_X))\boldsymbol{V}_{i,X} = 0$.
6:     $M_{t,X} = \sum_{i=1}^{t} \boldsymbol{V}_{i,X}\boldsymbol{V}_{i,X}^{\intercal}$.
7: **end for**
```

the pseudo log-likelihood function $L_{t,X}(\boldsymbol{\theta}_X)$ instead, which is constructed by integrating the gradient of it defined by $\triangledown_{\boldsymbol{\theta}_X} L_{t,X}(\boldsymbol{\theta}_X) = \sum_{i=1}^{t}\left[X^i - f_X(\boldsymbol{V}_{i,X}^{\intercal}\boldsymbol{\theta}_X)\right]\boldsymbol{V}_{i,X}$. Actually, this pseudo log-likelihood function is used in line 5 of Algorithm 2. The following lemma presents the result for the learning problem as the first technical component of the regret analysis. Let $M_{t,X}$ and $\hat{\boldsymbol{\theta}}_{t,X}$ be as defined in Algorithm 2, and also note that the definition of $\hat{\boldsymbol{\theta}}_{t,X}$ is equivalent to $\hat{\boldsymbol{\theta}}_{t,X} = \operatorname{argmax}_{\boldsymbol{\theta}_X} L_{t,X}(\boldsymbol{\theta}_X)$. Let $\lambda_{\min}(M)$ denote the minimum eigenvalue of matrix M.

Lemma 1 (Learning Problem for BGLM)**.** *Suppose that Assumptions 1 and 2 hold. Moreover, given $\delta \in (0,1)$, assume that*

$$\lambda_{\min}(M_{t,X}) \geq \frac{512|\boldsymbol{Pa}(X)|\left(L_{f_X}^{(2)}\right)^2}{\kappa^4}\left(|\boldsymbol{Pa}(X)|^2 + \ln\frac{1}{\delta}\right). \quad (5)$$

Then with probability at least $1 - 3\delta$, the maximum-likelihood estimator satisfies , for any $\boldsymbol{v} \in \mathbb{R}^{|\boldsymbol{Pa}(X)|}$,

$$\left|\boldsymbol{v}^{\intercal}(\hat{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}_X^*)\right| \leq \frac{3}{\kappa}\sqrt{\log(1/\delta)}\,\|\boldsymbol{v}\|_{M_{t,X}^{-1}},$$

where the probability is taken from the randomness of all data collected from round 1 to round t.

The proof of the above lemma is adapted from (Li, Lu, and Zhou 2017), and is included in Appendix C.2. Note that the initialization phase of the algorithm together with Assumption 3 and the Lecué and Mendelson's inequality would show the condition on $\lambda_{\min}(M_{t,X})$ in Eq.(5), and the design of the initialization phase, the summarization of Assumption 3 and the analysis to show Eq.(5) together form one of our key technical contributions in this section.

For the second component showing that a small change in parameters leads to a small change in the reward, we adapt the group observation modulated (GOM) bounded smoothness property for the LT model (Li et al. 2020) to show a GOM bounded smoothness property for BGLM. To do so, we define an equivalent form of BGLM as a threshold model as follows. For each node X, we randomly sample a threshold γ_X uniformly from $[0,1]$, i.e. $\gamma_X \sim \mathcal{U}[0,1]$, and if $f_X(\boldsymbol{Pa}(X) \cdot \boldsymbol{\theta}_X^*) + \varepsilon_X \geq \gamma_X$, X is activated (i.e. set to 1); if not, X is not activated (i.e. set to 0). Suppose $X_1, X_2, \cdots, X_{n-1}, Y$ is a topological order of nodes in $\boldsymbol{X} \cup \{Y\}$, then at time step 1, only X_1 is tried to be activated; at time step 2, only X_2 is tried to be activated by X_1; $\ldots$; at time step n, only Y is tried to be activated by activated nodes in $\boldsymbol{Pa}(Y)$. The above view of the propagating process is equivalent to BGLM, but it shows that BGLM is a general form of the LT model (Kempe, Kleinberg, and Tardos 2003) on DAG. Thus we can show a result below similar to Theorem 1 in (Li et al. 2020) for the LT model. For completeness, we include the proof in Appendix C.3. Henceforth, we use $\sigma(\boldsymbol{S}, \boldsymbol{\theta})$ to represent the reward function $\mathbb{E}[Y|do(\boldsymbol{S})]$ under parameter $\boldsymbol{\theta}$, to make the parameters of the reward explicit. We use $\boldsymbol{\gamma}$ to represent the vector $(\gamma_X)_{X \in \boldsymbol{X} \cup Y}$.

Lemma 2 (GOM Bounded Smoothness of BGLM)**.** *For any two weight vectors $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2 \in \Theta$ for a BGLM G, the difference of their expected reward for any intervened set $\boldsymbol{S}$ can be bounded as*

$$\left|\sigma(\boldsymbol{S}, \boldsymbol{\theta}^1) - \sigma(\boldsymbol{S}, \boldsymbol{\theta}^2)\right| \leq \mathbb{E}_{\boldsymbol{\varepsilon}, \boldsymbol{\gamma}}\left[\sum_{X \in \boldsymbol{X}_{\boldsymbol{S},Y}} \left|\boldsymbol{V}_X^{\intercal}(\boldsymbol{\theta}_X^1 - \boldsymbol{\theta}_X^2)\right| L_{f_X}^{(1)}\right], \quad (6)$$

where $\boldsymbol{X}_{\boldsymbol{S},Y}$ is the set of nodes in paths from $\boldsymbol{S}$ to Y excluding $\boldsymbol{S}$, and $\boldsymbol{V}_X$ is the propagation result of the parents of X under parameter $\boldsymbol{\theta}^2$. The expectation is taken over the randomness of the thresholds $\boldsymbol{\gamma}$ and the noises $\boldsymbol{\varepsilon}$.

We can now prove the regret bound of Algorithm 1. In the proof of Theorem 1, we use Lecué and Mendelson's inequality (Nie 2022) to prove that our initialization step has a very high probability to meet the needs of Lemma 1. Then we use Lemma 2 to transform the regret to the sum of $\|\boldsymbol{V}_{t,X}\|_{M_{t-1,X}^{-1}}$, which can be bounded using a similar lemma of Lemma 2 in (Li, Lu, and Zhou 2017). The details of the proof is put in Appendix C.4 for completeness.

Theorem 1 (Regret Bound of BGLM-OFU)**.** *Under Assumptions 1, 2 and 3, the regret of BGLM-OFU (Algorithms 1 and 2) is bounded as*

$$R(T) = O\left(\frac{1}{\kappa} n L_{\max}^{(1)} \sqrt{DT}\log T\right), \quad (7)$$

where the terms of $o(\sqrt{T})$ are omitted.

Remarks. The leading term of the regret in terms of T is in the order of $O(\sqrt{T}\log T)$, which is commonly seen in confidence ellipsoid based bandit or combinatorial bandit algorithms (e.g. (Abbasi-Yadkori, Pál, and Szepesvári 2011; Li et al. 2020; Zhang et al. 2022)). Also, it matches the regret of previous causal bandits algorithm, C-UCB in (Lu et al. 2020), which works on the atomic setting. The term $L_{\max}^{(1)}$ reflects the rate of changes in f_X's, and intuitively, the higher rate of changes in f_X's, the larger the regret since the online learning algorithm inevitably leads to some error in the parameter estimation, which will be amplified by the rate of changes in f_X's. Technically, $L_{\max}^{(1)}$ comes from the $L_{f_X}^{(1)}$ term in the GOM condition (Eq.(6)). Term n is to relax the sum over $\boldsymbol{X}_{\boldsymbol{S},Y}$ in Eq.(6), and it could be made tighter in causal graphs where $|\boldsymbol{X}_{\boldsymbol{S},Y}|$ is significantly smaller than n, and intuitively it means that all nodes on the path from $\boldsymbol{S}$ to Y would contribute to the regret. Term $\sqrt{D}$ implies that the regret depends on the number of parents of nodes, and technically it is because the scale of $\|\boldsymbol{V}_{t,X}\|_{M_{t-1,X}^{-1}}$ is approximately $\sqrt{|\boldsymbol{Pa}(X)|} \le \sqrt{D}$, and we bound the regret as the sum of $\|\boldsymbol{V}_{t,X}\|_{M_{t-1,X}^{-1}}$'s as we explained earlier. Term $\frac{1}{\kappa}$ comes from the learning problem (Lemma 1), which is also adopted in the regret bound of UCB-GLM of (Li, Lu, and Zhou 2017) using a similar learning problem. For the budget K, it does not appear in our regret because it is not directly related to the number of parameters we are estimating.

While our algorithm and analysis are based on several past studies (Li, Lu, and Zhou 2017; Zhang et al. 2022; Li et al. 2020), our innovation includes (a) the initialization phase and its corresponding Assumption 3 and its more involved analysis, because in our model we do not have direct observations of one-step immediate causal effect; and (b) the integration of the techniques from these separate studies, such as the maximum likelihood based analysis of (Li, Lu, and Zhou 2017), the pseudo log-likelihood function of (Zhang et al. 2022), and the GOM condition analysis of (Li et al. 2020), whereas each of these studies alone is not enough to achieve our result.

In line 9 of Algorithm 1, the argmax operator needs a simultaneous optimization over both the intervention set $\boldsymbol{S}$ and parameters $\boldsymbol{\theta}'$. This could be a computationally hard problem for large-scale problems, but since we focus on the online learning aspect of the problem, we leave the computationally-efficient solution for the full problem as future work, and such treatment is often seen in other combinatorial online learning problems (e.g. (Combes et al. 2015; Li et al. 2020)).

5 Algorithms for BLM with Hidden Variables

Previous sections consider only Markovian models without hidden variables. In many causal models, hidden variables exist to model the latent confounding factors. In this section, we present results on CCB under the linear model BLM with hidden variables.

5.1 Transforming the Model with Hidden Variables to the one without Hidden Variables

To address hidden variables, we first show how to reduce the BLM with hidden variables to a corresponding one without the hidden variables. Suppose the hidden variables in $G = (\boldsymbol{U} \cup \boldsymbol{X} \cup \{Y\}, E)$ are $\mathbf{U} = \{U_0, U_1, U_2, \cdots\}$, and we use X_i, X_j's to represent observed variables. Without loss of generality, we let U_0 always be 1 and it only has out-edges pointing to other observed or unobserved nodes, to model the self-activations of nodes. We allow various types of connections involving the hidden nodes, including edges from observed nodes to hidden nodes and edges among hidden nodes, but we disallow the situation where a hidden node U_s with $s > 0$ has two paths to X_i and X_i's descendant X_j and the paths contain only hidden nodes except the end points X_i and X_j. Figure 1 is an example of a causal graph allowed for this section.

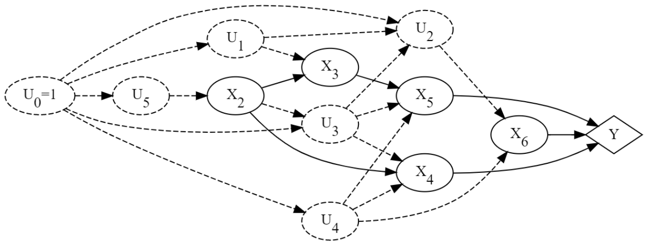

Figure 1: An Example of BLM with Hidden Variables

Our idea is to transform such a general causal model into an equivalent Markovian model $G' = (\{X_1\} \cup \mathbf{X} \cup \{Y\}, E')$. For convenience, we assume X_1 is not in the original observed variable set $\boldsymbol{X} \cup \{Y\}$. Henceforth, all notations with $'$, such as $\Pr', \mathbb{E}', \boldsymbol{\theta}^{*'}, \boldsymbol{Pa}'(X)$ correspond to the new Markovian model G', and the notations $\Pr, \mathbb{E}$ without $'$ refer to the original model G. For any two observed nodes X_i, X_j, a hidden path P from X_i to X_j is a directed path from X_i to X_j where all intermediate nodes are hidden or there are no intermediate nodes. We define θ_P^* to be the multiplication of weights on path P. Let $\mathcal{P}_{X_i,X_j}^u$ be the set of hidden paths from X_i to X_j. If $\mathcal{P}_{X_i,X_j}^u \neq \emptyset$, then we add edge (X_i, X_j) into E', and its weight $\theta_{X_i,X_j}^{*'} = \sum_{P \in \mathcal{P}_{X_i,X_j}^u} \theta_P^*$. As in the Markovian model, X_1 is always set to 1, and for each $X_i \in \boldsymbol{X} \cup \{Y\}$, we add edge (X_1, X_i) into E', with weight $\theta_{X_1,X_i}^{*'} = \Pr\{X_i = 1 | do(\mathbf{X} \cup \{Y\} \setminus \{X_i\} = \mathbf{0})\}$. The noise variables $\boldsymbol{\varepsilon}$ are carried over to G' without change.

The following lemma shows that the new model G' has the same parent-child conditional probability as the original model G. The proof of this lemma utilizes the do-calculus rules for causal inference (Pearl 2009).

Lemma 3. *For any $X \in \mathbf{X} \cup \{Y\}$, any $\boldsymbol{S} \subseteq \boldsymbol{X}$, any value $\boldsymbol{pa}'(X) \in \{0,1\}^{|\boldsymbol{Pa}'(X)|}$, any value $\boldsymbol{s} \in \{0,1\}^{|\boldsymbol{S}|}$ ($\boldsymbol{s}$ is consistent with $\boldsymbol{pa}'(X)$ on values in $\boldsymbol{S} \cap \boldsymbol{Pa}'(X)$), we have*

$$\Pr\left\{X = 1 | \boldsymbol{Pa}'(X) \setminus \{X_1\} = \boldsymbol{pa}'(X) \setminus \{x_1\}, do(\boldsymbol{S} = \boldsymbol{s})\right\}$$
$$= \Pr{}'\left\{X = 1 | \boldsymbol{Pa}'(X) = \boldsymbol{pa}'(X), do(\boldsymbol{S} = \boldsymbol{s})\right\}.$$

The next lemma shows that the objective function is also the same for G' and G.

Lemma 4. *For any $\boldsymbol{S} \subseteq \boldsymbol{X}$, any value $\boldsymbol{s} \in \{0,1\}^{|\boldsymbol{S}|}$, we have $\mathbb{E}[Y | do(\mathbf{S} = \mathbf{s})] = \mathbb{E}'[Y | do(\mathbf{S} = \mathbf{s})]$.*

The above two lemmas show that from G to G' the parent-child conditional probability and the reward function are all remain unchanged.

5.2 Applying BGLM-OFU on BLM

With the transformation described in Section 5.1 and its properties summarized in Lemmas 3 and 4, we can apply Algorithm 1 on G' to achieve the learning result for G. More precisely, we can use the observed data of X and $\boldsymbol{Pa}'(X)$ to estimate parameters $\boldsymbol{\theta}_X^{*'}$ and minimize the regret on the reward $\mathbb{E}'[Y|do(\mathbf{S})]$, which is the same as $\mathbb{E}[Y|do(\mathbf{S})]$. Furthermore, under the linear model BLM, Assumptions 1 and 2 hold with $L_{f_X}^{(1)} = \kappa = 1$ and $L_{f_X}^{(2)}$ could be any constant greater than 0. We still need Assumption 3, but we change the $\boldsymbol{Pa}(X)$'s in the assumption to $\boldsymbol{Pa}'(X)$'s. Then we can have the following regret bound.

Theorem 2 (Regret Bound of Algorithm 1 for BLM)**.** *Under Assumption 3, Algorithm 1 has the following regret bound when running on BLM with hidden variables:*

$$R(T) = O\left(n\sqrt{DT}\log T\right), \tag{8}$$

where n is the number of nodes in G', and $D = \max_{X \in \boldsymbol{X} \cup \{Y\}} |\boldsymbol{Pa}'(X)|$ is the maximum in-degree in G'.

Remarks. We compare our regret bound to the one in (Li et al. 2020) for the online influence maximization problem under the LT model, which is a special case of our BLM model with $\varepsilon_X = 0$ for all X's. Our regret bound is $O(n^{\frac{3}{2}}\sqrt{T}\ln T)$ while theirs is $O(n^{\frac{9}{2}}\sqrt{T}\ln T)$. The n^3 factor saving comes from three sources: (a) our reward is of scale $[0, 1]$ while theirs is $[0, n]$, and this saves one n factor; (b) our BLM is a DAG, and thus we can equivalently fix the activation steps as described before Lemma 2, causing our GOM condition (Lemma 2) to save another factor of n; and (c) we use MLE, which requires an initialization phase and Assumption 3 to give an accurate estimate of $\boldsymbol{\theta}^*$, while their algorithm uses linear regression without an initialization phase, which means we tradeoff a factor of n with an additional Assumption 3.

5.3 Algorithm for BLM based on Linear Regression

Now we have already introduced how to use Algorithm 1 and 2 to solve the online BLM CCB problem with hidden variables. However, in order to meet the needs of Lemma 1, we have to process an initialization phase. Assumption 3 needs to hold for the Markovian model G' after the transformation. We can remove the initialization phase and the dependency on Assumption 3, by noticing that our MLE based on pseudo log-likelihood maximization is equivalent to linear regression when adopted on BLMs. In particular, we use Lemma 1 in (Li et al. 2020) to replace Lemma 1 for MLE. We rewrite it as Lemma 11 in Appendix E.

Based on the above result, we introduce Algorithm 3 using the linear regression. Algorithm 3 is designed for Markovian BLMs. For a BLM G with more general hidden variables, we can apply the transformation described in Section 5.1 to transform the model into a Markovian model G' first. The following theorem shows the regret bound of Algorithm 3.

Theorem 3 (Regret Bound of Algorithm 3)**.** *The regret of BLM-LR (Algorithm 3) running on BLM with hidden variables is bounded as*

$$R(T) = O\left(n^2\sqrt{DT}\log T\right).$$

The proof of this theorem is adapted from the proof of Theorem 2 in (Li et al. 2020). For completeness, the proof is put in Appendix E. Comparing the algorithm and the regret bound of BLM-LR with those of BGLM-OFU, we can see a tradeoff between using the MLE method and the linear regression method: When we use the linear regression method with the BLM-LR algorithm, we do not need to have an initialization phase so Assumption 3 is not required anymore. However, the regret bound of BLM-LR (Theorem 3) has an extra factor of n in regret bound, comparing to the regret bound of BGLM-OFU on BLM (Theorem 2).

Algorithm 3: BLM-LR for BLM CCB Problem

```
Input: Graph $G = (\boldsymbol{X} \cup \{Y\}, E)$, intervention budget $K \in \mathbb{N}$.
Initialize $M_{0,X} \leftarrow \mathbf{I} \in \mathbb{R}^{|\boldsymbol{Pa}(X)| \times |\boldsymbol{Pa}(X)|}$, $\boldsymbol{b}_{0,X} \leftarrow \mathbf{0}^{|\boldsymbol{Pa}(X)|}$ for all $X \in \boldsymbol{X} \cup \{Y\}$, $\hat{\boldsymbol{\theta}}_{0,X} \leftarrow 0 \in \mathbb{R}^{|\boldsymbol{Pa}(X)|}$ for all $X \in \boldsymbol{X} \cup \{Y\}$, $\delta \leftarrow \frac{1}{n\sqrt{T}}$ and $\rho_t \leftarrow \sqrt{n\log(1+tn) + 2\log\frac{1}{\delta}} + \sqrt{n}$ for $t = 0, 1, 2, \cdots, T$.
for $t = 1, 2, \cdots, T$ do
    Compute the confidence ellipsoid $\mathcal{C}_{t,X} = \{\boldsymbol{\theta}'_X \in [0,1]^{|\boldsymbol{Pa}(X)|} : \left\|\boldsymbol{\theta}'_X - \hat{\boldsymbol{\theta}}_{t-1,X}\right\|_{M_{t-1,X}} \leq \rho_{t-1}\}$ for any node $X \in \boldsymbol{X} \cup \{Y\}$.
    $(\boldsymbol{S}_t, \tilde{\boldsymbol{\theta}}_t) = \operatorname{argmax}_{\boldsymbol{S} \subseteq \boldsymbol{X}, |\boldsymbol{S}| \leq K, \boldsymbol{\theta}'_{t,X} \in \mathcal{C}_{t,X}} \mathbb{E}[Y|do(\boldsymbol{S})]$.
    Intervene all the nodes in $\boldsymbol{S}_t$ to 1 and observe the feedback $(\boldsymbol{X}_t, Y_t)$.
    for $X \in \boldsymbol{X} \cup \{Y\}$ do
        Construct data pair $(\boldsymbol{V}_{t,X}, X^t)$ with $\boldsymbol{V}_{t,X}$ the parent value vector of $X$ in round $t$, and $X^t$ the value of $X$ in round $t$ if $X \notin S_t$.
        $M_{t,X} = M_{t-1,X} + \boldsymbol{V}_{t,X}\boldsymbol{V}_{t,X}^\intercal$, $\boldsymbol{b}_{t,X} = \boldsymbol{b}_{t-1,X} + X^t\boldsymbol{V}_{t,X}$, $\hat{\boldsymbol{\theta}}_{t,X} = M_{t,X}^{-1}\boldsymbol{b}_{t,X}$.
    end for
end for
```

6 Conclusion and Future Work

In this paper, we propose the combinatorial causal bandit (CCB) framework, and provide a solution for CCB under the BGLM. We further study a special model, the linear model. We show that our algorithm would work for models with many types of hidden variables. We further provide an algorithm for linear model not relying on Assumption 3 based on the linear regression.

There are many open problems and future directions to extend this work. For the BGLM, one could study how to remove some assumptions (e.g. Assumption 3), how to include hidden variables, or how to make the computation more efficient. For the linear model, one could consider how to remove the constraint on the hidden variable structure that does not allow a hidden variable to connect to an observed variable and its observed descendant via hidden variables. More generally, one could consider classes of causal models other than the BGLM.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and constructive feedback.

References

Abbasi-Yadkori, Y.; Pál, D.; and Szepesvári, C. 2011. Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 24.

Aitkin, M.; Francis, B.; Hinde, J.; and Darnell, R. 2009. *Statistical Modelling in R*. Oxford University Press Oxford.

Arnold, K. F.; Davies, V.; de Kamps, M.; Tennant, P. W.; Mbotwa, J.; and Gilthorpe, M. S. 2020. Reflection on modern methods: generalized linear models for prognosis and intervention—theory, practice and implications for machine learning. *International Journal of Epidemiology*, 49(6): 2074–2082.

Chen, W.; Wang, Y.; and Yuan, Y. 2013. Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning*, 151–159. PMLR.

Chen, W.; Wang, Y.; Yuan, Y.; and Wang, Q. 2016. Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. *The Journal of Machine Learning Research*, 17(1): 1746–1778.

Combes, R.; Shahi, M. S. T. M.; Proutiere, A.; et al. 2015. Combinatorial bandits revisited. In *Advances in Neural Information Processing Systems*, 2116–2124.

De la Fuente, A. 2000. *Mathematical Methods and Models for Economists*. Cambridge University Press.

Filippi, S.; Cappe, O.; Garivier, A.; and Szepesvári, C. 2010. Parametric bandits: The generalized linear case. *Advances in Neural Information Processing Systems*, 23.

Fisher, R. A. 1992. Statistical methods for research workers. In *Breakthroughs in Statistics*, 66–70. Springer.

Garcia-Huidobro, D.; and Michael Oakes, J. 2017. Squeezing observational data for better causal inference: Methods and examples for prevention research. *International Journal of Psychology*, 52(2): 96–105.

Han, B.; Yu, H.; and Friedberg, M. W. 2017. Evaluating the impact of parent-reported medical home status on children's health care utilization, expenditures, and quality: a difference-in-differences analysis with causal inference methods. *Health Services Research*, 52(2): 786–806.

Hastie, T. J.; and Pregibon, D. 2017. Generalized linear models. In *Statistical Models in S*, 195–247. Routledge.

Hernán, M. A.; and Robins, J. M. 2010. Causal inference.

Hilbe, J. M. 2011. Logistic regression. *International Encyclopedia of Statistical Science*, 1: 15–32.

Hoeffding, W. 1994. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, 409–426. Springer.

Karp, R. M. 1972. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, 85–103. Springer.

Kempe, D.; Kleinberg, J.; and Tardos, É. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137–146.

Lattimore, F.; Lattimore, T.; and Reid, M. D. 2016. Causal bandits: learning good interventions via causal inference. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 1189–1197.

Lattimore, T.; and Szepesvári, C. 2020. *Bandit Algorithms*. Cambridge University Press.

Lee, S.; and Bareinboim, E. 2018. Structural causal bandits: where to intervene? *Advances in Neural Information Processing Systems*, 31.

Lee, S.; and Bareinboim, E. 2019. Structural causal bandits with non-manipulable variables. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, 4146–4172.

Lee, S.; and Bareinboim, E. 2020. Characterizing optimal mixed policies: Where to intervene and what to observe. *Advances in Neural Information Processing Systems*, 33: 8565–8576.

Li, L.; Lu, Y.; and Zhou, D. 2017. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*, 2071–2080. PMLR.

Li, S.; Kong, F.; Tang, K.; Li, Q.; and Chen, W. 2020. Online influence maximization under linear threshold model. *Advances in Neural Information Processing Systems*, 33: 1192–1204.

Lu, Y.; Meisami, A.; and Tewari, A. 2021. Causal bandits with unknown graph structure. *Advances in Neural Information Processing Systems*, 34.

Lu, Y.; Meisami, A.; Tewari, A.; and Yan, W. 2020. Regret analysis of bandit problems with causal background knowledge. In *Conference on Uncertainty in Artificial Intelligence*, 141–150. PMLR.

Maiti, A.; Nair, V.; and Sinha, G. 2021. Causal Bandits on General Graphs. *arXiv preprint arXiv:2107.02772*.

Nair, V.; Patil, V.; and Sinha, G. 2021. Budgeted and non-budgeted causal bandits. In *International Conference on Artificial Intelligence and Statistics*, 2017–2025. PMLR.

Nie, Z. 2022. Matrix anti-concentration inequalities with applications. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, 568–581.

Pearl, J. 2009. *Causality*. Cambridge University Press. 2nd Edition.

Pearl, J. 2012. The do-calculus revisited. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, 3–11.

Russell, S.; and Norvig, P. 2021. Artificial Intelligence: A Modern Approach, Global Edition 4th. *Foundations*, 19: 23.

Sakate, D.; and Kashid, D. 2014. Comparison of Estimators in GLM with Binary Data. *Journal of Modern Applied Statistical Methods*, 13(2): 10.

Sen, R.; Shanmugam, K.; Dimakis, A. G.; and Shakkottai, S. 2017. Identifying best interventions through online importance sampling. In *International Conference on Machine Learning*, 3057–3066. PMLR.

Vansteelandt, S.; and Dukes, O. 2020. Assumption-lean inference for generalised linear model parameters. *arXiv preprint arXiv:2006.08402*.

Wu, W.; Du, H.; Wang, H.; Wu, L.; Duan, Z.; and Tian, C. 2018. On general threshold and general cascade models of social influence. *Journal of Combinatorial Optimization*, 35(1): 209–215.

Yabe, A.; Hatano, D.; Sumita, H.; Ito, S.; Kakimura, N.; Fukunaga, T.; and Kawarabayashi, K.-i. 2018. Causal bandits with propagating inference. In *International Conference on Machine Learning*, 5512–5520. PMLR.

Zhang, Z.; Chen, W.; Sun, X.; and Zhang, J. 2022. Online influence maximization with node-level feedback using standard offline oracles. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 9153–9161.

Appendix

A Notations

The main notations used in this paper are listed below.

Notations	Meanings
c	the constant in Lecué and Mendelson's inequality
$\boldsymbol{Ch}(X)$	set of children of X
D	the maximum in-degree of nodes in G
D_{out}	the maximum out-degree of nodes in Markovian BGLM G except X_1
E	the set of edges in G
f_X	the function that defines the activating probability of X in BGLM
G	a causal model $G = \{\mathbf{X} \cup \{Y\}, E\}$ with other hidden nodes
G'	the transformed Markovian model from G in Section 5
$J_{i,X}$	the number of created data pairs in the i^{th} round for node X
K	the maximal number of nodes that can be chosen to intervene in a round
$L^{(1)}_{f_X}$	the upper bound of $\dot{f}_X$
$L^{(1)}_{\max}$	$\max_{X \in \mathbf{X} \cup \{Y\}} L^{(1)}_{f_X}$
$L^{(2)}_{f_X}$	the upper bound of $\ddot{f}_X$
$M_{t,X}$	the observation matrix ($M_{t,X} = \sum_{i=1}^{t} \sum_{j=1}^{J_{i,X}} \boldsymbol{V}_{i,j,X} \boldsymbol{V}_{i,j,X}^{\intercal}$)
m	number of edges in G
n	number of observed nodes in G
pdf	probability density function
P	used to define the propagating rule of BGLMs by $P(X \mid \boldsymbol{Pa}(X))$
$\boldsymbol{Pa}(X)$	set of parents of X
$\mathcal{P}_{X_i,X_j}$	set of paths from X_i to X_j
$\mathcal{P}^{u}_{X_i,X_j}$	set of hidden paths from X_i to X_j
R	a constant in online algorithms
R_t	the regret in the t^{th} round
$\mathbf{S}_t$	the set of nodes we perform $do(\mathbf{S}_t = \mathbf{s}_t)$ in round t
T	total number of rounds in the online CCB problem
$\mathbf{U}$	the set of hidden variables in G
$\boldsymbol{V}_X$	the propagating result of $\boldsymbol{Pa}(X)$
$\mathbf{X}$	node set $\{X_1, X_2, \cdots, X_{n-1}\}$ such that $X_1 = 1$ for Markovian models
$(\mathbf{X}_t, Y_t)$	the observed propagating result of nodes in round t
X^i	propagating result of X at the i^{th} round
Y	the reward node in G
$\boldsymbol{\gamma}$	random thresholds in BGLMs
δ	a constant in the online algorithms
ε_X	the noise for activating X
ε'_X	a variable used in the proof of Lemma 1
ζ	a constant in Assumption 3 for BGLMs
$\boldsymbol{\theta}^*$	the real parameters
$\hat{\boldsymbol{\theta}}$	estimated parameters
$\tilde{\boldsymbol{\theta}}$	parameter with the upper confidence bound or from the OFU ellipsoid
$\lambda_{\min}(M)$	minimum eigenvalue of matrix M
κ	a constant in Assumption 2
ρ	a constant in the input of our online algorithms
υ	a constant in the justification of Assumption 3 (Appendix B)
$\sigma(\mathbf{S}, \boldsymbol{\theta})$	the expected reward under parameter $\boldsymbol{\theta}$ and intervention $do(\mathbf{S})$

B A Justification of Assumption 3

In order to show that our assumption is reasonable, we give a possible valuation of ζ here for general Markovian BGLMs under certain conditions. First, we use the following alternative assumption:

Assumption 4. *There exists a constant $\upsilon > 0$ such that for every $X \in \boldsymbol{X} \cup \{Y\} \setminus \{X_1\}$ and $\boldsymbol{pa}(X) \in \{0,1\}^{|\boldsymbol{Pa}(X)|}$,*

$$\mathbb{E}_{\boldsymbol{\varepsilon}}[P(X = 1|\boldsymbol{Pa}(X) = \boldsymbol{pa}(X))|\boldsymbol{\varepsilon}] \geq \upsilon, \tag{9}$$

$$\mathbb{E}_{\boldsymbol{\varepsilon}}[P(X = 0|\boldsymbol{Pa}(X) = \boldsymbol{pa}(X))|\boldsymbol{\varepsilon}] \geq \upsilon. \tag{10}$$

This assumption means that even if all the parents of X have fixed their values, X still has a constant probability of being either 0 or 1. The inequality for $X = 0$ is already adopted as Assumption 1 in (Zhang et al. 2022). In Markovian BGLMs, since $\Pr_{\boldsymbol{\varepsilon},X}\{X = 1|\boldsymbol{Pa}(X) = \boldsymbol{pa}(X)\} = \mathbb{E}_{\boldsymbol{\varepsilon},X}[X|\boldsymbol{Pa}(X) = \boldsymbol{pa}(X)] = \mathbb{E}_{\boldsymbol{\varepsilon}}[\mathbb{E}_X[f(\boldsymbol{\theta}^*_X \cdot \boldsymbol{pa}(X)) + \varepsilon_X|\boldsymbol{\varepsilon}]] = f(\boldsymbol{\theta}^*_X \cdot \boldsymbol{pa}(X)) \geq f(\theta^*_{X_1,X})$, condition $f(\theta^*_{X_1,X}) \geq \upsilon$ is suffice to satisfy Inequality (9). Similarly, $\Pr_{\boldsymbol{\varepsilon},X}\{X = 0|\boldsymbol{Pa}(X) = \boldsymbol{pa}(X)\} = 1 - \mathbb{E}_{\boldsymbol{\varepsilon},X}[X|\boldsymbol{Pa}(X) = \boldsymbol{pa}(X)] = 1 - f(\boldsymbol{\theta}^*_X \cdot \boldsymbol{pa}(X)) \geq 1 - f(||\boldsymbol{\theta}^*_X||_1)$. Thus, condition $f(||\boldsymbol{\theta}^*_X||_1) \leq 1 - \upsilon$ is suffice to satisfy Inequality (10). This means that the existence of υ can be reasonably achieved.

Denote the children of X by $\boldsymbol{Ch}(X)$. The following lemma shows that Assumption 4 leads to the following general result:

Lemma 5. *For every node $X \in \boldsymbol{X} \cup \{Y\} \setminus \{X_1\}$ and every value $\boldsymbol{v} \in \{0,1\}^{|\boldsymbol{X}|}$, we have*

$$\left(\frac{\upsilon}{1-\upsilon}\right)^{|\boldsymbol{Ch}(X)|+1} \leq \frac{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}(X = 1|\boldsymbol{X} \cup \{Y\} \setminus \{X\} = \boldsymbol{v})}{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}(X = 0|\boldsymbol{X} \cup \{Y\} \setminus \{X\} = \boldsymbol{v})} \leq \left(\frac{1-\upsilon}{\upsilon}\right)^{|\boldsymbol{Ch}(X)|+1}. \tag{11}$$

Proof. Without loss of generality, we only prove the left hand side inequality then the other side can be proved symmetrically. Let v_Z denote the entry in $\boldsymbol{v}$ corresponding to Z, and $\boldsymbol{v}_{\boldsymbol{Pa}(Z)}$ denote the sub-vector of $\boldsymbol{v}$ corresponding to $\boldsymbol{Pa}(Z)$. Concretely, we have

$$\begin{aligned}
&\frac{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\{X = 1|\boldsymbol{X} \cup \{Y\} \setminus \{X\} = \boldsymbol{v}\}}{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\{X = 0|\boldsymbol{X} \cup \{Y\} \setminus \{X\} = \boldsymbol{v}\}} \\
&= \frac{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\{X = 1, \boldsymbol{X} \cup \{Y\} \setminus \{X\} = \boldsymbol{v}\}}{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\{X = 0, \boldsymbol{X} \cup \{Y\} \setminus \{X\} = \boldsymbol{v}\}} \\
&= \frac{\mathbb{E}_{\boldsymbol{\varepsilon}}[\Pr_{\boldsymbol{X},Y}\{X = 1, \boldsymbol{X} \cup \{Y\} \setminus \{X\} = \boldsymbol{v}\}|\boldsymbol{\varepsilon}]}{\mathbb{E}_{\boldsymbol{\varepsilon}}[\Pr_{\boldsymbol{X},Y}\{X = 0, \boldsymbol{X} \cup \{Y\} \setminus \{X\} = \boldsymbol{v}\}|\boldsymbol{\varepsilon}]} \\
&= \frac{\prod_{Z \in \boldsymbol{X} \cup \{Y\} \setminus (\{X\} \cup \boldsymbol{Ch}(X))} \mathbb{E}_{\varepsilon_Z}[P(Z = v_Z|\boldsymbol{Pa}(Z) = \boldsymbol{v}_{\boldsymbol{Pa}(Z)})|\varepsilon_Z]}{\prod_{Z \in \boldsymbol{X} \cup \{Y\} \setminus (\{X\} \cup \boldsymbol{Ch}(X))} \mathbb{E}_{\varepsilon_Z}[P(Z = v_Z|\boldsymbol{Pa}(Z) = \boldsymbol{v}_{\boldsymbol{Pa}(Z)})|\varepsilon_Z]} \times \frac{\mathbb{E}_{\varepsilon_X}[P(X = 1|\boldsymbol{Pa}(X) = \boldsymbol{v}_{\boldsymbol{Pa}(X)})|\varepsilon_X]}{\mathbb{E}_{\varepsilon_X}[P(X = 0|\boldsymbol{Pa}(X) = \boldsymbol{v}_{\boldsymbol{Pa}(X)})|\varepsilon_X]} \\
&\quad \times \frac{\prod_{Z \in \boldsymbol{Ch}(X)} \mathbb{E}_{\varepsilon_Z}[P(Z = v_Z|\boldsymbol{Pa}(Z) \setminus \{X\} = \boldsymbol{v}_{\boldsymbol{Pa}(Z)\setminus X}, X = 1)|\varepsilon_Z]}{\prod_{Z \in \boldsymbol{Ch}(X)} \mathbb{E}_{\varepsilon_Z}[P(Z = v_Z|\boldsymbol{Pa}(Z) \setminus \{X\} = \boldsymbol{v}_{\boldsymbol{Pa}(Z)\setminus X}, X = 0)|\varepsilon_Z]} \qquad (12) \\
&\geq \frac{\upsilon}{1-\upsilon} \times \left(\frac{\upsilon}{1-\upsilon}\right)^{|\boldsymbol{Ch}(X)|} \\
&= \left(\frac{\upsilon}{1-\upsilon}\right)^{|\boldsymbol{Ch}(X)|+1},
\end{aligned}$$

where Eq. (12) applies the decomposition rule of the Bayesian causal model (Russell and Norvig 2021), and then the mutual independence of ε_X's for all $X \in \boldsymbol{X} \cup \{Y\}$. □

Now we can prove the following lemma which shows that the ζ constant in Assumption 3 exists.

Lemma 6. *With Assumption 4 in Markovian BGLM $G = (\boldsymbol{X} \cup \{Y\}, E)$, Assumption 3 holds for*

$$\zeta = \frac{\upsilon^{D_{out}+1}}{\upsilon^{D_{out}+1} + (1-\upsilon)^{D_{out}+1}},$$

where $D_{out} = \max_{X \in \boldsymbol{X} \cup \{Y\} \setminus \{X_1\}} |\boldsymbol{Ch}(X)|$.

Proof. With Lemma 5, we can deduce that for any $X \in \boldsymbol{X} \cup \{Y\} \setminus \{X_1\}$ and $X' \in \boldsymbol{Pa}(X)$, we have

$$\begin{aligned}
&\frac{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\{X' = 1|\boldsymbol{Pa}(X) \setminus \{X'\} = \boldsymbol{pa}(X) \setminus \{x'\}\}}{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\{X' = 0|\boldsymbol{Pa}(X) \setminus \{X'\} = \boldsymbol{pa}(X) \setminus \{x'\}\}} \\
&= \frac{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\{X' = 1, \boldsymbol{Pa}(X) \setminus \{X'\} = \boldsymbol{pa}(X) \setminus \{x'\}\}}{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\{X' = 0, \boldsymbol{Pa}(X) \setminus \{X'\} = \boldsymbol{pa}(X) \setminus \{x'\}\}} \\
&= \frac{\sum_{\boldsymbol{x} \cup \{y\} \setminus \boldsymbol{pa}(X) \in \{0,1\}^{n-|\boldsymbol{Pa}(X)|}} \Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\{X' = 1, \boldsymbol{X} \cup \{Y\} \setminus \{X'\} = \boldsymbol{x} \cup \{y\} \setminus \{x'\}\}}{\sum_{\boldsymbol{x} \cup \{y\} \setminus \boldsymbol{pa}(X) \in \{0,1\}^{n-|\boldsymbol{Pa}(X)|}} \Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\{X' = 0, \boldsymbol{X} \cup \{Y\} \setminus \{X'\} = \boldsymbol{x} \cup \{y\} \setminus \{x'\}\}}
\end{aligned}$$

$$
\begin{aligned}
&\geq \min_{\boldsymbol{x}\cup\{y\}\setminus \boldsymbol{pa}(X)\in\{0,1\}^{n-|\boldsymbol{Pa}(X)|}} \frac{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{X'=1, \boldsymbol{X}\cup\{Y\}\setminus\{X'\}=\boldsymbol{x}\cup\{y\}\setminus\{x'\}\right\}}{\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{X'=0, \boldsymbol{X}\cup\{Y\}\setminus\{X'\}=\boldsymbol{x}\cup\{y\}\setminus\{x'\}\right\}} \\
&\geq \left(\frac{\upsilon}{1-\upsilon}\right)^{\left|\boldsymbol{Ch}(X')\right|+1}.
\end{aligned}
$$

Similarly, we can prove the upper bound $\left(\frac{1-\upsilon}{\upsilon}\right)^{\left|\boldsymbol{Ch}(X')\right|+1}$ of it. Now we can deduce that $\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{X'=1|\boldsymbol{Pa}(X)\setminus\{X'\}=\boldsymbol{pa}(X)\setminus\{x'\}\right\}$ and $\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{X'=0|\boldsymbol{Pa}(X)\setminus\{X'\}=\boldsymbol{pa}(X)\setminus\{x'\}\right\}$ are both no less than $\frac{\upsilon^{|\boldsymbol{Ch}(X')|+1}}{\upsilon^{|\boldsymbol{Ch}(X')|+1}+(1-\upsilon)^{|\boldsymbol{Ch}(X')|+1}}$. This means that we can take $\zeta=\frac{\upsilon^{D_{\text{out}}+1}}{\upsilon^{D_{\text{out}}+1}+(1-\upsilon)^{D_{\text{out}}+1}}$ in Assumption 3. □

D_{out} is the maximum out-degree in the causal graph G. Lemma 6 shows that when D_{out} is small, we can have a reasonable ζ value for Assumption 3. Note that small D_{out} is a sufficient but not necessary condition for a reasonable ζ value in Assumption 3. Intuitively, as long as no node dominates or is dominated by other nodes, Assumption 3 is likely to hold.

C Proofs for the Online Algorithm of BGLM CCB Problem (Section 4)

C.1 A Technical Lemma

We first proof the following technical lemma, as a consequence Assumption 3. The lemma enables the use of Lecué and Mendelson's inequality in our later theoretical analysis.

Let $\mathit{Sphere}(d)$ denote the sphere of the d-dimensional unit ball.

Lemma 7. *For any $\boldsymbol{v}=(v_1,v_2,\cdots,v_{|\boldsymbol{Pa}(X)|})\in \mathit{Sphere}(|\boldsymbol{Pa}(X)|)$ and any $X\in\boldsymbol{X}\cup\{Y\}$ in a BGLM that satisfies Assumption 3, we have*

$$
\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{|\boldsymbol{Pa}(X)\cdot\boldsymbol{v}|\geq\frac{1}{\sqrt{4n^2-8n+1}}\right\}\geq\zeta,
$$

where $\boldsymbol{Pa}(X)$ is the random vector generated by the natural Bayesian propagation in BGLM G with no interventions (except for setting X_1 to 1).

Proof. The lemma is proved using the idea of Pigeonhole principle. Let $\boldsymbol{Pa}(X)=(X_{i_1}=X_1,X_{i_2},X_{i_3},\cdots,X_{i_{|\boldsymbol{Pa}(X)|}})$ as the random vector and $\boldsymbol{pa}(X)=(x_1=1,x_{i_1},x_{i_2},x_{i_3},\cdots,x_{i_{|\boldsymbol{Pa}(X)|}})$ as a possible valuation of $\boldsymbol{Pa}(X)$. Without loss of generality, we suppose that $|v_2|\geq|v_3|\geq\cdots\geq\left|v_{|\boldsymbol{Pa}(X)|}\right|$. For simplicity, we denote $n_0=\sqrt{n-2}+\frac{1}{2\sqrt{n-2}}$. If $|v_1|\geq\frac{n_0}{\sqrt{n_0^2+1}}$, we can deduce that

$$
\begin{aligned}
|\boldsymbol{pa}(X)\cdot\boldsymbol{v}| &\geq |v_1|-|v_2|-|v_3|-\cdots-\left|v_{|\boldsymbol{Pa}(X)|}\right| \\
&\geq \frac{n_0}{\sqrt{n_0^2+1}}-\sqrt{(n-2)\left(|v_2|^2+|v_3|^2+\cdots\left|v_{|\boldsymbol{Pa}(X)|}\right|^2\right)} \qquad (13)\\
&\geq \frac{n_0}{\sqrt{n_0^2+1}}-\sqrt{(n-2)\left(1-\frac{n_0^2}{n_0^2+1}\right)} \qquad (14)\\
&= \frac{1}{2\sqrt{(n_0^2+1)(n-2)}}=\frac{1}{\sqrt{4n^2-8n+1}},
\end{aligned}
$$

where Inequality (13) is by the Cauchy-Schwarz inequality and the fact that $|\boldsymbol{Pa}(X)|\leq n-1$, and Inequality (14) uses the fact that $\boldsymbol{v}\in \mathit{Sphere}(|\boldsymbol{Pa}(X)|)$. Thus, when $|v_1|\geq\frac{n_0}{\sqrt{n_0^2+1}}$, the event $|\boldsymbol{Pa}(X)\cdot\boldsymbol{v}|\geq\frac{1}{\sqrt{4n^2-8n+1}}$ holds deterministically. Otherwise, when $|v_1|<\frac{n_0}{\sqrt{n_0^2+1}}$, we use the fact that $|v_2|$ is the largest among $|v_2|,|v_3|,\ldots$ and deduce that

$$
|v_2|\geq\frac{1}{\sqrt{n-2}}\sqrt{|v_2|^2+|v_3|^2+\cdots}\geq\frac{\sqrt{1-\left(\frac{n_0}{\sqrt{n_0^2+1}}\right)^2}}{\sqrt{n-2}}=\frac{2}{\sqrt{4n^2-8n+1}}. \qquad (15)
$$

Therefore, using the fact that

$$
\begin{aligned}
&\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{X_{i_1}=1,X_{i_2}=x_{i_2},X_{i_3}=x_{i_3},\cdots\right\}= \\
&\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{X_{i_2}=x_{i_2}|X_{i_1}=1,X_{i_3}=x_{i_3},\cdots\right\}\cdot\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{(X_{i_1}=1,X_{i_3}=x_{i_3},\cdots\right\}\geq\zeta\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{X_{i_1}=1,X_{i_3}=x_{i_3},\cdots\right\}
\end{aligned}
$$

and $\sum_{x_{i_3},x_{i_4},\cdots}\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{X_{i_1}=1,X_{i_3}=x_{i_3},\cdots\right\}=1$, we have

$$
\begin{aligned}
&\Pr_{\boldsymbol{\varepsilon},\boldsymbol{X},Y}\left\{|\boldsymbol{Pa}(X)\cdot\boldsymbol{v}|\geq\frac{1}{\sqrt{4n^2-8n+1}}\right\}\\
&=\sum_{x_{i_3},x_{i_4},\cdots}\Pr\{X_{i_1}=1,X_{i_2}=1,X_{i_3}=x_{i_3},\cdots\}\cdot\mathbb{I}\left\{|(1,1,x_{i_3},x_{i_4},\cdots)\cdot(v_1,v_2,v_3,\cdots)|\geq\frac{1}{\sqrt{4n^2-8n+1}}\right\}\\
&\quad+\sum_{x_{i_3},x_{i_4},\cdots}\Pr\{X_{i_1}=1,X_{i_2}=0,X_{i_3}=x_{i_3},\cdots\}\cdot\mathbb{I}\left\{|(1,0,x_{i_3},x_{i_4},\cdots)\cdot(v_1,v_2,v_3,\cdots)|\geq\frac{1}{\sqrt{4n^2-8n+1}}\right\}\\
&\geq\sum_{x_{i_3},x_{i_4},\cdots}\zeta\Pr\{X_{i_1}=1,X_{i_3}=x_{i_3},X_{i_4}=x_{i_4}\cdots\}\cdot\mathbb{I}\left\{|(1,1,x_{i_3},x_{i_4},\cdots)\cdot(v_1,v_2,v_3,\cdots)|\geq\frac{1}{\sqrt{4n^2-8n+1}}\right\}\\
&\quad+\sum_{x_{i_3},x_{i_4},\cdots}\zeta\Pr\{X_{i_1}=1,X_{i_3}=x_{i_3},X_{i_4}=x_{i_4},\cdots\}\cdot\mathbb{I}\left\{|(1,0,x_{i_3},x_{i_4},\cdots)\cdot(v_1,v_2,v_3,\cdots)|\geq\frac{1}{\sqrt{4n^2-8n+1}}\right\}\\
&=\zeta\cdot\sum_{x_{i_3},x_{i_4},\cdots}\Pr\{X_{i_1}=1,X_{i_3}=x_{i_3},X_{i_4}=x_{i_4},\cdots\}\Bigg(\mathbb{I}\left\{|(1,1,x_{i_3},x_{i_4},\cdots)\cdot(v_1,v_2,v_3,\cdots)|\geq\frac{1}{\sqrt{4n^2-8n+1}}\right\}\\
&\quad+\mathbb{I}\left\{|(1,0,x_{i_3},x_{i_4},\cdots)\cdot(v_1,v_2,v_3,\cdots)|\geq\frac{1}{\sqrt{4n^2-8n+1}}\right\}\Bigg)\\
&\geq\zeta\sum_{x_{i_3},x_{i_4},\cdots}\Pr\{X_{i_1}=1,X_{i_3}=x_{i_3},X_{i_4}=x_{i_4},\cdots\}\qquad\qquad(16)\\
&=\zeta,
\end{aligned}
$$

which is exactly what we want to prove. Inequality (16) holds because otherwise, at least for some $x_{i_3},x_{i_4},\ldots$, both indicators on the left-hand side of the inequality have to be 0, which implies that

$$
|(1,1,x_{i_3},x_{i_4},\cdots)\cdot(v_1,v_2,v_3,\cdots)-(1,0,x_{i_3},x_{i_4},\cdots)\cdot(v_1,v_2,v_3,\cdots)|=|v_2|<\frac{2}{\sqrt{4n^2-8n+1}}, \tag{17}
$$

but this contradicts to Inequality (15). □

C.2 Proof of the Learning Problem for BGLM CCB Problem (Lemma 1)

Lemma 1 (Learning Problem for BGLM)**.** *Suppose that Assumptions 1 and 2 hold. Moreover, given $\delta\in(0,1)$, assume that*

$$
\lambda_{\min}(M_{t,X})\geq\frac{512|\boldsymbol{Pa}(X)|\left(L_{f_X}^{(2)}\right)^2}{\kappa^4}\left(|\boldsymbol{Pa}(X)|^2+\ln\frac{1}{\delta}\right). \tag{5}
$$

Then with probability at least $1-3\delta$, the maximum-likelihood estimator satisfies , for any $\boldsymbol{v}\in\mathbb{R}^{|\boldsymbol{Pa}(X)|}$,

$$
\left|\boldsymbol{v}^\intercal(\hat{\boldsymbol{\theta}}_{t,X}-\boldsymbol{\theta}_X^*)\right|\leq\frac{3}{\kappa}\sqrt{\log(1/\delta)}\,\|\boldsymbol{v}\|_{M_{t,X}^{-1}},
$$

where the probability is taken from the randomness of all data collected from round 1 to round t.

Proof. Note that $\hat{\boldsymbol{\theta}}_{t,X}$ satisfies that $\nabla L_{t,X}(\hat{\boldsymbol{\theta}}_X)=0$, where the gradient is

$$
\nabla L_{t,X}(\boldsymbol{\theta}_X)=\sum_{i=1}^t[X^t-f_X(\boldsymbol{V}_{i,X}^\intercal\boldsymbol{\theta}_X)]\boldsymbol{V}_{i,X}.
$$

Define $G(\boldsymbol{\theta}_X)=\sum_{i=1}^t(f_X(\boldsymbol{V}_{i,X}^\intercal\boldsymbol{\theta}_X)-f_X(\boldsymbol{V}_{i,X}^\intercal\boldsymbol{\theta}_X^*))\boldsymbol{V}_{i,X}$. Then we have $G(\boldsymbol{\theta}_X^*)=0$ and $G(\hat{\boldsymbol{\theta}}_{t,X})=\sum_{i=1}^t\varepsilon'_{i,X}\boldsymbol{V}_{i,X}$, where $\varepsilon'_{i,X}=X^i-f_X(\boldsymbol{V}_{i,X}^\intercal\boldsymbol{\theta}_X^*)$. Note that $\mathbb{E}[\varepsilon'_{i,X}|\boldsymbol{V}_{i,X}]=0$ and $\varepsilon'_{i,X}=X^i-f_X(\boldsymbol{V}_{i,X}^\intercal\boldsymbol{\theta}_X^*)\in[-1,1]$ since $X^i\in\{0,1\}$ and $f_X(\boldsymbol{V}_{i,X}^\intercal\boldsymbol{\theta}_X^*)\in[0,1]$. Therefore, $\varepsilon'_{i,X}$ is 1-sub-Gaussian. Furthermore, define $Z=G(\hat{\boldsymbol{\theta}}_{t,X})=\sum_{i=1}^t\varepsilon'_{i,X}\boldsymbol{V}_{i,X}$ for convenience. All the remaining notations in this proof are corresponding to some $X\in\mathbf{X}\cup\{Y\}$.

Step 1: Consistency of $\hat{\boldsymbol{\theta}}_{t,X}$. We first prove the consistency of $\hat{\boldsymbol{\theta}}_{t,X}$. For any $\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\mathbb{R}^{|\boldsymbol{Pa}(X)|}$, by the mean value theorem, $\exists\overline{\boldsymbol{\theta}}=s\boldsymbol{\theta}_1+(1-s)\boldsymbol{\theta}_2, 0<s<1$ such that

$$
G(\boldsymbol{\theta}_1)-G(\boldsymbol{\theta}_2)=\left[\sum_{i=1}^t\dot{f}_X(\boldsymbol{V}_{i,X}^\intercal\overline{\boldsymbol{\theta}})\boldsymbol{V}_{i,X}\boldsymbol{V}_{i,X}^\intercal\right](\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)
$$

$$\triangleq F(\overline{\boldsymbol{\theta}})(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2).$$

Since $\dot{f}_X \geq \kappa > 0$ as we state in Assumption 1 and $\lambda_{\min}(M_{t,X}) > 0$, we have $F(\overline{\boldsymbol{\theta}}) \succ \kappa M_{t,X}$ and for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, we have $(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^\intercal (G(\boldsymbol{\theta}_1) - G(\boldsymbol{\theta}_2)) \geq (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^\intercal (\kappa M_{t,X})(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) > 0$. Now we deduce that $G(\boldsymbol{\theta})$ is an injection from $\mathbb{R}^{|\boldsymbol{Pa}(X)|}$ to $\mathbb{R}^{|\boldsymbol{Pa}(X)|}$ and therefore G^{-1} is well defined. We have $\hat{\boldsymbol{\theta}}_{t,X} = G^{-1}(Z)$.

For any $\boldsymbol{\theta} \in \Theta$, now we have a $\overline{\boldsymbol{\theta}}_X$ such that

$$\begin{aligned}
\|Z\|^2_{M^{-1}_{t,X}} &= \left\| G(\hat{\boldsymbol{\theta}}_{t,X}) \right\|^2_{M^{-1}_{t,X}} \\
&= \left\| G(\hat{\boldsymbol{\theta}}_{t,X}) - G(\boldsymbol{\theta}^*_X) \right\|^2_{M^{-1}_{t,X}} \\
&= (\hat{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}^*)^\intercal F(\overline{\boldsymbol{\theta}}) M^{-1}_{t,X} F(\overline{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}^*_X) \\
&\geq \kappa^2 \lambda_{\min}(M_{t,X}) \left\| \hat{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}^*_X \right\|^2 .
\end{aligned}$$

Then we need a bound for $\|Z\|^2_{M^{-1}_{t,X}}$ so that we can get a bound for $\left\| \hat{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}^*_X \right\|$. The next lemma solves this.

Lemma 8 ((Zhang et al. 2022))**.** *For any $\delta > 0$, define the following event $\mathcal{E}_G = \{\|Z\|_{M^{-1}_{t,X}} \leq 4\sqrt{|\boldsymbol{Pa}(X)| + \log\frac{1}{\delta}}\}$. Then $\mathcal{E}_G$ holds with probability at least $1 - \delta$.*

Proof. The proof of this lemma is a simple rewriting of the proof of Lemma 10 in (Zhang et al. 2022). □

Combining this lemma and the deduction we made just now, we can get

$$\left\| \hat{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}^*_X \right\| \leq \frac{\|Z\|_{M^{-1}_{t,X}}}{\kappa\sqrt{\lambda_{\min}(M_{t,X})}} \leq \frac{4}{\kappa}\sqrt{\frac{|\boldsymbol{Pa}(X)| + \log\frac{1}{\delta}}{\lambda_{\min}(M_{t,X})}} \leq 1.$$

Step 2: Normality of $\hat{\boldsymbol{\theta}}_X$. Now we assume that $\mathcal{E}$ holds in the following proof. Define $\Delta = \hat{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}^*_X$, then $\exists s \in [0,1]$ such that $Z = G(\hat{\boldsymbol{\theta}}_{t,X}) - G(\boldsymbol{\theta}^*_X) = (H + E)\Delta$, where $\overline{\boldsymbol{\theta}} = s\boldsymbol{\theta}^*_X + (1-s)\hat{\boldsymbol{\theta}}_{t,X}$, $H = F(\boldsymbol{\theta}^*_X) = \sum_{i=1}^t \dot{f}_X(\boldsymbol{V}^\intercal_{i,X}\boldsymbol{\theta}^*_X)\boldsymbol{V}_{i,X}\boldsymbol{V}^\intercal_{i,X}$ and $E = F(\overline{\boldsymbol{\theta}}) - F(\boldsymbol{\theta}^*_X)$. According to the mean value theorem, we have

$$\begin{aligned}
E &= \sum_{i=1}^t (\dot{f}_X(\boldsymbol{V}_{i,X} \cdot \overline{\boldsymbol{\theta}}) - \dot{f}_X(\boldsymbol{V}_{i,X} \cdot \boldsymbol{\theta}^*_X))\boldsymbol{V}_{i,X}\boldsymbol{V}^\intercal_{i,X} \\
&= \sum_{i=1}^t \ddot{f}_X(r_i)\boldsymbol{V}^\intercal_{i,X}\Delta\boldsymbol{V}_{i,X}\boldsymbol{V}^\intercal_{i,X}
\end{aligned}$$

for some $r_i \in \mathbb{R}$. Because $\ddot{f}_X \leq L^{(2)}_{f_X}$ and $s \in [0,1]$, for any $\boldsymbol{v} \in \mathbb{R}^{|\boldsymbol{Pa}(X)|} \setminus \{\boldsymbol{0}\}$, we have

$$\begin{aligned}
\boldsymbol{v}^T H^{-1/2} E H^{-1/2} \boldsymbol{v} &= (1-s)\sum_{i=1}^t \ddot{f}_X(r_i)\boldsymbol{V}^\intercal_{i,X}\Delta \left\| \boldsymbol{v}^\intercal H^{-1/2}\boldsymbol{V}_{i,X} \right\| \\
&\leq \sum_{i=1}^t L^{(2)}_{f_X} \|\boldsymbol{V}_{i,X}\| \|\Delta\| \left\| \boldsymbol{v}^\intercal H^{-1/2}\boldsymbol{V}_{i,X} \right\|^2 \\
&\leq L^{(2)}_{f_X}\sqrt{|\boldsymbol{Pa}(X)|}\,\|\Delta\|\,(\boldsymbol{v}^\intercal H^{-1/2}(\sum_{i=1}^t \boldsymbol{V}_{i,X}\boldsymbol{V}^\intercal_{i,X})H^{-1/2}\boldsymbol{v}) \\
&\leq \frac{L^{(2)}_{f_X}\sqrt{|\boldsymbol{Pa}(X)|}}{\kappa}\|\Delta\|\,\|\boldsymbol{v}\|^2 ,
\end{aligned}$$

where we have use the fact that $\|\boldsymbol{V}_{i,X}\| \leq \sqrt{|\boldsymbol{Pa}(X)|}$ for the second inequality. Therefore, we have

$$\left\| H^{-1/2} E H^{-1/2} \right\| \leq \frac{L^{(2)}_{f_X}\sqrt{|\boldsymbol{Pa}(X)|}}{\kappa}\|\Delta\|$$

$$\leq \frac{4L_{f_X}^{(2)}\sqrt{|\boldsymbol{Pa}(X)|}}{\kappa^2}\sqrt{\frac{|\boldsymbol{Pa}(X)|+\ln\frac{1}{\delta}}{\lambda_{\min}(M_{t,X})}}.$$

When $\lambda_{\min}(M_{t,X}) \geq 64\left(L_{f_X}^{(2)}\right)^2 |\boldsymbol{Pa}(X)|\left(|\boldsymbol{Pa}(X)|+\ln\frac{1}{\delta}\right)/\kappa^4$, we have $\left\|H^{-1/2}EH^{-1/2}\right\| \leq \frac{1}{2}$.

Now we can prove this theorem. For any $\boldsymbol{v} \in \mathbb{R}^{|\boldsymbol{Pa}(X)|}$, we have

$$\begin{aligned}\boldsymbol{v}^\intercal(\hat{\boldsymbol{\theta}}_{t,X}-\boldsymbol{\theta}_X^*) &= \boldsymbol{v}^\intercal(H+E)^{-1}Z \\ &= \boldsymbol{v}^\intercal H^{-1}Z - \boldsymbol{v}^\intercal H^{-1}E(H+E)^{-1}Z.\end{aligned}$$

Note that $(H+E)^{-1}$ exists since $H+E = F(\overline{\boldsymbol{\theta}}) \succ \kappa M_{t,X} \succ 0$.

For the first term, define $D \triangleq (\boldsymbol{V}_{1,X}, \boldsymbol{V}_{2,X}, \cdots, \boldsymbol{V}_{t,X})^\intercal \in \mathbb{R}^{t\times|\boldsymbol{Pa}(X)|}$. Note that $D^\intercal D = \sum_{i=1}^t \boldsymbol{V}_{i,X}\boldsymbol{V}_{i,X}^\intercal = M_{t,X}$. By the Hoeffding's inequality (Hoeffding 1994),

$$\begin{aligned}\Pr(\left|\boldsymbol{v}^\intercal H^{-1}Z \geq a\right|) &\leq \exp\left(-\frac{a^2}{2\left\|\boldsymbol{v}^\intercal H^{-1}D^\intercal\right\|^2}\right) \\ &= \exp\left(-\frac{a^2}{2\boldsymbol{v}^\intercal H^{-1}D^\intercal DH^{-1}\boldsymbol{v}}\right) \\ &\leq \exp\left(-\frac{a^2\kappa^2}{2\left\|\boldsymbol{v}\right\|_{M_{t,X}^{-1}}^2}\right).\end{aligned}$$

The last inequality holds because $H \succeq \kappa M_{t,X} = \kappa D^\intercal D$. From this we deduce that with probability at least $1-2\delta$, $\left|\boldsymbol{v}^\intercal H^{-1}Z\right| \leq \frac{\sqrt{2\ln 1/\delta}}{\kappa}\left\|\boldsymbol{v}\right\|_{M_{t,X}^{-1}}$.

For the second term, we have

$$\begin{aligned}\left|\boldsymbol{v}^\intercal H^{-1}E(H+E)^{-1}Z\right| &\leq \left\|\boldsymbol{v}\right\|_{H^{-1}}\left\|H^{-\frac{1}{2}}E(H+E)^{-1}Z\right\| \\ &\leq \left\|\boldsymbol{v}\right\|_{H^{-1}}\left\|H^{-\frac{1}{2}}E(H+E)^{-1}H^{\frac{1}{2}}\right\|\left\|Z\right\|_{H^{-1}} \\ &\leq \frac{1}{\kappa}\left\|\boldsymbol{v}\right\|_{M_{t,X}^{-1}}\left\|H^{-\frac{1}{2}}E(H+E)^{-1}H^{\frac{1}{2}}\right\|\left\|Z\right\|_{M_{t,X}^{-1}}.\end{aligned}$$

where the last inequality is due to the fact that $H \succeq \kappa M_{t,X}$. Since $(H+E)^{-1} = H^{-1} - H^{-1}E(H+E)^{-1}$, we have

$$\begin{aligned}\left\|H^{-\frac{1}{2}}E(H+E)^{-1}H^{\frac{1}{2}}\right\| &= \left\|H^{-\frac{1}{2}}E(H^{-1}-H^{-1}E(H+E)^{-1})^{-1}H^{\frac{1}{2}}\right\| \\ &= \left\|H^{-\frac{1}{2}}EH^{\frac{1}{2}} + H^{-\frac{1}{2}}EH^{-1}E(H+E)^{-1}H^{\frac{1}{2}}\right\| \\ &\leq \left\|H^{-\frac{1}{2}}EH^{\frac{1}{2}}\right\| + \left\|H^{-\frac{1}{2}}EH^{-\frac{1}{2}}\right\|\left\|H^{-\frac{1}{2}}E(H+E)^{-1}H^{\frac{1}{2}}\right\|.\end{aligned}$$

By solving this inequality we get

$$\begin{aligned}\left\|H^{-\frac{1}{2}}E(H+E)^{-1}H^{\frac{1}{2}}\right\| &\leq \frac{\left\|H^{-\frac{1}{2}}EH^{-\frac{1}{2}}\right\|}{1-\left\|H^{-\frac{1}{2}}EH^{-\frac{1}{2}}\right\|} \\ &\leq 2\left\|H^{-\frac{1}{2}}EH^{-\frac{1}{2}}\right\| \\ &\leq \frac{8L_{f_X}^{(2)}}{\kappa^2}\sqrt{\frac{|\boldsymbol{Pa}(X)|(|\boldsymbol{Pa}(X)+\log\frac{1}{\delta}|)}{\lambda_{\min}(M_{t,X})}}.\end{aligned}$$

Therefore, we have

$$\left|v^\intercal H^{-1}E(H+E)^{-1}Z\right| \leq \frac{32L_{f_X}^{(2)}\sqrt{|\boldsymbol{Pa}(X)|}(|\boldsymbol{Pa}(X)|+\log\frac{1}{\delta})}{\kappa^3\sqrt{\lambda_{\min}(M_{t,X})}}\left\|\boldsymbol{v}\right\|_{M_{t,X}^{-1}}.$$

Thus, we have

$$\left|\boldsymbol{v}^\intercal(\hat{\boldsymbol{\theta}}_{t,X}-\boldsymbol{\theta}^*_X)\right| \le \left(\frac{32L^{(2)}_{f_X}\sqrt{|\boldsymbol{Pa}(X)|}(|\boldsymbol{Pa}(X)|+\log\frac{1}{\delta})}{\kappa^3\sqrt{\lambda_{\min}(M_{t,X})}}+\frac{\sqrt{2\ln 1/\delta}}{\kappa}\right)\|\boldsymbol{v}\|_{M^{-1}_{t,X}}$$
$$\le \frac{3}{\kappa}\sqrt{\log(1/\delta)}\,\|\boldsymbol{v}\|_{M^{-1}_{t,X}}\,,$$

when

$$\lambda_{\min}(M_{t,X}) \ge \frac{512|\boldsymbol{Pa}(X)|(L^{(2)}_{f_X})^2}{\kappa^4}\left(|\boldsymbol{Pa}(X)|^2+\ln\frac{1}{\delta}\right).$$

□

C.3 Proof of the GOM Bounded Smoothness Condition for BGLM (Lemma 2)

Lemma 2 (GOM Bounded Smoothness of BGLM)**.** *For any two weight vectors $\boldsymbol{\theta}^1,\boldsymbol{\theta}^2\in\Theta$ for a BGLM G, the difference of their expected reward for any intervened set $\boldsymbol{S}$ can be bounded as*

$$\left|\sigma(\boldsymbol{S},\boldsymbol{\theta}^1)-\sigma(\boldsymbol{S},\boldsymbol{\theta}^2)\right| \le \mathbb{E}_{\boldsymbol{\varepsilon},\boldsymbol{\gamma}}\left[\sum_{X\in\boldsymbol{X}_{\boldsymbol{S},Y}}\left|\boldsymbol{V}_X^\intercal(\boldsymbol{\theta}^1_X-\boldsymbol{\theta}^2_X)\right|L^{(1)}_{f_X}\right], \tag{6}$$

where $\boldsymbol{X}_{\boldsymbol{S},Y}$ is the set of nodes in paths from $\boldsymbol{S}$ to Y excluding $\boldsymbol{S}$, and $\boldsymbol{V}_X$ is the propagation result of the parents of X under parameter $\boldsymbol{\theta}^2$. The expectation is taken over the randomness of the thresholds $\boldsymbol{\gamma}$ and the noises $\boldsymbol{\varepsilon}$.

Proof. Firstly, we have

$$\left|\sigma(\mathbf{S},\boldsymbol{\theta}^1)-\sigma(\mathbf{S},\boldsymbol{\theta}^2)\right| = \mathbb{E}_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n,\boldsymbol{\varepsilon}}\left[\mathbb{I}\{Y\text{ is influenced under }\boldsymbol{\theta}^1,\boldsymbol{\gamma}\}\ne\mathbb{I}\{Y\text{ is influenced under }\boldsymbol{\theta}^2,\boldsymbol{\gamma}\}\right].$$

Then we define the following event $\mathcal{E}^{\boldsymbol{\varepsilon}}_0(X)$ as below:

$$\mathcal{E}^{\boldsymbol{\varepsilon}}_0(X)=\left\{\boldsymbol{\gamma}|\mathbb{I}\{X\text{ is activated under }\boldsymbol{\gamma},\boldsymbol{\theta}^1\}\ne\mathbb{I}\{X\text{ is activated under }\boldsymbol{\gamma},\boldsymbol{\theta}^2\}\right\}.$$

Thus we have

$$\left|\sigma(\mathbf{S},\boldsymbol{\theta}^1)-\sigma(\mathbf{S},\boldsymbol{\theta}^2)\right| \le \mathbb{E}_{\boldsymbol{\varepsilon}}\left[\Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_0(Y)\}\right].$$

Let $\phi^{\boldsymbol{\varepsilon}}(\boldsymbol{\theta},\boldsymbol{\gamma})=(\phi^{\boldsymbol{\varepsilon}}_0(\boldsymbol{\theta},\boldsymbol{\gamma})=S,\phi^{\boldsymbol{\varepsilon}}_1(\boldsymbol{\theta},\boldsymbol{\gamma}),\cdots,\phi^{\boldsymbol{\varepsilon}}_n(\boldsymbol{\theta},\boldsymbol{\gamma}))$ be the sequence of activation sets given weight vector $\boldsymbol{\theta}$, 0-mean noise $\boldsymbol{\varepsilon}$ and threshold factor $\boldsymbol{\gamma}$. More specifically, $\phi_i(\boldsymbol{\theta},\boldsymbol{\gamma})$ is the set of nodes activated by time step i, i.e. $\phi^{\boldsymbol{\varepsilon}}_i(\boldsymbol{\theta},\boldsymbol{\gamma})\backslash\phi^{\boldsymbol{\varepsilon}}_{i-1}(\boldsymbol{\theta},\boldsymbol{\gamma})\in\{\emptyset,\{X_i\}\}$. For every node $X\in\mathbf{X}_{\mathbf{S},Y}$, we define the event that X is the first node that has different activation under $\boldsymbol{\theta}^1$ and $\boldsymbol{\theta}^2$ as below:

$$\mathcal{E}^{\boldsymbol{\varepsilon}}_1(X)=\{\boldsymbol{\gamma}|\exists\tau\in[n],\forall\tau'<\tau,\phi^{\boldsymbol{\varepsilon}}_{\tau'}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})=\phi^{\boldsymbol{\varepsilon}}_{\tau'}(\boldsymbol{\theta}^2,\boldsymbol{\gamma}),X\in(\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})\setminus\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^2,\boldsymbol{\gamma})\cup(\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^2,\boldsymbol{\gamma})\setminus\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})))\}.$$

Then we have $\mathcal{E}^{\boldsymbol{\varepsilon}}_0(Y)\subseteq\cup_{X\in\mathbf{X}_{\mathbf{S},Y}}\mathcal{E}^{\boldsymbol{\varepsilon}}_1(X)$. Now we define some other events as below:

$$\begin{aligned}
&\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau)=\{\boldsymbol{\gamma}|\forall\tau'<\tau,\phi^{\boldsymbol{\varepsilon}}_{\tau'}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})=\phi^{\boldsymbol{\varepsilon}}_{\tau'}(\boldsymbol{\theta}^2,\boldsymbol{\gamma}),X\notin\phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})\},\\
&\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau)=\{\boldsymbol{\gamma}|\forall\tau'<\tau,\phi^{\boldsymbol{\varepsilon}}_{\tau'}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})=\phi^{\boldsymbol{\varepsilon}}_{\tau'}(\boldsymbol{\theta}^2,\boldsymbol{\gamma}),X\in\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})\setminus\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^2,\boldsymbol{\gamma})\},\\
&\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,2}(X,\tau)=\{\boldsymbol{\gamma}|\forall\tau'<\tau,\phi^{\boldsymbol{\varepsilon}}_{\tau'}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})=\phi^{\boldsymbol{\varepsilon}}_{\tau'}(\boldsymbol{\theta}^2,\boldsymbol{\gamma}),X\in\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^2,\boldsymbol{\gamma})\setminus\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})\},\\
&\mathcal{E}^{\boldsymbol{\varepsilon}}_{3,1}(X,\tau)=\{\boldsymbol{\gamma}|X\in\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})\setminus\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^2,\boldsymbol{\gamma})\},\mathcal{E}^{\boldsymbol{\varepsilon}}_{3,2}(X,\tau)=\{\boldsymbol{\gamma}|X\in\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^2,\boldsymbol{\gamma})\setminus\phi^{\boldsymbol{\varepsilon}}_{\tau}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})\}.
\end{aligned}$$

Because $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1},\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,2}$ are mutually exclusive, naturally we have

$$\Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_1(X)\}=\sum_{\tau=1}^n\Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau)\}+\sum_{\tau=1}^n\Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,2}(X,\tau)\}.$$

We first bound $\Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau)\}$. Now suppose that $\boldsymbol{\gamma}_{-X}$ is the vector of γ such that all the entries are fixed by entries of $\boldsymbol{\gamma}$ except γ_X. Furthermore, the corresponding sub-event of $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(u,\tau)$ is defined as $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau,\boldsymbol{\gamma}_{-X})\subseteq\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau)$. Similarly, we can define $\mathcal{E}^{\epsilon}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})\subseteq\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau)$ and $\mathcal{E}^{\boldsymbol{\varepsilon}}_{3,1}(X,\tau,\boldsymbol{\gamma}_{-X})\subseteq\mathcal{E}^{\boldsymbol{\varepsilon}}_{3,1}(X,\tau)$.

According to the definitions, it is easy to observe that $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau,\boldsymbol{\gamma}_{-X}) = \mathcal{E}^{\boldsymbol{\varepsilon}}_{3,1}(X,\tau,\boldsymbol{\gamma}_{-X}) \cup \mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})$ and $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,2}(X,\tau,\boldsymbol{\gamma}_{-X}) = \mathcal{E}^{\boldsymbol{\varepsilon}}_{3,2}(X,\tau,\boldsymbol{\gamma}_{-X}) \cup \mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})$. So,

$$\Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau)\} = \Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau)\} \cdot \Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{3,1}(X,\tau)|\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau)\}.$$

Then we also have

$$\Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau,\boldsymbol{\gamma}_{-X})\} = \Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})\} \cdot \Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{3,1}(X,\tau,\boldsymbol{\gamma}_{-X})|\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})\}. \tag{18}$$

Similar equations also holds for $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,2}(X,\tau,\boldsymbol{\gamma}_{-X})$. By the monotonicity of BGLM, obviously, in $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})$, the entry on γ_X must be an interval from some lowest value to 1. Let $\omega^{\boldsymbol{\varepsilon}}_{X,2,0}(\tau,\boldsymbol{\gamma}_{-X})$ to be the lowest value of this interval, then we have

$$\Pr_{\gamma_X\sim\mathcal{U}[0,1]}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})\} = 1 - \omega^{\boldsymbol{\varepsilon}}_{X,2,0}(\tau,\boldsymbol{\gamma}_{-X}). \tag{19}$$

Then we denote that the set of nodes activated by time step i under $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})$ as $\phi^{\boldsymbol{\varepsilon}}_i(\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}))$. Moreover, we first assume that $\omega^{\boldsymbol{\varepsilon}}_{X,2,0}(\tau,\boldsymbol{\gamma}_{-X}) < 1$ or $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}) \neq \emptyset$.

Now we consider the value of

$$\Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{3,1}(X,\tau,\boldsymbol{\gamma}_{-X})|\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})\}.$$

This conditional probability means that conditioned on $\gamma_X \geq \omega^{\boldsymbol{\varepsilon}}_{X,2,0}(\tau,\boldsymbol{\gamma}_{-X})$ and a fixed activated set $\phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}))$ by time $\tau-1$, the probability that X is activated at step τ under one of $\boldsymbol{\theta}^1$ and $\boldsymbol{\theta}^2$ but not both. Then if the event of this conditional probability holds, we have the following inequalities:

$$f_X\left(\sum_{X'\in\phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}))\cap N(X)} \boldsymbol{\theta}^1_{X',X}\right) + \epsilon_X < \gamma_X \leq f_X\left(\sum_{X'\in\phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}))\cap N(X)} \boldsymbol{\theta}^2_{X',X}\right) + \epsilon_X$$

or

$$f_X\left(\sum_{X'\in\phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}))\cap N(X)} \boldsymbol{\theta}^1_{X',X}\right) + \epsilon_X \geq \gamma_X > f_X\left(\sum_{X'\in\phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}))\cap N(X)} \boldsymbol{\theta}^2_{X',X}\right) + \epsilon_X.$$

Therefore, we can get

$$\begin{aligned}&\Pr_{\gamma_X\sim\mathcal{U}[0,1]}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{3,1}(X,\tau,\boldsymbol{\gamma}_{-X}) \cup \mathcal{E}^{\boldsymbol{\varepsilon}}_{3,2}(X,\tau,\boldsymbol{\gamma}_{-X})|\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})\}\\ &= \frac{\left|f_X\left(\sum_{X'\in\phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}))\cap N(X)} \boldsymbol{\theta}^1_{X',X}\right) - f_X\left(\sum_{X'\in\phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}))\cap N(X)} \boldsymbol{\theta}^2_{X',X}\right)\right|}{1-\omega^{\boldsymbol{\varepsilon}}_{X,2,0}(\tau,\boldsymbol{\gamma}_{-X})}\end{aligned}$$

so according to Eq. (18) and Eq. (19), we get

$$\Pr_{\gamma_X\sim\mathcal{U}[0,1]}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau,\boldsymbol{\gamma}_{-X}) \cup \mathcal{E}^{\boldsymbol{\varepsilon}}_{2,2}(X,\tau,\boldsymbol{\gamma}_{-X})\} \leq L^{(1)}_{f_X}\left|\sum_{X'\in\phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}))\cap N(X)} (\boldsymbol{\theta}^1_{X',X} - \boldsymbol{\theta}^2_{X',X})\right|. \tag{20}$$

When $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}) = \emptyset$, both the right side and the left side of the above inequality is zero, so we have this inequality holds in general.

Now we define $\mathcal{E}^{\boldsymbol{\varepsilon}}_{4,0}(X,\tau,\boldsymbol{\gamma}_{-X}) = \{\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{-X},\gamma_X)|X \notin \phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\boldsymbol{\theta}^1,\boldsymbol{\gamma})\}$. Then obviously, we have $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}) \subseteq \mathcal{E}^{\boldsymbol{\varepsilon}}_{4,0}(X,\tau,\boldsymbol{\gamma}_{-X})$ and if $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}) \neq \emptyset$, for $i \leq \tau-1$, $\phi^{\boldsymbol{\varepsilon}}_i(\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X})) = \phi^{\boldsymbol{\varepsilon}}_i(\mathcal{E}^{\epsilon}_{4,0}(X,\tau,\boldsymbol{\gamma}_{-X}))$. Therefore, we can relax Eq. (20) by

$$\Pr_{\gamma_X\sim\mathcal{U}[0,1]}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau,\boldsymbol{\gamma}_{-X}) \cup \mathcal{E}^{\boldsymbol{\varepsilon}}_{2,2}(X,\tau,\boldsymbol{\gamma}_{-X})\} \leq L^{(1)}_{f_X}\left|\sum_{X'\in\phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{4,0}(X,\tau,\boldsymbol{\gamma}_{-X}))\cap N(X)} (\boldsymbol{\theta}^1_{X',X} - \boldsymbol{\theta}^2_{X',X})\right|.$$

This also holds for $\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,0}(X,\tau,\boldsymbol{\gamma}_{-X}) = \emptyset$.

Now we can deduce that

$$\Pr_{\boldsymbol{\gamma}\sim(\mathcal{U}[0,1])^n}\{\mathcal{E}^{\epsilon}_1(X)\} = \int_{\boldsymbol{\gamma}_{-X}\in[0,1]^{n-1}} \sum_{\tau=1}^{n} \Pr_{\gamma_X\sim\mathcal{U}[0,1]}\{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X,\tau,\boldsymbol{\gamma}_{-X}) \cup \mathcal{E}^{\boldsymbol{\varepsilon}}_{2,2}(X,\tau,\boldsymbol{\gamma}_{-X})\}\mathrm{d}\boldsymbol{\gamma}_{-X}$$

$$
\begin{aligned}
&= \sum_{\tau=1}^{n} \int_{\boldsymbol{\gamma}_{-X} \in [0,1]^{n-1}} \Pr_{\gamma_X \sim \mathcal{U}[0,1]} \{\mathcal{E}^{\boldsymbol{\varepsilon}}_{2,1}(X, \tau, \boldsymbol{\gamma}_{-X}) \cup \mathcal{E}^{\boldsymbol{\varepsilon}}_{2,2}(X, \tau, \boldsymbol{\gamma}_{-X})\} \mathrm{d}\boldsymbol{\gamma}_{-X} \\
&\le \sum_{\tau=1}^{n} \int_{\boldsymbol{\gamma}_{-X} \in [0,1]^{n-1}} \left| \sum_{X' \in \phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{4,0}(X,\tau,\boldsymbol{\gamma}_{-X})) \cap N(X)} (\theta^1_{X',X} - \theta^2_{X',X}) \right| L^{(1)}_{f_X} \mathrm{d}\boldsymbol{\gamma}_{-X} \\
&= \sum_{\tau=1}^{n} \mathbb{E}_{\boldsymbol{\gamma}_{-X} \sim (\mathcal{U}[0,1])^{n-1}} \left[\left| \sum_{X' \in \phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{4,0}(X,\tau,\boldsymbol{\gamma}_{-X})) \cap N(X)} (\theta^1_{X',X} - \theta^2_{X',X}) \right| \right] L^{(1)}_{f_X}.
\end{aligned}
$$

Combining this with the fact $\mathcal{E}^{\boldsymbol{\varepsilon}}_0 \subseteq \cup_{X \in \mathbf{X}_{\mathbf{S},Y}} \mathcal{E}^{\boldsymbol{\varepsilon}}_1(X)$, we have

$$
\begin{aligned}
&\left|\sigma(\mathbf{S}, \boldsymbol{\theta}^1) - \sigma(\mathbf{S}, \boldsymbol{\theta}^2)\right| \\
&\le \mathbb{E}_{\boldsymbol{\varepsilon}} \left[\sum_{X \in \mathbf{X}_{\mathbf{S},Y}} \sum_{\tau=1}^{n} \mathbb{E}_{\boldsymbol{\gamma}_{-X} \sim (\mathcal{U}[0,1])^{n-1}} \left[\left| \sum_{X' \in \phi^{\boldsymbol{\varepsilon}}_{\tau-1}(\mathcal{E}^{\boldsymbol{\varepsilon}}_{4,0}(X,\tau,\boldsymbol{\gamma}_{-X})) \cap N(X)} (\theta^1_{X',X} - \theta^2_{X',X}) \right| \right] L^{(1)}_{f_X} \right] \\
&= \mathbb{E} \left[\sum_{X \in \mathbf{X}_{\mathbf{S},Y}} \left| \boldsymbol{V}_X (\boldsymbol{\theta}^1_X - \boldsymbol{\theta}^2_X) \right| L^{(1)}_{f_X} \right]
\end{aligned}
$$

which is what we want. □

C.4 Proof of Regret Bound for Algorithm 1 (Theorem 1)

Before the proof, we propose a lemma in order to bound the sum of $\|\boldsymbol{V}_{t,X}\|_{M^{-1}_{t-1,X}}$ at first. This lemma holds for general sequences of vectors, so can be used not only in this particular proof.

Lemma 9. *Let $\{\boldsymbol{W}_t\}_{t=1}^{\infty}$ be a sequence in $\mathbb{R}^d$ satisfying $\|\boldsymbol{W}_t\| \le \sqrt{d}$. Define $\boldsymbol{W}_0 = \mathbf{0}$ and $M_t = \sum_{i=0}^{t} \boldsymbol{W}_i \boldsymbol{W}_i^{\intercal}$. Suppose there is an integer t_1 such that $\lambda_{\min}(M_{t_1+1}) \ge 1$, then for all $t_2 > 0$,*

$$
\sum_{t=t_1}^{t_1+t_2} \|\boldsymbol{W}_t\|_{M^{-1}_{t-1}} \le \sqrt{2 t_2 d \log(t_2 d + t_1)}.
$$

Proof. As a direct application of Lemma 11 in (Abbasi-Yadkori, Pál, and Szepesvári 2011), we have

$$
\sum_{t=t_1+1}^{t_1+t_2} \|\boldsymbol{W}_t\|^2_{M^{-1}_{t-1}} \le 2 \log \frac{\det M_{t_1+t_2}}{\det M_{t_1}} \le 2d \log\left(\frac{\mathrm{tr}(M_{t_1}) + t_2 d^2}{d}\right) - 2 \log \det M_{t_1}.
$$

Note that $\mathrm{tr}(M_{t_1}) = \sum_{t=1}^{t_1} \mathrm{tr}(\boldsymbol{W}_t \boldsymbol{W}_t^{\intercal}) = \sum_{t=1}^{t_1} \|\boldsymbol{W}_t\|^2 \le t_1 d$ and that $\det(M_{t_1}) = \prod_{i=1}^{d} \lambda_i \ge \lambda^d_{\min}(M_{t_1}) \ge 1$ where $\{\lambda_i\}$ are the eigenvalues of M_{t_1}. Applying Cauchy-Schwarz inequality yields

$$
\sum_{t=t_1+1}^{t_1+t_2} \|\boldsymbol{W}_t\|_{M^{-1}_{t-1}} \le \sqrt{t_2 \sum_{t=t_1+1}^{t_1+t_2} \|\boldsymbol{W}_t\|^2_{M^{-1}_{t-1}}} \le \sqrt{2 t_2 d \log(t_2 d + t_1)}
$$

which is exactly what we want. □

Meanwhile, for the sake of clarity, we reclaim Lecué and Mendelson's inequality here, which is presented in (Nie 2022).

Lemma 10 (Lecué and Mendelson's Inequality)**.** *There exists an absolute constant $c > 0$ such that the following statement holds. Let $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k$ be independent copies of a random vector $\boldsymbol{v} \in \mathbb{R}^d$. Suppose that $\alpha \ge 0$ and $0 < \beta \le 1$ are two real numbers, such that the small-ball probability*

$$
\Pr\left\{ |\boldsymbol{v}^{\intercal} \boldsymbol{z}| > \alpha^{\frac{1}{2}} \right\} \ge \beta
$$

holds for any $\boldsymbol{z}$ in $Sphere(d)$. Suppose that

$$
k \ge \frac{cd}{\beta^2}.
$$

Then we have

$$
\Pr\left\{ \lambda_{\min}\left(\frac{1}{k} \sum_{i=1}^{k} \boldsymbol{v}_i \boldsymbol{v}_i^{\intercal} \right) \le \frac{\alpha\beta}{2} \right\} \le \exp\left(-\frac{k\beta^2}{c} \right).
$$

Then with Lemma 9 and this powerful inequality, we can prove Theorem 1 to bound the regret of Algorithms 1 and 2.

Theorem 1 (Regret Bound of BGLM-OFU)**.** *Under Assumptions 1, 2 and 3, the regret of BGLM-OFU (Algorithms 1 and 2) is bounded as*

$$R(T) = O\left(\frac{1}{\kappa} n L_{\max}^{(1)} \sqrt{DT} \log T\right), \tag{7}$$

where the terms of $o(\sqrt{T})$ are omitted.

Proof. Let H_t be the history of the first t rounds and R_t be the regret in the t^{th} round. By the definition of BGLM and our Algorithm 1, we can deduce that for any $t \leq T_0$, $R_t \leq 1$. Now we consider the case of $t > T_0$. When $t > T_0$, we have

$$\mathbb{E}[R_t | H_{t-1}] = \mathbb{E}[\sigma(\mathbf{S}^{\text{opt}}, \boldsymbol{\theta}^*) - \sigma(\mathbf{S}_t, \boldsymbol{\theta}^*) | H_{t-1}] \tag{21}$$

such that the expectation is taken over the randomness of $\mathbf{S}_t$. Then for $T_0 < t \leq T$, we can define $\xi_{t-1,X}$ for $X \in \mathbf{X} \cup \{Y\}$ as $\xi_{t-1,X} = \left\{\left|\boldsymbol{v}^T(\hat{\boldsymbol{\theta}}_{t-1,X} - \boldsymbol{\theta}^*_X)\right| \leq \rho \cdot \|\boldsymbol{v}\|_{M^{-1}_{t-1,X}}, \forall \boldsymbol{v} \in \mathbb{R}^{|\boldsymbol{Pa}(X)|}\right\}$. According to the settings in Algorithm 1, we can deduce that $\lambda_{\min}(M_{t-1,X}) \geq \lambda_{\min}(M_{T_0,X})$ and by Lecué and Mendelson's inequality (Nie 2022) (conditions of this inequality satisfied according to Lemma 7), we have $\Pr\{\lambda_{\min}(M_{T_0,X}) < R\} \leq \exp(-\frac{T_0\zeta^2}{c})$ where c is a constant. Then we can define $\xi_{t-1} = \wedge_{X \in \mathbf{X} \cup \{Y\}} \xi_{t-1,X}$ and let $\overline{\xi_{t-1}}$ be its complement so by Lemma 1 we have $\Pr\left\{\overline{\xi_{t-1}}\right\} \leq \left(3\delta + \exp\left(-\frac{T_0\zeta^2}{c}\right) + 3\delta \exp\left(-\frac{T_0\zeta^2}{c}\right)\right) n$.

Because under ξ_{t-1}, for any $X \in \mathbf{X} \cup \{Y\}$ and $\boldsymbol{v} \in \mathbb{R}^{|\boldsymbol{Pa}(X)|}$, we have $\left|\boldsymbol{v}^T(\hat{\boldsymbol{\theta}}_{t-1,X} - \boldsymbol{\theta}^*_X)\right| \leq \rho \cdot \|\boldsymbol{v}\|_{M^{-1}_{t-1,X}}$. Therefore, by the definition of $\tilde{\boldsymbol{\theta}}_t$, we have $\sigma(\mathbf{S}_t, \tilde{\boldsymbol{\theta}}_t) \geq \sigma(\mathbf{S}^{\text{opt}}, \boldsymbol{\theta}^*)$ because $\boldsymbol{\theta}^*$ is in our confidence ellipsoid. Therefore,

$$\begin{aligned}
\mathbb{E}[R_t] &\leq \Pr\{\xi_{t-1}\} \cdot \mathbb{E}[\sigma(\mathbf{S}^{\text{opt}}, \boldsymbol{\theta}^*) - \sigma(\mathbf{S}_t, \boldsymbol{\theta}^*)] + \Pr(\overline{\xi_{t-1}}) \\
&\leq \mathbb{E}[\sigma(\mathbf{S}^{\text{opt}}, \boldsymbol{\theta}^*) - \sigma(\mathbf{S}_t, \boldsymbol{\theta}^*)] + \left(3\delta + \exp\left(-\frac{T_0\zeta^2}{c}\right) + 3\delta \exp\left(-\frac{T_0\zeta^2}{c}\right)\right) n \\
&\leq \mathbb{E}[\sigma(\mathbf{S}_t, \tilde{\boldsymbol{\theta}}_t) - \sigma(\mathbf{S}_t, \boldsymbol{\theta}^*)] + \left(3\delta + \exp\left(-\frac{T_0\zeta^2}{c}\right) + 3\delta \exp\left(-\frac{T_0\zeta^2}{c}\right)\right) n.
\end{aligned}$$

Then we need to bound $\sigma(\mathbf{S}_t, \tilde{\boldsymbol{\theta}}_t) - \sigma(\mathbf{S}_t, \boldsymbol{\theta}^*)$ carefully.

According to Lemma 1 and Lemma 2, we can deduce that

$$\begin{aligned}
\mathbb{E}[R_t] &\leq \mathbb{E}\left[\sum_{X \in \mathbf{X}_{\mathbf{S}_t,Y}} \left|\boldsymbol{V}_{t,X}(\tilde{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}^*_X)\right| L_{f_X}^{(1)}\right] + \left(3\delta + \exp\left(-\frac{T_0\zeta^2}{c}\right) + 3\delta \exp\left(-\frac{T_0\zeta^2}{c}\right)\right) n \\
&\leq \mathbb{E}\left[\sum_{X \in \mathbf{X}_{\mathbf{S}_t,Y}} \|\boldsymbol{V}_{t,X}\|_{M^{-1}_{t-1,X}} \left\|\tilde{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}^*_X\right\|_{M_{t-1,X}} L_{f_X}^{(1)}\right] + \left(3\delta + \exp\left(-\frac{T_0\zeta^2}{c}\right) + 3\delta \exp\left(-\frac{T_0\zeta^2}{c}\right)\right) n \\
&\leq 2\rho \cdot \mathbb{E}\left[\sum_{X \in \mathbf{X}_{\mathbf{S}_t,Y}} \|\boldsymbol{V}_{t,X}\|_{M^{-1}_{t-1,X}} L_{f_X}^{(1)}\right] + \left(3\delta + \exp\left(-\frac{T_0\zeta^2}{c}\right) + 3\delta \exp\left(-\frac{T_0\zeta^2}{c}\right)\right) n.
\end{aligned}$$

The last inequality holds because $\left\|\tilde{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}^*_X\right\|_{M_{t-1,X}} \leq \left\|\tilde{\boldsymbol{\theta}}_{t,X} - \hat{\boldsymbol{\theta}}_{t-1,X}\right\|_{M_{t-1,X}} + \left\|\hat{\boldsymbol{\theta}}_{t-1,X} - \boldsymbol{\theta}^*_X\right\|_{M_{t-1,X}} \leq 2\rho$.

Therefore, the total regret can be bounded as

$$R(T) \leq 2\rho \cdot \mathbb{E}\left[\sum_{t=T_0+1}^{T} \sum_{X \in \mathbf{X}_{\mathbf{S}_t,Y}} \|\boldsymbol{V}_{t,X}\|_{M^{-1}_{t-1,X}} L_{f_X}^{(1)}\right] + \left(6\delta + \exp\left(-\frac{T_0\zeta^2}{c}\right)\right) n(T - T_0) + T_0.$$

For convenience, we define $\boldsymbol{W}_{t,X}$ as a vector such that if $X \in S_t$, $\boldsymbol{W}_{t,X} = \mathbf{0}^{|\boldsymbol{Pa}(X)|}$; if $X \notin S_t$, $\boldsymbol{W}_{t,X} = \boldsymbol{V}_{t,X}$. Using Lemma 9, we can get the result:

$$R(T) \leq 2\rho\mathbb{E}\left[\sum_{t=T_0+1}^{T} \sum_{X \in \mathbf{X}_{\mathbf{S}_t,Y}} \|\boldsymbol{V}_{t,X}\|_{M^{-1}_{t-1,X}} L_{f_X}^{(1)}\right] + \left(6\delta + \exp\left(-\frac{T_0\zeta^2}{c}\right)\right) n(T - T_0) + T_0$$

$$
\begin{aligned}
&\le 2\rho\mathbb{E}\left[\sum_{t=T_0+1}^{T}\sum_{X\in\boldsymbol{X}\cup\{Y\}}\|\boldsymbol{W}_{t,X}\|_{M_{t-1,X}^{-1}} L_{f_X}^{(1)}\right]+\left(6\delta+\exp\left(-\frac{T_0\zeta^2}{c}\right)\right)n(T-T_0)+T_0\\
&\le 2\rho\cdot\max_{X\in\mathbf{X}\cup\{Y\}}(L_{f_X}^{(1)})\mathbb{E}\left[\sum_{X\in\boldsymbol{X}\cup\{Y\}}\sqrt{2(T-T_0)|\boldsymbol{Pa}(X)|\log\left((T-T_0)|\boldsymbol{Pa}(X)|+T_0\right)}\right]\\
&\quad+\left(6\delta+\exp\left(-\frac{T_0\zeta^2}{c}\right)\right)n(T-T_0)+T_0\\
&=O\left(\frac{1}{\kappa}n\sqrt{TD}L_{\max}^{(1)}\ln T\right)=\tilde{O}\left(\frac{1}{\kappa}n\sqrt{TD}L_{\max}^{(1)}\right)
\end{aligned}
$$

because $\rho=\frac{3}{\kappa}\sqrt{\log(1/\delta)}$.

□

D Proofs of Lemmas for BLM (Section 5)

Lemma 4. *For any $\boldsymbol{S}\subseteq\boldsymbol{X}$, any value $\mathbf{s}\in\{0,1\}^{|\boldsymbol{S}|}$, we have $\mathbb{E}[Y|do(\mathbf{S}=\mathbf{s})]=\mathbb{E}'[Y|do(\mathbf{S}=\mathbf{s})]$.*

Proof. The lemma can be extended to three more detailed equations as below:

$$
\begin{aligned}
\mathbb{E}[Y|do(\mathbf{S}=\mathbf{s})]&=\sum_{X\in\mathbf{S}}s_X\sum_{P\in\mathcal{P}_{X,Y},P\cap\mathbf{S}=\{X\}}\prod_{e\in P}\theta_e^*+\sum_{P\in\mathcal{P}_{U_0,Y},P\cap\mathbf{S}=\emptyset}\prod_{e\in P}\theta_e^* &(22)\\
&=\sum_{X\in\mathbf{S}}s_X\sum_{P\in\mathcal{P}'_{X,Y},P\cap\mathbf{S}=\{X\}}\prod_{e\in P}\theta_e^{*'}+\sum_{P\in\mathcal{P}'_{X_1,Y},P\cap\mathbf{S}=\emptyset}\prod_{e\in P}\theta_e^{*'} &(23)\\
&=\mathbb{E}'[Y|do(\mathbf{S}=\mathbf{s})], &(24)
\end{aligned}
$$

where the notation $P\cap\mathbf{S}$ means that intersection of the node set of the path P and the node set $\boldsymbol{S}$.

In order to prove these three equations, we firstly need to prove that for an arbitrary node $X\in\boldsymbol{U}\cup\mathbf{X}\cup\{Y\}$, we have

$$
\mathbb{E}[X|do(\mathbf{S}=\mathbf{s})]=\begin{cases}\sum_{Z\in\boldsymbol{Pa}(X)}\mathbb{E}[Z|do(\mathbf{S}=\mathbf{s})]\cdot\theta_{Z,X}^* & X\notin\boldsymbol{S},\\ s_X & X\in\boldsymbol{S},\end{cases} \tag{25}
$$

where the case of $X\notin\boldsymbol{S}$ is because

$$
\begin{aligned}
\mathbb{E}[X|do(\mathbf{S}=\mathbf{s})]&=\Pr\left\{X=1|do(\mathbf{S}=\mathbf{s})\right\}\\
&=\mathbb{E}[\boldsymbol{pa}(X)\cdot\theta_X^*|do(\mathbf{S}=\mathbf{s})]\\
&=\sum_{Z\in\boldsymbol{Pa}(X)}\mathbb{E}[Z|do(\mathbf{S}=\mathbf{s})]\cdot\theta_{Z,X}^*.
\end{aligned}
$$

By recursively using Eq. (25) to replace the expectation of a node by the expectations of its parents in the expression of $\mathbb{E}[Y|do(\mathbf{S}=\mathbf{s})]$, we can get Eq. (22). Similarly, the Eq. (24) can be obtained.

Finally, the Eq. (23) can be proved by

$$
\begin{aligned}
&\sum_{X\in\mathbf{S}}s_X\sum_{P\in\mathcal{P}_{X,Y},P\cap\mathbf{S}=\{X\}}\prod_{e\in P}\theta_e^*+\sum_{P\in\mathcal{P}_{U_0,Y},P\cap\mathbf{S}=\emptyset}\prod_{e\in P}\theta_e^*\\
&=\sum_{X\in\mathbf{S}}s_X\sum_{P\in\mathcal{P}'_{X,Y},P\cap\mathbf{S}=\{X\}}\prod_{e\in P}\theta_e^{*'}+\sum_{X\in\mathbf{X}\cup\{Y\}\setminus\mathbf{S}}\Pr\left\{X=1|do\left(\mathbf{X}\cup\{Y\}\setminus\{X\}=\mathbf{0}\right)\right\}\sum_{P\in\mathcal{P}'_{X,Y},P\cap\mathbf{S}=\emptyset}\prod_{e\in P}\theta_e^{*'} &(26)\\
&=\sum_{X\in\mathbf{S}}s_X\sum_{P\in\mathcal{P}'_{X,Y},P\cap\mathbf{S}=\{X\}}\prod_{e\in P}\theta_e^{*'}+\sum_{P\in\mathcal{P}'_{X_1,Y},P\cap\mathbf{S}=\emptyset}\prod_{e\in P}\theta_e^{*'}. &(27)
\end{aligned}
$$

Now we explain why Eq. (26) and Eq. (27) hold. Firstly, we illustrate why for every $X\in\boldsymbol{S}$,

$$
\sum_{P\in\mathcal{P}_{X,Y},P\cap\mathbf{S}=\{X\}}\prod_{e\in P}\theta_e^*=\sum_{P\in\mathcal{P}'_{X,Y},P\cap\mathbf{S}=\{X\}}\prod_{e\in P}\theta_e^{*'}. \tag{28}
$$

To show the above equality, we first look at the right-hand side, and notice that for each path $P'\in\mathcal{P}'_{X,Y}$, it corresponds to a collection of paths from X to Y in G, that is, $\mathcal{P}_{X,Y}$. Then each $\theta_{P'}^{*'}=\prod_{e\in P'}\theta_e^{*'}$ is the sum of the θ_P^* values for some paths

$P \in \mathcal{P}_{X,Y}$ according to our G' construction. This means that the RHS $\sum_{P' \in \mathcal{P}'_{X,Y}, P' \cap \mathbf{S} = \{X\}} \prod_{e \in P'} \theta_e^{*'}$ is a summation of θ_P^* values for some paths $P \in \mathcal{P}_{X,Y}$. Moreover, these paths only intersect with $\boldsymbol{S}$ at the starting node X, in both G and G'.

Next, Suppose P_0 is an arbitrary path in $\mathcal{P}_{X,Y}$ such that $P_0 \cap \boldsymbol{S} = \{X\}$. With the above observation, we only need to show that in $\sum_{P \in \mathcal{P}'_{X,Y}, P \cap \mathbf{S} = \{X\}} \prod_{e \in P} \theta_e^{*'}$, $\theta_{P_0}^* = \prod_{e \in P} \theta_e^*$ has been counted exactly once. Suppose P_0 has the form of $X_{i_1} \to U_{1,1} \to \cdots \to U_{1,j_1} \to X_{i_2} \to U_{2,1} \to \cdots \to U_{2,j_2} \to \cdots \to X_{i_k}$, where X's represent observed variables and U's represent unobserved variables. Then according to our transformation, $\theta_{P_0}^*$ is contained in the expansion of $\theta_{X_{i_1}, X_{i_2}}^{*'} \cdots \theta_{X_{i_{k-1}}, X_{i_k}}^{*'}$, namely $(\sum_{P \in \mathcal{P}_{X_{i_1}, X_{i_2}}} \theta_P^*) \cdots (\sum_{P \in \mathcal{P}_{X_{i_{k-1}}, X_{i_k}}} \theta_P^*)$. Moreover, we know that the paths in $\mathcal{P}'_{X,Y}$ only contain observed nodes, so in other terms of the expansion of $\sum_{P \in \mathcal{P}'_{X,Y}, P \cap \mathbf{S} = \{X\}} \prod_{e \in P} \theta_e^{*'}$, we can not find another $\theta_{P_0}^*$. Therefore, Eq. (28) holds.

Furthermore, according to Eq. (22), we have for every $X \in \boldsymbol{X} \cup \{Y\}$,

$$\Pr\{X = 1 | do\,(\mathbf{X} \cup \{Y\} \setminus \{X\} = \mathbf{0})\} = \sum_{P \in \mathcal{P}_{U_0, X}, P \cap (\boldsymbol{X} \cup \{Y\}) = \{X\}} \theta_P^*.$$

For an arbitrary path P_0 in $\mathcal{P}_{U_0,Y}$ and $P_0 \cap \boldsymbol{S} = \emptyset$, we claim that it will be added exactly once in

$$\sum_{X \in \mathbf{X} \cup \{Y\} \setminus \mathbf{S}} \Pr\{X = 1 | do\,(\mathbf{X} \cup \{Y\} \setminus \{X\} = \mathbf{0})\} \sum_{P \in \mathcal{P}'_{X,Y}, P \cap \mathbf{S} = \emptyset} \prod_{e \in P} \theta_e^{*'}. \tag{29}$$

In fact, suppose X_i is the first observed node in P_0. Then $X_i \in \boldsymbol{X} \cup \{Y\} \setminus \boldsymbol{S}$, and P_0 will be added in $\Pr\{X_i = 1 | do\,(\mathbf{X} \cup \{Y\} \setminus \{X_i\} = \mathbf{0})\} \sum_{P \in \mathcal{P}'_{X_i,Y}, P \cap \mathbf{S} = \emptyset} \prod_{e \in P} \theta_e^{*'}$ for the similar reason of the proof of Eq. (28) by expansions. Therefore, up to now , we have proved Eq. (26).

Now we consider how Eq. (29) equals to $\sum_{P \in \mathcal{P}'_{X_1,Y}, P \cap S = \emptyset} \theta_P^{*'}$. Actually, for a path P_0 in G' from X_1 to Y, it must have the form $X_1 = X_{i_1} \to X_{i_2} \to \cdots \to Y$. Therefore, $\theta_{P_0}^{*'}$ is equal to $\theta_{X_1, X_{i_2}}^{*'} \theta_{X_{i_2} \to \cdots Y}^{*'}$. We already know that $\theta_{X_1, X_{i_2}}^{*'} = \Pr\{X_{i_2} = 1 | do\,(\mathbf{X} \cup \{Y\} \setminus \{X_{i_2}\} = \mathbf{0})\}$ by the definition of our transformation. Simultaneously, $\theta_{X_{i_2} \to \cdots Y}^{*'}$ is included exactly once in $\sum_{P \in \mathcal{P}'_{X_{i_2},Y}, P \cap \mathbf{S} = \emptyset} \prod_{e \in P} \theta_e^{*'}$, so we have Eq. (27) proved. □

Lemma 3. *For any $X \in \mathbf{X} \cup \{Y\}$, any $\boldsymbol{S} \subseteq \boldsymbol{X}$, any value $\boldsymbol{pa}'(X) \in \{0,1\}^{|\boldsymbol{Pa}'(X)|}$, any value $\boldsymbol{s} \in \{0,1\}^{|\boldsymbol{S}|}$ ($\boldsymbol{s}$ is consistent with $\boldsymbol{pa}'(X)$ on values in $\boldsymbol{S} \cap \boldsymbol{Pa}'(X)$), we have*

$$\begin{aligned}&\Pr\left\{X = 1 | \boldsymbol{Pa}'(X) \setminus \{X_1\} = \boldsymbol{pa}'(X) \setminus \{x_1\}, do(\boldsymbol{S} = \boldsymbol{s})\right\} \\ &= \Pr{}'\left\{X = 1 | \boldsymbol{Pa}'(X) = \boldsymbol{pa}'(X), do(\boldsymbol{S} = \boldsymbol{s})\right\}.\end{aligned}$$

Proof. In order to prove this, we firstly prove that

$$\mathbb{E}\left[X | do\left(\boldsymbol{Pa}'(X) \setminus \{X_1\} = \boldsymbol{pa}'(X) \setminus \{x_1\}, \boldsymbol{S} = \boldsymbol{s}\right)\right] = \mathbb{E}'\left[X | do\left(\boldsymbol{Pa}'(X) = \boldsymbol{pa}'(X), \boldsymbol{S} = \boldsymbol{s}\right)\right], \tag{30}$$

In fact, this can be seen as a direct application of Lemma 4 if we replace the Y in Lemma 4 by the node X here.

Next, we want to apply the second rule of do-calculus (Pearl 2009, 2012) to show the following:

$$\begin{aligned}&\mathbb{E}'\left[X | do\left(\boldsymbol{Pa}'(X) = \boldsymbol{pa}'(X), \boldsymbol{S} = \boldsymbol{s}\right)\right] \\ &= \Pr{}'\left\{X = 1 | do\left(\boldsymbol{Pa}'(X) = \boldsymbol{pa}'(X), \boldsymbol{S} = \boldsymbol{s}\right)\right\} \\ &= \Pr{}'\left\{X = 1 | \boldsymbol{Pa}'(X) = \boldsymbol{pa}'(X), do(\boldsymbol{S} = \boldsymbol{s})\right\},\end{aligned} \tag{31}$$

where the second equality applies the do-calculus rule. According to the rule, the equality holds when X and $\boldsymbol{Pa}'(X)$ is independent in the causal graph G' after removing all the outgoing edges of $\boldsymbol{Pa}(X)$ and all the incoming edges of $\boldsymbol{S}$. Since G' is Markovian with only observed nodes, after removing all outgoing edges of the parent nodes of X, it is certainly true that X is independent of $\boldsymbol{Pa}(X)$.

Finally, we want to show the following, again by the the second rule of do-calculus:

$$\begin{aligned}&\mathbb{E}\left[X | do\left(\boldsymbol{Pa}'(X) \setminus \{X_1\} = \boldsymbol{pa}'(X) \setminus \{x_1\}, \boldsymbol{S} = \boldsymbol{s}\right)\right] \\ &= \Pr\left\{X = 1 | do\left(\boldsymbol{Pa}'(X) \setminus \{X_1\} = \boldsymbol{pa}'(X) \setminus \{x_1\}, \boldsymbol{S} = \boldsymbol{s}\right)\right\} \\ &= \Pr\left\{X = 1 | \boldsymbol{Pa}'(X) \setminus \{X_1\} = \boldsymbol{pa}'(X) \setminus \{x_1\}, do(\boldsymbol{S} = \boldsymbol{s})\right\}.\end{aligned} \tag{32}$$

Note that the above is on the original graph G. To show the above equality according to the second rule of do-calculus, we need to show that X and $\boldsymbol{Pa}'(X) \setminus \{X_1\}$ are independent in the graph G after removing the outgoing edges of $\boldsymbol{Pa}'(X) \setminus \{X_1\}$

and the incoming edges of $\boldsymbol{S}$. Call this trimmed graph $\tilde{G}$. According to the properties of the causal diagram (Pearl 2009), if X and $\boldsymbol{Pa}'(X) \setminus \{X_1\}$ are not independent in $\tilde{G}$, there must exist an active path P between X and one of the nodes $Z \in \boldsymbol{Pa}'(X) \setminus \{X_1\}$. Since Z's outgoing edges are cut off, the edge connecting to Z in path P must be pointing towards Z. On the path P, it cannot be that all the edges pointing in the direction from X to Z, since this would violate the DAG assumption of G. Then, there is a first node W on the path from Z to X such that from W to Z all the edges pointing in the direction from W to Z, and on the segment from W to X, the edge connecting to W is also leaving W and pointing towards X, i.e. W is a fork (Definition 1.2.3 of (Pearl 2009)). If on the segment from W to X, there is a node V with two incoming edges pointing to V on the path, this means V is a collider (Pearl 2009). Since V cannot be in $\boldsymbol{S}$, otherwise, the incoming edges of V would have been trimmed in $\tilde{G}$, then this collider V would block the path making P not an active path. Therefore, no such collider exists on the path from W to X, and all edges on W to X point in the direction from W to X. If there is an observed variable X_i on the path from W to X other than X itself, then find the X_i that is closest to X on P. This means that the segment of P from X_i to X is a hidden path, which implies that X_i would become a parent of X in G', i.e. $X_i \in \boldsymbol{Pa}'(X) \setminus \{X_1\}$. But this implies that the outgoing edges of X_i should have been cut off, a contradiction. Therefore all nodes on the path segment from W to X except X are hidden variables. Then let U be the hidden variable W, and on the path segment from U to Z, let X_i be the first observed variable. We thus find an unobserved variable U connecting to both X_i and X with paths consisting of only hidden variables, except X_i and X, and X is a descendant of X_i. But in Section 5 we already state that we exclude such situation in graph G. Hence, no active path can exist between Z and W, and thus Eq. (32) holds.

The lemma is proved with Eqs. (30), (31) and (32). □

E Proof of Regret Bound for Algorithm 3 (Theorem 3)

We rewrite Lemma 1 in (Li et al. 2020) as Lemma 11 as below.

Lemma 11 (Lemma 1 in (Li et al. 2020))**.** *Given $\{\boldsymbol{V}_t, X^t\}_{t=1}^{\infty}$ with $\boldsymbol{V}_t \in \{0,1\}^d$ and $X^t \in \{0,1\}$ as a Bernoulli random variable with $\mathbb{E}[X^t|\boldsymbol{V}_1, X^1, \cdots, \boldsymbol{V}_{t-1}, X^{t-1}, \boldsymbol{V}_t] = \boldsymbol{V}_t^{\intercal}\boldsymbol{\theta}$ where $\boldsymbol{\theta} \in [0,1]^d$, let $M_t = \mathbf{I} + \sum_{i=1}^{t} \boldsymbol{V}_i \boldsymbol{V}_i^{\intercal}$ and $\hat{\boldsymbol{\theta}}_t = M_t^{-1}(\sum_{i=1}^{t} \boldsymbol{V}_i X^i)$ be the linear regression estimator. Then with probability at least $1-\delta$, for all $t \geq 1$, it holds that $\boldsymbol{\theta}$ lies in the confidence region*

$$\left\{\boldsymbol{\theta}' \in [0,1]^d : \left\|\boldsymbol{\theta}' - \hat{\boldsymbol{\theta}}_t\right\|_{M_t} \leq \sqrt{d\log(1+td) + 2\log\frac{1}{\delta}} + \sqrt{d}\right\}.$$

With Lemma 11, we are able to prove the regret bound of Algorithm 3.

Theorem 3 (Regret Bound of Algorithm 3)**.** *The regret of BLM-LR (Algorithm 3) running on BLM with hidden variables is bounded as*

$$R(T) = O\left(n^2\sqrt{DT}\log T\right).$$

Proof. In Section 5, we have already shown Lemmas 3 and 4, therefore, we can deduce that by seeing G as the Markovian graph G', Lemma 1 in (Li et al. 2020) and Theorem 2 still holds for $\boldsymbol{\theta}^{*'}$.

With probability at most $n\delta$, event $\left\{\exists t \leq T, x \in \boldsymbol{X} \cup \{Y\} : \left\|\boldsymbol{\theta}_X^{*'} - \hat{\boldsymbol{\theta}}_{t,X}\right\| > \rho_t\right\}$ occurs. Next we bound the regret conditioned on the absence of this event. In detail, according to Theorem 1 in (Li et al. 2020) and Theorem 2, we can deduce that

$$\begin{aligned}
\mathbb{E}\left[R_t\right] &= \mathbb{E}\left[\sigma'(\boldsymbol{S}^{\mathrm{opt}}, \boldsymbol{\theta}^{*'}) - \sigma'(\boldsymbol{S}_t, \boldsymbol{\theta}^{*'})\right] \\
&\leq \mathbb{E}\left[\sigma'(\boldsymbol{S}_t, \tilde{\boldsymbol{\theta}}_t) - \sigma'(\boldsymbol{S}_t, \boldsymbol{\theta}^{*'})\right] \\
&\leq \mathbb{E}\left[\sum_{X \in \boldsymbol{X}_{\boldsymbol{S}_t, Y}} \left|\boldsymbol{V}_{t,X}^{\intercal}(\tilde{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}_X^{*'})\right|\right] \\
&\leq \mathbb{E}\left[\sum_{X \in \boldsymbol{X}_{\boldsymbol{S}_t, Y}} \left\|\boldsymbol{V}_{t,X}\right\|_{M_{t-1,X}^{-1}} \left\|\tilde{\boldsymbol{\theta}}_{t,X} - \boldsymbol{\theta}_X^{*'}\right\|_{M_{t-1,X}}\right] \\
&\leq \mathbb{E}\left[\sum_{X \in \boldsymbol{X}_{\boldsymbol{S}_t, Y}} 2\rho_{t-1} \left\|\boldsymbol{V}_{t,X}\right\|_{M_{t-1,X}^{-1}}\right],
\end{aligned}$$

since $\tilde{\boldsymbol{\theta}}_{t,X}, \boldsymbol{\theta}^*_X$ are both in the confidence set. Thus, we have

$$R(T) = \mathbb{E}\left[\sum_{t=1}^{T} R_t\right] \leq 2\rho_T \cdot \mathbb{E}\left[\sum_{t=1}^{T} \sum_{X \in \boldsymbol{X}_{\boldsymbol{S}_t,Y}} \|\boldsymbol{V}_{t,X}\|_{M_{t-1,X}^{-1}}\right].$$

For convenience, we define $\boldsymbol{W}_{t,X}$ as a vector such that if $X \in S_t$, $\boldsymbol{W}_{t,X} = \mathbf{0}^{|\boldsymbol{Pa}(X)|}$; if $X \notin S_t$, $\boldsymbol{W}_{t,X} = \boldsymbol{V}_{t,X}$. According to Cauchy-Schwarz inequality, we have

$$\begin{aligned} R(T) &\leq 2\rho_T \cdot \mathbb{E}\left[\sum_{t=1}^{T} \sum_{X \in \boldsymbol{X} \cup \{Y\}} \|\boldsymbol{W}_{t,X}\|_{M_{t-1,X}^{-1}}\right] \\ &\leq 2\rho_T \cdot \mathbb{E}\left[\sqrt{T} \cdot \sum_{X \in \boldsymbol{X} \cup \{Y\}} \sqrt{\sum_{t=1}^{T} \|\boldsymbol{W}_{t,X}\|^2_{M_{t-1,X}^{-1}}}\right]. \end{aligned}$$

Note that $M_{t,X} = M_{t-1,X} + \boldsymbol{W}_{t,X}\boldsymbol{W}_{t,X}^{\intercal}$ and therefore, $\det(M_{t,X}) = \det(M_{t-1,X})\left(1 + \|\boldsymbol{W}_{t,X}\|^2_{M_{t-1,X}^{-1}}\right)$, we have

$$\begin{aligned} \sum_{t=1}^{T} \|\boldsymbol{W}_{t,X}\|^2_{M_{t-1,X}^{-1}} &\leq \sum_{t=1}^{T} \frac{n}{\log(n+1)} \cdot \log\left(1 + \|\boldsymbol{W}_{t,X}\|^2_{M_{t-1,X}^{-1}}\right) \\ &\leq \frac{n}{\log(n+1)} \cdot \log \frac{\det(M_{T,X})}{\det(\mathbf{I})} \\ &\leq \frac{n|\boldsymbol{Pa}(X)|}{\log(n+1)} \cdot \log \frac{\operatorname{tr}(M_{T,X})}{|\boldsymbol{Pa}(X)|} \\ &\leq \frac{n|\boldsymbol{Pa}(X)|}{\log(n+1)} \cdot \log\left(1 + \sum_{t=1}^{T} \frac{\|\boldsymbol{W}_{t,X}\|_2^2}{|\boldsymbol{Pa}(X)|}\right) \\ &\leq \frac{nD}{\log(n+1)} \log(1+T). \end{aligned}$$

Therefore, the final regret $R(T)$ is bounded by

$$\begin{aligned} R(T) &\leq 2\rho_T n \sqrt{T \frac{nD}{\log(n+1)} \log(1+T)} \\ &= O\left(n^2 \sqrt{DT} \log T\right), \end{aligned}$$

because $\rho_T = \sqrt{n \log(1 + Tn) + 2 \log \frac{1}{\delta}} + \sqrt{n}$. When $\left\{\exists t \leq T, x \in \boldsymbol{X} \cup \{Y\} : \left\|\boldsymbol{\theta}^{*'}_X - \hat{\boldsymbol{\theta}}_{t,X}\right\| > \rho_t\right\}$ does occur, the regret is no more than T. Therefore, the total regret is still $O\left(n^2\sqrt{DT}\log T\right)$. □

F Extensions to Causal Model With Continuous Variables

We consider linear models with continuous variables. In linear models, the propagation is defined as $X = \boldsymbol{\theta}^*_X \cdot \boldsymbol{pa}(X) + \varepsilon_X$ instead of $P(X = 1|\boldsymbol{Pa}(X) = \boldsymbol{pa}(X)) = \boldsymbol{\theta}^*_X \cdot \boldsymbol{pa}(X) + \varepsilon_X$. We still need ε_X to be a zero-mean sub-Gaussian noise that ensures that $X \in [0, 1]$. According to Theorem 2 in (Abbasi-Yadkori, Pál, and Szepesvári 2011), Lemma 11 still holds for continuous $\boldsymbol{V}_t$'s and X^t's. Formally, we have Lemma 12.

Lemma 12. *Given $\{\boldsymbol{V}_t, X^t\}_{t=1}^{\infty}$ with $\boldsymbol{V}_t \in [0, 1]^d$ and $X^t \in [0, 1]$ as a Bernoulli random variable with $\mathbb{E}[X^t|\boldsymbol{V}_1, X^1, \cdots, \boldsymbol{V}_{t-1}, X^{t-1}, \boldsymbol{V}_t] = \boldsymbol{V}_t^{\intercal}\boldsymbol{\theta}$ where $\boldsymbol{\theta} \in [0, 1]^d$, let $M_t = \mathbf{I} + \sum_{i=1}^{t} \boldsymbol{V}_i \boldsymbol{V}_i^{\intercal}$ and $\hat{\boldsymbol{\theta}}_t = M_t^{-1}(\sum_{i=1}^{t} \boldsymbol{V}_i X^i)$ be the linear regression estimator. Then with probability at least $1 - \delta$, for all $t \geq 1$, it holds that $\boldsymbol{\theta}$ lies in the confidence region*

$$\left\{\boldsymbol{\theta}' \in [0, 1]^d : \left\|\boldsymbol{\theta}' - \hat{\boldsymbol{\theta}}_t\right\|_{M_t} \leq \sqrt{d \log(1 + td) + 2 \log \frac{1}{\delta}} + \sqrt{d}\right\}.$$

Simultaneously, Lemma 3 and Lemma 4 also hold for linear models without any modification on their proofs. Therefore, we can still use BLM-LR (Algorithm 3) on continuous linear models and get the same regret guarantee. Formally, we have the following theorem.

Theorem 4 (Regret Bound of Algorithm 3 on Linear Models)**.** *theorem The regret of BLM-LR (Algorithm 3) running on linear model with hidden variables is bounded as*

$$R(T) = O\left(n^2\sqrt{DT}\log T\right).$$

The proof of Theorem 4 is completely the same as the proof of Theorem 3, which is given in Appendix E. It is worthy noting that the GOM bounded smoothness condition still holds for linear models. This is because the expectation of reward node Y in a linear model is the same comparing to the BLM with same parameters and skeleton.

G Hardness of Offline CCB Problems

In order to illustrate the reasonableness of our offline oracles which are used in the online algorithms, we show the hardness of CCB problems for BGLMs here. As a preparation, we already know that Maximum k-Vertex Cover (Max k-VC) problem is a NP-hard problem because it contains the vertex covering problem as a sub-problem, which is in Karp's original list of 21 NP-complete problems (Karp 1972). In the following theorem, we prove that the offline version of BGLM CCB problem is NP hard by reducing Max k-VC problem to this problem.

Theorem 5 (NP Hardness of Offline BGLM CCB Problem)**.** *With Assumptions 1 and 2 holds for the BGLM G, the offline version of BGLM CCB problem is NP hard.*

Proof. We consider an arbitrary instance of the set covering problem $(\mathcal{S}, \boldsymbol{X}', k)$. Suppose all the elements are $X_1', \cdots, X_l'$, all the sets are $\boldsymbol{S}_1, \cdots, \boldsymbol{S}_r$ and the final target is to find k sets such that the number of covered elements is maximized. Then we can create a two bipartite graph $G = (\mathbf{V}, E)$ such that $X_1, \cdots, X_l$ are on the right side and $Z_1, \cdots, Z_r$ are on the left side. Although the model is Markovian, the endogenesis distribution for all of the nodes is set as a zero distribution, i.e. without interventions, all the nodes will always be 0. For every X_i, the function that decides it is set by $f_{X_i}(\boldsymbol{pa}(X_i) \cdot \boldsymbol{\theta}^*_{X_i}) = 1 - \frac{1}{\alpha pa(X_i) \cdot \boldsymbol{\theta}^*_{X_i} + 1}$ and the j^{th} entry of $\boldsymbol{\theta}^*_{X_i}$ is 1 if and only if $X_i \in \boldsymbol{S}_j$ (otherwise, the entry is set by 0). This function obviously satisfies the two assumptions when $\kappa \leq \frac{\alpha}{(\alpha n+1)^2}$, $L^{(1)}_{f_{X_i}} >= \alpha$ and $L^{(2)}_{f_{X_i}} >= 0$. For the reward node Y, we define $f_Y((x_1, \cdots, x_l) \cdot \boldsymbol{\theta}^*_Y) = \frac{1}{l}(x_1 + \cdots + x_l)$, which means that $\boldsymbol{\theta}^*_Y = \mathbf{1}^l$. Until now, we get an instance (G, k) for the offline BGLM CCB problem.

If the optimal solution of the BGLM CCB problem instance is $X_{i_1}, \cdots, X_{i_t}$ and $Z_{j_1}, \cdots, Z_{j_{k-t}}$. We choose the sets corresponding to this nodes in $(\mathcal{S}, \boldsymbol{X}', k)$ and replace $X_{i_1}, \cdots, X_{i_t}$ by the corresponding t sets that contain $X'_{i_1}, \cdots, X'_{i_t}$ in the set covering instance. Suppose the number of covered elements of this strategy is K. Then the expected reward $\mathbb{E}[Y]$ when do interventions on $X_{i_1}, \cdots, X_{i_t}$ and $Z_{j_1}, \cdots, Z_{j_{k-t}}$ is at most $\frac{K}{l}$. Assume that K is not the optimal solution for the set covering instance $(\mathcal{S}, \boldsymbol{X}', k)$, then there exist k sets such that the number of covered elements is at least $K+1$. Doing interventions to 1 on corresponding Z_j's of these sets in the BGLM CCB problem instance (G, k), the expected reward is at least $\frac{1}{l}(K+1)(1 - \frac{1}{\alpha+1})$. Therefore, if $\frac{1}{l}(K+1)(1-\frac{1}{\alpha+1}) > \frac{K}{l}$, we get a contradiction, which can be realized by setting $\alpha = n + 1 > K$. Therefore, we can use an oracle of the offline BGLM CCB problem to solve the set covering problem. □

Finally, we prove that BGLM CCB problem has monotonicity. Formally, we propose Theorem 6.

Theorem 6. *In the offline version of BGLM CCB problem, the expected reward $\mathbb{E}[Y]$ is monotonous with respect to the interventions.*

Proof. Because each f_X is monotonous, we can deduce that BGLM is in the family of general cascade model (general threshold model is equivalent to general cascade model (Kempe, Kleinberg, and Tardos 2003)) (Wu et al. 2018). Therefore, we can create live-edge graphs.

For an arbitrary live-edge graph, if $\boldsymbol{S}$ are intervened to 1 and Y can be reached from one of these nodes through a directed active path, then Y can be obviously reached from nodes in $\boldsymbol{S} \cup \{X\}$ through the same path. Therefore, $\mathbb{E}[Y | do(\boldsymbol{S}) = \mathbf{1}^{|\boldsymbol{S}|}] \leq \mathbb{E}[Y | do(\boldsymbol{S}) = \mathbf{1}^{|\boldsymbol{S}|}, do(X) = 1]$. Moreover, if we intervene one node to 0, it is equivalent to cut off all the paths passing by that node to Y, so the value of Y can only decrease. Until now, we have proved the monotonicity of BGLM CCB problem. □

H Simulations

H.1 Efficient Pair-Oracle for BLMs

Before the introduction of our settings, we propose a computational efficient algorithm to replace the pair-oracle (line 9 of Algorithm 1) for binary linear models without unobserved nodes. For convenience, suppose $X_1, X_2, \cdots, X_{n-1}, Y$ is a topological order. The implementation is of $O(\binom{n}{K} D^3 n)$ time complexity. See Algorithm 4 for details.

In order to apply Lagrange multiplier method (De la Fuente 2000), we remove the $[0,1]$-boundaries of $\boldsymbol{\theta}'_{t,X}$ in the definition of ellipsoid $\mathcal{C}_{t,X}$ defined in Algorithm 3 in this section. This does not have impact on our regret analysis and in this section, we adopt $\mathcal{C}_{t,X} = \left\{\boldsymbol{\theta}'_X \in \mathbb{R}^{|\boldsymbol{Pa}(X)|} : \left\|\boldsymbol{\theta}'_X - \hat{\boldsymbol{\theta}}_{t-1,X}\right\|_{M_{t-1,X}} \leq \rho_{t-1}\right\}$. Then we have the following theorem to make sure correctness of Algorithm 4.

Algorithm 4: An Implementation of Pair-Oracle for BLMs

```
**Input:** Graph $G = (\boldsymbol{X} \cup \{Y\}, E)$, intervention budget $K \in \mathbb{N}$, estimated parameters $\hat{\boldsymbol{\theta}}_{t-1}$, current $\rho_{t-1}$, observation matrices $M_{t-1,X}$ for $X \in \boldsymbol{X} \cup \{Y\}$.
**Output:** Intervened set in the $t^{th}$ round $\boldsymbol{S}_t$.
for $\boldsymbol{S} \subseteq \boldsymbol{X}$ such that $|\boldsymbol{S}| = K$ do
  for $X = X_2, X_3, \cdots, X_{n-1}, Y$ do
    $\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[\boldsymbol{Pa}(X)|do(\boldsymbol{S})] = \left(\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X_{i_1}|do(\boldsymbol{S})], \cdots, \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X_{i_{|\boldsymbol{Pa}(X)|}}|do(\boldsymbol{S})]\right)^\intercal$ where $\boldsymbol{Pa}(X) = \left\{X_{i_1}, \cdots, X_{i_{|\boldsymbol{Pa}(X)|}}\right\}$.
    if $X \notin \boldsymbol{S}$ then
      $\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X|do(\boldsymbol{S})] = \rho_{t-1}\sqrt{\left(\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[\boldsymbol{Pa}(X)|do(\boldsymbol{S})]\right)^\intercal M_{t-1,X}^{-1}\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[\boldsymbol{Pa}(X)|do(\boldsymbol{S})]} + \left(\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[\boldsymbol{Pa}(X)|do(\boldsymbol{S})]\right)^\intercal \hat{\boldsymbol{\theta}}_{t-1,X}$.
    else
      $\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X|do(\boldsymbol{S})] = 1$.
    end if
  end for
end for
return $\boldsymbol{S}_t = \operatorname{argmax}_{\boldsymbol{S}} \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[Y|do(\boldsymbol{S})]$.
```

Theorem 7 (Correctness of Algorithm 4)**.** *The intervened set S_t we get in Algorithm 4 is exactly what we get from* $\operatorname{argmax}_{\boldsymbol{S}\subseteq\boldsymbol{X},|\boldsymbol{S}|\le K,\boldsymbol{\theta}'_{t,X}\in\mathcal{C}_{t,X}} \mathbb{E}[Y|do(\boldsymbol{S})]$ *for Algorithm 3. Moreover, for any $i = 1, 2, \cdots, n$ and $\boldsymbol{S} \subseteq \boldsymbol{X}$, we have*

$$\max_{\boldsymbol{\theta}'_{t,X_j}\in\mathcal{C}_{t,X_j},j\le i} \mathbb{E}[X_i|do(\boldsymbol{S})] = \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X_i|do(\boldsymbol{S})] \tag{33}$$

for $\widetilde{\boldsymbol{\theta}}_t$ corresponding to each $\boldsymbol{S}$ in the first loop of Algorithm 4. Hence, when $\boldsymbol{S}$ equals the selected seed node set $\boldsymbol{S}_t$, the corresponding $\widetilde{\boldsymbol{\theta}}_t$ in the loop of Algorithm 4 is a feasible solution of $\tilde{\boldsymbol{\theta}}_t$ for Algorithm 3.

Proof. Initially, we prove that $\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X_k|do(\boldsymbol{S})] = \max_{\boldsymbol{\theta}'_{t,X_k}\in\mathcal{C}_{t,X_k}} \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_{t,X_i},i<k}[X_k|do(\boldsymbol{S})]$. Here, the subscript of $\mathbb{E}$ on the right-hand side means that $\widetilde{\boldsymbol{\theta}}_{t,X_i}$ are already fixed for $i < k$. This part is similar to Appendix B.5 in (Li et al. 2020). In order to determine $\operatorname{argmax}_{\boldsymbol{\theta}'_{t,X_k}\in\mathcal{C}_{t,X_k}} \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_{t,X_i},i<k}[X_k|do(\boldsymbol{S})]$, we use the method of Lagrange multipliers to solve this optimization problem. The only constraint on $\boldsymbol{\theta}'_{t,X_k}$ is $\left\|\boldsymbol{\theta}'_{t,X_k} - \hat{\boldsymbol{\theta}}_{t-1,X_k}\right\|_{M_{t-1,X_k}} \le \rho_{t-1}$, therefore, the optimized $\boldsymbol{\theta}'_{t,X_k}$ we want should be a solution of

$$\frac{\partial \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_{t,X_i},i\le k-1,\boldsymbol{\theta}'_{i,X_k}}[X_k|do(\boldsymbol{S})]}{\partial \boldsymbol{\theta}'_{t,X_k}} - \lambda \frac{\partial\left(\left\|\boldsymbol{\theta}'_{t,X_k} - \hat{\boldsymbol{\theta}}_{t-1,X_k}\right\|^2_{M_{t-1,X_k}} - \rho^2_{t-1}\right)}{\partial \boldsymbol{\theta}'_{t,X_k}} = 0,$$

which indicates that

$$\left(\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_{t,X_i},i\le k-1}[X_{i_1}|do(\boldsymbol{S})], \cdots, \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_{t,X_i},i\le k-1}[X_{i_{|\boldsymbol{Pa}(X_k)|}}|do(\boldsymbol{S})]\right)^\intercal = 2\lambda M_{t-1,X_k}\left(\boldsymbol{\theta}'_{t,X_k} - \hat{\boldsymbol{\theta}}_{t-1,X_k}\right),$$

where $X_{i_1}, \cdots, X_{i_{|\boldsymbol{Pa}(X_k)|}}$ are parents of X_k. Therefore, we can deduce that

$$\boldsymbol{\theta}'_{t,X_k} = \frac{1}{2\lambda} M^{-1}_{t,X_k}\left(\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_{t,X_i},i\le k-1}\left[X_{i_1}|do(\boldsymbol{S})\right], \cdots, \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_{t,X_i},i\le k-1}[X_{i_{|\boldsymbol{Pa}(X_k)|}}|do(\boldsymbol{S})]\right)^\intercal + \hat{\boldsymbol{\theta}}_{t-1,X_k}.$$

Meanwhile, we know that when $\boldsymbol{\theta}'_{t,X_k}$ is optimized, it should be on the boundary of the confidence ellipsoid, so λ can be solved out as

$$\frac{1}{2\lambda} = \frac{\rho_{t-1}}{\sqrt{\left(\mathbb{E}_{\widetilde{\theta}_{t,X_i},i\le k-1}[\boldsymbol{Pa}(X_k)|do(\boldsymbol{S})]\right)^\intercal M^{-1}_{t-1,X_k}\mathbb{E}_{\widetilde{\theta}_{t,X_i},i\le k-1}[\boldsymbol{Pa}(X_k)|do(\boldsymbol{S})]}}.$$

Until now, we have shown that

$$\max_{\boldsymbol{\theta}'_{t,X_k}\in\mathcal{C}_{t,X_k}} \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_{t,X_i},i<k}[X_k|do(\boldsymbol{S})] = \frac{\rho_{t-1}M^{-1}_{t-1,X_k}\mathbb{E}_{\widetilde{\theta}_{t,X_i},i\le k-1}[\boldsymbol{Pa}(X_k)|do(\boldsymbol{S})]}{\sqrt{\left(\mathbb{E}_{\widetilde{\theta}_{t,X_i},i\le k-1}[\boldsymbol{Pa}(X_k)|do(\boldsymbol{S})]\right)^\intercal M^{-1}_{t-1,X_k}\mathbb{E}_{\widetilde{\theta}_{t,X_i},i\le k-1}[\boldsymbol{Pa}(X_k)|do(\boldsymbol{S})]}}$$

$$
\begin{aligned}
&+ \left(\mathbb{E}_{\widetilde{\theta}_{t,X_i}, i \le k-1}[\boldsymbol{Pa}(X_k)|do(\boldsymbol{S})]\right)^{\mathsf{T}} \hat{\boldsymbol{\theta}}_{t-1,X_k} \\
&= \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X_k|do(\boldsymbol{S})]. \qquad (34)
\end{aligned}
$$

Next, we prove Eq. (33) by induction on i. If $i = 2$, it is trivial that $\max_{\boldsymbol{\theta}'_{t,X_2} \in \mathcal{C}_{t,X_2}} \mathbb{E}[X_2|do(\boldsymbol{S})] = \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X_2|do(\boldsymbol{S})]$ according to Eq. (34).

Now we suppose that for any $i \le k-1$, we already know that $\max_{\boldsymbol{\theta}'_{t,X_j} \in \mathcal{C}_{t,X_j}, j \le i} \mathbb{E}[X_i|do(\boldsymbol{S})] = \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X_i|do(\boldsymbol{S})]$. We want to prove that $\max_{\boldsymbol{\theta}'_{t,X_j} \in \mathcal{C}_{t,X_j}, j \le k} \mathbb{E}[X_k|do(\boldsymbol{S})] = \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X_k|do(\boldsymbol{S})]$. Actually, we can deduce that

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}'_{t,X_j} \in \mathcal{C}_{t,X_j}, j \le k}[X_k|do(\boldsymbol{S})] &\le \max_{\boldsymbol{\theta}'_{t,X_j} \in \mathcal{C}_{t,X_j}, j \le k} \sum_{Z \in \boldsymbol{Pa}(X_k)} \mathbb{E}[Z|do(\boldsymbol{S})]\theta'_{t,Z,X_k} \\
&\le \max_{\boldsymbol{\theta}'_{t,X_k} \in \mathcal{C}_{t,X_k}} \sum_{Z \in \boldsymbol{Pa}(X_k)} \left(\max_{\boldsymbol{\theta}'_{t,X_j} \in \mathcal{C}_{t,X_j}, j \le k-1} \mathbb{E}[Z|do(\boldsymbol{S})] \right) \theta'_{t,Z,X_k} \\
&= \max_{\boldsymbol{\theta}'_{t,X_k} \in \mathcal{C}_{t,X_k}} \sum_{Z \in \boldsymbol{Pa}(X_k)} \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[Z|do(\boldsymbol{S})]\theta'_{t,Z,X_k} \\
&\le \sum_{Z \in \boldsymbol{Pa}(X_k)} \widetilde{\theta}_{t,Z,X_k} \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[Z|do(\boldsymbol{S})] \\
&= \mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X_k|do(\boldsymbol{S})],
\end{aligned}
$$

which is exactly what we want while $\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[X_k|do(\boldsymbol{S})]$ can be easily achieved because $\widetilde{\boldsymbol{\theta}}_{t,X_k} \in \mathcal{C}_{t,X_k}$ for any $X_k \in \boldsymbol{X} \cup \{Y\}$. Therefore, Eq. (33) holds and because $\boldsymbol{S}_t$ in Algorithm 4 is selected according to the largest $\mathbb{E}_{\widetilde{\boldsymbol{\theta}}_t}[Y|do(\boldsymbol{S})] = \max_{\boldsymbol{\theta}'_{t,X_k} \in \mathcal{C}_{t,X_k}, k=2,3,\cdots,n} \mathbb{E}[Y|do(\boldsymbol{S})]$, it is exactly the intervened set $\boldsymbol{S}_t$ we define in Algorithm 1. □

When conducting experiments on BLMs, Algorithm 4 works well in practice. It runs 30 times faster than the ϵ-net method proposed in (Li et al. 2020) when adopting $\epsilon = 0.01$.

H.2 Simulations on BLMs

In this section, we run simulations on some typical BLMs with our two algorithms based on maximum likelihood estimation, linear regression (BGLM-OFU and BLM-LR) and two baseline algorithms (UCB and ϵ-greedy). We call BGLM-OFU as BLM-OFU when it is adopted on BLMs. We show that our algorithms have much smaller best-in-hindsight regrets than the baseline algorithms (UCB and ϵ-greedy). This is consistent with our theoretical analysis that BLM-OFU and BLM-LR promise better regrets which are polynomial with respect to the size of graph in combinatorial settings. Also, we further show this by compare the performances of these algorithms when adopted on graphs with different sizes. When the graph becomes larger, the performance gaps between our algorithms and the baselines become larger.

Because our round number is limited, we adopt ρ_t, ρ to be $\frac{1}{10}$ times of our original parameters setting in BLM-OFU and BLM-LR. We use the implementation of pair-oracle introduced in the former section (Algorithm 4). About the initialization phase of BLM-OFU, we set $T_0 = \frac{1}{100}T$ for convenience (second order derivative of a linear function is 0, so $L^{(2)}_{f_X}$ in BGLM-OFU can be arbitrarily small). For the UCB algorithm, we adopt the common used upper confidence bound $\sqrt{\frac{\ln t}{n_{i,t}}}$ where t is the number of current round and $n_{i,t}$ is the number of playing times of arm i until the t^{th} round. For fairness, we also simulate on UCB with the upper confidence bound 10 times smaller than the standard one. The result of this heuristic algorithm is labeled by "UCB (scaled)". For ϵ-greedy algorithm, we adopt $\epsilon = 0.1$ and $\epsilon = 0.01$. We have tried other settings for these two baselines, and our choices are close to optimal for all tested BLMs. For both of the two baselines, we treat each possible K node intervention set as an arm so there are $\binom{n-2}{K}$ arms in total (one can intervene $X_2, \cdots, X_{n-1}$). For each experiment, we run average regrets of 30 repeated simulations in the same settings. Then we draw the 95% confidence intervals (Fisher 1992) of these average regrets by repeating the 30 repeats for 20 times. In total, we simulate 600 times for each setting of each experiment. All of our experiments are run on multi-threadedly on 4 performance-cores of Intel Core™ i7-12700H Processor at 4.30GHz with 32GB DDR5 SDRAM. Code is available in our supplementary material.

Simulations on Parallel Graphs In the first experiment, we set round number $T = 10000$, $K = 3$ and $n = 8$. The simulated model G_1 is a parallel BLM such that X_1 points to $X_2, X_3, \cdots, X_7$ and $X_2, X_3, \cdots, X_7$ all point to Y. X_1 is always 1, all the other paramters are

$$
\begin{aligned}
&\theta^*_{X_1,X_2} = 0.3, \theta^*_{X_1,X_3} = 0.4, \theta^*_{X_1,X_4} = 0.2, \theta^*_{X_1,X_5} = 0.1, \theta^*_{X_1,X_6} = 0.6, \theta^*_{X_1,X_7} = 0.5, \\
&\theta^*_{X_2,Y} = 0.1, \theta^*_{X_3,Y} = 0.3, \theta^*_{X_4,Y} = 0.2, \theta^*_{X_5,Y} = 0.2, \theta^*_{X_6,Y} = 0.1, \theta^*_{X_7,Y} = 0.1.
\end{aligned}
$$

The best intervention for G_1 is $\{do(X_3 = 1), do(X_4 = 1), do(X_5 = 1)\}$. One can find the graph structure and parameters on G_1 in Figure 2. The total running time of this experiment is 3714 seconds.

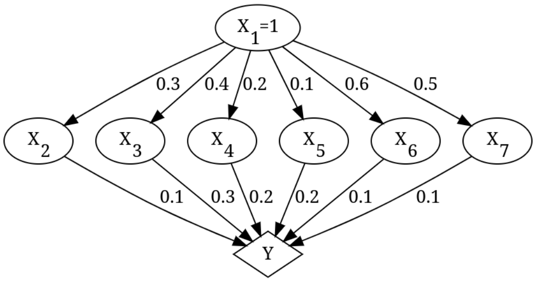

Figure 2: Structure and Parameters of G_1.

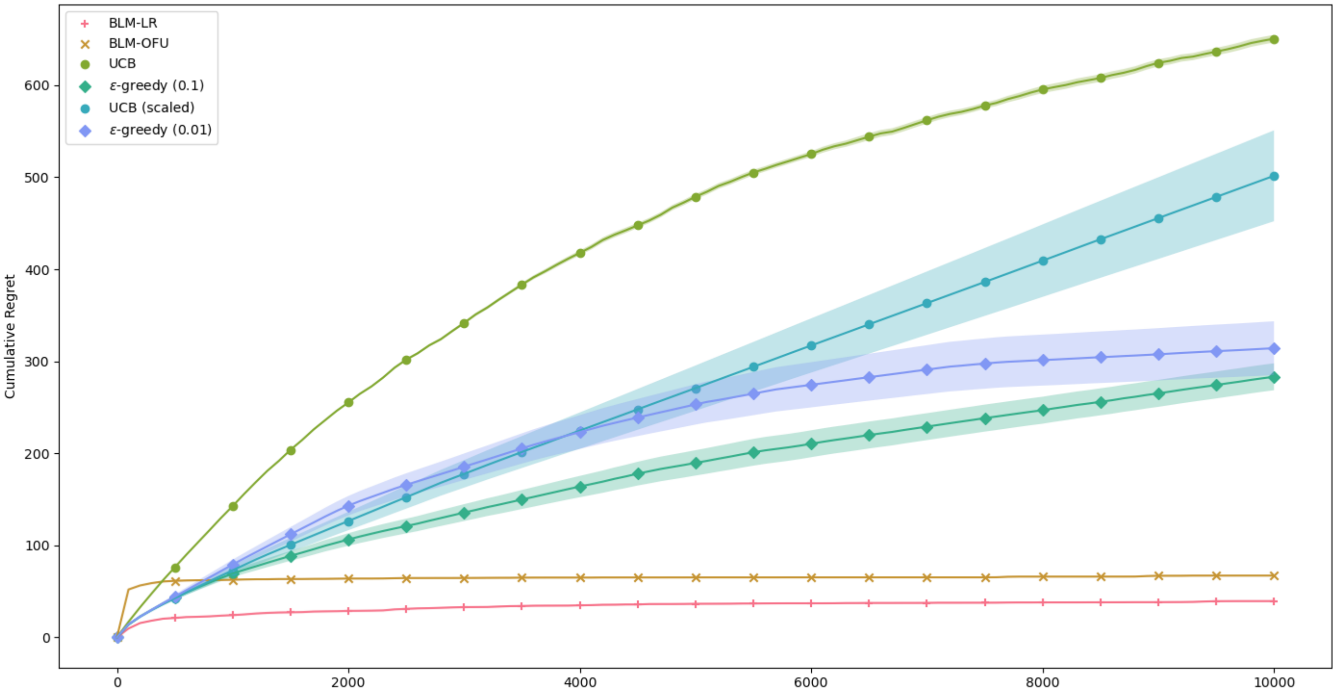

Figure 3: Regrets of Algorithms Run on G_1.

Figure 3 shows that regrets of BLM-LR and BLM-OFU are much smaller than all settings of UCB and ϵ-greedy algorithms. Moreover, regrets of BLM-LR and BLM-OFU remain nearly unchanged after some rounds. That may be because parameters are already estimated well so deviations from the best intervention are not likely to occur. The large regrets of UCB and ϵ-greedy are due to the large amount of arms (20 arms) and the hardness of estimating expected payoffs.

In the second experiment, we set round number $T = 2000$ and $K = 2$. We run this experiment on three parallel graphs G_2, G_3 and G_4. G_2 has 10 nodes in total. X_1 points to $X_2, X_3, \cdots, X_9$ and $X_2, X_3, \cdots, X_9$ point to Y. The parameters on G_2 are

$$\theta^*_{X_1,X_2} = 0.2, \theta^*_{X_1,X_3} = 0.2, \theta^*_{X_1,X_4} = 0.6, \theta^*_{X_1,X_5} = 0.6, \theta^*_{X_1,X_6} = 0.6, \theta^*_{X_1,X_7} = 0.6, \theta^*_{X_1,X_8} = 0.6, \theta^*_{X_1,X_9} = 0.6,$$
$$\theta^*_{X_2,Y} = 0.2, \theta^*_{X_3,Y} = 0.2, \theta^*_{X_4,Y} = 0.1, \theta^*_{X_5,Y} = 0.1, \theta^*_{X_6,Y} = 0.1, \theta^*_{X_7,Y} = 0.1, \theta^*_{X_8,Y} = 0.1, \theta^*_{X_9,Y} = 0.1.$$

The best intervention for G_2 is $\{do(X_2 = 1), do(X_3 = 1)\}$. One can find the graph structure and parameters on G_2 in Figure 4.

G_3 has 8 nodes, which is exactly G_2 without two nodes X_8 and X_9. G_4 has 6 nodes, which is exactly G_2 without four nodes X_6, X_7, X_8 and X_9. The best intervention for G_3 and G_4 is also $\{do(X_2 = 1), do(X_3 = 1)\}$. The total running time of this experiment is 603 seconds.

Figures 5, 6 and 7 also show that regrets of BLM-LR and BLM-OFU are much smaller than all settings of UCB and ϵ-greedy algorithms. Additionally, regrets of BLM-LR and BLM-OFU remain below 20 on each of the three BLMs G_2, G_3 and G_4,

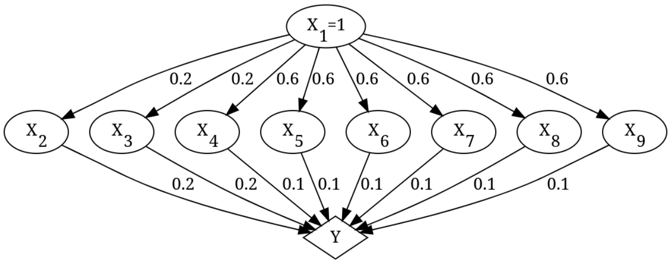

Figure 4: Structure and Parameters of G_2.

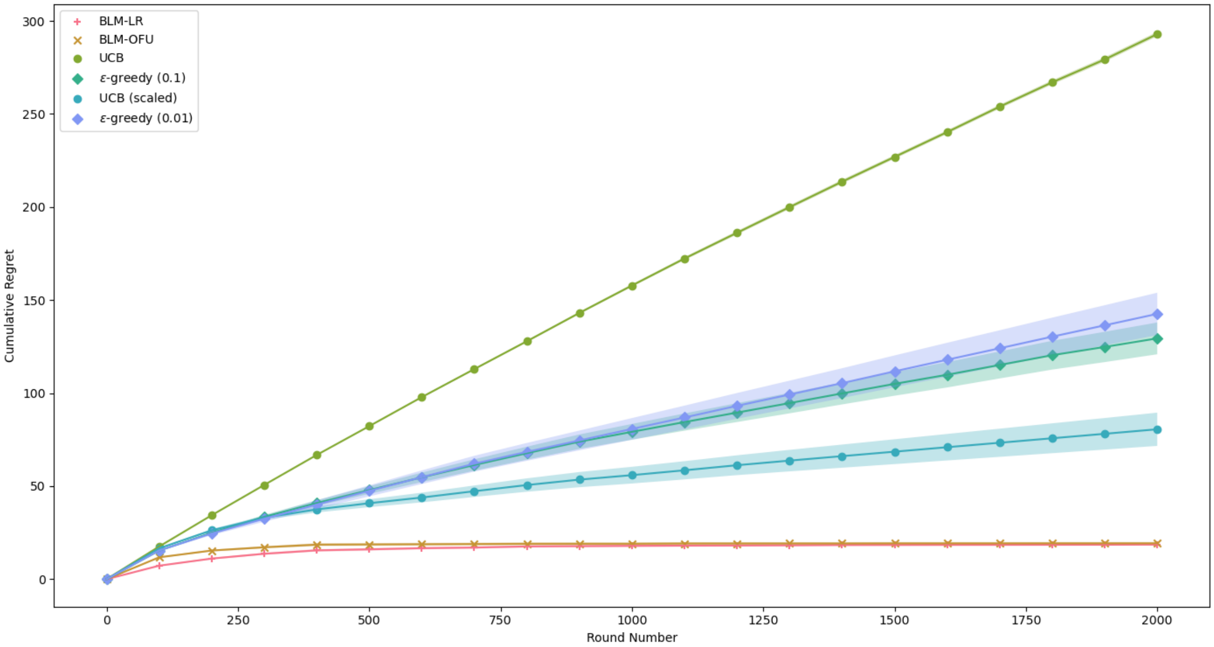

Figure 5: Regrets of Algorithms Run on G_2.

which are not sensitive to the graph size. However, the regrets of UCB and ϵ-greedy algorithms increase in proportion to the number of arms in each BLM (G_2, G_3 and G_4 have $28, 15$ and 6 arms respectively). This comparison shows that BLM-LR and BLM-OFU are able to overcome the exponentially large space of arms in combinatorial settings for causal bandits problem.

Simulations on Two-Layer Graph Besides parallel BLMs, we also conduct an experiment on a two-layer BLM. We set round number $T = 10000$, $K = 2$ and $n = 7$. G_5 is a two-layer graph such that X_1 points to $X_2, X_3, \cdots, X_6$ and both X_2, X_3 point to all of X_4, X_5, X_6 and X_4, X_5, X_6 all point to Y. The parameters on G_5 are

$$\theta^*_{X_1,X_2} = 0.1, \theta^*_{X_1,X_3} = 0.1, \theta^*_{X_1,X_4} = 0.1, \theta^*_{X_1,X_5} = 0.1, \theta^*_{X_1,X_6} = 0.1,$$
$$\theta^*_{X_2,X_4} = 0.1, \theta^*_{X_2,X_5} = 0.7, \theta^*_{X_2,X_6} = 0.7, \theta^*_{X_3,X_4} = 0.2, \theta^*_{X_3,X_5} = 0.1, \theta^*_{X_3,X_6} = 0.1,$$
$$\theta^*_{X_4,Y} = 0.6, \theta^*_{X_5,Y} = 0.1, \theta^*_{X_6,Y} = 0.1.$$

The best intervention for G_5 is $\{do(X_2 = 1), do(X_4 = 1)\}$. One can find the graph structure and parameters on G_5 in Figure 8. The total running time of this experiment is 355 seconds.

Figure 9 shows that regrets of BLM-LR and BLM-OFU are much smaller than all settings of UCB and ϵ-greedy algorithms not only on parallel graphs but also on more general graphs. Furthermore, because the difference between expected payoffs of the best intervention and other interventions is larger than that in G_1, BLM-LR estimates the parameters accurately enough even faster. Hence, BLM-LR achieves even much smaller regret than BLM-OFU.

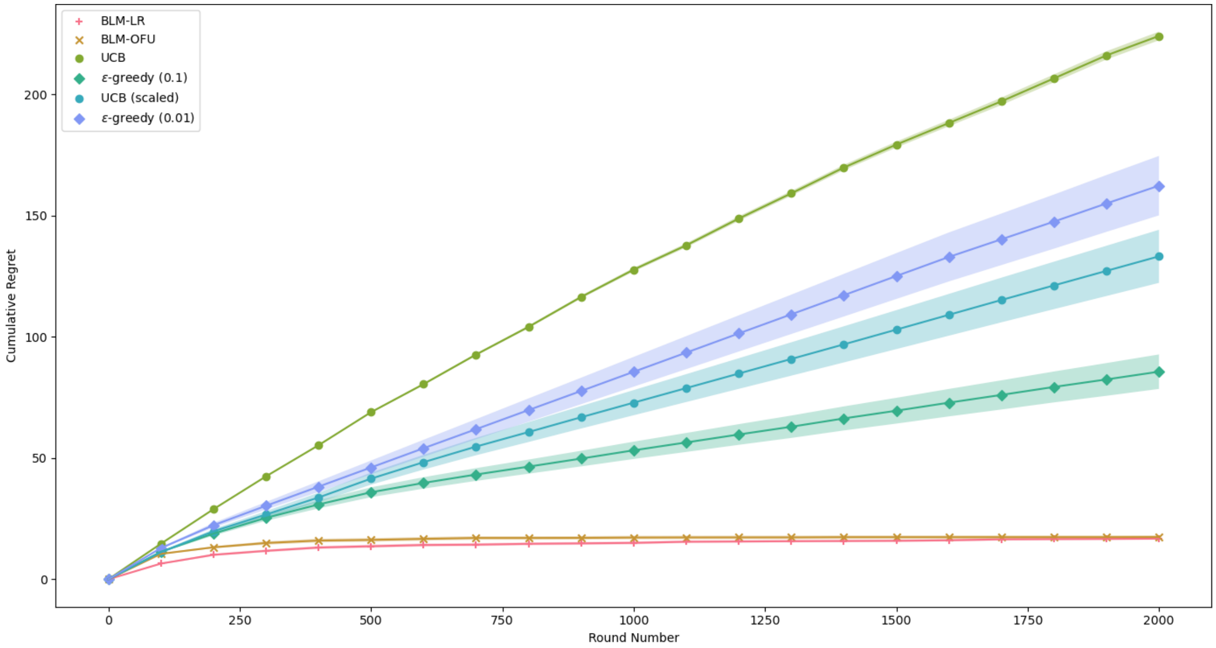

Figure 6: Regrets of Algorithms Run on G_3.

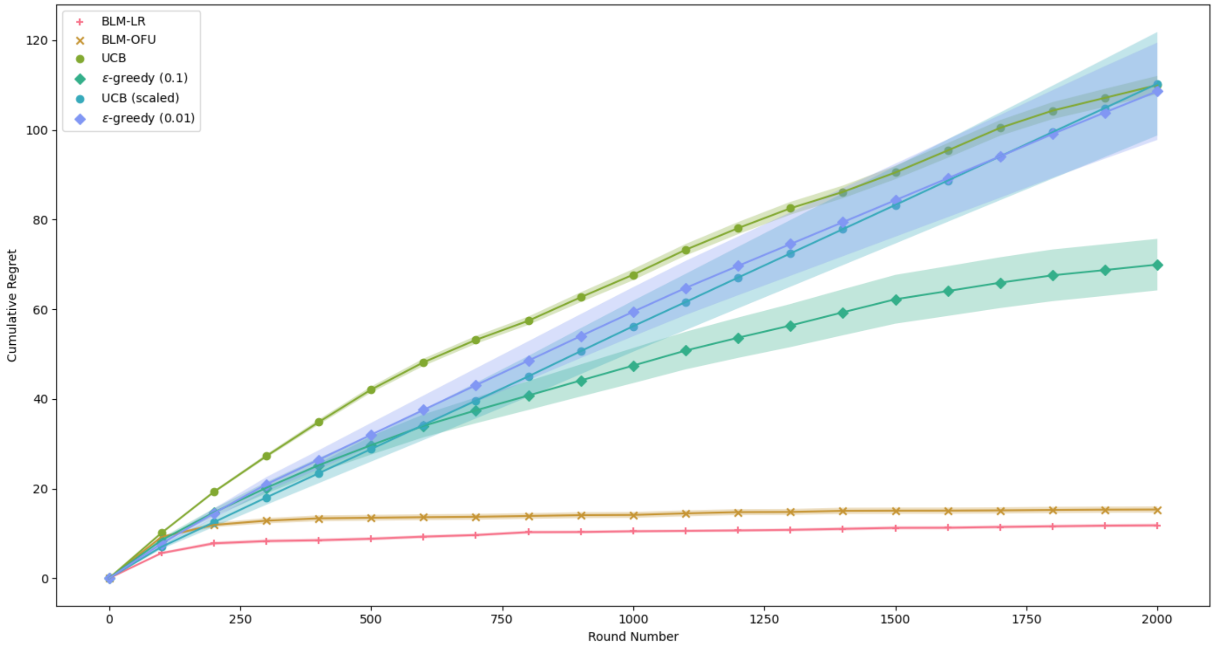

Figure 7: Regrets of Algorithms Run on G_4.

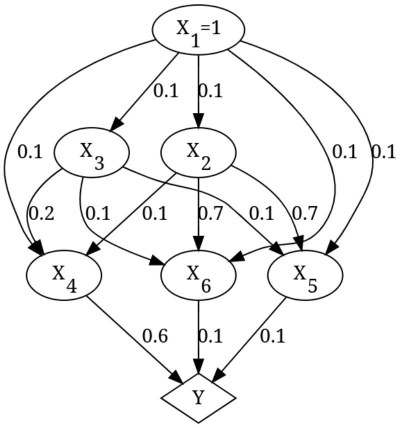

Figure 8: Structure and Parameters of G_5.

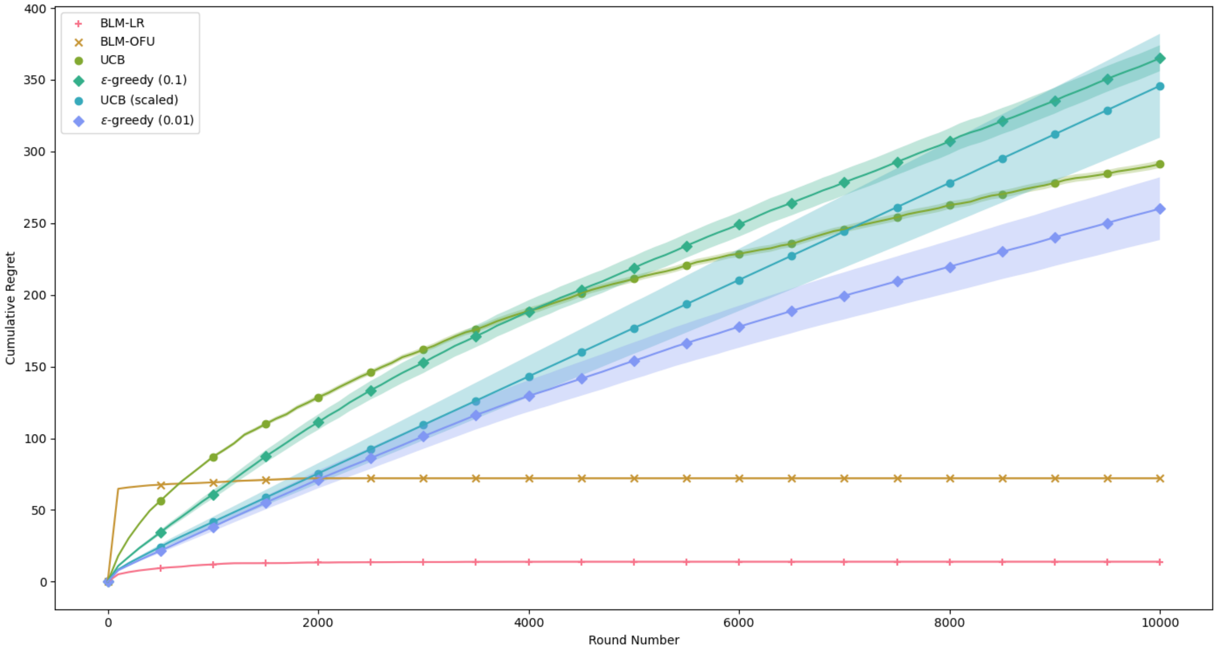

Figure 9: Regrets of Algorithms Run on G_5.