Combinatorial Causal Bandits

AUTHORS
Naser Kazemi, naserkazemi@sharif.edu
Mahyar Afshinmehr, mahyar.afshinmehr@sharif.edu

AFFILIATIONS
This poster is designed for the course "Reinforcement Learning", lectured by Prof. Rohban, at Sharif University of Technology. The work represented here is based on the researches of Shi Feng, Nuoya Xiong, and Wei Chen.

BACKGROUND

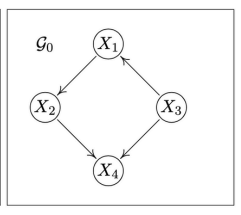
Definition 1: A *structural equation model* (SEM) is a tuple $\mathcal{S} := (\mathcal{S}, \mathbb{P}^N)$, where $\mathcal{S} = (\mathcal{S}_1, \dots, \mathcal{S}_p)$ consists of p equations:

$$\mathcal{S}_j : \quad X_j = f_j(\mathbf{Pa}(X_j), N_j), \quad j = 1, \dots, p,$$

with $\mathbf{Pa}(X_j) \subseteq \{X_1, \dots, X_p\} \setminus \{X_j\}$ as the *parents* of X_j, and $\mathbb{P}^N = \mathbb{P}_{N_1,\dots,N_p}$ as the joint distribution of independent noise variables. The graph of an SEM is obtained by drawing edges from each parent to its effects, and it is assumed to be acyclic.

$$\begin{aligned} X_1 &= f_1(X_3, N_1) \\ X_2 &= f_2(X_1, N_2) \\ X_3 &= f_3(N_3) \\ X_4 &= f_4(X_2, X_3, N_4) \end{aligned}$$

- N_i jointly independent
- $\mathcal{G}_0$ has no cycles



Definition 2 [Intervention Distribution]: Given a distribution $\mathbb{P}^X$ from an SEM $\mathcal{S} := (\mathcal{S}, \mathbb{P}^N)$, replacing structural equations without cycles creates a new SEM $\widetilde{\mathcal{S}}$. These *intervention distributions* are denoted by:

$$\mathbb{P}^X_{\widetilde{\mathcal{S}}} = \mathbb{P}_{\mathcal{S}}^{X|do(X_j = \widetilde{f}(\widetilde{\mathrm{Pa}}(X_j), \widetilde{N}_j))}.$$

The noise variables in $\widetilde{\mathcal{S}}$ include both new $\widetilde{N}$'s and old N's, which are mutually independent. If $\widetilde{f}(\widetilde{\mathrm{PA}}_j, \widetilde{N}_j)$ places a point mass on a value a, we write $\mathbb{P}^X_{\mathcal{S}} \mid do(X_j = a)$ and call this a *perfect intervention*. An intervention with $\widetilde{\mathrm{PA}}_j = \mathrm{PA}_j$ is an *imperfect intervention*, a special case of a *stochastic intervention* with positive variance in the marginal distribution.

INTRODUCTION

Causal Models

A *causal graph* $G = (X \cup Y, E)$ is a directed acyclic graph with intervenable variables X, target node Y, and edges E. Variables in $X \cup \{Y\}$ are binary and observable, but edges in E are unknown. *Parents* of X are $\mathbf{Pa}(X)$, with values $\mathbf{pa}(X)$. We focus on *Markovian causal graphs* with no hidden variables. X_1 is a constant parent for all nodes to model self-activation.

Binary Generalized Linear Models (BGLMs):

$$P(X = 1 | \mathbf{Pa}(X) = \mathbf{pa}(X)) = f_X(\theta^*_X \cdot \mathbf{pa}(X)) + \epsilon_X$$

Binary Linear Models (BLMs):

$$P(X = 1 | \mathbf{Pa}(X) = \mathbf{pa}(X)) = \theta^*_X \cdot \mathbf{pa}(X) + \epsilon_X$$

Notations:

- f_X: A monotone increasing function.
- θ^*_X: A weight vector.
- $\epsilon_X \le 1 - \max_{\mathbf{pa}(X) \in \{0,1\}^{|\mathrm{Pa}(X)|}} f_X(\mathbf{pa}(X) \cdot \theta^*_X)$: A sub-Gaussian noise.
- $\theta^*_{X',X}$: The entry in the vector θ^*_X that corresponds to node $X' \in \mathbf{Pa}(X)$.
- θ^*: The vector of all the weights, and Θ to denote the feasible domain for the weights.
- ϵ: Represents all noise random variables $(\epsilon_X)_{X \in X \cup Y}$.
- $\theta^*_{\min}$: Indicates difficulty in discovering edges. Large gaps ensure accurate discovery; small gaps may lead to inaccuracies.

$$\theta^*_{\min} = \min_{(X',X) \in E} \theta^*_{X',X}$$

Combinatorial Causal Bandits (CCB)

CCB involves interventions on all variables except X_1 and Y, using *do* operators. The action set $\mathcal{A} = \{do(S = s) | S \subseteq X \setminus \{X_1\}, s \in \{0,1\}^S\}$. The expected reward Y under intervention on $S \subseteq X \setminus \{X_1\}$ is $\mathbb{E}[Y | do(S = s)]$. A learning agent runs algorithm π for T rounds to output intervention set S_t. The goal is to minimize cumulative regret:

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

Where $do(S^* = s^*) \in \arg\max_{do(S=s) \in \mathcal{A}} \mathbb{E}[Y \mid do(S)]$. We assume $\mathrm{dx}(X) = 1$, indicating interventions are needed to identify the causal structure.

LOWER BOUND ON GENERAL BINARY CAUSAL MODEL

Theorem 1 (Binary Model Lower Bound). Recall that $n = |X|$. For any algorithm, when $T \ge \frac{16(2^n - 1)}{3}$, there exists a precise bandit instance of general binary causal model $\mathcal{T}$ such that

$$\mathbb{E}_{\mathcal{T}}[R(T)] \ge \frac{\sqrt{2^n T}}{8e}.$$

Moreover, when $T \le \frac{16(2^n-1)}{3}$, there exists a precise bandit instance of general binary causal model $\mathcal{T}$ that

$$\mathbb{E}_{\mathcal{T}}[R(T)] \ge \frac{T}{16e}.$$

Assumption 1. For every $X \in \mathbf{X} \cup \{Y\}$, 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 $\kappa = \inf_{X \in \mathbf{X} \cup \{Y\}, \mathbf{v} \in [0,1]^{|\mathrm{Pa}(X)|}, \|\theta - \theta^*_X\| \le 1} \dot{f}_X(\mathbf{v} \cdot \theta)$.

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

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

$$\Pr_{\epsilon, X, Y}(X' = 1 \mid \mathrm{Anc}(X) \setminus \{X', X_1\} = \mathbf{v}) \ge \zeta$$

$$\Pr_{\epsilon, X, Y}(X' = 0 \mid \mathrm{Anc}(X) \setminus \{X', X_1\} = \mathbf{v}) \ge \zeta$$

BGLM CCB GRAPH SKELETON

Algorithm 1 BGLM-OFU for BGLM CCB Problem

1: **Input:** Graph $G = (X \cup \{Y\}, E)$, intervention budget $K \in \mathbb{N}$, parameter $L^{(1)}_f, L^{(2)}_f, \kappa, \zeta$ in Assumption 1, 2 and 3.
2: Initialize $M_{0,X} \leftarrow 0 \in \mathbb{R}^{|\mathrm{Pa}(X)| \times |\mathrm{Pa}(X)|}$ for all $X \in X \cup \{Y\}$, $\delta \leftarrow \frac{1}{3n\sqrt{T}}$, $R \leftarrow \left\lceil \frac{512n(L^{(2)}_f)^2}{\kappa^4}(n^2 + \ln\frac{1}{\delta}) \right\rceil$, $T_0 \leftarrow \max\left\{\frac{c}{\zeta^2}\ln\frac{1}{\delta}, \frac{(8n^2-16n+2)R}{\zeta}\right\}$ 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 $(X_t, Y_t), 1 \le t \le T_0$.
5: **/* Iterative Phase: */**
6: **for** $t = T_0 + 1, T_0 + 2, \cdots, T$ **do**
7: $\quad\{\hat{\theta}_{t-1,X}, M_{t-1,X}\}_{X \in X \cup \{Y\}} \leftarrow$ BGLM-Estimate$((X_1, Y_1), \cdots, (X_{t-1}, Y_{t-1}))$ (see Algorithm 2).
8: $\quad$Compute the confidence ellipsoid $C_{t,X} = \{\theta'_X \in [0,1]^{|\mathbf{Pa}(X)|} : \|\theta'_X - \hat{\theta}_{t-1,X}\|_{M_{t-1,X}} \le \rho\}$ for any node $X \in X \cup \{Y\}$.
9: $\quad(S_t, \hat{\theta}_t) = \arg\max_{S \subseteq X, |S| \le K, \theta_t, x \in C_{t,x}} \mathbb{E}[Y | do(S)]$.
10: $\quad$Intervene all the nodes in S_t to 1 and observe the feedback (X_t, Y_t).
11: **end for**

Algorithm 2 BGLM-Estimate

1: **Input:** All observations $((X_1, Y_1), \cdots, (X_t, Y_t))$ until round t.
2: **Output:** $\{\hat{\theta}_{t,X}, M_{t,X}\}_{X \in X \cup \{Y\}}$
3: For each $X \in X \cup \{Y\}, i \in [t]$, construct data pair $(V_{i,X}, X^i)$ with $V_{i,X}$ 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** $X \in X \cup \{Y\}$ **do**
5: $\quad$Calculate the maximum-likelihood estimator $\hat{\theta}_{t,X}$ by solving the equation $\sum_{i=1}^{t}(X^i - f_X(V_{i,X}^T \theta_X))V_{i,X} = 0$.
6: $\quad M_{t,X} = \sum_{i=1}^{t} V_{i,X} V_{i,X}^T$.
7: **end for**

BGLM AND BGLM-UNKNOWN REGRET BOUND THEOREMS

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

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

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

Lemma 1: Let G be a BGLM with parameter θ^* that satisfies Assumption 2. Recall that $\theta^*_{\min} = \min_{(X',X) \in E} \theta^*_{X',X}$. If $X_i \in \mathrm{Pa}(X_j)$, we have $\mathbb{E}[X_j | do(X_i = 1)] - \mathbb{E}[X_j | do(X_i = 0)] \ge \kappa\theta^*_{X_i,X_j} \ge \kappa\theta^*_{\min}$; if X_i is not an ancestor of X_j, we have $\mathbb{E}[X_j | do(X_i = 1)] = \mathbb{E}[X_j | do(X_i = 0)]$.

Theorem 3 (Regret Bound of BGLM-OFU-Unknown): Let $L^{(1)}_{\max} = \max_{X \in X \cup \{Y\}} L^{(1)}_{f_X}$. Under Assumptions 1, 2, and 3, the regret of BGLM-OFU-Unknown (Algorithms 2, 3, and 4) is bounded by

$$R(T) = O\left(\frac{1}{\kappa} n^{3/2} L^{(1)}_{\max} \sqrt{T} \log T\right),$$

where the terms of $o(\sqrt{T}\ln T)$ are omitted. The big O notation holds for $T \ge 32\left(\frac{c_1}{\kappa\theta^*_{\min}}\right)^5$.

BGLM CCB WITHOUT GRAPH SKELETON BUT WITH MINIMUM WEIGHT GAP

Algorithm 3 BGLM-OFU-Unknown for BGLM CCB Problem

1: **Input:** Graph $G = (X \cup \{Y\}, E)$, action set A, parameters $L^{(1)}_f, L^{(2)}_f, \kappa, \zeta$ in Assumption 1, 2 and 3, c in Lecué and Mendelson's inequality, positive constants c_0 and c_1 for initialization phase such that $c_0\sqrt{T} \in \mathbb{N}^+$.
2: **/* Initialization Phase: */**
3: Initialize $T_0 \leftarrow 2(n-1)c_0 T^{1/2}$.
4: Do each intervention among $do(X_2 = 1), do(X_2 = 0), \dots, do(X_n = 1), do(X_n = 0)$ for $c_0 T^{1/2}$ times in order and observe the feedback $(X_t, Y_t), 1 \le t \le T_0$.
5: Compute the ancestors $\widehat{\mathbf{Anc}}(X)$, $X \in X \cup \{Y\}$ by BGLM-Ancestors$((X_1, Y_1), \dots, (X_{T_0}, Y_{T_0}), c_0, c_1)$ (see Algorithm 4).
6: **/* Parameters Initialization: */**
7: Exactly the same as Algorithm 1, with $\widehat{\mathbf{Anc}}(X)$ instead of $\mathbf{Pa}(X)$ and $T_1 \leftarrow T_0 + \max\left\{\frac{c}{\zeta^2}\ln\frac{1}{\delta}, \frac{(8n^2-6)R}{\zeta}\right\}$
8: **/* Iterative Phase: */**
9: Do the same as Algorithm 1, with $\widehat{\mathbf{Anc}}(X)$ instead of $\mathbf{Pa}(X)$. (Iteration starts from T_1)

Algorithm 4 BGLM-Ancestors

1: **Input:** Observations $((X_1, Y_1), \dots, (X_{T_0}, Y_{T_0}))$, positive constants c_0 and c_1.
2: **Output:** $\widehat{\mathbf{Anc}}(X)$, ancestors of X, $X \in X \cup \{Y\}$.
3: For all $X \in X$, $\widehat{\mathbf{Anc}}(X) = \emptyset$, $\widehat{\mathbf{Anc}}(Y) = X$.
4: **for** $i \in \{2, 3, \dots, n\}$ **do**
5: $\quad$**for** $j \in \{2, 3, \dots, n\} \setminus i$ **do**
6: $\quad\quad$**if** $\frac{1}{c_0\sqrt{T}} \sum_{k=1}^{c_0\sqrt{T}} \left(X_j^{(2ic_0\sqrt{T}+k)} - X_j^{((2i+1)c_0\sqrt{T}+k)}\right) > c_0 c_1 T^{3/10}$ **then**
7: $\quad\quad\quad$Add X_i into $\widehat{\mathbf{Anc}}(X_j)$.
8: $\quad\quad$**end if**
9: $\quad$**end for**
10: **end for**
11: Recompute the transitive closure of $\widehat{\mathbf{Anc}}(\cdot)$, i.e., if $X_i \in \widehat{\mathbf{Anc}}(X_j)$ and $X_j \in \widehat{\mathbf{Anc}}(X_\ell)$, then add X_i to $\widehat{\mathbf{Anc}}(X_\ell)$.

ALGORITHM FOR BLM BASED ON LINEAR REGRESSION

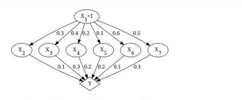


Figure 2: Structure and Parameters of G_1.

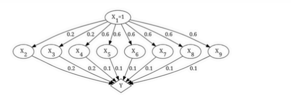


Figure 3: Structure and Parameters of G_2.

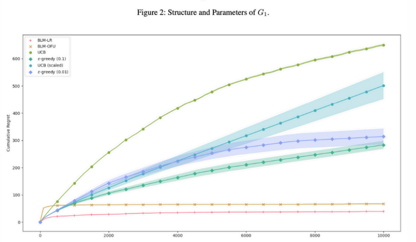


Figure 3: Regrets of Algorithms Run on G_1.

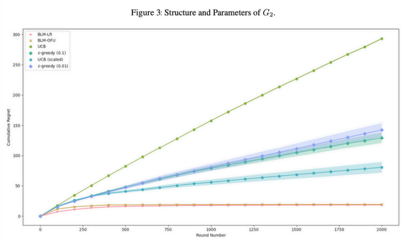


Figure 4: Regrets of Algorithms Run on G_2.

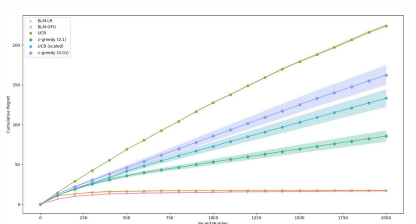


Figure 6: Regrets of Algorithms Run on G_3.

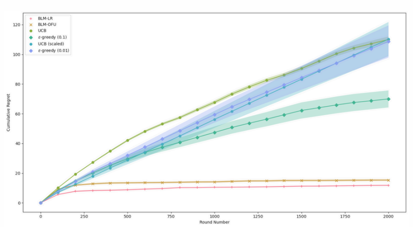


Figure 7: Regrets of Algorithms Run on G_4.

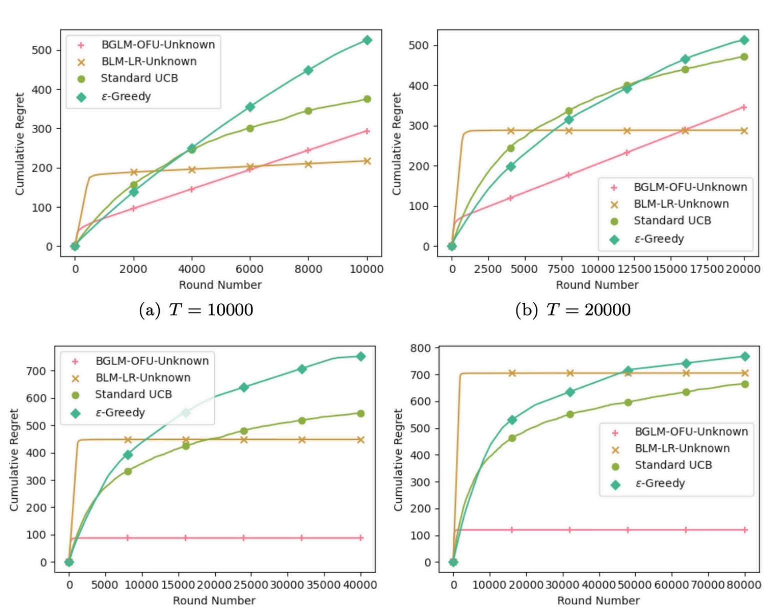
Algorithm 5 BLM-LR for BLM CCB Problem

1: **Input:** Graph $G = (X \cup \{Y\}, E)$, intervention budget $K \in \mathbb{N}$.
2: **/* Initialization Phase: */**
3: Initialize $M_{0,X} \leftarrow I \in \mathbb{R}^{|\mathrm{Pa}(X)| \times |\mathrm{Pa}(X)|}$, $b_{0,X} \leftarrow 0 \in \mathbb{R}^{|\mathrm{Pa}(X)|}$ for all $X \in X \cup \{Y\}$, $\hat{\theta}_{0,X} \leftarrow 0 \in \mathbb{R}^{|\mathrm{Pa}(X)|}$ for all $X \in X \cup \{Y\}$, $\delta \leftarrow \frac{1}{n\sqrt{T}}$ and $\rho_t \leftarrow \sqrt{n\log(1 + tn) + 2\log\frac{1}{\delta}} + \sqrt{n}$ for $t = 0, 1, 2, \cdots, T$.
4: **/* Iterative Phase: */**
5: **for** $t = 1, 2, \cdots, T$ **do**
6: $\quad$Compute the confidence ellipsoid $C_{t,X} = \{\theta'_X \in [0,1]^{|\mathrm{Pa}(X)|} : \|\theta'_X - \hat{\theta}_{t-1,X}\|_{M_{t-1,X}} \le \rho_{t-1}\}$ for any node $X \in X \cup \{Y\}$.
7: $\quad(S_t, \hat{\theta}_t) = \arg\max_{S \subseteq X, |S| \le K, \theta'_t, x \in C_{t,x}} \mathbb{E}[Y | do(S)]$.
8: $\quad$Intervene all the nodes in S_t to 1 and observe the feedback (X_t, Y_t).
9: $\quad$**for** $X \in X \cup \{Y\}$ **do**
10: $\quad\quad$Construct data pair $(V_{t,X}, X^t)$ with $V_{t,X}$ the parent value vector of X in round t, and X^t the value of X in round t if $X \notin S_t$.
11: $\quad\quad M_{t,X} = M_{t-1,X} + V_{t,X}V_{t,X}^T$, $b_{t,X} = b_{t-1,X} + X^t V_{t,X}$, $\hat{\theta}_{t,X} = M_{t,X}^{-1} b_{t,X}$.
12: $\quad$**end for**
13: **end for**

Theorem 4 (Regret Bound of Algorithm 5): The regret of BLM-LR (Algorithm 5) running on BLM with hidden variables is bounded as

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

BLM CCB WITHOUT GRAPH SKELETON AND WEIGHT GAP ASSUMPTION



(a) $T = 10000$ (b) $T = 20000$ (c) $T = 40000$ (d) $T = 80000$

Figure 7: Average Regrets for All Four Evaluated Algorithms on Our BLM

Algorithm 6 BLM-LR-Unknown for BLM CCB Problem without Weight Gap

1: **Input:** Graph $G = (X \cup \{Y\}, E)$, action set A, positive constants c_0 and c_1 for initialization phase.
2: **/* Initialization Phase: */**
3: Initialize $T_0 \leftarrow 2(n-1)c_0 T^{2/3}\log(T)$.
4: Do each intervention among $do(X_2 = 1), do(X_2 = 0), \dots, do(X_n = 1), do(X_n = 0)$ for $c_0 T^{2/3}$ times in order and observe the feedback (X_t, Y_t) for $1 \le t \le T_0$.
5: Compute the ancestors $\mathrm{Anc}(X)$, $X \in X \cup \{Y\}$ by Nogap-BLM-Ancestors$((X_1, Y_1), \dots, (X_{T_0}, Y_{T_0}), c_0, c_1)$ (see Algorithm **??**).
6: **/* Parameters Initialization: */**
7: Exactly the same as Algorithm 5, with $\widehat{\mathbf{Anc}}(X)$ instead of $\mathbf{Pa}(X)$.
8: **/* Iterative Phase: */**
9: Do the same as Algorithm 5, with $\widehat{\mathbf{Anc}}(X)$ instead of $\mathbf{Pa}(X)$. (Iteration starts from T_0)

Algorithm 7 Nogap-BLM-Ancestors

1: **Input:** Observations $((X_1, Y_1), \dots, (X_{T_0}, Y_{T_0}))$, positive constants c_0 and c_1.
2: **Output:** For all $X \in X \cup \{Y\}$, $\widehat{\mathbf{Anc}}(X)$.
3: For all $X \in X$, $\widehat{\mathbf{Anc}}(X) = \emptyset$, $\widehat{\mathbf{Anc}}(Y) = X$.
4: **for** $i \in \{2, 3, \dots, n\}$ **do**
5: $\quad$**for** $j \in \{2, 3, \dots, n\} \backslash \{i\}$ **do**
6: $\quad\quad$**if** $\sum_{k=1}^{c_0 T^{2/3}} \left(X_j\left(c_0(2i)T^{2/3} + k\right) - X_j\left(c_0(2i+1)T^{2/3} + k\right)\right) > c_0 c_1 T^{1/3}\log(T^2)$ **then**
7: $\quad\quad\quad$Add X_i into $\widehat{\mathbf{Anc}}(X_j)$.
8: $\quad\quad$**end if**
9: $\quad$**end for**
10: **end for**
11: Recompute the transitive closure of $\widehat{\mathbf{Anc}}(\cdot)$.

Theorem 5 If $c_0 \ge \max\left\{\frac{1}{c_1^2}, \frac{1}{(1-c_1)^2}\right\}$, the regret of Algorithm 6 running on BLM is upper bounded as

$$R(T) = O((n^3 T^{2/3})\log T).$$

REFERENCES

[1] Feng, Sh., & Chen, W. (2022). Combinatorial Causal Bandits.

[2] Feng, Sh., Chen, W., & Xiong, N. (2023). Combinatorial Causal Bandits without Graph Skeleton.

[3] Peters, J. (2015). *Causality Lecture Notes.* ETH Zurich.