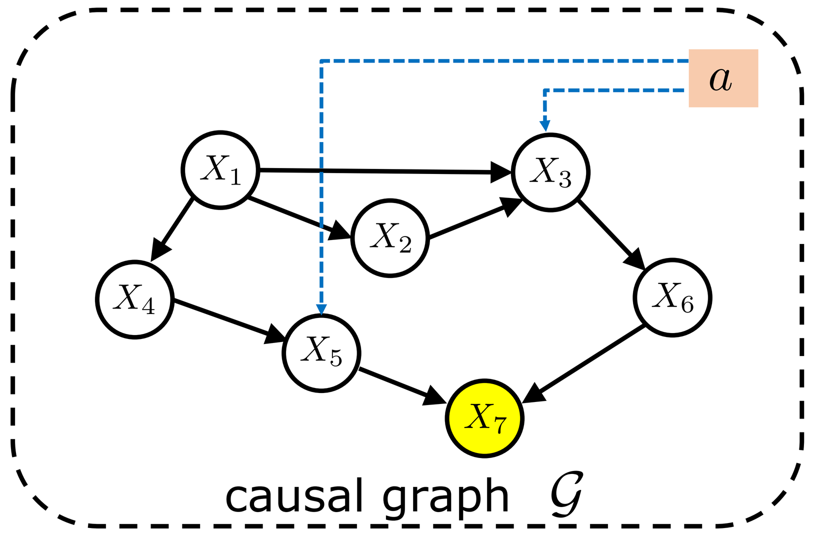
Causal Bandits for Linear Structural Equation Models

Burak Varıcı[1] Karthikeyan Shanmugam[2] Prasanna Sattigeri[3] Ali Tajer[4]

[1]Carnegie Mellon University [2]Google DeepMind India [3]IBM Research [4]Rensselaer Polytechnic Institute

Journal-to-conference track

Causal Bandits



$\mathcal{G}$:directed acyclic graph

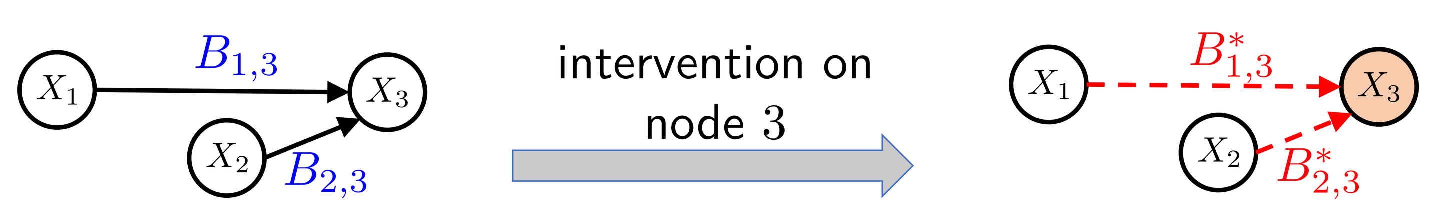
d :max. in-degree

L :max. causal path length

$X = (X_1, \dots, X_N)^\top$

Linear SEM: $X = \mathbf{B}^\top X + \epsilon$

- $\mathcal{A}$: set of possible interventions
- $\mathbf{B}$: unknown weight matrix, $\epsilon = (\epsilon_1, \dots, \epsilon_N)$: bounded noise
- **Soft intervention** on node i: changes cond. dist. of X_i

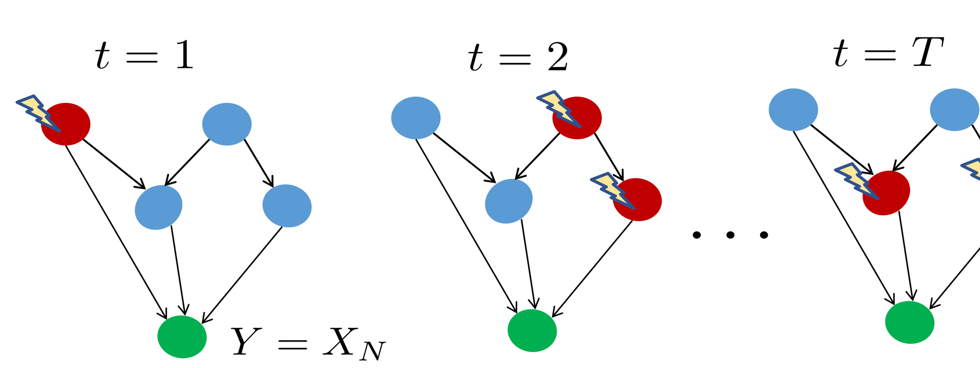


- Discrete setting: changes the column $[\mathbf{B}]_i$ to fixed $[\mathbf{B}^*]_i$.
- **Intervention space** $\mathcal{A} = 2^{[N]}$, so $|\mathcal{A}| = 2^N$, denote $a \in \mathcal{A}$
- Intervention weights $\mathbf{B}_a$: $[\mathbf{B}_a]_i = \begin{cases} [\mathbf{B}^*]_i, & \text{if } i \in a \\ [\mathbf{B}]_i, & \text{otherwise} \end{cases}$

objective: maximize a utility $g(X; a)$ of the network over $\mathcal{A}$

(w.l.o.g) **reward** Y = sink node X_N

sequential design of interventions $\rightarrow$ bandit framework!



Bandit Problem: expected reward : $\mu_a \triangleq \mathbb{E}_a[X_N]$

best intervention : $a^* \triangleq \arg\max_{a \in \mathcal{A}} \mu_a$

at time t: select a_t based on $\{X(s), a_s : s \in [t-1]\}$

cumulative regret: $\mathbb{E}[R(T)] \triangleq T\mu_{a^*} - \sum_{t=1}^T \mu_{a_t}$

Causal Information in Algorithm Design

- **Causal graph** $\mathcal{G}$: known vs. unknown
- **Interventional distributions** $\{\mathbb{P}_a : a \in \mathcal{A}\}$: known vs. unknown
- **Intervention models**: atomic vs. multi-node, *do* vs. soft
- **Random variables**: binary or continuous

main setting	$\{\mathbb{P}_a : a \in \mathcal{A}\}$ (partially) known	unknown $\{\mathbb{P}_a : a \in \mathcal{A}\}$
known $\mathcal{G}$	Lattimore et al., 2016 Sen et al., 2017 Lu et al., 2020 Nair et al., 2021	Yabe et al., 2018 (simple regret, binary RV) Maiti et al., 2022 (atomic interventions, binary RV) Feng and Chen, 2023 (binary RV) **THIS PAPER, JMLR'23**
unknown $\mathcal{G}$	Bilodeau et al., 2022	de Kroon et al., 2022 (no regret guarantees) Lu et al., 2021 (atomic interventions) Bilodeau et al., 2022 (regret scales with $\lvert\mathcal{A}\rvert$)

goal: use only $\mathcal{G}$ and **achieve optimal regret** $\mathcal{O}(\sqrt{T})$

main contribution: regret upper bound $\tilde{\mathcal{O}}(d^{L+\frac{1}{2}}\sqrt{NT})$

continuous RV, large space $|\mathcal{A}| = \exp(N)$, soft interventions!

LinSEM-UCB Overview

Expected rewards: $\mu_a = \langle f(\mathbf{B}_a), \nu \rangle$, $f(\mathbf{B}_a) = \sum_{\ell=0}^{L} [\mathbf{B}_a^\ell]_N$, $\nu = \mathbb{E}[\epsilon]$

share information

$a = \{1\} \rightarrow \mathbf{B}_a : \big[[\mathbf{B}^*]_1 \; \boxed{[\mathbf{B}]_2} \; [\mathbf{B}]_3 \; \boxed{[\mathbf{B}]_4} \big]$

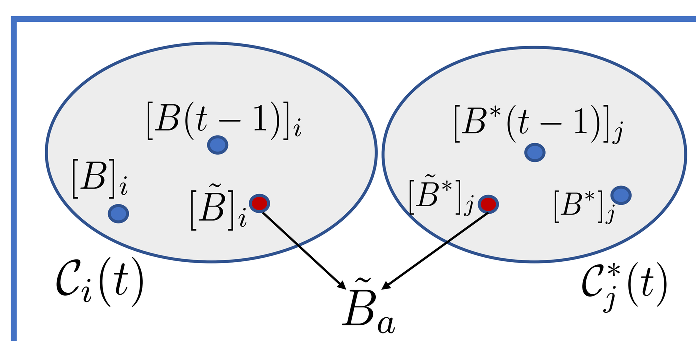
$a = \{3\} \rightarrow \mathbf{B}_a : \big[[\mathbf{B}]_1 \; \boxed{[\mathbf{B}]_2} \; [\mathbf{B}^*]_3 \; \boxed{[\mathbf{B}]_4} \big]$

- $2N$ vectors $\rightarrow 2^N$ intervention distributions

$[\mathbf{B}(t)]_i$ and $[\mathbf{B}^*(t)]_i$: via least-squares $X_i(1:t)$ & $X_{\mathrm{pa}(i)}(1:t)$

- **Upper confidence bound** (UCB)-based strategy
- Form **confidence intervals** $\mathcal{C}_i(t)$, choose a with largest upside:

$$a_t = \mathrm{argmax}_{a \in \mathcal{A}} \mathrm{UCB}_a(t)$$



$\mathrm{UCB}_a(t) = \max_{\tilde{\mathbf{B}}_a} \langle f(\tilde{\mathbf{B}}_a), \nu \rangle$

s.t. $[\tilde{\mathbf{B}}_a]_i \in \mathbb{I}(i \notin a) \cdot \mathcal{C}_i(t) + \mathbb{I}(i \in a) \cdot \mathcal{C}^*_i(t)$

Regret Bounds

Theorem (upper bound): The regret of LinSEM-UCB

$$\mathbb{E}[R(T)] = \mathcal{O}(d^{L+\frac{1}{2}}\sqrt{NT})$$

- **optimal** dependence on horizon $\sqrt{T}$
- **no explicit dependence** on cardinality $|\mathcal{A}|$
- estimation errors aggregate along $\mathcal{G}$: d parents, L deep

non-causal algorithms $\mathcal{O}(\sqrt{|\mathcal{A}|T})$ $\rightarrow$ unknown $\mathbb{P}_a$ $\tilde{\mathcal{O}}(d^{L+\frac{1}{2}}\sqrt{NT})$ $\leftarrow$ known $\mathbb{P}_a$ $\mathcal{O}(d\sqrt{T})$

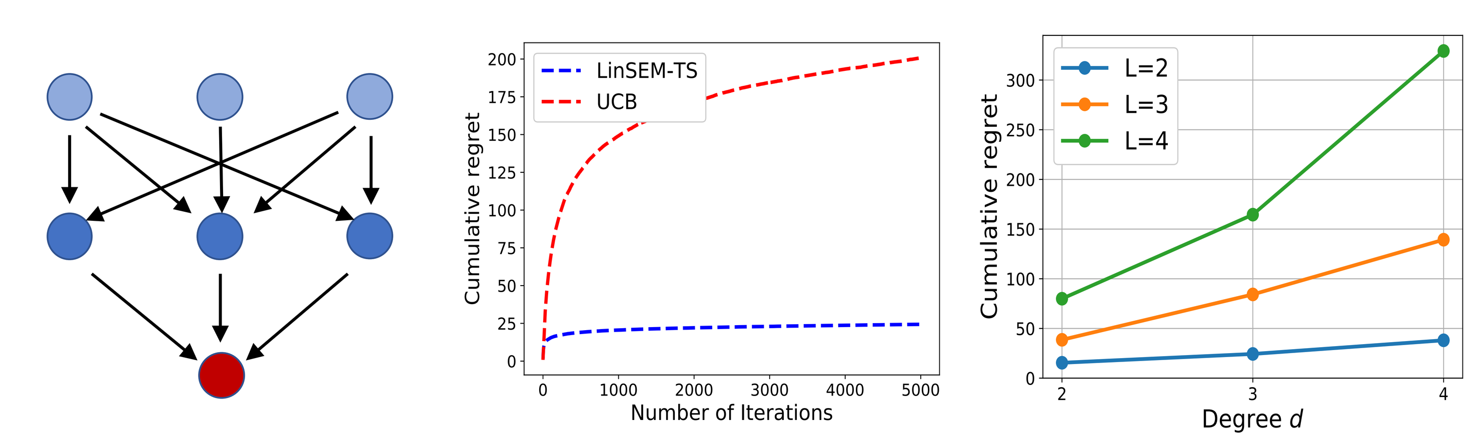
Theorem (minimax lower bound): Regret of any algorithm

$$\mathbb{E}[R(T)] = \Omega(d^{\frac{L}{2}+2}\sqrt{T})$$

- why? max. number of causal paths $\approx d^L$
- factor $\sqrt{N}$ in upper-bound: relaxed in follow-up

Simulations

- Linear Gaussian SEMs, hierarchical graphs $|\mathcal{A}| = 2^{d(L-1)}$
- LinSEM-Thompson sampling (ours) vs. non-causal UCB
- Varying d and L: effect is consistent with theory.



Check out follow-up papers on causal bandits

- *Linear Causal Bandits: Unknown Graph and Soft Interventions*, NeurIPS 2024. Unknown graph, improved upper bound matches lower bound (Yan and Tajer)
- *Robust Causal Bandits for Linear Models*, JSAIT 2024. Extends results to nonstationary linear SEMs (Yan, Mukherjee, Varıcı, Tajer)
- *Causal Bandits with General Causal Models and Interventions*, AISTATS 2024 Known graph, general SCMs, similar regret results (Yan, Wei, Katz, Sattigeri, Tajer)

contact: bvarici@andrew.cmu.edu