When Are Linear Stochastic Bandits Attackable?

Huazheng Wang, Haifeng Xu, Hongning Wang

ICML 2022

PRINCETON UNIVERSITY

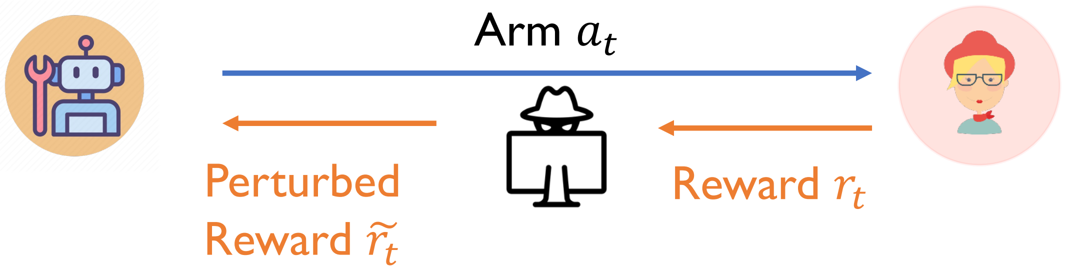
UNIVERSITY of VIRGINIA

Linear stochastic bandits

- Many real-world applications
  - Recommender system, advertisement, clinical trials, …
- Linear reward assumption: $r_t = x_{a_t}^T \theta^* + \eta_t$
- Minimize Regret $R(T) = \sum_{t=1}^{T} (\mathbb{E}[r^*] - \mathbb{E}[r_t])$
  - Equivalent to maximize the reward (rounds of pulling best arm)

Data poisoning attack

- Adversarial attack is a serious concern to ML systems
- Attacker : promote a target (suboptimal) arm $\tilde{x}$ by feeding perturbed rewards $\tilde{r}_t$ to the system
  - E.g., fake clicks, negative reviews to competitor's product
- Goal: fool the bandits to pull $\tilde{x}$ linear times using sublinear cost
  - Cost $C(T) = \sum_{t=1}^{T} |\tilde{r}_t - r_t|$



Attackability of a bandit environment

- Definition (informal): A bandit environment $\langle \mathcal{A}, \theta^* \rangle$ is attackable w.r.t. target arm $\tilde{x}$ if for any no-regret algorithm, the exists an attack method fools the bandits to pull $\tilde{x}$ $T - o(T)$ times using $o(T)$ cost for any large enough T

- Attackability is the property of an environment, not algorithm-specific
- Any MAB environment is attackable [Liu & Shroff, 2019]

- When Are Linear Stochastic Bandits Attackable?

Characterization of attackability

- Result 1: A bandit environment $\langle \mathcal{A}, \tilde{x}, \theta^* \rangle$ is attackable if and only if the following CQP's optimal objective $\epsilon^* > 0$

Perturbed reward

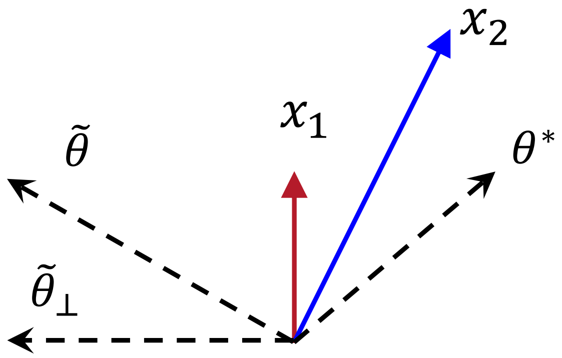
$$
\begin{aligned}
\max \quad & \epsilon \\
s.t. \quad & \tilde{x}^\top \theta_\parallel^* \geq \epsilon + \boxed{x_a^\top (\theta_\parallel^* + \tilde{\theta}_\perp)}, \qquad \forall x_a \neq \tilde{x} \\
& \tilde{x}^\top \tilde{\theta}_\perp = 0 \\
& \|\theta_\parallel^* + \tilde{\theta}_\perp\|_2 \leq 1
\end{aligned}
$$

- Key idea: decrease non-target arms' rewards in the null space of $\tilde{x}$ by $\tilde{\theta} = \theta_\parallel^* + \tilde{\theta}_\perp$ to make target arm the best
  - Increase target arm's reward requires linear cost [Feng et al., 2020]

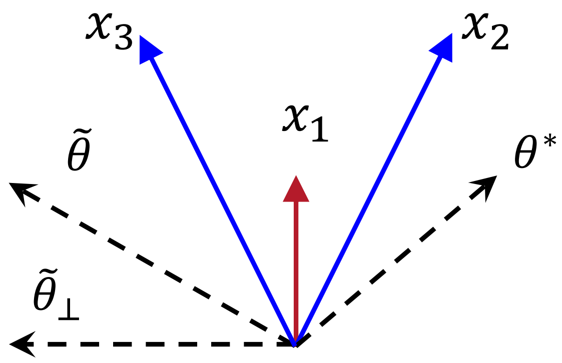
Unattackable environment: an example

- $\theta^* = (1,1)$
- $A = \{x_1 = (0,1), x_2 = (1,2)\}$
  - $r_1 = 1, r_2 = 3$
- Target arm $\tilde{x} = x_1$
- Attack according to $\tilde{\theta} = (-2,1)$
  - $\tilde{r}_1 = 1, \tilde{r}_2 = -1$

- $\theta^* = (1,1)$
- $A = \{x_1 = (0,1), x_2 = (1,2), x_3 = (-1,2)\}$
  - $r_1 = 1, r_2 = 3, r_3 = 1$
- Target arm $\tilde{x} = x_1$
- Attack according to $\tilde{\theta} = (?,1)$
  - Cannot find such $\tilde{\theta}$ to make x_1 the best



Insight of attackability: geometry (correlation) among arm features

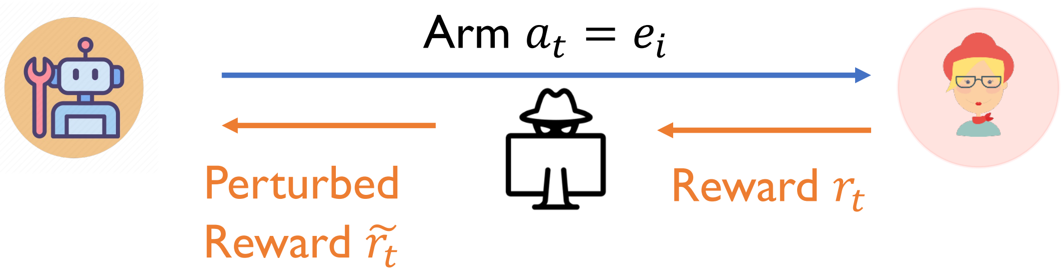


Attackability of MAB

- Since stochastic MAB is a special instance where $\mathcal{A} = \{e_1, \dots, e_K\}$, we have the following corollary:

For stochastic MAB, CQP is always feasible and the environment is always attackable for any target arm.

- Insight: reward estimates are independent for orthogonal arms
  - Attacker can arbitrarily decrease rewards of non-target arms
  - Recover similar attacks in [Jun et al., 2018 and Liu & Shroff, 2019] for MAB and [Garcelonet al., 2020] for k-armed linear contextual bandits



Oracle attack with known θ^*

- Following Theorem 1, we design Oracle Null Space Attack (with known θ^*) if environment is attackable

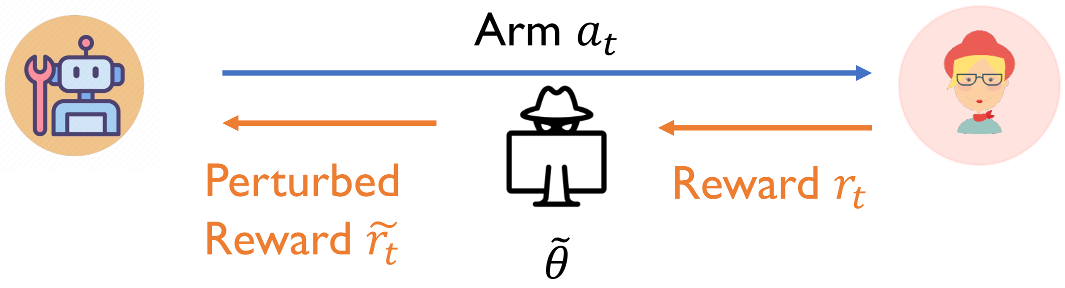
$$\tilde{r}_t = x_{a_t}^\top \tilde{\theta} + \eta_t, \text{where } \tilde{\theta} = \theta_{\parallel}^* + \tilde{\theta}_{\perp}$$

  - Any no-regret algorithm can be attacked with sublinear cost

- But in practice θ^* is unknown and can only be estimated online

Practical attack without knowing θ^*

- Two-stage Null Space Attack
  - First stage (T_1 rounds): collect rewards, estimate θ^*
  - Test attackability via CQP and compute $\tilde{\theta}$
  - Second stage: attack non-target arms according to $\tilde{\theta}$
    - Also compensate for rewards collected in the first stage



Practical attack without knowing θ^*

- Two-stage Null Space Attack
  - First stage (T_1 rounds): collect rewards, estimate θ^*
  - Test attackability via CQP and compute $\tilde{\theta}$
  - Second stage: attack non-target arms according to $\tilde{\theta}$
    - Also compensate for rewards collected in the first stage
- Result 2:

Target Algorithm	First stage rounds T_1	Cost $C(T)$	Non-target arm pulls
LinUCB [Li et al., 2010, Abbasi-yadkori et al., 2011]	$\sqrt{T}$	$\tilde{O}(T^{\frac{3}{4}})$	$\tilde{O}(T^{\frac{3}{4}})$
Robust Phase Elimination [Bogunovic et al., 2021]	$T^{\frac{2}{5}}$	$\tilde{O}(T^{\frac{4}{5}})$	$\tilde{O}(T^{\frac{4}{5}})$

Thank you

Check our paper for details!

ICML 2022 / https://arxiv.org/abs/2110.09008