Thresholded Linear Bandits

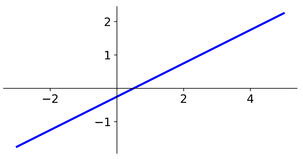
Nishant Mehta — University of Victoria
Junpei Komiyama — New York University
Vamsi Potluru — JPMorgan AI Research
Andrea Nguyen, Mica Grant-Hagen — University of Victoria

Linear Bandits: The Vanilla, The Generalized, and The Thresholded

Linear bandits

Unknown parameter $\theta^* \in \mathbb{R}^d$

Expected reward: $\mathbb{E}[r(a)] = \langle \theta^*, a \rangle$

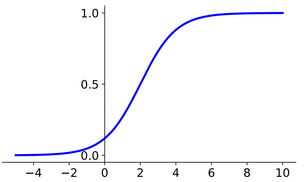


Generalized Linear Bandits

$\mathbb{E}[r(a)] = \sigma(\langle \theta^*, a \rangle - \tau)$

inverse link function (differentiable)

Logistic inverse link: $\sigma(z) = \dfrac{1}{1+e^{-z}}$



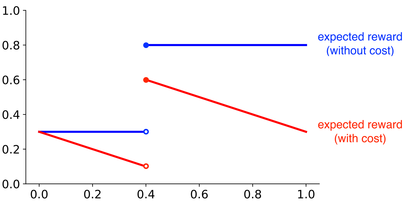
Thresholded Linear Bandits

Unknowns: threshold $\tau \in [0,1]$, $0 \le p_0 \le p_1 \le 1$

known cost vector

$\mathbb{E}[r(a)] = p_0 + \underbrace{(p_1 - p_0)}_{\text{gap } \Delta} 1[\langle \theta^*, a \rangle \ge \tau] - \langle v, a \rangle$

discontinuous!

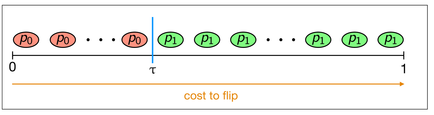


Toy Problem

Infinite sequence of coins in unit interval

Probability of heads: piecewise constant with a jump at threshold τ

Pay \$$x$ to flip coin $x \in [0,1]$. Collect \$1 if Heads and \$0 if Tails



Question: If sequentially flipping coins over T rounds, what is a good strategy to maximize expected cumulative reward?

One-Dimensional Case: Leftist

Characterization of optimal solution
Either arm 0 is optimal or arm τ is optimal

Possible to estimate threshold τ if $\Theta(\Delta)$ lower bound on Δ is known (use Noisy Binary Search)

Estimating Δ is easy:

Pull arm 0 (negative) and arm 1 (positive) in batches of doubling size

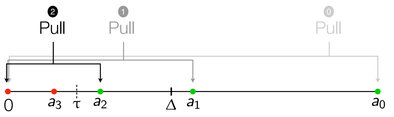
Once batch size of order $\frac{1}{\Delta^2}$, can find $\Theta(\Delta)$ lower bound on Δ

Problem: Pulling arm 1 for order $\frac{1}{\Delta^2}$ rounds can give $\frac{1}{\Delta^2}$ regret!

Smarter strategy:

In any batch r, consider pulling a smaller arm a_r in place of arm 1.

But τ unknown, so smaller arm might also be negative!



Leftist

For epochs $r = 0, 1, 2, \ldots$
  Set $a_r = 2^{-r}$ and $\varepsilon_r = a_r/8$
  Pull arms 0 and a_r for $n_r = \tilde{\Theta}(1/\varepsilon_r^2)$ rounds
  Set $\hat{\Delta}$ to empirical mean of arm a_r minus empirical mean of arm 0
  If $\mathrm{LCB}(\Delta) \ge \varepsilon_r$
    Use Noisy Binary Search to find approximate τ with $\hat{\tau} \ge \tau$
    Call UCB on arms $\{0, \hat{\tau}\}$

Key Lemma

Let $\Delta \ge \tau$. Then
(i) If Leftist is still running in epoch r, then arm a_r is positive.
(ii) After $O(\log(1/\Delta))$ epochs, Leftist finds a lower bound on Δ.

Regret Bounds for One-Dimensional Case

Theorem (Nice world)

If $\Delta \ge \tau$, then Leftist has regret of order at most

$$\frac{\log^2 T}{\Delta} + \min\left\{\sqrt{T \log T}, \frac{\log T}{|\Delta - \tau|}\right\}$$

Theorem (Fallback guarantee)

Unconditionally, Leftist's regret is always of order at most

$$\log(T)\sqrt{T} + \min\left\{\sqrt{T \log T}, \frac{\log T}{|\Delta - \tau|}\right\}$$

Theorem (Lower bound)

For any algorithm, there is a problem instance on which it must have $\Omega(\sqrt{T})$ regret.

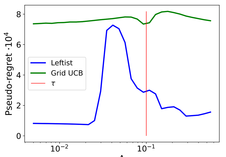
Conjecture: No algorithm can always obtain $\frac{\mathrm{polylog}(T)}{\Delta}$ regret.

Simulations for One-Dimensional Case

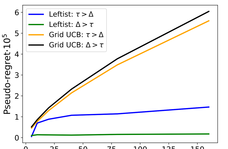
Compare to baseline Grid-UCB: UCB run on grid of $\sqrt{T}$ arms

Why Grid-UCB? Claim: Grid-UCB has worst-case regret $\tilde{O}(\sqrt{T})$

$\tau = 0.1$ varying Δ

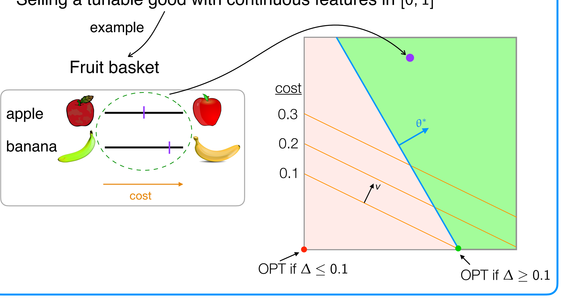


Case $\tau > \Delta$: $\Delta = 0.05$, $\tau = 0.1$
Case $\Delta > \tau$: $\Delta = 0.30$, $\tau = 0.2$



Multi-Dimensional Applications

Selling a tunable good with continuous features in $[0,1]$



Multidimensional Case

Characterization of optimal solution
Optimal arm is either arm **0** or minimum cost positive arm

How to find minimum cost positive arm?
Solve problem: Minimize cost $\langle v, a \rangle$ subject to utility $\langle \theta^*, a \rangle \ge \tau$

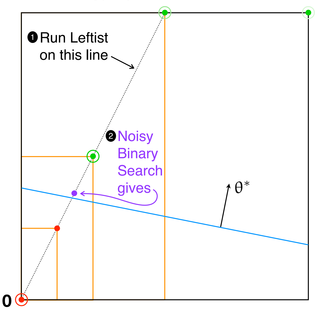
$\Updownarrow$

Dual of fractional knapsack!

Greedy solution if problem instance known:
(1) Sort coordinates by decreasing ratio $\frac{\theta_j^*}{v_j}$
(2) In order, fill each coordinate until $\langle \theta^*, a \rangle = \tau$

MultiDimLeftist (MD-Leftist)

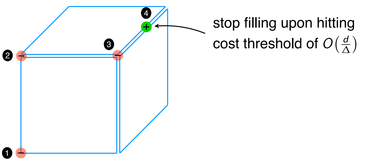
(1) Find point close to hyperplane



(2) Approximately sort coordinates via "coordinate dueling" (technical)

coordinates in sorted order: (↑ , → , ↗)

(3) Use MultiGreedy to find approximately minimum cost positive arm $\hat{a}$



(4) Call UCB on arms $\{\mathbf{0}, \hat{a}\}$

Regret Bounds for Multidimensional Case

Regret bounds below rely on Regularity Assumption:

There is a constant $\theta^*_{\min} > 0$ such that $\theta^*_j > 0 \Rightarrow \theta^* \ge \theta_{\min}$

Theorem (Nice world)

Suppose optimal arm is positive and $\Delta \le v_{\min}/16$ and Regularity Assumption holds. Then MD-Leftist has regret or order at most

$$\log(T) \cdot \left(\frac{\Delta + \|v\|_1}{v_{\min}^2} + \frac{\max\{d \log T, d^2 \log d\}}{\Delta}\right) + \min\left\{\sqrt{T \log T}, \frac{\log T}{|\Delta - v^*|}\right\}$$

Theorem (Fallback guarantee)

Suppose $\Delta \le v_{\min}/16$ and Regularity Assumption holds.
Then MD-Leftist has regret or order at most

$$\log(T) \cdot \frac{\Delta + \|v\|_1}{v_{\min}^2} + \max\{\sqrt{d} \log T, d^{1.5} \log d\}\sqrt{T} + \min\left\{\sqrt{T \log T}, \frac{\log T}{|\Delta - v^*|}\right\}$$