Online Learning with Composite Loss Functions

Ofer Dekel (MSR) Jian Ding (UChicago)
Tomer Koren (Technion) Yuval Peres (MSR)

Conference on Learning Theory
Barcelona, June 2014

Oblivious Multiarmed Bandit

Player plays a repeated game against an *oblivious adversary*.
Players action set is the set of *arms*: $[k] = \{1, \ldots, k\}$.

- adversary defines loss functions $f_1, \ldots, f_T$, where $f_t : [k] \mapsto [0, 1]$
- for $t = 1, \ldots, T$
  1. player (randomly) chooses an arm $X_t \in [k]$
  2. player incurs loss $f_t(X_t)$
  3. bandit feedback: player observes $f_t(X_t)$

Measuring the Difficulty of the Game

Define: The player's regret

$$R(T) \;=\; \mathbb{E}\left[\sum_{t=1}^{T} f_t(X_t)\right] - \min_{x\in[k]} \sum_{t=1}^{T} f_t(x) \quad .$$

Define: The game's minimax regret

$$\mathcal{R}(T) \;=\; \inf_{\text{player}} \; \sup_{\text{adversary}} R(T) \quad .$$

Theorem (ACFS02, AB09) $\mathcal{R}(T) \;=\; \Theta(\sqrt{T}) \quad .$

We say that "oblivious multiarmed bandit is an easy problem"

1-Memory Multiarmed Bandit

Player plays a repeated game against a *1-memory adversary*.
Players action set is the set of *arms*: $[k] = \{1, \ldots, k\}$.

- adversary defines loss functions $f_1, \ldots, f_T$, where $f_t : [k] \times [k] \mapsto [0, 1]$
- for $t = 1, \ldots, T$
  1. player (randomly) chooses an arm $X_t \in [k]$
  2. player incurs loss $f_t(X_{t-1}, X_t)$
  3. bandit feedback: player observes $f_t(X_{t-1}, X_t)$

Measuring the Difficulty of the Game

Define: The player's regret

$$R(T) = \mathbb{E}\left[\sum_{t=1}^{T} f_t(X_{t-1}, X_t)\right] - \min_{x \in [k]} \sum_{t=1}^{T} f_t(x, x) \quad .$$

Define: The game's minimax regret

$$\mathcal{R}(T) = \inf_{\text{player}} \sup_{\text{adversary}} R(T) \quad .$$

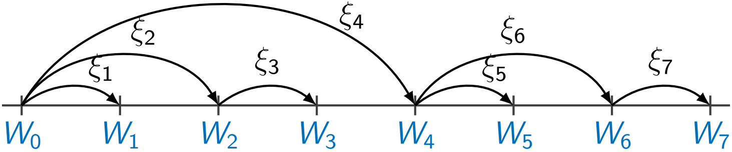
Theorem (DDKP14) $\mathcal{R}(T) = \widetilde{\Theta}\left(T^{2/3}\right)$.

"1-memory multiarmed bandit is hard, but still learnable"

Analysis: Multiscale Random Walk

Assume $k = 2$. Define the loss of arm 1:

- Draw $\xi_1, \ldots, \xi_T$ i.i.d. Gaussians
- Recursively define: $W_0 = \frac{1}{2}$ and $W_t = W_{\rho(t)} + \xi_t$, with $\rho(t) = t - gcd(t, 2^T)$

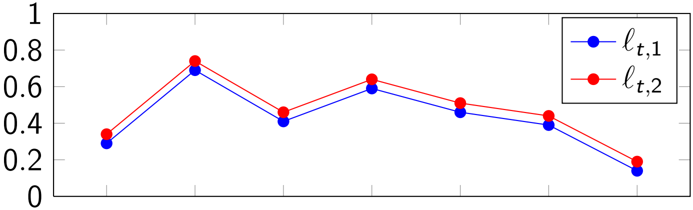


Define the loss of arm 2 to be either ϵ greater or ϵ less.

Analysis: Bandits and Switching

The result: $\ell_1, \ldots, \ell_T \;\; (\ell_t \in [0,1]^2)$ such that

- Each pull of suboptimal arm adds ϵ to the regret
- Player has to switch $\frac{1}{\epsilon^2}$ times to identify better arm



Theorem (DDKP14) Bandit with switching costs, $f_t(X_{t\text{-}1}, X_t) = \ell_t(X_t) + \mathbb{1}_{X_t \neq X_{t\text{-}1}}$, has $\mathcal{R}(T) = \widetilde{\Theta}(T^{2/3})$

Composite Loss Functions

Assume the adversary defines loss functions as follows

- adversary chooses oblivious functions $\ell_1, \ldots, \ell_T$ (namely, $\ell_t : [k] \mapsto [0, 1]$)
- adversary applies a known *loss combining function* $g : [0, 1] \times [0, 1] \mapsto [0, 1]$ to define

$$f_t(X_{t-1}, X_t) \; = \; g\Big(\ell_{t-1}(X_{t-1}), \ell_t(X_t)\Big)$$

- Feedback: player may even observe $\ell_t(X_t)$

Examples $g(a, b) = \min(a, b), \quad g(a, b) = \max(a, b)$

Results

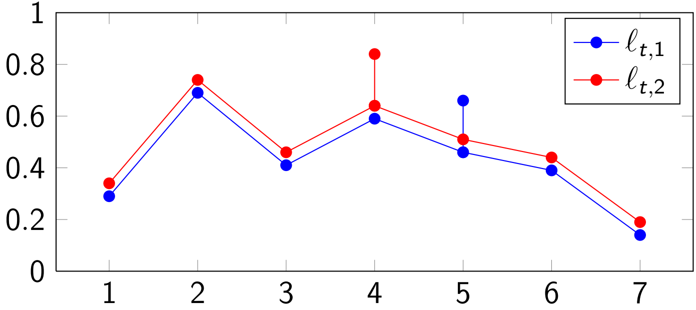
Theorem

- $g(a, b)$ is linear $\quad \Rightarrow \quad \mathcal{R}(T) = \Theta(\sqrt{T})$ (easy)
- $g(a, b) = \min(a, b) \quad \Rightarrow \quad \mathcal{R}(T) = \widetilde{\Theta}(T^{2/3})$ (hard)
- $g(a, b) = \max(a, b) \quad \Rightarrow \quad \mathcal{R}(T) = \widetilde{\Theta}(T^{2/3})$ (hard)

Analysis: Spiking the Loss

Randomly add pairs of spikes to simulate a switching cost

- They are too small to be detected by the player
- Each switch adds $\widetilde{\Theta}(1)$ to the regret
- When no-switch, spikes have negligible effect



More details at our poster.