Online Learning over a Finite Action Set with Limited Switching

Jason Altschuler & Kunal Talwar

COLT 2018

Online learning over a finite action set: classical setup

- T-iteration repeated game between algorithm & adversary:

In each iteration $t \in \{1,\ldots,T\}$

Choice. Simultaneously:

Algorithm (randomly) chooses action $I_t \in \{1,\ldots,n\}$

Adversary chooses losses $\ell_t : \{1,\ldots,n\} \to [0,1]$

Feedback. Either:

Prediction from Experts (PFE) setting:
Algorithm observes all losses ℓ_t

Multi-Armed Bandit (MAB) setting:
Algorithm observes only $\ell_t(I_t)$

- Classical goal: min cumulative loss w.r.t. meaningful baseline:

$$\text{Regret}_T := \sum_{t=1}^{T} \ell_t(I_t) - \min_{i^* \in \{1,\ldots,n\}} \sum_{t=1}^{T} \ell_t(i^*)$$

Switching as a resource

- **Switching between actions is bad** in applications

$$\text{Switches}_T := \sum_{t=2}^{T} \mathbf{1}\{I_t = I_{t-1}\}$$

- Many such applications [see paper for long list…]
- This motivates viewing switching as a *resource*.
- Leads to a *bi-criteria* optimization problem. Formalize by:

Switching-cost: incur additional loss of c every switch.
[Expensive but unlimited.]

Switching-budget: limited to S total switches in game.
[Free but limited.]

Our goal: understand tradeoff between Regret & Switches.

Our Contributions

1. Present the **first PFE algorithms which w.h.p. achieve the optimal order for both Regret and Switches**, resolving COLT 2013 open problem of Devroye, Lugosi, and Neu.
   - Many existing algorithms work in expectation, but no h.p. guarantees.
   - Efficiently extendable to online combinatorial optimization with limited switching.
2. Using the above and several reductions, we unify previous work and **completely characterize the complexity of the switching-budget problem** (up to small polylog factors): for both the PFE and MAB problems, for all switching budgets, and for both expectation and h.p. guarantees.
   - Shows qualitatively different behaviors for full-info & partial-info settings.
   - Implies duality between switching costs & switching budgets (a priori, only one reduction is trivial).

Contribution 1: first h.p. algorithms for switching-cost PFE

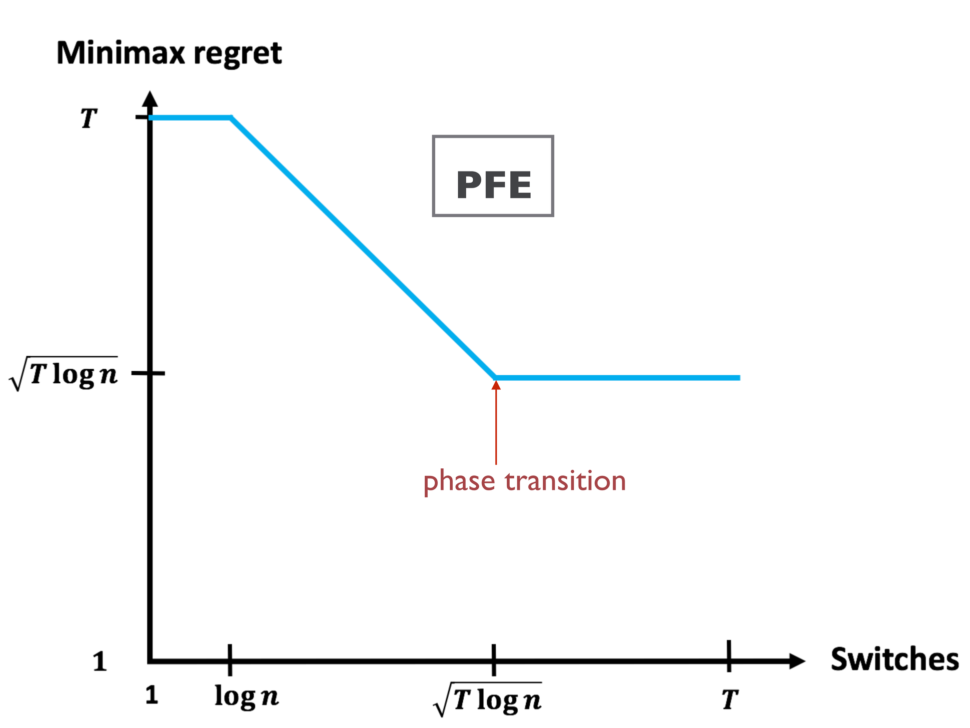
- General framework to convert an algorithm with optimal Regret & Switches *expectation guarantees*, into an algorithm with analogous *h.p. guarantees*:

```
while in iteration ≤ T do
    Run A with fresh randomness. Stop when use S' = O(√(T log n / log(1/δ))) switches.
end
```

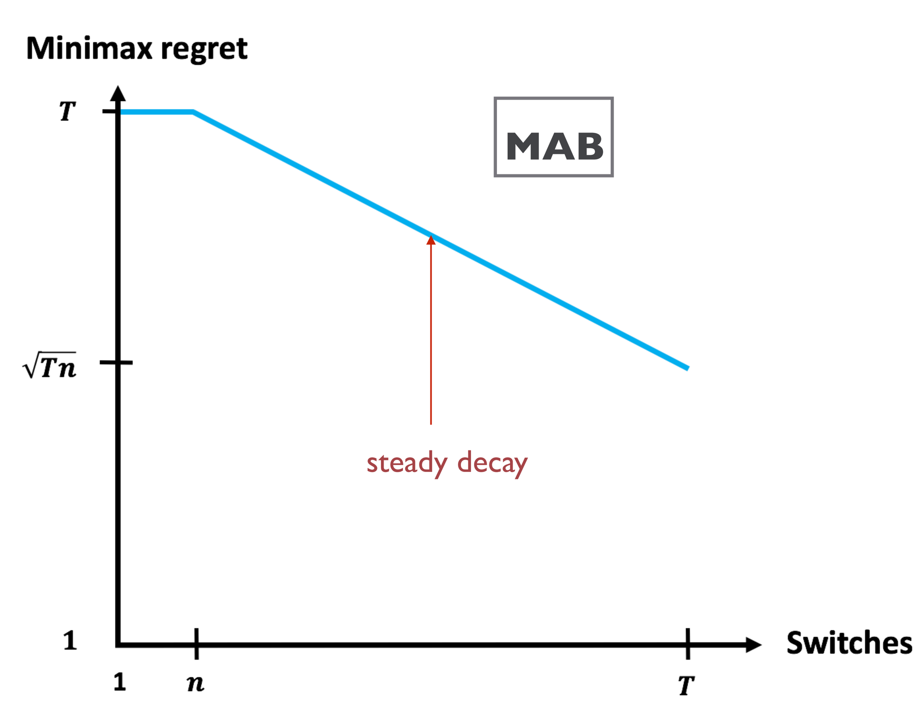
- In words: split T iterations into $N \approx \log \frac{1}{\delta}$ variable-length epochs. Restart epoch once uses S' switches, with fresh randomness.
  - Variable-length epoch is (provably) essential.
- Analysis is broken into 2 parts:
  1. **H.p. switching guarantee:** show $\mathbb{P}(\#\text{ epochs} > N) \leq e^{-N}$
     - Can prove in black-box manner with just E[switching] bounds for A (no other info on A needed)
  2. **H.p. regret guarantee:** show cumulative regret concentrates around (# epochs) x (E[Regret] in single epoch).
     - Can do for FPL-based algorithms.
     - This part of the analysis is not black-boxed as it depends on the algorithm A used.
- Examples of algorithms A that work with our framework:
  - Multiplicative Follow the Perturbed Leader [Kalai and Vempala, 2005]
  - Prediction by Random Walk Perturbation (+ combinatorial version) [Devroye, Lugosi, and Neu, 2013]

Open question: *uniform* h.p. algorithms?

Contribution 2: complexity landscape of online learning with limited switching



	LB on $\mathbb{E}[\text{Regret}]$	UB on $\mathbb{E}[\text{Regret}]$	High probability UB
Unconstrained switching	$\sqrt{T \log n}$	$\sqrt{T \log n}$	$\sqrt{T \log \frac{n}{\delta}}$
c **switching cost**	$\sqrt{cT \log n}$	$\sqrt{cT \log n}$	$\sqrt{cT \log n \log \frac{1}{\delta}}$
$S = \Omega(\sqrt{T \log n})$ **switching budget**	$\sqrt{T \log n}$	$\sqrt{T \log n} \log T$	$\sqrt{T \log n} \log \frac{1}{\delta}$
$S = O(\sqrt{T \log n})$ **switching budget**	$\frac{T \log n}{S}$	$\frac{T \log n}{S} \log T$	$\frac{T \log n}{S} \log \frac{1}{\delta}$



	LB on $\mathbb{E}[\text{Regret}]$	UB on $\mathbb{E}[\text{Regret}]$	High probability UB
Unconstrained switching	$\sqrt{Tn}$	$\sqrt{Tn}$	$\sqrt{Tn} \frac{\log \frac{n}{\delta}}{\sqrt{\log n}}$
c **switching cost**	$\frac{c^{1/3} T^{2/3} n^{1/3}}{\log T}$	$c^{1/3} T^{2/3} n^{1/3}$	$c^{1/3} T^{2/3} n^{1/3} \frac{\log^{2/3} \frac{n}{\delta}}{\log^{1/3} n}$
S **switching budget**	$\frac{T\sqrt{n}}{\sqrt{S} \log^{3/2} T}$	$\frac{T\sqrt{n}}{\sqrt{S}}$	$\frac{T\sqrt{n}}{\sqrt{S}} \frac{\log \frac{n}{\delta}}{\sqrt{\log n}}$