More Adaptive Algorithms for Adversarial Bandits

Chen-Yu Wei and Haipeng Luo

University of Southern California

Multi-Armed Bandit

- For $t = 1, \ldots, T$,
 - Player picks arm $i_t \in \{1, \dots, K\}$
 - Adversary reveals the loss of arm $i_t: \ell_{t,i_t} \in [0,1]$ (but not $\ell_{t,i}$ for $i \neq i_t$)
 - Player suffers loss ℓ_{t,i_t} in this round

Multi-Armed Bandit

- For $t = 1, \ldots, T$,
 - Player picks arm $i_t \in \{1, \dots, K\}$
 - Adversary reveals the loss of arm $i_t: \ell_{t,i_t} \in [0,1]$ (but not $\ell_{t,i}$ for $i \neq i_t$)
 - Player suffers loss ℓ_{t,i_t} in this round
- Goal: minimize the regret against the best arm:

regret =
$$\sum_{t=1}^{T} \ell_{t,i_t} - \min_i \sum_{t=1}^{T} \ell_{t,i}$$

Multi-Armed Bandit

- For $t = 1, \ldots, T$,
 - Player picks arm $i_t \in \{1, \dots, K\}$
 - Adversary reveals the loss of arm i_t : $\ell_{t,i_t} \in [0,1]$ (but not $\ell_{t,i}$ for $i \neq i_t$)
 - Player suffers loss ℓ_{t,i_t} in this round
- Goal: minimize the regret against the best arm:

regret =
$$\sum_{t=1}^{T} \ell_{t,i_t} - \min_i \sum_{t=1}^{T} \ell_{t,i}$$

Target of this work: designing algorithms that always have (nearly) minimax regret guarantee $(\mathcal{O}(\sqrt{KT}))$ but are much better when data is easy.

- Using a SINGLE algorithm when losses are **i.i.d.** $\Rightarrow \mathcal{O}\left(\frac{K \log T}{\Delta}\right)$ when losses are **adversarial** $\Rightarrow \tilde{\mathcal{O}}\left(\sqrt{KL^*}\right)$
 - Δ : gap between the mean of best arm and the 2nd-best arm • $L^* = \sum_{t=1}^{T} \ell_{t,i^*}$: best arm's total loss

• Using a SINGLE algorithm when losses are **i.i.d.** $\Rightarrow \mathcal{O}\left(\frac{K \log T}{\Delta}\right)$

when losses are **adversarial** $\Rightarrow \tilde{O}\left(\sqrt{\kappa L^*}\right)$

- Δ : gap between the mean of best arm and the 2nd-best arm • $L^* = \sum_{t=1}^{T} \ell_{t,i^*}$: best arm's total loss
- Similar to

• Using a SINGLE algorithm when losses are **i.i.d.** $\Rightarrow \mathcal{O}\left(\frac{K \log T}{\Delta}\right)$

when losses are **adversarial** $\Rightarrow \tilde{\mathcal{O}}\left(\sqrt{\mathcal{KL}^*}\right)$

- Δ : gap between the mean of best arm and the 2nd-best arm • $L^* = \sum_{t=1}^{T} \ell_{t,i^*}$: best arm's total loss
- Similar to

[Bubeck&Slivkins'12, Seldin&Slivkins'14, Auer&Chiang'16, Seldin&Lugosi'17]

• Benefits of our algorithm:

• Using a SINGLE algorithm when losses are **i.i.d.** $\Rightarrow \mathcal{O}\left(\frac{K \log T}{\Delta}\right)$

when losses are **adversarial** $\Rightarrow \tilde{\mathcal{O}}\left(\sqrt{\kappa L^*}\right)$

- Δ : gap between the mean of best arm and the 2nd-best arm • $L^* = \sum_{t=1}^{T} \ell_{t,i^*}$: best arm's total loss
- Similar to

- Benefits of our algorithm:
 - In the adversarial setting: $\sqrt{KT} \rightarrow \sqrt{KL^*}$

• Using a SINGLE algorithm when losses are **i.i.d.** $\Rightarrow \mathcal{O}\left(\frac{K \log T}{\Delta}\right)$

when losses are **adversarial** $\Rightarrow \tilde{\mathcal{O}}\left(\sqrt{\kappa L^*}\right)$

- Δ : gap between the mean of best arm and the 2nd-best arm • $L^* = \sum_{t=1}^{T} \ell_{t,i^*}$: best arm's total loss
- Similar to

- Benefits of our algorithm:
 - In the adversarial setting: $\sqrt{\textit{KT}} \rightarrow \sqrt{\textit{KL}^*}$
 - In the stochastic setting: ^{K log T}/_Δ bound holds under weaker assumption: E_t[ℓ_{t,i*}] ≤ E_t[ℓ_{t,i}] + Δ (can be neither independent nor identical)

• Using a SINGLE algorithm when losses are **i.i.d.** $\Rightarrow \mathcal{O}\left(\frac{K \log T}{\Delta}\right)$

when losses are **adversarial** $\Rightarrow \tilde{\mathcal{O}}\left(\sqrt{\mathcal{KL}^*}\right)$

- Δ : gap between the mean of best arm and the 2nd-best arm
- $L^* = \sum_{t=1}^{T} \ell_{t,i^*}$: best arm's total loss
- Similar to

- Benefits of our algorithm:
 - In the adversarial setting: $\sqrt{\textit{KT}} \rightarrow \sqrt{\textit{KL}^*}$
 - In the stochastic setting: $\frac{K \log T}{\Delta}$ bound holds under weaker assumption: $\mathbb{E}_t[\ell_{t,i^*}] \leq \mathbb{E}_t[\ell_{t,i}] + \Delta$ (can be neither independent nor identical)
 - Much SIMPLER algorithm and analysis: no extra statistical tests are required

• when losses have small empirical variance

•
$$Q_i = \sum_{t=1}^{T} (\ell_{t,i} - \mu_i)^2$$
, where $\mu_i = \frac{1}{T} \sum_{t=1}^{T} \ell_{t,i}$.

• when losses have small empirical variance

•
$$Q_i = \sum_{t=1}^{T} (\ell_{t,i} - \mu_i)^2$$
, where $\mu_i = \frac{1}{T} \sum_{t=1}^{T} \ell_{t,i}$.

tull-into	bandıt
$\sqrt{\max_i Q_i}$ [Hazan&Kale'08]	$\sqrt{\sum_i Q_i}$ [Bubeck et al.'18]
$\sqrt{Q_{i^*}}$ [Steinhardt&Liang'14]	

۲

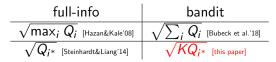
• when losses have small empirical variance

$$Q_{i} = \sum_{t=1}^{T} (\ell_{t,i} - \mu_{i})^{2}, \text{ where } \mu_{i} = \frac{1}{T} \sum_{t=1}^{T} \ell_{t,i}.$$

$$\frac{\text{full-info}}{\sqrt{\max_{i} Q_{i}}} \frac{\text{bandit}}{\sqrt{\sum_{i} Q_{i}}} \frac{\sqrt{\sum_{i} Q_{i}}}{(\text{Bubeck et al.}^{18})}$$

• when losses have small empirical variance

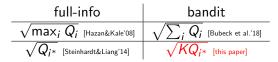
•
$$Q_i = \sum_{t=1}^{T} (\ell_{t,i} - \mu_i)^2$$
, where $\mu_i = \frac{1}{T} \sum_{t=1}^{T} \ell_{t,i}$.



• when losses have small path length

• when losses have small empirical variance

•
$$Q_i = \sum_{t=1}^{T} (\ell_{t,i} - \mu_i)^2$$
, where $\mu_i = \frac{1}{T} \sum_{t=1}^{T} \ell_{t,i}$.

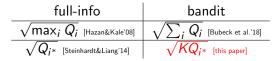


• when losses have small path length

•
$$D_i = \sum_t (\ell_{t,i} - \ell_{t-1,i})^2$$

• when losses have small empirical variance

•
$$Q_i = \sum_{t=1}^{T} (\ell_{t,i} - \mu_i)^2$$
, where $\mu_i = \frac{1}{T} \sum_{t=1}^{T} \ell_{t,i}$.



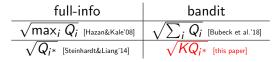
when losses have small path length

•
$$D_i = \sum_t (\ell_{t,i} - \ell_{t-1,i})^2$$

full-info	bandit
$\sqrt{\max_i D_i}$ [Chiang et al. 2012]	
$\sqrt{D_{i^*}}$ [Steinhardt&Liang 2014]	

• when losses have small empirical variance

•
$$Q_i = \sum_{t=1}^{T} (\ell_{t,i} - \mu_i)^2$$
, where $\mu_i = \frac{1}{T} \sum_{t=1}^{T} \ell_{t,i}$.



when losses have small path length

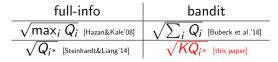
•
$$D_i = \sum_t (\ell_{t,i} - \ell_{t-1,i})^2 \longrightarrow V_i = \sum_t |\ell_{t,i} - \ell_{t-1,i}|$$

$$\frac{\text{full-info} \qquad \text{bandit}}{\sqrt{\max_i D_i} \text{ [Chiang et al. 2012]}}$$

$$\frac{\sqrt{D_i^*} \text{ [Steinhardt&Liang 2014]}}$$

• when losses have small empirical variance

•
$$Q_i = \sum_{t=1}^{T} (\ell_{t,i} - \mu_i)^2$$
, where $\mu_i = \frac{1}{T} \sum_{t=1}^{T} \ell_{t,i}$.



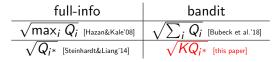
when losses have small path length

•
$$D_i = \sum_t (\ell_{t,i} - \ell_{t-1,i})^2 \longrightarrow V_i = \sum_t |\ell_{t,i} - \ell_{t-1,i}|$$

$$\frac{\text{full-info}}{\sqrt{\max_i D_i}} \frac{\text{bandit}}{[\text{Chiang et al. 2012}]} \frac{\sqrt{K \sum_i V_i}}{\sqrt{V_i^*} \text{ [this paper]}}$$

• when losses have small empirical variance

•
$$Q_i = \sum_{t=1}^{T} (\ell_{t,i} - \mu_i)^2$$
, where $\mu_i = \frac{1}{T} \sum_{t=1}^{T} \ell_{t,i}$.



• when losses have small path length

•
$$D_i = \sum_t (\ell_{t,i} - \ell_{t-1,i})^2 \longrightarrow V_i = \sum_t |\ell_{t,i} - \ell_{t-1,i}|$$

full-info
bandit

 $\frac{\sqrt{\max_i D_i} \text{ [Chiang et al. 2012]}}{\sqrt{D_i^*} \text{ [Steinhardt&Liang 2014]}} \frac{\sqrt{K \sum_i V_i} \text{ [this paper]}}{K \sqrt{V_i^*} \text{ [this paper]}}$

Application: faster convergence (1/T³/₄) for multi-player games with bandit feedback (~[Rakhlin&Sridharan'13, Syrgkanis et al.'15, Abernethy et al.'18]). Typical bandit algorithm: 1/√T.

• BROAD=Barrier-Regularized with Optimism and ADaptivity

- BROAD=Barrier-Regularized with Optimism and ADaptivity
- Online Mirror Descent (OMD):

Sample $i_t \sim p_t$ $p_{t+1} = \arg \min_p \left\{ \langle p, \hat{\ell}_t \rangle + D_{\psi_t}(p, p_t) \right\}$

- BROAD=Barrier-Regularized with Optimism and ADaptivity
- Optimistic OMD [Rakhlin&Sridharan'13]:

Sample
$$i_t \sim p_t$$

 $p'_{t+1} = \arg \min_p \left\{ \langle p, \hat{\ell}_t \rangle + D_{\psi_t}(p, p'_t) \right\}$
 $p_{t+1} = \arg \min_p \left\{ \langle p, m_{t+1} \rangle + D_{\psi_{t+1}}(p, p'_{t+1}) \right\},$

- BROAD=Barrier-Regularized with Optimism and ADaptivity
- OMD with Adaptivity and Optimism [~ Steinhardt&Liang'14]:

Sample
$$i_t \sim p_t$$

 $p'_{t+1} = \arg \min_p \left\{ \langle p, \hat{\ell}_t + a_t \rangle + D_{\psi_t}(p, p'_t) \right\}$
 $p_{t+1} = \arg \min_p \left\{ \langle p, m_{t+1} \rangle + D_{\psi_{t+1}}(p, p'_{t+1}) \right\},$

- BROAD=Barrier-Regularized with Optimism and ADaptivity
- OMD with Adaptivity and Optimism [~ Steinhardt&Liang'14]:

Sample
$$i_t \sim p_t$$

 $p'_{t+1} = \arg \min_p \left\{ \langle p, \hat{\ell}_t + a_t \rangle + D_{\psi_t}(p, p'_t) \right\}$
 $p_{t+1} = \arg \min_p \left\{ \langle p, m_{t+1} \rangle + D_{\psi_{t+1}}(p, p'_{t+1}) \right\},$

where ψ_t is a time-varying **log-barrier** [Foster et al.'16]:

$$\psi_t(p) = \sum_{i=1}^K \frac{1}{\eta_{t,i}} \log \frac{1}{p_i}$$

- BROAD=Barrier-Regularized with Optimism and ADaptivity
- OMD with Adaptivity and Optimism [~ Steinhardt&Liang'14]:

Sample
$$i_t \sim p_t$$

 $p'_{t+1} = \arg \min_p \left\{ \langle p, \hat{\ell}_t + a_t \rangle + D_{\psi_t}(p, p'_t) \right\}$
 $p_{t+1} = \arg \min_p \left\{ \langle p, m_{t+1} \rangle + D_{\psi_{t+1}}(p, p'_{t+1}) \right\},$

where ψ_t is a time-varying **log-barrier** [Foster et al.'16]:

$$\psi_t(p) = \sum_{i=1}^K \frac{1}{\eta_{t,i}} \log \frac{1}{p_i}$$

• Set $a_t = 0$ with appropriately chosen m_t to achieve $\sqrt{K \sum_i V_i}$ and best of both worlds.

- BROAD=Barrier-Regularized with Optimism and ADaptivity
- OMD with Adaptivity and Optimism [~ Steinhardt&Liang'14]:

Sample
$$i_t \sim p_t$$

 $p'_{t+1} = \arg \min_p \left\{ \langle p, \hat{\ell}_t + a_t \rangle + D_{\psi_t}(p, p'_t) \right\}$
 $p_{t+1} = \arg \min_p \left\{ \langle p, m_{t+1} \rangle + D_{\psi_{t+1}}(p, p'_{t+1}) \right\},$

where ψ_t is a time-varying **log-barrier** [Foster et al.'16]:

$$\psi_t(p) = \sum_{i=1}^K \frac{1}{\eta_{t,i}} \log \frac{1}{p_i}$$

- Set $a_t = 0$ with appropriately chosen m_t to achieve $\sqrt{K \sum_i V_i}$ and best of both worlds.
- Set $a_{t,i} = 6\eta_{t,i} p_{t,i} (\hat{\ell}_{t,i} m_{t,i})^2$ with appropriately chosen m_t to adapt to the best arm: $\sqrt{KQ_{i^*}}$ and $K\sqrt{V_{i^*}}$

Other Elements / Open Problems

- To get some of the results, **increasing learning rates** are required; for some other results, **decreasing learning rates** are required.
- Most of our results can be generalized to **combinatorial bandits** with semi-bandit feedback.

Other Elements / Open Problems

- To get some of the results, **increasing learning rates** are required; for some other results, **decreasing learning rates** are required.
- Most of our results can be generalized to **combinatorial bandits** with semi-bandit feedback.

Open Problems:

- Parameter-free algorithms that achieve $\sqrt{KQ_{i^*}}$ and $K\sqrt{V_{i^*}}$.
- Second-order path-length bound for bandit
- Extensions to other bandit settings (e.g., linear/contextual)