Stability-penalty-adaptive follow-the-regularized-leader: Sparsity, game-dependency, and best-of-both-worlds

Taira Tsuchiya [1], Shinji Ito [2,3], Junya Honda [4,3]

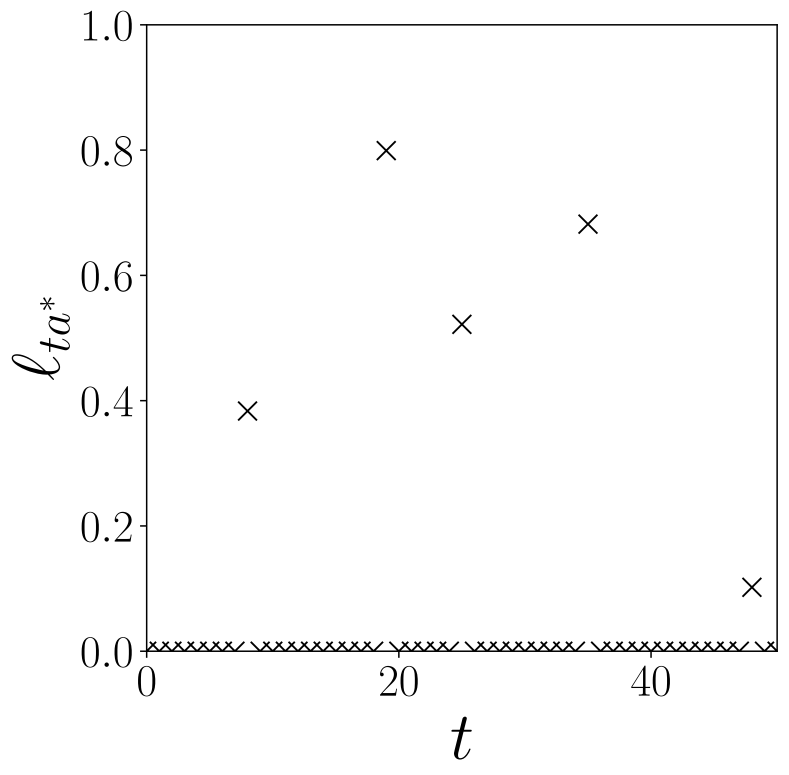
1. The University of Tokyo, 2. NEC, 3. RIKEN, 4. Kyoto University

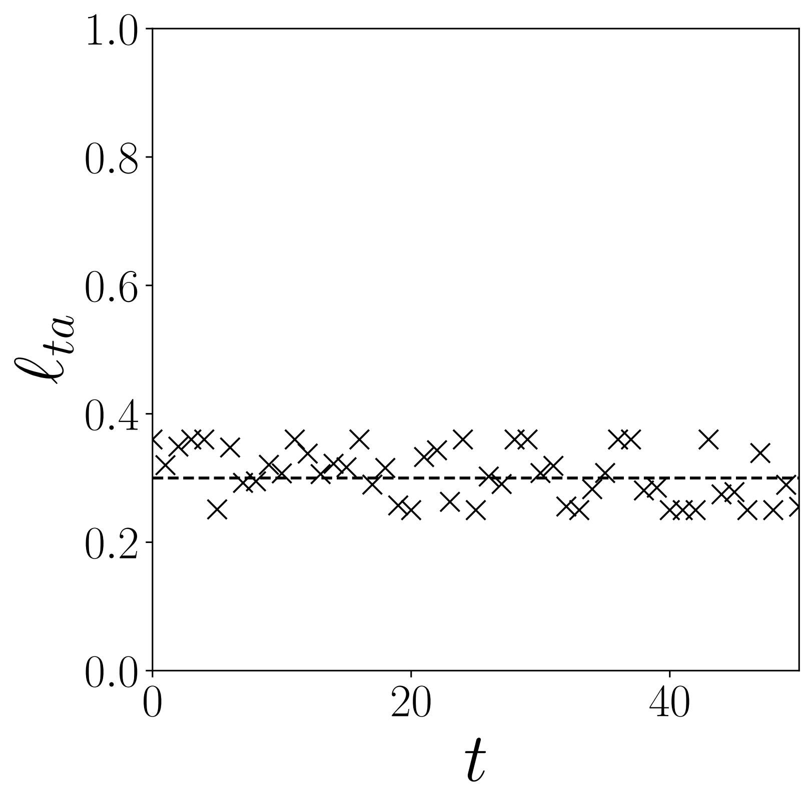
Neural Information Processing Systems (NeurIPS 2023)

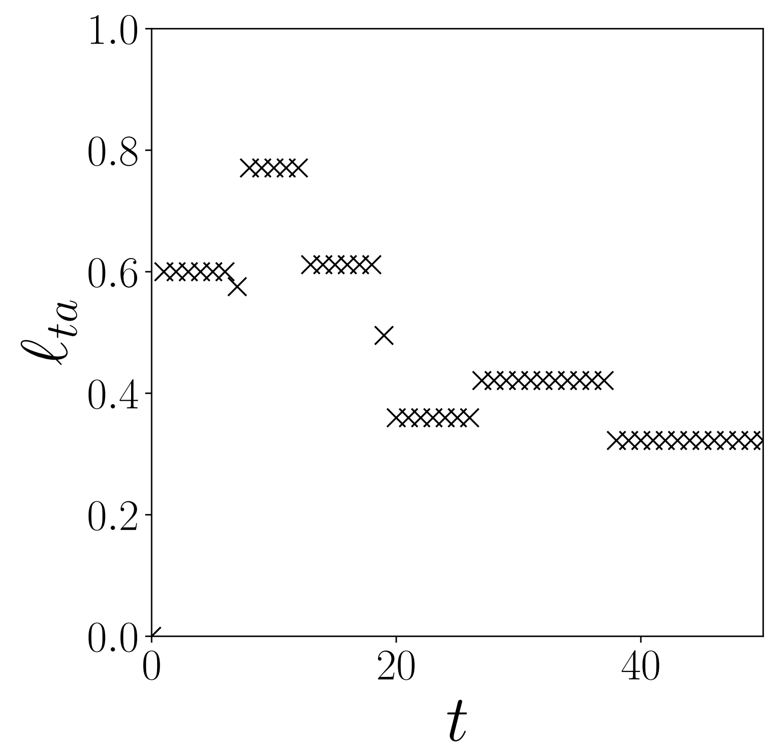
Environment adaptivity in online learning and bandits

Consider regret minimization for given T rounds

- **Data-dependent bounds** in adversarial environments [Allenberg-Auer-Györfi-Ottucsák 2006]
  - Regret bounds exploiting the property of the underlying environment
  - e.g., First-order / second-order / path-length bounds







- **Best-of-both-worlds** [Bubeck & Slivkins 2012]
  - Knowing if the environment is stochastic or adversarial in advance is challenging
  - Aiming to achieve optimality in both stochastic and adversarial environments simultaneously e.g., $O(\log T)$ in stochastic environments and $O(\sqrt{T})$ in adversarial environments for T rounds

C. Allenberg, P. Auer, L. Györfi, and G. Ottucsák. Hannan consistency in on-line learning in case of unbounded losses under partial monitoring. In ALT 2006.
S. Bubeck and A. Slivkins. The best of both worlds: Stochastic and adversarial bandits. In COLT 2012.

Can we make FTRL more adaptive?

- **Follow-the-regularized-leader (FTRL)** can achieve these environment adaptivities
- For FTRL with the Shannon entropy regularizer with learning rate $(\eta_t)_{t=1}^T$, a main part of the regret is bounded by $\mathbb{E}\left[\widehat{\mathrm{Reg}}_T^{\mathrm{SP}}\right]$ for

$$\widehat{\mathrm{Reg}}_T^{\mathrm{SP}} = \sum_{t=1}^{T} \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right) \underbrace{h_{t+1}}_{\text{penalty}} + \sum_{t=1}^{T} \eta_t \underbrace{z_t}_{\text{stability}}$$

- Existing adaptive learning rates $(\eta_t)_{t=1}^T$ in FTRL depend on either the (empirical) penalty or stability terms
  - With **empirical** stability $(z_s)_{s=1}^{t-1}$ and **worst-case** penalty terms $h_{\max} \geq \max_{t \in [T]} h_t$, we get **data-dependent bounds** [McMahan 2011; Lattimore & Szepesvári 2020, and so many!]
  - With **empirical** penalty $(h_s)_{s=1}^{t-1}$ and **worst-case** stability $\bar{z} \geq \max_{t \in [T]} z_t$, we get **best-of-both-worlds** [Ito-Tsuchiya-Honda 2022, Tsuchiya-Ito-Honda 2023]

Q. Can we construct learning rates jointly dependent on the **empirical** stability and penalty?

Stability-penalty-adaptive (SPA) learning rate

Definition (informal)

A sequence of learning rates $(\eta_t)_{t=1}^{T}$ is *stability-penalty-adaptive (SPA) learning rate* if the update is written with a certain non-negative reals $((h_t, z_t, \bar{z}_t))_{t=1}^{T}$ as follows:

$$\beta_t = \frac{1}{\eta_t}, \quad \beta_1 > 0\,, \quad \beta_{t+1} = \beta_t + \frac{c_1 z_t}{\sqrt{c_2 + \bar{z} h_1 + \sum_{s=1}^{t-1} \boxed{z_s h_{s+1}}}}$$

update jointly dependent on stability z_s & penalty h_{s+1}

Theorem (informal)

Let $(\eta_t)_{t=1}^{T}$ be a SPA learning rate. Then under a certain condition on $((h_t, z_t, \bar{z}_t))_{t=1}^{T}$,

$$\widehat{\mathrm{Reg}}_T^{\mathrm{SP}} = \tilde{O}\left(\sqrt{c_2 + \bar{z}_t h_1 + \boxed{\sum_{t=1}^{T} z_t h_{t+1}}}\right)$$

regret bound jointly dependent on stability z_s & penalty h_{s+1}

Q. Can we simultaneously achieve BOBW and data-dependent bounds?
→ check in **multi-armed bandits** and **partial monitoring**

1. Sparsity and BOBW in multi-armed bandits

- Sparsity level of losses $\ell_1, \ldots, \ell_T \in [0,1]^k$ is defined as $s = \max_{t \in [T]} \|\ell_t\|_0 \leq k$
- Sparsity-dependent bounds: data-dependent bounds considering the sparsity level $s \ll k$
  - Lower bound: $\Omega(\sqrt{sT})$, Upper bound: $\tilde{O}(\sqrt{sT})$ [Kwon & Perchet 2016, Bubeck-Cohen-Li 2018]
- Appropriately setting the stability and penalty terms in the **SPA learning rate** yields

with some important techniques

Theorem (informal)

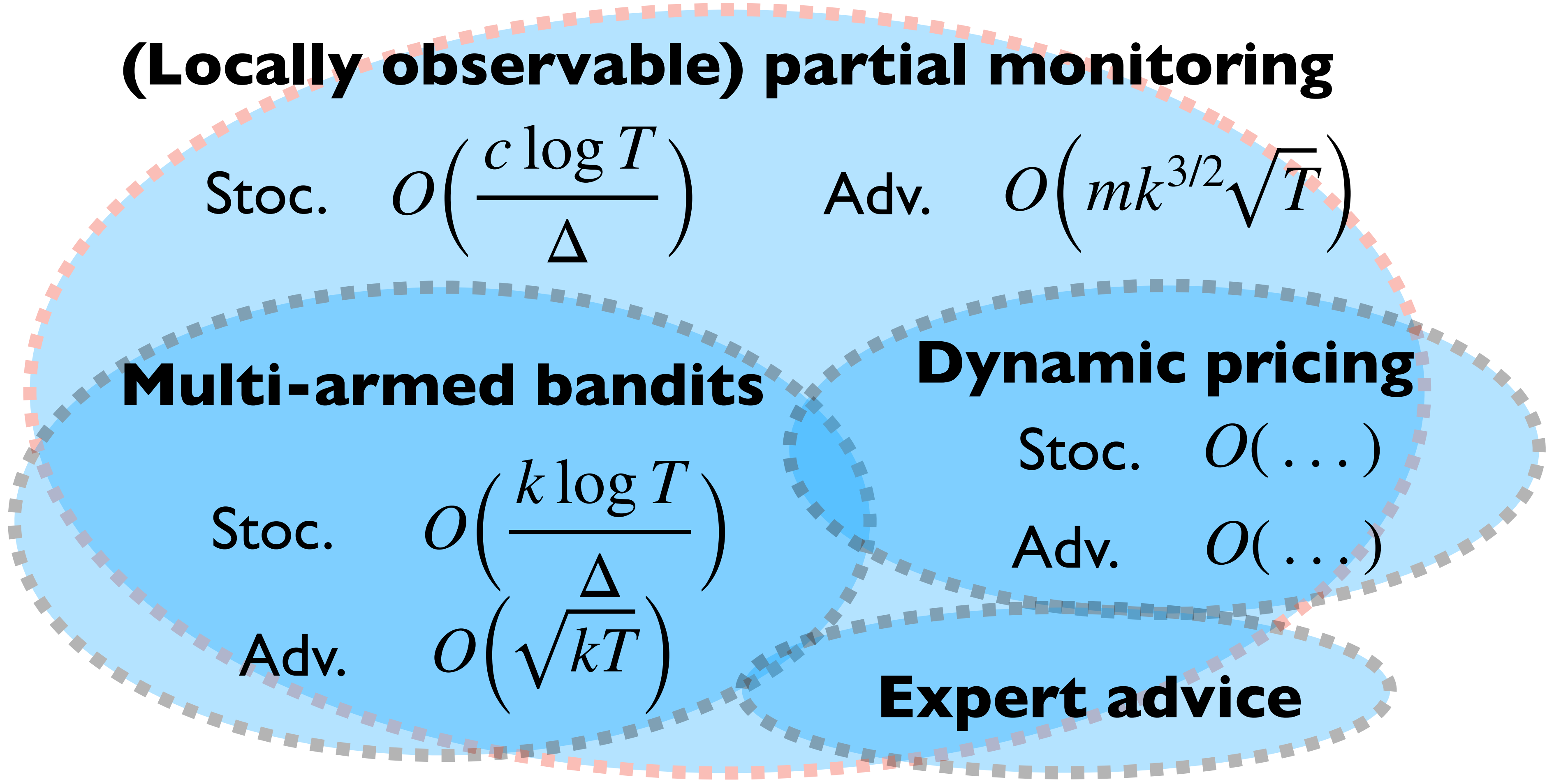
Corrupted Stochastic Env. $$R_T = O\left(\frac{s\log(T)\log(kT)}{\Delta_{\min}} + \sqrt{\frac{Cs\log(T)\log(kT)}{\Delta_{\min}}}\right)$$ best-of-both-worlds

Adversarial Env. $R_T = O(\sqrt{sT\log(k)\,\log(T)})$ sparsity-dependent bound

J. Kwon and V. Perchet. Gains and losses are fundamentally different in regret minimization: The sparse case. JMLR, 2016.
S. Bubeck, M. Cohen, and Y. Li. Sparsity, variance and curvature in multi-armed bandits. In ALT 2018.

2. Game-dependency and BOBW in partial monitoring

Hierarchical structure of problem classes



(Locally observable) partial monitoring

Stoc. $O\left(\frac{c\log T}{\Delta}\right)$ Adv. $O\left(mk^{3/2}\sqrt{T}\right)$

Multi-armed bandits

Stoc. $O\left(\frac{k\log T}{\Delta}\right)$

Adv. $O\left(\sqrt{kT}\right)$

Dynamic pricing

Stoc. $O(\dots)$

Adv. $O(\dots)$

Expert advice

Partial monitoring = very general online decision-making problems
Tend to be pessimistic

Desirable to automatically achieve regret that depends on
the inherent difficulty of the problem being solved
→ game-dependent bounds [Lattimore & Szepesvári 2020]

Theorem (informal) For locally observable partial monitoring games, by SPA learning rate,

Adversarial Env. $$R_T \le \mathbb{E}\left[\sqrt{\sum_{t=1}^{T} \boxed{V'_t}\log(k)\log(1+T)}\right] + o(\log T)$$

$V'_t, \bar{V}$: variables dependent on problem's inherent difficulty

Corrupted Stochastic Env. $$R_T = O\left(\frac{r_{\mathcal{M}}\boxed{\bar{V}}\log(T)\log(kT)}{\Delta_{\min}} + \sqrt{\frac{Cr_{\mathcal{M}}\boxed{\bar{V}}\log(T)\log(kT)}{\Delta_{\min}}}\right) + o(\log T)$$

T. Lattimore and Cs. Szepesvári. Exploration by optimisation in partial monitoring. In COLT 2020.

Summary

The main term of regret upper bound of FTRL

$$\widehat{\mathrm{Reg}}^{\mathrm{SP}}_T = \sum_{t=1}^{T} \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) \underset{\text{penalty}}{h_{t+1}} + \lambda \sum_{t=1}^{T} \underset{\text{stability}}{\eta_t z_t} \quad \text{for some} \quad \lambda > 0$$

Stability-penalty-adaptive learning rate

$$\beta_{t+1} = \beta_t + \frac{c_1 z_t}{\sqrt{c_2 + \bar{z} h_1 + \sum_{s=1}^{t-1} z_s h_{s+1}}}$$

Regret bound jointly dependent on stability and penalty

$$\widehat{\mathrm{Reg}}^{\mathrm{SP}}_T = \tilde{O}\left(\sqrt{c_2 + \bar{z}_t h_1 + \sum_{t=1}^{T} z_t h_{t+1}} \right)$$

1. Multi-armed bandits

Sparsity-dependent bound
and best-of-both-worlds guarantee

2. Partial monitoring

Game-dependent bound
and best-of-both-worlds guarantee