Data-Dependent Bounds for Online Portfolio Selection Without Lipschitzness and Smoothness

Chung-En Tsai, Ying-Ting Lin, Yen-Huan Li (National Taiwan University)

NEURAL INFORMATION PROCESSING SYSTEMS

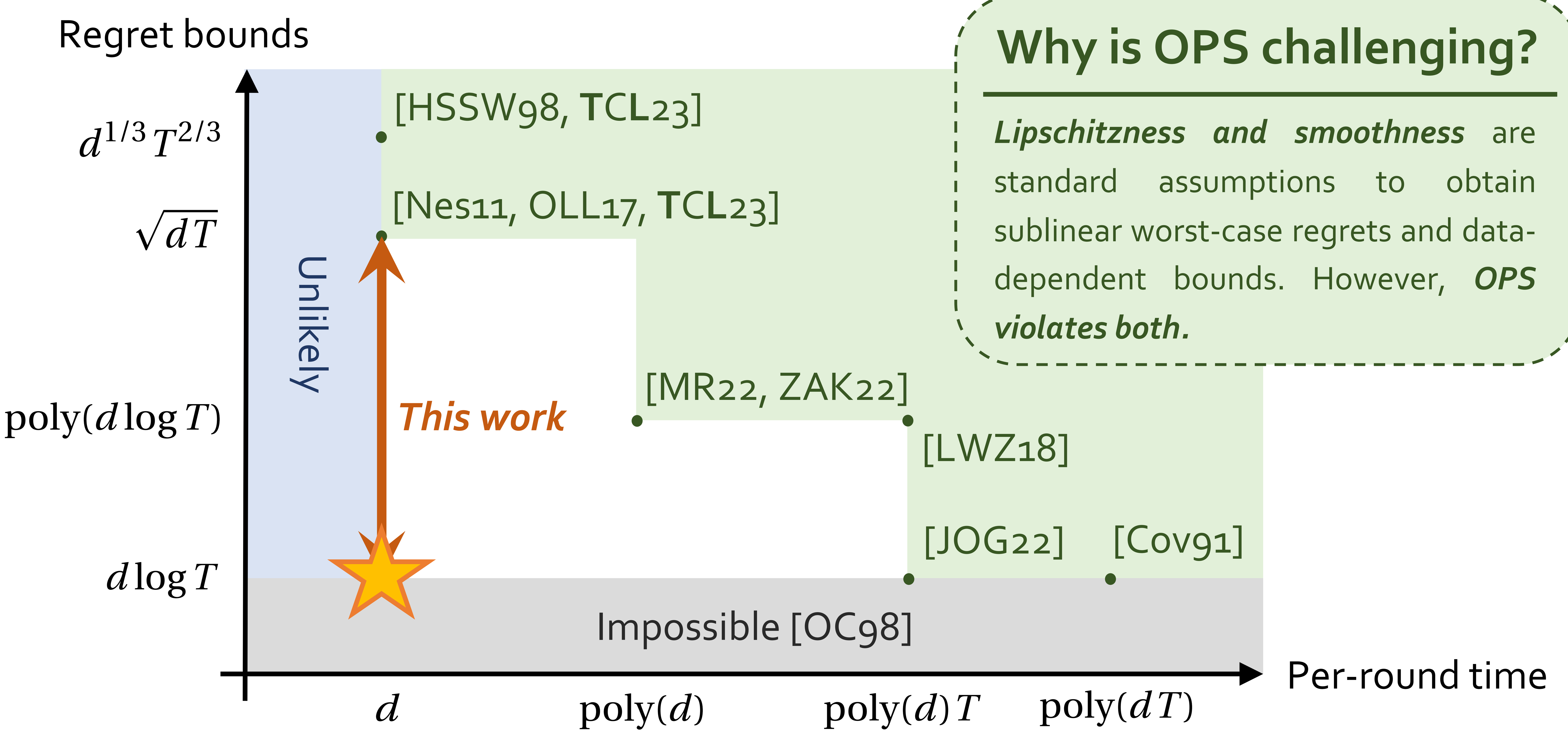
Contributions

- ***First data-dependent bounds for OPS.***
- ***First data-dependent bounds for non-Lipschitz, non-smooth losses.***
- Novel smoothness characterizations of log-loss.
- An implicitly defined optimistic LB-FTRL.
- A general analysis of Optimistic FTRL (OFTRL) with self-concordant regularizers.

Online Portfolio Selection (OPS)

- ***"The single most iconic online learning problem."***
  At the t-th round,
  1. Investor chooses a portfolio $x_t \in \Delta_d$;
  2. Market reveals a price relative $a_t \in [0,\infty)^d$;
  3. Investor suffers $f_t(x_t) := -\log\langle a_t, x_t\rangle$.
- **Regret:** $R_T(a_{1:T}) = \sum_{t=1}^{T} f_t(x_t) - \min_{x\in\Delta_d} \sum_{t=1}^{T} f_t(x)$.

The Holy Grail: Achievability of ★ ?



Why is OPS challenging?

Lipschitzness and smoothness are standard assumptions to obtain sublinear worst-case regrets and data-dependent bounds. However, ***OPS violates both.***

First Data-Dependent Bounds

There exist two algorithms satisfying

$$R_T(a_{1:T}) \le O\left(d\log^2 T + \sqrt{dL_T^\star}\log T\right)$$

$$R_T(a_{1:T}) \le O\left(d\log T + \sqrt{dV_T}\log T\right),$$

respectively, where

$$L_T^\star := \min_{x\in\Delta_d}\sum_{t=1}^{T} f_t(x) - \sum_{t=1}^{T}\min_{x\in\Delta_d} f_t(x)$$

$$V_T := \sum_{t=2}^{T}\max_{x\in\Delta_d}\|\nabla f_t(x) - \nabla f_{t-1}(x)\|_{x,*}^2.$$

Smoothness Characterizations

$$\|x\circ\nabla f_t(x) - y\circ\nabla f_t(y)\|_2 \le 4\|x-y\|_x$$

$$\|\nabla f_t(x) + \alpha_x(\nabla f_t(x))\mathbf{1}\|_{x,*}^2 \le 4\left(f_t(x) - \min_{y\in\Delta_d} f_t(y)\right).$$

Implicitly Defined LB-OFTRL

Let $h(x) := \sum_{i=1}^{d} -\log x(i)$ be the log-barrier.
Choose $p_{t+1} \in -\Delta_d$ and solve $(x_{t+1}, \hat{g}_{t+1})$:

$$\begin{cases} x_{t+1} \in \operatorname*{arg\,min}_{x\in\Delta_d} \sum_{\tau=1}^{t}\langle\nabla f_\tau(x_\tau), x\rangle + \langle\hat{g}_{t+1}, x\rangle + \frac{1}{\eta_t}h(x) \\ x_{t+1}\circ\hat{g}_{t+1} = p_{t+1}. \end{cases}$$