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

Chung-En Tsai, Ying-Ting Lin, and Yen-Huan Li

Department of Computer Science and Information Engineering, National Taiwan University

arXiv: 2305.13946

Background

1. OPS is a model of long-term investment ***without any statistical assumptions***.
2. Designing efficient and worst-case-optimal algorithms for Online Portfolio Selection (OPS) is a three-decade-old open problem.
3. Existing regret bounds for OPS are all worst-case. They do not exploit patterns of the data.

Online Portfolio Selection

On the t-th day,

1. Investor chooses a portfolio $x_t \in \Delta \subseteq \mathbb{R}^d$.
2. Then, Market reveals a price relatives $r_t \in \mathbb{R}^d_+$.
3. Then, Investor suffers a loss $f_t(x_t) := -\log\langle r_t, x_t\rangle$.

- Regret: $R_T := \sum_{t=1}^T f_t(x_t) - \min_{x\in\Delta} \sum_{t=1}^T f_t(x)$
- ***No assumptions on the behavior of Market!***

Contributions

1. ***First*** data-dependent bounds for OPS.
2. ***First*** data-dependent bounds with non-Lipschitz and non-smooth losses.
3. General analysis of Optimistic Follow-the-Regularized-Leader (FTRL) with self-concordant regularizers.
4. Two novel ***smoothness characterizations*** of the log-loss.

Challenges

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

Prior Work

Algorithms	Best-case	Worst-case	Per-round Time
(1991) UPS	×	$O(d\log T)$	$O(d^4 T^{14})$
(2011-) LB-FTRL, Soft-Bayes, LB-OMD	×	$\tilde{O}(\sqrt{dT})$	$\tilde{O}(d)$
(2018) ADA-BARRONS	×	$O(d^2 \mathrm{polylog}(T))$	$O(d^{2.5} T)$
(2022) PAE+DONS, BISONS	×	$O(d^2 \mathrm{polylog}(T))$	$O(\mathrm{poly}(d))$
(2022) VB-FTRL	×	$\tilde{O}(d\log T)$	$O(d^2 T)$
(2023) This work	$O(d\log T)$	$\tilde{O}(\sqrt{dT})$	$\tilde{O}(d)$

Main Results

There exist two algorithms that achieve

$$R_T = O(\sqrt{dL_T}\log T + d\log^2 T)$$
$$R_T = O(\sqrt{dV_T}\log T + d\log T)$$

respectively, where

$$L_T := \min_{x\in\Delta} \sum_{t=1}^T (f_t(x) - f_t^\star)$$
$$V_T := \sum_{t=2}^T \|\nabla f_{t-1}(x_{t-1}) - \nabla f_t(x_{t-1})\|^2_{x_{t-1},*}$$

Base Algorithm

Implicit Optimistic LB-FTRL

Input: a sequence of learning rates $\{\eta_t\}_t$.

Initialize $x_1 = e/d$, $\hat{g}_1 = \mathbf{0}$. Let $h(x) := -\sum_{i=1}^d \log x_i$.

For $t = 1, 2, \dots, T$,

1. Announce x_t and receive r_t.
2. Choose a vector p_{t+1}.
3. Compute $(\hat{g}_{t+1}, x_{t+1})$ such that

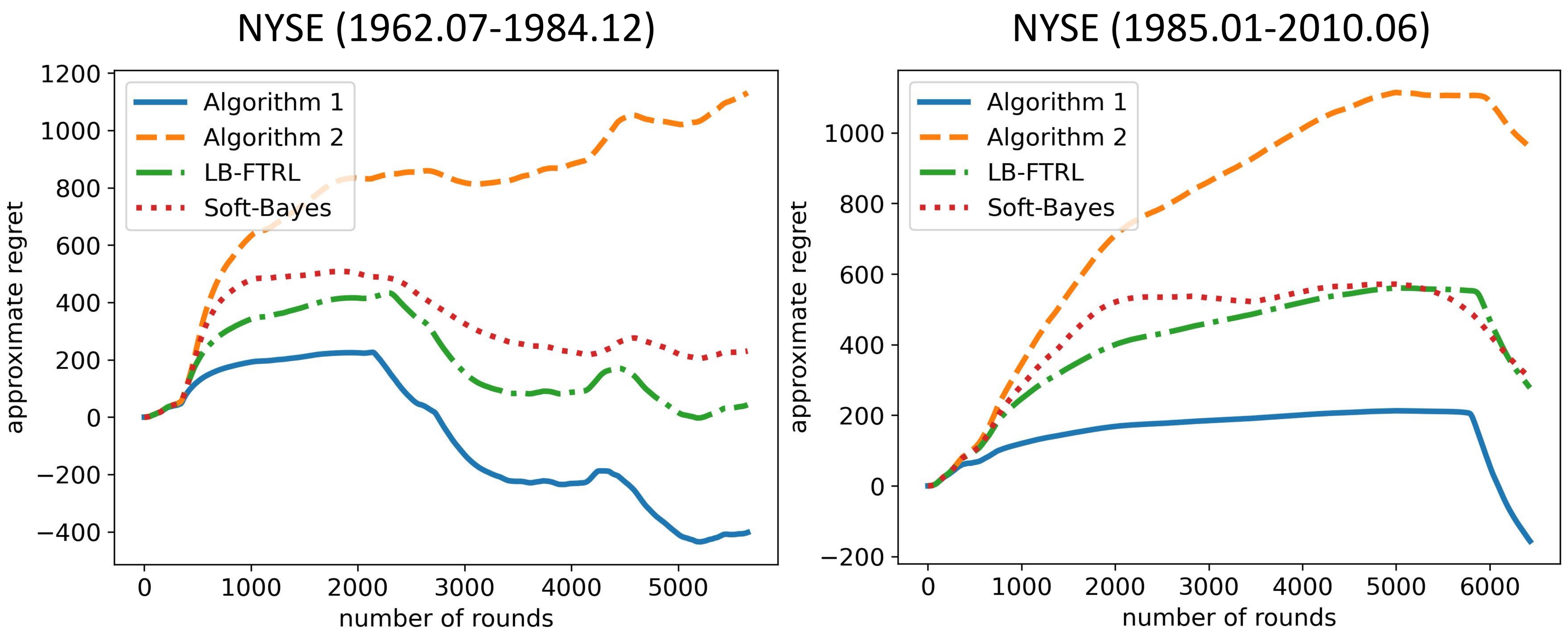
$$\begin{cases} \hat{g}_{t+1} = p_{t+1} \oslash x_{t+1} \\ x_{t+1} \in \operatorname{arg\,min}_{x\in\Delta} \sum_{\tau=1}^t \langle \nabla f_\tau(x_\tau), x\rangle + \langle \hat{g}_{t+1}, x\rangle + \eta_t^{-1} h(x) \end{cases}$$

Key Ideas

- The loss functions are "Lipschitz" under the local norm associated with h.
- The loss functions are "locally smooth" and smooth relative to h.
- Two smoothness characterizations:

$$\|x \odot \nabla f(x) - y \odot \nabla f(y)\|_2 \leq 4\|x - y\|_x$$
$$\|\nabla f(x) + \alpha(x) e\|^2_{x,*} \leq 4(f(x) - \min_{y\in\Delta} f(y))$$

Numerical Experiments



Research Directions

1. Better data-dependent bounds.
2. Generalization of our analyses.

國立臺灣大學資訊工程學系