* UNIVERSITY OF OXFORD † Google Research

Non-stationary Bandit Convex Optimization: A Comprehensive Study

Xiaoqi Liu* Dorian Baudry* Julian Zimmert† Patrick Rebeschini* Arya Akhavan*

Non-stationary Bandit Convex Optimization

- Adversary fixes convex loss functions $f_1, f_2, \ldots, f_T : \mathbb{R}^d \to [-1, 1]$
- For $t \geq 1$, learner
  - Selects action $\boldsymbol{z}_t$ from continuous (convex & compact) arm set $\Theta \subseteq \mathbb{R}^d$
  - Incurs loss $f_t(\boldsymbol{z}_t)$, observes *bandit feedback* with sub-Gaussian noise ξ_t

$$y_t = f_t(\boldsymbol{z}_t) + \xi_t$$

- *Regret* as benchmark with comparators $\boldsymbol{u}_{1:T} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T\}$:

$$R(T, \boldsymbol{u}_{1:T}) := \sum_{t=1}^{T} \mathbb{E}\left[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u}_t)\right]$$

Non-stationarity measures:

- Number of switches: $S(\boldsymbol{u}_{1:T}) := 1 + \sum_{t=2}^{T} \mathbf{1}\{\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}\} \leq S$
- Path-length: $P(\boldsymbol{u}_{1:T}) := \sum_{t=2}^{T} \|\boldsymbol{u}_t - \boldsymbol{u}_{t-1}\| \leq P$
- Total variation: $\Delta(f_{1:T}) := \sum_{t=2}^{T} \max_{\boldsymbol{z} \in \Theta} |f_t(\boldsymbol{z}) - f_{t-1}(\boldsymbol{z})| \leq \Delta$

Regret notions: switching, path-length, and dynamic regret:

$$R^{\text{swi}}(T, S) := \max_{\boldsymbol{u}_{1:T} : S(\boldsymbol{u}_{1:T}) \leq S} R(T, \boldsymbol{u}_{1:T}), \quad R^{\text{path}}(T, P) := \max_{\boldsymbol{u}_{1:T} : P(\boldsymbol{u}_{1:T}) \leq P} R(T, \boldsymbol{u}_{1:T}),$$

$$R^{\text{dyn}}(T, \Delta) := \sup_{f_{1:T} : \Delta(f_{1:T}) \leq \Delta} \sum_{t=1}^{T} \mathbb{E}\left[f_t(\boldsymbol{z}_t) - \min_{\boldsymbol{z} \in \Theta} f_t(\boldsymbol{z})\right].$$

Our goals

- Unified treatment for non-stationary BCO (previous work: only $R^{\text{dyn}}(T, \Delta)$ [1, 2] or only $R^{\text{path}}(T, P)$ [3, 4])
- Design algorithms with optimal sublinear regret w.r.t. T, S, Δ and P

Main results

Regret bounds for $R^{\text{swi}}(T, S)$, $R^{\text{dyn}}(T, \Delta)$ and $R^{\text{path}}(T, P)$, respectively, for algorithms *tuned with known S, Δ and P*:

Algo.:	**TEWA-SE**	**cExO**
General convex (GC)	$\sqrt{d} S^{\frac{1}{4}} T^{\frac{3}{4}}$, $d^{\frac{2}{5}} \Delta^{\frac{1}{5}} T^{\frac{4}{5}}$, $d^{\frac{2}{5}} P^{\frac{1}{5}} T^{\frac{4}{5}}$	$\underline{d^{\frac{5}{2}}\sqrt{ST}}$, $d^{\frac{5}{3}}\underline{\Delta^{\frac{1}{3}} T^{\frac{2}{3}}}$, $d^{\frac{5}{3}}\, \underset{\sim}{P^{\frac{1}{3}} T^{\frac{2}{3}}}$
Strongly convex (SC)	$\underline{d\sqrt{ST}}$, $\underline{d^{\frac{2}{3}} \Delta^{\frac{1}{3}} T^{\frac{2}{3}}}$, $\underset{\sim}{d^{\frac{2}{3}} P^{\frac{1}{3}} T^{\frac{2}{3}}}$	
Comp. complexity	polynomial	exponential

- Straight underline: minimax-optimal rates.
- Wavy underline: result is either new to the literature (SC case) or improves on the best-known $P^{\frac{1}{4}} T^{\frac{3}{4}}$ rate [3] (GC case).

Algorithm 1: TEWA-SE

Tilted Exponentially Weighted Average with Sleeping Experts

Input: perturbation step-size $h = \sqrt{d}\mathrm{B}^{-\frac{1}{4}}$, expert algorithm $E(I, \eta)$ is *online gradient descent* over interval I with step-size parameterized by η

for $t = 1, 2, \ldots, T$ **do**
  for Active expert $E_i \equiv E_i(I_i, \eta_i) \in \{E_1, E_2, \ldots, E_{n_t}\}$ **do**
    Receive action $\boldsymbol{x}_{t,I_i}^{\eta_i}$ from expert E_i
  end for
  Set meta-action using TEWA:

$$\boldsymbol{x}_t = \sum_{i=1}^{n_t} \frac{\eta_i e^{-L_{t-1,I_i}^{\eta_i}}}{\sum_{j=1}^{n_t} \eta_j e^{-L_{t-1,I_j}^{\eta_j}}} \boldsymbol{x}_{t,I_i}^{\eta_i} \qquad \triangleright \text{aggregation}$$

  Sample $\boldsymbol{\zeta}_t$ uniformly from unit sphere $\partial\mathbb{B}^d$
  Query point $\boldsymbol{z}_t = \boldsymbol{x}_t + h\boldsymbol{\zeta}_t$ to obtain $y_t = f_t(\boldsymbol{z}_t) + \xi_t$
  Construct gradient estimate $\boldsymbol{g}_t = (d/h) y_t \boldsymbol{\zeta}_t$
  for $i = 1, 2, \ldots, n_t$ **do**
    Send meta-action $\boldsymbol{x}_t$ and $\boldsymbol{g}_t$ to E_i
    Increment loss $L_{t,I_i}^{\eta_i} = L_{t-1,I_i}^{\eta_i} + \ell_t^{\eta_i}(\boldsymbol{x}_{t,I_i}^{\eta_i})$ $\triangleright$ update experts
  end for
end for

Our contribution: Construct SC surrogate losses with one-point gradient estimates:

$\because$ Simple linear surrogate loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\boldsymbol{x}_t - \boldsymbol{x}) \Rightarrow \sqrt{|I|}$ expert static regret $\Rightarrow$linear $R^{\text{ada}}(\mathrm{B}, T)$.
$\therefore$ We instead use the SC surrogate loss

$$\ell_t^{\eta}(\boldsymbol{x}) = -\eta \boldsymbol{g}_t^\top(\boldsymbol{x}_t - \boldsymbol{x}) + \eta^2 G^2 \|\boldsymbol{x}_t - \boldsymbol{x}\|^2, \quad \forall \boldsymbol{x} \in \mathbb{R}^d$$

where w.h.p. $\|\boldsymbol{g}_t\| \leq G$ $\forall t \in [T]$ $\Rightarrow$ $\log|I|$ expert static regret $\Rightarrow$ sublinear $R^{\text{ada}}(\mathrm{B}, T)$.

Gradient estimate $\boldsymbol{g}_t$ satisfies $\mathbb{E}[\boldsymbol{g}_t | \boldsymbol{x}_t] = \nabla \hat{f}_t(\boldsymbol{x}_t)$ where $\hat{f}_t(\boldsymbol{x}) = \mathbb{E}\left[f_t(\boldsymbol{x} + h\tilde{\boldsymbol{\zeta}})\right]$ is smoothed loss ($\tilde{\boldsymbol{\zeta}}$ uniform on unit ball $\mathbb{B}^d$)
Handle bias-variance tradeoff in regret upper bound by tuning h

Tools from prior work:

- Sleeping experts on geometric intervals with geometric step-size
- TEWA aggregation to adapt to unknown loss curvature [5–7]

Theorem: For known B, $R^{\text{ada}}(\mathrm{B}, T) \lesssim \begin{cases} \sqrt{d}\mathrm{B}^{\frac{3}{4}} & \text{(GC)} \\ \frac{d}{\alpha}\sqrt{\mathrm{B}} & (SC) \end{cases}$

Minimax-optimal lower bounds

$d\sqrt{ST}$	$d^{\frac{2}{3}} \Delta^{\frac{1}{3}} T^{\frac{2}{3}}$	$d^{\frac{4}{5}} P^{\frac{2}{5}} T^{\frac{3}{5}}$

- For $d = 1$, rates w.r.t. T, S, Δ match those for multi-armed bandits
- Path-length bound improves on the only existing $d\sqrt{PT}$ from [3]

Algorithm 2: cExO clipped Exploration by Optimization

Vanilla ExO from [8] + clipping.
Input: a finite covering set $\mathcal{C} \subset \Theta$ of Θ, and $\tilde{\Delta} = \Delta(\mathcal{C}) \cap [\gamma, 1]^{|\mathcal{C}|}$

for $t = 1, \ldots, T$ **do**
  Compute *reference dist.*:

$$\boldsymbol{q}_t = \Pi_{\tilde{\Delta}}(\tilde{\boldsymbol{q}}_t), \quad \tilde{q}_t(\boldsymbol{x}) = \frac{e^{-\eta L_{t-1}(\boldsymbol{x})}}{\sum_{\boldsymbol{x}' \in \mathcal{C}} e^{-\eta L_{t-1}(\boldsymbol{x}')}} \quad \forall \boldsymbol{x} \in \mathcal{C}.$$

  Select *sampling dist.* $\boldsymbol{p}_t \in \Delta(\mathcal{C})$ & *loss estimator* E_t by solving

$$\underset{\substack{\boldsymbol{p} \in \Delta(\mathcal{C}), \\ E : \mathcal{C} \times [-1,1] \to \mathbb{R}^{|\mathcal{C}|}}}{\arg\min} \Lambda(\boldsymbol{q}_t, \boldsymbol{p}, E) \qquad \triangleright \text{ExO step}$$

  Sample $\boldsymbol{z}_t \sim \boldsymbol{p}_t$, observe $f_t(\boldsymbol{z}_t)$
  Set $\boldsymbol{\ell}_t = E_t(\boldsymbol{z}_t, f_t(\boldsymbol{z}_t))$, $L_t(\boldsymbol{x}) = L_{t-1}(\boldsymbol{x}) + \ell_t(\boldsymbol{x})$ $\forall \boldsymbol{x} \in \mathcal{C}$.
end for

$$\Lambda(\boldsymbol{q}_t, \boldsymbol{p}, E) = \sup_{\boldsymbol{p}^*, f} \mathbb{E}_{\boldsymbol{z} \sim \boldsymbol{p}}\Big[\underbrace{\overbrace{\langle \boldsymbol{p} - \boldsymbol{p}^*, f \rangle}^{\text{true loss}} - \overbrace{\langle \boldsymbol{q} - \boldsymbol{p}^*, E(\boldsymbol{z}, f(\boldsymbol{z})) \rangle}^{\text{surrogate loss}}}_{\text{bias}} + \underbrace{\frac{1}{\eta} V_{\boldsymbol{q}_t}(\eta E(\boldsymbol{z}, f(\boldsymbol{z})))}_{\text{variance}}\Big]$$

Solving $\arg\min_{\boldsymbol{p}, E} \Lambda(\boldsymbol{q}_t, \boldsymbol{p}, E)$ in ExO step is statistically-sharp but exponential in computational complexity

Theorem: For known B, $R^{\text{ada}}(\mathrm{B}, T) \lesssim d^{\frac{5}{2}}\sqrt{\mathrm{B}}$ (GC)

Side result: Conversions between regrets

Define adaptive regret: for $\mathrm{B} \in [T]$,

$$R^{\text{ada}}(\mathrm{B}, T) := \max_{\substack{p, q \in [T], \\ 0 < q - p \leq \mathrm{B}}} \max_{\boldsymbol{u} \in \Theta} \sum_{t=p}^{q} \mathbb{E}\left[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u})\right].$$

$$\begin{tikzcd} & & R^{\text{dyn}}(T, \Delta) \\ R^{\text{ada}}(\mathrm{B}, T) \arrow[r] & R^{\text{swi}}(T, S) \arrow[ur] \arrow[dr, "\text{new proof}"'] & \\ & & R^{\text{path}}(T, P) \end{tikzcd}$$

Legend: $R_1 \longrightarrow R_2$ means that if regret R_1 is sublinear in T (or B), then regret R_2 is also sublinear in T, by tuning B based on S, Δ, P.

References

[1] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015.
[2] Yining Wang. On adaptivity in nonstationary stochastic optimization with bandit feedback. *Operations Research*, 73(2):819–828, 2025.
[3] Peng Zhao, Guanghui Wang, Lijun Zhang, and Zhi-Hua Zhou. Bandit convex optimization in non-stationary environments. *Journal of Machine Learning Research*, 22(125):1–45, 2021.
[4] Tianyi Chen and Georgios B Giannakis. Bandit convex optimization for scalable and dynamic IoT management. *IEEE Internet of Things Journal*, 6(1):1276–1286, 2018.
[5] Lijun Zhang, Guanghui Wang, Wei-Wei Tu, Wei Jiang, and Zhi-Hua Zhou. Dual adaptivity: a universal algorithm for minimizing the adaptive regret of convex functions. In *International Conference on Neural Information Processing Systems*, 2021.
[6] Guanghui Wang, Shiyin Lu, and Lijun Zhang. Adaptivity and optimality: A universal algorithm for online convex optimization. In *Proceedings of Uncertainty in Artificial Intelligence Conference*, volume 115, pages 659–668. PMLR, 2020.
[7] Tim van Erven, Wouter M. Koolen, and Dirk van der Hoeven. Metagrad: Adaptation using multiple learning rates in online learning. *Journal of Machine Learning Research*, 22(161):1–61, 2021.
[8] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.