Non-Stationary Bandit Convex Optimization: Algorithms and Links to Coin Betting

Xiaoqi (Shirley) Liu, Dorian Baudry, Julian Zimmert,
Patrick Rebeschini, Arya Akhavan

Algorithmic Statistics Workshop, University of Oxford

28 November 2025

Bandit Convex Optimization (BCO)

Bandit Convex Optimization (BCO)

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

$$f_t(\boldsymbol{x}_t) + \xi_t$$

Bandit Convex Optimization (BCO)

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

$$f_t(\boldsymbol{x}_t) + \xi_t$$

- Regret against comparators $\boldsymbol{u}_{1:T} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_T\}$ chosen by adversary:

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

Bandit Convex Optimization (BCO)

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

$$f_t(\boldsymbol{x}_t) + \xi_t$$

- Regret against comparators $\boldsymbol{u}_{1:T} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T\}$ chosen by adversary:

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

  - Static regret:

$$\max_{\boldsymbol{u}_{1:T}:\boldsymbol{u}_1=\cdots=\boldsymbol{u}_T} R(T, \boldsymbol{u}_{1:T}) = \mathbb{E}\left[\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \min_{\boldsymbol{u}\in\Theta} \sum_{t=1}^{T} f_t(\boldsymbol{u})\right]$$

Bandit Convex Optimization (BCO)

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

$$f_t(\boldsymbol{x}_t) + \xi_t$$

- Regret against comparators $\boldsymbol{u}_{1:T} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T\}$ chosen by adversary:

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

  - Static regret:

$$\max_{\boldsymbol{u}_{1:T}:\boldsymbol{u}_1=\cdots=\boldsymbol{u}_T} R(T, \boldsymbol{u}_{1:T}) = \mathbb{E}\left[\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \min_{\boldsymbol{u}\in\Theta} \sum_{t=1}^{T} f_t(\boldsymbol{u})\right]$$

  - Non-stationary regrets?

Non-stationary regrets

Non-stationary regrets

Focus on switching regret:

$$R^{\mathrm{swi}}(T,S) = \max_{\boldsymbol{u}_{1:T}:S(\boldsymbol{u}_{1:T})\leq S} R(T,\boldsymbol{u}_{1:T}), \quad \text{where} \quad S(\boldsymbol{u}_{1:T}) := 1 + \sum_{t=2}^{T} \mathbf{1}\{\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}\}$$

Non-stationary regrets

Focus on switching regret:

$$R^{\mathrm{swi}}(T,S) = \max_{\boldsymbol{u}_{1:T} : S(\boldsymbol{u}_{1:T}) \le S} R(T, \boldsymbol{u}_{1:T}), \quad \text{where} \quad S(\boldsymbol{u}_{1:T}) := 1 + \sum_{t=2}^{T} \mathbf{1}\{\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}\}$$

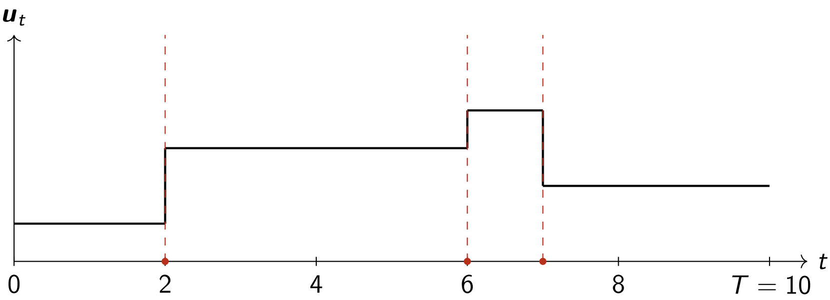
Example:

Non-stationary regrets

Focus on switching regret:

$$R^{\mathrm{swi}}(T,S) = \max_{\boldsymbol{u}_{1:T} : S(\boldsymbol{u}_{1:T}) \le S} R(T, \boldsymbol{u}_{1:T}), \quad \text{where} \quad S(\boldsymbol{u}_{1:T}) := 1 + \sum_{t=2}^{T} \mathbf{1}\{\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}\}$$

Example:

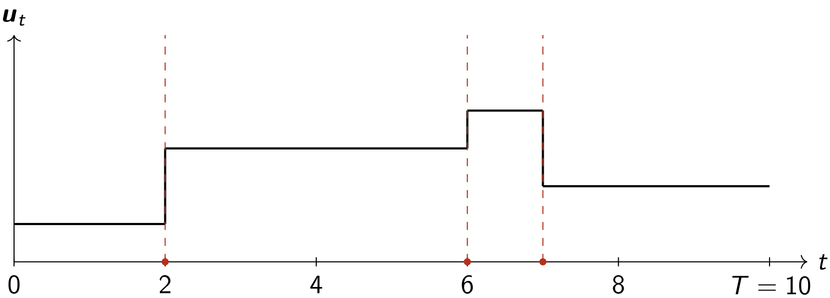


Non-stationary regrets

Focus on switching regret:

$$R^{\text{swi}}(T,S) = \max_{\boldsymbol{u}_{1:T}: S(\boldsymbol{u}_{1:T}) \le S} R(T, \boldsymbol{u}_{1:T}), \quad \text{where} \quad S(\boldsymbol{u}_{1:T}) := 1 + \sum_{t=2}^{T} \mathbf{1}\{\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}\}$$

Example:



Reduction to adaptive regret:

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

Main results & related work

Main results & related work

Main results & related work

- Unified treatment for non-stationary Bandit Convex Optimization (BCO); previous work handled only isolated aspects [Besbes et al., 2015, Wang, 2025, Zhao et al., 2021, Chen and Giannakis, 2018]

Main results & related work

- Unified treatment for non-stationary Bandit Convex Optimization (BCO); previous work handled only isolated aspects [Besbes et al., 2015, Wang, 2025, Zhao et al., 2021, Chen and Giannakis, 2018]
- Built on Online Convex Optimization (OCO) (observe $f_t(\cdot)$ or at least $\nabla f_t(\boldsymbol{x}_t)$ at t):

Main results & related work

- Unified treatment for non-stationary Bandit Convex Optimization (BCO); previous work handled only isolated aspects [Besbes et al., 2015, Wang, 2025, Zhao et al., 2021, Chen and Giannakis, 2018]
- Built on Online Convex Optimization (OCO) (observe $f_t(\cdot)$ or at least $\nabla f_t(\boldsymbol{x}_t)$ at t):
  - **Expert tracking methods** [Herbster and Warmuth, 1998, Littlestone and Warmuth, 1994, Cesa-Bianchi et al., 1997, Vovk, 1998, Freund et al., 1997]

Main results & related work

- Unified treatment for non-stationary Bandit Convex Optimization (BCO); previous work handled only isolated aspects [Besbes et al., 2015, Wang, 2025, Zhao et al., 2021, Chen and Giannakis, 2018]
- Built on Online Convex Optimization (OCO) (observe $f_t(\cdot)$ or at least $\nabla f_t(\boldsymbol{x}_t)$ at t):
  - **Expert tracking methods** [Herbster and Warmuth, 1998, Littlestone and Warmuth, 1994, Cesa-Bianchi et al., 1997, Vovk, 1998, Freund et al., 1997]
  - **Study of adaptive regret** [Hazan and Seshadhri, 2009, Daniely et al., 2015, Adamskiy et al., 2016, Jun et al., 2017, Cutkosky, 2020, Wang et al., 2018, Zhao et al., 2022, Zhang et al., 2021, Zhang et al., 2018a, Yang et al., 2024]

Main results & related work

- Unified treatment for non-stationary Bandit Convex Optimization (BCO); previous work handled only isolated aspects [Besbes et al., 2015, Wang, 2025, Zhao et al., 2021, Chen and Giannakis, 2018]
- Built on Online Convex Optimization (OCO) (observe $f_t(\cdot)$ or at least $\nabla f_t(\boldsymbol{x}_t)$ at t):
  - **Expert tracking methods** [Herbster and Warmuth, 1998, Littlestone and Warmuth, 1994, Cesa-Bianchi et al., 1997, Vovk, 1998, Freund et al., 1997]
  - **Study of adaptive regret** [Hazan and Seshadhri, 2009, Daniely et al., 2015, Adamskiy et al., 2016, Jun et al., 2017, Cutkosky, 2020, Wang et al., 2018, Zhao et al., 2022, Zhang et al., 2021, Zhang et al., 2018a, Yang et al., 2024]

Algorithm	**clipped Exploration by Optimization (cExO)**	**Tilted-EW Average with Sleeping Experts (TEWA-SE)**
General convex (GC)	$\underline{d^{\frac{5}{2}}\sqrt{ST}}$	$\sqrt{d}\, S^{\frac{1}{4}} T^{\frac{3}{4}}$
Strongly convex (SC)		$\underline{d\sqrt{ST}}$

Underline: minimax-optimal rates. [Liu et al., 2025]

Main results & related work

Algorithm	clipped Exploration by Optimization (cExO)	Tilted-EW Average with Sleeping Experts (TEWA-SE)
General convex (GC)	$d^{\frac{5}{2}}\underline{\sqrt{ST}}$	$\sqrt{d}\, S^{\frac{1}{4}}\, T^{\frac{3}{4}}$
Strongly convex (SC)		$\underline{d\sqrt{ST}}$

Underline: minimax-optimal rates. [Liu et al., 2025]

Main results & related work

ExO: choose surrogate loss and exploratory distribution by optimizing worst-case bias-variance tradeoff [Lattimore and Gyorgy, 2021]

Algorithm	**clipped Exploration by Optimization (cExO)**	**Tilted-EW Average with Sleeping Experts (TEWA-SE)**
General convex (GC)	$d^{\frac{5}{2}}\underline{\sqrt{ST}}$	$\sqrt{d}\, S^{\frac{1}{4}}\, T^{\frac{3}{4}}$
Strongly convex (SC)		$\underline{d\sqrt{ST}}$

Underline: minimax-optimal rates. [Liu et al., 2025]

Main results & related work

clipping: to enforce exploration

ExO: choose surrogate loss and exploratory distribution by optimizing worst-case bias-variance tradeoff [Lattimore and Gyorgy, 2021]

Algorithm	**clipped Exploration by Optimization (cExO)**	**Tilted-EW Average with Sleeping Experts (TEWA-SE)**
General convex (GC)	$d^{\frac{5}{2}}\underline{\sqrt{ST}}$	$\sqrt{d}\, S^{\frac{1}{4}} T^{\frac{3}{4}}$
Strongly convex (SC)	$d^{\frac{5}{2}}\underline{\sqrt{ST}}$	$\underline{d\sqrt{ST}}$

Underline: minimax-optimal rates. [Liu et al., 2025]

Main results & related work

ExO: choose surrogate loss and exploratory distribution by optimizing worst-case bias-variance tradeoff [Lattimore and Gyorgy, 2021]

clipping: to enforce exploration

Algorithm	**clipped Exploration by Optimization (cExO)**	**Tilted-EW Average with Sleeping Experts (TEWA-SE)**
General convex (GC) Strongly convex (SC)	$d^{\frac{5}{2}}\underline{\sqrt{ST}}$	$\sqrt{d}\,S^{\frac{1}{4}}\,T^{\frac{3}{4}}$ $\underline{d\sqrt{ST}}$
Features	• Conceptual design • Minimax-optimal in S, T • Not poly-time computable	

Underline: minimax-optimal rates. [Liu et al., 2025]

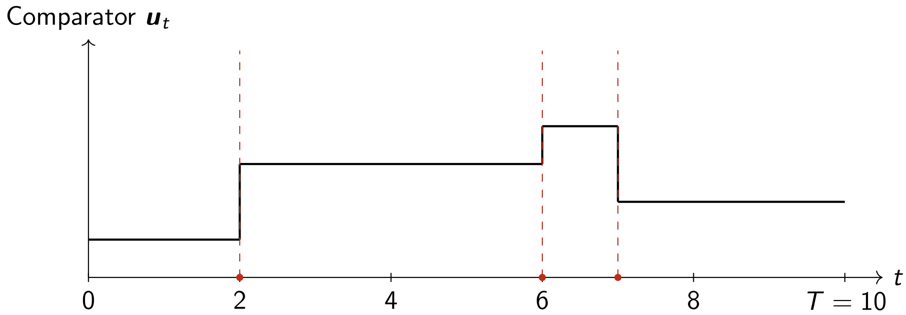
Main results & related work

clipping: to enforce exploration

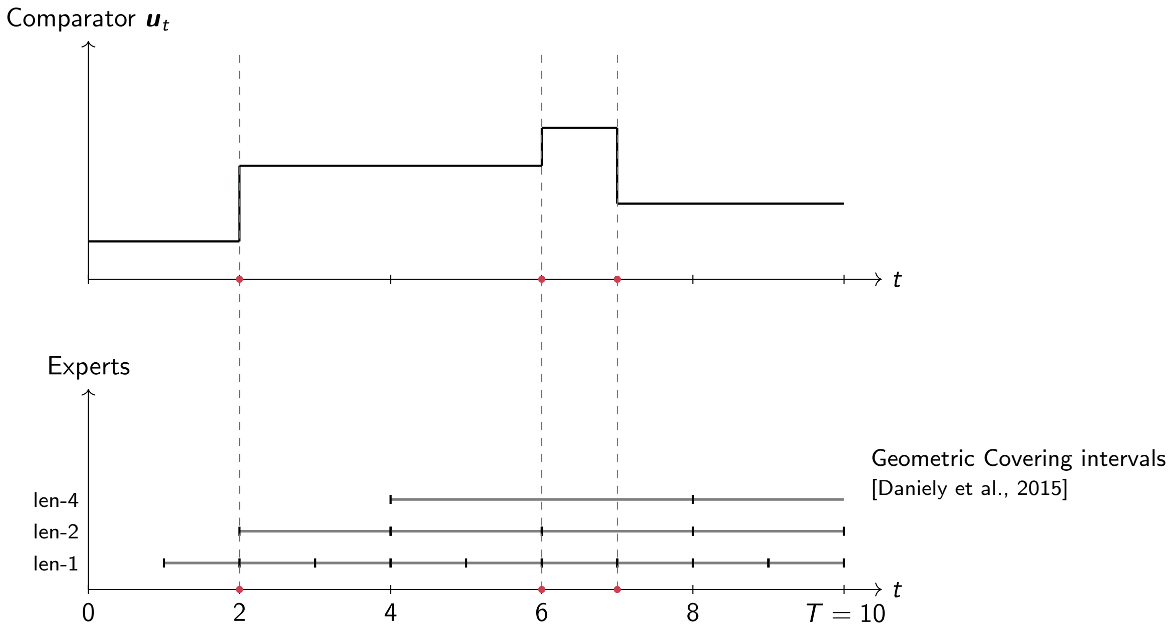
ExO: choose surrogate loss and exploratory distribution by optimizing worst-case bias-variance tradeoff [Lattimore and Gyorgy, 2021]

Algorithm	**clipped Exploration by Optimization (cExO)**	**Tilted-EW Average with Sleeping Experts (TEWA-SE)**
General convex (GC)	$d^{\frac{5}{2}}\underline{\sqrt{ST}}$	$\sqrt{d}\, S^{\frac{1}{4}}\, T^{\frac{3}{4}}$
Strongly convex (SC)		$\underline{d\sqrt{ST}}$
Features	• Conceptual design • Minimax-optimal in S, T • Not poly-time computable	• Instantiates design principles • Adaptive to curvature α • Poly-time

Underline: minimax-optimal rates. [Liu et al., 2025]



Comparator $\boldsymbol{u}_t$
t
0
2
4
6
8
$T = 10$


Comparator $\boldsymbol{u}_t$
t
Experts
Geometric Covering intervals
[Daniely et al., 2015]
len-4
len-2
len-1
t
0
2
4
6
8
$T = 10$

Sleeping experts (full-information)

Sleeping experts (full-information)

for $t = 1, 2, \ldots, T$ **do**

Let $\{E_1, \ldots, E_{n_t}\}$ be the active experts at t with $E_i \equiv E_i(I_i)$.

for $i = 1, 2, \ldots, n_t$ **do**

Expert E_i outputs action $\boldsymbol{x}_{t,I_i}$

end for

Play action using exponential-weights (EW):

$$\boldsymbol{x}_t = \sum_{i=1}^{n_t} \frac{e^{-L_{t-1,I_i}}}{\sum_{j=1}^{n_t} e^{-L_{t-1,I_j}}} \boldsymbol{x}_{t,I_i}$$

Observe entire loss function $f_t(\cdot)$

for $i = 1, 2, \ldots, n_t$ **do**

Send $f_t(\cdot)$ to E_i

Increment cumulative loss $L_{t,I_i} = L_{t-1,I_i} + f_t(\boldsymbol{x}_{t,I_i})$

end for

end for

[Daniely et al., 2015, Jun et al., 2017, Zhang et al., 2018b]

Sleeping experts (full-information)

for $t = 1, 2, \ldots, T$ **do**

Let $\{E_1, \ldots, E_{n_t}\}$ be the active experts at t with $E_i \equiv E_i(I_i)$.

for $i = 1, 2, \ldots, n_t$ **do**

Expert E_i outputs action $\boldsymbol{x}_{t,I_i}$

end for

Play action using exponential-weights (EW):

$$\boldsymbol{x}_t = \sum_{i=1}^{n_t} \frac{e^{-L_{t-1,I_i}}}{\sum_{j=1}^{n_t} e^{-L_{t-1,I_j}}} \boldsymbol{x}_{t,I_i}$$

Observe entire loss function $f_t(\cdot)$

for $i = 1, 2, \ldots, n_t$ **do**

Send $f_t(\cdot)$ to E_i

Increment cumulative loss $L_{t,I_i} = L_{t-1,I_i} + f_t(\boldsymbol{x}_{t,I_i})$

end for

end for

$$\sum_{t\in I} \left(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{u})\right) = \underbrace{\sum_{t\in I} \left(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_{t,I})\right)}_{\text{meta-regret}} + \underbrace{\sum_{t\in I} \left(f_t(\boldsymbol{x}_{t,I}) - f_t(\boldsymbol{u})\right)}_{\text{expert-regret}}$$

[Daniely et al., 2015, Jun et al., 2017, Zhang et al., 2018b]

Sleeping experts (full-information)

for $t = 1, 2, \ldots, T$ **do**

Let $\{E_1, \ldots, E_{n_t}\}$ be the active experts at t with $E_i \equiv E_i(I_i)$.

for $i = 1, 2, \ldots, n_t$ **do**

Expert E_i outputs action $\boldsymbol{x}_{t,I_i}$

end for

Play action using exponential-weights (EW):

$$\boldsymbol{x}_t = \sum_{i=1}^{n_t} \frac{e^{-L_{t-1,I_i}}}{\sum_{j=1}^{n_t} e^{-L_{t-1,I_j}}} \boldsymbol{x}_{t,I_i}$$

Observe entire loss function $f_t(\cdot)$

for $i = 1, 2, \ldots, n_t$ **do**

Send $f_t(\cdot)$ to E_i

Increment cumulative loss $L_{t,I_i} = L_{t-1,I_i} + f_t(\boldsymbol{x}_{t,I_i})$

end for

end for

$$\sum_{t \in I} \big(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{u})\big) = \underbrace{\sum_{t \in I} \big(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_{t,I})\big)}_{\text{meta-regret}} + \underbrace{\sum_{t \in I} \big(f_t(\boldsymbol{x}_{t,I}) - f_t(\boldsymbol{u})\big)}_{\text{expert-regret}}$$

$$\Longrightarrow \quad R^{\text{ada}}(\mathrm{B}, T) \lesssim \begin{cases} \sqrt{\mathrm{B}} & \text{(GC)} \\ \frac{1}{\alpha} \log \mathrm{B} & \text{(SC)} \end{cases}.$$

[Daniely et al., 2015, Jun et al., 2017, Zhang et al., 2018b]

From full-information to bandit feedback

From full-information to bandit feedback

for $t = 1, 2, \dots, T$ **do**

Let $\{E_1, \dots, E_{n_t}\}$ be the active experts at t with $E_i \equiv E_i(I_i)$.

for $i = 1, 2, \dots, n_t$ **do**

Expert E_i outputs action $\boldsymbol{x}_{t,I_i}$

end for

Play action using EW:

$$\boldsymbol{x}_t = \sum_{i=1}^{n_t} \frac{e^{-L_{t-1,I_i}}}{\sum_{j=1}^{n_t} e^{-L_{t-1,I_j}}} \boldsymbol{x}_{t,I_i}$$

Observe bandit feedback $f_t(\boldsymbol{x}_t) + \xi_t$

for $i = 1, 2, \dots, n_t$ **do**

Increment surrogate loss $L_{t,I_i} = L_{t-1,I_i} + \ell_t(\boldsymbol{x}_{t,I_i})$

end for

end for

From full-information to bandit feedback

for $t = 1, 2, \ldots, T$ **do**

Let $\{E_1, \ldots, E_{n_t}\}$ be the active experts at t with $E_i \equiv E_i(I_i)$.

for $i = 1, 2, \ldots, n_t$ **do**

Expert E_i outputs action $\boldsymbol{x}_{t,I_i}$

end for

Play action using EW: $\zeta_t \sim$ uniform on unit sphere

$$\boldsymbol{x}_t = \tilde{\boldsymbol{x}}_t + h\boldsymbol{\zeta}_t, \quad \tilde{\boldsymbol{x}}_t = \sum_{i=1}^{n_t} \frac{e^{-L_{t-1,I_i}}}{\sum_{j=1}^{n_t} e^{-L_{t-1,I_j}}} \boldsymbol{x}_{t,I_i}$$

Observe bandit feedback $f_t(\boldsymbol{x}_t) + \xi_t$

Construct gradient estimate $\boldsymbol{g}_t = (d/h)(f_t(\boldsymbol{x}_t) + \xi_t)\boldsymbol{\zeta}_t$

for $i = 1, 2, \ldots, n_t$ **do**

Increment surrogate loss $L_{t,I_i} = L_{t-1,I_i} + \ell_t(\boldsymbol{x}_{t,I_i})$

end for

end for

From full-information to bandit feedback

for $t = 1, 2, \ldots, T$ **do**

Let $\{E_1, \ldots, E_{n_t}\}$ be the active experts at t with $E_i \equiv E_i(I_i)$.

for $i = 1, 2, \ldots, n_t$ **do**

Expert E_i outputs action $\boldsymbol{x}_{t,I_i}$

end for

Play action using EW: $\quad \boldsymbol{\zeta}_t \sim$ uniform on unit sphere

$$\boldsymbol{x}_t = \tilde{\boldsymbol{x}}_t + h\boldsymbol{\zeta}_t, \quad \tilde{\boldsymbol{x}}_t = \sum_{i=1}^{n_t} \frac{e^{-L_{t-1,I_i}}}{\sum_{j=1}^{n_t} e^{-L_{t-1,I_j}}} \boldsymbol{x}_{t,I_i}$$

Observe bandit feedback $f_t(\boldsymbol{x}_t) + \xi_t$

Construct gradient estimate $\boldsymbol{g}_t = (d/h)(f_t(\boldsymbol{x}_t) + \xi_t)\boldsymbol{\zeta}_t$

for $i = 1, 2, \ldots, n_t$ **do**

Send $\boldsymbol{x}_t$ and $\boldsymbol{g}_t$ to E_i

Increment surrogate loss $L_{t,I_i} = L_{t-1,I_i} + \ell_t(\boldsymbol{x}_{t,I_i})$

end for

end for

From full-information to bandit feedback

for $t = 1, 2, \ldots, T$ **do**

Let $\{E_1, \ldots, E_{n_t}\}$ be the active experts at t with $E_i \equiv E_i(I_i)$.

for $i = 1, 2, \ldots, n_t$ **do**

Expert E_i outputs action $\boldsymbol{x}_{t,I_i}$

end for

Play action using EW: $\zeta_t \sim$ uniform on unit sphere

$$\boldsymbol{x}_t = \tilde{\boldsymbol{x}}_t + h\zeta_t, \quad \tilde{\boldsymbol{x}}_t = \sum_{i=1}^{n_t} \frac{e^{-L_{t-1,I_i}}}{\sum_{j=1}^{n_t} e^{-L_{t-1,I_j}}} \boldsymbol{x}_{t,I_i}$$

Observe bandit feedback $f_t(\boldsymbol{x}_t) + \xi_t$

Construct gradient estimate $\boldsymbol{g}_t = (d/h)(f_t(\boldsymbol{x}_t) + \xi_t)\zeta_t$

for $i = 1, 2, \ldots, n_t$ **do**

Send $\boldsymbol{x}_t$ and $\boldsymbol{g}_t$ to E_i

Increment surrogate loss $L_{t,I_i} = L_{t-1,I_i} + \ell_t(\boldsymbol{x}_{t,I_i})$

end for

end for

- $\mathbb{E}[\boldsymbol{g}_t | \tilde{\boldsymbol{x}}_t] = \nabla \hat{f}_t(\tilde{\boldsymbol{x}}_t)$ where $\hat{f}_t(\cdot)$ is a smoothed version of $f_t(\cdot)$, also convex
- Bias $= \mathbb{E}[\hat{f}_t(\boldsymbol{u}) - f_t(\boldsymbol{u})] \lesssim h$, Variance-proxy $\|\boldsymbol{g}_t\| \leq G \asymp d/h$

[Flaxman et al., 2005, Kleinberg, 2004]

From full-information to bandit feedback

for $t = 1, 2, \dots, T$ **do**

Let $\{E_1, \dots, E_{n_t}\}$ be the active experts at t with $E_i \equiv E_i(I_i)$.

for $i = 1, 2, \dots, n_t$ **do**

Expert E_i outputs action $\boldsymbol{x}_{t,I_i}$

end for

Play action using EW: $\zeta_t \sim$ uniform on unit sphere

$$\boldsymbol{x}_t = \tilde{\boldsymbol{x}}_t + h\zeta_t, \quad \tilde{\boldsymbol{x}}_t = \sum_{i=1}^{n_t} \frac{e^{-L_{t-1,I_i}}}{\sum_{j=1}^{n_t} e^{-L_{t-1,I_j}}} \boldsymbol{x}_{t,I_i}$$

Observe bandit feedback $f_t(\boldsymbol{x}_t) + \xi_t$

Construct gradient estimate $\boldsymbol{g}_t = (d/h)(f_t(\boldsymbol{x}_t) + \xi_t)\zeta_t$

for $i = 1, 2, \dots, n_t$ **do**

Send $\boldsymbol{x}_t$ and $\boldsymbol{g}_t$ to E_i

Increment surrogate loss $L_{t,I_i} = L_{t-1,I_i} + \ell_t(\boldsymbol{x}_{t,I_i})$

end for

end for

- $\mathbb{E}[\boldsymbol{g}_t | \tilde{\boldsymbol{x}}_t] = \nabla \hat{f}_t(\tilde{\boldsymbol{x}}_t)$ where $\hat{f}_t(\cdot)$ is a smoothed version of $f_t(\cdot)$, also convex
- Bias $= \mathbb{E}[\hat{f}_t(\boldsymbol{u}) - f_t(\boldsymbol{u})] \lesssim h$, Variance-proxy $\|\boldsymbol{g}_t\| \le G \asymp d/h$

How to design $\ell_t(\cdot)$?

[Flaxman et al., 2005, Kleinberg, 2004]

Designing surrogate loss $\ell_t(\cdot)$

- Linear loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\tilde{\boldsymbol{x}}_t - \boldsymbol{x})$

Designing surrogate loss $\ell_t(\cdot)$

- Linear loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\tilde{\boldsymbol{x}}_t - \boldsymbol{x})$ ▶ $R^{\text{ada}}(\mathsf{B}, T) \lesssim \sqrt{\mathsf{B}}$ in OCO [Wang et al., 2018]

Designing surrogate loss $\ell_t(\cdot)$

- Linear loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\tilde{\boldsymbol{x}}_t - \boldsymbol{x})$
  - ▶ $R^{\text{ada}}(\mathrm{B}, T) \lesssim \sqrt{\mathrm{B}}$ in OCO [Wang et al., 2018]
  - ▶ expert-regret for $E_I \lesssim \sqrt{|I|} \Rightarrow R^{\text{ada}}(\mathrm{B}, T) \lesssim B$ in BCO

Designing surrogate loss $\ell_t(\cdot)$

- Linear loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\tilde{\boldsymbol{x}}_t - \boldsymbol{x})$
  - ▶ $R^{\text{ada}}(\mathrm{B}, T) \lesssim \sqrt{\mathrm{B}}$ in OCO [Wang et al., 2018]
  - ▶ expert-regret for $E_I \lesssim \sqrt{|I|} \Rightarrow R^{\text{ada}}(\mathrm{B}, T) \lesssim B$ in BCO
- Inspired by [Chernov and Vovk, 2009, van Erven et al., 2021, Wang et al., 2020, Zhang et al., 2021], propose quadratic surrogate loss

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

Designing surrogate loss $\ell_t(\cdot)$

- Linear loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\tilde{\boldsymbol{x}}_t - \boldsymbol{x})$
  - ▶ $R^{\mathrm{ada}}(\mathrm{B}, T) \lesssim \sqrt{\mathrm{B}}$ in OCO [Wang et al., 2018]
  - ▶ expert-regret for $E_I \lesssim \sqrt{|I|} \Rightarrow R^{\mathrm{ada}}(\mathrm{B}, T) \lesssim B$ in BCO
- Inspired by [Chernov and Vovk, 2009, van Erven et al., 2021, Wang et al., 2020, Zhang et al., 2021], propose quadratic surrogate loss

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

(i) Absorbs the variance of $\boldsymbol{g}_t$

(ii) Together with η, adapts to unknown curvature α

Designing surrogate loss $\ell_t(\cdot)$

- Linear loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\tilde{\boldsymbol{x}}_t - \boldsymbol{x})$
  - $R^{\text{ada}}(\mathrm{B}, T) \lesssim \sqrt{\mathrm{B}}$ in OCO [Wang et al., 2018]
  - expert-regret for $E_I \lesssim \sqrt{|I|} \Rightarrow R^{\text{ada}}(\mathrm{B}, T) \lesssim B$ in BCO
- Inspired by [Chernov and Vovk, 2009, van Erven et al., 2021, Wang et al., 2020, Zhang et al., 2021], propose quadratic surrogate loss

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

Learning rate η:

- Multiple experts with different η's on each interval
- EW tilted by η locks onto the best η for unknown curvature

(i) Absorbs the variance of $\boldsymbol{g}_t$

(ii) Together with η, adapts to unknown curvature α

Designing surrogate loss $\ell_t(\cdot)$

- Linear loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\tilde{\boldsymbol{x}}_t - \boldsymbol{x})$
  - ▶ $R^{\text{ada}}(\mathrm{B}, T) \lesssim \sqrt{\mathrm{B}}$ in OCO [Wang et al., 2018]
  - ▶ expert-regret for $E_I \lesssim \sqrt{|I|} \Rightarrow R^{\text{ada}}(\mathrm{B}, T) \lesssim B$ in BCO
- Inspired by [Chernov and Vovk, 2009, van Erven et al., 2021, Wang et al., 2020, Zhang et al., 2021], propose quadratic surrogate loss

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

Designing surrogate loss $\ell_t(\cdot)$

- Linear loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\tilde{\boldsymbol{x}}_t - \boldsymbol{x})$
  - ▶ $R^{\text{ada}}(\mathrm{B}, T) \lesssim \sqrt{\mathrm{B}}$ in OCO [Wang et al., 2018]
  - ▶ expert-regret for $E_I \lesssim \sqrt{|I|} \Rightarrow R^{\text{ada}}(\mathrm{B}, T) \lesssim B$ in BCO
- Inspired by [Chernov and Vovk, 2009, van Erven et al., 2021, Wang et al., 2020, Zhang et al., 2021], propose quadratic surrogate loss

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

$$\Downarrow \text{ by definition}$$

$$\underbrace{\sum_{t \in I} \langle \mathbb{E}[\boldsymbol{g}_t | \tilde{\boldsymbol{x}}_t], \tilde{\boldsymbol{x}}_t - \boldsymbol{u} \rangle}_{\text{linearized regret associated with } \hat{f}_t} \leq \frac{1}{\eta} \underbrace{\sum_{t \in I} \mathbb{E}\left[\ell_t^\eta(\tilde{\boldsymbol{x}}_t) - \ell_t^\eta(\boldsymbol{u}) | \tilde{\boldsymbol{x}}_t\right]}_{:=\bigstar \lesssim \log |I|} + \eta G^2 \sum_{t \in I} \|\tilde{\boldsymbol{x}}_t - \boldsymbol{u}\|^2$$

Designing surrogate loss $\ell_t(\cdot)$

- Linear loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\tilde{\boldsymbol{x}}_t - \boldsymbol{x})$
  - ▶ $R^{\mathrm{ada}}(\mathrm{B}, T) \lesssim \sqrt{\mathrm{B}}$ in OCO [Wang et al., 2018]
  - ▶ expert-regret for $E_I \lesssim \sqrt{|I|} \Rightarrow R^{\mathrm{ada}}(\mathrm{B}, T) \lesssim B$ in BCO
- Inspired by [Chernov and Vovk, 2009, van Erven et al., 2021, Wang et al., 2020, Zhang et al., 2021], propose quadratic surrogate loss

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

$$\Downarrow \text{ by definition}$$

$$\underbrace{\sum_{t\in I} \langle \mathbb{E}[\boldsymbol{g}_t | \tilde{\boldsymbol{x}}_t], \tilde{\boldsymbol{x}}_t - \boldsymbol{u}\rangle}_{\text{linearized regret associated with } \hat{f}_t} \leq \frac{1}{\eta} \underbrace{\sum_{t\in I} \mathbb{E}\left[\ell_t^\eta(\tilde{\boldsymbol{x}}_t) - \ell_t^\eta(\boldsymbol{u}) | \tilde{\boldsymbol{x}}_t\right]}_{:=\bigstar \lesssim \log|I|} + \eta G^2 \sum_{t\in I} \|\tilde{\boldsymbol{x}}_t - \boldsymbol{u}\|^2$$

$$\Downarrow \text{ by convexity of } \hat{f}_t$$

$$\underbrace{\sum_{t\in I} \mathbb{E}\left[\hat{f}_t(\tilde{\boldsymbol{x}}_t) - \hat{f}_t(\boldsymbol{u})\right]}_{\text{regret associated with } \hat{f}_t} \leq \underbrace{\mathbb{E}\left[\frac{1}{\eta}\bigstar + \left(\eta G^2 - \frac{\alpha}{2}\right) \sum_{t\in I} \|\tilde{\boldsymbol{x}}_t - \boldsymbol{u}\|^2\right]}_{\substack{\text{grid over } \eta \text{ to hedge} \\ \text{against unknown curvature } \alpha}}$$

Main results: Tilted-EW Average with Sleeping Experts (TEWA-SE)

Main results: Tilted-EW Average with Sleeping Experts (TEWA-SE)

Theorem (informal) [Liu et al., 2025]

For any $T \in \mathbb{N}^+$ and $\mathrm{B} \in [T]$, TEWA-SE with $h = \sqrt{d}\mathrm{B}^{-\frac{1}{4}}$ & geometric grid for η, satisfies

$$R^{\mathrm{ada}}(\mathrm{B}, T) \lesssim \sqrt{d}\mathrm{B}^{\frac{3}{4}},$$

and if f_t is α-strongly-convex with $\arg\min_{\boldsymbol{x} \in \mathbb{R}^d} f_t(\boldsymbol{x}) \in \Theta$ for all $t \in [T]$, it furthermore holds that

$$R^{\mathrm{ada}}(\mathrm{B}, T) \lesssim \frac{d}{\alpha}\sqrt{\mathrm{B}}.$$

Main results: Tilted-EW Average with Sleeping Experts (TEWA-SE)

Theorem (informal) [Liu et al., 2025]

For any $T \in \mathbb{N}^+$ and $\mathrm{B} \in [T]$, TEWA-SE with $h = \sqrt{d}\mathrm{B}^{-\frac{1}{4}}$ & geometric grid for η, satisfies

$$R^{\text{ada}}(\mathrm{B}, T) \lesssim \sqrt{d}\mathrm{B}^{\frac{3}{4}},$$

and if f_t is α-strongly-convex with $\arg\min_{\boldsymbol{x} \in \mathbb{R}^d} f_t(\boldsymbol{x}) \in \Theta$ for all $t \in [T]$, it furthermore holds that

$$R^{\text{ada}}(\mathrm{B}, T) \lesssim \frac{d}{\alpha}\sqrt{\mathrm{B}}.$$

Corollary

For known S, setting $\mathrm{B} = \frac{T}{S}$ yields

$$R^{\text{swi}}(T, S) \lesssim \begin{cases} \sqrt{d}S^{\frac{1}{4}}T^{\frac{3}{4}} & \text{(GC)} \\ \underline{d\sqrt{ST}} & \text{(SC)} \end{cases}$$

Underline: minimax-optimal rates.

Towards parameter-free guarantees (e.g., unknown S)

- **Bandit-over-Bandit** wrapper to hedge unknown S yields suboptimal $R^{\mathrm{swi}}(T, S)$
  [Liu et al., 2025]

Towards parameter-free guarantees (e.g., unknown S)

- **Bandit-over-Bandit** wrapper to hedge unknown S yields suboptimal $R^{\mathsf{swi}}(T, S)$ [Liu et al., 2025]
- **Links to coin betting:**

Towards parameter-free guarantees (e.g., unknown S)

- **Bandit-over-Bandit** wrapper to hedge unknown S yields suboptimal $R^{\text{swi}}(T, S)$
  [Liu et al., 2025]
- **Links to coin betting:**
  - Coin-betting with KT potential yields optimal parameter-free $R^{\text{swi}}(T, S)$ for OCO
    [Jun et al., 2017, Orabona and Pál, 2016]

Towards parameter-free guarantees (e.g., unknown S)

- **Bandit-over-Bandit** wrapper to hedge unknown S yields suboptimal $R^{\mathrm{swi}}(T, S)$ [Liu et al., 2025]
- **Links to coin betting:**
  - Coin-betting with KT potential yields optimal parameter-free $R^{\mathrm{swi}}(T, S)$ for OCO [Jun et al., 2017, Orabona and Pál, 2016]
  - TEWA as a coin-betting scheme with the log-sum-exp potential for BCO

Summary & discussions

Summary & discussions

Algorithm	clipped Exploration by Optimization (cExO)	Tilted-EW Average with Sleeping Experts (TEWA-SE)
General convex (GC) Strongly convex (SC)	$\underline{d^{\frac{5}{2}}\sqrt{ST}}$	$\sqrt{d}\, S^{\frac{1}{4}} T^{\frac{3}{4}}$ $\underline{d\sqrt{ST}}$
Features	• Conceptual design • Minimax-optimal in S, T • Not poly-time computable	• Adaptive to curvature α • Optimal in S, T for SC losses • Poly-time • Natural betting interpretation

Summary & discussions

Algorithm	clipped Exploration by Optimization (cExO)	Tilted-EW Average with Sleeping Experts (TEWA-SE)
General convex (GC) Strongly convex (SC)	$\underline{d^{\frac{5}{2}}\sqrt{ST}}$	$\sqrt{d}\,S^{\frac{1}{4}}T^{\frac{3}{4}}$ $\underline{d\sqrt{ST}}$
Features	• Conceptual design • Minimax-optimal in S, T • Not poly-time computable	• Adaptive to curvature α • Optimal in S, T for SC losses • Poly-time • Natural betting interpretation

Future directions: Towards minimax-optimal, efficient, parameter-free algorithms.

- Combine TEWA-SE with coin betting [Jun et al., 2017]?
- Leverage second-order information like online Newton methods from [Fokkema et al., 2024, Suggala et al., 2024] that achieve state-of-the-art $\sqrt{T}$ static regrets?

References I

Adamskiy, D., Koolen, W. M., Chernov, A., and Vovk, V. (2016).
A closer look at adaptive regret.
Journal of Machine Learning Research, 17(23):1–21.

Besbes, O., Gur, Y., and Zeevi, A. (2015).
Non-stationary stochastic optimization.
Operations Research, 63(5):1227–1244.

Cesa-Bianchi, N., Freund, Y., Haussler, D., Helmbold, D. P., Schapire, R. E., and Warmuth, M. K. (1997).
How to use expert advice.
Journal of the ACM, 44(3):427–485.

Chen, T. and Giannakis, G. B. (2018).
Bandit convex optimization for scalable and dynamic IoT management.
IEEE Internet of Things Journal, 6(1):1276–1286.

Chernov, A. and Vovk, V. (2009).
Prediction with expert evaluators' advice.
In *International Conference on Algorithmic Learning Theory*, pages 8–22. Springer.

References II

Cutkosky, A. (2020).
Parameter-free, dynamic, and strongly-adaptive online learning.
In *International Conference on Machine Learning*, volume 119, pages 2250–2259. PMLR.

Daniely, A., Gonen, A., and Shalev-Shwartz, S. (2015).
Strongly adaptive online learning.
In *International Conference on Machine Learning*, pages 1405–1411. PMLR.

Flaxman, A., Kalai, A. T., and McMahan, H. B. (2005).
Online convex optimization in the bandit setting: gradient descent without a gradient.
In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 385–394. SIAM.

Fokkema, H., van der Hoeven, D., Lattimore, T., and Mayo, J. J. (2024).
Online newton method for bandit convex optimisation.
In *Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 1713–1714. PMLR.

Freund, Y., Schapire, R., Singer, Y., and Warmuth, M. (1997).
Using and combining predictors that specialize.
Conference Proceedings of the Annual ACM Symposium on Theory of Computing.

References III

Hazan, E. and Seshadhri, C. (2009).
Efficient learning algorithms for changing environments.
In *International Conference on Machine Learning*, volume 382 of *ACM International Conference Proceeding Series*, pages 393–400. ACM.

Herbster, M. and Warmuth, M. K. (1998).
Tracking the Best Expert.
Machine Learning, 32(2):151–178.

Jun, K.-S., Orabona, F., Wright, S., and Willett, R. (2017).
Online learning for changing environments using coin betting.
arXiv preprint arXiv:1711.02545.

Kleinberg, R. (2004).
Nearly tight bounds for the continuum-armed bandit problem.
In *International Conference on Neural Information Processing Systems*, page 697–704. MIT Press.

Lattimore, T. and Gyorgy, A. (2021).
Mirror descent and the information ratio.
In *Conference on Learning Theory*, pages 2965–2992. PMLR.

References IV

Littlestone, N. and Warmuth, M. (1994).
The weighted majority algorithm.
Information and Computation, 108(2):212–261.

Liu, X., Baudry, D., Zimmert, J., Rebeschini, P., and Akhavan, A. (2025).
Non-stationary bandit convex optimization: A comprehensive study.
Advances in Neural Information Processing Systems.
arXiv:2506.02980.

Orabona, F. and Pál, D. (2016).
Coin betting and parameter-free online learning.
Advances in Neural Information Processing Systems, 29.

Suggala, A., Sun, Y. J., Netrapalli, P., and Hazan, E. (2024).
Second order methods for bandit optimization and control.
In *The Thirty Seventh Annual Conference on Learning Theory*, pages 4691–4763. PMLR.

van Erven, T., Koolen, W. M., and van der Hoeven, D. (2021).
Metagrad: Adaptation using multiple learning rates in online learning.
Journal of Machine Learning Research, 22(161):1–61.

References V

Vovk, V. (1998).
A game of prediction with expert advice.
Journal of Computer and System Sciences, 56(2):153–173.

Wang, G., Lu, S., and Zhang, L. (2020).
Adaptivity and optimality: A universal algorithm for online convex optimization.
In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 659–668. PMLR.

Wang, G., Zhao, D., and Zhang, L. (2018).
Minimizing adaptive regret with one gradient per iteration.
In *International Joint Conference on Artificial Intelligence*, IJCAI'18, page 2762–2768. AAAI Press.

Wang, Y. (2025).
On adaptivity in nonstationary stochastic optimization with bandit feedback.
Operations Research, 73(2):819–828.

Yang, W., Wang, Y., Zhao, P., and Zhang, L. (2024).
Universal online convex optimization with 1 projection per round.
In *Advances in Neural Information Processing Systems*, volume 37, pages 31438–31472. Curran Associates, Inc.

References VI

Zhang, L., Lu, S., and Zhou, Z.-H. (2018a).
Adaptive online learning in dynamic environments.
In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 1330–1340. Curran Associates Inc.

Zhang, L., Wang, G., Tu, W.-W., Jiang, W., and Zhou, Z.-H. (2021).
Dual adaptivity: a universal algorithm for minimizing the adaptive regret of convex functions.
In *International Conference on Neural Information Processing Systems*. Curran Associates Inc.

Zhang, L., Yang, T., Zhou, Z.-H., et al. (2018b).
Dynamic regret of strongly adaptive methods.
In *International conference on machine learning*, pages 5882–5891. PMLR.

Zhao, P., Wang, G., Zhang, L., and Zhou, Z.-H. (2021).
Bandit convex optimization in non-stationary environments.
Journal of Machine Learning Research, 22(125):1–45.

Zhao, P., Xie, Y.-F., Zhang, L., and Zhou, Z.-H. (2022).
Efficient methods for non-stationary online learning.
Advances in Neural Information Processing Systems, 35:11573–11585.