ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΠΑΙΓΝΙΩΝ ΚΑΙ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ

ΕΠΑΝΑΛΑΜΒΑΝΟΜΕΝΗ ΚΥΡΤΗ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

Παναγιώτης Μερτικόπουλος

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Τμήμα Μαθηματικών

Χειμερινό Εξάμηνο, 2023–2024

Outline

Setting

Sequence of events: Online convex optimization (OCO)

Require: convex action set $\mathcal{X} \subseteq \mathbb{R}^d$; convex loss functions $\ell_t \colon \mathcal{X} \to \mathbb{R}$, $t = 1, 2, \dots$

repeat

At each epoch $t = 1, 2, \dots$ **do**

Choose ***action*** $x_t \in \mathcal{X}$ # action selection

Encounter ***loss function*** $\ell_t \colon \mathcal{X} \to \mathbb{R}$ # Nature plays

Incur ***cost*** $c_t = \ell_t(x_t)$ # reward phase

Observe loss function ℓ_t # feedback phase

until end

Defining elements

- ***Time:*** discrete
- ***Players:*** single
- ***Actions:*** continuous
- ***Losses:*** exogenous
- ***Feedback:*** **depends** (function-based, gradient-based, loss-based, ...)

Setting

Sequence of events: Online convex optimization (OCO)

Require: convex action set $\mathcal{X} \subseteq \mathbb{R}^d$; convex loss functions $\ell_t \colon \mathcal{X} \to \mathbb{R}$, $t = 1, 2, \dots$

repeat

At each epoch $t = 1, 2, \dots$ **do**

Choose ***action*** $x_t \in \mathcal{X}$ # action selection

Encounter ***loss function*** $\ell_t \colon \mathcal{X} \to \mathbb{R}$ # Nature plays

Incur ***cost*** $c_t = \ell_t(x_t)$ # reward phase

Observe ***gradient*** $g_t = \nabla \ell_t(x_t)$ # feedback phase

until end

Defining elements

- ***Time:*** discrete
- ***Players:*** single
- ***Actions:*** continuous
- ***Losses:*** exogenous
- ***Feedback:*** **depends** (function-based, ***gradient-based,*** loss-based, ...)

Setting

Sequence of events: Online convex optimization (OCO)

Require: convex **action set** $\mathcal{X} \subseteq \mathbb{R}^d$; convex **loss functions** $\ell_t \colon \mathcal{X} \to \mathbb{R}$, $t = 1, 2, \ldots$

repeat

At each epoch $t = 1, 2, \ldots$ **do**

Choose ***action*** $x_t \in \mathcal{X}$ — # action selection

Encounter ***loss function*** $\ell_t \colon \mathcal{X} \to \mathbb{R}$ — # Nature plays

Incur ***cost*** $c_t = \ell_t(x_t)$ — # reward phase

Observe **cost** $c_t = \ell_t(x_t)$ — # feedback phase

until end

Defining elements

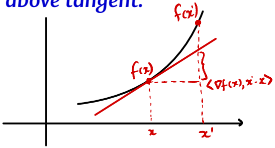
- ***Time:*** discrete
- ***Players:*** single
- ***Actions:*** continuous
- ***Losses:*** exogenous
- ***Feedback:*** **depends** (function-based, gradient-based, **loss-based,** ...)

Convex analysis cheatsheet

If ℓ is convex:

1. ***Local minima = global minima = stationary points*** # stationarity = optimality

2. ***Graph above tangent:*** # consistent linear estimates



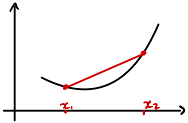
$$f(x') \geq f(x) + \langle \nabla f(x), x' - x \rangle$$

subgradient: $f(x') \geq f(x) + \langle g, x' - x \rangle$

3. ***First-order stationarity:***

$$x^* \text{ is a minimizer of } f \iff \langle \nabla f(x^*), x - x^* \rangle \geq 0 \quad \text{for all } x \in \mathcal{X}$$
$$\iff \langle \nabla f(x), x - x^* \rangle \geq 0 \quad \text{for all } x \in \mathcal{X}$$

4. ***Jensen's inequality:*** # mean value exceeds value of the mean



$$f\left(\sum_{i=1}^m \lambda_i x_i\right) \leq \sum_{i=1}^m \lambda_i f(x_i) \qquad \text{for all } x_i \in \mathcal{X}, \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1.$$

Feedback

Types of feedback

From best to worst (more to less info):

- ***Full information:*** observe entire loss function $\ell_t : \mathcal{X} \to \mathbb{R}$ # deterministic function feedback
- ***First-order info, exact:*** observe (~~sub~~)gradient $\left(g_t \in \partial \ell_t(x_t)\right)$ $g_t = \nabla \ell_t(x_t)$ # deterministic vector feedback
- ***First-order info, inexact***: observe noisy estimate of g_t # stochastic vector feedback
- ***Zeroth-order info (bandit):*** observe only incurred cost $c_t = \ell_t(x_t)$ # deterministic scalar feedback

Feedback

Types of feedback

From best to worst (more to less info):

- *Full information:* observe entire loss function $\ell_t \colon \mathcal{X} \to \mathbb{R}$ # deterministic function feedback
- ***First-order info, exact:*** observe (sub)gradient $g_t \in \partial \ell_t(x_t)$ # deterministic vector feedback
- ***First-order info, inexact***: observe noisy estimate of g_t # stochastic vector feedback
- *Zeroth-order info (bandit):* observe only incurred cost $c_t = \ell_t(x_t)$ # deterministic scalar feedback

The oracle model

A ***stochastic first-order oracle (SFO)*** for $g_t \in \partial \ell_t(x_t)$ is a random vector of the form

$$\hat{g}_t = g_t + U_t + b_t \tag{SFO}$$

where U_t is **zero-mean** and $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - g_t$ is the **bias** of $\hat{g}_t$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$[\ell_t(x_t) - \ell_t(p)]$$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]$$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\max_{p \in \mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]$$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\operatorname{Reg}(T) = \max_{p \in \mathcal{X}} \underbrace{\sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]}_{\operatorname{Reg}_p(T)} = \sum_{t=1}^{T} \ell_t(x_t) - \min_{p \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(p)$$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\mathrm{Reg}(T) = \max_{p \in \mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)] = \sum_{t=1}^{T} \ell_t(x_t) - \min_{p \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(p)$$

- ***No regret:*** $\mathrm{Reg}(T) = o(T)$
- ***Adversarial framework:*** minimize regret against **any** given sequence ℓ_t

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\mathrm{Reg}(T) = \max_{p \in \mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)] = \sum_{t=1}^{T} \ell_t(x_t) - \min_{p \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(p)$$

- ***No regret:*** $\mathrm{Reg}(T) = o(T)$
- ***Adversarial framework:*** minimize regret against **any** given sequence ℓ_t
- ***Expected regret:***

$$\mathbb{E}[\mathrm{Reg}(T)] = \mathbb{E}\left[\max_{p \in \mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]\right]$$

- ***Pseudo-regret:***

$$\overline{\mathrm{Reg}}(T) = \max_{p \in \mathcal{X}} \mathbb{E}\left[\sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]\right]$$

Regret

Performance measured by the agent's ***regret*** (loss formulation):

$$\mathrm{Reg}(T) = \max_{p\in\mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)] = \sum_{t=1}^{T} \ell_t(x_t) - \min_{p\in\mathcal{X}} \sum_{t=1}^{T} \ell_t(p)$$

- ***No regret:*** $\mathrm{Reg}(T) = o(T)$
- ***Adversarial framework:*** minimize regret against **any** given sequence ℓ_t
- ***Expected regret:***

$$\mathbb{E}[\mathrm{Reg}(T)] = \mathbb{E}\left[\max_{p\in\mathcal{X}} \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]\right]$$

- ***Pseudo-regret:***

$$\overline{\mathrm{Reg}}(T) = \max_{p\in\mathcal{X}} \mathbb{E}\left[\sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)]\right]$$

- $\overline{\mathrm{Reg}}(T) \le \mathbb{E}[\mathrm{Reg}(T)]$: bounds do not translate "as is" but "almost"

Cesa-Bianchi & Lugosi, 2006, Bubeck & Cesa-Bianchi, 2012, Lattimore & Szepesvári, 2020

Outline

Be the leader

- Suppose ℓ_t is observed ***before*** playing x_t
- Then the agent can try to ***be the leader (BTL)***

$$x_t \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t} \ell_s(x) \tag{BTL}$$

Be the leader

- Suppose ℓ_t is observed ***before*** playing x_t
- Then the agent can try to ***be the leader (BTL)***

$$x_t \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t} \ell_s(x) \tag{BTL}$$

Regret of BTL

☞ Under (BTL), the learner incurs $\mathrm{Reg}(T) = 0$.

Be the leader

- Suppose ℓ_t is observed ***before*** playing x_t
- Then the agent can try to ***be the leader (BTL)***

$$x_t \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t} \ell_s(x) \tag{BTL}$$

Regret of BTL

☞ Under (BTL), the learner incurs $\mathrm{Reg}(T) = 0$.

...unrealistic

Follow the leader

- Suppose ℓ_t is observed ***after*** playing x_t
- Then the agent can try to ***follow the leader (FTL)***

$$x_{t+1} \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t} \ell_s(x) \tag{FTL}$$

Follow the leader

- Suppose ℓ_t is observed ***after*** playing x_t
- Then the agent can try to ***follow the leader (FTL)***

$$x_{t+1} \in \arg\min_{x\in\mathcal{X}} \sum_{s=1}^{t} \ell_s(x) \tag{FTL}$$

$$x_1 \in \arg\min_{x\in\mathcal{X}} \cancelto{0}{\sum_{s=1}^{0} \ell_s(x)} = \arg\min_{x\in\mathcal{X}} 0 = \mathcal{X}$$

Does (FTL) lead to no regret?

Template bound for FTL

FTL regret bound

For all $p \in \mathcal{X}$, the regret of (FTL) can be bounded as

$$\mathrm{Reg}_p(T) = \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(p)] \leq \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(x_{t+1})]$$

Small when x_t close to x_{t+1}

"Stability"

"Following"

"One-lag-behind"

Template bound for FTL

FTL regret bound

For all $p \in \mathcal{X}$, the regret of (FTL) can be bounded as

$$\mathrm{Reg}_p(T) = \sum_{t=1}^{T}[\ell_t(x_t) - \ell_t(p)] \leq \sum_{t=1}^{T}[\ell_t(x_t) - \ell_t(x_{t+1})]$$

Proof.

By induction, assume that $\sum_{t=1}^{T-1} \ell_t(x_{t+1}) \leq \sum_{t=1}^{T-1} \ell_t(p)$ $\forall p$. Want to show: $\sum_{t=1}^{T} \ell_t(x_{t+1}) \leq \sum_{t=1}^{T} \ell_t(p)$

$\hookrightarrow \sum_{t=1}^{T} \ell_t(x_{t+1}) = \sum_{t=1}^{T-1} \ell_t(x_{t+1}) + \ell_T(x_{T+1}) \leq \sum_{t=1}^{T-1} \ell_t(p) + \ell_T(x_{T+1})$

(**) For $p \leftarrow x_{T+1}$: $\sum_{t=1}^{T} \ell_t(x_{t+1}) \leq \sum_{t=1}^{T} \ell_t(x_{T+1}) \underset{\text{(FTL)}}{=} \min_{x \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(x) \underset{\text{from min}}{\leq} \sum_{t=1}^{T} \ell_t(p)$ $\forall p$ ✓

□

References I

[1] Abernethy, J., Bartlett, P. L., Rakhlin, A., and Tewari, A. Optimal strategies and minimax lower bounds for online convex games. In *COLT '08: Proceedings of the 21st Annual Conference on Learning Theory*, 2008.

[2] Bubeck, S. and Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

[3] Cesa-Bianchi, N. and Lugosi, G. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[4] Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, Cambridge, UK, 2020.

[5] Nesterov, Y. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120(1):221–259, 2009.

[6] Shalev-Shwartz, S. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011.

[7] Shalev-Shwartz, S. and Singer, Y. Convex repeated games and Fenchel duality. In *NIPS' 06: Proceedings of the 19th Annual Conference on Neural Information Processing Systems*, pp. 1265–1272. MIT Press, 2006.

[8] Xiao, L. Dual averaging methods for regularized stochastic learning and online optimization. *Journal of Machine Learning Research*, 11: 2543–2596, October 2010.