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** $\ell_t\colon \mathcal{X} \to \mathbb{R}$, $t = 1, 2, \dots$

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 **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:***

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

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 $\ell_t\colon \mathcal{X} \to \mathbb{R}$ # 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

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):

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

Regret

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

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

- ***No regret:*** $\text{Reg}(T) = o(T)$
- ***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:***

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

- $\overline{\mathrm{Reg}}(T) \leq \mathbb{E}[\mathrm{Reg}(T)]$: 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)***

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

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

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

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.

□

FTL against quadratic losses

Test (FTL) in an ***online quadratic optimization (OQO)*** problem:

$$\ell_t(x) = \frac{1}{2}\|x - p_t\|^2 \quad \text{for some sequence of center points } p_t,\ t = 1, 2, \ldots \tag{OQO}$$

FTL against quadratic losses

Test (FTL) in an ***online quadratic optimization (OQO)*** problem:

$$\ell_t(x) = \frac{1}{2}\|x - p_t\|^2 \quad \text{for some sequence of center points } p_t,\ t = 1, 2, \ldots \tag{OQO}$$

Regret of FTL in quadratic problems

☞ ***Assume:*** (FTL) is run against (OQO) with $\sup_t \|p_t\| \leq R$

✓ **Then:** $\operatorname{Reg}(T) \leq 4R^2(1 + \log T)$

FTL against quadratic losses

Test (FTL) in an ***online quadratic optimization (OQO)*** problem:

$$\ell_t(x) = \frac{1}{2}\|x - p_t\|^2 \quad \text{for some sequence of center points } p_t,\ t = 1, 2, \dots \tag{OQO}$$

Regret of FTL in quadratic problems

- ☞ ***Assume:*** (FTL) is run against (OQO) with $\sup_t \|p_t\| \leq R$
- ✓ **Then:** $\operatorname{Reg}(T) \leq 4R^2(1 + \log T)$

Proof.

□

FTL against linear losses

Test (FTL) in an ***online linear optimization (OLO)*** problem:

$$\ell_t(x) = \langle w_t, x \rangle \quad \text{for some sequence of loss vectors } w_t \in \mathbb{R}^d,\ t = 1, 2, \dots \tag{OLO}$$

FTL against linear losses

Test (FTL) in an ***online linear optimization (OLO)*** problem:

$$\ell_t(x) = \langle w_t, x \rangle \quad \text{for some sequence of loss vectors } w_t \in \mathbb{R}^d,\ t = 1, 2, \dots \tag{OLO}$$

Chasing the leader

☞ ***Assume:*** $\mathcal{X} = [-1, 1]$ and (FTL) is run against (OLO) with $w_1 = -1/2$ and $w_t = (-1)^t$ otherwise

⚠ ***What is the incurred regret?***

Follow the regularized leader

Add a fictitious "day zero loss" $\Longrightarrow$ ***follow the regularized leader (FTRL)***

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \Big\{ \sum_{s=1}^{t} \ell_s(x) + \underbrace{\lambda h(x)}_{\text{“}\ell_0(x)\text{”}} \Big\} \tag{FTRL}$$

where

- The ***regularization function*** $h \colon \mathcal{X} \to \mathbb{R}$ is strongly convex   # $h(x) - (K/2)\|x\|^2$ convex for some $K > 0$
- The ***regularization weight*** $\lambda > 0$ can be tuned by the optimizer

Main idea: Regularization $\Longrightarrow$ Stability $\Longrightarrow$ Less regret

Algorithm due to Shalev-Shwartz & Singer, 2006, Shalev-Shwartz, 2011

Example 1: Euclidean regularization

- **Setup:** $\mathcal{X} = \mathbb{R}^d$, linear losses $\ell_t(x) = \langle w_t, x \rangle$
- **Regularizer:**

$$h(x) = \tfrac{1}{2}\|x\|^2$$

- **Algorithm:**

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \langle w_s, x \rangle + \frac{\lambda}{2}\|x\|^2 \right\}$$

Example 1: Euclidean regularization

- **Setup:** $\mathcal{X} = \mathbb{R}^d$, linear losses $\ell_t(x) = \langle w_t, x \rangle$
- **Regularizer:**

$$h(x) = \tfrac{1}{2}\|x\|^2$$

- **Algorithm:**

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \langle w_s, x \rangle + \frac{\lambda}{2}\|x\|^2 \right\} = -\frac{1}{\lambda} \sum_{s=1}^{t} w_s = x_t - (1/\lambda) w_t$$

Example 1: Euclidean regularization

- **Setup:** $\mathcal{X} = \mathbb{R}^d$, linear losses $\ell_t(x) = \langle w_t, x \rangle$
- **Regularizer:**

$$h(x) = \tfrac{1}{2}\|x\|^2$$

- **Algorithm:**

$$x_{t+1} = \arg\min_{x\in\mathcal{X}} \left\{ \sum_{s=1}^{t} \langle w_s, x\rangle + \frac{\lambda}{2}\|x\|^2 \right\} = -\frac{1}{\lambda}\sum_{s=1}^{t} w_s = x_t - (1/\lambda) w_t$$

- Euclidean regularization + linear losses $(w_t = \nabla \ell_t(x_t)) \implies$ ***gradient descent:***

$$x_{t+1} = x_t - \underbrace{\eta}_{1/\lambda} \nabla \ell_t(x_t) \tag{GD}$$

Example 2: Entropic regularization

- **Setup:** $\mathcal{X} = \Delta(\mathcal{A})$, linear payoffs $u_t(x) = \langle v_t, x \rangle$

☞ payoffs instead of costs

- **Regularizer:**

$$h(x) = \sum_{\alpha \in \mathcal{A}} x_\alpha \log x_\alpha$$

- **Algorithm:**

$$x_{t+1} = \arg\max_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \langle v_s, x \rangle - \lambda \sum_{\alpha \in \mathcal{A}} x_\alpha \log x_\alpha \right\}$$

Example 2: Entropic regularization

- **Setup:** $\mathcal{X} = \Delta(\mathcal{A})$, linear payoffs $u_t(x) = \langle v_t, x \rangle$ ☞ payoffs instead of costs
- **Regularizer:**
$$h(x) = \sum_{\alpha \in \mathcal{A}} x_\alpha \log x_\alpha$$
- **Algorithm:**
$$x_{t+1} = \arg\max_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \langle v_s, x \rangle - \lambda \sum_{\alpha \in \mathcal{A}} x_\alpha \log x_\alpha \right\} = \frac{\exp(\sum_{s=1}^{t} v_{\alpha,s}/\lambda)}{\sum_{\beta \in \mathcal{A}} \exp(\sum_{s=1}^{t} v_{\beta,s}/\lambda)}$$

Example 2: Entropic regularization

- **Setup:** $\mathcal{X} = \Delta(\mathcal{A})$, linear payoffs $u_t(x) = \langle v_t, x\rangle$ ☞ payoffs instead of costs
- **Regularizer:**

$$h(x) = \sum_{\alpha\in\mathcal{A}} x_\alpha \log x_\alpha$$

- **Algorithm:**

$$x_{t+1} = \arg\max_{x\in\mathcal{X}} \left\{ \sum_{s=1}^{t} \langle v_s, x\rangle - \lambda \sum_{\alpha\in\mathcal{A}} x_\alpha \log x_\alpha \right\} = \frac{\exp(\sum_{s=1}^{t} v_{\alpha,s}/\lambda)}{\sum_{\beta\in\mathcal{A}} \exp(\sum_{s=1}^{t} v_{\beta,s}/\lambda)}$$

- Entropic regularization + linear payoffs $\implies$ ***exponential weights:***

$$\begin{aligned} y_{t+1} &= y_t + \overbrace{\eta}^{1/\lambda} v_t \\ x_{t+1} &= \underbrace{\Lambda(y_{t+1})}_{\text{logit map}} \end{aligned} \tag{EW}$$

Template bound for FTRL

FTRL regret bound

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

$$\operatorname{Reg}_p(T) \leq \lambda[h(p) - h(x_1)] + \sum_{t=1}^{T}[\ell_t(x_t) - \ell_t(x_{t+1})]$$

Template bound for FTRL

FTRL regret bound

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

$$\mathrm{Reg}_p(T) \leq \lambda[h(p) - h(x_1)] + \sum_{t=1}^{T} [\ell_t(x_t) - \ell_t(x_{t+1})]$$

Proof.

□

Variability bound for FTRL

Variability of FTRL

- ☞ *Assume:* h is K-strongly convex; each ℓ_t is G_t-Lipschitz continuous
- ✓ **Then:**

$$\ell_t(x_t) - \ell_t(x_{t+1}) \leq G_t \|x_{t+1} - x_t\| \leq G_t^2/(\lambda K)$$

Variability bound for FTRL

Variability of FTRL

☞ *Assume:* h is K-strongly convex; each ℓ_t is G_t-Lipschitz continuous

✓ **Then:**

$$\ell_t(x_t) - \ell_t(x_{t+1}) \leq G_t \|x_{t+1} - x_t\| \leq G_t^2/(\lambda K)$$

Proof.

□

Regret of FTRL

Theorem (Shalev-Shwartz & Singer, 2006; Shalev-Shwartz, 2011)

- **Assume:** *h is K-strongly convex; each ℓ_t is G-Lipschitz continuous*
- ✓ ***Then:*** (FTRL) *enjoys the regret bound*

$$\mathrm{Reg}_p(T) \leq \lambda[h(p) - \min h] + \frac{G^2}{\lambda K} T$$

Regret of FTRL

Theorem (Shalev-Shwartz & Singer, 2006; Shalev-Shwartz, 2011)

☞ **Assume:** *h is K-strongly convex; each ℓ_t is G-Lipschitz continuous*

✓ ***Then:*** (FTRL) *enjoys the regret bound*

$$\mathrm{Reg}_p(T) \leq \lambda[h(p) - \min h] + \frac{G^2}{\lambda K} T$$

Corollary

With assumptions as above, $H = \max h - \min h$ and $\lambda = G\sqrt{T/(2KH)}$, (FTRL) *enjoys the bound*

$$\mathrm{Reg}(T) \leq G\sqrt{(2H/K)\,T} = \mathcal{O}(\sqrt{T})$$

Regret of FTRL

Theorem (Shalev-Shwartz & Singer, 2006; Shalev-Shwartz, 2011)

- **Assume:** *h is K-strongly convex; each ℓ_t is G-Lipschitz continuous*
- ✓ **Then:** (FTRL) *enjoys the regret bound*

$$\mathrm{Reg}_p(T) \leq \lambda[h(p) - \min h] + \frac{G^2}{\lambda K} T$$

Corollary

With assumptions as above, $H = \max h - \min h$ and $\lambda = G\sqrt{T/(2KH)}$, (FTRL) *enjoys the bound*

$$\mathrm{Reg}(T) \leq G\sqrt{(2H/K)\,T} = \mathcal{O}(\sqrt{T})$$

Remarks:

- The bound is tight in T — Abernethy et al., 2008
- Requires full information and tuning in terms of T — # can relax

Outline

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

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 $g_t \in \partial \ell_t(x_t)$ # 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 $b_t = \mathbb{E}[\hat{g}_t \mid \mathcal{F}_t] - v(x_t)$ is the **bias** of $\hat{g}_t$

Follow the linearized leader

Can we relax the full information requirement of FTRL?

- Replace ℓ_t with first-order surrogate

$$\hat{\ell}_t(x) = \ell_t(x_t) + \langle g_t, x - x_t \rangle \qquad g_t \in \partial \ell_t(x_t)$$

- Plug into (FTRL)

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \hat{\ell}_s(x) + \underbrace{\lambda}_{1/\eta} h(x) \right\} = \arg\min_{x \in \mathcal{X}} \left\{ \eta \sum_{s=1}^{t} \langle g_s, x - x_s \rangle + h(x) \right\}$$

Follow the linearized leader

Can we relax the full information requirement of FTRL?

- Replace ℓ_t with first-order surrogate

$$\hat{\ell}_t(x) = \ell_t(x_t) + \langle g_t, x - x_t \rangle \qquad g_t \in \partial \ell_t(x_t)$$

- Plug into (FTRL)

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t} \hat{\ell}_s(x) + \underbrace{\lambda}_{1/\eta} h(x) \right\} = \arg\min_{x \in \mathcal{X}} \left\{ \eta \sum_{s=1}^{t} \langle g_s, x - x_s \rangle + h(x) \right\}$$

- ***Follow the linearized leader (FTLL)***

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{ \eta \sum_{s=1}^{t} \langle g_s, x \rangle + h(x) \right\} \tag{FTLL}$$

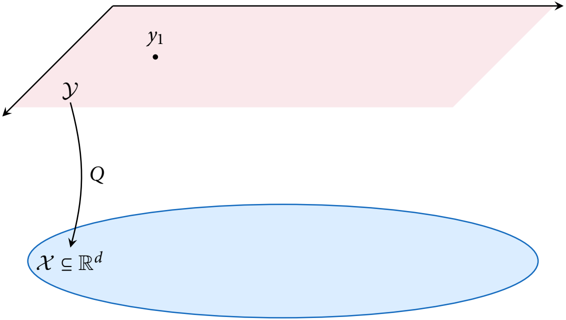
Dual averaging

Dual averaging (DA) formulation of FTLL

Nesterov, 2009; Xiao, 2010

$$y_{t+1} = y_t - \eta g_t$$
$$x_{t+1} = Q(y_{t+1}) \tag{DA}$$

where $Q(y) = \arg\max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$ is the ***mirror map*** associated to h

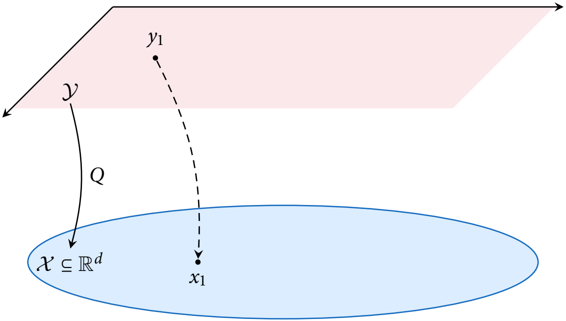


Dual averaging

Dual averaging (DA) formulation of FTLL — Nesterov, 2009; Xiao, 2010

$$\begin{aligned} y_{t+1} &= y_t - \eta g_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{DA}$$

where $Q(y) = \arg\max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$ is the ***mirror map*** associated to h



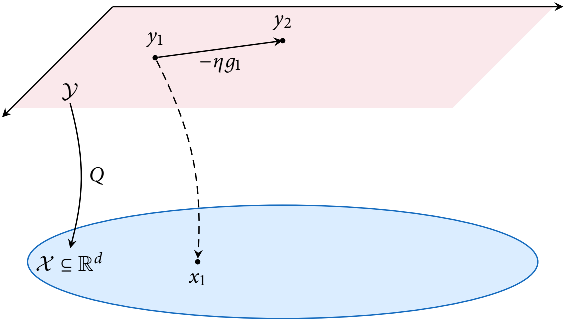
Dual averaging

Dual averaging (DA) formulation of FTLL

Nesterov, 2009; Xiao, 2010

$$\begin{aligned} y_{t+1} &= y_t - \eta g_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{DA}$$

where $Q(y) = \arg\max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$ is the ***mirror map*** associated to h

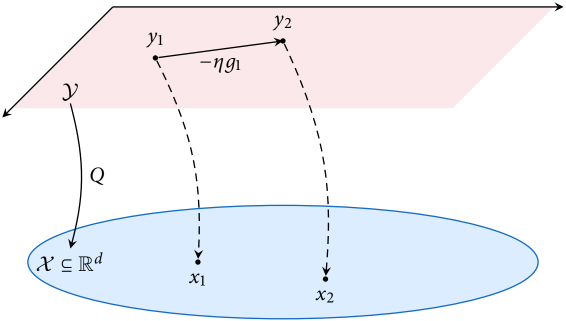


Dual averaging

Dual averaging (DA) formulation of FTLL — Nesterov, 2009; Xiao, 2010

$$\begin{aligned} y_{t+1} &= y_t - \eta g_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{DA}$$

where $Q(y) = \arg\max_{x\in\mathcal{X}}\{\langle y, x\rangle - h(x)\}$ is the ***mirror map*** associated to h

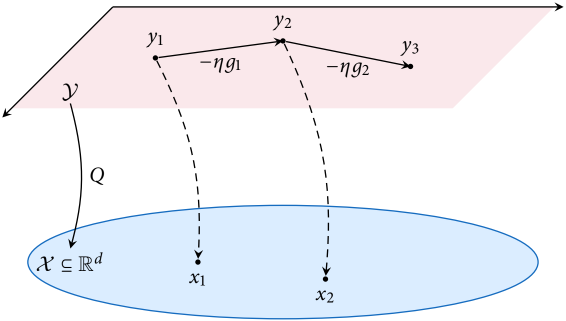


Dual averaging

Dual averaging (DA) formulation of FTLL ➾ Nesterov, 2009; Xiao, 2010

$$\begin{aligned} y_{t+1} &= y_t - \eta g_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{DA}$$

where $Q(y) = \arg\max_{x\in\mathcal{X}}\{\langle y, x\rangle - h(x)\}$ is the ***mirror map*** associated to h

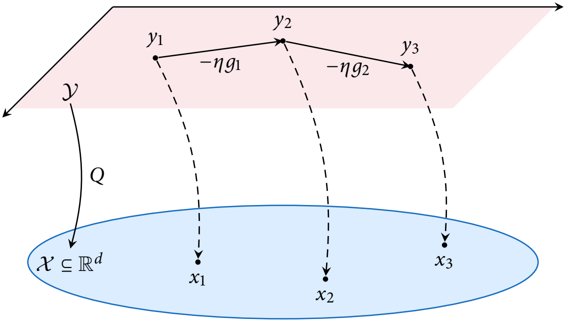


Dual averaging

Dual averaging (DA) formulation of FTLL ⇨ Nesterov, 2009; Xiao, 2010

$$\begin{aligned} y_{t+1} &= y_t - \eta g_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{DA}$$

where $Q(y) = \arg\max_{x\in\mathcal{X}}\{\langle y, x\rangle - h(x)\}$ is the ***mirror map*** associated to h



Example: online gradient descent

Special case when $h(x) = (1/2)\|x\|_2^2 \rightsquigarrow$ ***online gradient descent (OGD)*** # lazy version

$$y_{t+1} = y - \eta g_t \qquad x_{t+1} = \Pi(y_{t+1}) \tag{OGD}$$

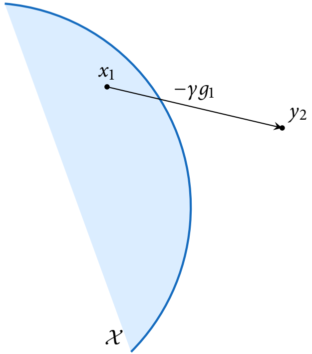


Figure: Schematics of (OGD)

Example: online gradient descent

Special case when $h(x) = (1/2)\|x\|_2^2 \rightsquigarrow$ ***online gradient descent (OGD)*** # lazy version

$$y_{t+1} = y - \eta g_t \qquad x_{t+1} = \Pi(y_{t+1}) \tag{OGD}$$

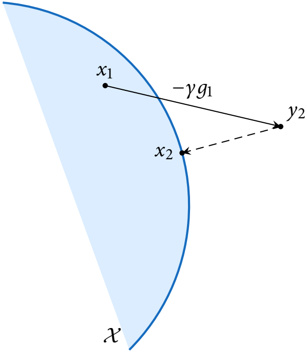


Figure: Schematics of (OGD)

Example: online gradient descent

Special case when $h(x) = (1/2)\|x\|_2^2 \rightsquigarrow$ ***online gradient descent (OGD)*** # lazy version

$$y_{t+1} = y - \eta g_t \qquad x_{t+1} = \Pi(y_{t+1}) \tag{OGD}$$

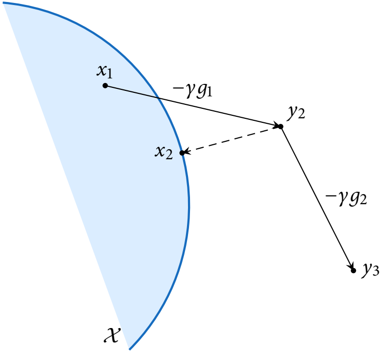


Figure: Schematics of (OGD)

Example: online gradient descent

Special case when $h(x) = (1/2)\|x\|_2^2 \rightsquigarrow$ ***online gradient descent (OGD)*** # lazy version

$$y_{t+1} = y - \eta g_t \qquad x_{t+1} = \Pi(y_{t+1}) \tag{OGD}$$

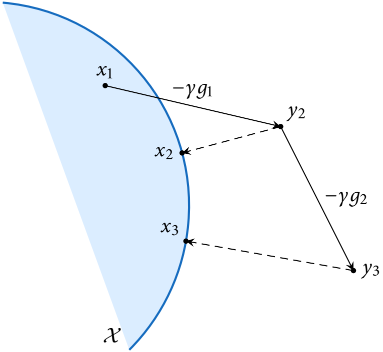


Figure: Schematics of (OGD)

Online mirror descent (deep dive)

- Gradient signals enter (DA) unweighted / unadjusted # post-adaptation
- Variable weights $\rightsquigarrow$ "lazy", primal-dual variant of **online mirror descent**

$$\begin{aligned} y_{t+1} &= y_t + \eta_t \hat{g}_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{OMD$_{\text{lazy}}$}$$

- Primal-primal ("eager") variant of (OMD$_{\text{lazy}}$)

$$x_{t+1} = P_{x_t}(\eta_t \hat{g}_t) \tag{OMD}$$

with the **Bregman proximal mapping** P defined as

$$P_x(w) = \arg\min_{x' \in \mathcal{X}} \{ \langle w, x - x' \rangle + D(x', x) \}$$

where $D(x', x) = h(x') - h(x) - \langle \nabla h(x'), x - x' \rangle$ is the **Bregman divergence** of h

Online mirror descent (deep dive)

- Gradient signals enter (DA) unweighted / unadjusted # post-adaptation
- Variable weights $\rightsquigarrow$ "lazy", primal-dual variant of **online mirror descent**

$$\begin{aligned} y_{t+1} &= y_t + \eta_t \hat{g}_t \\ x_{t+1} &= Q(y_{t+1}) \end{aligned} \tag{OMD$_{\text{lazy}}$}$$

- Primal-primal ("eager") variant of (OMD$_{\text{lazy}}$)

$$x_{t+1} = P_{x_t}(\eta_t \hat{g}_t) \tag{OMD}$$

with the **Bregman proximal mapping** P defined as

$$P_x(w) = \arg\min_{x' \in \mathcal{X}} \{\langle w, x - x' \rangle + D(x', x)\}$$

where $D(x', x) = h(x') - h(x) - \langle \nabla h(x'), x - x' \rangle$ is the **Bregman divergence** of h

Proposition

The iterates of (OMD$_{\text{lazy}}$) *and* (OMD) *coincide whenever* $\operatorname{dom} \partial h = \operatorname{ri} \mathcal{X}$

Regret under dual averaging

- ***Gradient trick:*** # linear model

$$\ell_t(x_t) - \ell_t(p) \leq \langle g_t, x_t - p \rangle \quad \text{for all } p \in \mathcal{X}$$

Regret under dual averaging

- ***Gradient trick:*** # linear model

$$\ell_t(x_t) - \ell_t(p) \leq \langle g_t, x_t - p \rangle \quad \text{for all } p \in \mathcal{X}$$

- ***Energy function:*** ⚠ take for granted

$$F_t = h(p) + h^*(y_t) - \langle y_t, p \rangle$$

where $h^*(y) = \max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$ is the ***potential*** of $Q \rightsquigarrow \nabla h^* = Q$

Regret under dual averaging

- ***Gradient trick:*** # linear model

$$\ell_t(x_t) - \ell_t(p) \leq \langle g_t, x_t - p \rangle \quad \text{for all } p \in \mathcal{X}$$

- ***Energy function:*** ⚠ take for granted

$$F_t = h(p) + h^*(y_t) - \langle y_t, p \rangle$$

where $h^*(y) = \max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$ is the **potential** of $Q \rightsquigarrow \nabla h^* = Q$

- ***Template inequality:*** ⚠ take for granted

$$F_{t+1} \leq F_t - \eta \langle g_t, x_t - p \rangle + \frac{\eta^2}{2K} \|g_t\|^2$$

Regret under dual averaging

- ***Gradient trick:*** # linear model

$$\ell_t(x_t) - \ell_t(p) \leq \langle g_t, x_t - p \rangle \quad \text{for all } p \in \mathcal{X}$$

- ***Energy function:*** ⚠ take for granted

$$F_t = h(p) + h^*(y_t) - \langle y_t, p \rangle$$

where $h^*(y) = \max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$ is the ***potential*** of $Q \rightsquigarrow \nabla h^* = Q$

- ***Template inequality:*** ⚠ take for granted

$$F_{t+1} \leq F_t - \eta \langle g_t, x_t - p \rangle + \frac{\eta^2}{2K} \|g_t\|^2$$

- ***Rearrange & telescope:*** # build the regret

$$\overline{\operatorname{Reg}}(T) \leq \frac{H}{\eta} + \frac{\eta}{2K} \sum_{t=1}^{T} G_t^2$$

Regret under dual averaging, cont'd

- Take $\eta = \sqrt{2KH / \sum_{t=1}^{T} G_t^2}$

⚠ Why?

$$\mathrm{Reg}(T) \leq \sqrt{(2H/K) \sum_{t=1}^{T} G_t^2}$$

Regret under dual averaging, cont'd

- Take $\eta = \sqrt{2KH / \sum_{t=1}^{T} G_t^2}$ ⚠ Why?

$$\mathrm{Reg}(T) \leq \sqrt{(2H/K) \sum_{t=1}^{T} G_t^2}$$

Theorem (Shalev-Shwartz, 2011)

☞ **Assume:** *h is K-strongly convex; each ℓ_t is G-Lipschitz continuous; $H = \max h - \min h$ and $\eta = G^{-1}\sqrt{2KH/T}$*

✓ ***Then:*** (DA) / (FTLL) *enjoys the regret bound*

$$\mathrm{Reg}_p(T) \leq G\sqrt{(2H/K)T}$$

Outline

Oracle feedback

The oracle model

A ***stochastic first-order oracle (SFO)*** model of g_t is a random vector $\hat{g}_t$ of the form

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

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

Oracle feedback

The oracle model

A ***stochastic first-order oracle (SFO)*** model of g_t is a random vector $\hat{g}_t$ of the form

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

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

Assumptions

- ***Bias:*** $\|b_t\|_\infty \leq B_t$
- ***Variance:*** $\mathbb{E}[\|U_t\|_\infty^2 \mid \mathcal{F}_t] \leq \sigma_t^2$
- ***Second moment:*** $\mathbb{E}[\|\hat{g}_t\|_\infty^2 \mid \mathcal{F}_t] \leq M_t^2$

Oracle feedback

The oracle model

A ***stochastic first-order oracle (SFO)*** model of g_t is a random vector $\hat{g}_t$ of the form

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

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

Algorithm Stochastic gradient descent (SGD) # OGD with stochastic feedback

---

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

Initialize: $y_1 \in \mathbb{R}^{\mathcal{A}}$

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

  play $x_t \leftarrow \Pi(y_t)$ # action selection

  incur $c_t = \ell_t(x_t)$ # incur cost

  observe estimate $\hat{g}_t$ of $g_t \in \partial \ell_t(x_t)$ # SFO feedback

  set $y_{t+1} \leftarrow y_t - \eta_t \hat{g}_t$ # update state

end for

Regret under OGD

- **Gradient trick:** # linear model

$$\ell_t(x_t) - \ell_t(p) \leq \langle g_t, x_t - p \rangle \quad \text{for all } p \in \mathcal{X}$$

- **Energy function:** # as before

$$F_t = \tfrac{1}{2}\|y_t - p\|^2 - \tfrac{1}{2}\|y_t - x_t\|^2$$

- **Energy inequality:** # $\hat{g}_t$ instead of g_t

$$F_{t+1} \leq F_t - \eta\langle \hat{g}_t, x_t - p \rangle + \frac{\eta^2}{2}\|\hat{g}_t\|^2$$

- **Expand and rearrange:**

$$\langle v_t, p - x_t \rangle \leq \frac{F_t - F_{t+1}}{\eta} - \langle U_t, x_t - p \rangle - \langle b_t, x_t - p \rangle + \frac{\eta}{2}\|\hat{g}_t\|_\infty^2$$

- **How to proceed?**

Regret analysis, cont'd

Bound each term separately:

Regret of SGD

Theorem

☞ *Assume:*

- *feedback of the form* (SFO)
- $\eta = \operatorname{diam}(\mathcal{X}) \Big/ \sqrt{\sum_{t=1}^{T} M_t^2}$

✓ *Then: for all $p \in \mathcal{X}$, the SGD algorithm enjoys the bound*

$$\mathbb{E}[\operatorname{Reg}_p(T)] \leq 2 \sum_{t=1}^{T} B_t + \operatorname{diam}(\mathcal{X}) \sqrt{\sum_{t=1}^{T} M_t^2}$$

Regret of SGD

Theorem

☞ *Assume:*

- *feedback of the form* (SFO)
- $\eta = \text{diam}(\mathcal{X}) \big/ \sqrt{\sum_{t=1}^{T} M_t^2}$

✓ *Then: for all $p \in \mathcal{X}$, the SGD algorithm enjoys the bound*

$$\mathbb{E}[\text{Reg}_p(T)] \leq 2 \sum_{t=1}^{T} B_t + \text{diam}(\mathcal{X}) \sqrt{\sum_{t=1}^{T} M_t^2}$$

Remarks:

- $\mathcal{O}(\sqrt{T})$ regret if feedback is unbiased ($b_t = 0$) and has finite variance ($M_t \leq M$)
- This bound is tight in T — Abernethy et al., 2008

Stochastic convex optimization

Stochastic convex optimization

$$\begin{aligned} \text{minimize} \quad & f(x) = \mathbb{E}_{\omega\sim P}[F(x;\omega)] \\ \text{subject to} \quad & x \in \mathcal{X} \end{aligned} \tag{Opt-S}$$

Stochastic convex optimization

Stochastic convex optimization

$$\begin{aligned} &\text{minimize} \quad f(x) = \mathbb{E}_{\omega \sim P}[F(x;\omega)] \\ &\text{subject to} \quad x \in \mathcal{X} \end{aligned} \tag{Opt-S}$$

- Important for data science ⇝ ***finite-sum objectives:***

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x)$$

- Special case of OCO:

$$\ell_t \leftarrow f \quad \text{for all } t = 1, 2, \dots$$

- Access to ***stochastic gradients***

$$\hat{g}_t \leftarrow \nabla F(x_t;\omega_t) \quad \text{with } \omega_t \text{ drawn i.i.d. from } P$$

Convergence rate of SGD

Theorem

☞ **Assume:** $\mathbb{E}[\|\hat{g}_t\|^2] \leq M^2$ *and SGD is run for* T *iterations with* $\eta = \operatorname{diam}(\mathcal{X})/(M\sqrt{T})$

✓ **Then:** *the ergodic average* $\bar{x}_T = (1/T)\sum_{t=1}^{T} x_t$ *of SGD enjoys the rate*

$$\mathbb{E}[f(\bar{x}_T) - \min f] \leq \frac{M \operatorname{diam}(\mathcal{X})}{\sqrt{T}}$$

Convergence rate of SGD

Theorem

☞ ***Assume:*** $\mathbb{E}[\|\hat{g}_t\|^2] \leq M^2$ *and SGD is run for* T *iterations with* $\eta = \operatorname{diam}(\mathcal{X})/(M\sqrt{T})$

✓ ***Then:*** *the ergodic average* $\bar{x}_T = (1/T)\sum_{t=1}^{T} x_t$ *of SGD enjoys the rate*

$$\mathbb{E}[f(\bar{x}_T) - \min f] \leq \frac{M \operatorname{diam}(\mathcal{X})}{\sqrt{T}}$$

Proof.

□

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.