On Adaptive Strategies and Convex Optimization Algorithms

Joon Kwon
joint work with Panayotis Mertikopoulos

Institut de Mathématiques de Jussieu
Université Pierre-et-Marie-Curie
Paris, France

Workshop on Algorithms and Dynamics for Games and Optimization
Playa Blanca, Tongoy, Chile

October 2013

Framework

$(V, \|\cdot\|)$ a normed space of finite dimension and $(V^*, \|\cdot\|_*)$ its dual
$C \subset V$ a convex compact set
Nature chooses a sequence $u_1, \ldots, u_n, \ldots \in V^*$

- choose $x_1 \in C$
- u_1 is revealed
- get payoff $\langle u_1 | x_1 \rangle$

  $\vdots$

- A stage $n+1$, knowing $u_1, \ldots, u_n$ choose $x_{n+1} \in C$
- u_{n+1} is revealed
- get payoff $\langle u_{n+1} | x_{n+1} \rangle$

$$\sigma_{n+1} : \begin{array}{ccc} (V^*)^n & \longrightarrow & C \\ (u_1 \ldots, u_n) & \longmapsto & x_{n+1} \end{array} \qquad \sigma = (\sigma_n)_{n \geqslant 1}$$

$$\textit{maximize} \qquad \sum_{k=1}^{n} \langle u_k | x_k \rangle$$

The Case of the simplex

- $V = V^* = \mathbb{R}^d$
- $C = \Delta_d = \left\{ x \in \mathbb{R}_+^d \,\middle|\, \sum_{i=1}^{d} x_i = 1 \right\} \quad \rightsquigarrow \quad$ prob. dist. on $\{1, \ldots, d\}$

- Choose $x_{n+1} \in \Delta_d$,
- Draw $i_{n+1} \in \{1, \ldots, d\}$ according to x_{n+1},
- Get payoff $u_{n+1}^{i_{n+1}}$.

$$\mathbb{E}\left[\sum_{k=1}^{n} u_k^{i_k}\right] = \sum_{k=1}^{n} \langle u_k | x_k \rangle$$

The Regret

Wish: A strategy σ such that:

$$\forall (u_n)_{n\geqslant 1}, \quad \limsup_{n\to+\infty} \left[\frac{1}{n} \left(\underbrace{\max_{x\in C} \sum_{k=1}^{n} \langle u_k | x \rangle - \sum_{k=1}^{n} \langle u_k | x_k \rangle}_{\text{Regret}} \right) \right] \leqslant 0$$

Speed of convergence?

Extension to convex losses

- $\ell_n : C \longrightarrow \mathbb{R}$ convex loss functions
- Loss: $\ell_n(x_n)$

$$\begin{aligned}
\sum_{k=1}^{n} \ell_k(x_k) - \min_{x \in C} \sum_{k=1}^{n} \ell_k(x) &= \max_{x \in C} \sum_{k=1}^{n} (\ell_k(x_k) - \ell_k(x)) \\
&\leqslant \max_{x \in C} \sum_{k=1}^{n} \langle \nabla \ell_k(x_k) | x_k - x \rangle \\
&= \max_{x \in C} \sum_{k=1}^{n} \langle -\nabla \ell_k(x_k) | x \rangle - \sum_{k=1}^{n} \langle -\nabla \ell_k(x_k) | x_k \rangle \\
&= \max_{x \in C} \sum_{k=1}^{n} \langle u_k | x \rangle - \sum_{k=1}^{n} \langle u_k | x_k \rangle
\end{aligned}$$

$$u_n = -\nabla \ell_n(x_n)$$

Convex optimization

- $f : C \longrightarrow \mathbb{R}$ convex function

$$\ell_n = f$$

$$\frac{1}{n}\sum_{k=1}^{n}\ell_k(x_k) - \min_{x\in C}\frac{1}{n}\sum_{k=1}^{n}\ell_k(x) = \frac{1}{n}\sum_{k=1}^{n}f(x_k) - \min_{x\in C}f(x)$$

A Family of strategies

$$u_1, u_2, \ldots, u_n \in V^*$$

$$\downarrow$$

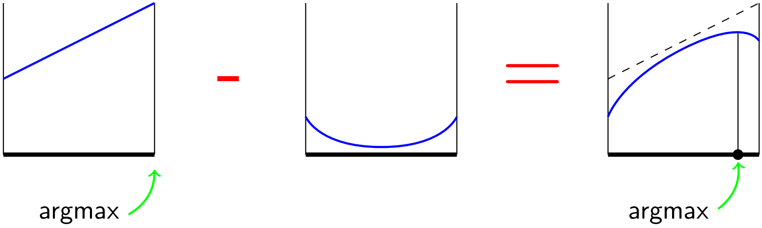
$$\sum_{k=1}^{n} u_k \in V^*$$

$$\downarrow$$

$$x_{n+1} = Q\left(\sum_{k=1}^{n} u_k\right)$$

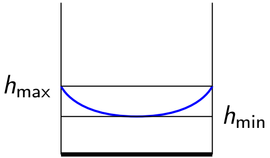
$$(Q : V^* \longrightarrow C)$$

$$\begin{array}{rcll} Q: & V^* & \longrightarrow & C \\ & y & \longmapsto & \underset{x\in C}{\arg\max}\,\{\langle y|x\rangle - h(x)\} \end{array}$$



$h : C \longrightarrow \mathbb{R}$ convex

- continous $\rightsquigarrow$ $Q_h(y)$ exists
- strictly convex $\rightsquigarrow$ $Q_h(y)$ is unique



$$x_{n+1} = Q_h\left(\eta_n \sum_{k=1}^{n} u_k\right) = Q_h(y_n) \quad \eta_n > 0 \text{ and } \searrow$$

Some known strategies and algorithms

- Exponential Weight Algorithm (EWA)
- $1/\sqrt{n}$-Exponential Weight Algorithm ($1/\sqrt{n}$-EWA)
- Vanishingly Smooth Fictitious Play (VSFP)
- Smooth Fictitious Play (SFP)
- Projected Subgradient Method (PSM)
- Mirror Descent (MD)
- Online Gradient Descent (OGD)
- Online Mirror Descent (OMD)
- Follow the Regularized Leader (FRL)

Exponential Weight Algorithm

- $C = \Delta_d$

$$x_{n+1,i} = \frac{\exp\left(\eta \sum_{k=1}^{n} u_{k,i}\right)}{\sum_{j=1}^{d} \exp\left(\eta \sum_{k=1}^{n} u_{k,j}\right)}.$$

$$h(x) = \sum_{i=1}^{d} x_i \log x_i \quad \longrightarrow \quad Q_h(y)_i = \frac{e^{y_i}}{\sum_{j=1}^{d} e^{y_j}}$$

$$x_{n+1} = Q_h\left(\eta \sum_{k=1}^{n} u_k\right)$$

Projected Subgradient Method

$$\begin{cases} y_n = -\sum_{k=1}^{n} \gamma_k \nabla f(x_k) \\ x_{n+1} = \arg\min_{x \in C} \|x - y_n\|_2 . \end{cases}$$

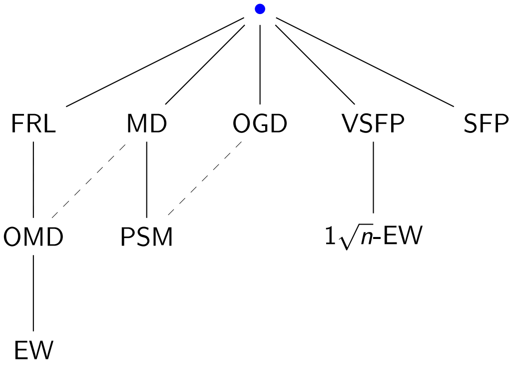
$$\begin{aligned} x_{n+1} &= \arg\min_{x \in C} \|x - y_n\|_2^2 \\ &= \arg\min_{x \in C} \left\{ \|x\|_2^2 - 2 \langle y_n | x \rangle + \|y_n\|_2^2 \right\} \\ &= \arg\max_{x \in C} \left\{ \langle y_n | x \rangle - \frac{1}{2} \|x\|_2^2 \right\} \end{aligned}$$

$$h(x) = \frac{1}{2} \|x\|_2^2$$

$$u_n = -\gamma_n \nabla f(x_n)$$

Name	C	h	η_n	u_n	$\|\cdot\|$	References
EW	Δ_d	$\sum_{i=1}^{d} x_i \log x_i$	η	–	$\|\cdot\|_1$	Littlestone, Warmuth 1994 Sorin 2009
$1/\sqrt{n}$-EW	Δ_d	$\sum_{i=1}^{d} x_i \log x_i$	$\frac{\eta}{\sqrt{n}}$	–	$\|\cdot\|_1$	Auer, Cesa-Bianchi, Gentile 2002
VSFP	Δ_d	any	ηn^{α} $\alpha \in (-1, 0)$	–	$\|\cdot\|_1$	Benaïm, Faure 2013
SFP	Δ_d	any	$\frac{\eta}{n}$	–	$\|\cdot\|_1$	Fudenberg, Levine 1995 Benaïm, Hofbauer, Sorin 2006
PSM	any	$\frac{1}{2}\|\cdot\|_2^2$	1	$-\gamma_n \nabla f(x_n)$	$\|\cdot\|_2$	Polyak 69?
MD	any	any	1	$-\gamma_n \nabla f(x_n)$	any	Nemirovski, Yudin 1983 Beck, Teboulle 2003
OGD	any	$\frac{1}{2}\|\cdot\|_2^2$	1	$-\gamma_n \nabla f_n(x_n)$	$\|\cdot\|_2$	Zinkevich 2003
OMD	any	any	η	$-\nabla f_n(x_n)$	any	Shalev-Shwartz 2007
FRL	any	any	η	–	any	Shalev-Shwartz 2007

Interrelations



The Continuous-Time Counterpart

$$u : \begin{array}{ccc} \mathbb{R}_+ & \longrightarrow & V^* \\ t & \longmapsto & u_t \end{array} \text{ meas.} \qquad \eta : \begin{array}{ccc} \mathbb{R}_+ & \longrightarrow & \mathbb{R}_+^* \\ t & \longmapsto & \eta_t \end{array} \text{ cont.,} \searrow$$

$$x_{n+1} = Q_h\left(\eta_n \sum_{k=1}^{n} u_k\right)$$

$$\tilde{x}_t = Q_h\left(\eta_t \int_0^t u_s \, ds\right) = Q_h(y_t)$$

Theorem

$\forall (u_t)_{t \in \mathbb{R}_+}$,

$$\forall t \geqslant 0, \qquad \max_{x \in C} \int_0^t \langle u_s | x \rangle \, ds - \int_0^t \langle u_s | \tilde{x}_s \rangle \, ds \leqslant \frac{h_{max} - h_{min}}{\eta_t}$$

The Analysis

$$\max_{x \in C} \int_0^t \langle u_s | x \rangle \, ds - \int_0^t \langle u_s | \tilde{x}_s \rangle \, ds \leqslant \frac{h_{\max} - h_{\min}}{\eta_t}$$

$$\begin{aligned}
\int_0^t \langle u_s | x \rangle \, ds = \frac{1}{\eta_t} \langle y_t | x \rangle &\leqslant \frac{h^*(y_t)}{\eta_t} + \frac{h(x)}{\eta_t} \\
&\leqslant \frac{h^*(0)}{\eta_0} + \int_0^t \underbrace{\frac{d}{ds}\left(\frac{h^*(y_s)}{\eta_s}\right)}_{\leqslant \langle u_s | \tilde{x}_s \rangle + h_{\min} \dot{\eta}_s / \eta_s^2} \, ds + \frac{h_{\max}}{\eta_t} \\
&\leqslant \frac{-h_{\min}}{\eta_0} + \int_0^t \langle u_s | \tilde{x}_s \rangle \, ds + h_{\min}\left(-\frac{1}{\eta_t} + \frac{1}{\eta_0}\right) + \frac{h_{\max}}{\eta_t} \\
&\leqslant \int_0^t \langle u_s | \tilde{x}_s \rangle \, ds + \frac{h_{\max} - h_{\min}}{\eta_t}
\end{aligned}$$

Back to Discrete Time

$$\max_{x \in C} \int_0^t \langle u_s | x \rangle \; ds - \int_0^t \langle u_s | \tilde{x}_s \rangle \; ds \leqslant \frac{h_{\max} - h_{\min}}{\eta_t}$$

$$(u_n)_{n \geqslant 1}, \quad h, \quad (\eta_n)_{n \geqslant 1}$$

$$\begin{cases} x_{n+1} = Q_h(y_n) \\ y_n = \eta_n \sum_{k=1}^{n} u_k \end{cases}$$

$$\max_{x \in C} \sum_{k=1}^{n} \langle u_k | x \rangle - \sum_{k=1}^{n} \langle u_k | x_k \rangle \leqslant \quad ?$$

$$\int_0^n \left\langle u_t | \tilde{x}_{\lfloor t \rfloor} \right\rangle \; dt$$

$u_t = u_{\lceil t \rceil}, \quad \eta_t$ cont. interp. of η_n

$$\begin{cases} \tilde{x}_t = Q_h(y_t) \\ y_t = \eta_t \int_0^t u_s \; ds \end{cases}$$

$$\max_{x \in C} \int_0^n \langle u_t | x \rangle \; dt - \int_0^n \langle u_t | \tilde{x}_t \rangle \; dt$$

$$\int_0^n \langle u_t | \tilde{x}_t \rangle \; dt$$

$$
\begin{aligned}
\left|\left\langle u_s \middle| \tilde{x}_{\lfloor s \rfloor} \right\rangle - \langle u_s | \tilde{x}_s \rangle\right| &= \left|\left\langle u_s \middle| \tilde{x}_{\lfloor s \rfloor} - \tilde{x}_s \right\rangle\right| \\
&\leqslant \qquad \left\| \tilde{x}_{\lfloor s \rfloor} - \tilde{x}_s \right\| \\
\\
&\leqslant \left\| Q_h(y_{\lfloor s \rfloor}) - Q_h(y_s) \right\| \\
&\leqslant K \left\| y_s - y_{\lfloor s \rfloor} \right\|_* \\
&\leqslant K \left\| \int_{\lfloor s \rfloor}^{s} \eta_v \int_0^v u + (-\dot{\eta}_v) u_v \, dv \right\|_* \\
&\leqslant K(\eta_s - s\dot{\eta}_s)
\end{aligned}
$$

$$Q_h = \nabla h^*$$

$$\nabla h^* \; K\text{-Lipschitz} \Longleftrightarrow h \; \frac{1}{K}\text{-strongly convex}$$

Definition

f is C-strongly convex wrt $\|\cdot\|$ if $\forall x, y, \; \forall \lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y) - \frac{C}{2}\lambda(1-\lambda)\left\|y - x\right\|^2$$

$$\sum_{i=1}^{d} x^i \log x^i \quad \text{is 1-strongly convex wrt } \|\cdot\|_1$$

$$\frac{1}{2}\|\cdot\|_2^2 \quad \text{is 1-strongly convex wrt } \|\cdot\|_2$$

Theorem

1. *h K-strongly convex on C wrt* $\|\cdot\|$
2. $(\eta_n)_{n\geqslant 1}$ *positive and nonincreasing*
3. η_t *a continuous and nonincresing interpolation*
4. $x_{n+1} = Q_h\left(\eta_n \sum_{k=1}^{n} u_k\right)$

Then, for every sequence $\|u_n\|_* \leqslant M$,

$$\max_{x\in C} \sum_{k=1}^{n} \langle u_k|x_k\rangle - \sum_{k=1}^{n} \langle u_k|x_k\rangle \leqslant \frac{h_{max} - h_{min}}{\eta_n} + \frac{M^2}{K}\int_0^n (\eta_t - t\dot{\eta}_t)\, dt.$$

Name	Assumption	Bound on the regret
EW	$\|u_n\|_\infty \leqslant 1$	$\dfrac{\log d}{\eta} + \eta n$
$1/\sqrt{n}$-EW	$\|u_n\|_\infty \leqslant 1$	$\left(\dfrac{\log d}{\eta} + 3\eta\right)\sqrt{n}$
VSFP	$\|u_n\|_\infty \leqslant 1$	$\dfrac{h_{\max} - h_{\min}}{\eta} n^{-\alpha} + \dfrac{\eta(1-\alpha)}{C(1+\alpha)} n^{\alpha+1}$
SFP	$\|u_n\|_\infty \leqslant 1$	$\dfrac{h_{\max} - h_{\min}}{\eta} n + \dfrac{\eta(1+\log n)}{K}$
PSM	$\|\nabla f\|_2 \leqslant M$	$\dfrac{\|C\|^2/2 + M^2 \sum_{k=1}^n \gamma_k^2}{\sum_{k=1}^n \gamma_k}$
MD	$\|\nabla f\|_* \leqslant M$	$\dfrac{h_{\max} - h_{\min} + M^2/(2K)\sum_{k=1}^n \gamma_k^2}{\sum_{k=1}^n \gamma_k}$
OGD	$\|\nabla f_n\|_2 \leqslant M$	$\dfrac{\|C\|^2/2 + M^2 \sum_{k=1}^n \gamma_k^2}{\sum_{k=1}^n \gamma_k}$
OMD	$\|\nabla f_n\|_* \leqslant M$	$\dfrac{h_{\max} - h_{\min}}{\eta} + \dfrac{\eta M^2}{K} n$
FRL	$\|u_n\|_* \leqslant M$	$\dfrac{h_{\max} - h_{\min}}{\eta} + \dfrac{\eta M^2}{K} n$