Online Linear Optimization
— Full feedback and Bandit Feedback

Source: Bandit Algorithms, L. & S. Chap 28.

Setting: Evolves over time steps $t = 1, 2, \ldots, n$.
At each time:

Player: Chooses action $a_t \in A \subseteq \mathbb{R}^d$ (A a convex set).

Adversary: Chooses loss $y_t \in \mathcal{Y} \subseteq \mathbb{R}^d$. y_t can depend on the player policy (same as the usual adversarial setting).

Observation: (Full feedback Setting, for now).

loss incurred at time $t = \langle a_t, y_t \rangle$

Player also gets to observe y_t

Regret (w.r.t. best fixed action in hindsight):

$$R_n = \max_{a \in A} \sum_{t=1}^{n} \langle a_t - a, y_t \rangle.$$

Note: Action is not randomized in the full feedback setting.

Two Related Algorithms: Mirror Descent and Follow the Regularized Leader (FTRL).

Mirror Descent: Parameters: n (time horizon), A (set of actions, $A \subseteq \mathbb{R}^d$), η (learning rate), $F: \mathbb{R}^d \to \mathbb{R}$, with domain$(F) = \Omega$ (Legendre function).

F: potential function

Initial: $a_1 = \underset{a \in A}{\arg\min}\ F(a)$.

$t \geq 1$: $a_{t+1} = \underset{a \in A}{\operatorname{argmin}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$

typically, A is a convex set

D_F: Bregman Divergence

FTRL (Follow the Regularized Leader): (parameters as above).

Initial: $a_1 = \underset{a \in A}{\operatorname{argmin}}\ F(a)$.

$t \geq 1$: $a_{t+1} = \underset{a \in A}{\operatorname{argmin}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$

Note 1: $g(x) = b^T x$ (linear function)
$\Rightarrow D_g(x,y) = 0$

FTRL implementation

$\therefore$ F Legendre, g linear $\Rightarrow D_{F+g}(x,y) = D_F(x,y)$.

Thus, for A convex, $\mathcal{D} = \mathbb{R}^d$, FTRL is equivalent to:

$$\tilde{a}_{t+1} = \underset{a}{\operatorname{argmin}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right).$$

more carefully, $a \in \mathcal{D}$

$$a_{t+1} = \underset{a \in A}{\operatorname{argmin}} D_F(a, \tilde{a}_{t+1})$$

if $\mathcal{D} = \mathbb{R}^d$, this reduces to the unconstrained problem.

Simplifying: (taking grad. and setting to zero).

$$\tilde{a}_{t+1} = \underset{a}{\operatorname{argmin}} \left(\langle a, \eta \sum_{s=1}^{t} y_s \rangle + F(a) \right).$$

$\therefore$ $\eta \sum_{s=1}^{t} y_s + \nabla F(\tilde{a}_{t+1}) = 0$

$$\Rightarrow \quad \tilde{a}_{t+1} = \nabla F^{-1}\left(-\eta \sum_{s=1}^{t} y_s\right)$$

$$= \nabla F^*\left(-\eta \sum_{s=1}^{t} y_s\right).$$

$$\boxed{a_{t+1} = \Pi_{F,A}\left(\nabla F^*\left(-\eta \sum_{s=1}^{t} y_s\right)\right)} \qquad \text{FTRL}$$

↳ Bregman Projection

Note 2: $\tilde{a}_{t+1} = \underset{a}{\operatorname{argmin}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$ — Mirror Descent Implementation

more carefully, $a \in D$

$$a_{t+1} = \underset{a \in A}{\operatorname{argmin}} \; D_F(a, \tilde{a}_{t+1})$$

Simplify: (Take gradient, and set to zero)

$$\eta y_t + \nabla F(\tilde{a}_{t+1}) - \nabla F(a_t) = 0$$

$$\Rightarrow \quad \tilde{a}_{t+1} = \nabla F^*\left(\nabla F(a_t) - \eta y_t\right).$$

$$\boxed{a_{t+1} = \Pi_{F,A}\left(\nabla F^*\left(\nabla F(a_t) - \eta y_t\right)\right)}$$

Note 3: Follow the Leader (FTL):

FTL is bad.

$$a_{t+1} = \underset{a \in A}{\operatorname{argmin}} \sum_{s=1}^{t} \langle a, y_t \rangle$$

This is a bad algorithm, because a_t can vary too much over time steps chasing the best action-to-date in hindsight, and can result in large regret. To see this,

$A = [-1, 1]$, $y_1 = \frac{1}{2}$, $y_2 = y_4 = y_6 = \cdots = -1$

$a_1 = 0$ (arbitrary) $y_3 = y_5 = y_7 = \cdots = +1$

$$a_2 = \underset{a \in [-1,1]}{\operatorname{argmin}} \langle a, y_1 \rangle = -1.$$ ($\langle a, y_1 \rangle = \frac{1}{2}$)

$a = -1$: $\langle a, y_1 \rangle = -\frac{1}{2}$.

$$a_3 = \underset{a \in [-1,1]}{\operatorname{argmin}} \left(\frac{a}{2} - a \right) = +1$$ ($\langle a, y_1 \rangle + \langle a, y_2 \rangle = \left(\frac{a}{2} - a\right) = -\frac{a}{2}$, $a = (+1)$)

$$a_4 = \underset{a \in [-1,1]}{\operatorname{argmin}} \left(\frac{a}{2} - a + a \right) = -1$$ ($\langle a, y_1 \rangle + \langle a, y_2 \rangle + \langle a, y_3 \rangle = \frac{a}{2} - a + a = \frac{a}{2}$, $a_4^* = -1$)

i.e., $\{a_2, a_3, a_4, \ldots\} = \{-1, +1, -1, +1, \ldots\}$

$$\text{Loss} = \langle a_2, y_2 \rangle + \langle a_3, y_3 \rangle + \cdots = \underbrace{1 + 1 + \cdots 1}_{(n-1)}$$

$$= (n-1) \qquad \text{(at time } n\text{)}$$

for the best fixed action: Choose $a = 1 \quad \forall\, t \geq 1$.

$$\text{Loss} = \tfrac{1}{2} - 1 + 1 - 1 + 1 \cdots = \begin{cases} 1/2 & \text{if } n \text{ odd} \\ -1/2 & \text{if is even} \end{cases}$$

$\Rightarrow$ Linear regret w.r.t best arm in hindsight!

Adding a regularizer prevents actions from changing too often, and thus, FTRL has much better regret scaling.

See also: Online Learning and Online Convex Optimization, by Shai Shalev-Shwartz, Found. and Trends in ML, Vol. 4, No. 2, 2011

Note 4: FTRL and Mirror Descent are related, and equivalent when $\mathcal{D} \subseteq \mathcal{A}$.

(Note this is not a typical setting; typically it is the other way around where $\mathcal{D} \supseteq \mathcal{A}$)

In this case, we have $\tilde{a}_{t+1} = a_{t+1}$, because the second step in the implementation (the projection onto A using Bregman divergence) is unneccessary.

In the calculation below, we are further assuming $A = D = \mathbb{R}^d$.

FTRL: $\tilde{a}_{t+1} = \nabla F^*\left(-\eta \sum_{s=1}^{t} y_s\right)$.

Mirror Descent: $\tilde{a}_{t+1} = \nabla F^*\left(\nabla F(a_t) - \eta y_t\right)$.

But $\nabla F(a_t) = \nabla F(\tilde{a}_t)$

$$= \nabla F\Big(\underbrace{\underbrace{\nabla F^*}_{=\nabla F^{-1}}\left(\nabla F(a_{t-1}) - \eta y_{t-1}\right)}_{\text{inverses.}}\Big)$$

$$= \nabla F(a_{t-1}) - \eta y_{t-1}$$

$$\Rightarrow \tilde{a}_{t+1} = \nabla F^*\left(\nabla F(a_{t-1}) - \eta y_{t-1} - \eta y_t\right)$$

$$\vdots$$

$$= \nabla F^*\left(-\eta \sum_{s=1}^{t} y_s\right)$$

($\because \nabla F(a_1) = 0$ because by defn., $a_1 = \operatorname{argmin}_a F(a)$)

Note 5: Exponential weights algorithm.

Suppose $A = \Theta^{d-1}$, the d-dimensional simplex of probability distributions, and $F(x) = \sum_{i=1}^{d}(x_i \ln x_i - x_i)$.

discrete pmf: $\sum_i x_i = 1$, $x_i \geq 0$.

Here $D_F(x,y) = \sum_{i=1}^{d} x_i \ln\left(\frac{x_i}{y_i}\right)$ (KL Divergence)

Mirror Descent: extends D_F to $\mathbb{R}^d_+$ $+ \sum_{i=1}^{d}(y_i - x_i)$

$$\tilde{a}_{t+1} = \operatorname*{argmin}_{a}\left(\eta \langle a, y_t\rangle + D_F(a, a_t)\right)$$

Take derivative and set to 0:

$$0 = \eta y_t + \nabla F(\tilde{a}_{t+1}) - \nabla F(a_t) \longrightarrow (*)$$

$$\nabla F(x)_i = \frac{\partial F(x)}{\partial x_i} = \ln x_i \Rightarrow \nabla F(x) = \ln(x) \quad (x \in \mathbb{R}^d)$$

Observe $(*)$ decouples across coordinates, i.e.,

$$\ln(\tilde{a}_{t+1,i}) = \ln(a_{t,i}) - \eta y_{t,i}$$

$$\Rightarrow \tilde{a}_{t+1,i} = e^{-\eta y_{t,i}} \cdot a_{t,i}, \quad i = 1, 2, \ldots, d.$$

Project: $a_{t+1} = \underset{a \in \Delta}{\operatorname{argmin}} D_F(a, \tilde{a}_{t+1})$

$$= \underset{a \in \Delta}{\operatorname{argmin}} \sum_{i=1}^{d} a_i \ln\left(\frac{a_i}{a_{t,i} e^{-\eta y_{t,i}}}\right)$$

Δ → Δ^{d-1}

constant wrt $a \in \Delta$

$$+ \sum_{i=1}^{d} \tilde{a}_{t+1,i} - \sum_{i=1}^{d} a_i = 1$$

∴ this can be ignored.

Claim: $a_i = \dfrac{a_{t,i} e^{-\eta y_{t,i}}}{\sum_j a_{t,j} e^{-\eta y_{t,j}}}$ is the minimizer to above.

Pf: Let $C = \sum_{j=1}^{d} a_{t,i} e^{-\eta y_{t,j}}$

Then observe that $\sum_{i=1}^{d} a_i \left(\ln\left(\dfrac{a_i}{a_{t,i} e^{-\eta y_{t,i}}}\right)\right)$

$$= \sum_{i=1}^{d} a_i \ln\left(a_i \cdot \frac{C}{a_{t,i} e^{-\eta y_{t,i}}} \cdot \frac{1}{C}\right)$$

$$= \sum_{i=1}^{d} a_i \ln\left(\frac{a_i}{\left(\frac{a_{t,i} e^{-\eta y_{t,i}}}{C}\right)}\right) + \left(\sum_{i=1}^{d} a_i\right) \ln\left(\frac{1}{C}\right)$$

$$= \sum_{i=1}^{d} a_i \ln\left(a_i \Big/ \left(\frac{a_{t,i} e^{-\eta y_{t,i}}}{C}\right)\right) + \ln\left(\frac{1}{C}\right).$$

≥ 0 for any $a_i \in \Delta^{d-1}$

with equality when $a_i = \left(\frac{a_{t,i} e^{-\eta y_{t,i}}}{C} \right)$

$$\therefore \quad a_{t+1,i} = \frac{a_{t,i} e^{-\eta y_{t,i}}}{\sum_j a_{t,i} e^{-\eta y_{t,i}}}, \quad i = 1, \dots d$$

results from Mirror Descent + KL projection on Simplex = Exp3 Rule

---

Note 6: Equivalently, we have for FTRL:

FTRL $\quad \tilde{a}_{t+1} = \underset{a}{\operatorname{argmin}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$.

$$a_{t+1} = \underset{a \in A}{\operatorname{argmin}} \; D_F(a, \tilde{a}_{t+1})$$

$$A = \mathcal{P}^{d-1}, \quad F(x) = \sum_{i=1}^{d} (x_i \ln x_i - x_i)$$

Then, similar calculation as above shows:

$$a_{t+1,i} = \frac{e^{-\eta \sum_{s=1}^{t} y_{s,i}}}{\sum_{j=1}^{d} e^{-\eta \sum_{s=1}^{t} y_{s,j}}}, \quad i=1,2,\dots,d.$$

i.e., we get back Exp3 algorithm ☒

Regret Bound

Mirror Descent:

Thm (28.4 in textbook): $\eta > 0$, and F Legendre, with domain D. A a non-empty convex set, with interior $(D) \cap A \neq \emptyset$. (assumed to exist)

Mirror descent actions: $\{a_1, a_2, \dots, a_{n+1}\}$ with

$$a_1 = \operatorname{argmin}_{a \in A} F(a), \quad \tilde{a}_{t+1} = \operatorname{argmin}_{a \in D} \eta \langle a, y_t \rangle + D_F(a, a_t)$$

$$a_{t+1} = \operatorname{argmin}_{a \in A} D_F(a, \tilde{a}_{t+1}).$$

For any $a \in A$,

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D_F(a_{t+1}, a_t)$$

Further, suppose $\nabla F(a) - \eta y \in \text{interior}\left(\text{Dom}(F^*)\right) \; \forall \; y \in \alpha$, $a \in A \cap D$.

Then,

$$R_n(a) \leq \frac{1}{\eta} \left(F(a) - F(a_1) + \sum_{t=1}^{n} D_F(a_t, \tilde{a}_{t+1}) \right).$$

Proof: Suppose $a \in A$, but $a \notin D$. Then $F(a) = \infty$, thus result holds. Thus, assume $a \in A \cap D$.

Now, by defn, $R_n(a) = \sum_{t=1}^{n} \langle a_t - a, y_t \rangle$.

Further,

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle \longrightarrow (1)$$

Recall: $a_{t+1} = \underset{b \in A}{\text{argmin}} \; \eta \langle b, y_t \rangle + D_F(b, a_t)$.

Further, $D_F(b, a_t) = F(b) - F(a_t) - \langle \nabla F(a_t), b - a_t \rangle$.

Now, $D_F(b, a_t)$ convex in b, and $\eta \langle b, y_t \rangle$ linear in b. Thus, $\eta \langle b, y_t \rangle + D_F(b, a_t)$ convex. From first order optimality, it follows that:

$$\left\langle \nabla \left(\eta \langle b, y_t \rangle + D_F(b, a_t) \right)\Big|_{b = a_{t+1}}, a - a_{t+1} \right\rangle \geq 0$$

for any $a \in A$.

i.e., $$\langle \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t), a - a_{t+1} \rangle \geq 0$$

Reordering this, we have:

$$\langle a_{t+1} - a, y_t \rangle \leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle$$

$$= \frac{1}{\eta} \left(D_F(a, a_t) - D_F(a, a_{t+1}) - D_F(a_{t+1}, a_t) \right)$$

$F(a) - F(a_t) - \langle \nabla F(a_t), a - a_t \rangle$

$F(a) - F(a_{t+1}) - \langle \nabla F(a_{t+1}), a - a_{t+1} \rangle$

$F(a_{t+1}) - F(a_t) - \langle \nabla F(a_t), a_{t+1} - a_t \rangle$

→ (2)

$\therefore$

$$R_n = \sum_{t=1}^{n} \langle a_t - a, y_t \rangle$$

$$\overset{(1)}{=} \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \sum_{t=1}^{n} \langle a_{t+1} - a, y_t \rangle$$

$$\leq \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \sum_{t=1}^{n} \Big(D_F(a, a_t) - D_F(a, a_{t+1}) - D_F(a_{t+1}, a_t) \Big) \quad (2)$$

collect/cancel terms

$$= \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \Big(\underbrace{D_F(a, a_1)}_{\leq F(a) - F(a_1)} - \underbrace{D_F(a, a_{n+1})}_{\geq 0} - \sum_{t=1}^{n} D_F(a_{t+1}, a_t) \Big) \quad \rightarrow (3)$$

Now, $D_F(a, a_{n+1}) \geq 0$ and $D_F(a, a_1) = F(a) - F(a_1) - \underbrace{\langle \nabla F(a_1), a - a_1 \rangle}_{\geq 0}$

However, $a_1 = \operatorname{argmin}_a F(a) \Rightarrow \langle \nabla F(a_1), a - a_1 \rangle \geq 0$

(2 first order optimality.)

$\therefore$ From (3) + above, $\rightarrow (4)$

$$R_n \leq \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{F(a) - F(a_1)}{\eta} - \frac{1}{\eta} \sum_{t=1}^{n} D_F(a_{t+1}, a_t)$$

Next, since $\nabla F(a) - \eta y \in \text{interior}\big(\text{Dom}(F^*)\big)$, and

$$\tilde{a}_{t+1} = \operatorname*{argmin}_{a \in D} \eta \langle a, y_t \rangle + D_F(a, a_t) \quad \rightarrow (1)$$

gradient at $\tilde{a}_{t+1}$ is zero

$$\Rightarrow \quad \eta y_t + \nabla F(\tilde{a}_{t+1}) - \nabla F(a_t) = 0$$

$$\Rightarrow \quad \langle a_t - a_{t+1}, y_t \rangle = \frac{1}{\eta} \langle a_t - a_{t+1}, \nabla F(a_t) - \nabla F(\tilde{a}_{t+1}) \rangle$$

$$= \frac{1}{\eta} \Big(D_F(a_{t+1}, a_t) + D_F(a_t, \tilde{a}_{t+1}) - D_F(a_{t+1}, \tilde{a}_{t+1}) \Big).$$

$$\leq \frac{1}{\eta} \Big(D_F(a_{t+1}, a_t) + D_F(a_t, \tilde{a}_{t+1}) \Big).$$

$\longrightarrow$ (5).

Substituting (5) in (4),

$$R_n(a) \leq \frac{1}{\eta} \Big(F(a) - F(a_1) + \sum_{t=1}^{n} D_F(a_t, \tilde{a}_{t+1}) \Big).$$

PG

FTRL Regret Bound:

Thm (28.5 in textbook): As before: $\eta > 0$, F Legendre, over D. A convex, $A \cap D \neq \emptyset$. $\{a_1, a_2, \cdots, a_{n+1}\}$ actions from FTRL. Then, for any $a \in A$, we have:

$$R_n(a) \le \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D_F(a_{t+1}, a_t)$$

Pf: Similar to above. ∎

Applications

① A = Unit ball in $\mathbb{R}^d$ $= \{a \in \mathbb{R}^d : \|a\|_2 \le 1\}$

$\alpha = \{y_t \in \mathbb{R}^d : \|y_t\|_2 \le 1\}$

(i.e., both player and adversary choose vectors from the unit ℓ_2-ball $\triangleq B_2^d$).

$$F(a) = \frac{1}{2}\|a\|_2^2 \qquad \eta = \frac{1}{\sqrt{n}}.$$

Prop (28.6 in book): Setting as above. Then,

$$R_n(a) \le \sqrt{n}$$

Proof: $D_F(x,y) = \frac{1}{2}\|x-y\|^2$

Further, $\tilde{a}_{t+1}$ s.t. $\eta y_t + \nabla F(\tilde{a}_{t+1}) - \nabla F(a_t) = 0$

$$\Rightarrow \quad \eta y_t + \tilde{a}_{t+1} - a_t = 0$$

$$\therefore \quad D_F(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|a_t - \tilde{a}_{t+1}\|^2 = \frac{\eta^2}{2}\|y_t\|^2$$

Also, $|F(a) - F(b)| \le \frac{1}{2} \quad \forall \quad a, b \in B_2^d$

$$\therefore \quad R_n(a) \le \frac{1}{2\eta} + \frac{1}{\eta} \cdot \sum_{t=1}^{n} \frac{\eta^2}{2} \underbrace{\|y_t\|^2}_{\le 1}$$

$$\le \frac{1}{2\eta} + \frac{\eta}{2} \cdot n$$

$$\eta = \frac{1}{\sqrt{n}} \Rightarrow \quad R_n(a) \le \sqrt{n}$$ ☒

② $A = P_{d-1}$ = d-dimensional simplex of prob. dist.

$$y_t \in \alpha = [0,1]^d \qquad F(x) = \sum_{i=1}^{d} (x_i \ln x_i - x_i),$$

$$D_f(x,y) = \sum_{i=1}^{d} x_i \ln\left(\frac{x_i}{y_i}\right) + \underbrace{\sum_{i=1}^{d} (y_i - x_i)}$$

$x, y \in P_{d-1} \Rightarrow \sum x_i = \sum y_i = 1.$

$$\eta = \sqrt{2\ln(d)/n}.$$

Prop. (Prop. 28.7 in textbook): $R_n \le \sqrt{2n \ln(d)}$ ✓ with mirror descent.

Proof:

$$R_n(a) \le \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D_R(a_{t+1}, a_t)$$

↓ Pinsker's inequality.

$$\le \frac{\ln(d)}{\eta} + \sum_{t=1}^{n} \underbrace{\|a_t - a_{t+1}\|_1}_{z} \underbrace{\|y_t\|_\infty}_{\le 1} - \frac{1}{2\eta} \sum_{t=1}^{n} \underbrace{\|a_{t+1} - a_t\|_1^2}_{z^2}$$

note ℓ_1 norm.

Note: $z - \frac{z^2}{2\eta} \le \frac{\eta}{2}$. $\forall z \in \mathbb{R}$.

$$\Rightarrow R_n(a) \le \frac{\ln(d)}{\eta} + \frac{n\eta}{2}$$

$\sqrt{\frac{2\ln(d)}{n}}$ ∎

A general bound (Corollary 28.8): F Legendre, $\|\cdot\|_\gamma$ a norm and $\|\cdot\|_{\gamma^*}$ the dual-norm, s.t.

$$D_F(a_{t+1}, a_t) \ge \frac{1}{2} \|a_{t+1} - a_t\|_\gamma^2$$

(i.e., F is 1-strongly convex with the $\|\cdot\|_\gamma$ norm)

Then, with either mirror descent or FTRL, we have

$$R_n \leq \frac{\text{diameter}_F(A)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_{r^*}^2$$

$$\|y\|_{r^*} = \max_{x : \|x\|_r \leq 1} \langle x, y \rangle$$

Bandit Feedback

Moving from full feedback (i.e. y_t is observed) to the bandit feedback setting (i.e., $\langle A_t, y_t \rangle$ is observed).

Setting: As in the usual adversarial setting, adversary aware of policy of player. Chooses loss sequence $\{y_1, y_2, \dots, y_n\}$, $y_t \in \mathcal{A} \subseteq \mathbb{R}^d$.

Player chooses actions: $\{A_1, A_2, \dots, A_n\}$, $A_t \in A$.

The actions now are randomized, hence notation of A_t instead of a_t.

Regret: $R_n(a) = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a, y_t \rangle\right]$.

$$R_n = \max_{a \in A} R_n(a).$$

Algorithm Ideas:

① Use importance sampled estimator to determine unbiased estimate of y_t

② Use mirror descent / FTRL to evolve the expected action to take at the next step. Actual action is sampled from a suitably chosen dist. whose mean is determined above by Mirror descent / FTRL.

Bandit Mirror Descent / FTRL Algorithm:

Stochastic M.D. / FTRL

Setting: F is Legendre, Learning rate $\eta > 0$, action set A, domain$(F) = D$.

"mean" action

Initial: $\overline{A}_1 = \operatorname{argmin}_{a \in A \cap D} F(a)$

P_t : Distribution over actions at time t, chosen such that

the variance of the IS estimator for $\hat{Y}_t$ does not blow up. (Note that if P_t were deterministic / delta function, the estimator variance will blow up.) P_t a function of past actions / observations.

For each $t = 1, 2, \ldots, n$:

(a) Sample action $A_t \sim P_t$, with $\mathbb{E}_{P_t}[A_t] = \bar{A}_t$, and observe $\langle A_t, y_t \rangle$.

(b) Construct $\hat{Y}_t$, an unbiased, finite variance estimate for y_t.

(c) Mirror Descent / FTRL:

(Mirror Descent) $\bar{A}_{t+1} = \underset{a \in \mathcal{A} \cap \mathcal{D}}{\operatorname{argmin}} \; \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t)$.

(FTRL) $\bar{A}_{t+1} = \underset{a \in \mathcal{A} \cap \mathcal{D}}{\operatorname{argmin}} \; \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a)$.

Thm (28.10 in textbook): F Legendre, $\mathcal{A}$ convex, $\eta > 0$, and s.t. $\mathbb{E}[\hat{Y}_t \mid \bar{A}_t] = y_t \;\; \forall t = 1, \ldots, n$, and $(\hat{Y}_t, \bar{A}_t)$ well defined.

Then, $$R_n(a) \le \mathbb{E}\left[\frac{F(a)-F(\bar{A}_1)}{\eta} + \sum_{t=1}^{n}\langle \bar{A}_t - \bar{A}_{t+1}, \hat{Y}_t\rangle - \frac{1}{\eta}\sum_{t=1}^{n} D_F(\bar{A}_{t+1}, \bar{A}_t)\right]$$

Further, for stochastic Mirror Descent where

$$\bar{A}_{t+1} = \underset{a \in A \cap D}{\operatorname{argmin}}\ \eta\langle a, \hat{Y}_t\rangle + D_F(a, \bar{A}_t),$$

and assuming $\bar{A}_{t+1}$ lies in the interior of $A \cap D$, for all $a \in A$ (a.s.), then

$$R_n \le \frac{\text{diameter}_F(A)}{\eta} + \frac{1}{\eta}\sum_{t=1}^{n}\mathbb{E}\left[D(\bar{A}_t, \bar{A}_{t+1})\right]$$

Proof:

$$\mathbb{E}\left[\langle A_t, y_t\rangle\right] = \mathbb{E}\left[\langle \bar{A}_t, y_t\rangle\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, y_t\rangle \,|\, \bar{A}_t\right]\right]$$

unbiased estimator

$$= \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, \hat{Y}_t\rangle \,|\, \bar{A}_t\right]\right]$$

$$\Rightarrow R_n(a) = \mathbb{E}\left[\sum_{t=1}^{n}\langle A_t, y_t\rangle - \langle a, y_t\rangle\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n} \langle \bar{A}_t - a, \hat{Y}_t \rangle \right].$$

Tracked by bandit M.D/FTRL

Rest of the proof is immediate from mirror descent / FTRL proof. ☐

Explicit Construction of $\hat{Y}_t, P_t$ for $A = B_2^d$:

Linear Bandits on a unit Ball.

Setting: $A = B_2^d$ = unit ball in $\mathbb{R}^d$. $= \mathcal{X}$.

$S_t \in \{0,1\}$, a selector r.v.

$U_t \in \{\pm e_1, \pm e_2, \cdots, \pm e_d\}$, over $2d$ unit vectors.

s.t.

Note: $\|\bar{A}_t\|_2 < 1$, because we will operate FTRL over the action $\tilde{A} = \{x : \|x\|_2 \le r\}$, $r = 1 - 2\eta d < 1$.

$$\mathbb{E}\left[S_t \mid A_1, A_2, \cdots, A_{t-1}\right] = 1 - \|\bar{A}_t\|_2$$

(≈ 0; $\approx r$; ≈ 1)

i.e., $P\left(S_t = 1 \mid A_1, \cdots, A_{t-1}\right) = 1 - \|\bar{A}_t\|_2$

(S_t, U_t cond. indep. of other r.vs. given $A_1, \cdots, A_{t-1}$)

$U_t \mid (A_1, \cdots, A_{t-1}) \sim \text{Uniform}\{\pm e_1, \cdots, \pm e_d\}$.

Choose $$A_t = S_t V_t + (1-S_t)\frac{\bar{A}_t}{\|\bar{A}_t\|_2}$$

i.e., w.p. $1-\|\bar{A}_t\|_2$, choose a random unit direction, and otherwise choose (normalized) action suggested by M.D./FTRL.

Observe $$E[A_t | A_1, \dots, A_{t-1}] = 0 + \bar{A}_t .$$

Finally, $$\hat{Y}_t = \left(\frac{d\, S_t \langle A_t, y_t\rangle}{1-\|\bar{A}_t\|_2}\right) A_t$$ ← unbiased estimator.

and $$F(a) = -\ln\left(1-\|a\|_2\right) - \|a\|$$

Algo: FTRL over $\tilde{A} = \{a \in \mathbb{R}^d : \|a\|_2 \le r\}$,

for $r = (1-2\eta d) < 1$. $\left(\eta = \sqrt{\frac{\ln(n)}{3dn}}\right)$

controls variance of estimator

i.e.,

$$\bar{A}_{t+1} = \underset{a \in \tilde{A} \cap D}{\operatorname{argmin}} \; \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a).$$

Facts: With above iteration and regularizer, and with

$\hat{L}_{t-1} \triangleq \sum_{s=1}^{t-1} \hat{Y}_s$, we have: $\bar{A}_t = \widetilde{\Pi}_2\left(\frac{\eta \hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}\right)$

where $\widetilde{\Pi}_2(x)$ is the ℓ_2 projection of x onto $\tilde{A}$.

Summary: At each time $t=1,2,\dots,n$:

① Compute mean: $\bar{A}_t = \widetilde{\Pi}_2\left(\frac{\eta \hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}\right)$, where

$\hat{L}_{t-1} = \sum_{s=1}^{t-1} \hat{Y}_s$

② $S_t \in \{0,1\}$, with $S_t = \begin{cases} 1 & \text{w.p. } 1-\|\bar{A}_t\|_2 \\ 0 & \text{w.p. } \|\bar{A}_t\|_2 \end{cases}$ ($=r<1$)

Sample: $U_t \in \{\pm e_i, i=1,2,\dots,d\}$ uniformly.

③ Play: $A_t = S_t U_t + \frac{(1-S_t)\bar{A}_t}{\|\bar{A}_t\|_2}$

④ Observe | Estimate: Observe $\langle A_t, y_t \rangle$, and update

$\hat{Y}_t = \left(\frac{d\, S_t \langle A_t, y_t \rangle}{1-\|\bar{A}_t\|_2}\right) A_t$

Theorem (28.11 in textbook): $\{y_t, t=1,\cdots,n\} \subseteq \mathcal{A} = B_2^d$.

Then, with $r = 1 - 2\eta d$, $\eta = \sqrt{\frac{\ln(n)}{3dn}}$, we have

$$R_n \leq 2\sqrt{3nd\ln(n)}$$

Note: This gives $\sqrt{d}$ scaling, whereas previous notes based on Kiefer-Wolfowitz had scaling linearly in d.

Proof: $a^* = \operatorname{argmin}_{a} \sum_{t=1}^{n} \langle a, y_t \rangle$, the optimal action.

$$R_n = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a^*, y_t \rangle\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - r a^*, y_t \rangle\right] + (1-r)\sum_{t=1}^{n} \langle a^*, y_t \rangle$$

($\leq \|a^*\| \|y_t\| \leq 1$)

$$\leq \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - r a^*, y_t \rangle\right] + (1-r)n.$$

Now, recall: ψ Legendre, $\eta > 0$, $x, y \in$ interior $(\mathcal{D})$,

(1) ψ twice differentiable. Let $z \in [x,y]$ s.t.

$D_\psi(x,y) = \frac{1}{2}\|x-y\|^2_{\nabla^2\psi(z)}$. Then, for any $u \in \mathbb{R}^d$,

$$\langle x-y, u\rangle - \frac{D_\psi(x,y)}{\eta} \le \frac{\eta}{2}\|u\|^2_{\nabla^2\psi(z)^{-1}}$$

(Thm 26.13 in textbook) : Uses Cauchy-Schwartz + Calculus.

---

② For FTRL, we have from Thm 28.5,

$$R_n(a) \le \frac{F(a)-F(a_1)}{\eta} + \sum_{t=1}^{n}\langle a_t - a_{t+1}, y_t\rangle - \frac{1}{\eta}\sum_{t=1}^{n} D_F(a_{t+1}, a_t)$$

$$\therefore R_n(ra^*) \triangleq E\left[\sum_{t=1}^{n}\langle A_t - ra^*, y_t\rangle\right]$$

look at page 22/23 in these notes.

$$= E\left[\sum_{t=1}^{n}\langle \bar{A}_t - ra^*, \hat{y}_t\rangle\right]$$

$\le \ln\left(\frac{1}{1-r}\right)$ $\qquad$ $E[\cdot] \le 2d$

$$\le \frac{\text{Diameter}_F(\tilde{A})}{\eta} + \frac{\eta}{2} E\left[\sum_{t=1}^{n}\|\hat{y}_t\|^2_{\nabla^2 F(z_t)^{-1}}\right],$$

for $z_t \in [\bar{A}_t, \bar{A}_{t+1}]$.

Further, $\eta \|\hat{\gamma}_t\|_2 \le \left(\frac{\eta d}{1-r}\right) = \frac{1}{2}$ $\left(\text{defn. of } r = 1-2\eta d\right)$

$\rightarrow = \alpha \bar{A}_t + (1-\alpha)\bar{A}_{t+1}$, for some $\alpha \in [0,1]$

$$\therefore \quad \frac{1-\|z_t\|_2}{1-\|\bar{A}_t\|_2} \le \sup_{\alpha\in[0,1]} \frac{1-\|\alpha\bar{A}_t + (1-\alpha)\bar{A}_{t+1}\|_2}{1-\|\bar{A}_t\|_2}$$

(by default, $\|\cdot\| = \|\cdot\|_2$, unless otherwise stated)

$$= \max\left\{1, \frac{1-\|\bar{A}_{t+1}\|_2}{1-\|\bar{A}_t\|_2} \chi_{\{\|\bar{A}_{t+1}\|_2 < \|\bar{A}_t\|_2\}}\right\}$$

?

$$\le \max\left\{1, \frac{1+\eta\|\hat{L}_{t-1}\|_2}{1+\eta\|\hat{L}_t\|_2}\right\}$$

! Recall:

$\bar{A}_t = \widetilde{\Pi}_2\left(\frac{\eta \hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}\right)$

$\hat{L}_t = \hat{L}_{t-1} + \hat{\gamma}_t$ - - - +

$\eta\|\hat{L}_{t+1}\|_2 \le \eta\|\hat{L}_{t-1}\| + \frac{1}{2}$

$$\le \max\left\{1, \frac{1+\eta\|\hat{L}_{t-1}\|_2}{\frac{1}{2}+\eta\|\hat{L}_{t-1}\|_2}\right\} \le 2.$$

→ (✗)

Next, $\nabla^2 F(a) = \frac{I}{1-\|a\|} + \frac{aa^\top}{\|a\|(1-\|a\|)^2} \ge \frac{I}{1-\|a\|}$

$$\Rightarrow \nabla^2 F(a)^{-1} \le (1-\|a\|) I$$

$$\Rightarrow \|\hat{\gamma}_t\|^2_{\nabla^2 F(a)^{-1}} = \hat{\gamma}_t^\top \nabla F(a)^{-1} \hat{\gamma}_t \le (1-\|a\|)\|\hat{\gamma}_t\|_2^2$$

$$\therefore \quad E\left[\|\hat{y}_t\|^2_{\nabla^2 F(z_t)^{-1}}\right] \leq E\left[(1-\|z_t\|)\|\hat{y}_t\|_2^2\right]$$

$$= d^2 E\left[\frac{(1-\|z_t\|)\, S_t \langle v_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right]$$

From ⊛:

$$\left(\frac{1-\|z_t\|}{1-\|\bar{A}_t\|}\right) \leq 2$$

$$\Rightarrow E\left[\|\hat{y}_t\|^2_{\nabla^2 F(z_t)^{-1}}\right] \leq 2d^2 \underbrace{E\left[\frac{S_t}{1-\|\bar{A}_t\|}\right]}_{=1} E\left[\langle v_t, y_t\rangle^2\right]$$

$$\leq 2d. \qquad = \sum_{i=1}^{d} \frac{y_i^2}{2d} + \sum_{i=1}^{d} \frac{(-y_i)^2}{2d}$$

$$= \frac{1}{d} \underbrace{\sum y_i^2}_{\leq 1} \leq \frac{1}{d}.$$

Finally, $\text{diameter}(\tilde{A}) \leq \ln\left(\frac{1}{1-\gamma}\right)$.

$$\therefore \quad R_n \leq (1-\gamma)n + \frac{1}{\eta}\ln\left(\frac{1}{1-\gamma}\right) + \eta n d$$

$$= \frac{1}{\eta}\ln\left(\frac{1}{2\eta d}\right) + 3\eta n d$$

$$\leq 2\sqrt{3nd\ln(n)}$$ ∎