# Lecture 14 — September 22

*Lecturer: Aditya Gopalan* *Scribe: Arvind Kumar*

## 14.1 RECAP

### 14.1.1 THE ONLINE CONVEX OPTIMIZATION MODEL

$\mathbb{K}$ Convex Set $\subseteq \mathbb{R}^d$
$\forall t \geq 1$ choose $w_t \in \mathbb{K}$
See $f_t : \mathbb{K} \to \mathbb{R}$
Suffer loss : $f_t(w_t)$

### 14.1.2 FOLLOW THE LEADER

Choose $W_1 \in \mathbb{K}$ arbitrarily,$\forall t \geq 1$

$$W_t = \min_{w \in \mathbb{K}} \sum_{s=1}^{t-1} fs(w)$$

FTL is very bad (linear regret ) in some cases.

**Example**- Linear loss function $\{f_t\}$
$\mathbb{K} = [-1,1]$ ,$f_t(x) = F_t x$, $\quad \{F_t\} = \{0.5, -1, 1, -1, ......\}$
"DARTBOARD GAME" : $K \in R^d$

$$f_t(x) = \|z_t - x\|_2^2$$

FTL gives $\mathcal{O}(\log T)$ regret

**Lemma 1 - [FTL Regret]**

$$\forall u \in \mathbb{K}: \quad \sum_{t=1}^{T} (f_t(w_t) - f_t(u)) \leq \sum_{t=1}^{T} [f_t(w_t) - f_t(w_{t+1})]$$

**Question** : How can we improve FTL or prevent "Oscillations" or "instability in its behavior"?

**Ans** Three broad streams of algorithms:

1. REGULARIZATION BASED ALGORITHM (Follow The Regularised Leader)

2. PROXIMAL POINT ALGORITHMS(e.g Projected Gradient Descent).

3. PERTURBATION BASED ALGORITHM (e.g Follow the Perturbed leader)

## 14.2 Follow The Regularised Leader (FTRL)

$R\colon \mathbb{K} \to \mathbb{R}$ is a strictly convex function.

$$w_t = \underset{w\in\mathbb{K}}{\operatorname{argmin}}\{\sum_{s=1}^{t-1}(f_s(w)+R(w))\}$$

### 14.2.1 Classic Examples

1. $\mathbb{K} = \mathbb{R}^d$ and linear loss function $f_t x = \langle z_t, x\rangle$
$R(x) = \frac{1}{2\eta}\|x\|_2^2$

$$w_t = \underset{w\in\mathbb{R}^d}{\operatorname{argmin}}\sum_{s=1}^{t-1}\left(\langle z_s, w\rangle + \frac{1}{2\eta}\|w\|_2^2\right)$$

Let

$$G(w) = \sum_{s=1}^{t-1}\left(\langle z_s, w\rangle + \frac{1}{2\eta}\|w\|_2^2\right)$$

$\bigtriangledown G(w)|_{w=w_t} = 0$
So

$$\iff \sum_{s=1}^{t-1} Z_s + \frac{1}{\eta} w_t = 0$$

$$w_t = -\eta \sum_{s=1}^{t-1} Z_s$$
$$= w_{t-1} - \eta Z_{t-1}$$

$$= w_{t-1} - \eta \bigtriangledown f_{t-1}(w_{t-1})$$

This is (Online and Unconstrained) Gradient Descent.

2. $\mathbb{K} = \triangle_d \quad \{(w_1, w_2, ...., w_d) : w_i \geq 0 \forall i \quad \sum_{i=1}^{d} w_i = 1\}$
Linear losses : $f_t(x) = \langle z_t, x \rangle$
Entropic Regularizer : $\forall x \in \mathbb{K}$

$$R(x) = \frac{\langle w, \log w \rangle}{\eta}$$
$$= \frac{\sum_{i=1}^{d} w_i \log w_i}{\eta}$$

$$w_t = \underset{w \in \mathbb{K}}{\operatorname{argmin}} \{ \sum_{s=1}^{t-1} \langle z_s, w \rangle + \frac{\sum_{i=1}^{d} w_i \log w_i}{\eta} \}$$

Its Solution turns out to be

$$w_t(i) = \frac{w_{t-1}(i) \exp(-\eta Z_{t-1}(i))}{\sum_{j=1}^{d} w_{t-1}(j) \exp(-\eta Z_{t-1}(j))}$$

### 14.2.2 Lemma: [FTRL Regret Bound]

If FTRL produces $w_1, w_2, ....., w_T$ then $\quad \forall u \in \mathbb{K}$

$$\sum_{t=1}^{T} (f_t(w_t) - f_t(u)) \leq R(u) - R(w_1) + \sum_{t=1}^{T} [f_t(w_t) - f_t(w_{t+1})]$$

**Proof:** Key Observations : Running FTRL on the sequence of loss functions $f_1, f_2, f_3, ......, f_T$ is equivalent to running FTL with loss function $f_0 \equiv R, f_1, f_2, ....., f_T$.

$$\sum_{t=0}^{T} (f_t(\,w_t) - f_t(u)) \leq \sum_{t=0}^{T} [f_t(w_t) - f_t(w_{t+1})]$$
$$= f_0(w_0) - f_0(u) + \sum_{t=1}^{T} f_t(w_t) - f_t(u)$$
$$= f_0(w_i) - f_0(w_1) + \sum_{t=0}^{T} [f_t(w_t) - f_t(w_{t+1})]$$

□

**Theorem 14.1.** *Let $f_t(x) = \langle z_t, x \rangle, \mathbb{K} = \mathbb{R}^d$ and $R(x) = \frac{1}{2\eta} \|x\|_2^2$*
*$W_t$ is computed by FTRL(O.G.D).*
*Then $\forall u \in \mathbb{R}^d$*

$$Regret_T{}^{FTRL}(u) \leq \frac{1}{2\eta} \|u\|_2^2 + \eta \times \sum_{t=1}^{T} \|z_t\|_2^2$$

**Proof:** By FTRL regret lemma ,

$$Regret_T{}^{FTRL}(u) \leq R(u) - R(w_1) + \sum_{t=1}^{T} [f_t(w_t) - f_t(w_{t+1})]$$

$$\leq \frac{1}{2\eta}\|u\|_2^2 + \sum_{t=1}^{T} \langle z_t, w_t - w_{t+1} \rangle$$

By OGD,

$$w_{t+1} = w_t - \eta \times z_t$$

$$= \frac{1}{2\eta}\|u\|_2^2 + \sum_{t=1}^{T} \langle z_t, \eta z_t \rangle$$

□

## 14.3 DEFINITION

LIPSCHITZ CONTINUITY :
$f : \mathbb{K} \to \mathbb{R}$ is L-Lipschitz continuous with respect to a norm $||\cdot||$, if $\forall x, y \in \mathbb{K}$

$$|f(x) - f(y) \leq L\|x - y\|$$

**Theorem 14.2.** *FTRL Regret with strongly convex Regularizer+ Lipschitz-continuous losses Let $f_1, f_2, ....$be such that $f_t : \mathbb{K} \to \mathbb{R}$ is $L_T$ - Lipsschitz continuous with respect to $\| \quad \|_\Box$ . Let $\mathbb{R}$ be $\sigma-$strongly convex w.r.t the same norm $\| \quad \|_\Box$ .Then $\forall u \in \mathbb{K}$*

$$Regret_T^F TRLu \leq R(u) - \min_{v \in \mathbb{K}} R(v) + \frac{1}{\sigma} \times \sum_{t=1}^{T} L_t^2$$

**Proof:** $\forall t$ Let

$$\phi_t(w) = \sum_{s=1}^{t-1} f_s(w) + R(w)$$

FTRL picks

$$w_t = \underset{w \in \mathbb{K}}{\text{argmin}} \quad \phi_t(w)$$

$\phi_t(w)$ is $\sigma-$strongly convex.
**Lemma :** $\phi_t$ is $\sigma-$ strongly convex over $\mathbb{K}$,

$$w_t = \underset{w \in \mathbb{K}}{\text{argmin}} \quad \phi_t(w)$$

Then $\forall v \in \mathbb{K}$:

$$\phi_t(v) - \phi_t(w_t) \geq \frac{\sigma}{2}\|v - w_t\|_\Box^2$$

**Proof:** From strong convexity

$$\phi_t(v) - \phi_t(w_t) \geq \langle \triangledown \phi_t(w_t), v - w_t \rangle + \frac{\sigma}{2} \|v - w_t\|_\square^2$$

$$\geq \frac{\sigma}{2} \|v - w_t\|_\square^2$$

□

$$\phi_t(w_{t+1}) - \phi_t(w_t) \geq \frac{\sigma}{2} \|w_{t+1} - w_t\|_\square^2 \tag{14.1}$$

In Lemma switching
$\phi_t \to \phi_{t+1}$ and $v \to w_t$

$$\phi_{t+1}(w_t) - \phi_{t+1}(w_{t+1}) \geq \frac{\sigma}{2} \|w_{t+1} - w_t\|_\square^2 \tag{14.2}$$

Adding inequalities 14.1 and 14.2

$$f_t(w_t) - f_t(w_{t+1}) \geq \sigma \|w_{t+1} - w_t\|_\square^2$$

Also since $f_t$ is $L_T$ - Lipsschitz continuous

$$\sigma \|w_{t+1} - w_t\|_\square^2 \leq L_t \|w_{t+1} - w_t\|_\square$$

$$\|w_{t+1} - w_t\|_\square \leq \frac{L_t}{\sigma}$$

Since

$$\sum_{t=1}^{T} (f_t(w_t) - f_t(u)) \leq R(u) - R(w_1) + \sum_{t=1}^{T} [f_t(w_t) - f_t(w_{t+1})]$$

$$\leq R(u) - R(w_1) + \sum_{t=1}^{T} L_t \times \frac{L_t}{\sigma}$$

$$\leq R(u) - R(w_1) + T \times \frac{L^2}{\sigma}$$

□