RFTL

$w_1 := \min_{w \in S} R(w)$

$w_{t+1} = \operatorname{argmin}_{w \in S} \; \eta \sum \nabla \ell_t(w) + R(w)$

Thm:

$R(\cdot)$ is β s.c. regularizer wrt $\|\cdot\|$

$\forall t \quad \|\ell_t(w_t)\|_* \leq L$

$D = \sqrt{\sup_{x,y \in S} R(x) - R(y)}$

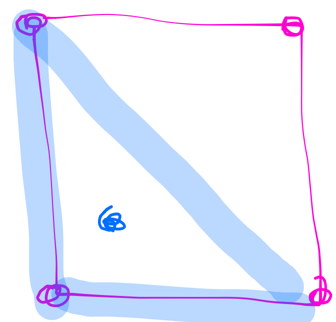
for an appropriate choice of η

$\forall u \in S \quad R_T(u) \leq \frac{2LD}{\sqrt{\beta}} \sqrt{T}$

Cool application: Approx Caratheodory Thm

C.T. $z \in \operatorname{conv}(V)$ where $V = \{v_1, \ldots, v_m\} \subseteq \mathbb{R}^n$

Then z can be written as convex combination of $n+1$ pts in V



best possible

Approximate C.T. Let $p \geq 2$

Let $V = \{v_1, \ldots, v_m\}$ $v_i \in \mathbb{R}^n$ with $\|v_i\|_p \leq 1$

Dimension free!

Then $\forall z \in \operatorname{conv}(V)$ $\exists$ set $V' \subseteq V$ s.t. $|V'| = O\left(\frac{p}{\varepsilon^2}\right)$

and $\exists z' \in \operatorname{conv}(V')$ s.t. $\|z - z'\|_p \leq \varepsilon$

Lots of cool apps

- computing NE
- algs for k-densest subgraph
- submodular minimization
- SVM training

Original proof: [Barman]

Use exact C.T. $z = \sum_{i=1}^{m} \lambda_i v_i$

Sample v_i's according to this distn

use concentration inequalities

[Mirrokni, Leme, Vladu, Wong]

- deterministic
- nearly linear time   using OMD = FTRL

Goal: find $x \in \mathbb{R}^m$ s.t. $z \approx \sum x_i v_i$ and x sparse

$\equiv \min_x \|Vx - z\|_p$ s.t. x sparse

$$\begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \approx \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

don't know how to deal w/ sparsity, so let's drop it for now

Recall $\|Vx - z\|_p = \max\{ (y, Vx-z) \mid \|y\|_q \le 1 \}$

so we want $\min_{x \in \Delta_m} \max_{y \in B_q} (y, Vx - z)$

Think of this as 2-player game $\Rightarrow$

$$\min_{x \in \Delta_m} \max_{y \in B_q} (y, Vx-z) = \max_{y \in B_q} \min_{x \in \Delta_m} (y, Vx-z)$$

apply Sion's minimax Thm

$X \subseteq \mathbb{R}^n$ convex & compact

$Y \subseteq \mathbb{R}^n$ convex & compact

$g : X \times Y \to \mathbb{R}$ fn s.t.

$\forall y \in Y$ $g(\cdot, y)$ convex & cont over X

$\forall x \in X \quad g(x,\cdot)$ concave & cont over Y

Then

$$\max_{y \in Y} \min_{x \in X} g(x,y) = \min_{x \in X} \max_{y \in Y} g(x,y)$$

$= $ saddle point $=$ Nash Eq (x^*, y^*)

$$g(x, y^*) \geq g(x^*, y^*) \geq g(x^*, y) \qquad \forall x \in X, y \in Y$$

$$\max_{y \in B_q} \min_{x \in \Delta_m} \langle y, Vx - z \rangle$$

$$= - \min_{y \in B_q} \boxed{\max_{x \in \Delta_m} \langle y, z - Vx \rangle}$$

$f(y)$ convex

(sup g affine)

Solve using linearized FTRL

need $\nabla f(y)$

$$\ell_t(w) = \langle \nabla f(w^t), w \rangle \quad \overset{w^t}{\rightleftarrows}$$

$$f(y) = \max_{x \in \Delta_m} \langle y, z - Vx \rangle$$

$$= \langle y, z - V x^*(y) \rangle$$

Envelope Thm $\quad \nabla f(y) = z - V x^*(y)$

$$\nabla f(y)_i = (z - Vx^*(y))_i - (y^T V)_i \frac{\partial x^*(y)}{\partial y_i} \rightarrow = 0$$

Crucial observation:

$$\ell_t(w) = \langle z - Vx^*(w^t), w \rangle$$

x^* has only one nonzero!

Aside $\qquad g(t) = \sup_{x \in X} f(x,t) \qquad$ say $t \in [0,1]$

$$X^*(t) = \{ x \in X \mid f(x,t) = g(t) \}$$

Envelope thms describe sufficient conditions for

$$g'(t) = \frac{\partial f(x,t)}{\partial t} \qquad \forall x \in X^*(t)$$

i.e. derivative is what you get holding maximizer fixed at optimal level

Player	"Adversary"
runs FTRL

$w^1 \in B_q \longrightarrow$

$\longleftarrow x^1$

$x^1 = \underset{x \in \Delta_m}{\operatorname{argmax}} \langle w^1, z - Vx \rangle$

$\ell_1(w) = \langle z - Vx^1, w \rangle$

$w^2 \in B_q \longrightarrow$

$\longleftarrow x^2$

$x^2 = \underset{x \in \Delta_m}{\operatorname{argmax}} \langle w^2, z - Vx \rangle$

$\ell_2(w) = \langle z - Vx^2, w \rangle$

$\vdots$

$w^t \in B_q \longrightarrow$

$\longleftarrow x^t$

$x^t = \underset{x \in \Delta_m}{\operatorname{argmax}} \langle w^t, z - Vx \rangle$

$\ell_t(w) = \langle z - Vx^t, w \rangle$

Observations:

1. x^i puts all mass on one column

$$\langle w, z - Vx \rangle = w^T z - \underline{w^T V} x$$
$$= \sum_{j=1}^{m} (w^T V)_j \, x_j$$

2. $\langle w^t, z - Vx^t \rangle \geq 0$ since $z \in \operatorname{conv}(V) \Rightarrow$ can make this 0 and we are maximizing

So if we don't need too many rounds, happy!

Guarantee of FTRL:

$$\frac{1}{T}\sum_{t=1}^{T}\left[\ell_t(w^t)-\ell_t(u)\right] \le \frac{2LD}{\sqrt{\beta T}} \longrightarrow \varepsilon \qquad \forall u\in B_q$$

$$LHS = \frac{1}{T}\sum_{t=1}^{T}(z-Vx^t, w^t-u) \ge \frac{1}{T}\sum_{t=1}^{T}(Vx^t-z, u) \qquad \text{by } ②$$

$$= \left(V\sum\frac{x^t}{T}-z, u\right)$$

Conclusion: $\left\|V\sum_{t=1}^{T}\frac{x^t}{T}-z\right\|_p \le \varepsilon$

convex comb of at most T elts of V!

Analysis

- Take $R(w)=\frac{1}{2}\|w\|_q^2$

  $R(w)$ is $\frac{1}{2(p-1)}$ s.c. wrt $\|\cdot\|_q$ for $1<q\le 2$

- What is L?

  $\ell_t(w) = (z-Vx^t, w)$

  $\nabla \ell_t(w) = z - Vx^t = z - v^t$ ← some col of V

  $\Rightarrow \|\nabla\ell_t\|_* = \|\nabla\ell_t(w_t)\|_p \le \|z\|_p + \|v^t\|_p \le 2 \qquad \Rightarrow L\le 2$

- What is D?

  $B_q=\{w \mid \|w\|_q\le 1\}$ $\qquad D=\sqrt{\max_{u,v\in B_q} \frac{1}{2}\|u\|^2-\frac{1}{2}\|v\|^2} = \frac{1}{\sqrt{2}}$

- Regret bound

$$\frac{2LD}{\sqrt{\beta T}} = \frac{4\sqrt{2(p-1)}}{\sqrt{2}\,\sqrt{T}} \leq \varepsilon$$

$\Rightarrow T = \Omega\left(\frac{p}{\varepsilon^2}\right)$ iterations suffice