Hannan–Kalai–Vempala algorithm.

of experts is exponentially large in the dimension of V

if:

- Strategies can be represented as vectors in a low-dimensional vector space, V

&

- costs can be represented by linear functions on that vector space

&

- ∃ an efficient algorithm for minimizing such a cost function, then

<u>Notation:</u> $\mathcal{Y}$, set of experts $\subseteq \mathbb{R}^n$ is bounded.

let D be the ℓ_1 diameter of $\mathcal{Y}$,

$$R \stackrel{\text{def}}{=} \sup_{\substack{1 \le t \le T \\ x,y \in \mathcal{Y}}} |c_t \cdot (x-y)| \qquad D = \sup_{x,y \in \mathcal{Y}} \|x-y\|,$$

an adversary specifies cost vectors: $c_1, c_2, \cdots, c_T$

let $A = \sup_{1 \le t \le T} \|c_t\|,$

strategies are $x_1, x_2, \cdots, x_T$

cost of strategy x @ time t is $c_t \cdot x$

$$C_{i..j} \stackrel{\text{def}}{=} \sum_{t=i}^{j} c_t$$

if ALG is an algorithm choosing strategies,

then $ALG(i..j) \stackrel{\text{def}}{=} \sum_{t=i}^{j} c_t \cdot x_t$

for a vector c,

$$M(c) \stackrel{\text{def}}{=} \arg\min_{x \in \mathcal{S}} c \cdot x$$

any $x \in \mathcal{S}$

s.t.

$x \in$

bands (#1)

$$E\left[\sum_{t=1}^{T} c_t \cdot x_t\right] \leq (1+O(\varepsilon A)) \cdot E\left[c_{1..T} \cdot M(c_{1..T})\right] + O\left(\frac{D}{\varepsilon}\log(n)\right)$$

(#2)

$$\leq E\left[c_{1..T} \cdot M(c_{1..T})\right] + O\left(\sqrt{DRAT}\right)$$

'follow the leader' strategy: FTL

$$x_t = M(c_{1..t-1})$$

can be defeated easily.

"follow the perturbed leader" FTL_p

let c_0 be sampled from p, some probability distribution on $\mathbb{R}^n$

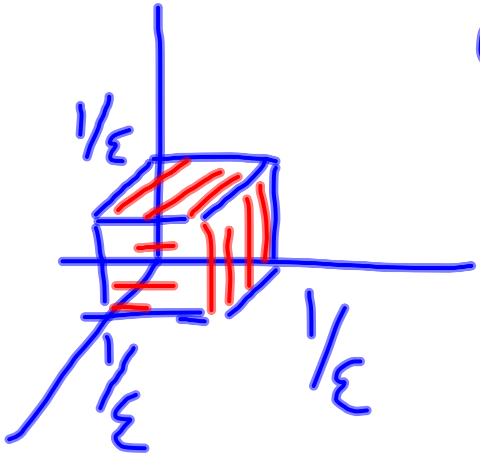
$$x_t = M(c_{0..t-1})$$

Let $\varepsilon > 0$ be a real.

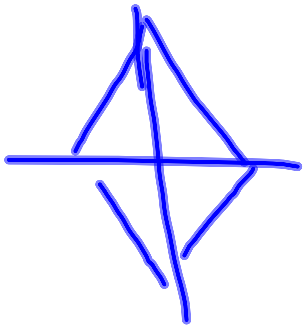
let:

$$d\mu(x) = \left(\frac{\varepsilon}{2}\right)^n e^{-\varepsilon \|x\|_1}$$

$$d\alpha(x) = \begin{cases} \left(\frac{\varepsilon}{2}\right)^n & \text{if } \|x\|_\infty \leq \frac{1}{\varepsilon} \\ 0 & \text{otherwise} \end{cases}$$



$d\alpha$



$d\mu$

how to sample from μ & α:

μ: draw y from the exponential distribution: $\Pr(y > r) = e^{-\varepsilon r}$ and change its sign from + to − w.p. $1/2$

α: sample each coordinate from uniform dist on $[-1/\varepsilon, 1/\varepsilon]$

can't do:

"be the perturbed leader:" BPL

$$x_t = M(c_{0..t})$$

let $OPT(1..T) = c_{1..T} \cdot M(c_{1..T})$

facts:

1. $BPL(1..T) \leq OPT(1..T) + E[c_0 \cdot (M(c_{1..T}) - M(c_0))]$

(2) FPL(1..T) is not much worse than BPL(1..T)

(3) $E[c_0 \cdot (x-y)]$ is small $\forall x, y \in \mathcal{Y}$

def: for two distributions p, q on $\mathbb{R}^n$

the multiplicative distance

$$d_x(p,q) \stackrel{\text{def}}{=} \text{the minimum } \delta \text{ s.t.}$$

$$dp(x) \leq (1+\delta)\, dq(x)$$

$$dq(x) \leq (1+\delta)\, dp(x)$$

$$\forall x$$

additive distance:

$d_+(p,q) \stackrel{\text{def}}{=}$ minimum ε s.t.

$\exists$ a prob. dist. μ on pairs $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$

s.t. $\mu(x \neq y) \leq \delta$

and for all measurable subsets $S \subseteq \mathbb{R}^n$,

$$\mu(x \in S) = p(S)$$

$$\mu(y \in S) = q(S)$$

$$BPL(1..T) \leq OPT(1..T) + E\left(c_0 \cdot [M(s_{1..T}) - M(c_0)]\right).$$

Bounding the expectation of $c_0 \cdot (x-y)$.

Lemma: If $\|x-y\|_1 \leq D$ and $E[\|c_0\|_\infty] \leq M$

then $E[c_0 \cdot (x-y)] \leq DM$.

$$\|x-y\|_1 = \sum_{i=1}^{n} |x_i - y_i| \leq D$$

$$E[\|c_0\|_\infty] = E\left[\max_i |(c_0)_i|\right] \leq M$$

Lemma: If c_0 is sampled from α

$$d\alpha(x) = \begin{cases} (\varepsilon/2)^n & \text{if } \|x\|_\infty \le \frac{1}{\varepsilon} \\ 0 & \text{o.w} \end{cases}$$

$E[\|c_0\|_\infty] \le \frac{1}{\varepsilon}$

c_0 " " " μ

$$d\mu(x) = \left(\frac{\varepsilon}{2}\right)^n e^{-\varepsilon\|x\|_1}$$

$E[\|c_0\|_\infty] \le O\left(\frac{\log n}{\varepsilon}\right)$

$n \ge 3.$

Proof (2): Abs values of coordinates of c_0 are independent exponentially distributed random vars with mean $1/\varepsilon$.

$$E\left(y - r \mid y > r\right) = \frac{1}{\varepsilon} \quad \forall r > 0.$$

$$y_i = |(c_0)_i|, \quad r = \frac{\ln(n)}{\varepsilon}$$

$$E\left(y_i - \frac{\ln(n)}{\varepsilon} \,\middle|\, y_i > \frac{\ln(n)}{\varepsilon}\right) = \frac{1}{\varepsilon}$$

$$\begin{aligned}
E\left(\|c_0\|_\infty\right) &= E\left(\max_i y_i\right) \\
&= \frac{\ln(n)}{\varepsilon} + E\left(\max_i \left(y_i - \frac{\ln(n)}{\varepsilon}\right)\right) \\
&\leq \frac{\ln(n)}{\varepsilon} + \sum_i E\left(\max\left\{y_i - \frac{\ln(n)}{\varepsilon}, 0\right\}\right)
\end{aligned}$$

$$E\left(\|c_0\|_\infty\right) \leq \frac{\ln(n)}{\varepsilon} + \sum_i \operatorname{Prob}\left(y_i > \frac{\ln(n)}{\varepsilon}\right).$$

$$E\left(y_i - \frac{\ln(n)}{\varepsilon} \,\middle|\, y_i > \frac{\ln(n)}{\varepsilon}\right)$$

$$= \frac{\ln(n)}{\varepsilon} + n \cdot \left(\frac{1}{n}\right) \cdot \frac{1}{\varepsilon}$$

$$\leq \frac{2\ln(n)}{\varepsilon} = O\left(\frac{\log n}{\varepsilon}\right), \quad n \geq 3.$$

FPL_α , $\varepsilon = \sqrt{D/RAT}$

Corollary 4: Suppose $d_t(c_0, c+c_0) \le \delta$

$\forall c \in \{c_1, \ldots, c_T\}$

Then $FPL(1..T) \le BPL(1..T) + \delta RT$

$d_x(r_0, c+c_0) \le \delta$ $c \cdot x \ge 0$

$\forall x \in \varphi$

$FPL(1..T) \le (1+\delta)BPL(1..T)$

$$FPL_\alpha(1..T) \lesssim BPL(1..T) + 2\varepsilon RAT \qquad \epsilon = \sqrt{D/RAT}$$

Lemma: c, s.t $\|c\|_1 \lesssim A$. c_0 is a random

Sample from α

$$d_+(c_0, c+c_0) \lesssim \epsilon A$$

$c_0 \ldots \mu$

$$d_\times(c_0, c+c_0) \lesssim e^{\varepsilon A} - 1.$$

$$FPL_\alpha \lesssim BPL_\alpha(1..T) + \delta RT$$

$$\lesssim BPL_\alpha(1..T) + \epsilon . ART$$

$$\epsilon = \sqrt{D/RAT}$$

$$FPL_\alpha(1..T) \le OPT(1..T) + \epsilon ART + \frac{D}{\epsilon}$$

$$= OPT(1..T) + O(\sqrt{DRAT})$$

$\forall \epsilon \le 1/A$

$$FPL_\mu(1..T) \le e^{\epsilon A} BPL_\mu(1..T)$$

$$\le (1 + O(\epsilon A))\, OPT(1..T)$$

$$+ \frac{2\ln(n)D}{\epsilon}.$$