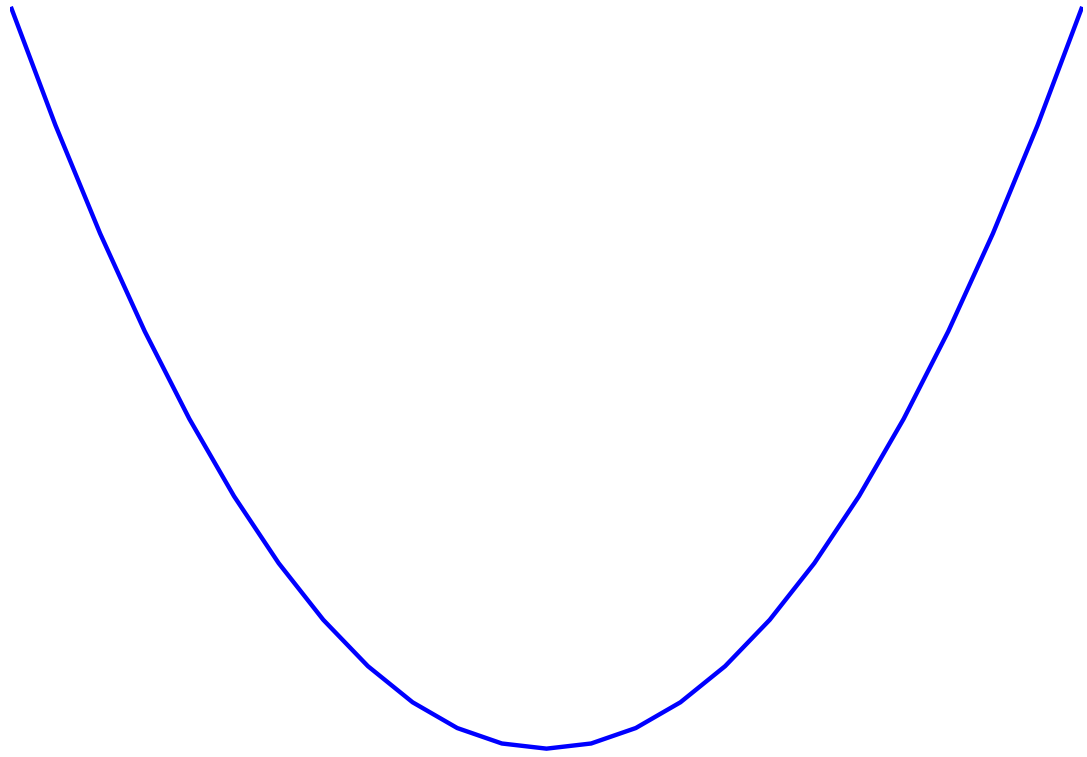
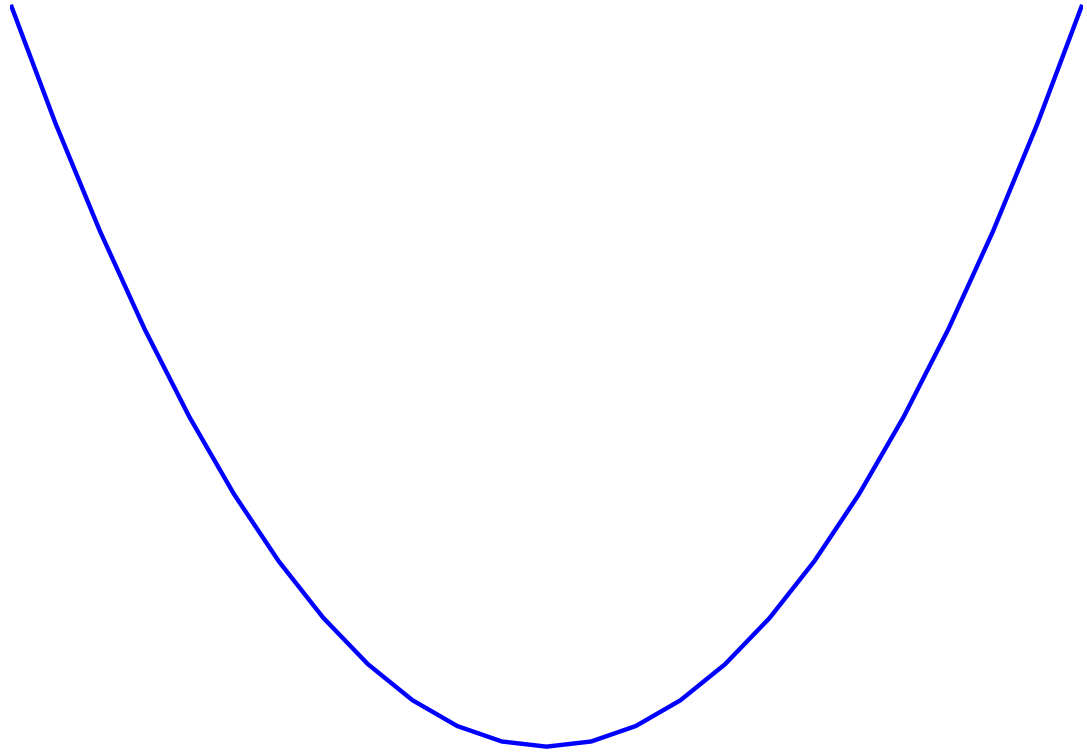
Bandit Algorithms (part 2)

Tor Lattimore

Convex functions



Convex functions



The sum of convex functions is convex

Strictly convex functions have unique minimizers

Online convex optimisation

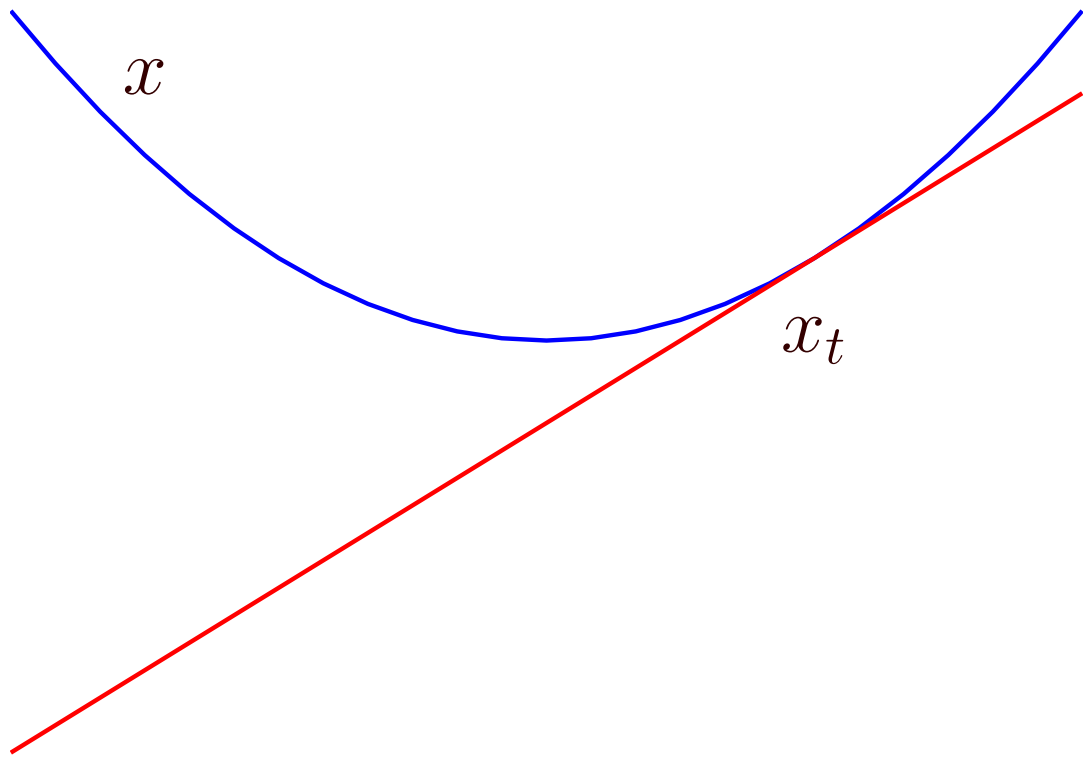
- A game over n rounds
- $\mathcal{K} \subset \mathbb{R}^d$ is convex
- Learner chooses $x_t \in \mathcal{K}$
- Adversary chooses convex $f_t : \mathcal{K} \to \mathbb{R}$
- Loss in round t is $f_t(x_t)$ and regret is

$$\mathfrak{R}_n(x) = \sum_{t=1}^{n} f_t(x_t) - \sum_{t=1}^{n} f_t(x)$$

- **Special case** $f_t(x) = \langle x, \ell_t \rangle$ with $\ell_t \in \mathbb{R}^d$

Linearisation

$$f_t(x) \geq f_t(x_t) + \langle x - x_t, \nabla f_t(x_t) \rangle$$



Rearranging, $f_t(x_t) - f_t(x) \leq \langle x_t - x, \nabla f_t(x_t) \rangle$

Linearisation and first order methods

- Regret is bounded by

$$\begin{aligned}\mathfrak{R}_n(x) &= \sum_{t=1}^{n} f_t(x_t) - \sum_{t=1}^{n} f_t(x) \\ &\leq \sum_{t=1}^{n} \langle x_t - x, \nabla f_t(x_t) \rangle\end{aligned}$$

- Reduction from nonlinear to linear
- Only uses first order information (the gradient)
- **Linear losses from now on** $f_t(x) = \langle x, \ell_t \rangle$
- Think of $\ell_t = \nabla f_t(x_t)$

Optimisation

- You probably know gradient descent

$$x_{t+1} = x_t - \eta \ell_t$$

- Game theorists might try fictitious play

$$x_{t+1} = \operatorname{argmin}_{x \in \mathcal{K}} \sum_{s=1}^{t} \langle x, \ell_s \rangle$$

- Or maybe you know about exponential weights

$$x_{t+1,i} = \frac{\exp\left(-\eta \sum_{s=1}^{t} \ell_{s,i}\right)}{\sum_{j=1}^{d} \exp\left(-\eta \sum_{s=1}^{t} \ell_{s,j}\right)}$$

Unified view

- **Follow the regularized leader**

$$x_{t+1} = \operatorname{argmin}_{x \in \mathcal{K}} \left(\eta \sum_{s=1}^{t} \langle x, \ell_s \rangle + F(x) \right)$$

- F is a convex function, the **regularizer** or **potential**
- $\eta > 0$ is the **learning rate**
- Different choices of F lead to different algorithms
- One clean analysis

Example – No regularization

- $F(x) = 0$
- $x_{t+1} = \operatorname{argmin}_{x \in \mathcal{K}} \sum_{s=1}^{t} \langle x, \ell_s \rangle$
- Fictitious play
- Also called **follow the leader**
- Does not work in general

Example – Gradient descent

- $\mathcal{K} = \mathbb{R}^d$ and $F(x) = \frac{1}{2}\|x\|_2^2$

$$x_{t+1} = \operatorname{argmin}_{x \in \mathcal{K}} \eta \sum_{s=1}^{t} \langle x, \ell_s \rangle + \frac{1}{2}\|x\|_2^2$$

Example – Gradient descent

- $\mathcal{K} = \mathbb{R}^d$ and $F(x) = \frac{1}{2}\|x\|_2^2$

$$x_{t+1} = \text{argmin}_{x \in \mathcal{K}} \eta \sum_{s=1}^{t} \langle x, \ell_s \rangle + \frac{1}{2}\|x\|_2^2$$

- Differentiating,

$$0 = \eta \sum_{s=1}^{t} \ell_s + x$$

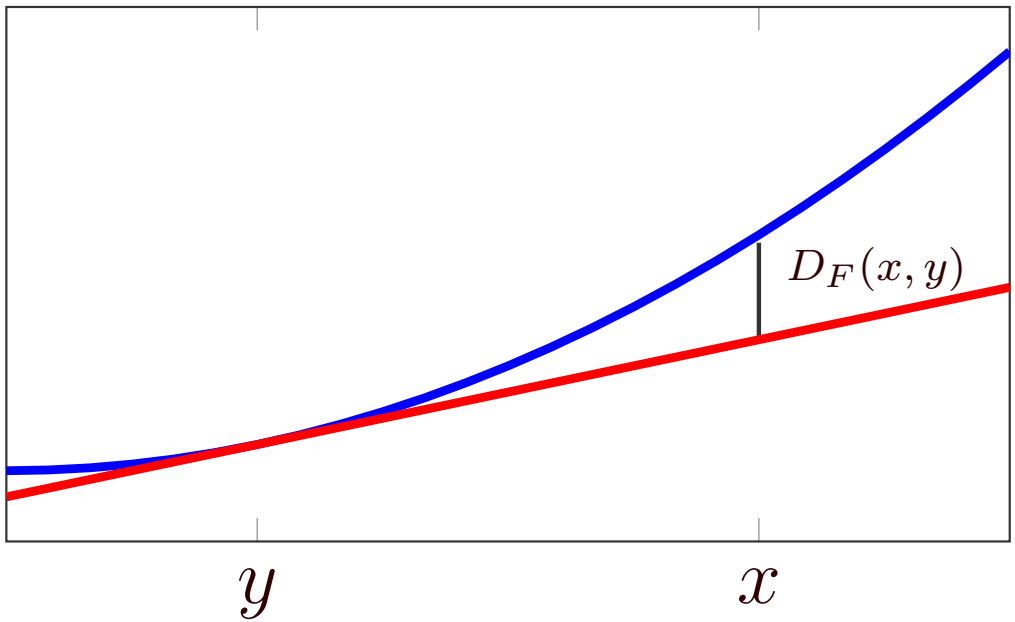
- $x_{t+1} = -\eta \sum_{s=1}^{t} \ell_s = x_t - \eta \ell_s$

A few tools

- Online convex optimization uses many tools from convex analysis
- Bregman divergence
- First-order optimality conditions
- Dual norms

Bregman divergence

For convex F, $D_F(x, y) = F(x) - F(y) - \langle \nabla F(y), x - y \rangle$



$F(x)$

$F(y) + \langle \nabla F(y), x - y \rangle$

- Bregman divergence is not a distance, but behaves a bit like one
- $D_F(x, y) \approx (x - y)^\top \nabla^2 F(y)(x - y) = \|x - y\|^2_{\nabla^2 F(y)}$

Examples

- **Quadratic** $F(x) = \frac{1}{2} \|x\|^2$
- **Negentropy** $F(x) = \sum_{i=1}^{d} x_i \log(x_i) - x_i$
- $F(x) = -2 \sum_{i=1}^{d} \sqrt{x_i}$

First order optimality conditions

- Let $\mathcal{K}$ be convex, $f : \mathcal{K} \to \mathbb{R}$ convex, differentiable

$$x^* = \operatorname{argmin}_{x \in \mathcal{K}} f(x) \Leftrightarrow \langle \nabla f(x^*), x - x^* \rangle \geq 0 \ \forall x \in \mathcal{K}$$

- **Interpretation** f is increasing in direction $x - x^*$ for all $x \in \mathcal{K}$

FTRL analysis

- Rewriting the regret

$$\begin{aligned}\mathfrak{R}_n(x) &= \sum_{t=1}^{n} \langle x_t - x, \ell_t \rangle \\ &= \sum_{t=1}^{n} \langle x_t - x_{t+1}, \ell_t \rangle + \sum_{t=1}^{n} \langle x_{t+1} - x, \ell_t \rangle\end{aligned}$$

- The second term needs some massaging

- $\Phi_t(x) = \dfrac{F(x)}{\eta} + \sum_{s=1}^{t} \langle x, \ell_s \rangle$
- Then

$$\sum_{t=1}^{n} \langle x_{t+1} - x, \ell_t \rangle$$

- $\Phi_t(x) = \dfrac{F(x)}{\eta} + \sum_{s=1}^{t} \langle x, \ell_s \rangle$
- Then

$$\sum_{t=1}^{n} \langle x_{t+1} - x, \ell_t \rangle = \sum_{t=1}^{n} \langle x_{t+1}, \ell_t \rangle - \Phi_n(x) + \frac{F(x)}{\eta}$$

- $\Phi_t(x) = \dfrac{F(x)}{\eta} + \sum_{s=1}^{t} \langle x, \ell_s \rangle$

- Then

$$\sum_{t=1}^{n} \langle x_{t+1} - x, \ell_t \rangle = \sum_{t=1}^{n} \langle x_{t+1}, \ell_t \rangle - \Phi_n(x) + \frac{F(x)}{\eta}$$

$$= \sum_{t=1}^{n} (-\Phi_{t-1}(x_{t+1}) + \Phi_t(x_{t+1})) - \Phi_n(x) + \frac{F(x)}{\eta}$$

- $\Phi_t(x) = \dfrac{F(x)}{\eta} + \sum_{s=1}^{t} \langle x, \ell_s \rangle$

- Then

$$\sum_{t=1}^{n} \langle x_{t+1} - x, \ell_t \rangle = \sum_{t=1}^{n} \langle x_{t+1}, \ell_t \rangle - \Phi_n(x) + \frac{F(x)}{\eta}$$

$$= \sum_{t=1}^{n} (-\Phi_{t-1}(x_{t+1}) + \Phi_t(x_{t+1})) - \Phi_n(x) + \frac{F(x)}{\eta}$$

$$\leq \frac{F(x) - F(x_1)}{\eta} - \frac{1}{\eta} \sum_{t=1}^{n} D_F(x_t, x_{t+1})$$

Putting it together

$$\mathfrak{R}_n(x) \leq \frac{F(x) - F(x_1)}{\eta} + \sum_{t=1}^{n} \langle x_t - x_{t+1}, \ell_t \rangle - \frac{D_F(x_t, x_{t+1})}{\eta}$$

- First term is “distance from start to goal”
- Second term ????

A reminder about dual norms

- Suppose we have a norm $\|\cdot\|$ on $\mathbb{R}^k$
- Dual norm is a norm on $\mathbb{R}^k$ (on the dual space, technically)

$$\|\ell\|_* = \sup_{x \in \mathbb{R}^k} \frac{\langle x, \ell \rangle}{\|x\|}$$

- $\|x\| = \|x\|_2 \iff \|\ell\|_* = \|\ell\|_2$
- $\|x\| = \|x\|_A = \sqrt{x^\top A x} \iff \|\ell\|_* = \|\ell\|_{A^{-1}}$
- Cauchy-Schwarz is immediate $\langle x, \ell \rangle \leq \|x\| \, \|\ell\|_*$

The second term

$$\langle x_t - x_{t+1}, \ell_t \rangle - \frac{D_F(x_{t+1}, x_t)}{\eta}$$

Suppose that $D_F(x_{t+1}, x_t) \geq \frac{1}{2} \|x_{t+1} - x_t\|_t^2$

The second term

$$\langle x_t - x_{t+1}, \ell_t \rangle - \frac{D_F(x_{t+1}, x_t)}{\eta}$$

Suppose that $D_F(x_{t+1}, x_t) \geq \frac{1}{2} \|x_{t+1} - x_t\|_t^2$

$$\begin{aligned}
&\langle x_{t+1} - x_t, \ell_t \rangle - \frac{1}{\eta} D_F(x_{t+1}, x_t) \\
&\quad \leq \langle x_{t+1} - x_t, \ell_t \rangle - \frac{1}{2\eta} \|x_{t+1} - x_t\|_t^2 \\
&\quad \leq \|x_{t+1} - x_t\|_t \|\ell_t\|_{t*} - \frac{1}{2\eta} \|x_{t+1} - x_t\|_t^2 \\
&\quad \leq \frac{\eta}{2} \|\ell_t\|_{t*}^2
\end{aligned}$$

$$ax - bx^2/2 \leq a^2/b$$

Final form

$$\mathfrak{R}_n(x) \leq \frac{F(x) - F(x_1)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|\ell_t\|_{t^*}^2$$

- Regret depends on distance from start to optimal
- Magnitude of the losses
- Learning rate needs careful tuning

Application 1: Online gradient descent

Assume $\mathcal{K} = \{x : \|x\|_2 \leq 1\}$ and $\ell_t \in \mathcal{K}$ for all t

$$D_F(x, y) = \frac{1}{2} \|x - y\|_2^2$$

Then

$$\mathfrak{R}_n(x) \leq \frac{\|x^*\|_2^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|\ell_t\|_2^2 \leq \frac{1}{2\eta} + \frac{\eta n}{2} \leq \sqrt{n}$$

Application 2: Exponential weights

Assume $\mathcal{K} = \{x \geq 0 : \|x\|_1 = 1\}$ and $\ell_t \in [0,1]^d$ for all t

$$F(x) = \sum_{i=1}^{d} x_i \log(x_i) - x_i$$

Bregman divergence

$$D_F(x,y) = \sum_{i=1}^{d} x_i \log\left(\frac{x_i}{y_i}\right) \approx \frac{1}{2}\sum_{i=1}^{d} \frac{(x_i - y_i)^2}{y_i}$$
$$= \frac{1}{2}\|x - y\|^2_{\mathrm{diag}(1/y)}$$

Dual norm of $\|\cdot\|_{\mathrm{diag}(1/y)}$ is $\|\cdot\|_{\mathrm{diag}(y)}$

Application 2: Exponential weights

Assume $\mathcal{K} = \{x \geq 0 : \|x\|_1 = 1\}$ and $\ell_t \in [0,1]^d$ for all t

$$x = \operatorname{argmin}_{x \in \mathcal{K}} \sum_{t=1}^{n} \langle x, \ell_t \rangle$$

Optimal action is a standard basis vector

$$\begin{aligned}
\mathfrak{R}_n(x) &\lesssim \frac{F(x) - F(x_1)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|\ell_t\|_{t*}^2 \\
&\leq \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \sum_{i=1}^{d} x_{t,i} \ell_{t,i}^2 \\
&\leq \frac{\log(d)}{\eta} + \frac{\eta n}{2} \leq \sqrt{2n \log(d)}
\end{aligned}$$

Application 3: Online gradient descent

Assume $\mathcal{K} = \{x \geq 0 : \|x\|_1 = 1\}$ and $\ell_t \in [0,1]^d$ for all t

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

$$\mathfrak{R}_n(x) \leq \frac{\|x\|_2^2}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|\ell_t\|_2^2$$

$$\leq \frac{1}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|\ell_t\|_2^2 \leq \frac{1}{\eta} + \frac{\eta dn}{2} \leq \sqrt{dn}$$

Adversarial bandits

- A different model
- No statistical assumptions!
- At the start of the game the **adversary** secretly chooses losses $\ell_1, \ldots, \ell_n$ with $\ell_t \in [0,1]^k$
- In each round the learner chooses $A_t \in [k]$
- Suffers loss ℓ_{t,A_t}
- Regret is $\mathfrak{R}_n = \max_a \mathbb{E}\left[\sum_{t=1}^n \ell_{t,A_t} - \ell_{t,a}\right]$
- **Surprising result** there exists an algorithm such that $\mathfrak{R}_n \leq \sqrt{2nk\log(k)}$ for any adversary

Adversarial bandits

- Randomization is crucial (like game theory)
- Given a deterministic algorithm
- Adversary chooses $\ell_{t,a} = \mathbb{1}(A_t = a)$
- Linear regret

Adversarial bandits

- Randomization is crucial (like game theory)
- Given a deterministic algorithm
- Adversary chooses $\ell_{t,a} = \mathbb{1}(A_t = a)$
- Linear regret
- **Core idea** Estimate the loss vector and apply algorithms for online linear optimization

Importance-weighted estimators

Algorithm chooses $A_t = a$ with probability $P_t(a)$

$$\hat{\ell}_{t,a} = \frac{\ell_{t,a} \mathbb{1}(A_t = a)}{P_t(a)}$$

Unbiased estimator,

$$\begin{aligned} \mathbb{E}\left[\hat{\ell}_{t,a} \mid P_t\right] &= \frac{\ell_{t,a}}{P_t(a)} \mathbb{E}[\mathbb{1}(A_t = a) \mid P_t] \\ &= \ell_{t,a} \end{aligned}$$

Large second moment: $\mathbb{E}\left[\hat{\ell}_{t,a}^2 \mid P_t\right] = \frac{\ell_{t,a}^2}{P_t(a)}$

Exp3 algorithm

- Estimate ℓ_t using importance-weighted estimator
- Then follow the regularized leader with negentropy
- FTRL recommends a point P_t in the simplex
- Algorithm chooses $A_t \sim P_t$

$$P_{t,a} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \hat{\ell}_{s,a}\right)}{\sum_{b=1}^{k} \exp\left(-\eta \sum_{s=1}^{t-1} \hat{\ell}_{s,b}\right)}$$

Back to bandits

$$\mathfrak{R}_n = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a^*, \ell_t \rangle\right]$$

Back to bandits

$$\mathfrak{R}_n = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a^*, \ell_t \rangle\right] = \mathbb{E}\left[\sum_{t=1}^{n} \langle P_t - a^*, \hat{\ell}_t \rangle\right]$$

Back to bandits

$$\mathfrak{R}_n = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a^*, \ell_t \rangle\right] = \mathbb{E}\left[\sum_{t=1}^{n} \langle P_t - a^*, \hat{\ell}_t \rangle\right]$$

$$\leq \frac{\log(k)}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n}\sum_{a=1}^{k} P_t(a)\hat{\ell}_{t,a}^2\right]$$

Back to bandits

$$\mathfrak{R}_n = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a^*, \ell_t \rangle\right] = \mathbb{E}\left[\sum_{t=1}^{n} \langle P_t - a^*, \hat{\ell}_t \rangle\right]$$

$$\leq \frac{\log(k)}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n}\sum_{a=1}^{k} P_t(a)\hat{\ell}_{t,a}^2\right]$$

$$= \frac{\log(k)}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n}\sum_{a=1}^{k} \ell_{t,a}^2\right]$$

Back to bandits

$$\mathfrak{R}_n = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a^*, \ell_t \rangle\right] = \mathbb{E}\left[\sum_{t=1}^{n} \langle P_t - a^*, \hat{\ell}_t \rangle\right]$$

$$\leq \frac{\log(k)}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n}\sum_{a=1}^{k} P_t(a)\hat{\ell}_{t,a}^2\right]$$

$$= \frac{\log(k)}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n}\sum_{a=1}^{k} \ell_{t,a}^2\right]$$

$$\leq \frac{\log(k)}{\eta} + \frac{\eta n k}{2}$$

Back to bandits

$$\mathfrak{R}_n = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a^*, \ell_t \rangle\right] = \mathbb{E}\left[\sum_{t=1}^{n} \langle P_t - a^*, \hat{\ell}_t \rangle\right]$$

$$\leq \frac{\log(k)}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n}\sum_{a=1}^{k} P_t(a)\hat{\ell}_{t,a}^2\right]$$

$$= \frac{\log(k)}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n}\sum_{a=1}^{k} \ell_{t,a}^2\right]$$

$$\leq \frac{\log(k)}{\eta} + \frac{\eta n k}{2} = \sqrt{2nk\log(k)}$$

Summary

- Online linear optimization
- Importance-weighted estimates
- Application to finite-armed bandits
- **Tomorrow** Lot’s more applications

Online mirror descent

- Online mirror descent

$$x_1 = \operatorname{argmin}_{x \in \mathcal{K}} f(x)$$
$$x_{t+1} = \operatorname{argmin}_{x \in \mathcal{K}} \eta \langle x, \ell_t \rangle + D_F(x, x_t)$$

- FTRL

$$x_{t+1} = \operatorname{argmin}_{x \in \mathcal{K}} \eta \sum_{s=1}^{t} \langle x, \ell_s \rangle + F(x)$$

- These algorithms are often the same

Mirror descent in continuous time

- Mirror descent
  $x_{t+1} = \operatorname{argmin}_{x \in \mathcal{K}} \eta \langle x, \ell_t \rangle + D_F(x, x_t)$
- What happens in continuous time?

$$D_F(x, x_t) \approx \frac{1}{2} \|x - x_t\|^2_{\nabla^2 F(x_t)}$$

- Then

$$x_{t+1} \approx x_t - \eta \nabla^2 F(x_t)^{-1} \ell_t$$

- Taking limits $\eta \to 0$, $\dot{x}(t) = -\nabla^2 F(x_t)^{-1} \ell(t)$

- Dynamical system $\dot{x}(t) = -\alpha \nabla^2 F(x_t)^{-1} \ell(t)$

$$\begin{aligned}
&\frac{d}{dt} D(x, x(t)) \\
&\qquad = \frac{d}{dt} \left(F(x) - F(x(t)) - \langle \nabla F(x(t)), x - x(t) \rangle \right) \\
&\qquad = -\langle \dot{x}(t) \nabla^2 F(x(t)), x - x(t) \rangle \\
&\qquad = -\alpha \langle \ell(t), x(t) - x \rangle
\end{aligned}$$

- Regret is

$$\begin{aligned}
\int_0^n \langle \ell(t), x(t) - x \rangle dt &= -\frac{1}{\alpha} \int_0^n \frac{d}{dt} D(x, x(t)) dt \\
&= \frac{D_F(x, x(0)) - D_F(x, x(n))}{\alpha}
\end{aligned}$$

From continuous time to discrete

$$\sum_{t=1}^{n} \langle x(t-1) - x, \ell(t) \rangle$$

$$= \sum_{t=0}^{n-1} \int_0^1 \langle x(t) - x(t+s), \ell(t+1) \rangle ds + \int_0^n \langle x(t), \ell(t) \rangle dt$$

$$\approx -\sum_{t=1}^{n} \int_0^1 s \langle \dot{x}(t), \ell(t) \rangle ds + \int_0^n \langle x(t), \ell(t) \rangle dt$$

$$\approx \frac{\alpha}{2} \sum_{t=1}^{n} \|\ell(t)\|_*^2 + \frac{D_F(x, x(0))}{\alpha}$$