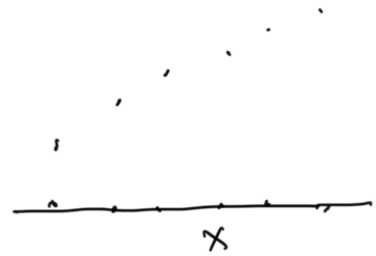
Linear Bandits

$$\text{reward } y_t = \underbrace{\langle x_t, \theta^* \rangle}_{} + \varepsilon_t \qquad \mu(x_t) = \langle x_t, \theta^* \rangle$$



$\theta^* \in$ d-dim parameters

$$x^* = \arg\max_x \langle x, \theta^* \rangle$$

Choose $x_1 \dots x_T$

Minimize cumulative regret
$$\sum_{t=1}^{T} \mu(x^*) - \mu(x_t)$$
$$= \sum_{t=1}^{T} \langle x^* - x_t, \theta^* \rangle$$
$$= E\left[\sum_{t=1}^{T} \langle x^* - x_t, \hat{\theta}_t \rangle \right]$$
↳ estimate (unbiased)

Successive elimination (finite arms)

Linear case

active arms $\mathcal{X}_1 \leftarrow \mathcal{X}$ all arms

while $|\mathcal{X}_t| > 1$

sample ideally $O(d)$ arms

G-opt expt design
- in practice sample $\frac{d(d+1)}{2} \sim d^2$ arms each sample
- $\lceil k_t w_i \rceil$ where $w_i \in$ weight from G-opt design
- $k_t \sim d\, 2^{2t} \log \frac{t|\mathcal{X}|}{\delta}$

construct $\hat{\theta}_t$

deactivate $x \in \mathcal{X}_t$ if

$UCB_t(x) < LCB_t(x')$

$\langle \hat{\theta}_t, x \rangle + \sigma_t < \langle \hat{\theta}_t, x' \rangle - \sigma_t$

$\sigma_t \sim$ confidence sequence

Independent arms

active $\mathcal{X}_1 \leftarrow \mathcal{X}$ all arms

while $|\mathcal{X}_t| > 1$

[sample each active arms x_i once $\in \mathcal{X}_t$

construct $\hat{\mu}_i$

deactivate $x \in \mathcal{X}_t$ if

$UCB_t(x) < LCB_t(x')$, $x' \in \mathcal{X}_t$

HW1, prob 3

cumulative regret $- \sqrt{dT} \quad \in \sqrt{\frac{d}{\pi}}$

UCB (Linear)

$\hat{\theta}_1 = 0$

$x_t = \arg\max_x \; UCB_t(x) = \hat{\mu}_t(x) + \sigma_t(x)$ ✓

$= \langle \hat{\theta}_t, x_t \rangle + \sigma_t(x)$

observe y_t

update $\hat{\theta}_t$

$UCB_t(x) = \max_{\theta \in C_t} \langle x, \theta \rangle$

$C_t \to$ confidence interval on parameter space

Confidence set $C_t = \{\theta : \|\theta - \hat{\theta}_{t-1}\|^2_{V_{t-1}} \le \beta_t\}$

Given $\{x_s, y_s\}_{s=1}^{t-1}$, $\hat{\theta}_t = \arg\min_\theta \sum_{s=1}^{t-1} (\langle x_s, \theta \rangle - y_s)^2 + \lambda \|\theta\|_2^2$

Ridge estimator

$= \left(\sum_{s=1}^{t-1} x_s x_s^T + \lambda I \right)^{-1} \sum_{s=1}^{t-1} x_s y_s$

$(XX^T + \lambda I)^{-1} X^T Y$

$V_{t-1} = \sum_{s=1}^{t-1} x_s x_s^T + \lambda I$

(A1) β_t - increasing seq^n, ≥ 1

(A2) whp $\theta^* \in C_t$

online confidence set for parameter vector

whp $\|\theta^* - \hat{\theta}_{t-1}\|^2_{V_{t-1}} \le \beta_t$

(A3) $\max_x \langle x_* - x, \theta_* \rangle \le 1$ bounded rewards

(A4) $\|x\|_2 \le L \quad \forall x$

✓ $\sqrt{dT} \equiv \sqrt{\frac{d}{n}}$

Bound cumulative regret.

$R(T) = \sum_{t=1}^{T} \langle x_* - x_t, \theta_* \rangle \underset{\text{will show}}{=} O(d\sqrt{T}\log T)$ under (A1) - (A4)

Proof:

$\langle x_*, \theta_* \rangle \le UCB_t(x^*) \underset{\text{Algo picks } x_t}{\le} UCB_t(x_t) = \langle x_t, \tilde{\theta}_t \rangle$ ☆ , $\tilde{\theta}_t = \arg\max_{\theta \in C_t} \langle x_t, \theta \rangle$

$\langle x_* - x_t, \theta_* \rangle = \langle x_*, \theta_* \rangle - \langle x_t, \theta_* \rangle$

$$= \underbrace{\langle x_*, \theta_* \rangle - \langle x_t, \tilde\theta_t \rangle} + \langle x_t, \tilde\theta_t \rangle - \langle x_t, \theta_* \rangle$$

$$\underset{*}{\leq} 0 + \langle x_t, \tilde\theta_t - \theta^* \rangle = x_t^T V_{t-1}^{-1/2} V_{t-1}^{1/2} (\tilde\theta_t - \theta^*)$$

$$\leq \|x_t\|_{V_{t-1}^{-1}} \|\tilde\theta_t - \theta_*\|_{V_{t-1}} \quad \text{Cauchy-Schwartz}$$

$$\leq \|x_t\|_{V_{t-1}^{-1}} \left(\underbrace{\|\tilde\theta_t - \hat\theta_t\|_{V_{t-1}}}_{\leq \sqrt{\beta_t}} + \underbrace{\|\hat\theta_t - \theta_*\|_{V_{t-1}}}_{\leq \sqrt{\beta_t}} \right) \quad \text{Triangle ineq.}$$

bcoz $\tilde\theta_t, \theta_* \in C_t$ (A2)

$$\leq 2\sqrt{\beta_t} \|x_t\|_{V_{t-1}^{-1}}$$

$$\langle x_* - x_t, \theta^* \rangle \leq \min(1, 2\sqrt{\beta_t} \|x_t\|_{V_{t-1}^{-1}}) \quad \text{(A3)}$$

$$\leq 2\sqrt{\beta_t} \min(1, \|x_t\|_{V_{t-1}^{-1}}) \quad \text{(A1)} \quad \beta_t \geq 1$$

Elliptical potential lemma

$u \geq 0$

$\min(u,1) \leq 2\log(1+u)$

$$\to \sum_{t=1}^{T} \min(1, \|x_t\|^2_{V_{t-1}}) \leq 2\sum_{t=1}^{T} \log(1 + \|x_t\|^2_{V_{t-1}})$$

$$\leq 2\log \frac{|V_T|}{|V_0|}$$

$$\sim d \log(T, d)$$

$V_0 = \lambda I$

$V_t = \sum_{s=1}^{t-1} x_s x_s^T + \lambda I$

$$V_t = V_{t-1} + x_t x_t^T$$

$$= V_{t-1}^{1/2} (I + V_{t-1}^{-1/2} x_t x_t^T V_{t-1}^{-1/2}) V_t^{1/2}$$

$|AB| = |A||B|$

$$|V_t| = |V_{t-1}| \, | I + \underbrace{V_{t-1}^{-1/2} x_t x_t^T V_t^{-1/2}}_{zz^T} |$$

$z = V_t^{-1/2} x_t$

$d\times 1 \quad d\times d \quad d\times 1$

$v^T M v = \lambda$

$\|v\|_2 = 1$

$$|A| = \prod_{i=1}^{d} \lambda_i$$

$v^T z z^T v = \|v^T z\|^2$

eval of $I + V_{t-1}^{-1/2} x_t x_t^T V_t^{-1/2} = \begin{cases} 1 + \|V_t^{-1/2} x_t\|^2 = 1 + \|x_t\|^2_{V_{t-1}} \\ 1 \quad d-1 \end{cases}$

$$\Rightarrow |V_t| = |V_{t-1}| \left(1 + \|x_t\|^2_{V_{t-1}}\right)$$

$$|V_T| = |V_0| \prod_{t=1}^{T} \left(1 + \|x_t\|^2_{V_{t-1}}\right)$$

$$\Rightarrow \log \frac{|V_T|}{|V_0|} = \sum_{t=1}^{T} \log \left(1 + \|x_t\|^2_{V_{t-1}}\right)$$

$$|V_T| = \prod_{i=1}^{d} \lambda_i \le \left(\frac{1}{d} \sum_i \lambda_i\right)^d = \left(\frac{tr(V_T)}{d}\right)^d \le \left(\frac{TL^2 + d\lambda}{d}\right)^d$$

geo $\le$ arith

$$\to V_T = \sum_{t=1}^{T} \underbrace{x_t x_t^T}_{} + \lambda I$$

$$eval(x_t x_t^T) = \begin{cases} \|x_t\|_2^2 \le L & 1 \\ 0 & d-1 \end{cases}$$

$$tr(V_T) \le TL^2 + d\lambda$$

$$|V_0| = |\lambda I| = \lambda^d$$

$$\Rightarrow \log \frac{|V_T|}{|V_0|} = d \log(T, d)$$

$$R(T) = \sum_{t=1}^{T} \langle x_* - x_t, \theta_* \rangle \le \sqrt{\sum_{t=1}^{T} 1 \cdot \sum_{t=1}^{T} \langle x_* - x_t, \theta_* \rangle^2}$$

$$= \sqrt{T \cdot \sum_{t=1}^{T} 4 \beta_t \min\left(1, \|x_t\|^2_{V_{t-1}^{-1}}\right)}$$

$$= O\left(\sqrt{T \beta_T \, d \log(T, d)}\right)$$

$\sqrt{dT\beta_T}$

$\beta_T \sim d$

$\underline{\underline{d\sqrt{T}}}$

Implementation

$$\max_x \quad UCB_t(x) = \max_{\theta \in C_t} \langle x, \theta \rangle$$

$$\max_x \quad UCB'_t(x) = \langle \hat{\theta}_{t-1}, x \rangle + \underbrace{\sqrt{\beta_t} \|x\|_{V_{t-1}^{-1}}}_{\sigma_t(x)}$$

easier to optimize

$$\langle x, \theta \rangle = \langle x, \hat{\theta}_{t-1} \rangle + \langle x, \theta - \hat{\theta}_{t-1} \rangle$$

$$\leq \langle x, \hat{\theta}_{t-1} \rangle + \|x\|_{V_{t-1}^{-1}} \underbrace{\|\theta - \hat{\theta}_{t-1}\|_{V_{t-1}}}_{\leq \sqrt{\beta_t}} \qquad \theta \in C_t$$

$$= \langle \hat{\theta}_{t-1}, x \rangle + \sqrt{\beta_t} \|x\|_{V_{t-1}^{-1}}$$