Fix a set of arms/vectors $\mathcal{X} \subset \mathbb{R}^d$ (e.g. $\mathcal{X} = (x_1, x_2, \ldots)$, $x_i \in \mathbb{R}^d$)

We assume $\exists\, \theta_* \in \mathbb{R}^d$ s.t. pulling arm $x \in \mathcal{X}$ results in observation $y = \langle x, \theta_* \rangle + \eta$, $\quad \mathbb{E}[\eta] = 0$

$$\mathbb{E}[\exp(\lambda \eta)] \leq \exp(\lambda^2/2)$$

for $t = 1, 2, \ldots, T$

  Player chooses $x_t \in \mathcal{X}$

  Nature reveals $y_t = \langle x_t, \theta_* \rangle + \eta_t$

Player aims to minimize regret $\max_{x \in \mathcal{X}} \sum_{t=1}^{T} \langle x - x_t, \theta_* \rangle$

Suppose you choose $x_1, \ldots, x_T \in \mathbb{R}^d$ and then observe $y_t = \langle x_t, \theta_* \rangle + \eta_t \quad \forall t$.

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{T} (\langle x_t, \theta \rangle - y_t)^2$$

$$X = \begin{pmatrix} - x_1^T - \\ \vdots \\ - x_T^T - \end{pmatrix} \qquad 2 \sum_t x_t (x_t^T \theta - y_t) = 0$$

$$\left(\sum x_t x_t^T\right) \theta = \sum_t x_t y_t$$

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix} = X\theta_* + \eta$$

$$\hat{\theta} = \left(\sum_t x_t x_t^T\right)^{-1} \sum_t x_t y_t$$

$$\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_T \end{pmatrix} \qquad = (X^T X)^{-1} X^T Y$$

$$= \theta_* + (X^T X)^{-1} X^T \eta$$

$$\mathbb{E}[\hat{\theta}] = \theta_* + \mathbb{E}[\underbrace{(X^T X)^{-1} X^T \eta}] = \theta_*$$

$$= \sum_{t=1}^{T} (X^T X)^{-1} x_t \eta_t$$

$$\mathbb{E}[(\hat{\theta} - \theta_*)(\hat{\theta} - \theta_*)^T] = \mathbb{E}[(X^T X)^{-1} X^T \eta \eta^T X (X^T X)^{-1}]$$

$A \prec B \quad B - A \succ 0$

$$= (X^T X)^{-1} X^T \mathbb{E}[\eta \eta^T] X (X^T X)^{-1}$$

$$\prec (X^T X)^{-1} X^T X (X^T X)^{-1}$$

$$= (X^T X)^{-1}$$

$(\eta \eta^T)_{i,j} = \eta_i \eta_j$

$(\eta \eta^T)_{i,i} = \eta_i^2$

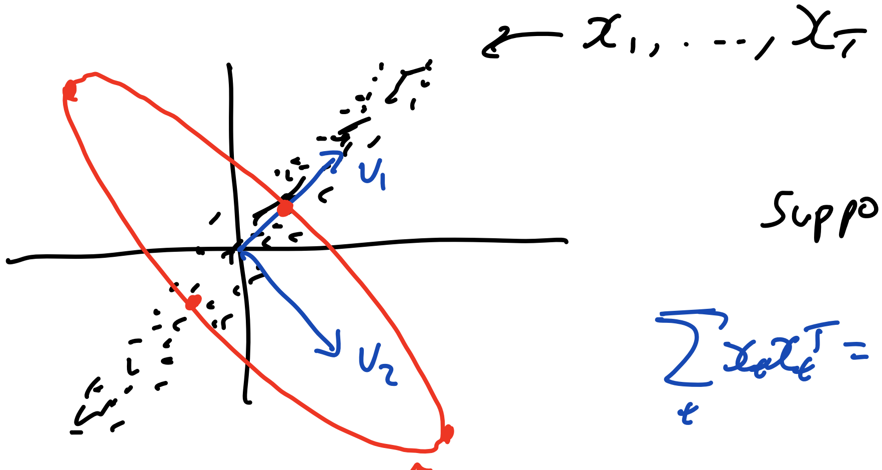
For any z, what is $z^T(\hat{\theta} - \theta_*)$

$$\mathbb{E}[z^T(\hat{\theta} - \theta_*)] = 0$$

$$\mathbb{E}[(z^T(\hat{\theta} - \theta_*))^2] = \mathbb{E}[z^T(\hat{\theta} - \theta_*)(\hat{\theta} - \theta_*)^T z]$$

$$= z^T (X^T X)^{-1} z$$

$z^T(\hat{\theta} - \theta_*)$ is approx $\mathcal{N}(0, z^T (X^T X)^{-1} z)$



Suppose z

$$\sum_t x_t x_t^T = \sum_{i=1}^{2} v_i v_i^T \sigma_i$$

$\{u : u^T (X^T X)^{-1} u = 1\}$ where $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, v_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$\sigma_1 >> \sigma_2$

$$z^T (X^T X)^{-1} z$$

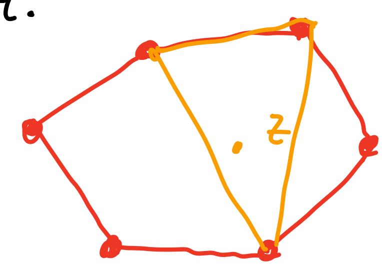
$$z^T \left(v_1 v_1^T \sigma_1 + v_2 v_2^T \sigma_2 \right)^{-1} z$$

$$= z^T \left(v_1 v_1^T \sigma_1^{-1} + v_2 v_2^T \sigma_2^{-1} \right) z$$

$$= (z^T v_1)^2 \sigma_1^{-1} + (z^T v_2)^2 \sigma_2^{-1}$$

For any $x_1, \dots, x_T \in \mathbb{R}^d$, $\exists \lambda \in \Delta_{\mathcal{X}}$ s.t.

$$\sum_{t=1}^{T} x_t x_t^T = T \cdot \underbrace{\sum_{x \in \mathcal{X}} \lambda_x x x^T}_{=: A(\lambda)}$$



"Design matrix"

Caratheodory: Fix $\mathcal{X} \subset \mathbb{R}^d$: For any z: $z = \sum_{x \in \mathcal{X}} p_x x$

$\exists \lambda \in \Delta_{\mathcal{X}}$ s.t. $z = \sum_{x \in \mathcal{X}} \lambda_x x$ w/ $\|\lambda\|_0 \le d+1$,

<u>Claim</u> For any matrix $\frac{1}{T}\sum_{t=1}^{T} x_t x_t^T$ $\exists \lambda \in \Delta_{\mathcal{X}}$ s.t. $\frac{1}{T}\sum_{t=1}^{T} x_t x_t^T = A(\lambda)$ and $\|\lambda\|_0 \leq \frac{d(d+1)}{2}$

Goal: Want to choose $x_1, \ldots, x_T$ s.t. when we construct $\hat{\theta}$ we have

$$\max_{x \in \mathcal{X}} x^T(\hat{\theta} - \theta_*) \leq \varepsilon.$$

Let $\bar{x} = \arg\max_{x \in \mathcal{X}} x^T A(\lambda)^{-1} x$

$\left[\nabla\, x^T A(\lambda)^{-1} x\right]_z = -x^T A(\lambda)^{-1} z z^T A(\lambda)^{-1} x$

Let $\hat{\lambda} = \underset{\lambda \in \Delta_{\mathcal{X}}}{\arg\min} \max_{x \in \mathcal{X}} x^T A(\lambda)^{-1} x$

and let $\|\hat{\lambda}\|_0 \leq \frac{d(d+1)}{2}$. For $\mathcal{T} \in \mathbb{N}$ measure $x \in \mathcal{X}$ exactly $\lceil \lambda_x \mathcal{T} \rceil$ times.

$\Rightarrow$ Total # meas $\leq \mathcal{T} + \frac{d(d+1)}{2}$

$$(X^T X)^{-1} \preceq (T \cdot A(\lambda))^{-1} = \frac{1}{T} A(\lambda)^{-1}$$

$$\Rightarrow \forall x \in \mathcal{X} \quad x^T(\hat{\theta} - \theta_*) \leq \|x\|_{A(\lambda)^{-1}} \cdot \sqrt{\frac{2\log(|\mathcal{X}|/\delta)}{T}}$$

$$\mathbb{E}\left[\exp\left(\lambda z^T(\hat{\theta}-\theta_*)\right)\right] = \mathbb{E}\left[\exp\left(\lambda \underbrace{z^T(X^TX)^{-1}X^T}_{=:\omega^T} \eta\right)\right]$$

$$= \mathbb{E}\left[\exp\left(\lambda \sum_{t=1}^{T} \omega_t \eta_t\right)\right]$$

$$= \mathbb{E}\left[\prod_{t=1}^{T} \exp(\lambda \omega_t \eta_t)\right]$$

$\mathbb{E}[\exp(s\eta_t)] \le \exp(s^2/2)$

$$= \prod_{t=1}^{T} \mathbb{E}\left[\exp(\lambda \omega_t \eta_t)\right]$$

$\|x\|_2^2 = x^T x$

$$\le \prod_{t=1}^{T} \exp\left(\lambda^2 \omega_t^2 / 2\right)$$

$$= \exp\left(\frac{\lambda^2}{2} \sum_{t=1}^{T} \omega_t^2\right)$$

$$= \exp\left(\lambda^2 \omega^T \omega / 2\right)$$

$$= \exp\left(\lambda^2 z^T (X^TX)^{-1} z / 2\right)$$

$\Rightarrow$ $z^T(\hat{\theta}-\theta_*)$ is $z^T(X^TX)^{-1}z$ -sub-Gaussian.

$$\Rightarrow \mathbb{P}\left(z^T(\hat{\theta}-\theta_*) > \varepsilon\right) \le \exp\left(-\varepsilon^2 z^T (X^TX)^{-1} z / 2\right)$$

$\Rightarrow$ w.p. $\ge 1-\delta$ $\quad z^T(\hat{\theta}-\theta_*) \le \sqrt{2\|z\|^2_{(X^TX)^{-1}} \log(1/\delta)}$

where $\|z\|_A^2 := z^T A z$

Lemma 11 (Kiefer-Wolfowitz (1960)). *For any $\mathcal{X}$ with $d = dim(span(\mathcal{X}))$, there exists a $\lambda^* \in \triangle_{\mathcal{X}}$ that*

$[\nabla g(\lambda)]_x = \|x\|^2_{A(\lambda)^{-1}}$

- $\max_\lambda g_D(\lambda) = g_D(\lambda^*)$

$g_D(\lambda) = \log|A(\lambda)|$

- $\min_\lambda f_G(\lambda) = f_G(\lambda^*)$

$f_G(\lambda) = \max_{x\in\mathcal{X}} \|x\|^2_{A(\lambda)^{-1}}$

- $f_G(\lambda^*) = d$
- $|support(\lambda^*)| = (d+1)d/2$

Proposition 2. *If λ^* is the G-optimal design for $\mathcal{X}$ then if we pull arm $x \in \mathcal{X}$ exactly $\lceil \tau\lambda^*_x \rceil$ times for some $\tau > 0$ and compute the least squares estimator $\widehat{\theta}$. Then for each $x \in \mathcal{X}$ we have with probability at least $1-\delta$*

$$\begin{aligned}
\langle x, \widehat{\theta} - \theta^* \rangle &\le \|x\|_{(\sum_{x\in\mathcal{X}} \lceil \tau\lambda^*_x \rceil x x^\top)^{-1}} \sqrt{2\log(1/\delta)} \\
&\le \frac{1}{\sqrt{\tau}} \|x\|_{(\sum_{x\in\mathcal{X}} \lambda^*_x x x^\top)^{-1}} \sqrt{2\log(1/\delta)} \\
&\le \sqrt{\frac{2d\log(1/\delta)}{\tau}}
\end{aligned}$$

and we have taken at most $\tau + \frac{d(d+1)}{2}$ pulls. Thus, for any $\delta' \in (0,1)$ we have $\mathbb{P}(\bigcup_{x\in\mathcal{X}}\{|\langle x, \widehat{\theta} - \theta^*\rangle| > \sqrt{\frac{2d\log(2|\mathcal{X}|/\delta')}{\tau}}\}) \le \delta'$.

Frank-Wolfe

Input $K \subset \mathbb{R}^d$, convex function $f: \mathbb{R}^d \to \mathbb{R}$

Init $x_0 \in K$

for $t = 1, 2, \ldots$

$$z_t = \arg\min_{z\in K} \langle z, \nabla f(x_t) \rangle$$

$$x_{t+1} = (1-\gamma_t)x_t + \gamma_t z_t$$

For D-optimal design

Init λ_0 sparsely.

$$x_t = \arg\max_{x\in\mathcal{X}} [\nabla g(\lambda_t)]_x = \arg\max_x \|x\|^2_{A(\lambda_t)^{-1}}$$

$$\lambda_{t+1} = (1-\gamma_t)\lambda_t + \gamma_t x_t$$

- **A-optimality**: minimize $f_A(\lambda) = \mathrm{Tr}(A_\lambda^{-1})$ minimizes $\mathbb{E}[\|\widehat{\theta} - \theta\|_2^2]$

- **E-optimality**: minimize $f_E(\lambda) = \max_{u:\|u\|\leq 1} u^\top A_\lambda^{-1} u$ minimizes $\max_{u:\|u\|\leq 1} \mathbb{E}[(\langle u, \widehat{\theta} - \theta\rangle)^2]$

- **D-optimality**: maximize $g_D(\lambda) = \log(|A_\lambda|)$ maximizes the entropy of distribution. Also, if $\mathcal{E}_\lambda = \{x : x^\top A_\lambda^{-1} x \leq d\}$ then D-optimality is the minimum volume ellipsoid that contains $\mathcal{X}$.

- **G-optimality**: minimize $f_G(\lambda) = \max_{x\in\mathcal{X}} x^\top A_\lambda^{-1} x$ minimizes $\max_{x\in\mathcal{X}} \mathbb{E}[(\langle x, \widehat{\theta} - \theta^*\rangle)^2]$

Input: Finite set $\mathcal{X} \subset \mathbb{R}^d$, confidence level $\delta \in (0,1)$.
Let $\mathcal{X}_1 \leftarrow \mathcal{X}, \ell \leftarrow 1$
while $|\mathcal{X}_\ell| > 1$ **do**
Let $\widehat{\lambda}_\ell \in \triangle_{\mathcal{X}_\ell}$ be a $\frac{d(d+1)}{2}$-sparse minimizer of $f(\lambda) = \max_{x \in \mathcal{X}_\ell} \|x\|^2_{(\sum_{x\in\mathcal{X}_\ell} \lambda_x x x^\top)^{-1}}$
$\epsilon_\ell = 2^{-\ell}$, $\tau_\ell = 2d\epsilon_\ell^{-2} \log(\underline{4\ell^2}|\mathcal{X}|/\delta)$
Pull arm $x \in \mathcal{X}$ exactly $\lceil \widehat{\lambda}_{\ell,x}\tau_\ell \rceil$ times and construct the least squares estimator $\widehat{\theta}_\ell$ using only the observations of this round
$\mathcal{X}_{\ell+1} \leftarrow \mathcal{X}_\ell \setminus \{x \in \mathcal{X}_\ell : \max_{x' \in \mathcal{X}_\ell} \langle x' - x, \widehat{\theta}_\ell \rangle > 2\epsilon_\ell\}$
$\ell \leftarrow \ell + 1$
Output: $\mathcal{X}_\ell$

Lemma 12. *Assume that* $\max_{x\in\mathcal{X}} \langle x^\star - x, \theta^* \rangle \le 4$. *With probability at least* $1-\delta$, *we have* $x^\star \in \mathcal{X}_\ell$ *and* $\max_{x\in\mathcal{X}_\ell} \langle x^\star - x, \theta^* \rangle \le 8\epsilon_\ell$ *for all* $\ell \in \mathbb{N}$.

For any $V \subset \mathcal{X}$ let $\hat{\theta}_\ell(V)$ be the least squares solution at stage ℓ if $\mathcal{X}_\ell = V$.

$$\mathcal{E}_{x,\ell}(V) = \{ |\langle \hat{\theta}_\ell(V) - \theta_*, x \rangle| \le \varepsilon_\ell \}$$

$$\mathbb{P}\Big(\bigcup_{\ell=1}^{\infty} \bigcup_{x\in\mathcal{X}_\ell} \mathcal{E}^c_{x,\ell}(\mathcal{X}_\ell)\Big) \le \sum_{\ell=1}^{\infty} \mathbb{P}\Big(\bigcup_{x\in\mathcal{X}_\ell} \mathcal{E}^c_{x,\ell}(\mathcal{X}_\ell)\Big)$$

$$= \sum_{\ell=1}^{\infty} \sum_{V\subset\mathcal{X}} \mathbb{P}\Big(\bigcup_{x\in V} \mathcal{E}^c_{x,\ell}(V), \mathcal{X}_\ell = V\Big)$$

$$= \sum_{\ell=1}^{\infty} \sum_{V\subset\mathcal{X}} \underbrace{\mathbb{P}\Big(\bigcup_{x\in V} \mathcal{E}_{x,\ell}(V)\Big)}_{\le \frac{\delta}{2\ell^2}} \mathbb{P}(\mathcal{X}_\ell = V)$$

$$\le \sum_{\ell=1}^{\infty} \frac{\delta}{2\ell^2} \cancel{\sum_{V\subset\mathcal{X}} \mathbb{P}(\mathcal{X}_\ell = V)}^{=1} \le \delta$$

$$\text{Regret} = \sum_{x \in \mathcal{X}} \Delta_x T_x \qquad \Delta_x = \langle \theta_*, x_* - x \rangle$$

$$x_* = \arg\max_{x \in \mathcal{X}} \langle x, \theta_* \rangle$$

$$= \sum_{x : \Delta_x \leq \nu} \Delta_x T_x + \sum_{x : \Delta_x \geq \nu} \Delta_x T_x$$

$$\leq \nu T + \sum_{x : \Delta_x > \nu} \sum_{\ell = 1}^{\infty} \overbrace{\Delta_x}^{\leq 8 \varepsilon_\ell} \lceil \lambda_x \zeta_\ell \rceil \mathbb{1}\{x \in \mathcal{X}_\ell\}$$

$$\Delta = \min_{x \neq x_*} \Delta_x$$

$$\leq \nu T + \sum_{x : \Delta_x > \nu} \sum_{\ell = 1}^{\log_2 \lceil 8 (\Delta \vee \nu)^{-1} \rceil} 8 \varepsilon_\ell \left(\lambda_x \zeta_\ell + \mathbb{1}\{x \in \operatorname{supp}(\lambda_\ell)\} \right)$$

$$\zeta_\ell = 2 d \varepsilon_\ell^{-2} \log(4 \ell^2 |\mathcal{X}| / \delta)$$

$$\ell \leq \log_2 \lceil 8 (\Delta \vee \nu)^{-1} \rceil$$

$$\leq \nu T + 8 \cdot \frac{d(d+1)}{2} + 16 d \log\left(4 \log \lceil 8 (\Delta \vee \nu)^{-1} \rceil^2 |\mathcal{X}| / \delta\right) \sum_{\ell \geq 1}^{\log_2(\cdot)} \varepsilon_\ell^{-1}$$

$$\leq \nu T + 4 d(d+1) + 128 d \log\left(4 \log \lceil 8 (\Delta \vee \nu)^{-1} \rceil |\mathcal{X}| / \delta\right) \cdot (\Delta \vee \nu)^{-1}$$

If $\nu = 0$ $\quad \text{Regret} \leq O\left(d^2 + \frac{d}{\Delta} \log\left(\frac{|\mathcal{X}|}{\delta} \log \Delta^{-1}\right)\right)$

If opt ν $\quad \text{Regret} \leq O\left(\sqrt{d T \log\left(\log(T/d) |\mathcal{X}| / \delta\right)}\right)$

What if $|\mathcal{X}| = \infty$

$$A = X^T X \qquad X = \begin{pmatrix} - x_1^T - \\ \vdots \\ - x_T^T - \end{pmatrix}$$

If chosen to be G-optimal then

$$\max_{x \in \mathcal{X}} |\langle \hat{\theta} - \theta_*, x \rangle| = \max_x |\langle A^{1/2} (\hat{\theta} - \theta_*), A^{-1/2} x \rangle|$$

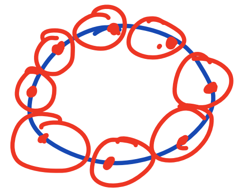
Cauchy-Schwartz $x, y \in \mathbb{R}^d$

$\langle x, y \rangle \leq \|x\|_2 \cdot \|y\|_2$

$$\leq \max_x \|A^{1/2} (\hat{\theta} - \theta_*)\|_2 \cdot \|A^{-1/2} x\|_2$$

$$= \underbrace{\|\hat{\theta} - \theta_*\|_A}_{\lesssim \sqrt{d \log(1/\delta)}} \cdot \underbrace{\max_{x \in \mathcal{X}} \|x\|_{A^{-1}}}_{\leq \sqrt{d/T}} \leq \sqrt{\frac{d^2 \log(1/\delta)}{T}}$$

$$\|\hat{\theta}-\theta_*\|_A = \|A^{1/2}(\hat{\theta}-\theta_*)\|_2 = \sup_{u:\|u\|_2=1} \langle u, A^{1/2}(\hat{\theta}-\theta_*)\rangle$$



Lemma $\exists$ cover $C \subset \mathbb{R}^d$: $|C| \le (3/\varepsilon)^d$ such that for all x w/ $\|x\|_2 = 1$ $\exists$ $y \in C$ s.t. $\|x-y\|_2 \le \varepsilon$.

$$\begin{aligned}
\|\hat{\theta}-\theta_*\|_A &= \sup_u \langle u, A^{1/2}(\hat{\theta}-\theta_*)\rangle \\
&= \sup_u \min_{y\in C} \langle u-y, A^{1/2}(\hat{\theta}-\theta_*)\rangle + \langle y, A^{1/2}(\hat{\theta}-\theta_*)\rangle \\
&\le \sup_u \underbrace{\min_{y\in C} \|u-y\|_2}_{\le \varepsilon} \cdot \underbrace{\|A^{1/2}(\hat{\theta}-\theta_*)\|_2}_{\|\hat{\theta}-\theta_*\|_A} + \underbrace{\langle y, A^{1/2}(\hat{\theta}-\theta_*)\rangle}_{\le \sqrt{2d\log(3/\varepsilon\delta)}} \\
&\le \varepsilon \|\hat{\theta}-\theta_*\|_A + \sqrt{2d\log(3/\varepsilon\delta)}
\end{aligned}$$

$$\Rightarrow \quad \|\hat{\theta}-\theta_*\|_A \le \frac{1}{1-\varepsilon} \cdot \sqrt{2d\log(3/\varepsilon\delta)}.$$

$$\mathcal{E} = \{\theta : \|\hat{\theta}-\theta\|_A \le \beta\}$$

For a fixed vector $z = A^{1/2}y$ we have

$$\langle z, \hat{\theta}-\theta_*\rangle \le \|z\|_{A^{-1}} \cdot \sqrt{2d\log(1/\delta)} \qquad \text{w.p.} \ge 1-\delta.$$

$$\begin{aligned}
\Rightarrow \quad \langle y, A^{1/2}(\hat{\theta}-\theta_*)\rangle &= \langle A^{1/2}y, \hat{\theta}-\theta_*\rangle \\
&\le \|A^{1/2}y\|_{A^{-1}} \cdot \sqrt{2\log(1/\delta)} \\
&= \|y\|_2 \cdot \sqrt{2\log(1/\delta)} \\
&= \sqrt{2\log(1/\delta)}
\end{aligned}$$

$$\Rightarrow \text{ w.p. } 1-\delta \qquad \max_{y\in C} \langle y, A^{1/2}(\hat{\theta}-\theta_*)\rangle \le \sqrt{2\log(|C|/\delta)}$$

$\leq \sqrt{2d \log(3/\varepsilon\delta)}$

UCB for linear bandits - "LinUCB"

Let $\mathcal{E}_t \subset \mathbb{R}^d$: $\theta_* \in \mathcal{E}_t$ $\forall t$ w.p. $\geq 1-\delta$

and define $UCB_t(x) = \underset{\theta \in \mathcal{E}_t}{\operatorname{argmax}} \langle x, \theta \rangle$

$$x_t = \underset{x \in \mathcal{X}}{\operatorname{argmax}} \; UCB_t(x)$$

$$\hat{\theta}_t = \underset{\theta}{\operatorname{argmin}} \sum_{s=1}^{t} (y_s - \langle \theta, x_s \rangle)^2 + \lambda \|\theta\|_2^2$$

$$= (X^T X + \lambda I)^{-1} X^T y$$

$$= V_t^{-1} S_t$$

$$V_0 = \lambda I$$

$$V_t = \sum_{s=1}^{t} x_s x_s^T + V_0$$

$$S_t = \sum_{s=1}^{t} x_s y_s$$

$\mathcal{E}_t = \{\theta : \|\theta - \hat{\theta}_{t-1}\|_{V_t}^2 \leq \beta_t\}$ for some increasing seq β_t.

Thm) If $\theta_* \in \mathcal{E}_t$ $\forall t$ w.p. $\geq 1-\delta$, then w.p. $\geq 1-\delta$

$$R_T = \sum_{t=1}^{T} \langle \theta_*, x_* - x_t \rangle \leq \sqrt{8T\beta_T \log\left(\frac{|V_T|}{|V_0|}\right)}.$$

$$\langle \theta_*, x_* \rangle \le UCB_t(x_*) \le UCB_t(x_t) =: \langle \tilde{\theta}_t, x_t \rangle$$

where $\tilde{\theta}_t = \underset{\theta \in \mathcal{E}_t}{\operatorname{argmax}} \langle \theta, x_t \rangle$

$$\begin{aligned}
\langle \theta_*, x_* - x_t \rangle &\le \langle \theta_*, x_* \rangle - \langle \theta_*, x_t \rangle \\
&\le \langle \tilde{\theta}_t, x_t \rangle - \langle \theta_*, x_t \rangle \\
&= \langle \tilde{\theta}_t - \theta_*, x_t \rangle \\
&= \langle V_t^{1/2}(\tilde{\theta} - \theta_*), V_t^{-1/2} x_t \rangle \\
&\le \|\tilde{\theta}_t - \theta_*\|_{V_t} \cdot \|x_t\|_{V_t^{-1}} \\
&\le \left(\|\tilde{\theta}_t - \hat{\theta}_{t-1}\|_{V_t} + \|\hat{\theta}_{t-1} - \theta_*\|_{V_t} \right) \cdot \|x_t\|_{V_t^{-1}} \\
&\le 2\sqrt{\beta_t} \cdot \|x_t\|_{V_t^{-1}}
\end{aligned}$$

Assume $|\langle \theta_*, x_* - x_t \rangle| \le 2$, $\beta_t \ge 1$

$$\langle \theta_*, x_* - x_t \rangle \le 2 \wedge 2\sqrt{\beta_t} \|x_t\|_{V_t^{-1}} \le 2\sqrt{\beta_t} \left(1 \wedge \|x_t\|_{V_t^{-1}}\right)$$

$$\begin{aligned}
R_T = \sum_{t=1}^{T} \langle \theta_*, x_* - x_t \rangle &\le \sqrt{T \cdot \sum_{t=1}^{T} \langle \theta_*, x_* - x_t \rangle^2} \\
&\le 2\sqrt{T \cdot \beta_T \cdot \sum_{t=1}^{T} \left(1 \wedge \|x_t\|^2_{V_t^{-1}}\right)}
\end{aligned}$$

Elliptic Potential Lemma

Fix V_0 and let $x_t \in \mathbb{R}^d$ be an arbitrary sequence and set $V_t = V_0 + \sum_{s=1}^{t} x_s x_s^T$ then for any $V_0 \succ 0$

$$\sum_{t=1}^{T} \left(1 \wedge \|x_t\|^2_{V_t^{-1}}\right) \leq 2 \log\left(\frac{|V_T|}{|V_0|}\right).$$

Moreover, if $\|x_t\|_2 \leq L \ \forall t$, $\leq 2d \log\left(\frac{Tr(V_0) + TL^2}{d|V_0|^{1/d}}\right)$

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

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

$|ABC| = |A||B||C|$

$|AB| = |A| \cdot |B|$

$$|V_t| = |V_{t-1}| \cdot \left|I + V_{t-1}^{-1/2} x_t x_t^T V_{t-1}^{-1/2}\right|$$

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

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

$$|I + \alpha e_1 e_1^T| = 1^{d-1} \cdot (1+\alpha) = 1 + \alpha$$

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

$$|V_T| = \prod_{i=1}^{d} \lambda_i \leq \left(\frac{1}{d}\sum_{i=1}^{d} \lambda_i\right)^d = \left(\frac{1}{d} Tr(V_T)\right)^d$$

$\leftarrow$ AMGM.