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

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 $t = 1, 2, \dots, T$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$(z z^T)_{i,j} = z_i z_j$

$(z z^T)_{i,i} = z_i^2$

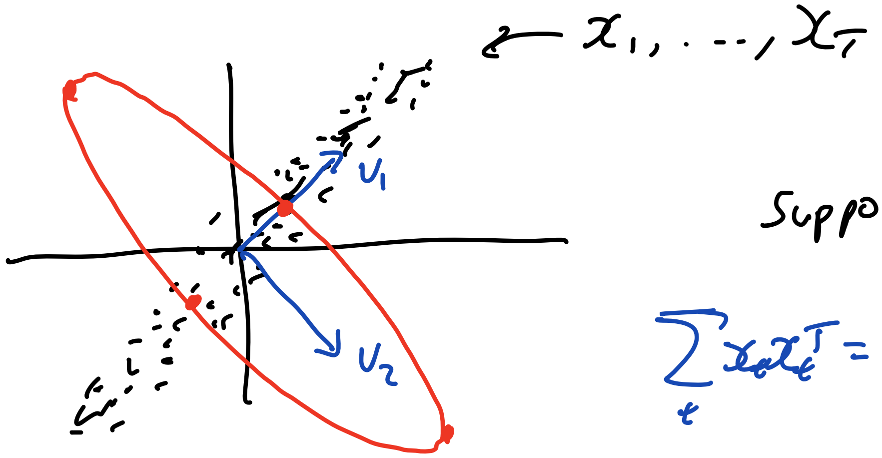
For any z, what is $z^T(\hat{\theta} - \theta_*)$

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

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

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

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



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 $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $v_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$\sigma_1 \gg \sigma_2$

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

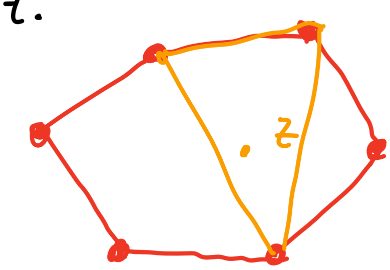
$$z^T \left(v_1 v_1^T \sigma_1 + v_2 v_2^T \sigma_2 \right)^{-1} z$$

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

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

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/ $\|\lambda\|_0 \le d+1$.

Claim 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 $x_1, \dots, x_T$ s.t. when we construct $\hat{\theta}$ we have

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

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 $\mathfrak{Z} \in \mathbb{N}$

measure $x \in \mathcal{X}$ exactly $\lceil \lambda_x \mathfrak{Z} \rceil$ times.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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