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 $\mathbb{P}(\bigcup_{x \in \mathcal{X}} \{ |\langle x, \widehat{\theta} - \theta^* \rangle| > \underbrace{\sqrt{\frac{2d \log(2|\mathcal{X}|/\delta')}{\tau}}}_{=\varepsilon} \}) \le \delta'$. $\Longrightarrow$ $\delta' = \exp\left(-\frac{\varepsilon^2 \tau}{2d}\right) \cdot 2|\mathcal{X}|$

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 $f(\lambda) = \max_{x \in \mathcal{X}_\ell} \|x\|^2_{(\sum_{x \in \mathcal{X}_\ell} \lambda_x x x^\top)^{-1}}$
$\epsilon_\ell = 2^{-\ell}$, $\tau_\ell = 2d\epsilon_\ell^{-2} \log(4\ell^2 |\mathcal{X}|/\delta)$
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: $\mathcal{X}_\ell$

Good events: $\mathcal{E}_{x,\ell} = \{ |\langle x, \hat{\theta}_\ell - \theta_* \rangle| \le \varepsilon_\ell \}$, Want $\bigcap_\ell \bigcap_{x \in \mathcal{X}_\ell} \mathcal{E}_{x,\ell}$ to hold.

!!

$\mathcal{E}_{x,\ell}(\mathcal{X}_\ell)$

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

$$= \sum_{\ell=1}^{\infty} \sum_{V \subseteq \mathcal{X}} \underbrace{\mathbb{P}\left(\bigcup_{x \in V} \mathcal{E}_{x,\ell}(V)^c \,\middle|\, \mathcal{X}_\ell = V \right)} \mathbb{P}(\mathcal{X}_\ell = V)$$

$\le 2|V| \exp\left(\frac{-\varepsilon_\ell^2 \tau_\ell}{2d}\right)$ by proposition above.

$\le 2|V| \cdot \frac{\delta}{4\ell^2 |\mathcal{X}|} \le \frac{\delta}{2\ell^2}$ Plug in τ_ℓ from alg.

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

$$= \sum_{\ell=1}^{\infty} \frac{\delta}{2\ell^2} \le \delta.$$

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

Lemma: W.p. $\geq 1-\delta$ we have $x_* \in \mathcal{X}_\ell$ for all $\ell \geq 1$

and $\max_{x \in \mathcal{X}_\ell} \langle x_* - x, \theta_* \rangle \leq 8\varepsilon_\ell$.

Proof: Assume $x_* \in \mathcal{X}_\ell$ and show $x_* \in \mathcal{X}_{\ell+1}$.

$x_* \notin \mathcal{X}_{\ell+1}$ if $\max_{x' \in \mathcal{X}_\ell} \langle x' - x_*, \hat{\theta}_\ell \rangle > \varepsilon_\ell$.

For $\underline{\text{any}}$ $x' \in \mathcal{X}$,

$$\begin{aligned}
\langle x' - x_*, \hat{\theta}_\ell \rangle &= \langle x' - x_*, \hat{\theta}_\ell - \theta_* \rangle + \langle x' - x_*, \theta_* \rangle \\
&= \langle x', \hat{\theta}_\ell - \theta_* \rangle - \langle x_*, \hat{\theta}_\ell - \theta_* \rangle + \langle x' - x_*, \theta_* \rangle \\
&\leq 2\varepsilon_\ell + \langle x' - x_*, \theta_* \rangle \\
&\leq 2\varepsilon_\ell \implies x_* \in \mathcal{X}_{\ell+1}
\end{aligned}$$

Now fix any $x \in \mathcal{X}_\ell$ w/ $\underline{\langle x_* - x, \theta_* \rangle > 4\varepsilon_\ell}$. Then

$$\begin{aligned}
\max_{x'} \langle x' - x, \hat{\theta}_\ell \rangle &\geq \langle x_* - x, \hat{\theta}_\ell \rangle \\
&= \langle x_*, \hat{\theta} - \theta_* \rangle - \langle x, \hat{\theta} - \theta_* \rangle + \langle x_* - x, \theta_* \rangle \\
&\geq -2\varepsilon_\ell + \langle x_* - x, \theta_* \rangle \\
&> -2\varepsilon_\ell + 4\varepsilon_\ell \\
&> 2\varepsilon_\ell. \implies x \notin \mathcal{X}_{\ell+1}
\end{aligned}$$

$4\varepsilon_\ell = 8\varepsilon_{\ell+1} \implies \max_{x \in \mathcal{X}_\ell} \langle x_* - x, \theta_* \rangle \leq 8\varepsilon_\ell$ ☑

Regret: $\mathbb{E}\left[\sum_{t=1}^{T} \langle x_*, \theta_* \rangle - \langle x_t, \theta_* \rangle\right]$

On $\bigcap_{\ell} \bigcap_{x \in \mathcal{X}_\ell} \mathcal{E}_{x,\ell}$ (which hold w.p. $\geq 1-\delta$) we have

$$\sum_{t=1}^{T} \langle x_* - x_t, \theta_* \rangle = \sum_{\ell=1}^{\infty} \sum_{x \in \mathcal{X}_\ell} T_x \cdot \langle x_* - x, \theta_* \rangle$$

$$\leq \sum_{\ell=1}^{\infty} \sum_{x \in \mathcal{X}_\ell} \lceil \lambda_{x,\ell} \zeta_\ell \rceil \cdot 8 \varepsilon_\ell$$

$\Delta_{min} = \min_{x \neq x_*} \langle x_* - x, \theta_* \rangle$

$$\leq \sum_{\ell=1}^{\log_2 (8\Delta_{min})^{-1}} d^2 8 \varepsilon_\ell + \sum_{x \in \mathcal{X}_\ell} \lambda_{x,\ell} \zeta_\ell \, 8 \varepsilon_\ell$$

$$= \sum_{\ell}^{\log_2 (8\Delta_{min})^{-1}} d^2 8 \varepsilon_\ell + 16 d \varepsilon_\ell^{-2} \log(4\ell^2 |\mathcal{X}|/\delta) \cdot \varepsilon_\ell$$

$\sum_{\ell=1}^{L} \varepsilon_\ell^{-1} = \sum_{\ell=1}^{L} 2^\ell \approx 2^{L+1}$

$$\leq c d^2 + \sum_{\ell=1}^{\log_2(8\Delta_{min}^{-1})} 16 d \varepsilon_\ell^{-1} \log(4\ell^2 |\mathcal{X}|/\delta)$$

$$\leq c d^2 + c' d \log\left(\frac{|\mathcal{X}| \log(\Delta_{min}^{-1})}{\delta}\right) \Delta_{min}^{-1}$$

Using same analysis as in multi-armed bandit:

we also have $\sum_{t} \langle x_* - x_t, \theta_* \rangle \leq c d^2 + \sqrt{c' d T \log(|\mathcal{X}| \log(T)/\delta)}$

$$\text{Regret} = \mathbb{E}\left[\sum_{t} \langle x_* - x_t, \theta_* \rangle\right] \leq d^2 + \min\left\{ \frac{d \log(|\mathcal{X}| \cdot T)}{\Delta_{min}}, \sqrt{d T \log(T \cdot |\mathcal{X}|)} \right\}$$

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

The above analysis treated as a R.V. for each $x \in \mathcal{X}$

Note $\langle x, \hat\theta_\ell - \theta_* \rangle = x^T (X^T X)^{-1/2} (X^T X)^{1/2} (\hat\theta_\ell - \theta_*)$

$$\le \sqrt{x^T (X^T X)^{-1} x} \cdot \sqrt{(\hat\theta_\ell - \theta_*)^T (X^T X) (\hat\theta - \theta_*)}$$

Cauchy Schwartz $\langle x, y \rangle \le \|x\|_2 \cdot \|y\|_2$

$$= \|x\|_{(X^T X)^{-1}} \cdot \|\hat\theta_\ell - \theta_*\|_{X^T X}$$

$$\le \sqrt{\frac{d}{\tau_\ell}} \cdot \|\hat\theta - \theta_*\|_{X^T X}$$

$$\|\hat\theta_\ell - \theta\|_{A(\lambda_\ell)} = \|(X^T X)^{1/2} (\hat\theta_\ell - \theta_*)\|_2$$

$$\overset{\text{dist.}}{=} \|(X^T X)^{1/2} (X^T X)^{-1/2} z\|_2 \qquad z \sim \mathcal{N}(0, I_d)$$

$$= \|z\|_2$$

$$\simeq \sqrt{d}.$$

$$\Longrightarrow \langle x, \hat\theta_\ell - \theta_* \rangle \lesssim \sqrt{\frac{d^2}{\tau_\ell}} \le \varepsilon_\ell \quad \text{if} \quad \tau_\ell \ge d^2 \varepsilon_\ell^{-2}$$

One can show w/ this τ_ℓ that $\mathbb{E}\left[\sum_t \langle x_* - x_t, \theta_* \rangle\right] \le d\sqrt{T \log(T)}$.

Proof Sketch

$$\sum_t \langle x_* - x_t, \theta_* \rangle \leq \min_L \varepsilon_L T + \sum_{\ell=1}^{L} g_\ell \varepsilon_\ell$$

$$\lesssim \min_L 2^{-L} T + \sum_{\ell=1}^{L} 2^{\ell} d \log(|\mathcal{X}|/\delta)$$

$$\lesssim \min_L 2^{-L} T + 2^{L} d \log(|\mathcal{X}|/\delta)$$

$$\lesssim \min_\alpha \alpha^{-1} T + \alpha d \log(|\mathcal{X}|/\delta)$$

$$\lesssim \sqrt{d T \log(|\mathcal{X}|/\delta)}$$