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*

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

- $\max_\lambda g_D(\lambda) = g_D(\lambda^*)$ $\quad g_D(\lambda) = \log|A(\lambda)|$
- $\min_\lambda f_G(\lambda) = f_G(\lambda^*)$ $\quad f_G(\lambda) = \max_{x\in\mathcal{X}} \|x\|_{A(\lambda)^{-1}}^2$
- $f_G(\lambda^*) = d$
- $|support(\lambda^*)| = (d+1)d/2$

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 $1-\delta$*

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

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

Frank-Wolfe

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

Init $x_0 \in K$

for $t = 1, 2, \ldots$

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

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

For D-optimal design

Init λ_0 sparsely.

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

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

- **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* $\max_{x \in \mathcal{X}} \langle x^\star - x, \theta^* \rangle \leq 4$. *With probability at least* $1-\delta$, *we have* $x^\star \in \mathcal{X}_\ell$ *and* $\max_{x \in \mathcal{X}_\ell} \langle x^\star - x, \theta^* \rangle \leq 8\epsilon_\ell$ *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 $\mathcal{X}_\ell = V$.

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

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

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

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

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

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

$$x_* = \operatorname*{argmax}_{x \in \mathcal{X}} \langle x, \theta_* \rangle$$

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

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

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

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

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

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

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

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

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