# APPROXIMATE SUPERMODULARITY BOUNDS FOR EXPERIMENTAL DESIGN

Luiz F. O. Chamon and Alejandro Ribeiro

e-mail: `luizf@seas.upenn.edu`, Electrical & Systems Engineering, University of Pennsylvania

Penn
UNIVERSITY of PENNSYLVANIA

## CONTRIBUTIONS

i. The greedy solution of the **A-optimal design problem** is $(1-e^{-\alpha})$-optimal with $\alpha \geq [1+\mathcal{O}(\text{SNR})]^{-1}$.

ii. The value of the greedy solution of an **E-optimal design problem** is at most $(1-e^{-1})(f(\mathcal{D}^\star)+k\epsilon)$, where $\epsilon \leq \mathcal{O}(\text{SNR})$.

iii. As the SNR of the experiments decreases, the performance guarantees for greedy A- and E-optimal designs approach $1-1/e$.

## INTRODUCTION

- Experimental design = select which experiments to run or measurements to observe to estimate a variable of interest
- Applications: designing experiments, semi-supervised learning, multivariate analysis, sketching, sensor placement...
- NP-hard in general, P approximations for *supermodular* objectives
- estimation MSE (A-optimality) and spectral norm of error covariance matrix (E-optimality) are not *supermodular*

## EXPERIMENTAL DESIGN

- Pool of experiments $\mathcal{E}$

$$y_e = \boldsymbol{A}_e^T \boldsymbol{\theta} + \boldsymbol{v}_e$$

$$\mathbb{E}\boldsymbol{\theta} = \bar{\boldsymbol{\theta}} \qquad \mathbb{E}(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})^T = \boldsymbol{R}_\theta \qquad \boldsymbol{v}_e \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_e)$$

- Design $\mathcal{D} \in \mathcal{P}(\mathcal{E})$ (multiset)
- Use the experiments in $\mathcal{D}$ to estimate

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{\theta}$$

**Proposition (Bayesian estimator)**

Given a design $\mathcal{D} \in \mathcal{P}(\mathcal{E})$, the unbiased affine estimator of $\boldsymbol{z}$ with the smallest error covariance matrix in the PSD cone is given by

$$\hat{\boldsymbol{z}}_{\mathcal{D}} = \mathbb{E}[\boldsymbol{z} \mid \boldsymbol{\theta}, \boldsymbol{v}_e] = [\text{long uninformative expression}]$$

with error covariance matrix

$$\boldsymbol{K}(\mathcal{D}) = \boldsymbol{H}\left[\boldsymbol{R}_\theta^{-1} + \sum_{e\in\mathcal{D}} \boldsymbol{A}_e^T \boldsymbol{R}_e^{-1} \boldsymbol{A}_e\right]^{-1} \boldsymbol{H}^T.$$

## OPTIMAL EXPERIMENTAL DESIGN

- A-optimal (NOT supermodular)

$$\underset{|\mathcal{D}|\leq k}{\text{minimize}} \quad \operatorname{Tr}\left[\boldsymbol{K}(\mathcal{D})\right] - C_A$$

- E-optimal (NOT supermodular)

$$\underset{|\mathcal{D}|\leq k}{\text{minimize}} \quad \lambda_{\max}\left[\boldsymbol{K}(\mathcal{D})\right] - C_E$$

- D-optimal (supermodular)

$$\underset{|\mathcal{D}|\leq k}{\text{minimize}} \quad \log\det\left[\boldsymbol{K}(\mathcal{D})\right] - C_D$$

## GREEDY EXPERIMENTAL DESIGN

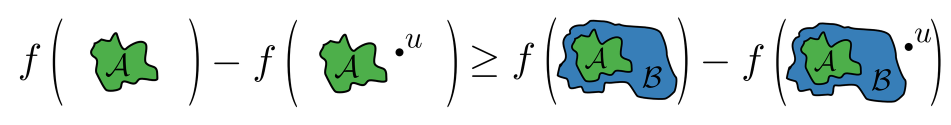
```
function GREEDY(ℓ)
  G_0 = {}
  for j = 1, ..., ℓ
    u = argmin_{e∈E} f(G_{j-1} ∪ {e})
    G_j = G_{j-1} ∪ {u}
  end
end
```

- Low complexity
- Sequential
- $(1-1/e)$-optimal for supermodular objectives

## SUPERMODULARITY

For $\mathcal{A}, \mathcal{B} \in \mathcal{P}(\mathcal{E})$, $\mathcal{A} \subseteq \mathcal{B}$,

$$f(\mathcal{A}) - f(\mathcal{A}\cup\{u\}) \geq f(\mathcal{B}) - f(\mathcal{B}\cup\{u\})$$




## $\alpha$-SUPERMODULARITY

For $\mathcal{A}, \mathcal{B} \in \mathcal{P}(\mathcal{E})$, $\mathcal{A} \subseteq \mathcal{B}$, and $\alpha \in [0,1]$

$$f(\mathcal{A}) - f(\mathcal{A}\cup\{u\}) \geq \alpha(\#\mathcal{A}, \#\mathcal{B})\left[f(\mathcal{B}) - f(\mathcal{B}\cup\{u\})\right]$$

**Theorem (Greedy approximately supermodular minimization)**

Let $f$ be a normalized, monotone decreasing, $\alpha$-supermodular multiset function. Then, for $\bar{\alpha} = \min_{a<\ell,\ b<\ell+k} \alpha(a,b)$,

$$f(\mathcal{G}_\ell) \leq \left[1 - \prod_{h=0}^{\ell-1}\left(1 - \frac{1}{\sum_{s=0}^{k-1}\alpha(h,h+s)^{-1}}\right)\right] f(\mathcal{D}^\star)$$

$$\leq (1-e^{-\bar{\alpha}\ell/k}) f(\mathcal{D}^\star)$$

- If $\alpha \equiv 1$, then $f$ is supermodular [$\ell = k \Rightarrow (1-1/e)$-optimality]
- If $\alpha < 1$, then $f$ is *approximately* supermodular [$\ell = \alpha^{-1}k \Rightarrow (1-1/e)$-optimality]

## $\epsilon$-SUPERMODULARITY

For $\mathcal{A}, \mathcal{B} \in \mathcal{P}(\mathcal{E})$, $\mathcal{A} \subseteq \mathcal{B}$, and $\epsilon \geq 0$

$$f(\mathcal{A}) - f(\mathcal{A}\cup\{u\}) \geq f(\mathcal{B}) - f(\mathcal{B}\cup\{u\}) - \epsilon(\#\mathcal{A}, \#\mathcal{B})$$

**Theorem (Greedy approximately supermodular minimization)**

Let $f$ be a normalized, monotone decreasing, $\epsilon$-supermodular multiset function. Then, for $\bar{\epsilon} = \max_{a<\ell,\ b<\ell+k} \epsilon(a,b)$,

$$f(\mathcal{G}_\ell) \leq \left[1 - \left(1 - \frac{1}{k}\right)^\ell\right] f(\mathcal{D}^\star) + \frac{1}{k}\sum_{s=0}^{k-1}\sum_{h=0}^{\ell-1} \epsilon(h,h+s)\left(1-\frac{1}{k}\right)^{\ell-1-h}$$

$$\leq (1-e^{-\ell/k})(f(\mathcal{D}^\star) + k\bar{\epsilon})$$

- If $\epsilon \equiv 0$, then $f$ is supermodular [$\ell = k \Rightarrow (1-1/e)$-optimality]
- If $\epsilon > 0$, then $f$ is approximately supermodular [$\epsilon < f(\mathcal{D}^\star)/3k,\ \ell = 3k \Rightarrow (1-1/e)$-optimality]

## NEAR-A-OPTIMAL DESIGN

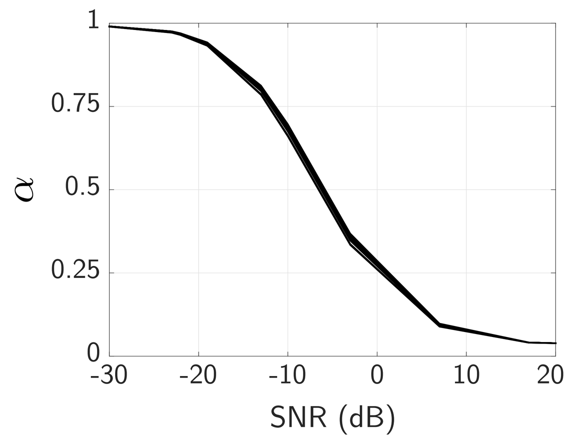
**Theorem (A-optimality is $\alpha$-supermodular)**

The objective of A-optimal design is $\alpha$-supermodular with

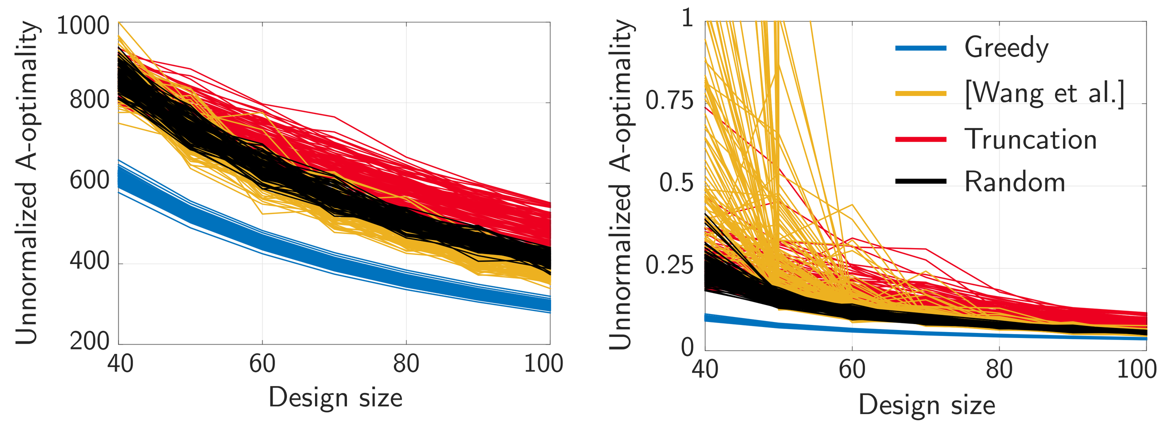
$$\alpha(a,b) \geq \frac{1}{\kappa(\boldsymbol{H})^2} \frac{\lambda_{\min}\left[\boldsymbol{R}_\theta^{-1}\right]}{\lambda_{\max}\left[\boldsymbol{R}_\theta^{-1}\right] + a\cdot\ell_{\max}}, \quad \text{for all } b,$$

where $\ell_{\max} = \max_{e\in\mathcal{E}} \lambda_{\max}(\boldsymbol{A}_e^T \boldsymbol{R}_e^{-1} \boldsymbol{A}_e)$. For $\boldsymbol{R}_\theta = \sigma_\theta^2 \boldsymbol{I}$, $\boldsymbol{H} = \boldsymbol{I}$, $\gamma = \max_{e\in\mathcal{E}} \operatorname{Tr}[\boldsymbol{A}_e^T \boldsymbol{R}_e^{-1} \boldsymbol{A}_e]$, and $\ell = k$,

$$\bar{\alpha} \geq \frac{1}{1+2k\sigma_\theta^2\gamma}.$$




- $\alpha \to 1$ as $\gamma \to 0$: effectively supermodular for low SNRs
- $\alpha \to 0$ as $\gamma \to \infty$ (high SNR)
- Low SNR is the regime of interest: in high SNR, even random designs perform well




(a) Low SNR (b) High SNR

## NEAR-E-OPTIMAL DESIGN

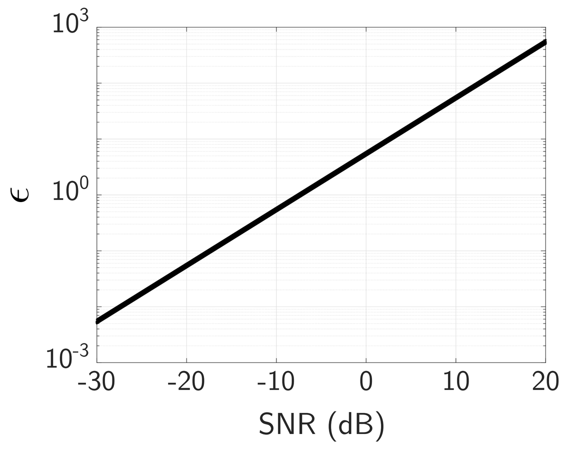
**Theorem (E-optimality is $\epsilon$-supermodular)**

The objective of E-optimal design is $\epsilon$-supermodular with

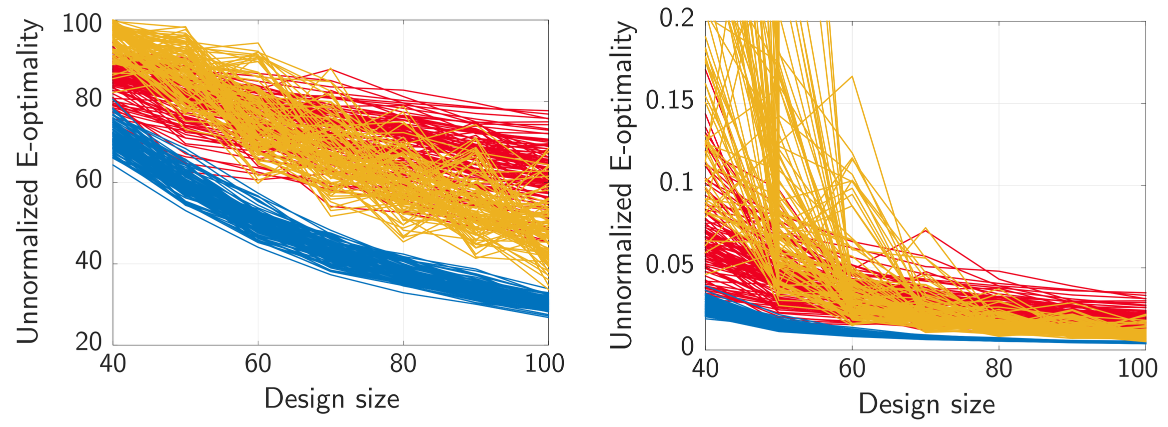
$$\epsilon(a,b) \leq (b-a)\,\sigma_{\max}(\boldsymbol{H})^2\,\lambda_{\max}(\boldsymbol{R}_\theta)^2\,\ell_{\max},$$

where $\ell_{\max} = \max_{e\in\mathcal{E}} \lambda_{\max}(\boldsymbol{A}_e^T \boldsymbol{R}_e^{-1} \boldsymbol{A}_e)$. For $\boldsymbol{R}_\theta = \sigma_\theta^2 \boldsymbol{I}$, $\boldsymbol{H} = \boldsymbol{I}$, $\gamma = \max_{e\in\mathcal{E}} \operatorname{Tr}[\boldsymbol{A}_e^T \boldsymbol{R}_e^{-1} \boldsymbol{A}_e]$, and $\ell = k$,

$$\bar{\epsilon} \leq 2k\sigma_\theta^4\gamma.$$




- $\epsilon \to 0$ as $\gamma \to 0$: effectively supermodular for low SNRs
- $\epsilon \to \infty$ as $\gamma \to \infty$ (high SNR)
- Better guarantees for smaller designs

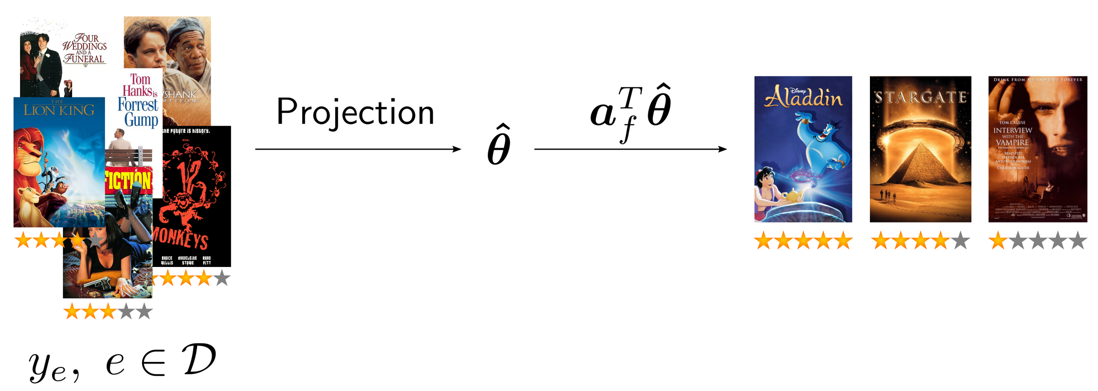


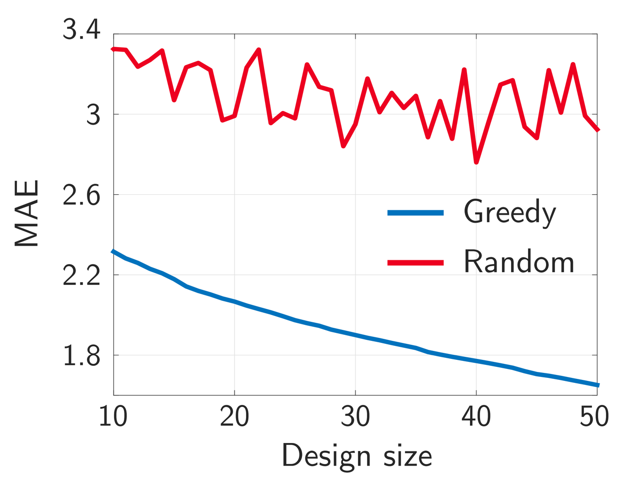

(a) Low SNR (b) High SNR

## NUMERICAL EXAMPLE

### Cold-start problem

- How to give recommendations when you don't know what people like? *Cold-start survey*
- $\boldsymbol{a}_e^T$: ratings of movie $e \in \mathcal{E}$ by users in the training set
- New user's rating for movie $e \in \mathcal{E}$ (experiments): $y_e = \boldsymbol{a}_e^T \boldsymbol{\theta} + v_e$




- EachMovie dataset (sample of 12.000 users)
- 75/25 split
- Favorite genre prediction: 25% misclassification with 100 movies

## RELATED WORK

### Optimal experimental design

- Convex relaxation (SDPs or sequential SOCPs) [Flaherty'06, Joshi'09, Sagnol'11, Horel'14, Wang'17]
- D-optimal design: $(1-1/e)$ guarantee using pipage rounding [Ageev'04, Horel'14]
- A-optimal design: near-optimal randomized schemes for large $k$ [Wang'17]

### Greedy non-submodular optimization

- $\epsilon$-supermodularity with constant $\epsilon$ [Krause'10]
- $\alpha$-supermodularity with constant $\alpha$ [Chamon'16]
- submodularity ratio ($\gamma$) bounds using RIP [Das'11], RSC [Elenberg'16], and spectral inequalities [Bian'17]
- [Chamon'16, Bian'17] do not account for multisets: $2.5\times10^{-6}$-optimality vs $0.1$-optimality (A-optimality, SNR = 0 dB)
- more stringent approximate submodularity ("$\delta$-submodularity"): function must be upper and lower bounded by a submodular function [Horel'16]
- *approximate submodularity* is sometimes called *weak submodularity* (e.g., [Elenberg'16]), not to be confused with weak submodularity in [Borodin'14]

## CONCLUSION

Greedy A- and E-optimal experimental design is guaranteed to work well despite the fact that their objectives are not supermodular.