Georgia Institute of Technology, Atlanta GA, USA

COMPUTATIONAL ASPECTS OF SELECTION OF EXPERIMENTS

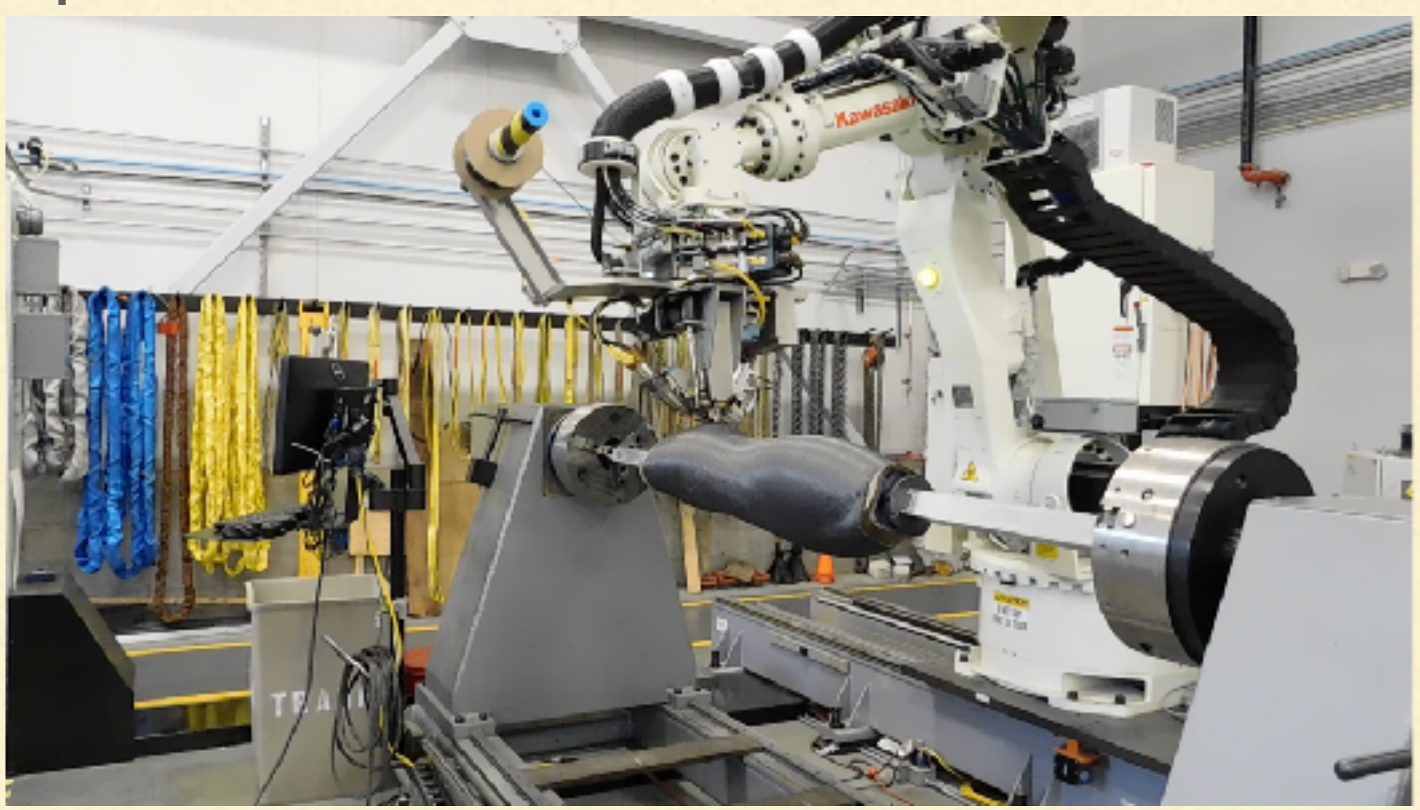
Yining Wang Machine Learning Department, Carnegie Mellon University

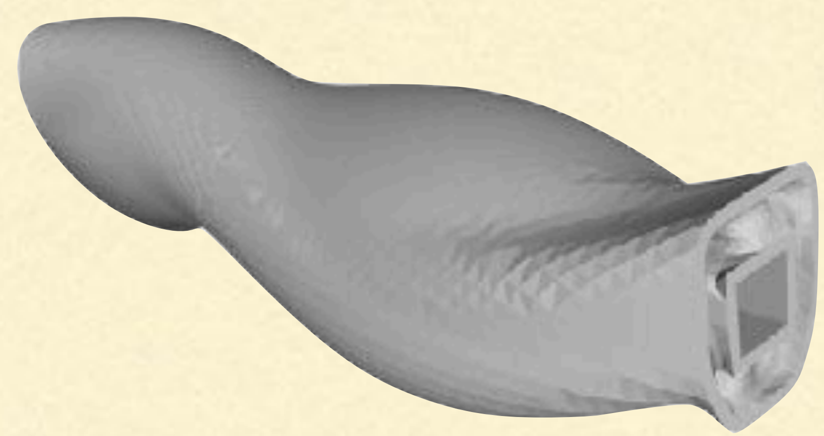
Arxiv:1711.05174

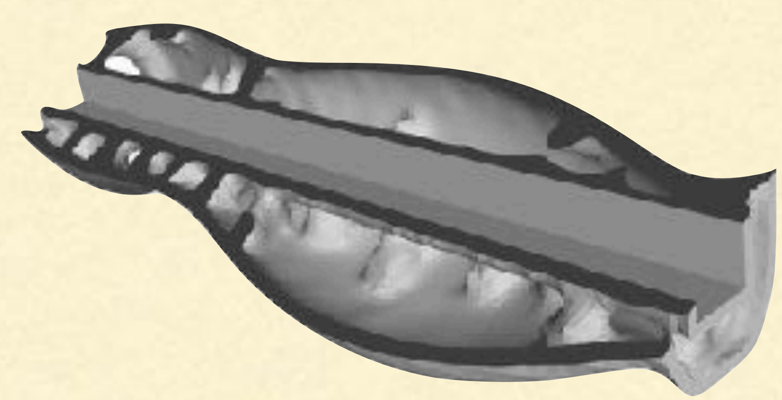
Joint work with Zeyuan Allen-Zhu, Yuanzhi Li *and* Aarti Singh

MOTIVATING APPLICATION

- Worst-case structural analysis
  - **Maximum** stress resulting from **worst-case** external forces
  - Example application: lightweight structural design in automated fiber process



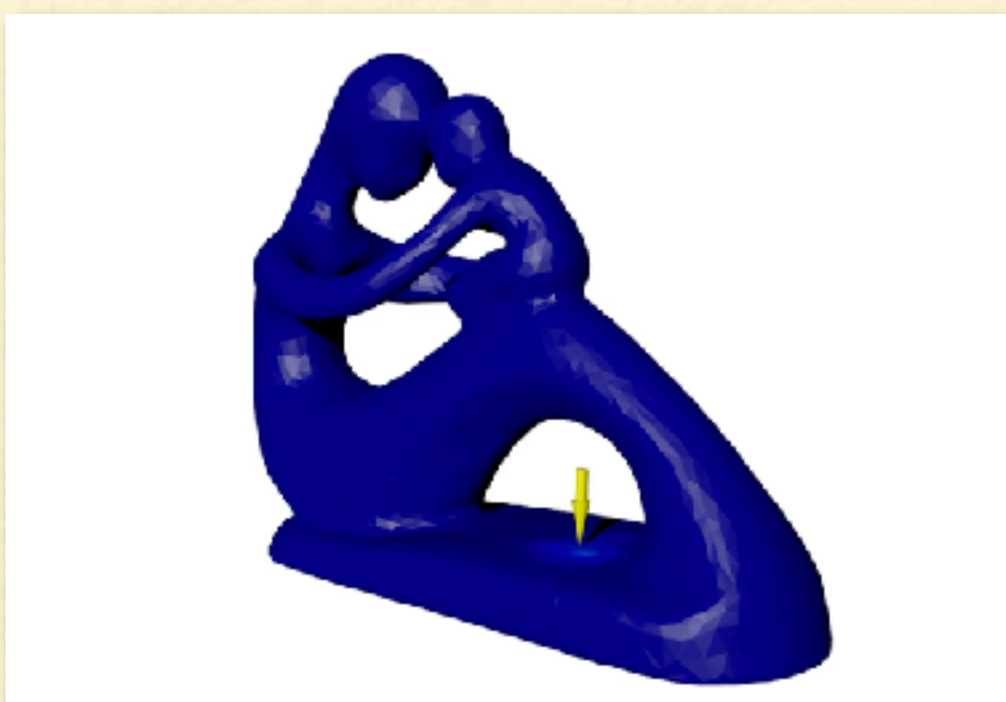




MOTIVATING APPLICATION

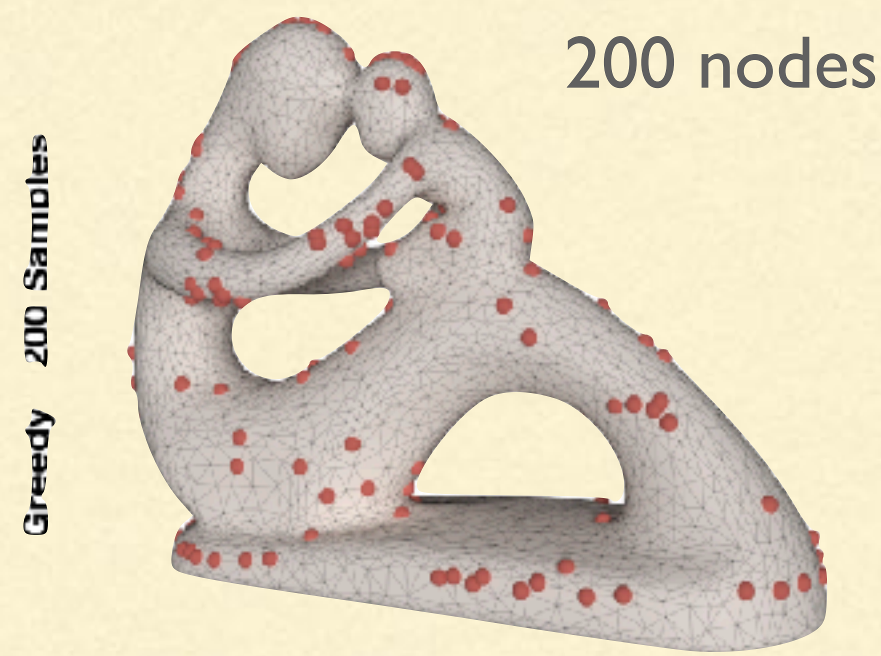
- Worst-case structural analysis
  - **<u>Challenges</u>**: Finite Element Analysis (FEA) for <u>*every*</u> external force locations would be computationally too expensive

Justification for *single, normal, compressive* load can be found in Ulu et al.'17, based on *Rockafellar's Theorem*



MOTIVATING APPLICATION

- Worst-case structural analysis
  - **Idea**: *Sample* a few "representative" force locations and build a *predictive model* for the rest locations
  - **Challenge**: How to determine the "best" representative locations



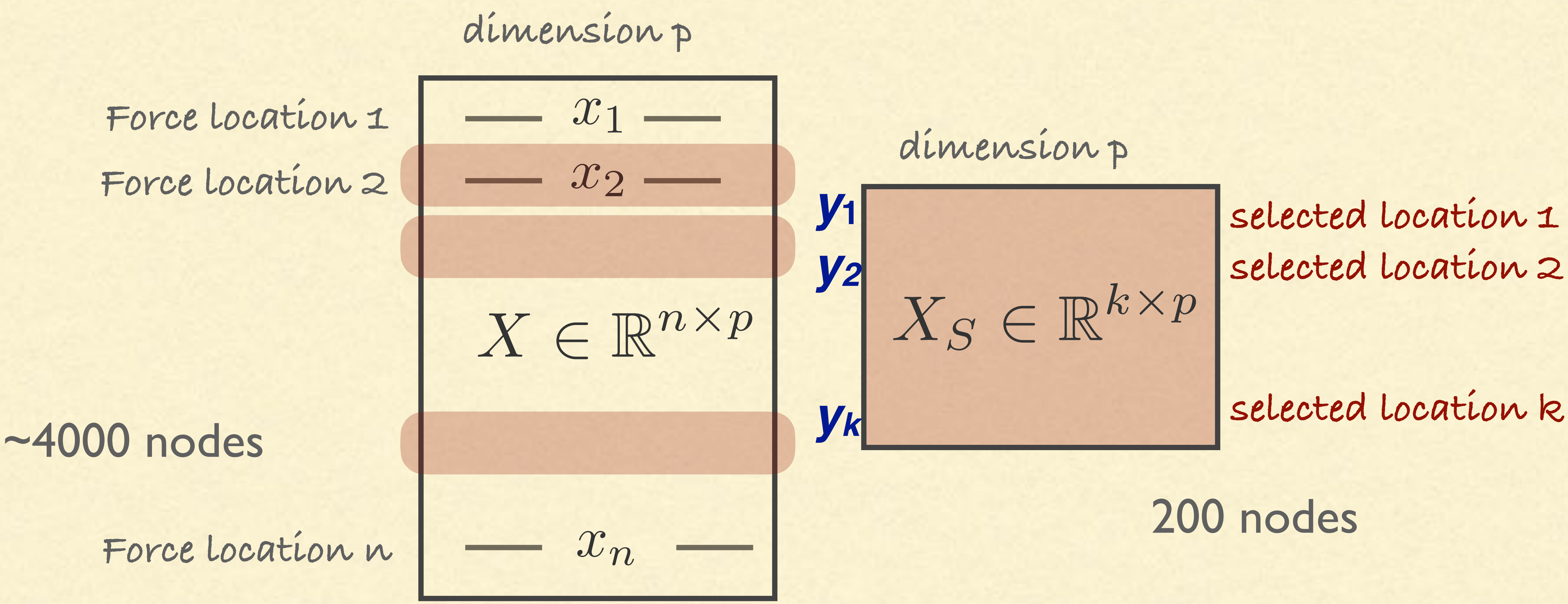
PROBLEM FORMULATION

max. stress response

- Linear regression model: $y_i = \langle x_i, \theta_0 \rangle + \varepsilon_i$ modeling error

top e-vec of surface Laplacian unknown regression model

- Experiment selection:



PROBLEM FORMULATION

- Linear regression model: $y_i = \langle x_i, \theta_0 \rangle + \varepsilon_i$
- Ordinary Least Squares: $\widehat{\theta} = (\sum_{i \in S} x_i x_i^\top)^{-1} (\sum_{i \in S} y_i x_i)$

  - By CLT: $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, (\sum_{i \in S} x_i x_i^\top)^{-1})$

(scaled) sample covariance,
Fisher's Information

Optimal experimental design

Find subset $S \subseteq [n]$, $|S| \leq k$ so as to minimize

$$f\left(\sum_{j \in S} x_j x_j^\top\right)$$

"optimality criteria"

PROBLEM FORMULATION

- Predictive model: $y_i = \langle x_i, \theta_0 \rangle + \varepsilon_i$

Optimal experimental design

Find subset $S \subseteq [n]$, $|S| \leq k$ so as to minimize

$$f\left(\sum_{j \in S} x_j x_j^\top\right)$$

"optimality criteria"

Example:

A-optimality $f_A(\Sigma) = \mathrm{tr}(\Sigma^{-1})/p$ MSE $\mathbb{E}\|\hat{\theta} - \theta_0\|_2^2$

D-optimality $f_D(\Sigma) = \det(\Sigma)^{-1/p}$ "scale invariant"

E-optimality $f_E(\Sigma) = 1/\|\Sigma^{-1}\|_{\mathrm{op}}$

V-optimality

....

PROBLEM FORMULATION

- Predictive model: $y_i = \langle x_i, \theta_0 \rangle + \varepsilon_i$

Optimal experimental design

Find subset $S \subseteq [n]$, $|S| \leq k$ so as to minimize

$$f\left(\sum_{j\in S} x_j x_j^\top\right)$$

Objective: efficient approximation algorithms

$$f\left(\sum_{j\in\widehat{S}} x_j x_j^\top\right) \leq \boxed{C(n,p)} \cdot \min_{|S|\leq k} f\left(\sum_{j\in S} x_j x_j^\top\right)$$

"approximation ratio"

EXISTING RESULTS

- Existing positive results
  - $O(1)$ approximation for D-optimality (*Nikolov & Singh,* STOC'15)
  - O(n/k) approximation for A-optimality (*Avron & Boutsidis,* SIMAX'13)
- Existing negative results
  - NP-Hard for exact optimization of D/E-optimality (*Summa et al., SODA*'15)
  - NP-Hard for $(1+\varepsilon)$ approximation for D-optimality when $\boldsymbol{k=p}$ (*Cerny & Hladik,* Comput. Optim. Appl.'12)

Applicable to only one or two criteria f

REGULAR CRITERIA

Optimal experimental design

Find subset $S \subseteq [n]$, $|S| \leq k$ so as to minimize

$$f\left(\sum_{j \in S} x_j x_j^\top\right)$$

"Regular" criteria:

(A1) **Convexity**: f (or its surrogate) is convex;

(A2) **Monotonicity**: $A \preceq B \implies f(A) \geq f(B)$

(A3) **Reciprocal linearity**: $f(tA) = t^{-1} f(A)$

All popular optimality criteria are "regular", e.g., A/D/E/V/G-optimality

OUR RESULT

Theorem. For all regular criteria *f*, there exists a polynomial time (1+ε) approximation algorithm provided that

#. of design subsets $$k = \Omega(p/\varepsilon^2)$$ #. of variables / dimension

- *Remark 1:* Concurrent to or after our works, 1+ε approx. for D/A-optimality are obtained under condition $k = \Omega(p/\varepsilon + 1/\varepsilon^2)$

  (*Singh & Xie,* SODA'18; Nikolov et al., arXiv'18)

- *Remark 2:* The $k = \Omega(p/\varepsilon^2)$ condition is **tight** for E-optimality and continuous relaxation type methods.

  (Nikolov et al., arXiv'18)

ALGORITHMIC FRAMEWORK

- *Continuous relaxation* of the discrete problem
- *Whitening* of candidate design points
- *Regret minimization* characterization of least eigenvalues
- *Greedy swapping* based on FTRL potential functions

ALGORITHMIC FRAMEWORK

- ***Continuous relaxation*** **of the discrete problem**
- *Whitening* of candidate design points
- *Regret minimization* characterization of least eigenvalues
- *Greedy swapping* based on FTRL potential functions

CONTINUOUS RELAXATION

Optimal experimental design

Find subset $S \subseteq [n]$, $|S| \leq k$ so as to minimize

$$f\left(\sum_{j \in S} x_j x_j^\top\right)$$

- Equivalent formulation:

Relaxation: $0 \leq s_i \leq 1$

$$\min_{s_1, \cdots, s_n} f\left(\sum_{i=1}^{n} s_i x_i x_i^\top\right) \qquad s.t. \sum_{i=1}^{n} s_i \leq k, \quad \boxed{s_i \in \{0, 1\}}$$

- Convex! Can be solved using classical methods (e.g., *projected gradient/ mirror descent*)

CONTINUOUS RELAXATION

Optimal experimental design

Find subset $S \subseteq [n]$, $|S| \leq k$ so as to minimize

$$f\left(\sum_{j \in S} x_j x_j^\top\right)$$

Relaxation: $0 \leq s_i \leq 1$

- Equivalent formulation:

$$\min_{s_1, \cdots, s_n} f\left(\sum_{i=1}^{n} s_i x_i x_i^\top\right) \qquad s.t. \sum_{i=1}^{n} s_i \leq k, \quad \boxed{s_i \in \{0,1\}}$$

- **Question:** Round $\{s_i\}$ to integer values

ALGORITHMIC FRAMEWORK

- *Continuous relaxation* of the discrete problem
- ***Whitening* of candidate design points**
- *Regret minimization* characterization of least eigenvalues
- *Greedy swapping* based on FTRL potential functions

WHITENING

Rounding problem. Given optimal *continuous* solution π , round it to $\widehat{s} \in \{0,1\}^n, \;\; \sum_i \widehat{s}_i \leq k$ such that

$$f(\textstyle\sum_i \widehat{s}_i x_i x_i^\top) \leq (1+O(\varepsilon)) \cdot f(\textstyle\sum_i \pi_i x_i x_i^\top)$$

- Whitening: $\widetilde{x}_i = W^{-1/2} x_i$ where $W = \sum_i \pi_i x_i x_i^\top$
- By monotonicity of *f*, the rounding problem is reduced to

$$\lambda_{\min}(\textstyle\sum_i \widehat{s}_i \widetilde{x}_i \widetilde{x}_i^\top) \geq 1 - O(\varepsilon)$$

ALGORITHMIC FRAMEWORK

- *Continuous relaxation* of the discrete problem
- *Whitening* of candidate design points
- ***Regret minimization* characterization of least eigenvalues**
- *Greedy swapping* based on FTRL potential functions

REGRET MINIMIZATION

- Matrix linear bandit/online learning:

  Action space $\Delta_p = \{A \succeq 0, \mathrm{tr}(A) = 1\}$

  - At each time *t* a player picks an **action** $A_t \in \Delta_p$, observes a **reference** F_t and suffers **loss** $\langle A_t, F_t \rangle$

  - Objective: minimize **regret** of the action sequences

$$R(A) := \sum_{t=1}^{T} \langle F_t, A_t \rangle - \underbrace{\inf_{U \in \Delta_p} \sum_{t=1}^{T} \langle F_t, \Delta \rangle}_{\text{precisely } \lambda_{\min}(\sum F_t)}$$

REGRET MINIMIZATION

- Matrix linear bandit/online learning:
  - At each time *t* a player picks an **action** $A_t \in \Delta_p$, observes a **reference** F_t and suffers **loss** $\langle A_t, F_t \rangle$
  - Objective: minimize **regret** of the action sequences *R*(*A*)
  - Follow-The-Regularized-Leader policy:

$$A_t = \arg\min_{A \in \Delta^p} \Big\{ \underbrace{w(A)}_{\text{“regularizer”}} + \alpha \cdot \sum_{\tau=1}^{t-1} \langle F_\tau, A \rangle \Big\}$$

Example regularizers:

1. MWU: $w(A) = \mathrm{tr}(A^\top (\log A - I))$ $\longrightarrow$ $A_t = \exp\left\{ cI - \alpha \sum_{\tau=1}^{t-1} F_\tau \right\}$
2. $l_{1/2}$-regularization: $w(A) = -2\mathrm{tr}(A^{1/2})$ $\longrightarrow$ $A_t = \left(cI - \alpha \sum_{\tau=1}^{t-1} F_\tau \right)^{-2}$

REGRET MINIMIZATION

swapping of two design points

Regret lemma. Suppose $F_t = u_t u_t^\top - v_t v_t^\top$. Then

$$\inf_{U \in \Delta_p} \sum_{t=0}^{k} \langle F_t, U \rangle \geq \sum_{t=1}^{k} \frac{u_t^\top A_t u_t}{1 + 2\alpha u_t^\top A_t^{1/2} u_t} - \frac{v_t^\top A_t v_t}{1 - 2\alpha v_t^\top A_t^{1/2} v_t} - \frac{2\sqrt{p}}{\alpha}$$

penalty parameter in FTRL

FTRL solution at time t

- Proved using classical analysis of regret of FTRL policies
- F_t: **swapping** of two design points from the pool.

ALGORITHMIC FRAMEWORK

- *Continuous relaxation* of the discrete problem
- *Whitening* of candidate design points
- *Regret minimization* characterization of least eigenvalues
- ***Greedy swapping* based on FTRL potential functions**

GREEDY SWAPPING

Regret lemma. Suppose $F_t = u_t u_t^\top - v_t v_t^\top$. Then

$$\inf_{U \in \Delta_p} \sum_{t=0}^{k} \langle F_t, U \rangle \geq \sum_{t=1}^{k} \frac{u_t^\top A_t u_t}{1 + 2\alpha u_t^\top A_t^{1/2} u_t} - \frac{v_t^\top A_t v_t}{1 - 2\alpha v_t^\top A_t^{1/2} v_t} - \frac{2\sqrt{p}}{\alpha}$$

- A "potential" function:

$$\psi(u, v; A) := \frac{u^\top A u}{1 + 2\alpha u^\top A^{1/2} u} - \frac{v^\top A v}{1 - 2\alpha v^\top A^{1/2} v}$$

GREEDY SWAPPING

Regret lemma. Suppose $F_t = u_t u_t^\top - v_t v_t^\top$. Then

$$\inf_{U \in \Delta_p} \sum_{t=0}^{k} \langle F_t, U \rangle \geq \sum_{t=1}^{k} \frac{u_t^\top A_t u_t}{1 + 2\alpha u_t^\top A_t^{1/2} u_t} - \frac{v_t^\top A_t v_t}{1 - 2\alpha v_t^\top A_t^{1/2} v_t} - \frac{2\sqrt{p}}{\alpha}$$

- The "greedy swapping" algorithm:
  - Start with an arbitrary set $S_0 \subseteq [n]$ of size *k*
  - At each *t*, find $i_t \in S_{t-1}, j_t \notin S_{t-1}$ that maximize $\psi(x_{j_t}, x_{i_t}; A_{t-1})$
  - Greedy swapping: $S_t \leftarrow S_{t-1} \cup \{j_t\} \backslash \{i_t\}$

GREEDY SWAPPING

Regret lemma. Suppose $F_t = u_t u_t^\top - v_t v_t^\top$. Then

$$\inf_{U \in \Delta_p} \sum_{t=0}^{k} \langle F_t, U \rangle \geq \sum_{t=1}^{k} \frac{u_t^\top A_t u_t}{1 + 2\alpha u_t^\top A_t^{1/2} u_t} - \frac{v_t^\top A_t v_t}{1 - 2\alpha v_t^\top A_t^{1/2} v_t} - \frac{2\sqrt{p}}{\alpha}$$

- Proof framework:
  - If $k \geq 5p/\varepsilon^2, \alpha = \sqrt{p}/\varepsilon$ then the "progress" of each swapping is lower bounded by ε/k until $\lambda_{\min} \geq 1 - O(\varepsilon)$
  - Repeat the swapping for at most $O(k/\varepsilon)$ iterations until we're done.

SUMMARY

- Summary of our result (re-cap):
  - Objective: discrete optimization

$$\min_s f\left(\sum_i s_i x_i x_i^\top\right) \quad s.t. \quad s_i \in \{0,1\}, \sum_i s_i \le k$$

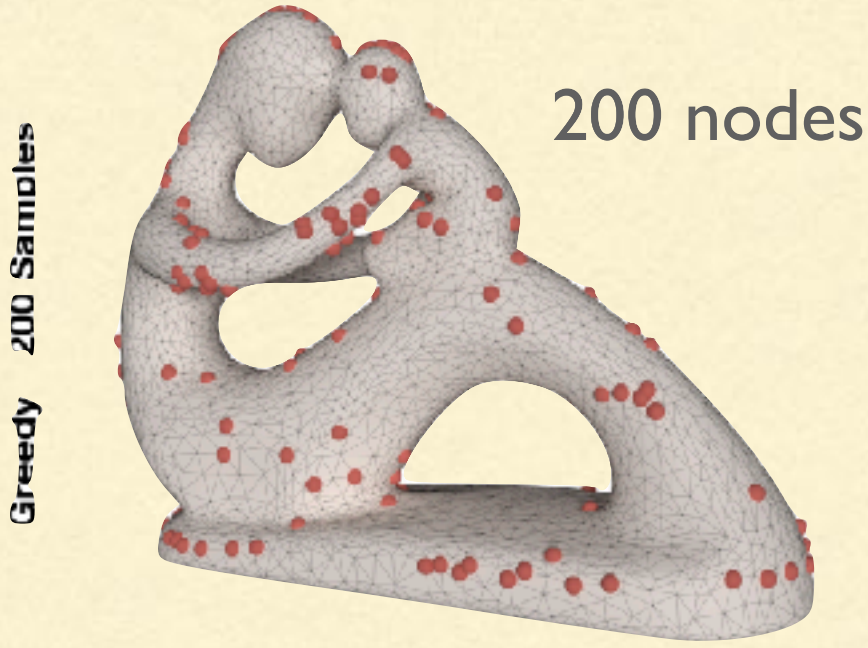
  - Regularity: *f* is "regular" if it is convex, monotonic and reciprocal linear
  - Method: continuous relaxation + greedy swapping

Theorem. For all regular criteria *f*, there exists a polynomial time (1+ε) approximation algorithm provided that

$$k = \Omega(p/\varepsilon^2)$$

APPLICATION

- Worst-case structural analysis
  - *Sample* a few “representative” force locations and build a *predictive model* for the rest locations



APPLICATION

- Worst-case structural analysis
    - *Sample* a few "representative" force locations and build a *predictive model* for the rest locations
    - Predictive model: *Laplacian (linear) smoothing*

max. stress response

$$y_i = \langle x_i, \theta_0 \rangle + \varepsilon_i$$

modeling error

top e-vec of surface Laplacian

unknown regression model

    - Typical problem parameter range:

$$n = 4000 \sim 6000, \quad p = 10 \sim 15, \quad k = 25 \sim 300$$

ALGORITHMIC FRAMEWORK

- Input: a structure with fixed boundary conditions (blue) and contact regions (red).

(a) Fertility (b) Rocking Chair (c) Shark

RESULTS

$$\boxed{n_F = 3914}$$

- Results for the "**Fertility**" model

	$n_{FL} =$	25	50	100	150	200	250	300	Total FEAs
$\delta = 0$	UNIFORM	316.5	149	78.5	37.5	98.5	42.5	39	178.5 ($n_{FL} = 100$)
	LEVSCORE	252.5	54.5	73.5	68.5	42.5	31	13.5	104.5 ($n_{FL} = 50$)
	K-MEANS	237	**25**	61	82	57	**17**	**16**	75 ($n_{FL} = 50$)
	SAMPLING	210.5	148.5	51	30	35.5	34	26.5	151 ($n_{FL} = 100$)
	GREEDY	**12**	26	**13**	**7**	**11**	25	**33**	$\boxed{\textbf{37}\ (n_{FL} = 25)}$
$\delta = 0.05$	UNIFORM	285	80.5	52	10	63	10	10	130.5 ($n_{FL} = 50$)
	LEVSCORE	175	26.5	55.5	59	17	10	7	76.5 ($n_{FL} = 50$)
	K-MEANS	144	**2**	19	22	14	**2**	**2**	52 ($n_{FL} = 50$)
	SAMPLING	202	113	10	**7**	11	8	6	110 ($n_{FL} = 100$)
	GREEDY	**4**	3	**4**	**7**	**5**	**2**	6	$\boxed{\textbf{29}\ (n_{FL} = 25)}$

Our algorithm

RESULTS

$n_F = 5348$

- Results for the "**RockingChair**" model

	$n_{FL} =$	25	50	100	150	200	250	300	Total FEAs
$\delta = 0$	UNIFORM	716	857	385.5	42	135.5	269.5	**36**	192 ($n_{FL} = 150$)
	LEVSCORE	764.5	208.5	**36**	**36**	**36**	**36**	**36**	136 ($n_{FL} = 100$)
	K-MEANS	4013	4400	4573	4301	4320	4620	4757	4038 ($n_{FL} = 25$)
	SAMPLING	672.5	282	38.5	38	38	**36**	**36**	138.5 ($n_{FL} = 100$)
	GREEDY	**36**	**35**	208	**35**	**36**	**36**	**36**	**61** ($n_{FL} = 25$)
$\delta = 0.05$	UNIFORM	404	466	201.5	20	88	93.5	18	170 ($n_{FL} = 150$)
	LEVSCORE	444	192.5	20	**18.5**	**18**	**18**	**18**	120 ($n_{FL} = 100$)
	K-MEANS	285	466	**14**	24	26	161	195	114 ($n_{FL} = 100$)
	SAMPLING	540	268	21.5	20.5	20.5	20	20	121.5 ($n_{FL} = 100$)
	GREEDY	**20**	**19**	200	20	20	20	20	**45** ($n_{FL} = 25$)

RESULTS

$n_F = 4281$

- Results for the "**Shark**" model

	$n_{FL} =$	25	50	100	150	200	250	300	Total FEAs
$\delta = 0$	UNIFORM	585	384	141.5	208.5	20	**9**	9.5	220 ($n_{FL} = 200$)
	LEVSCORE	478.5	**9**	**9**	**9**	**9**	**9**	**9**	59 ($n_{FL} = 50$)
	K-MEANS	133	102	**9**	**9**	**9**	**9**	**9**	109 ($n_{FL} = 100$)
	SAMPLING	963.5	87	**9**	**9**	**9**	**9**	**9**	109 ($n_{FL} = 100$)
	GREEDY	**9**	171	**9**	**9**	**9**	**9**	**9**	**34** ($n_{FL} = 25$)
$\delta = 0.01$	UNIFORM	568.5	341	131.5	156	15	**4**	4.5	215 ($n_{FL} = 200$)
	LEVSCORE	416	**4**	**4**	**4**	**4**	**4**	**4**	54 ($n_{FL} = 50$)
	K-MEANS	129	84	**4**	**4**	**4**	**4**	**4**	104 ($n_{FL} = 100$)
	SAMPLING	872.5	69	**4**	**4**	**4**	**4**	**4**	104 ($n_{FL} = 100$)
	GREEDY	**4**	115	**4**	**4**	**4**	**4**	**4**	**29** ($n_{FL} = 25$)

SAMPLING ALGORITHM

- **Results**: comparison with equi-distance sampling

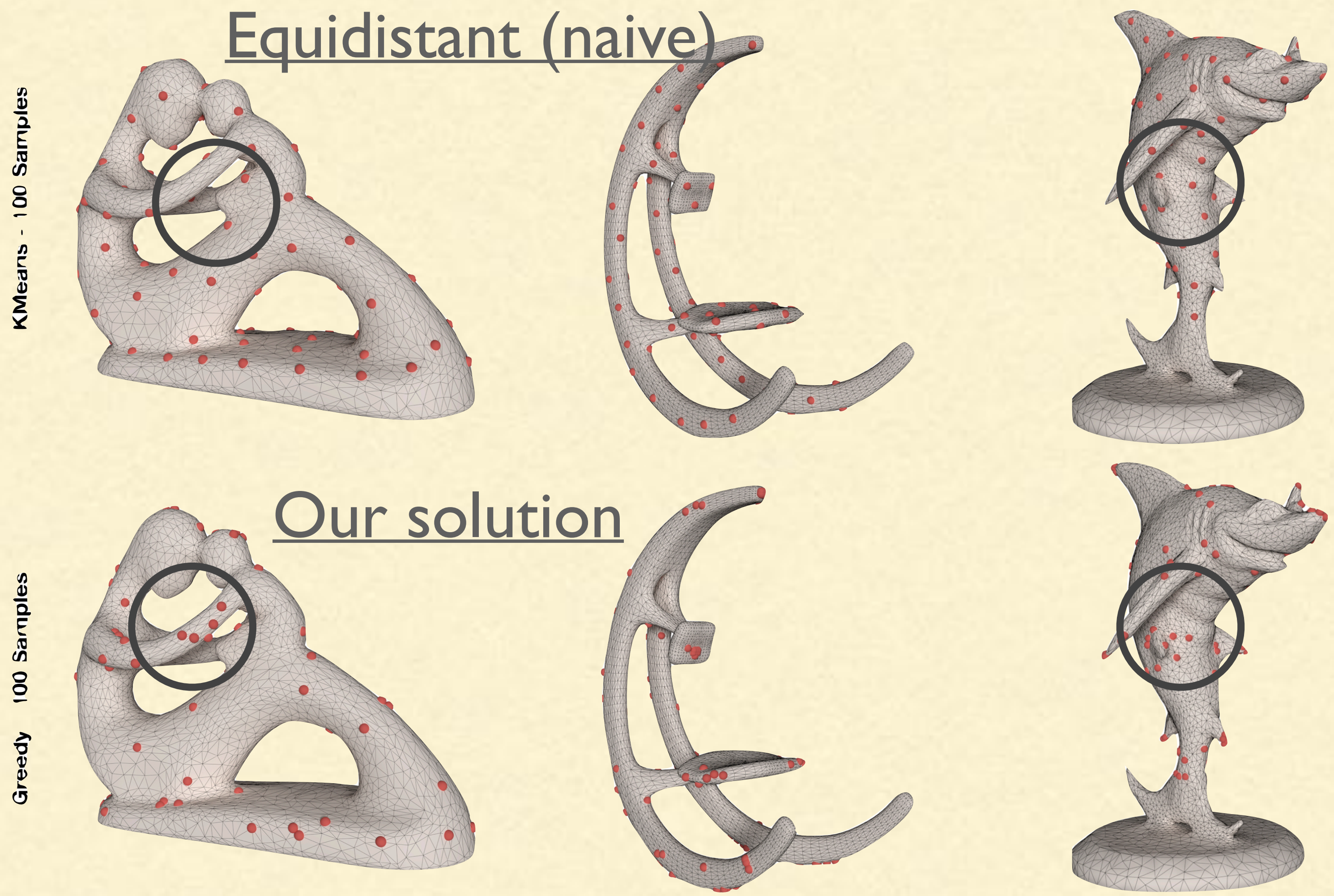
  - K=100 sampling points

"Sensitive" regions (e.g., arms, wingtips) more sampled

"Easy" regions less sampled

Equidistant (naive)

Our solution



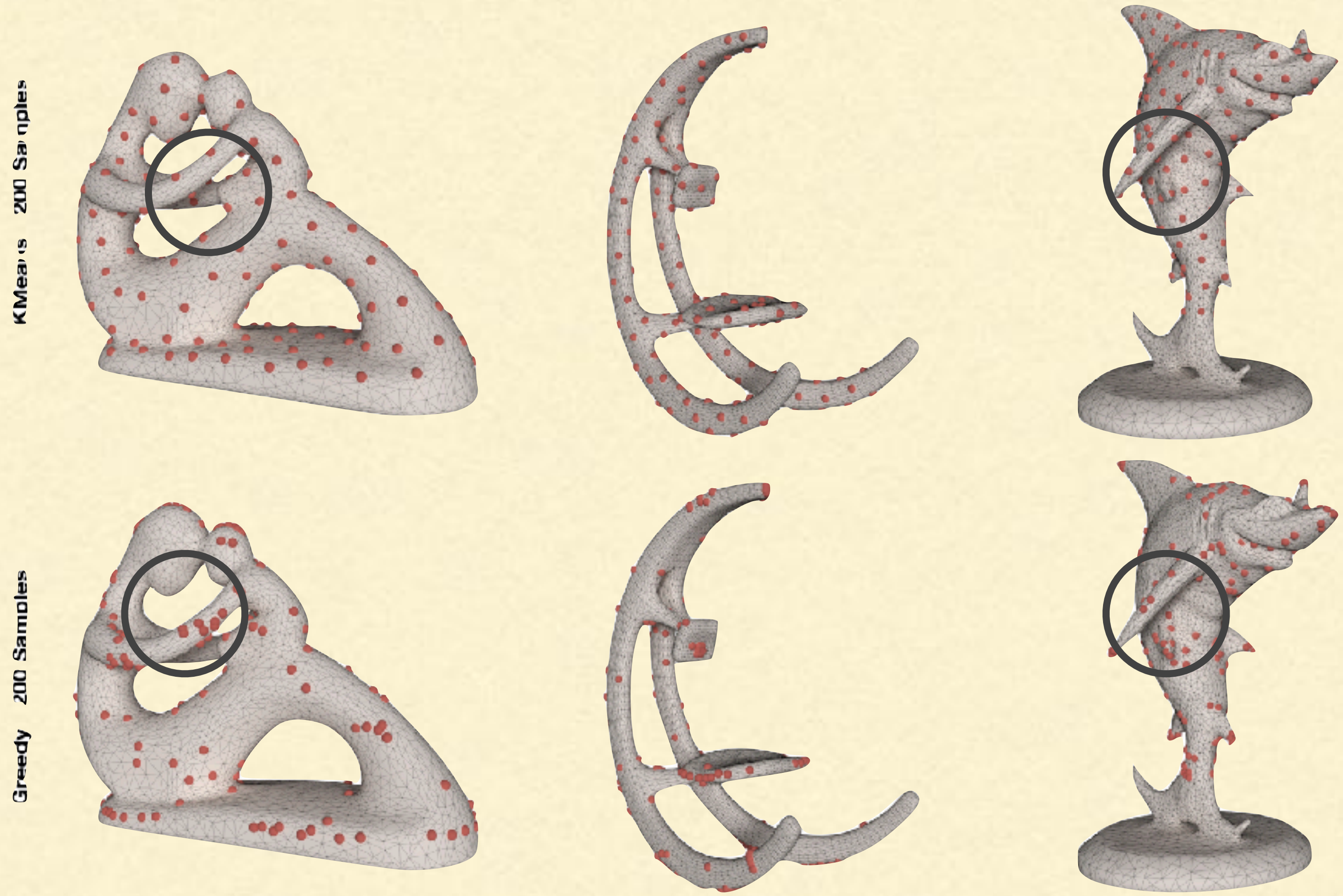
SAMPLING ALGORITHM

- **Results**: comparison with equidistance sampling

  - *K*=200 sampling points

"Sensitive" regions (e.g., arms, wingtips) more sampled

"Easy" regions less sampled



EXTENSIONS

- “Robust” experimental design
  - Design points are subject to adversarial “perturbations”
  - Example discrete optimization problem:

Adversarial perturbations

$$\min_{s_1,\cdots,s_n} \max_{\xi_1,\cdots,\xi_n} f\left(\sum_{i=1}^{n}(x_i+\xi_i)(x_i+\xi_i)^\top\right)$$

$$s.t. \quad s_i \in \{0,1\},\ \sum_i s_i \le k,\ \|\xi_i\|_2 \le \delta$$

EXTENSIONS

- "Random design" linear regression
  - Random designs $(x_i, y_i) \sim D$, but $\mathbb{E}[y_i|x_i] \neq \langle x_i, \beta_0 \rangle$
  - Worst-case optimal designs:

$$\min_S \sup_{D \in \mathcal{D}} \mathbb{E}_{D,S} \left[\|\widehat{\beta}_S - \beta_0\|_2^2 \right]$$

S: selected design subset
β_0: best linear predictor w.r.t. D
β_S: OLS on X_S

Variance: $f(\sum_{i \in S} x_i x_i^\top)$

Bias: dependent on D

Thank you!