Optimal experimental design for inverse problems via column subset selection

Arvind K. Saibaba[a]

March 6, 2026

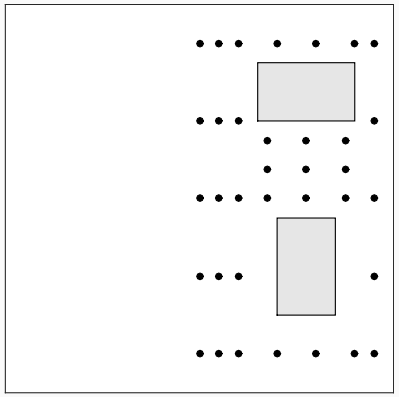
Department of Mathematics
North Carolina State University

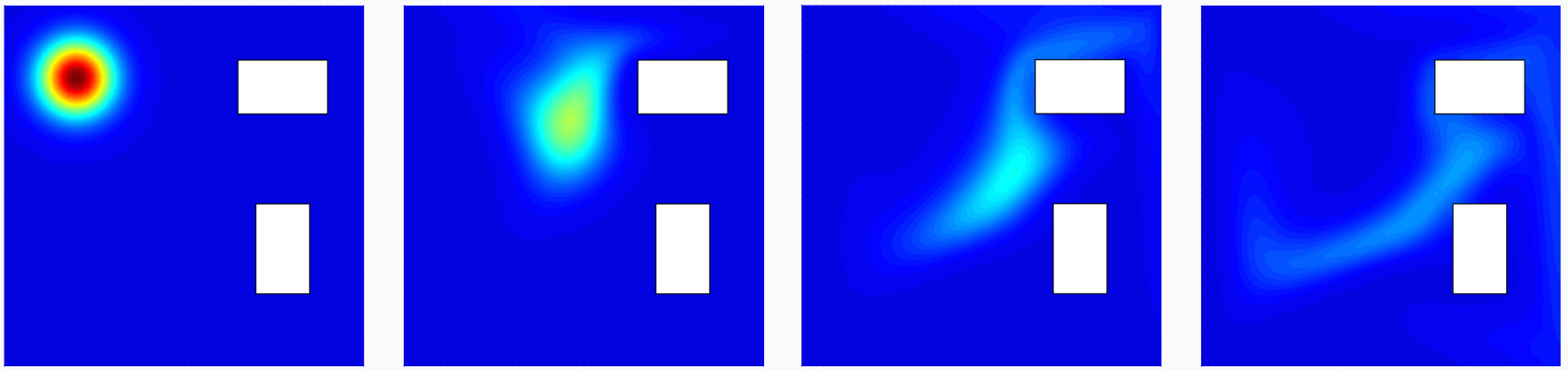
[a]Thanks to the National Science Foundation and Department of Energy for funding support.

Model Problem: 2D contaminant source identification

Forward problem: Contaminant Transport

$$
\begin{aligned}
u_t - \kappa \Delta u + \mathbf{v} \cdot \nabla u &= 0 && \text{in } \mathcal{D} \times [0, T] \\
u(0, \boldsymbol{x}) &= m(\boldsymbol{x}) && \text{in } \mathcal{D} \\
\kappa \nabla u \cdot \mathbf{n} &= 0 && \text{on } \partial\mathcal{D} \times [0, T]
\end{aligned}
$$





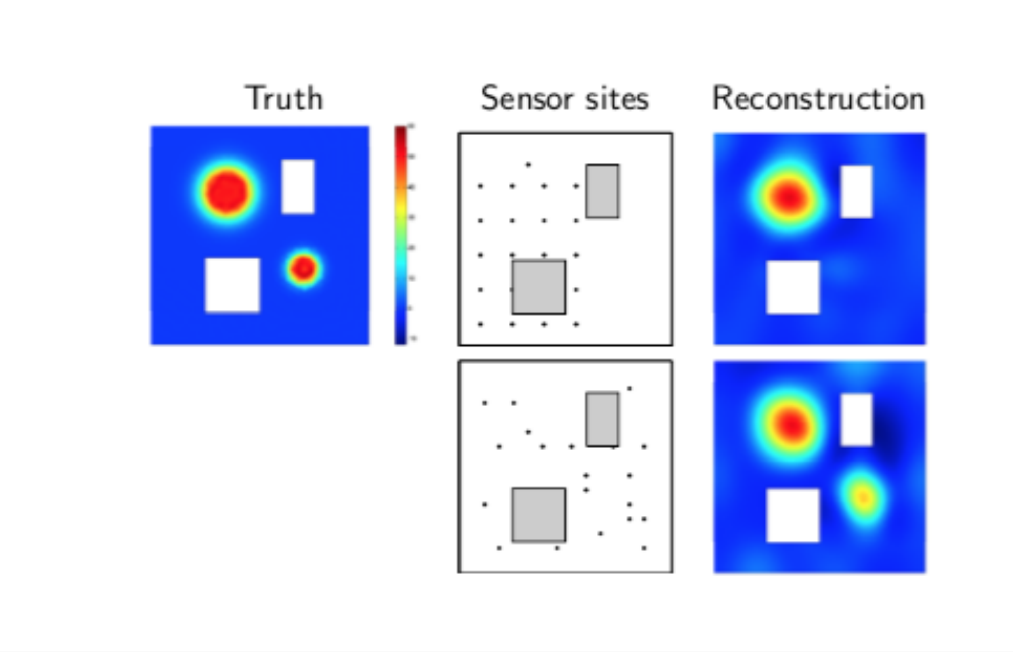
Inverse Problem: Recover initial conditions from sensor measurements.

This assumes data has been collected.

Graphics courtesy of Alen Alexanderian.

Importance of a good design

Data acquisition can be limited by physics/budget. A poorly designed experiment may miss important information.



Goal: How to optimally collect limited data to reduce uncertainty in the reconstructions?

Our solution: Relate it to the column subset selection problem in NLA.

Graphics courtesy of Alen Alexanderian.

Sensor placement in Linear Bayesian Inverse Problems

Bayesian Inverse Problem

Data Acquisition:

$$\boldsymbol{d} = \boldsymbol{F}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \eta^2 \boldsymbol{I})$$

Prior distribution is Gaussian with $\boldsymbol{\theta} \sim \mu_{\text{pr}} = \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{pr}})$

Posterior distribution is Gaussian $\boldsymbol{\theta}|\boldsymbol{d} \sim \mu^{\boldsymbol{d}}_{\text{post}} = \mathcal{N}(\boldsymbol{\theta}^{\boldsymbol{d}}_{\text{post}}, \boldsymbol{\Gamma}_{\text{post}})$, where

$$\boldsymbol{\Gamma}_{\text{post}} = (\eta^{-2}\boldsymbol{F}^\top \boldsymbol{F} + \boldsymbol{\Gamma}_{\text{pr}}^{-1})^{-1} \qquad \boldsymbol{\theta}^{\boldsymbol{d}}_{\text{post}} = \boldsymbol{\Gamma}_{\text{post}}(\eta^{-2}\boldsymbol{F}^\top \boldsymbol{d}).$$

Optimality criteria

D-optimality criterion/Expected information Gain

The D-optimality criterion is

$$\begin{aligned}\phi_D &\equiv \mathbb{E}_{\boldsymbol{d}}[\mathrm{KL}(\mu^{\boldsymbol{d}}_{\mathrm{post}}||\mu_{\mathrm{pr}})] \\ &= -\log\det(\boldsymbol{\Gamma}_{\mathrm{post}}) + \log\det(\boldsymbol{\Gamma}_{\mathrm{pr}}).\end{aligned}$$

Some remarks:

1. Measure of information gain from the prior to the posterior
2. Measure of uncertainty of the reconstruction
3. Mutual information between data and parameters

Other criteria (A/C/T/Goal-Oriented/...) are available

D-optimality criterion: Reformulation

Reformulate the D-optimality criterion:

$$\phi_D = \text{logdet}(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{A}^\top) \qquad \boldsymbol{A} \equiv \eta^{-1}\boldsymbol{\Gamma}_{\text{pr}}^{1/2}\boldsymbol{F}^\top.$$

1. Columns of $\boldsymbol{A}$ map to sensors or design variables
2. $\boldsymbol{A}$ is the preconditioned adjoint operator, and
3. $\boldsymbol{A}$ has decaying singular values in many instances

Optimization viewpoint

Binary Optimization

$$\max_{\boldsymbol{w}\in\{0,1\}^m} \operatorname{logdet}(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^\top) \qquad \boldsymbol{W} = \operatorname{diag}(w_1, \ldots, w_m),$$

subject to the constraint $\sum_{j=1}^{k} w_j \leq k$.

- A weight of 1 means place sensor there, 0 means don't place sensor.
- Large search space with $\binom{m}{k}$ possibilities
- This optimization problem is either solved greedily, with relaxation+rounding, etc.

Prior work: see review papers

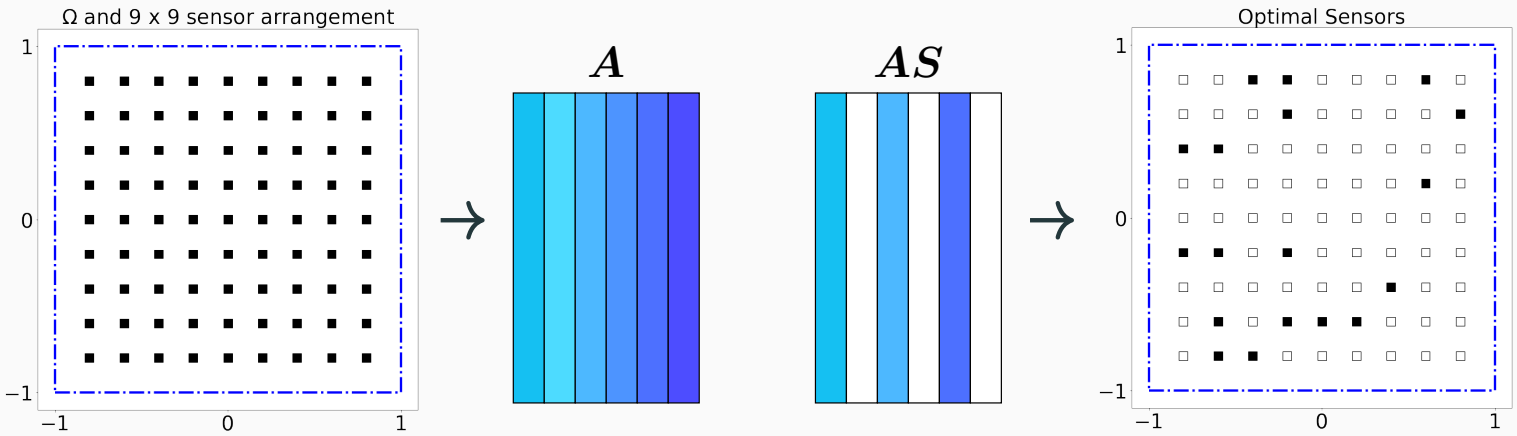
- Chaloner, Verdinelli (1995), Alexanderian (2021), Huan, Jagalur, Marzouk (2024)

OED via Column subset selection

Joint work with Srinivas Eswar, Vishwas Rao (both at Argonne)

Bayesian D-optimal experimental designs via column subset selection. To appear, SIAM Journal on Scientific Computing, arXiv:5428327, 2026.

Our approach in one graphic



Candidate sensor locations map to columns, and we perform column subset selection to identify optimal sensor placement.

Column selection perspective

Our approach: Optimize over $m \times m$ permutation matrices $\boldsymbol{\Pi}$

$$\max_{\boldsymbol{\Pi} \in \mathbb{R}^{m \times m}} \operatorname{logdet}(\boldsymbol{I} + \boldsymbol{AS}(\boldsymbol{AS})^\top) \qquad \boldsymbol{S} = \boldsymbol{\Pi}(:, 1:k).$$

We also use $S = \{i_1, \ldots, i_k\}$ to denote the selected indices.

Alternatively, select k "best" columns, $\boldsymbol{AS}$, from $\boldsymbol{A}$.

Computational Complexity

This is an NP-Hard problem (reduction to Exact Cover by 3-sets).

Similar reduction in Civril, Magdon-Ismail, 2009, for other subset selection problems.

Golub-Klema-Stewart approach

Want to find a permutation matrix $\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 \end{bmatrix}$ such that

$$\boldsymbol{A} \begin{bmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{T} \end{bmatrix}.$$

Main idea: plug in SVD

$$\boldsymbol{A}\boldsymbol{\Pi} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{\Pi}, \qquad \boldsymbol{V}^\top\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix}.$$

Finding most "independent" columns of $\boldsymbol{A} \Leftrightarrow$ a submatrix $\boldsymbol{V}_{11}$ with maximum volume:

$$\frac{\sigma_j(\boldsymbol{A})}{\|\boldsymbol{V}_{11}^{-1}\|_2} \leq \sigma_j(\boldsymbol{C}) \leq \sigma_j(\boldsymbol{A}) \qquad 1 \leq j \leq k.$$

Golub-Klema-Stewart approach

Two stage approach:

1. Compute truncated SVD: $\boldsymbol{A} \approx \boldsymbol{U}_k \boldsymbol{\Sigma}_k \boldsymbol{V}_k^\top$.
2. Compute pivoted QR on $\boldsymbol{V}_k^\top$ $(k \times m)$

$$\boldsymbol{V}_k^\top \begin{bmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 \end{bmatrix} = \boldsymbol{Q}_1 \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \end{bmatrix}.$$

Set selection operator $\boldsymbol{S} = \boldsymbol{\Pi}_1$.

Golub-Klema-Stewart approach

Two stage approach:

1. Compute truncated SVD: $\boldsymbol{A} \approx \boldsymbol{U}_k \boldsymbol{\Sigma}_k \boldsymbol{V}_k^\top$.
2. Compute pivoted QR on $\boldsymbol{V}_k^\top$ $(k \times m)$

$$\boldsymbol{V}_k^\top \begin{bmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 \end{bmatrix} = \boldsymbol{Q}_1 \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \end{bmatrix}.$$

Set selection operator $\boldsymbol{S} = \boldsymbol{\Pi}_1$.

MATLAB code:

```
A = Gamma_pr_sqrt*F'/eta; % Form preconditioned operator
[~,~,v] = svd(A,'econ');  % SVD
[~,~,p] = qr(v(:,1:k)', 0); % Pivoted QR
p = p(1:k); % Select indices
```

Performance guarantees

D-optimality criterion $\phi_D(S) = \operatorname{logdet}(\boldsymbol{I} + \boldsymbol{AS}(\boldsymbol{AS})^\top)$

Performance of GKS (Eswar, Rao, S.)

We obtain the following bound for the D-optimality

$$\operatorname{logdet}(\boldsymbol{I} + \boldsymbol{\Sigma}_k^2/q(m,k)^2) \leq \phi_D(S) \leq \phi_D(S_{\text{opt}}) \leq \operatorname{logdet}(\boldsymbol{I} + \boldsymbol{\Sigma}_k^2),$$

where $q(m,k)$ depends on the technique used to select S.

Method	$q(m,k)$	Cost (flops)
sRRQR	$\sqrt{1 + f^2 k(m-k)}$	$O(mk^2 \log_f(m))$
Pivoted QR	$\sqrt{m-k} \cdot 2^k$	$O(mk^2)$

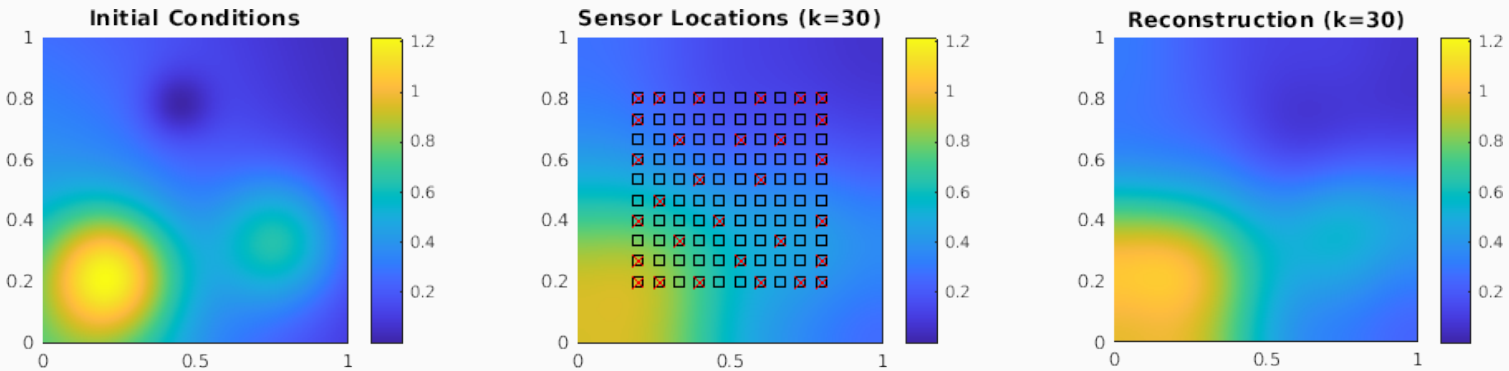
Here $f \geq 1$ is a user-specified parameter in strong rank-revealing QR (sRRQR). In numerical experiments, use column pivoted QR.

2D Heat Problem

Problem setup: Recover initial conditions from partial measurements at final time $T = 0.01$

$$\frac{\partial u}{\partial t} = \Delta u \qquad u(\boldsymbol{x}, 0) = \theta(\boldsymbol{x})$$

with homogeneous Neumann bcs, 2% noise is added to data, Whittle-Matérn prior.

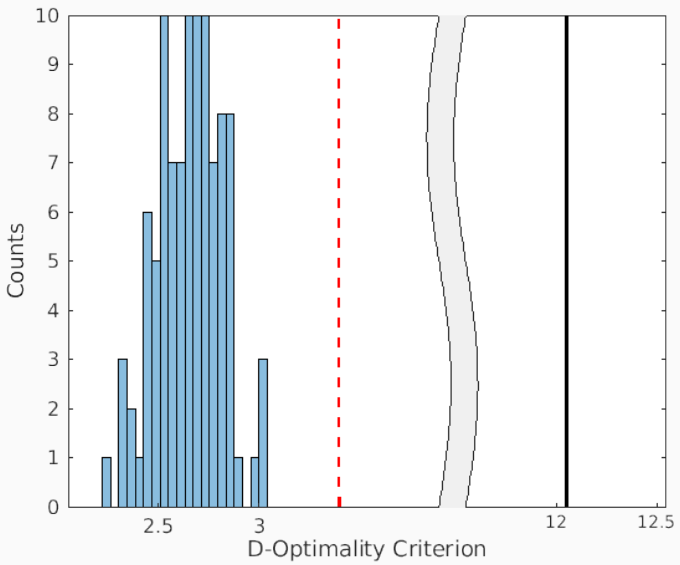


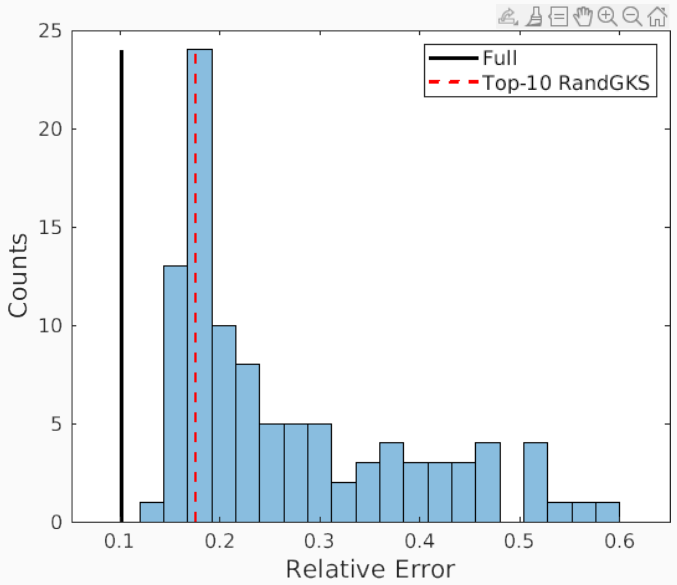
Goal: Pick the best $k = 10/30$ sensor locations out of 100.

Implementation: we used GKS with randomized SVD + pivoted QR.

2D Heat Problem: $k = 10$ sensors

We compare two metrics: D-optimality and 2-norm relative error

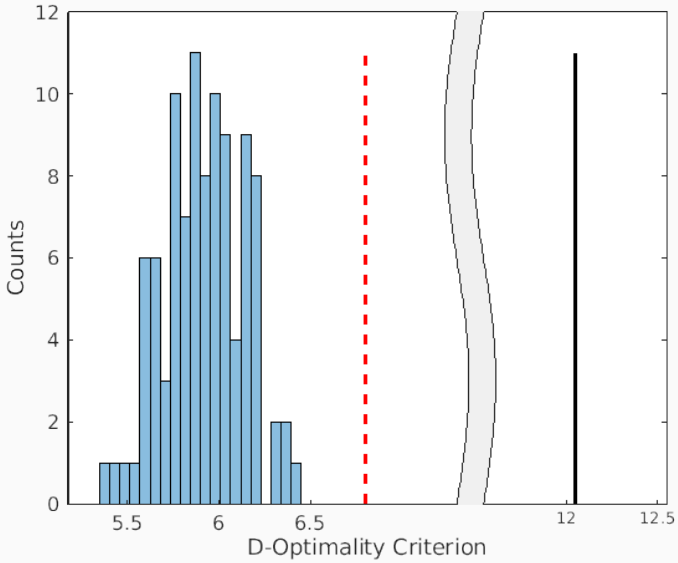


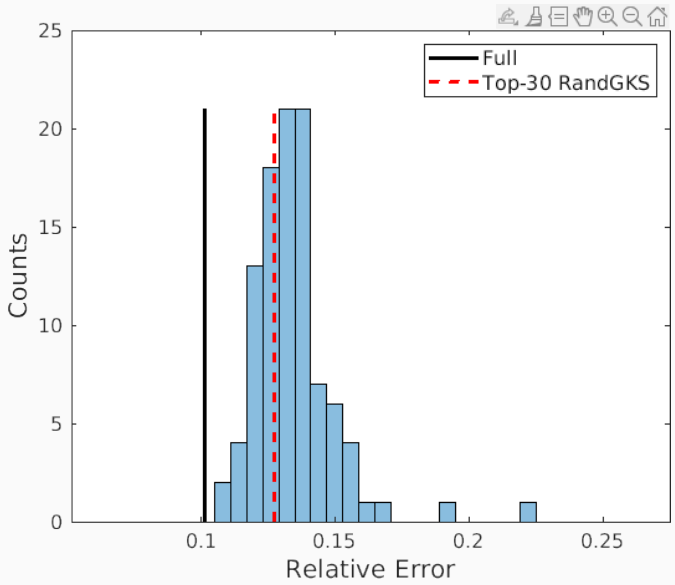


CSSP beats random designs every time in D-optimality, does okay in terms of relative error in MAP.

2D Heat Problem: $k = 30$ sensors

We compare two metrics: D-optimality and 2-norm relative error





CSSP beats random designs every time in D-optimality, does okay in terms of relative error in MAP.

Randomized OED algorithms

Two major costs of the GKS approach:

1. Cost of SVD of $\boldsymbol{A}$: T_{SVD}
2. Cost of the pivoted QR of $\boldsymbol{V}_k^\top$: $O(mk^2)$ flops

For some applications, both can be very expensive.

Randomized OED algorithms

Two major costs of the GKS approach:

1. Cost of SVD of $\boldsymbol{A}$: T_{SVD}
2. Cost of the pivoted QR of $\boldsymbol{V}_k^\top$: $O(mk^2)$ flops

For some applications, both can be very expensive.

Overview of randomized methods:

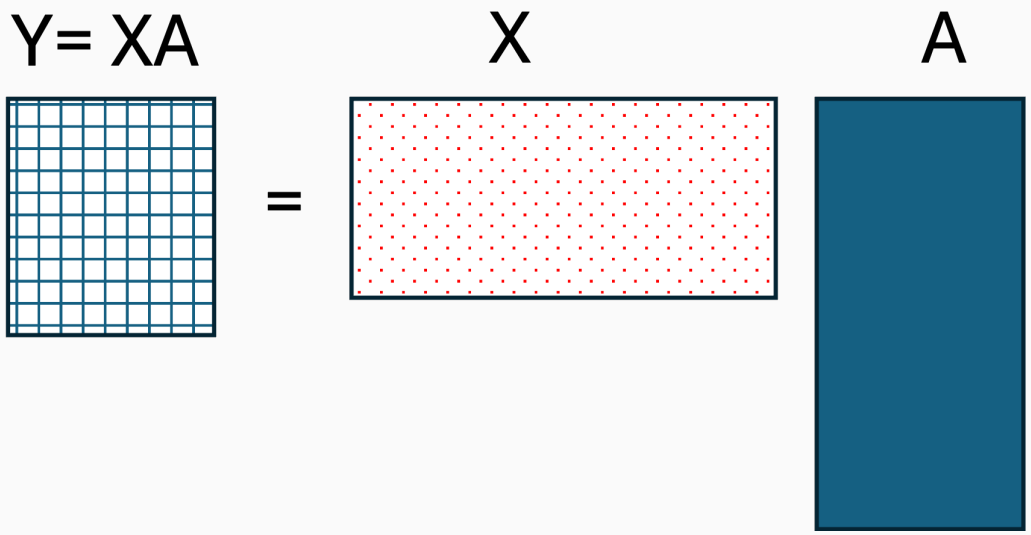
1. Computing the basis $\boldsymbol{V}_k$ can be accelerated using randomized SVD

2. Sampling from the index set $\{1, \ldots, m\}$
   - Leverage score sampling: proportional to squared row norms of $\boldsymbol{V}_k$
   - Hybrid approach: Leverage score sampling + deterministic truncation

3. Randomized adjoint-free method

Randomized adjoint-free OED (RAF-OED)

Sketch the matrix $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A} \in \mathbb{R}^{d \times m}$, do subset selection on $\boldsymbol{Y}$



1. The sketching matrix $\boldsymbol{X}$ is a subspace embedding:
   - e.g., iid entries $\mathcal{N}(0, 1/d)$, $d \sim k$
2. The matrix $\boldsymbol{Y}$ has few rows, so computationally more efficient
3. Can be implemented matrix-free and adjoint-free since

$$\boldsymbol{Y}^\top = \boldsymbol{A}^\top \boldsymbol{X}^\top = \eta^{-1} \boldsymbol{F}(\boldsymbol{\Gamma}_{\text{pr}}^{1/2} \boldsymbol{X})$$

Performance guarantees: RAF-OED

Recall: $\phi_D(S) = \operatorname{logdet}(\boldsymbol{I} + (\boldsymbol{AS})(\boldsymbol{AS})^\top)$

D-optimality

For RAF-OED, with probability at least $1 - \delta$,

$$\operatorname{logdet}(\boldsymbol{I} + \boldsymbol{\Sigma}_k^2/q(m,k)^2) \le \phi_D(S) \le \phi_D(S_{\text{opt}}) \le \operatorname{logdet}(\boldsymbol{I} + \boldsymbol{\Sigma}_k^2),$$

where $d = k + p$, $p \ge 2$, and

$$q(m,k) = \underbrace{\sqrt{1 + f^2 k(m-k)}}_{\text{sRRQR}} \cdot \underbrace{\frac{e\sqrt{d}}{p+1}\left(\frac{2}{\delta}\right)^{1/(p+1)}\left(\sqrt{n} + \sqrt{d} + \sqrt{2\ln(2/\delta)}\right)}_{\text{Randomization}}.$$

This gives a weaker guarantee than sRRQR but comparable numerical performance.

Comparison of methods: Heat Problem

Algorithm	D-optimality	Relative Error	$\|\boldsymbol{V}_{11}^{-1}\|_2$	Time
Full	12.0491	0.1015	—	—
RandGKS	6.8017	0.1274	5.8193	124.59 s
Hybrid	6.6621	0.1194	6.6528	122.87 s
RAF	6.6441	0.1263	10.413	41.63 s
Greedy	7.0857	0.1319	∞	1348.89 s

RAF is nearly 30$\times$ faster than Greedy, with comparable performance in terms of D-optimality.

Big selling point: $\sim k$ forward operator applies, adjoint-free

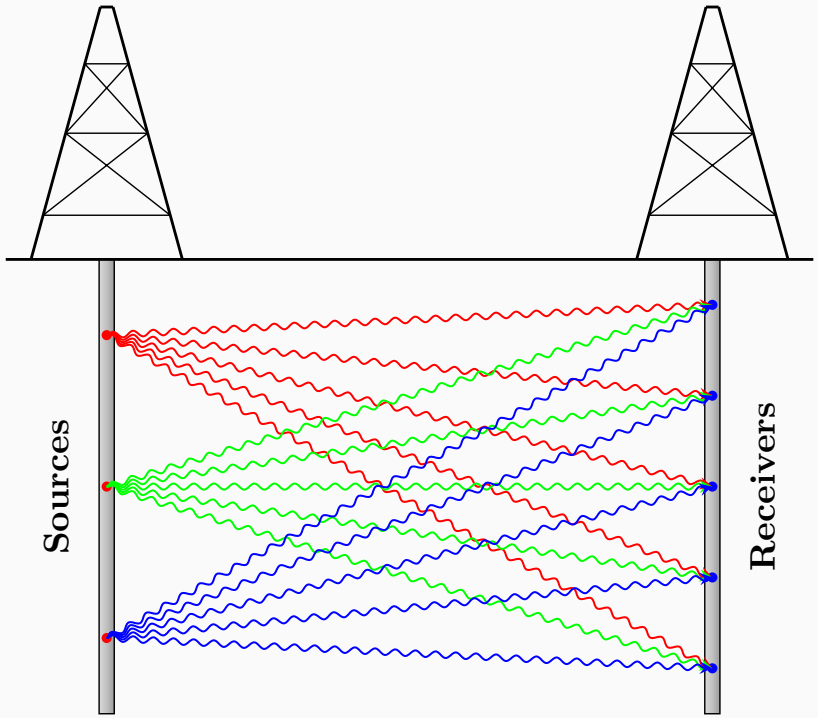
Structured OED

Joint work with Hugo Díaz (NC State), Srinivas Eswar, Vishwas Rao, Zichao Wendy Di (Argonne).

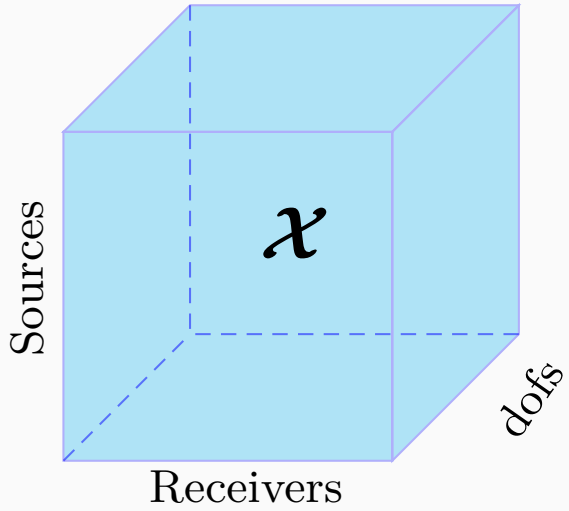
Structured Column Subset Selection for Bayesian Optimal Experimental Design. arXiv preprint arXiv:2506.0033 (2025)

Motivating application: Seismic imaging

Goal: Select k_1 out of m_1 sources and k_2 out of m_2 receivers.



Seismic Data



Idea: View the matrix $\boldsymbol{A}$ as a $m_1 \times m_2 \times n$ tensor.

Tensor-based reformulation

Problem setting: We have $m_1 \times \cdots \times m_d$ experiments, and want to choose $k_1 \times \cdots \times k_d$ experiments.

Our Approach

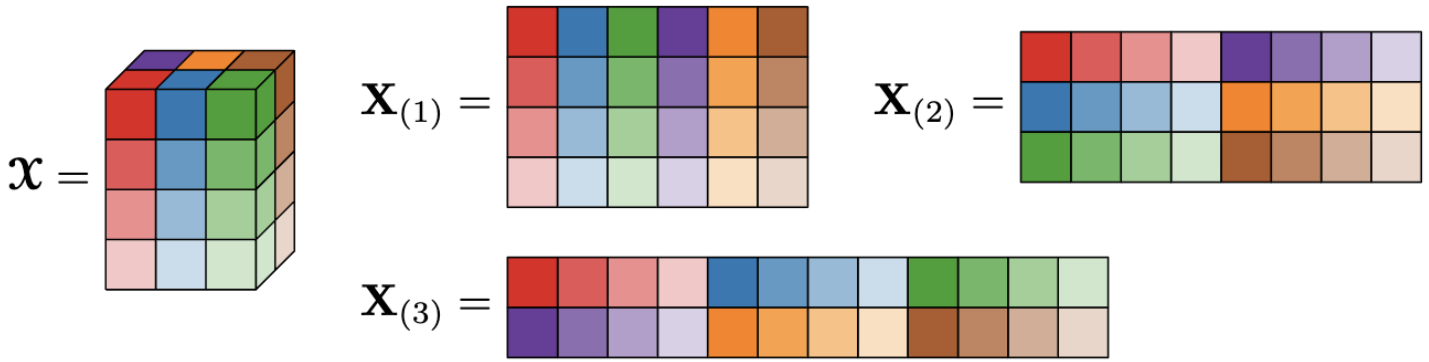
Reshape $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ to a $(d+1)-$order tensor

$$\boldsymbol{\mathcal{X}} : m_1 \times \cdots \times m_d \times n, \qquad m = \prod_{j=1}^{d} m_j.$$

We have d experimental variables and want to select $(k_1, \ldots, k_d)$ experiments.

Tensor unfoldings

In general, for a $d+1$-way tensor, we have $d+1$ unfoldings.



Idea: do row selection on each unfolding.

Image credit: Ballard and Kolda, Tensor Decompositions for Data Science.

Overview: Three ways of selections

1. **IndSelect**: Independently select rows from each mode unfolding

2. **SeqSelect**: Sequentially select rows from each mode unfolding, by incorporating previous selections

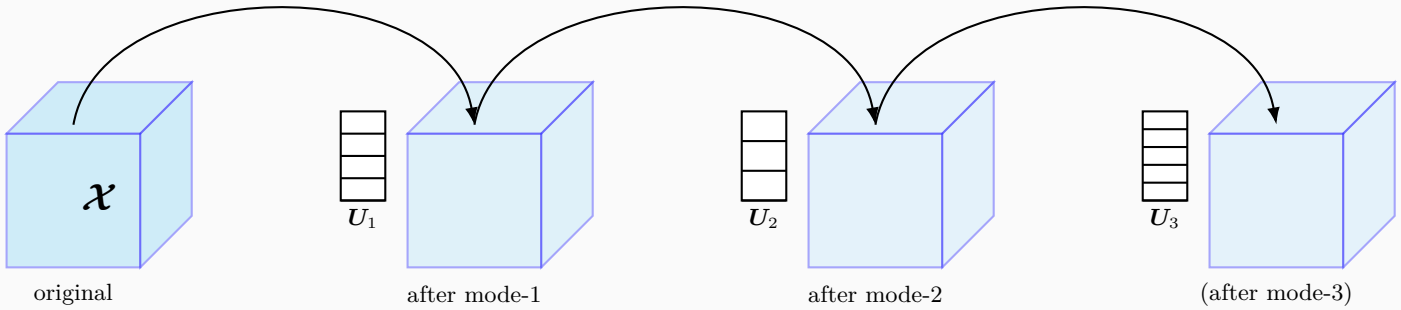
3. **IterSelect**: Start with random initialization. Iteratively select rows from a particular mode unfolding, with all other modes being fixed

Computational cost is but roughly in decreasing order. We have derived greedy versions of these algorithms as well.

Tensor-based OED

1. IndSelect

Selects indices independently in each mode. Analogy: **HOSVD**.

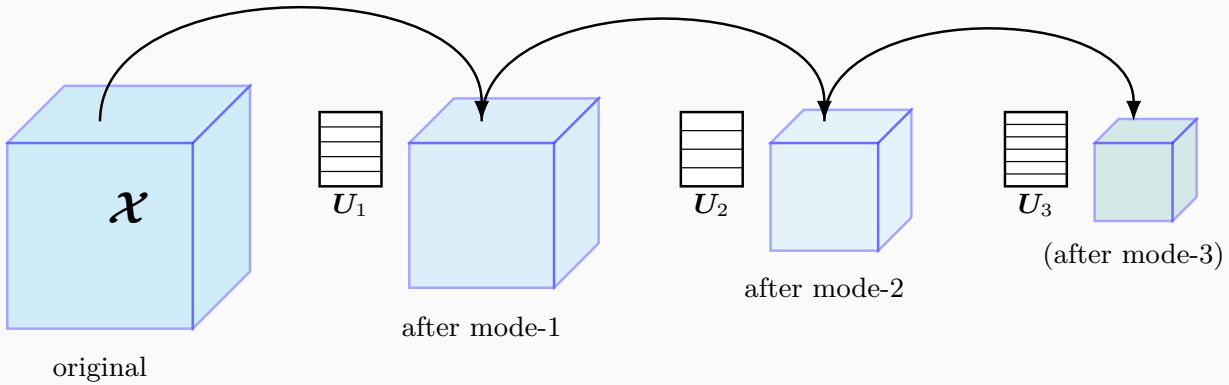


- Here $\boldsymbol{U}_1, \boldsymbol{U}_2$, and $\boldsymbol{U}_3$ are left singular vectors of the mode unfoldings.
- We perform pivoted QR on $\boldsymbol{U}_i^\top$ for $1 \leq i \leq 3$

Tensor-based OED

2. SeqSelect

Sequential selection strategy that updates the choices mode-by-mode.
Analogy: **ST-HOSVD**.



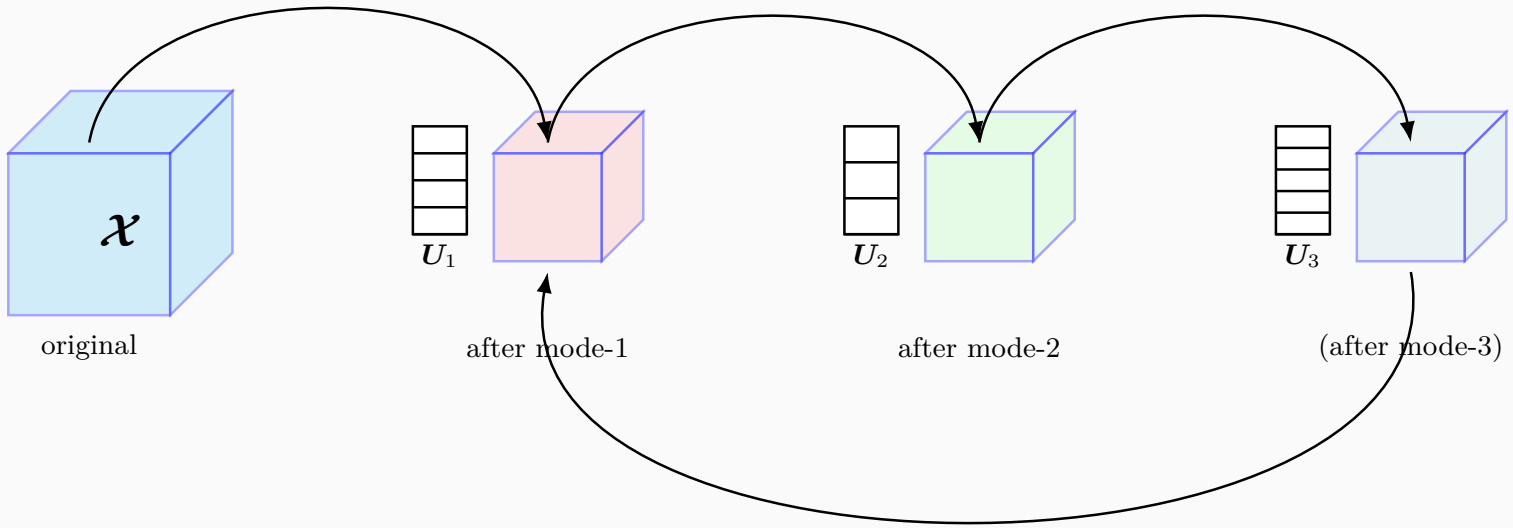
- Here $\boldsymbol{U}_1, \boldsymbol{U}_2$, and $\boldsymbol{U}_3$ are left singular vectors of the mode unfoldings.
- We perform pivoted QR on $\boldsymbol{U}_i^\top$ for $1 \leq i \leq 3$

Tensor-based OED

3. IterSelect

Iterative refinement method that improves the selections across modes.

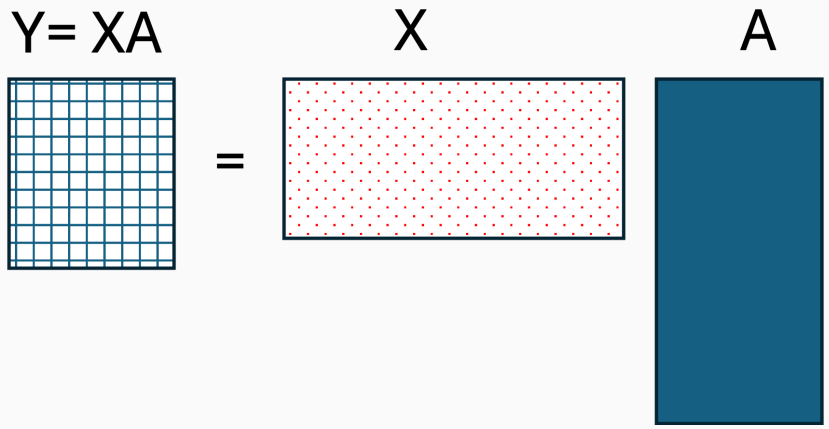
Analogy: **HOOI**. Starts from a small random initial selection.



- Here $\boldsymbol{U}_1, \boldsymbol{U}_2$, and $\boldsymbol{U}_3$ are left singular vectors of the mode unfoldings.
- We perform pivoted QR on $\boldsymbol{U}_i^\top$ for $1 \leq i \leq 3$

Randomized Adjoint-free Approach

Sketch : $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A} \in \mathbb{R}^{s \times m}$ Reshape $\mathcal{Y} \in \mathbb{R}^{m_1 \times \cdots \times m_d \times s}$



Remarks:

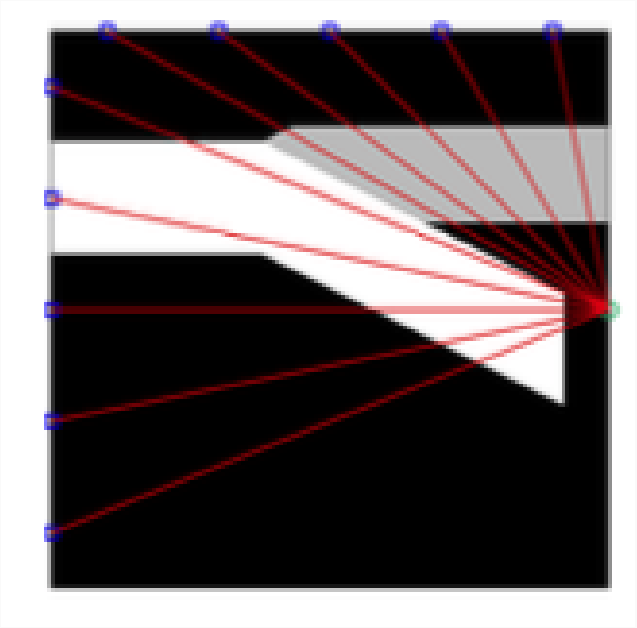
1. The sketching matrix $\boldsymbol{X}$ is a subspace embedding
2. We can apply established techniques to sketched tensor $\mathcal{Y}$
3. Can be implemented matrix-free and adjoint-free!

Seismic tomography: Problem Setup

Imaging technique that uses seismic waves

Problem settings:

- Sources: $s = 32$ sources on right boundary
- Receivers: $p = 45$ receivers on top and left boundary
- Grid size 64×64
- 2% observational error
- Whittle-Matérn prior



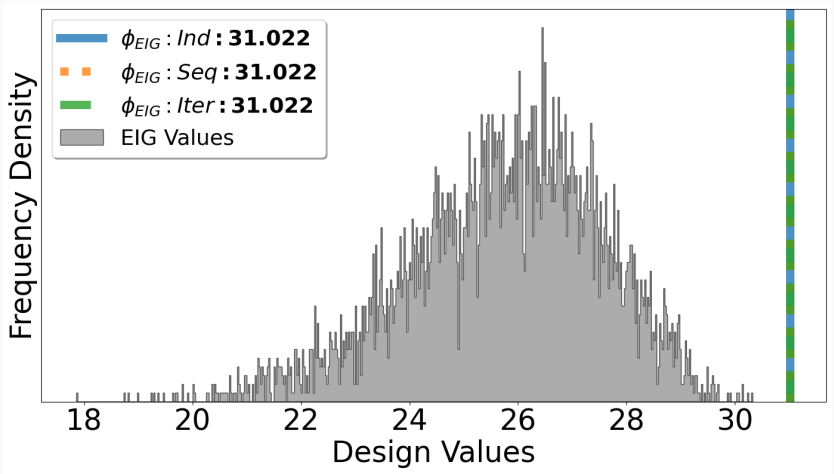
OED Goal: Pick the best $k_1 = 10$ sources and $k_2 = 10$ receivers.

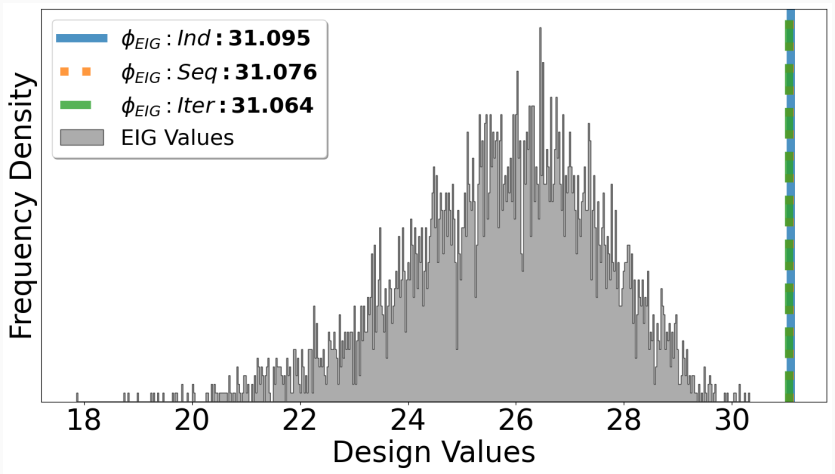
Metrics: D-optimal criterion

IR Tools: MATLAB Package. Gazzola et al. 2018.

Seismic tomography: Results

OED Goal: Pick the best $k_1 = 10$ (out of 32) sources and $k_2 = 10$ (out of 45) receivers.





All 12 methods: (three GKS-based + greedy) + randomized variants perform very similarly.

Timing results

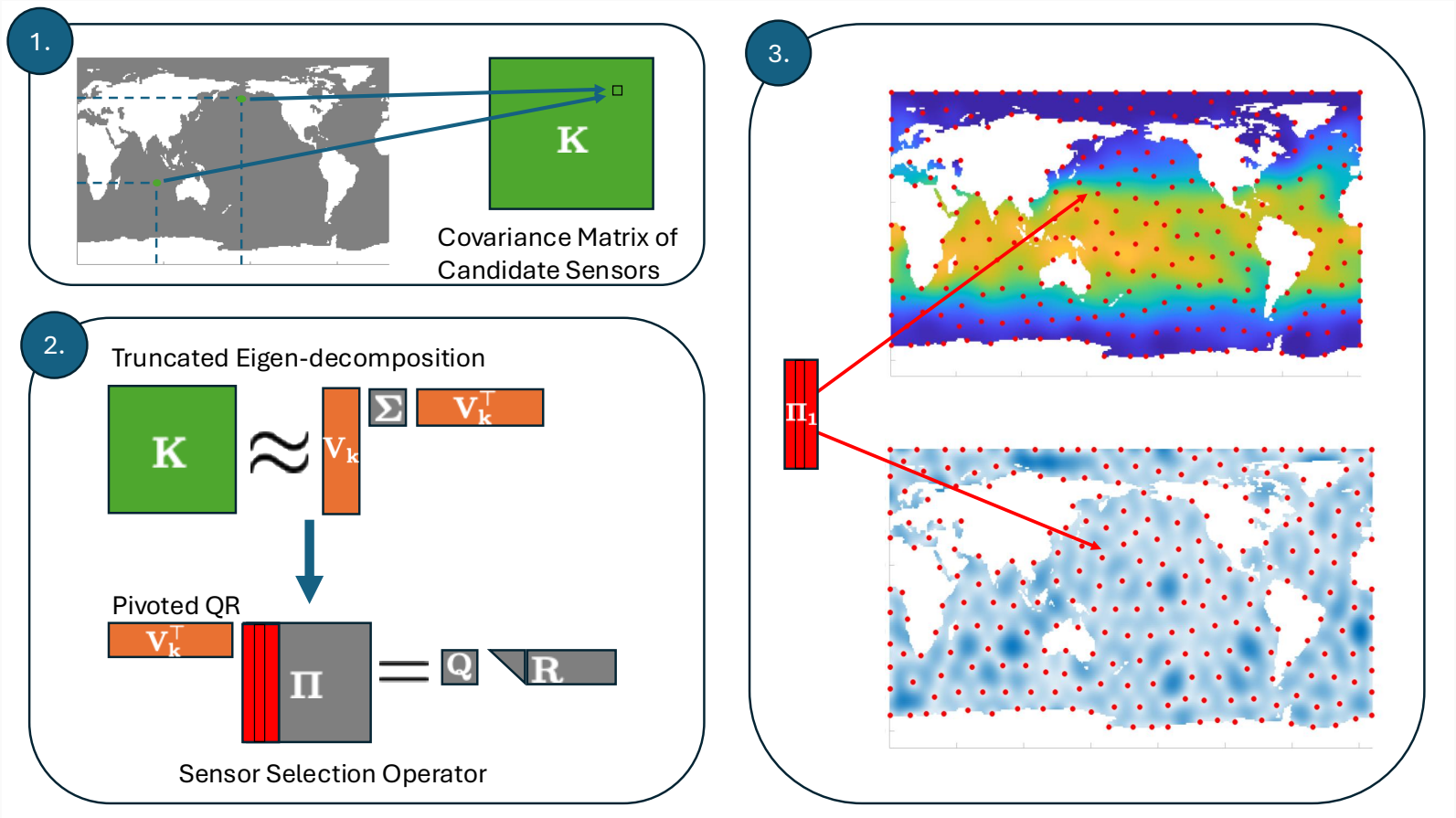
Method	GKS	GKS (Sketch)	Greedy	Greedy (Sketch)
IndSelect	8.63×10^{-1}	1.42×10^{-2}	18.45	0.58
SeqSelect	4.30×10^{-1}	6.68×10^{-3}	11.78	0.38
IterSelect	3.10×10^{-1}	1.08×10^{-2}	8.29	0.38

Runtimes (in seconds) for selection methods using GKS and Greedy approaches, with and without sketching.

Related Work

Sensor Placement in Gaussian Processes

Criterion $\phi_D(S) = \text{logdet}(\boldsymbol{I} + \eta^{-2}\boldsymbol{S}^\top \boldsymbol{K}\boldsymbol{S})$

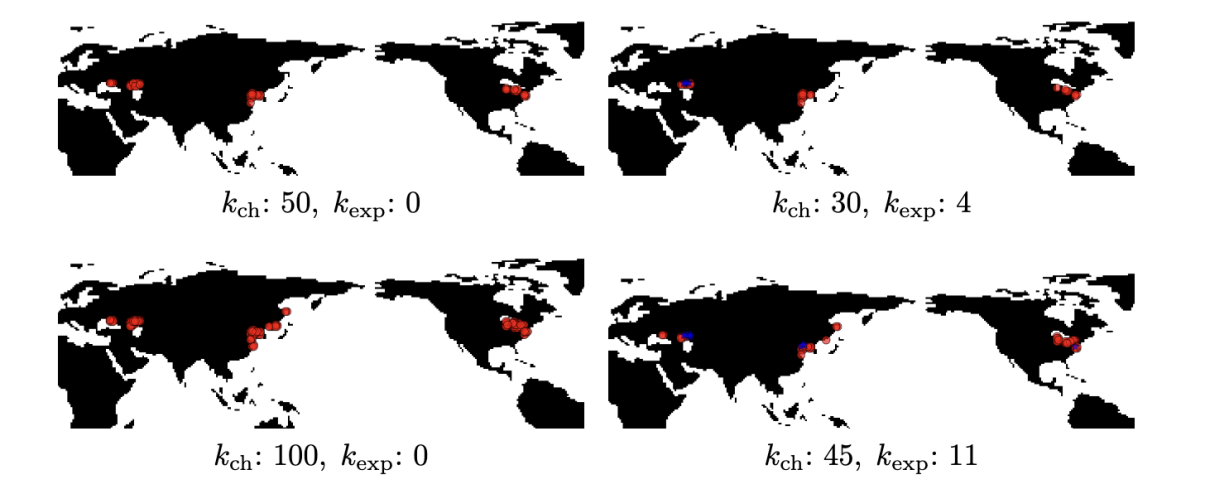


Chen, Ji, and **Saibaba**. "Optimal Sensor Placement in Gaussian Processes via Column Subset Selection." arXiv:2601.20781 (2026).

Multifidelity sensor placement

Given: Place two types of sensor within a fixed budget

- Cheap, but with low signal-to-noise ratio
- Expensive, but with high signal-to-noise ratio



Ramon, Sarnoski, Tumuluri, Díaz, **Saibaba** (2026). Multifidelity sensor placement in Bayesian state estimation problems. arXiv:2602.07269.

Other ongoing work

1. EIG for Nonlinear OED problems
    - Combination of sampling and linearization
    - Use ideas from linear OED as heuristics

2. Exchange algorithms for A-optimal/correlated noise/goal-oriented
    - Column subset selection are good initializations
    - Use exchange algorithms with efficient evaluation of the criteria to improve on the initializations

3. BasisSubset.jl: Julia package on column subset selection
    - CPU/GPU implementation of many algorithms
    - Beyond OED, applicable to clustering, model reduction, etc

Contributions

Contributions:

1. Reformulated sensor placement as column subset selection
2. Can utilize efficient NLA software implementations
3. Randomized algorithms make the algorithms computational efficient with guarantees
4. New structured column subset selection with tensor decompositions

Products:

1. Eswar, Rao, **Saibaba**, To appear, SIAM Journal on Scientific Computing, arXiv:5428327, 2026.
2. Díaz, **Saibaba**, Eswar, Rao, Di. arXiv preprint arXiv:2506.0033 (2025).
3. Software: https://github.com/RandomizedOED/css4oed.

Thank you!