Transport approaches for Bayesian experimental design across inference and prediction

Tiangang Cui[1], Sergey Dolgov[2], Roland Herzog[3], Karina Koval[3], Robert Scheichl[3]

[1]School of Mathematics and Statistics, University of Sydney, Australia

[2]Applied Mathematics, University of Bath, England

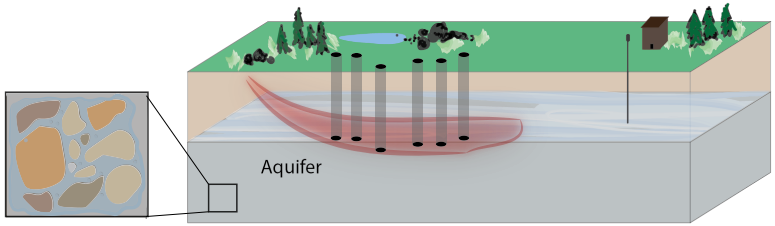
[3]Interdisciplinary Center for Scientific Computing, Heidelberg University, Germany

mODa14, Drongen Abbey, Belgium — June 16, 2026

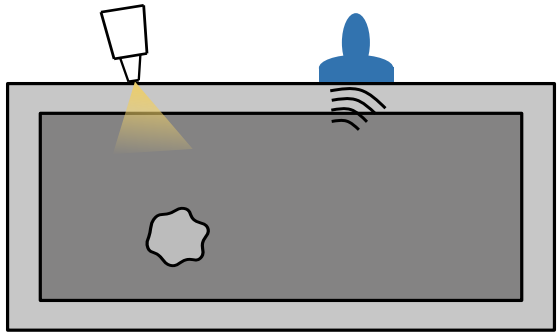
Financial support from the Carl Zeiss Foundation

Outline

Motivating examples



(sciencefriday.com)

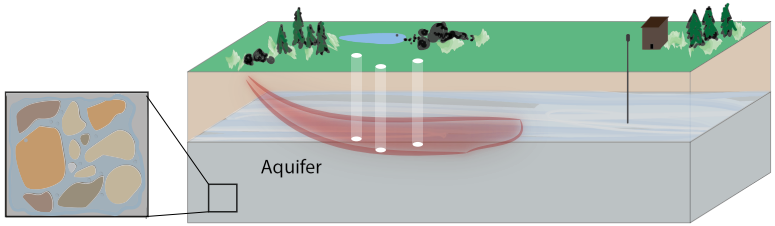


- Models for real-world problems involve unknown/unobservable inputs
- Accurate simulation/prediction requires knowledge of unknowns

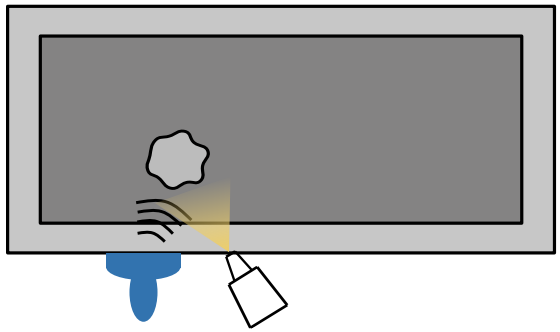
Inverse problem: Given model + observations, estimate unknown

- Data is noisy, sparse, expensive to obtain

Motivating examples



(sciencefriday.com)



- Models for real-world problems involve unknown/unobservable inputs
- Accurate simulation/prediction requires knowledge of unknowns

Inverse problem: Given model + observations, estimate unknown

- Data is noisy, sparse, expensive to obtain

Optimal experimental design (OED):

How to acquire data to *optimally* estimate the unknown?

e.g., where to drill wells, where to place sensors. . .

Bayesian inverse problems with design

$$\mathbf{d}(\mathbf{e}) = \mathcal{F}(\mathbf{m}, \mathbf{e}) + \boldsymbol{\eta}(\mathbf{e})$$

- $\mathbf{m} \in \mathbb{R}^n$: unknown parameters
- $\mathbf{e} \in \mathcal{E}$: design parameters
- $\mathcal{F}$: parameter-to-observable map
- $\boldsymbol{\eta}(\mathbf{e}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}(\mathbf{e}))$: noise
- $\mathbf{d}(\mathbf{e}) \in \mathbb{R}^d$: data

Bayesian approach: prior $\mathbf{m} \sim \pi(\mathbf{m})$, Bayes' rule $\rightsquigarrow$ design-dependent posterior

$$\pi(\mathbf{m}|\mathbf{e}, \mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\mathbf{m}, \mathbf{e})\, \pi(\mathbf{m}),$$

$$\mathcal{L}(\mathbf{d}|\mathbf{m}, \mathbf{e}) \propto \exp\left[-\tfrac{1}{2}\|\mathcal{F}(\mathbf{m}, \mathbf{e}) - \mathbf{d}\|^2_{\boldsymbol{\Gamma}(\mathbf{e})^{-1}}\right]$$

Bayesian optimal experimental design

Goal: choose design $\mathbf{e}$ to maximize information in data

$$\mathbf{e}^* = \arg\max_{\mathbf{e}\in\mathcal{E}} \Psi(\mathbf{e}), \qquad \Psi(\mathbf{e}) = \mathbb{E}_{\mathbf{d}|\mathbf{e}}[\mathcal{D}_{\mathsf{KL}}(\pi(\cdot|\mathbf{e},\mathbf{d})||\pi(\cdot))]$$

Ψ is a complicated function of intractable densities

1. $\pi(\mathbf{m}|\mathbf{e},\mathbf{d})$ intractable, high-dimensional
2. Many evaluations of Ψ required for optimization
3. PTO map $\mathcal{F}$ expensive (PDE solves)

This talk: Transport map approaches in two BOED settings:

1: Sequential design

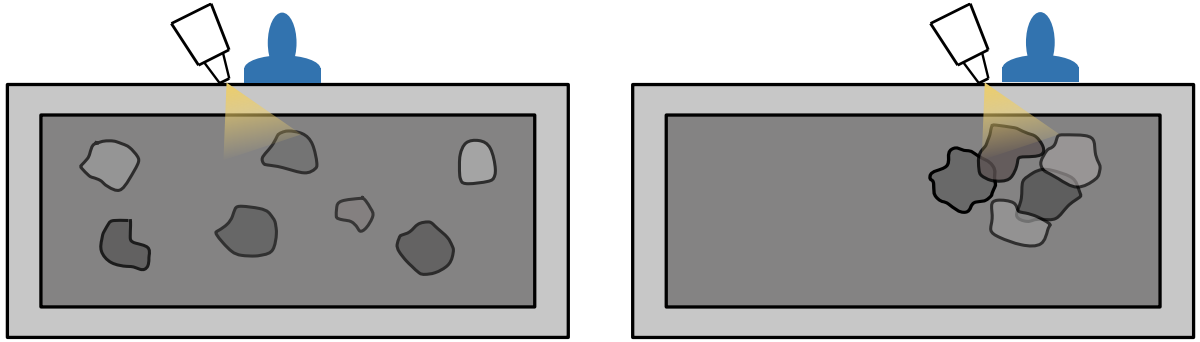
Infer $\mathbf{m}$ over K experiments
Access to density

2: Design for prediction

Infer a QoI $\mathbf{q} = \mathcal{G}(\mathbf{m})$.
Sample access only (density-free).

Outline

Greedy sequential design of experiments



$k = 1$ $k = 2$

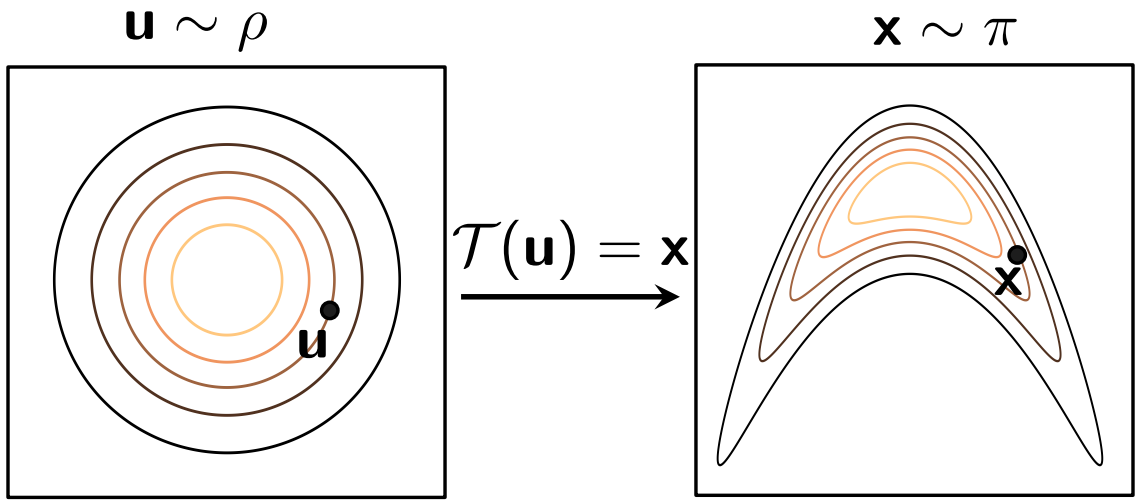
- Designs chosen in sequence (stages $k = 1, \ldots, K$)
- Previous results $\mathbf{H}_{k-1} = [\mathbf{e}_1, \mathbf{d}_1, \ldots, \mathbf{e}_{k-1}, \mathbf{d}_{k-1}]$ guide next choice
- Stage k prior: stage $k-1$ posterior $\pi(\mathbf{m}|\mathbf{H}_{k-1})$
- Greedy approach: maximize *incremental* EIG at each stage

$$\mathbf{e}_k^* = \underset{\mathbf{e}_k \in \mathcal{E}}{\arg\max}\ \Psi_k(\mathbf{e}_k), \quad \Psi_k(\mathbf{e}_k) = \mathbb{E}_{\mathbf{d}_k|\mathbf{e}_k,\mathbf{H}_{k-1}}[\mathcal{D}_{\mathrm{KL}}(\pi(\cdot|\mathbf{e}_k, \mathbf{d}_k, \mathbf{H}_{k-1})||\pi(\cdot|\mathbf{H}_{k-1}))]$$

Additional challenges in the sequential setting

- Both $\pi(\mathbf{m}|\mathbf{e}_k, \mathbf{d}_k, \mathbf{H}_{k-1})$ and $\pi(\mathbf{m}|\mathbf{H}_{k-1})$ intractable, high dimensional
- Ψ_k must be optimized quickly (designs often time-sensitive)

Transport maps



- $\mathcal{T}$ couples reference ρ (standard Gaussian) and target π
- **Key:** given $\mathcal{T}$, can evaluate and sample π cheaply

$$\mathbf{u} \sim \rho \;\Rightarrow\; \mathcal{T}(\mathbf{u}) = \mathbf{x} \sim \pi$$

$$\pi(\mathbf{x}) = (\mathcal{T})_\sharp \rho(\mathbf{x}) = \rho \circ \mathcal{T}^{-1}(\mathbf{x}) |\nabla \mathcal{T}^{-1}(\mathbf{x})|$$

$$\rho(\mathbf{u}) = (\mathcal{T})^\sharp \pi(\mathbf{u}) = (\pi \circ \mathcal{T})(\mathbf{u}) |\nabla \mathcal{T}(\mathbf{u})|$$

sOED via iEIG surrogate

Let $\mathcal{T}^{k-1}$ s.t. $(\mathcal{T}^{k-1})^\sharp \pi(\mathbf{u}|\mathbf{H}_{k-1}) = \rho(\mathbf{u})$

Using properties of the KL divergence and results from Li et al.*:

$$\Psi_k(\mathbf{e}_k) \;\leq\; \Psi_k^{\mathrm{UB}}(\mathbf{e}_k) := \tfrac{1}{2}\log\det\big(\mathbf{I} + H(\mathbf{e}_k, \mathcal{T}^{k-1})\big)$$

$$H(\mathbf{e}_k, \mathcal{T}^{k-1}) = \mathbb{E}_\rho\big[\nabla\mathcal{T}^{k-1}(\mathbf{u})^\top \nabla\mathcal{F}(\mathbf{m}, \mathbf{e}_k)^* \boldsymbol{\Gamma}(\mathbf{e}_k)^{-1} \nabla\mathcal{F}(\mathbf{m}, \mathbf{e}_k)\nabla\mathcal{T}^{k-1}(\mathbf{u})\big]$$

- Ψ_k^{UB} easier to evaluate using Monte Carlo than Ψ_k
- Use Ψ_k^{UB} to guide design selection

Key question: How to construct sequence of maps $\mathcal{T}^1, \ldots$ efficiently?

*Li et al., 2024, Thm. 9.

Amortised inference via conditional transport

Knothe–Rosenblatt (KR) rearrangement: unique lower-triangular map $\mathcal{S} = \mathcal{T}^{-1}$

$$\mathcal{S}(\mathbf{x}) = \begin{bmatrix} \mathcal{S}_{x_1}(x_1) \\ \mathcal{S}_{x_2|x_1}(x_1, x_2) \\ \vdots \\ \mathcal{S}_{x_n|x_{1:n-1}}(x_1, \ldots, x_n) \end{bmatrix} = \mathbf{u}, \qquad \mathbf{x} \sim \pi \;\Rightarrow\; \mathbf{u} \sim \rho$$

- Components defined via conditional CDFs of $\pi \;\rightsquigarrow$ triangular structure *exposes conditionals* $\pi(x_k|\mathbf{x}_{1:k-1})$

Amortised inference via conditional transport

Knothe–Rosenblatt (KR) rearrangement: unique lower-triangular map $\mathcal{S} = \mathcal{T}^{-1}$

$$\mathcal{S}(\mathbf{x}) = \begin{bmatrix} \mathcal{S}_{x_1}(x_1) \\ \mathcal{S}_{x_2|x_1}(x_1, x_2) \\ \vdots \\ \mathcal{S}_{x_n|x_{1:n-1}}(x_1, \ldots, x_n) \end{bmatrix} = \mathbf{u}, \qquad \mathbf{x} \sim \pi \;\Rightarrow\; \mathbf{u} \sim \rho$$

- Components defined via conditional CDFs of $\pi \;\rightsquigarrow$ triangular structure *exposes conditionals* $\pi(x_k|\mathbf{x}_{1:k-1})$

In our setting: Build KR map over data and parameters *while experiment k is being conducted*,

$$\mathcal{T}^k_{\mathbf{d},\mathbf{m}}(\mathbf{u}) = \begin{bmatrix} \mathcal{T}^k_{\mathbf{d}}(\mathbf{u}_d) \\ \mathcal{T}^k_{\mathbf{m}|\mathbf{d}}(\mathbf{u}_d, \mathbf{u}_m) \end{bmatrix} \quad \text{s.t.} \quad \left(\mathcal{T}^k_{\mathbf{d},\mathbf{m}}\right)_\sharp \rho \approx \pi(\mathbf{d}_k, \mathbf{m}|\mathbf{e}^*_k, \mathbf{H}_{k-1})$$

Once data $\mathbf{d}_k$ collected, **immediately restrict**:

$$\mathcal{T}^k(\cdot) \;=\; \mathcal{T}^k_{\mathbf{m}|\mathbf{d}}\left((\mathcal{T}^k_{\mathbf{d}})^{-1}(\mathbf{d}_k), \cdot\right) \quad \Rightarrow \quad (\mathcal{T}^k)_\sharp \rho \approx \pi(\mathbf{m}|\mathbf{H}_k)$$

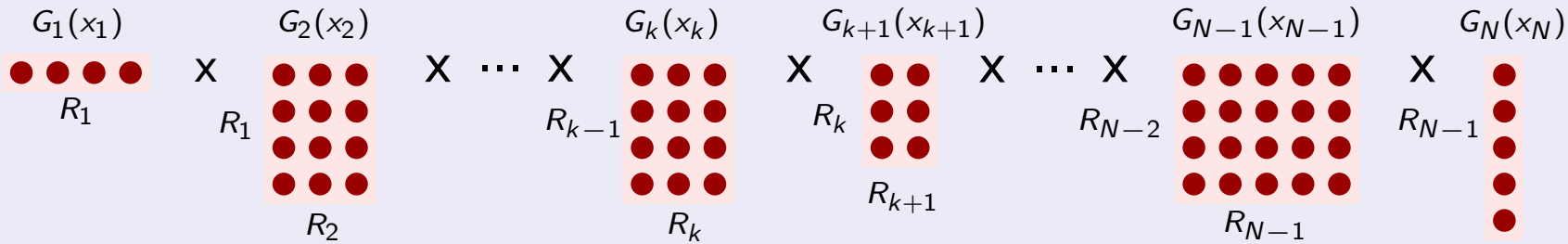
Bulk of computation done while waiting for data

Constructing $\mathcal{T}$ via tensor train (TT) decomposition†

KR components require evaluating CDFs of $\pi(x_k|\mathbf{x}_{1:k-1})$

Approximate π via TT decomposition

$$\sqrt{\pi(\mathbf{x})} \approx g(\mathbf{x}) = \prod_{i=1}^{N} \mathbf{G}_i(x_i), \qquad \pi \approx p \propto g^2$$



- Separable form enables evaluation of CDFs $\rightsquigarrow \mathcal{T} = \mathcal{S}^{-1}$
- Built using *cross interpolation* methods*
  - Requires evaluations of π (PDE solves)
  - TT ranks R control accuracy: $\mathcal{D}_H(\pi, p) \leq \epsilon$
  - **Cost: $\mathcal{O}(N\,R^2)$ PDE solves**, $N = n + d$ — expensive for large n and R

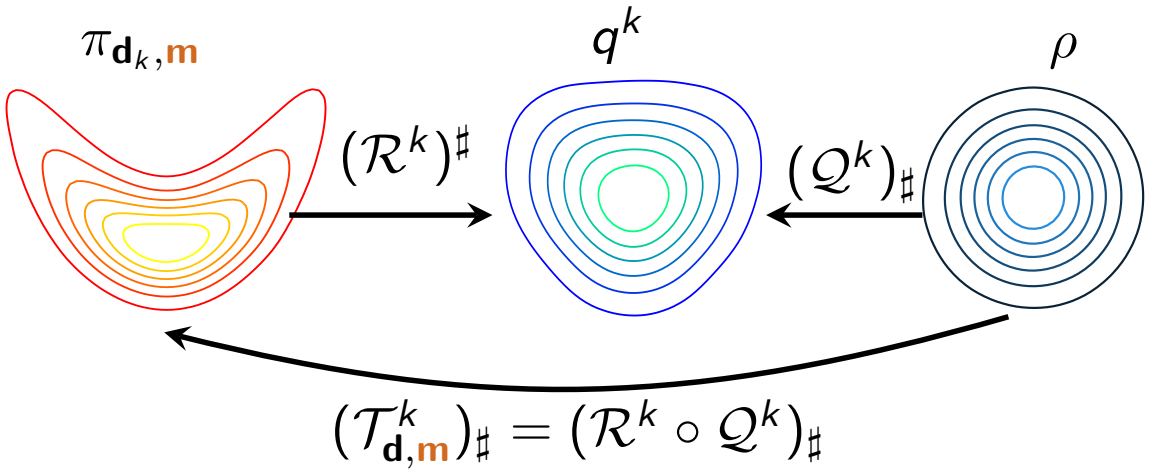
*Oseledets and Tyrtyshnikov, 2010.

†Cui, Dolgov, and Zahm, 2023.

Scalable construction of $\mathcal{T}$

Challenge: Building $\mathcal{T}^k_{\mathbf{d},\mathbf{m}}$ requires $\mathcal{O}(N\,R^2)$ PDE solves, expensive for large N, R

(i) Preconditioning: $\mathcal{T}^k_{\mathbf{d},\mathbf{m}} = \mathcal{R}^{k-1}_{\mathbf{d},\mathbf{m}} \circ \mathcal{Q}^k_{\mathbf{d},\mathbf{m}}$



- Consecutive posteriors are close $\rightsquigarrow$ use $\mathcal{T}^{k-1}$ as preconditioner $\mathcal{R}^{k-1}$
- $\mathcal{Q}^k$ corrects a small residual $\rightsquigarrow$ **much lower TT ranks (R) needed**

(ii) Likelihood-informed subspace (LIS):*

- q^k differs from ρ along only $r \ll n$ directions $\implies$ build $\mathcal{Q}^k$ in $\mathbb{R}^{d+r}$
- Dimension reduction via SVD of $H(\mathbf{e}^*_k, \mathcal{T}^{k-1})$
- **Significantly lower dimension** $N_r \ll N$

*Cui and Zahm, 2021; Li et al., 2024.

Numerical example — Photoacoustic imaging

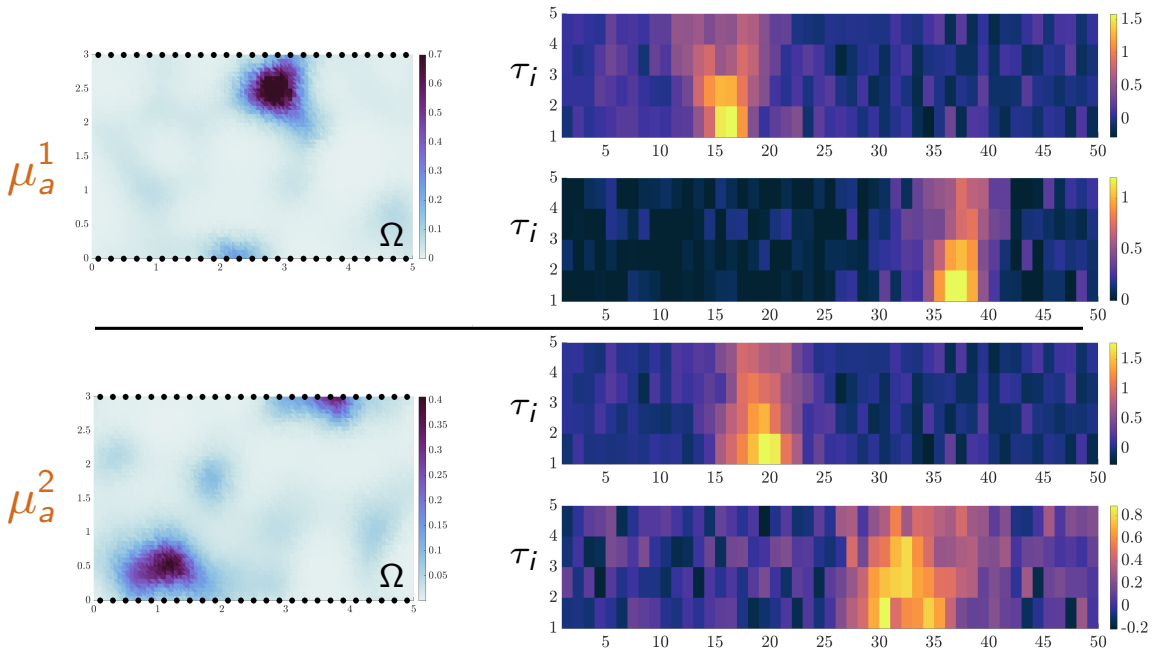
Optic PDE:

$$\mu_a \phi - \nabla \cdot (\kappa(\mu_a) \nabla \phi) = 0 \quad \text{in } \Omega,$$
$$\phi/\pi + \kappa(\mu_a) \nabla \phi \cdot n/2 = S(e_1) \quad \text{on } \Gamma,$$

Acoustic PDE:

$$p_{tt}/c^2 - \Delta p = 0 \quad \text{in } \Omega \times (0, T],$$
$$p(t=0) = p_0(e_1, \mu_a) \quad \text{in } \Omega,$$
$$p_t(t=0) = 0 \quad \text{in } \Omega,$$
$$\nabla p \cdot \mathbf{n} - p_t/c = 0 \quad \text{on } \Gamma \times (0, T]$$

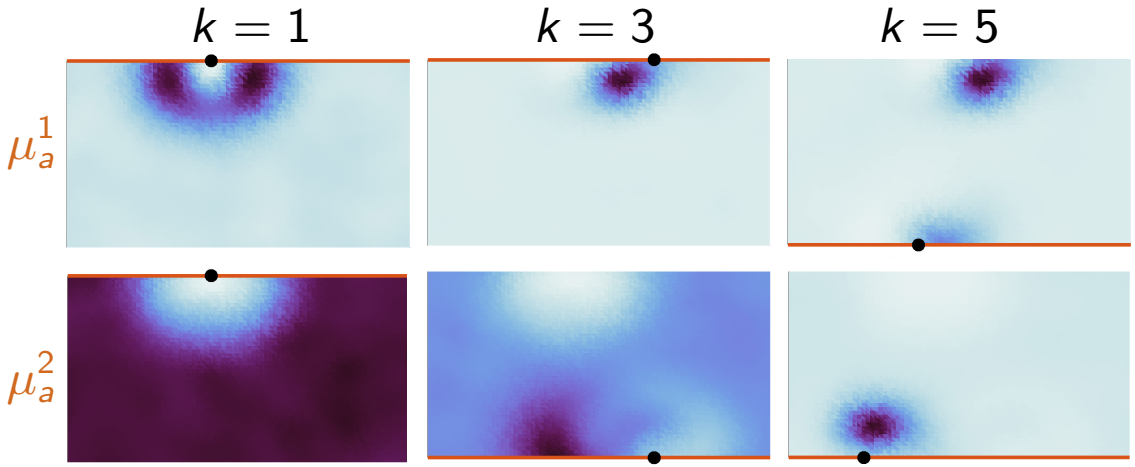
- e_1: location of light source (top or bottom)
- e_2 chooses one sensor, $\mathbf{d}(\mathbf{e}) \in \mathbb{R}^5$
- $\mu_a = \exp(\mathbf{m})$, $\mathbb{R}^{3969} \ni \mathbf{m} \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_{mm})$



Goal: Choose location e_2 and illumination source e_1 for optimal inference of μ_a

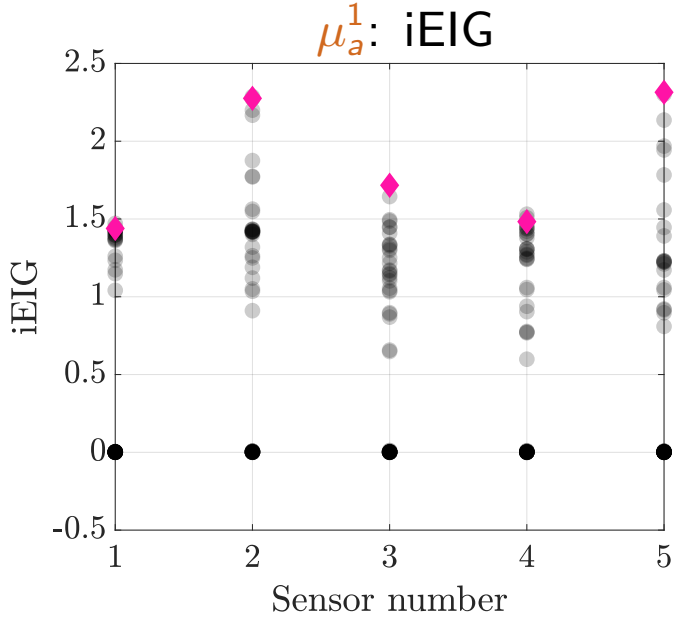
sOED results

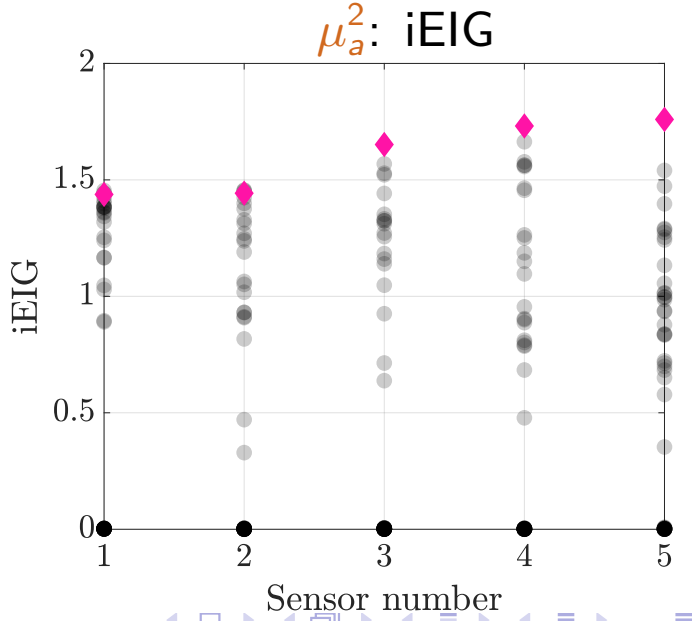
Posterior mean comparison:



- $N = 500$ samples used to approximate Ψ_k^{UB}
- Dimension of $\mathbf{m}$, $n = 3969$
- LIS size $r \leq 50$ for all k

Evaluation of designs:





Outline

What changes when the goal is prediction?

So far: Sequential inference of $\mathbf{m}$ with density-based KR/TT maps.

- Often, estimating $\mathbf{m}$ is not the ultimate goal
- Inferring $\mathbf{m}$ is a step toward predicting a **quantity of interest (QoI)**

$$\mathbf{q} = \mathcal{G}(\mathbf{m})$$

Two key differences:

1. **Objective.** Maximize information about $\mathbf{q}$, not $\mathbf{m}$:

$$\Psi(\mathbf{e}) = \mathbb{E}_{\mathbf{d}|\mathbf{e}}\left[\mathcal{D}_{\mathrm{KL}}(\pi_{\mathbf{q}|\mathbf{d},\mathbf{e}} || \pi_{\mathbf{q}})\right]$$

2. **Access.** Predictive densities $\pi_{\mathbf{q}}, \pi_{\mathbf{q}|\mathbf{d},\mathbf{e}}, \pi_{\mathbf{d}|\mathbf{q},\mathbf{e}}$ intractable
   $\implies$ iEIG bound & previous map construction not feasible

What we do have: Sampling from the joint $\pi_{\mathbf{q},\mathbf{d}|\mathbf{e}}$ is easy

$$\mathbf{m}^i \sim \pi_{\mathbf{m}},\ \mathbf{d}^i \sim \mathcal{L}(\cdot|\mathbf{m}^i, \mathbf{e}),\ \mathbf{q}^i = \mathcal{G}(\mathbf{m}^i)$$

$\rightsquigarrow$ Need a density-free way to estimate and optimize Ψ

Estimating the predictive EIG via transport maps

Rewrite the predictive EIG as an entropy difference:

$$\Psi(\mathbf{e}) = h(\mathbf{d}|\mathbf{e}) - \mathbb{E}_{\mathbf{q}}[h(\mathbf{d}|\mathbf{q},\mathbf{e})], \quad h(\mathbf{x}) = -\mathbb{E}_{\mathbf{x}}[log\pi(\mathbf{x})]$$

How to estimate $h(\mathbf{d}|\mathbf{e})$, $h(\mathbf{d}|\mathbf{q},\mathbf{e})$ from samples $\{(\mathbf{q}^i,\mathbf{d}^i)\}_{i=1}^N \sim \pi_{\mathbf{q},\mathbf{d}|\mathbf{e}}$?
Given $(\mathcal{S}^{\mathbf{d}|\mathbf{e}})_\sharp\rho \approx \pi_{\mathbf{d}|\mathbf{e}}$, $(\mathcal{S}^{\mathbf{d}|\mathbf{q},\mathbf{e}}(\mathbf{q},\cdot))_\sharp\rho \approx \pi_{\mathbf{d}|\mathbf{q},\mathbf{e}}$, this is easy:

$$h(\mathbf{d}|\mathbf{e}) = -\mathbb{E}_{\mathbf{d}|\mathbf{e}}\left[\log(\pi_{\mathbf{d}|\mathbf{e}})\right] \approx -\frac{1}{N}\sum_{i=1}^{N}\log\left((\mathcal{S}^{\mathbf{d}|\mathbf{e}})_\sharp\rho(\mathbf{d}^i)\right) =: L(\mathcal{S}^{\mathbf{d}|\mathbf{e}})$$

$$\mathbb{E}_{\mathbf{q}}[h(\mathbf{d}|\mathbf{q},\mathbf{e})] = -\mathbb{E}_{\mathbf{q},\mathbf{d}|\mathbf{e}}\left[\log(\pi_{\mathbf{d}|\mathbf{q},\mathbf{e}})\right] \approx L(\mathcal{S}^{\mathbf{d}|\mathbf{q},\mathbf{e}})$$

$$\widehat{\Psi}_N(\mathbf{e}) = L(\mathcal{S}^{\mathbf{d}|\mathbf{e}}) - L(\mathcal{S}^{\mathbf{d}|\mathbf{q},\mathbf{e}})$$

$L(\mathcal{S})$ **is exactly what's used to train transport maps from samples!**

Properties of the loss-based EIG estimator

Error bound: Estimator error controlled by map approximation quality,

$$-\mathcal{D}_{\mathrm{KL}}(\pi_{\mathbf{d}|\mathbf{e}}||(\mathcal{S}^{\mathbf{d}|\mathbf{e}})_\sharp\rho) \;\leq\; \Psi(\mathbf{e}) - \widehat{\Psi}(\mathbf{e}) \;\leq\; \mathbb{E}_{\mathbf{q}}\left[\mathcal{D}_{\mathrm{KL}}(\pi_{\mathbf{d}|\mathbf{q},\mathbf{e}}||(\mathcal{S}^{\mathbf{d}|\mathbf{q},\mathbf{e}})_\sharp\rho)\right]$$

$$\widehat{\Psi}(\mathbf{e}) = \underbrace{-\mathbb{E}_{\mathbf{d}|\mathbf{e}}\left[\log((\mathcal{S}^{\mathbf{d}|\mathbf{e}})_\sharp\rho)\right]}_{\widehat{h}(\mathbf{d}|\mathbf{e})} + \underbrace{\mathbb{E}_{\mathbf{d},\mathbf{q}|\mathbf{e}}\left[\log((\mathcal{S}^{\mathbf{d}|\mathbf{q},\mathbf{e}})_\sharp\rho)\right]}_{-\mathbb{E}_{\mathbf{q}}\left[\widehat{h}(\mathbf{d}|\mathbf{q},\mathbf{e})\right]}$$

- Estimator is neither an upper nor lower bound
- Improve the maps $\Rightarrow$ improve the estimate

Map parametrization:

- We use a **TT-based parametrization** for both maps
- Richer parametrization $\Rightarrow$ better map approximation $\Rightarrow$ tighter EIG estimate

EIG surrogate over the design space

Treat design $\mathbf{e}$ as an additional input to the maps. A *single* training run $\rightsquigarrow$ surrogate valid over the *entire* design space:

1. Samples $(\mathbf{e}^i, \mathbf{q}^i, \mathbf{d}^i) \sim \pi_{\mathbf{e},\mathbf{q},\mathbf{d}}$

$$(\mathbf{e}^i, \mathbf{m}^i) \sim \pi_{\mathbf{e}}\pi_{\mathbf{m}},\ \mathbf{d}^i \sim \mathcal{L}(\cdot|\mathbf{m}^i, \mathbf{e}^i),\ \mathbf{q}^i = \mathcal{G}(\mathbf{m}^i)$$

2. Train $\mathcal{S}^{\mathbf{d}|\mathbf{e}}(\mathbf{e}, \mathbf{d})$ and $\mathcal{S}^{\mathbf{d}|\mathbf{q},\mathbf{e}}(\mathbf{e}, \mathbf{q}, \mathbf{d})$ once on this joint dataset

Evaluating $\widehat{\Psi}_N(\mathbf{e})$ at any new design:

- Generate data at $\mathbf{e}$ using the trained map:

$$\mathbf{z}^i \sim \rho \quad \rightsquigarrow \quad \mathbf{d}^i(\mathbf{e}) = \mathcal{S}^{\mathbf{d}|\mathbf{q},\mathbf{e}}(\mathbf{e}, \mathbf{q}^i, \mathbf{z}^i) \quad \rightsquigarrow \quad (\mathbf{q}^i, \mathbf{d}^i(\mathbf{e})) \sim \pi_{\mathbf{q},\mathbf{d}|\mathbf{e}}$$

- Evaluate $\widehat{\Psi}_N(\mathbf{e}) = L(\mathcal{S}^{\mathbf{d}|\mathbf{e}}) - L(\mathcal{S}^{\mathbf{d}|\mathbf{q},\mathbf{e}})$ on samples $\{(\mathbf{e}, \mathbf{q}^i, \mathbf{d}^i(\mathbf{e}))\}$

Example 1: scalar nonlinear model

Scalar model*

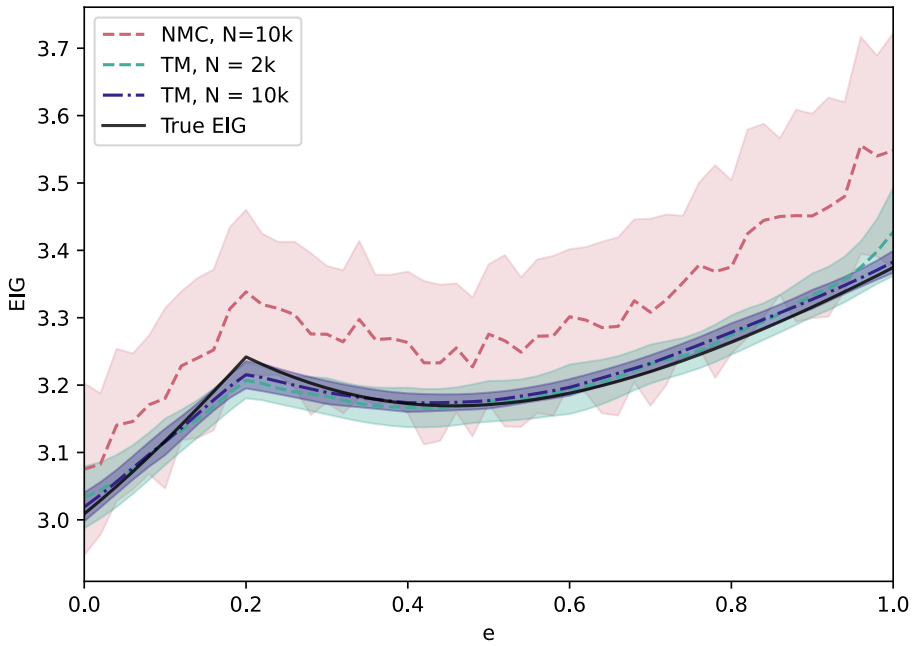
$$d = e^2 m^3 + m \exp(-|0.2 - e|) + \eta, \quad m \sim \mathcal{U}([0,1]), \quad \eta \sim \mathcal{N}(0, 10^{-4})$$

QoI is the unknown m itself; design $e \in [0,1]$ continuous.

Compared with NMC estimator

$$\Psi_{\mathrm{NMC}}(\mathbf{e}) = \frac{1}{N_{\mathrm{out}}} \sum_{i=1}^{N_{\mathrm{out}}} \left[\log \mathcal{L}(d^i | e, m^i) - \log \left(\sum_{j=1}^{N_{\mathrm{in}}} \mathcal{L}(d^i | e, m^{i,j}) \right) \right]$$

- Ψ_N built with $N = 2000, 10000$
- Ψ_{NMC} built with $N_{\mathrm{in}} N_{\mathrm{out}} = 10000$ per $\mathbf{e}$
- 50 runs of each



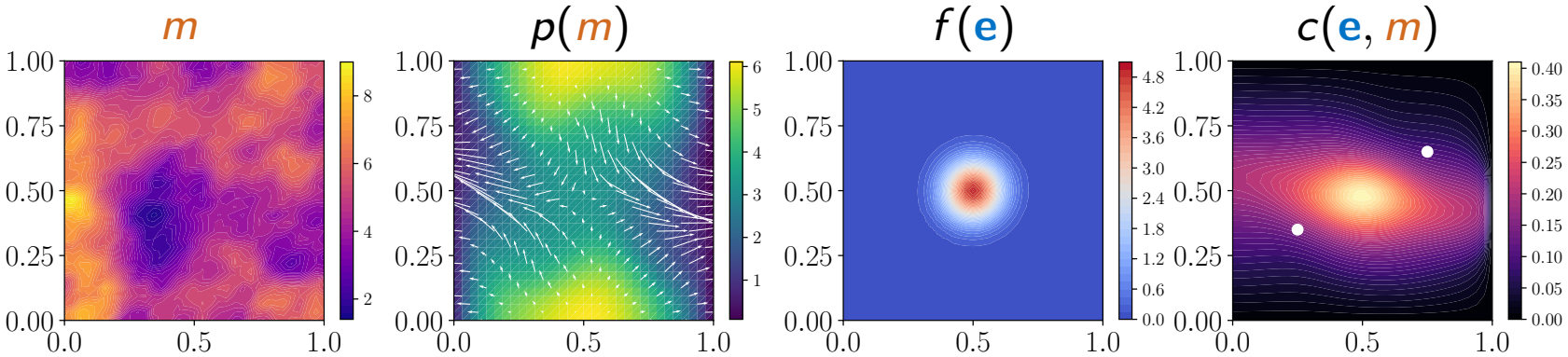
*Huan and Marzouk, 2013.

Example 2: PDE-based problem

How to inject a tracer to best predict the discharge out of boundary q?*

$$-D\Delta c + \nabla \cdot (\mathbf{v}(m)c) = f(\mathbf{e})$$
$$\mathbf{v}(m) = -\kappa \nabla p(m)$$
$$-\nabla \cdot (\kappa \nabla p) = m$$

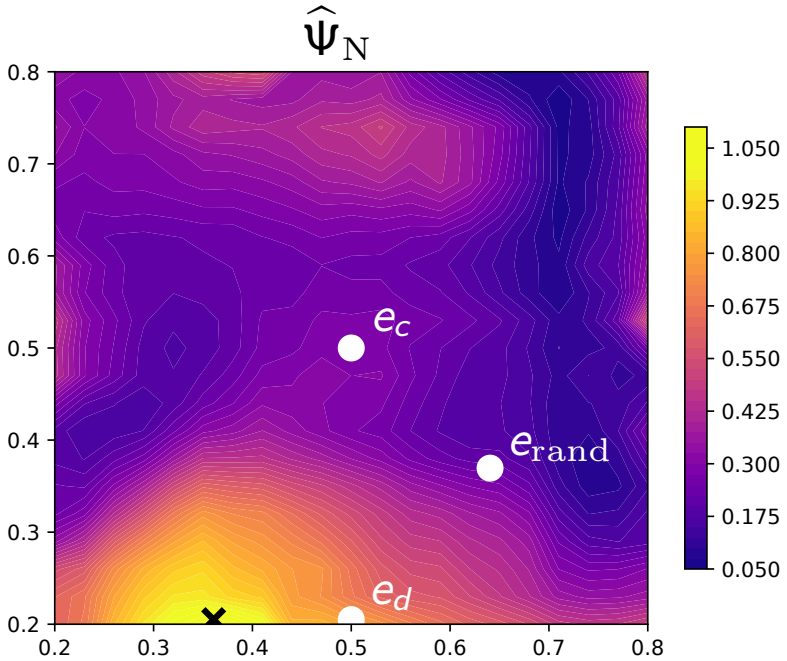
- m: source term field (high-dim.)
- $q = \int_{\Gamma_R} \mathbf{v}(m) \cdot \mathbf{n} \, \mathrm{d}S$
- $\mathbf{e} \in [0.2, 0.8]^2$: source location

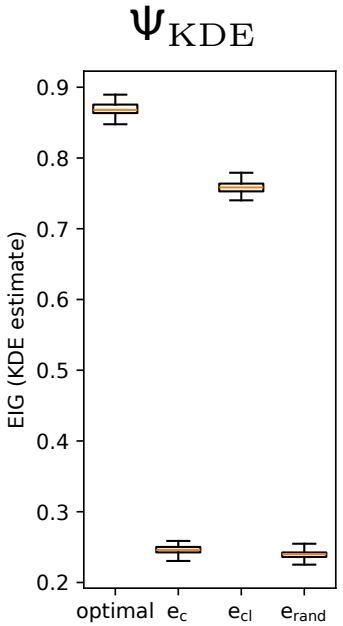


*Neuberger, Alexanderian, and Bloemen Waanders, 2025.

Example 2: PDE-based problem EIG estimate

- $N = 10000$ samples used to build $\widehat{\Psi}_{\mathrm{N}}(\mathbf{e})$ valid for all $\mathbf{e}$
- Sanity check using KDE-based estimator using 10000 samples per $\mathbf{e}$





Outline

Summary

Transport maps are a powerful, flexible tool for OED — key ideas used across very different design problems:

- **A cheap EIG surrogate:** map losses or upper bound replaces expensive nested-loop EIG estimation
- **Amortized inference:** conditional (triangular) transport grants access to any conditional through single map

Problem structure dictates how the maps are built:

- **What can you access?**
  - Density evaluations $\rightsquigarrow$ TT decomposition + cross interpolation *(sOED)*
  - Samples only $\rightsquigarrow$ minimize an empirical loss *(predictive OED)*
- **High dimensions** $\rightsquigarrow$ dimension reduction (LIS / low-rank) for scalability
- **Sequential design** $\rightsquigarrow$ preconditioning with the previous-stage map

Current work: sequential design for prediction, combining both approaches

References

Thank you! Preprint (sOED): arXiv:2502.20086, appearing in JUQ soon


Cui, T., S. Dolgov, and O. Zahm (2023). "Scalable conditional deep inverse Rosenblatt transports using tensor trains and gradient-based dimension reduction". In: *Journal of Computational Physics* 485, p. 112103.


Cui, T. and O. Zahm (2021). "Data-free likelihood-informed dimension reduction of Bayesian inverse problems". In: *Inverse Problems* 37.4, p. 045009.


Huan, X. and Y. M. Marzouk (2013). "Simulation-based optimal Bayesian experimental design for nonlinear systems". In: *Journal of Computational Physics* 232.1, pp. 288–317.


Li, M. T., T. Cui, F. Li, Y. Marzouk, and O. Zahm (2024). "Sharp detection of low-dimensional structure in probability measures via dimensional logarithmic Sobolev inequalities". In: *arXiv preprint arXiv:2406.13036*.


Neuberger, J. N., A. Alexanderian, and B. van Bloemen Waanders (Oct. 2025). "Goal Oriented Optimal Design of Infinite-Dimensional Bayesian Inverse Problems using Quadratic Approximations". In: *Journal of Scientific Computing* 105.2.


Oseledets, I. and E. Tyrtyshnikov (2010). "TT-cross approximation for multidimensional arrays". In: *Linear Algebra and its Applications* 432.1, pp. 70–88.

Sequential Bayesian inference with design — details

Stage k inverse problem:

$$\mathbf{d}_k(\mathbf{e}_k) = \mathcal{F}(\mathbf{m}, \mathbf{e}_k) + \boldsymbol{\eta}_k(\mathbf{e}_k)$$

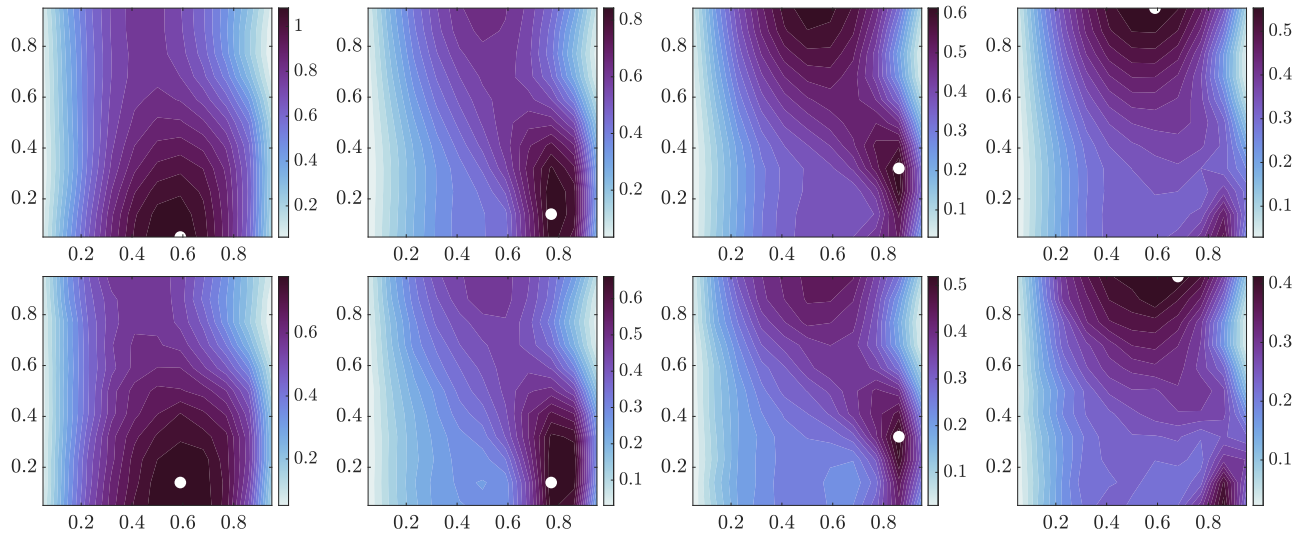
- Current knowledge: $\pi(\mathbf{m}|\mathbf{H}_{k-1})$, $\mathbf{H}_{k-1} = [\mathbf{e}_1, \mathbf{d}_1, \ldots, \mathbf{e}_{k-1}, \mathbf{d}_{k-1}]$
- Bayes' rule $\rightsquigarrow$ update:

$$\pi(\mathbf{m}|\mathbf{e}_k, \mathbf{d}_k, \mathbf{H}_{k-1}) \propto \mathcal{L}(\mathbf{d}_k|\mathbf{m}, \mathbf{e}_k)\, \pi(\mathbf{m}|\mathbf{H}_{k-1})$$

iEIG upper bound — accuracy

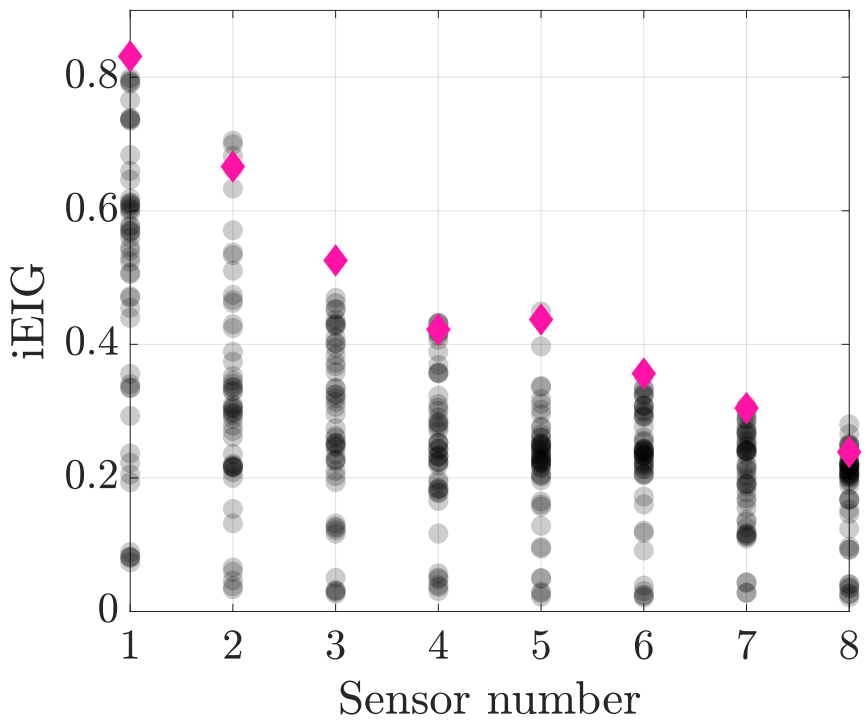
Ψ_k^{UB}

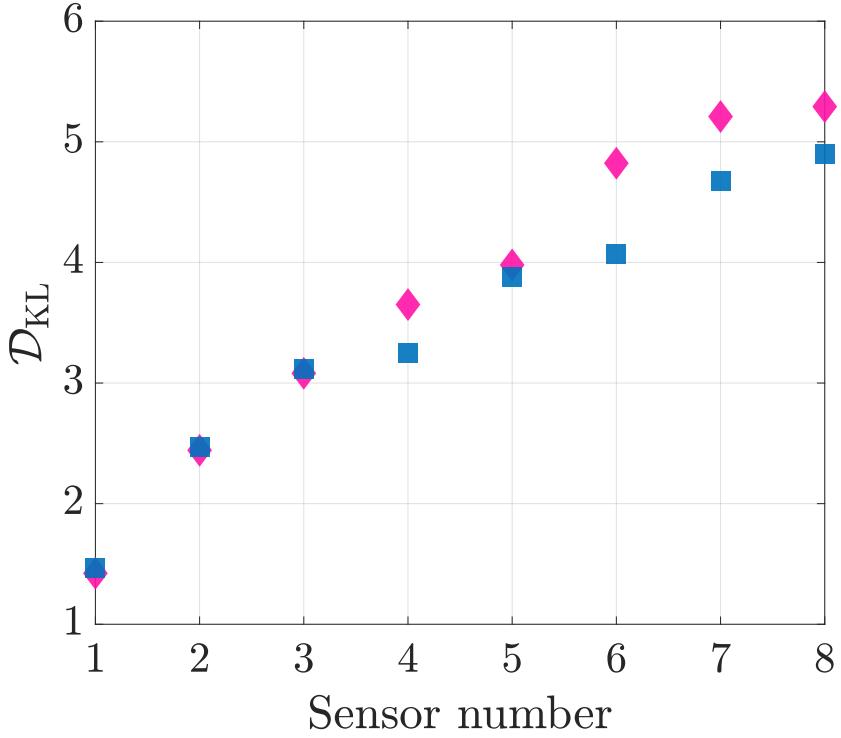
Ψ_k^{NMC}



- 100 samples for H; 10 000 samples for nested Monte Carlo (NMC)

Further sOED comparisons





- Blue squares: Gaussian approximation to each posterior

TT approximation, some details

Given arbitrary $g\colon \mathcal{X} \to \mathbb{R}$, with $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_n$,

$$g(\mathbf{x}) \approx \widehat{g}(\mathbf{x}) := \mathbf{G}_1(x_i) \cdots \mathbf{G}_k(x_k) \cdots \mathbf{G}_n(x_n), \tag{1}$$

TT cores: $\mathbf{G}_k\colon \mathcal{X}_k \to \mathbb{R}^{r_{k-1} \times r_k}$ represented using m_k basis functions $\{\phi_k^{(\ell)}\}_{\ell=1}^{m_k}$.

$$[\mathbf{G}_k(x_k)]_{i,j} = \sum_{\ell=1}^{m_k} \phi_k^{(\ell)}(x_k)\mathbf{A}_k[i,\ell,j], \quad \mathbf{A}_k \in \mathbb{R}^{r_{k-1} \times m_k \times r_k}.$$

- Coefficients $\mathbf{A}_k$ are chosen via interpolation to control L_2 approximation error: $\|g - \widehat{g}\|_{L_2}$
- Given collocation points $\{x_k^{(\ell)}\}_{\ell=1}^{m_k}$, *TT cross* iterates over cores and computes $\mathbf{A}_k$ by ensuring:

$$g(\mathbf{x}^{(j)}) = \widehat{g}(\mathbf{x}^{(j)}) \quad \forall \mathbf{x}^{(j)} \in X_{<k>}$$

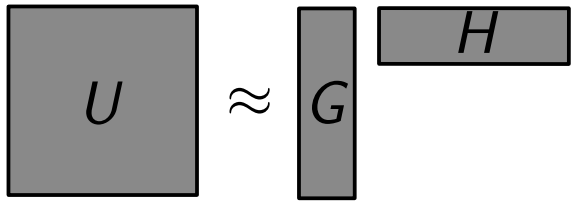
for a set $X_{<k>}$ of cardinality $r_{k-1} m_k r_k$.

TT approximation, some details

- Procedure for choosing sets $X_{<k>}$ very technical, but uses ideas similar to cross approximation for matrices...

2D example: Consider $u(x_1, x_2)$ evaluated on a grid of values collected into matrix

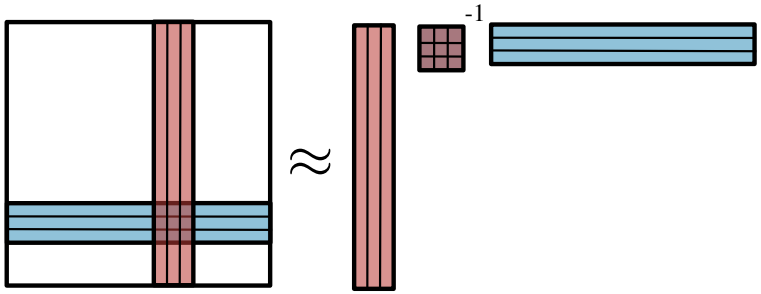
$$U \approx \sum_{\alpha=1}^{r} G_{i,\alpha} H_{\alpha,j}$$



Cross interpolation methods find G and H without computing SVD

- Instead, find best interpolation indices $\mathcal{I}, \mathcal{J}$

$$U \approx U(:, \mathcal{J}) U^{-1}(\mathcal{I}, \mathcal{J}) U(\mathcal{I}, :)$$

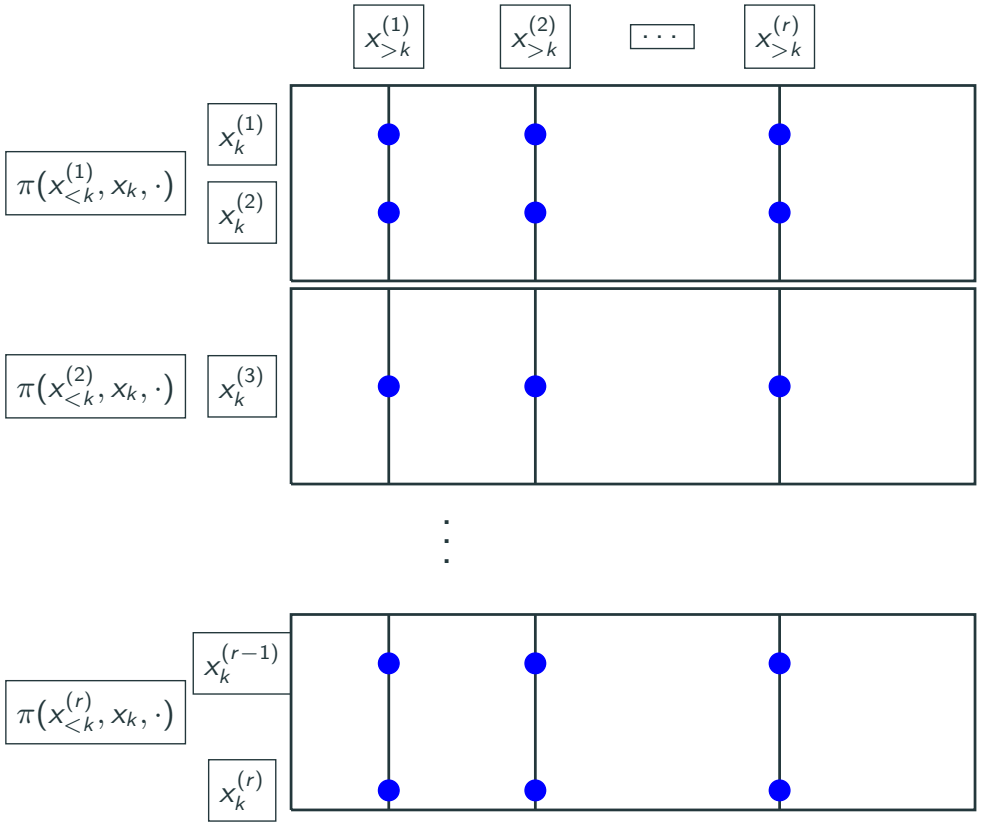


- Requires maximizing $|\det(U(\mathcal{I}, \mathcal{J}))|$ - NP hard!
- Instead alternate between finding $\mathcal{I}$ and $\mathcal{J}$ - approximates optimal indices

Functional TT cross

Grouped coordinates: $x_{<k} = (x_1, x_2, \ldots, x_{k-1})$, $x_{>k} = (x_{k+1}, \ldots, x_{d-1}, x_d)$

Left and right interpolation sets: $\mathcal{I}_{<k} = \{x_{<k}^{(i)}\}_{i=1}^{r}$, $\mathcal{J}_{>k} = \{x_{>k}^{(j)}\}_{j=1}^{r}$



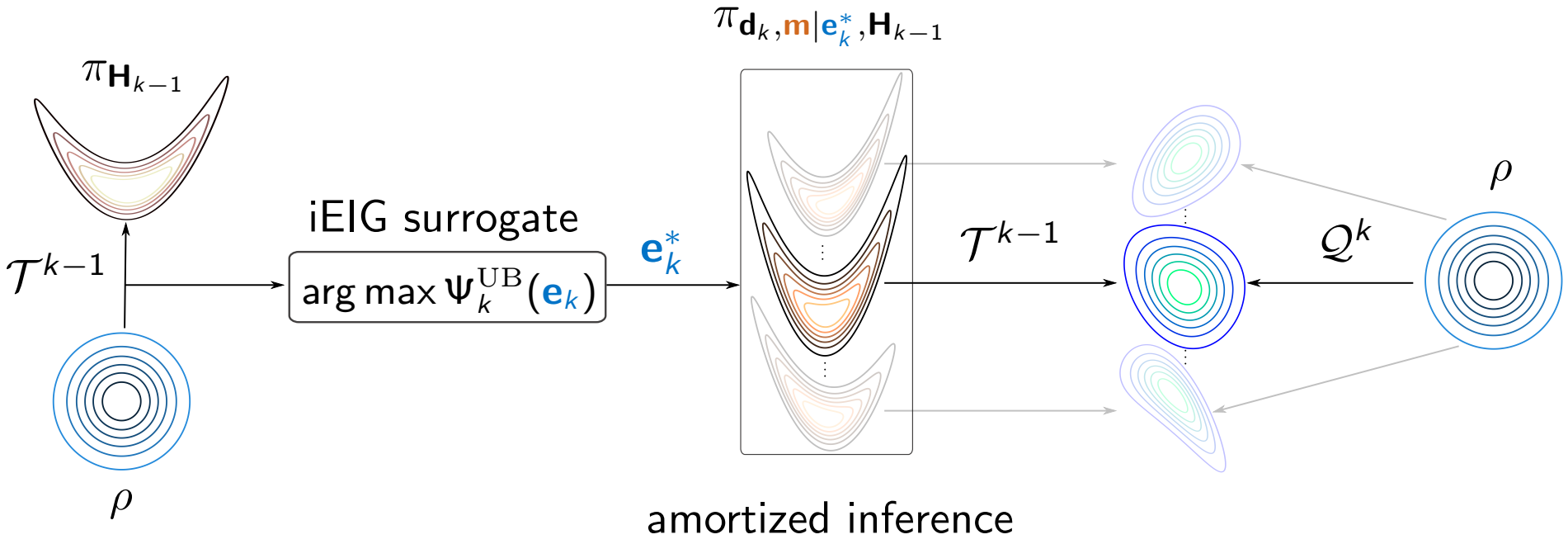
Iterate to next sets:

* Discretize x_k: $\pi(\mathcal{I}_{<k}, \underline{x_k}, \mathcal{J}_{>k})$ for all $x_k \in \mathcal{I}_k$
* $\mathcal{I}_{<k+1} = \texttt{MaxVol} \subset \mathcal{I}_{<k} \otimes \mathcal{I}_k$
* $\mathcal{J}_{>k+1} = \{x_{>k+1}^{(j)}\}_{j=1}^{r}$
* $k \to k+1$, move to the x_{k+1}.
* $k = d$, switch direction and update $\mathcal{J}_{>k}$.
* Stop if converged, repeat otherwise.
* $O(dr^2)$ function evaluations.

TT cross: [Oseledets, Tyrtyshnikov, '10], [Gorodetsky, Karaman, Marzouk, '18]

Empirical interpolation: [Maday, Chaturantabut, Sorensen ...]

Overall procedure



At each stage k:

1. If $k = 1 : \mathcal{T}^0 = \mathcal{I}$
   Else: Obtain data $\mathbf{d}_{k-1}$, restrict $\mathcal{T}^{k-1}(\cdot) = \mathcal{T}^{k-1}_{\mathbf{m}|\mathbf{d}}((\mathcal{T}^{k-1}_{\mathbf{d}})^{-1}(\mathbf{d}_{k-1}), \cdot)$
2. $\mathbf{e}_k^* = \arg\max \frac{1}{2} \log\det\left(\mathbf{I} + H(\mathbf{e}_k, \mathcal{T}^{k-1})\right)$; obtain LIS $\mathbf{U}_k$ from $H(\mathbf{e}_k^*, \mathcal{T}^{k-1})$
3. Build $\mathcal{T}^k_{\mathbf{d},\mathbf{m}}$ (preconditioning + LIS) while waiting for data

A few other tricks...

Map construction

What do we want?

- Invertible expressive map $\mathcal{S}\colon \mathbb{R}^d \to \mathbb{R}^d$ with cheap Jacobian

Our choice: Compose d scalar maps, each mapping one coordinate to reference

$$\mathcal{S} = R \circ t_1 \circ R \circ \cdots \circ R \circ t_{d-1} \circ R \circ t_d$$

(t_k: acts on last coordinate, identity on rest; R: cyclic shift)

Each t_k is a **functional TT with built-in monotonicity**:

$$t_k(\mathbf{x}) = \sum_{i=1}^{n} \underbrace{\sigma(\mathbf{G}_1(x_1)\cdots\mathbf{G}_{d-1}(x_{d-1})\mathbf{f}_i)}_{\geq 0,\ \text{TT coefficients}} \underbrace{c_i(x_d)}_{\nearrow\ \text{monotone}}$$

Key properties:

- Jacobian at each step is a scalar: $|\partial t_k/\partial x_d|$
- Training: sequential, one scalar optimization per coordinate
- Cost per component: $\mathcal{O}(d \cdot n \cdot r^2)$, linear in d

Estimating the predictive EIG via transport maps

From loss-based training to entropy estimation: Target $\pi_{\mathbf{x}}$, fit $\mathcal{S}_F = \mathcal{T}_F^{-1}$ so that $(\mathcal{S}_F)_\sharp \rho \approx \pi_{\mathbf{x}}$ by minimizing $\mathcal{D}_{\mathrm{KL}}(\pi_{\mathbf{x}} || (\mathcal{S}_F)_\sharp \rho)$

- Given samples $\mathbf{x}^i \sim \pi_{\mathbf{x}}$, achieved by minimizing empirical loss:

$$L(F) = \frac{1}{N} \sum_{i=1}^{N} -\log\left[(\mathcal{S}_F)_\sharp \rho(\mathbf{x}^i)\right] \quad \xrightarrow{F^*} \quad L(F^*) \approx h(\mathbf{x}) := -\mathbb{E}_{\mathbf{x}}\left[\log(\pi(\mathbf{x}))\right]$$

Outlook: sequential design for prediction

Naturally combining two threads from today:

sequential structure + *density-free, goal-oriented estimation*

- **Goal:** choose designs over stages to maximize information about a QoI $\mathbf{q}$

$$\Psi_k(\mathbf{e}) = \mathbb{E}_{\mathbf{d}|\mathbf{H}_{k-1}}\left[\mathcal{D}_{\mathrm{KL}}(\pi_{\mathbf{q}|\mathbf{d},\mathbf{e},\mathbf{H}_{k-1}} || \pi_{\mathbf{q}|\mathbf{H}_{k-1}})\right]$$

  - Needs samples $(\mathbf{d}^i(\mathbf{e}), \mathbf{q}^i) \sim \pi_{\mathbf{q},\mathbf{d}|\mathbf{e},\mathbf{H}_{k-1}}$ at each stage

- **How to generate these samples?** Learn $\pi_{\mathbf{m}|\mathbf{H}_{k-1}}$, push through $\mathcal{G}$ + likelihood
  - Extending ideas from sOED: LIS (& important for $\mathbf{q}$?) + preconditioning
  - *Open question:* $\mathbf{m}$ is just intermediate, EIG needs only the low-dim joint $\pi_{\mathbf{q},\mathbf{d}|\mathbf{e},\mathbf{H}_{k-1}}$... Can we bypass it? What maps should we learn?

- **Bias correction:** samples generated with imperfect, incremental maps $\rightsquigarrow$ importance weighting / debiasing to keep samples on target