Optimal experimental design (OED)

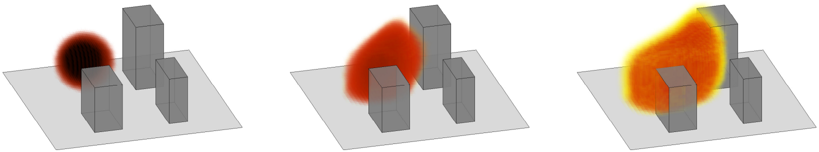
Other names: design of experiments; optimal data collection

- **Main question:** How/where to (optimally) collect observations in inverse problems/data assimilation?
- Typically decided upfront (i.e., experiment planning).
- Particulary important when experiments are **costly/slow/dangerous**.
- **Experimental designs:** Sensor locations, projection angles, excitation frequencies, . . .

Classic research question in electrical engineering, drug testing etc.
Also related to *active learning* in ML.

Diffusive contaminant transport, unknown init. cond.

A linear inverse problem with time-dependent forward model



- **Forward model** f: advection-diffusion equation
- **Inversion parameter**: m initial concentration field
- **Inverse problem**: Use a vector $\boldsymbol{d}$ of point measurements of concentration to infer distribution of m
- **Optimal experimental design problem**: Find sensor placements (to collect the data $\boldsymbol{d}$) that minimize the posterior uncertainty in m

Challenges for linear inverse problem

- Infinite-dimensional inference
- Need proper discretization, and high-dimensional after discretization
- Expensive forward/adjoint solves
- Need posterior covariance (inverse of Hessian, large, dense)
- Optimal design: Combinatorial problem

Reference:

A. Alexanderian, N. Petra, G. Stadler and O. Ghattas. "A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized l_0-sparsification", SIAM J. Sci. Comput., 36(5), A2122-A2148 (2014).

Bayesian inference in Hilbert spaces

- $\mathcal{D}$: bounded domain $\qquad \mathscr{V} = L^2(\mathcal{D}) \qquad m \in \mathscr{V}$: parameter
- Linear parameter-to-observable map: $\mathcal{F} : \mathscr{V} \to \mathbb{R}^q$
- Additive Gaussian noise:

$$\boldsymbol{d} = \mathcal{F} m + \boldsymbol{\eta}, \qquad \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\text{noise}})$$

- Likelihood:

$$\pi_{\text{like}}(\boldsymbol{d}|m) \propto \exp\left\{ -\frac{1}{2}(\mathcal{F} m - \boldsymbol{d})^* \boldsymbol{\Gamma}_{\text{noise}}^{-1} (\mathcal{F} m - \boldsymbol{d}) \right\}$$

- Measurable space: $(\mathscr{V}, \mathcal{B}(\mathscr{V})) \qquad \mu_0$: prior $\qquad \mu_{\text{post}}^{\boldsymbol{d}}$: posterior
- Bayes Theorem (in infinite dimension):

$$\frac{d\mu_{\text{post}}^{\boldsymbol{d}}}{d\mu_{\text{pr}}} \propto \pi_{\text{like}}(\boldsymbol{d}|m) \qquad \left(\text{“} d\mu_{\text{post}}^{\boldsymbol{d}} \propto \pi_{\text{like}}(\boldsymbol{d}|m)\, d\mu_{\text{pr}} \text{”} \right)$$

Bayesian inference in Hilbert spaces

- Prior measure: $\mu_{\text{pr}} = \mathcal{N}(m_{\text{pr}}, \mathcal{C}_{\text{pr}})$

$$\mathcal{C}_{\text{pr}} : \mathscr{V} \to \mathscr{V} \qquad \text{(trace-class operator)}$$

- Gaussian prior/noise and **linear** parameter-to-observable map implies:

$$\mu_{\text{post}}^{\boldsymbol{d}} = \mathcal{N}(m_{\text{post}}, \mathcal{C}_{\text{post}})$$

- Posterior covariance:

$$\mathcal{C}_{\text{post}} = (\mathcal{F}^* \boldsymbol{\Gamma}_{\text{noise}}^{-1} \mathcal{F} + \mathcal{C}_{\text{pr}}^{-1})^{-1} \text{ (independent of } m \text{ and } \boldsymbol{d}\text{!)}$$

- Posterior mean: $m_{\text{post}} = \mathcal{C}_{\text{post}}(\mathcal{F}^* \boldsymbol{\Gamma}_{\text{noise}}^{-1} \boldsymbol{d} + \mathcal{C}_{\text{pr}}^{-1} m_{\text{pr}})$

Optimal experimental design (OED)

- A-optimal design:

Minimize "average variance" of m

- Covariance function: $c(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Cov}\,\{m(\boldsymbol{x}), m(\boldsymbol{y})\}$

$$\text{average variance} = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} c(\boldsymbol{x}, \boldsymbol{x})\, d\boldsymbol{x}$$

- Covariance operator:

$$[\mathcal{C}_{\text{post}} u](\boldsymbol{x}) = \int_{\mathcal{D}} c(\boldsymbol{x}, \boldsymbol{y}) u(\boldsymbol{y})\, d\boldsymbol{y}$$

- Mercer's Theorem:

$$\int_{\mathcal{D}} c(\boldsymbol{x}, \boldsymbol{x})\, d\boldsymbol{x} = \mathrm{tr}(\mathcal{C}_{\text{post}})$$

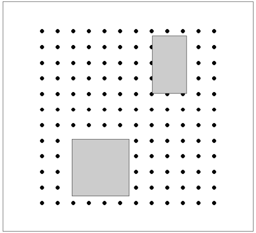
- Optimal design criterion (A-optimal design):

Choose a "design" to minimize $\mathrm{tr}(\mathcal{C}_{\textbf{post}})$

The design and a weighted inference problem

Finite-dimensional sensor domain

$$\text{design} := \begin{Bmatrix} \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_s} \\ w_1, \ldots, w_{n_s} \end{Bmatrix}$$



- $\boldsymbol{x}_i$: candidate sensor locations
- w_i: weights
- Ideally: $w_i \in \{0, 1\}$
- Relax: $0 \le w_i \le 1$

- $\boldsymbol{w}$-weighted data-likelihood:

$$\pi_{\text{like}}(\boldsymbol{d}|m; \boldsymbol{w}) \propto \exp\left\{ -\frac{1}{2\sigma^2_{noise}} (\mathcal{F}m - \boldsymbol{d})^T \mathbf{W} (\mathcal{F}m - \boldsymbol{d}) \right\}$$

- $\mathbf{W}$: diagonal matrix with $\boldsymbol{w}$ on its diagonal; posterior covariance operator:

$$\mathcal{C}_{\text{post}}(\boldsymbol{w}) = \left(\frac{1}{\sigma^2_{noise}} \mathcal{F}^* \mathbf{W} \mathcal{F} + \mathcal{C}_{\text{pr}}^{-1} \right)^{-1}$$

A-optimal design with sparsity control

$$\begin{aligned} \underset{\boldsymbol{w}\in\mathbb{R}^{n_s}}{\textit{minimize}} \quad & \mathrm{tr}\left[\boldsymbol{\Gamma}_{\mathrm{post}}(\boldsymbol{w})\right] + \gamma P(\boldsymbol{w}) \\ \textit{subject to} \quad & \mathbf{0} \le \boldsymbol{w} \le \mathbf{1} \end{aligned}$$

- $P(\boldsymbol{w})$: penalty term
- Discretized posterior covariance operator:

$$\boldsymbol{\Gamma}_{\mathrm{post}}(\boldsymbol{w}) = \Big(\underbrace{\frac{1}{\sigma^2_{noise}}\mathbf{F}^*\mathbf{W}\mathbf{F}}_{\mathcal{H}_{\mathrm{misfit}}} + \boldsymbol{\Gamma}_{\mathrm{pr}}^{-1}\Big)^{-1}$$

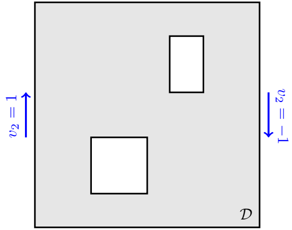
- Numerical optimization: interior-point (BFGS approx to Hessian of OED objective function)

Model problem

Time dependent advection-diffusion

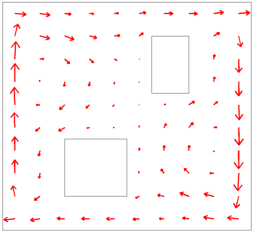
- Forward problem:

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



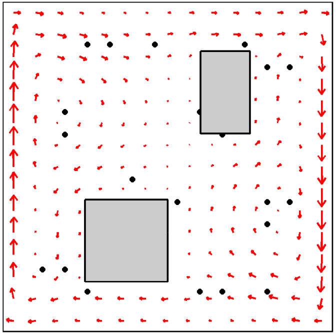
- Adjoint problem:

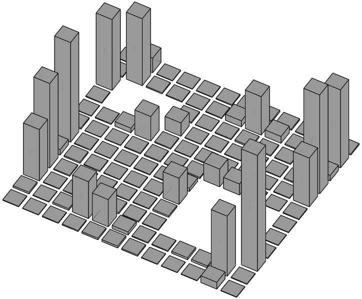
$$\begin{aligned} -p_t - \nabla \cdot (p\boldsymbol{v}) - \kappa \Delta p &= -\mathcal{B}^* \boldsymbol{\Gamma}_{\text{noise}}^{-1} (\mathcal{B}u - \boldsymbol{d}) \\ p(T) &= 0 \\ (\boldsymbol{v}p + \kappa \nabla p) \cdot \boldsymbol{n} &= 0 \end{aligned}$$



- m: *Unknown* initial condition
- $\boldsymbol{v}$: Velocity "wind" field — here: assumed known

A-optimal design: ℓ^1-sparsity control





Towards 0–1 design vectors

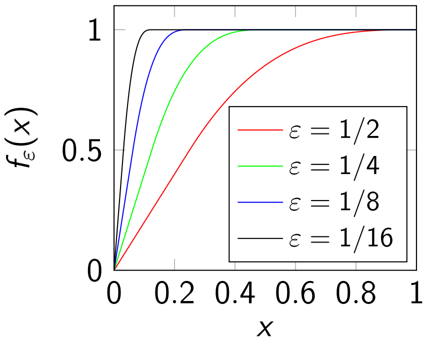
A family of penalty functions and continuation strategy

$$\begin{aligned} \underset{\boldsymbol{w} \in \mathbb{R}^{n_s}}{\text{minimize}} \quad & \text{tr}\left[\boldsymbol{\Gamma}_{\text{post}}(\boldsymbol{w})\right] + \gamma P_\varepsilon(\boldsymbol{w}) \\ \text{subject to} \quad & \mathbf{0} \le \boldsymbol{w} \le \mathbf{1} \end{aligned}$$

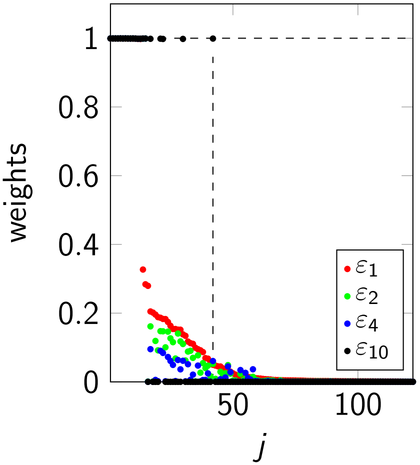
- Motivated by continuation ideas from topology optimization
- $P_\varepsilon(\boldsymbol{w}) := \sum_{i=1}^{n_s} f_\varepsilon(w_i)$

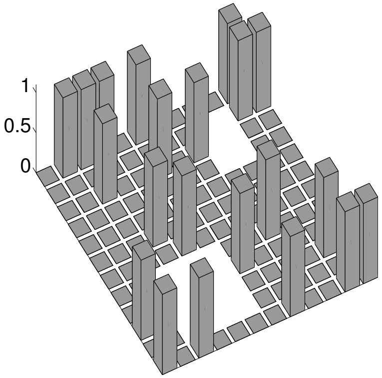
$$f_\varepsilon(x) = \begin{cases} \dfrac{x}{\varepsilon}, & 0 \le x \le \frac{1}{2}\varepsilon \\ p_\varepsilon(x), & \frac{1}{2}\varepsilon < x \le 2\varepsilon \\ 1, & 2\varepsilon < x \le 1 \end{cases}$$

- p_ε: cubic polynomial computed such that f_ε is C^1

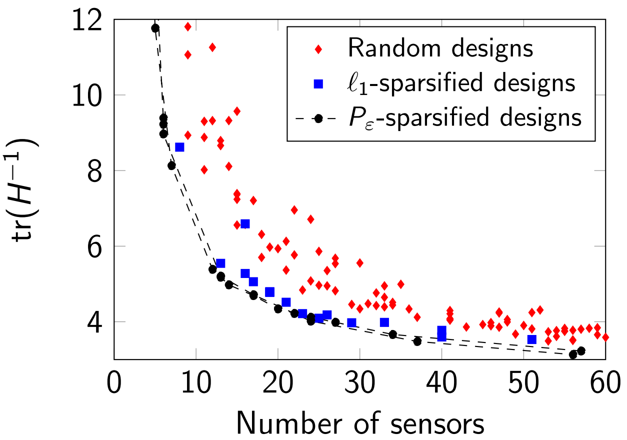


0–1 designs with continuation



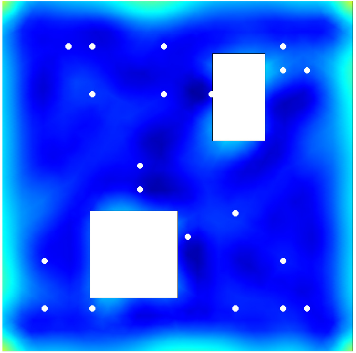


Comparing ℓ^1 sparsification vs P_ε

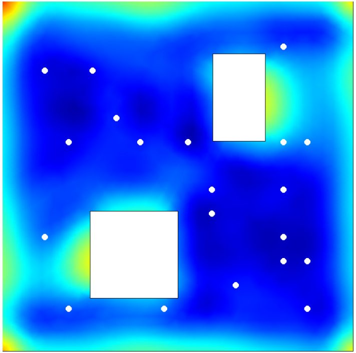


- OED improves significantly over random designs
- P_ε-sparsified designs better than ℓ_1-sparsified designs

A-optimal design: the variance field

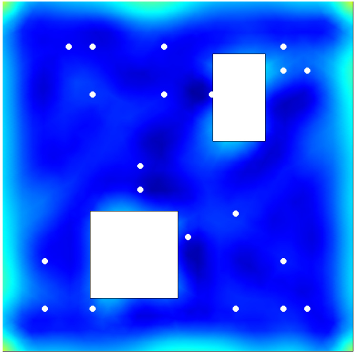


Optimal

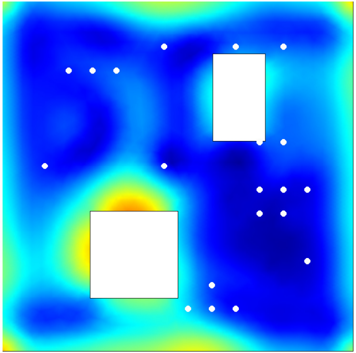


Sub-optimal

A-optimal design: the variance field

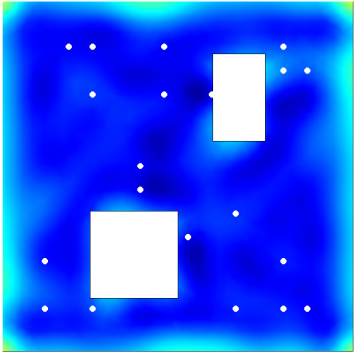


Optimal

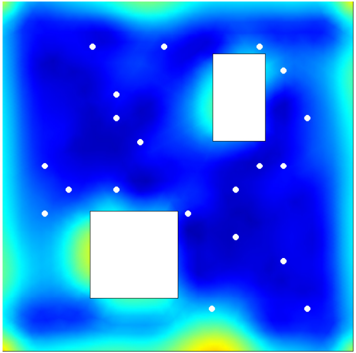


Sub-optimal

A-optimal design: the variance field



Optimal



Sub-optimal

Challenges for linear inverse problem

- Infinite-dimensional inference
- Need proper discretization, and high-dimensional after discretization

- Expensive forward/adjoint solves
- Need posterior covariance (inverse of Hessian, large, dense)

- Optimal design: Combinatorial problem

Challenges for linear inverse problem

- Infinite-dimensional inference (formulation of inference in inf. dim)
- Need proper discretization, and high-dimensional after discretization (Matern priors using Laplace-like PDE operators; randomized estimation of trace)
- Expensive forward/adjoint solves (matrix-free using adjoints)
- Need posterior covariance (inverse of Hessian, large, dense) (low-rank approximation of parameter-to-data map; Sherman-Woodbury)
- Optimal design: Combinatorial problem (relaxation and penalization; or greedy approach)

Optimal experimental design (OED)

Above we used *A-optimal design*, i.e., minimizing the average variance (trace of the posterior covariance matrix/operator).

Alternative measures:

- *D-optimal design*: Expected information gain from prior to posterior using Kullback-Leibler divergence (discretized, becomes determinant of covariance matrix)
- *E-optimal design*: Minimizes the maximal eigenvalue of the covariance matrix
- ...

Optimal experimental design (OED)

Above we used *A-optimal design*, i.e., minimizing the average variance (trace of the posterior covariance matrix/operator).

Alternative measures:

- *D-optimal design*: Expected information gain from prior to posterior using Kullback-Leibler divergence (discretized, becomes determinant of covariance matrix)
- *E-optimal design*: Minimizes the maximal eigenvalue of the covariance matrix
- ...

Different design criteria typically give different designs; some are tricky to interpret in infinite dimensions.

Optimal experimental design (OED)

Challenges (tomorrow's lecture):

- Computing trace of posterior efficiently (posterior is typically not assembled!)
- Alternatives to solve the ℓ_0 optimization problem; greedy approaches (how suboptimal are they?); Alternatives: norm-reweighting, splitting approaches, etc
- This is for *linear* inverse problems. Things are much less clear for nonlinear inverse problems as covariance is typically not directly accessible
- Here we were certain of our mathematical models–but models can be uncertain (OED under uncertainty)