Optimal design of experiments for inverse problems:

Computational aspects, extension to nonlinear problems and uncertain models

Georg Stadler

Courant Institute, New York University
stadler@cims.nyu.edu

Lecture material at: https://cims.nyu.edu/~stadler/oed/

June 30, 2023

Optimal experimental design (OED)

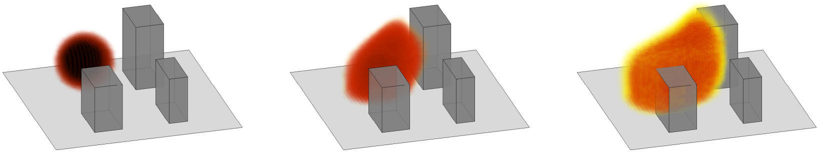
Other names: design of experiments; optimal data collection

- **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

Outline

Computation of posterior trace, other OED optimality criteria

OED with model errors

OED for nonlinear model

A-optimal design with sparsity control

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

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

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

Main challenges:

- Choice of penalty function to achieve binary weights
- Re-computation of inverse covariance trace
- Gradient of posterior trace with respect to $\boldsymbol{w}$

A-optimal design: the cost and gradient

- Randomized trace estimator:

$$\operatorname{tr}\left[\boldsymbol{\Gamma}_{\text{post}}(\boldsymbol{w})\right] \approx \frac{1}{n_{\text{tr}}} \sum_{i=1}^{n_{\text{tr}}} \langle \boldsymbol{z}_i, \boldsymbol{\Gamma}_{\text{post}}(\boldsymbol{w}) \boldsymbol{z}_i \rangle =: \phi(\boldsymbol{w})$$

- $\boldsymbol{z}_i$ random vectors (e.g., Gaussian)
- Computation of cost and gradient:

$$\phi(\boldsymbol{w}) = \frac{1}{n_{\text{tr}}} \sum_{i=1}^{n_{\text{tr}}} \langle \boldsymbol{z}_i, \boldsymbol{q}_i \rangle \qquad \frac{\partial \phi}{\partial w_j} = -\frac{1}{n_{\text{tr}}} \sum_{i=1}^{n_{\text{tr}}} \langle \boldsymbol{q}_i, \partial_j \mathcal{H}_{\text{misfit}} \boldsymbol{q}_i \rangle$$

- Need $\boldsymbol{q}_i = \boldsymbol{\Gamma}_{\text{post}}(\boldsymbol{w}) \boldsymbol{z}_i$, $i = 1, \ldots, n_{\text{tr}}$
- Need $\mathbf{F}\boldsymbol{q}_i$, $i = 1, \ldots, n_{\text{tr}}$ (for computing $\frac{\partial \phi}{\partial w_j}$)

SVD surrogate for the parameter-to-observable map

- Need many applications of $\mathbf{F}$ in the optimization process
- Idea: $\mathbf{F}$ is low-rank (often) $\Longrightarrow$ compute low-rank SVD for $\mathbf{F}$
- Better idea: Compute SVD surrogate for $\tilde{\mathbf{F}} = \mathbf{F}\boldsymbol{\Gamma}_{\text{pr}}^{1/2}$
- Use randomized SVD*
- SVD surrogate for $\tilde{\mathbf{F}}$:

  no forward/adjoint PDE solves in OED algorithm

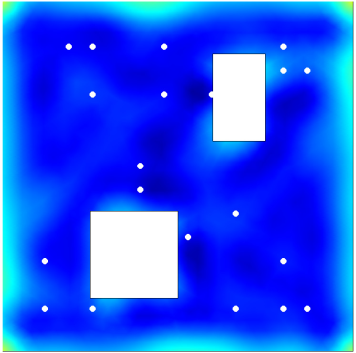
*N. Halko, P.G. Martinsson, J.A. Tropp, Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. SIAM Review (2011).

Alternative: Rewrite posterior coariance in observation space

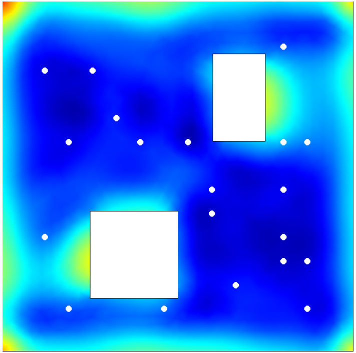
- Reformulate posterior covariance trace in observation space
- Has advantages when we consider a moderate number of observation location candidates
- Requires inverse only in the observation space dimension

Koval, Alexanderian, Stadler: Inverse Problems 2020. Alexanderian: Inverse Problems 2021.

A-optimal design: the variance field

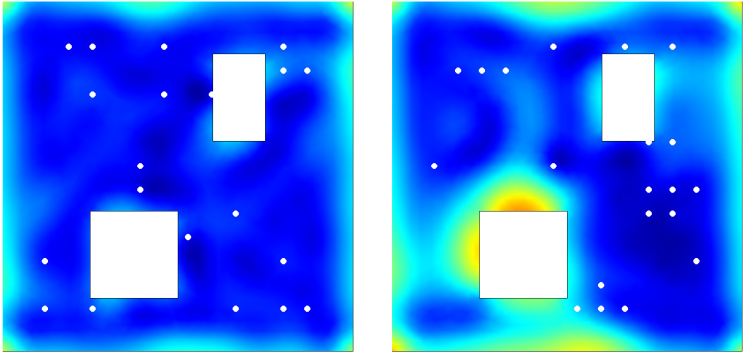


Optimal



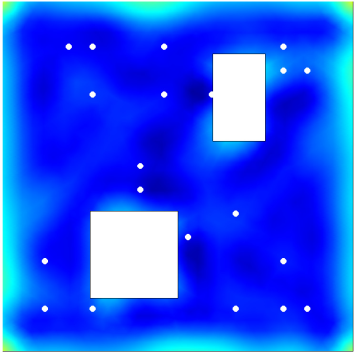
Sub-optimal

A-optimal design: the variance field

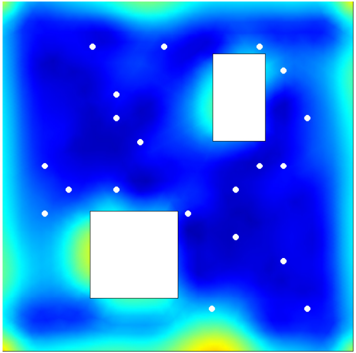


Optimal Sub-optimal

A-optimal design: the variance field



Optimal



Sub-optimal

Alternative design criteria

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.

Outline

Computation of posterior trace, other OED optimality criteria

OED with model errors

OED for nonlinear model

Extension I: Forward model with uncertainty

Forward problem with uncertain advection $\mathbf{v} = \mathbf{v}(\xi)$

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

Extension I: Forward model with uncertainty

Forward problem with uncertain advection $\mathbf{v} = \mathbf{v}(\xi)$

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

Corresponding posterior covariance matrix:

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

Extension I: Forward model with uncertainty

Forward problem with uncertain advection $\mathbf{v} = \mathbf{v}(\xi)$

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

Corresponding posterior covariance matrix:

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

OED formulation:

$$\min_{\boldsymbol{w} \in [0,1]^d} \quad \operatorname{tr}\left(\mathcal{C}_{\text{post}}(\boldsymbol{w}, \xi)\right) \qquad + \gamma P_\varepsilon(\boldsymbol{w})$$

Extension I: Forward model with uncertainty

Forward problem with uncertain advection $\mathbf{v} = \mathbf{v}(\xi)$

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

Corresponding posterior covariance matrix:

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

OED formulation:

$$
\min_{\boldsymbol{w} \in [0,1]^d} \int_{\Omega} \operatorname{tr}\left(\mathcal{C}_{\text{post}}(\boldsymbol{w}, \xi) \right) P(d\xi) + \gamma P_{\varepsilon}(\boldsymbol{w})
$$

Reference:

K. Koval, A. Alexanderian and G. Stadler, Optimal experimental design under irreducible uncertainty for inverse problems governed by PDEs, Inverse Problems, 2020.

(1) Computation of trace, "measurement space approach"

$$\phi(\xi, \boldsymbol{w}) = \text{tr}\left[\left(\frac{1}{\sigma^2}\mathcal{F}(\xi)^*\boldsymbol{W}\mathcal{F}(\xi) + \mathcal{C}_0{}^{-1}\right)^{-1}\right]$$

- Avoid having to approximate trace of an infinite-dimensional operator
- We can show that we can rewrite $\phi(\xi, \boldsymbol{w})$ as:

$$\phi(\xi, \boldsymbol{w}) = \text{tr}\,(\mathcal{C}_0) - \text{tr}\left[\frac{1}{\sigma^2}\mathcal{C}_0\mathcal{F}^*(\xi)\mathcal{S}^{-1}(\xi, \boldsymbol{w})\boldsymbol{W}\mathcal{F}(\xi)\mathcal{C}_0\right]$$

where $\left(\mathcal{I} + \frac{1}{\sigma^2}\boldsymbol{W}\mathcal{F}(\xi)\mathcal{C}_0\mathcal{F}^*(\xi)\right) = \mathcal{S}(\xi, \boldsymbol{w})$

(1) Computation of trace, "measurement space approach"

- Rearranging terms in the trace, we have:

$$\begin{aligned}\phi(\xi, \boldsymbol{w}) &= \mathrm{tr}\,(\mathcal{C}_0) - \mathrm{tr}\left[\frac{1}{\sigma^2}\mathcal{S}^{-1}(\xi, \boldsymbol{w})\boldsymbol{W}\mathcal{F}(\xi)\mathcal{C}_0^2\mathcal{F}^*(\xi)\right] \\ &= \mathrm{tr}\,(\mathcal{C}_0) - \mathrm{tr}\,[\mathcal{K}(\xi, \boldsymbol{w})]\end{aligned}$$

- First term independent of $\boldsymbol{w}$ $\Longrightarrow$ drop it
- Optimal design now satisfies:

$$\boldsymbol{w}^* = \mathrm{argmin}_{\boldsymbol{w}\in[0,1]^d} - \int_\Omega \mathrm{tr}\,[\mathcal{K}(\xi, \boldsymbol{w})]\, P(d\xi) + \gamma\psi\,(\boldsymbol{w})$$

- The dimensionality of the operator is reduced in this formulation
- Trace of an operator in observation space (finite)

(2) Discretization of the uncertainty

- Approximate the expected value of the average pointwise posterior variance
- Assuming we can sample $\xi_i \in \Omega$, we use SAA to approximate the integral

$$\int_\Omega \operatorname{tr}\left[\mathcal{K}(\xi, \boldsymbol{w})\right] P(d\xi) \approx \frac{1}{N} \sum_{i=1}^{N} \operatorname{tr}\left[\mathcal{K}(\xi_i, \boldsymbol{w})\right]$$

(3) Elimination of PDEs from the minimization

OEDUU objective (with discretized operators):

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{tr}\left[\boldsymbol{K}(\xi_i, \boldsymbol{w})\right] = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{d}\langle \boldsymbol{e}_j, \boldsymbol{K}(\xi_i, \boldsymbol{w})\boldsymbol{e}_j\rangle\right)$$

where

$$\boldsymbol{K}(\xi_i, \boldsymbol{w}) = \frac{1}{\sigma^2}\boldsymbol{S}^{-1}(\xi, \boldsymbol{w})\boldsymbol{W}\boldsymbol{F}_i\boldsymbol{C}_0^2\boldsymbol{F}_i^*$$

$$\boldsymbol{S}(\xi_i, \boldsymbol{w}) = \left(\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{W}\boldsymbol{F}_i\boldsymbol{C}_0\boldsymbol{F}_i^*\right)$$

Expensive to optimize:

- in each step of minimization need to compute trace for each sample
- evaluation of trace requires d applications of $\boldsymbol{K}(\xi_i, \boldsymbol{w})$
- each application requires 2 PDE solves plus applying the inverse of an operator containing 2 PDE solves

(3) Elimination of PDEs from the minimization

Find a low-rank approximation to $\mathcal{C}_0^{\frac{1}{2}}\boldsymbol{F}(\xi_i) = \widetilde{\boldsymbol{F}}_i$

- Eigenspectrum decays quickly, thus feasible
- Done via randomized range-finder techniques (Halko et al., 2011)

Storing separate basis vectors for each $\widetilde{\boldsymbol{F}}_i$ is infeasible

- Different choices of $\mathbf{v}(\xi)$ may have similar learning directions
- **Solution**: find a space that captures the "effective" composite range space

Find $\boldsymbol{Q} \in \mathbb{R}^{d\times k}$ and $\widehat{\boldsymbol{Q}} \in \mathbb{R}^{m\times k}$ (k small) such that $\forall\, i \in [1, \cdots, N]$:

$$\widetilde{\boldsymbol{F}}_i \approx \boldsymbol{Q}\boldsymbol{Q}^*\widetilde{\boldsymbol{F}}_i\widehat{\boldsymbol{Q}}\widehat{\boldsymbol{Q}}^*$$

(4) Reduced basis and clustering

Joint basis algorithm:

Algorithm 1 Composite randomized range-finder algorithm

1: N random matrices $\boldsymbol{\Omega}_i \in \mathbb{R}^{d\times r}$, $\widehat{\boldsymbol{\Omega}}_i \in \mathbb{R}^{n\times r}$, $i = 1, \ldots, N$.
2: $\boldsymbol{Y}_i = \widetilde{\boldsymbol{F}}_i \boldsymbol{\Omega}_i$, $\widehat{\boldsymbol{Y}}_i = \widetilde{\boldsymbol{F}}_i^T \widehat{\boldsymbol{\Omega}}_i$.
3: Set $\boldsymbol{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N]$ and $\widehat{\boldsymbol{Y}} = [\widehat{\boldsymbol{Y}}_1, \ldots, \widehat{\boldsymbol{Y}}_N]$.
4: Compute SVDs of $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ and of $\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{V}}^T$
5: Truncate SVDs up to desired tolerance
6: $\boldsymbol{Q} = \boldsymbol{U}[:, 1:k]$ and $\widehat{\boldsymbol{Q}} = \widehat{\boldsymbol{U}}[:, 1:k]$.
7: **return** $\boldsymbol{Q}$ and $\widehat{\boldsymbol{Q}}$

(4) Reduced basis and clustering

Potential issues:

- Still need to apply $\widetilde{\boldsymbol{F}}_i$ to k vectors to compute $\boldsymbol{Q}^* \widetilde{\boldsymbol{F}}_i \widehat{\boldsymbol{Q}}$
  - Requires minimal-residual solution to

$$\boldsymbol{B}_i \widehat{\boldsymbol{Q}}^T \boldsymbol{\Omega}_i = \boldsymbol{Q}^T \boldsymbol{Y}_i, \tag{1}$$

$$\boldsymbol{B}_i{}^T \boldsymbol{Q}^T \widehat{\boldsymbol{\Omega}}_i = \widehat{\boldsymbol{Q}}^T \widehat{\boldsymbol{Y}}_i \tag{2}$$

- Memory usage $\uparrow$ as $N \uparrow$
  - Cluster "similar" uncertain parameters ξ together
  - Definition of "similarity" metric problem specific
  - For each cluster, compute a composite basis and store it
  - Reduces basis size and PDE solves needed

(4) Reduced basis and clustering

Potential issues:

- Still need to apply $\widetilde{\boldsymbol{F}}_i$ to k vectors to compute $\boldsymbol{Q}^*\widetilde{\boldsymbol{F}}_i\widehat{\boldsymbol{Q}}$
  - Find $\boldsymbol{B}_i \approx \boldsymbol{Q}^*\widetilde{\boldsymbol{F}}_i\widehat{\boldsymbol{Q}}$
  - Requires minimal-residual solution to

$$\boldsymbol{B}_i\widehat{\boldsymbol{Q}}^T\boldsymbol{\Omega}_i = \boldsymbol{Q}^T\boldsymbol{Y}_i, \tag{1}$$

$$\boldsymbol{B}_i{}^T\boldsymbol{Q}^T\widehat{\boldsymbol{\Omega}}_i = \widehat{\boldsymbol{Q}}^T\widehat{\boldsymbol{Y}}_i \tag{2}$$

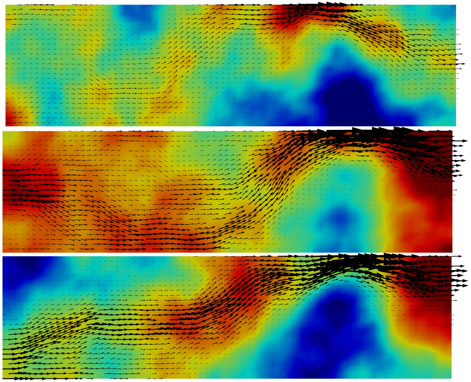
- Memory usage $\uparrow$ as N $\uparrow$

Synthesizing a distribution for $\boldsymbol{v}$

Solve Bayesian inverse problem for ξ using synthetic data

- Sample $\xi_i \rightsquigarrow$ sample $\boldsymbol{v}_i$

$$
\begin{aligned}
-\nabla \cdot (e^{\xi} \nabla p) &= 0 \quad \text{in } D, \\
p &= 0 \quad \text{on } \Gamma_L, \\
p &= 1 \quad \text{on } \Gamma_R, \\
e^{\xi} \nabla p \cdot \boldsymbol{n} &= 0 \quad \text{on } \Gamma_N
\end{aligned}
$$



- Leads to $\pi_{\xi}(\xi)$
- $\pi_{\xi}(\xi)$: Gaussian approximation at MAP

Subsurface flow OEDUU

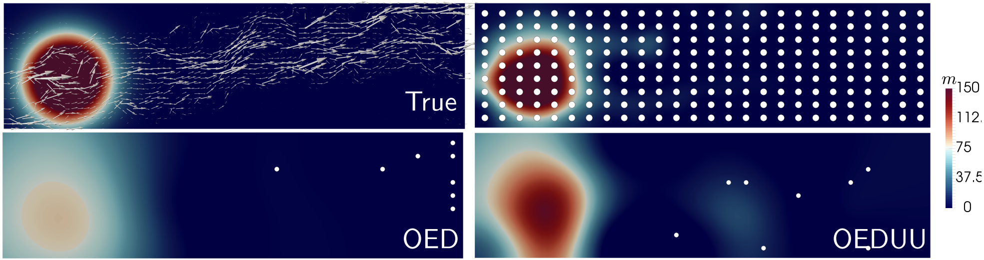
$$\boldsymbol{w}^* = \text{argmin}_{\boldsymbol{w}\in[0,1]^d} -\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{K}(\boldsymbol{v}_i, \boldsymbol{w}) + \gamma\psi(\boldsymbol{w})$$

- Solve minimization problem many times with different $\psi(\boldsymbol{w})$
  - Change $\psi(\boldsymbol{w})$ to approximate ℓ_0 "norm" better in each iteration
- Each minimization solved with gradient-based method (projected BFGS)
- Number of minimization problems needed to solve ≈ 10

Design comparisons

How well do the designs do for initial condition inversion using the "true" $\boldsymbol{v}$?

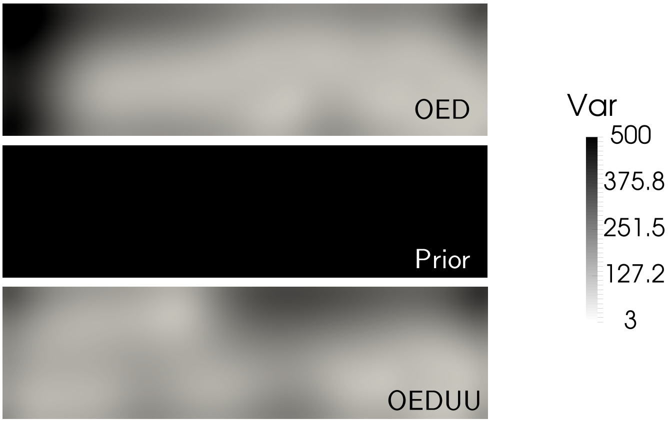
MAP points



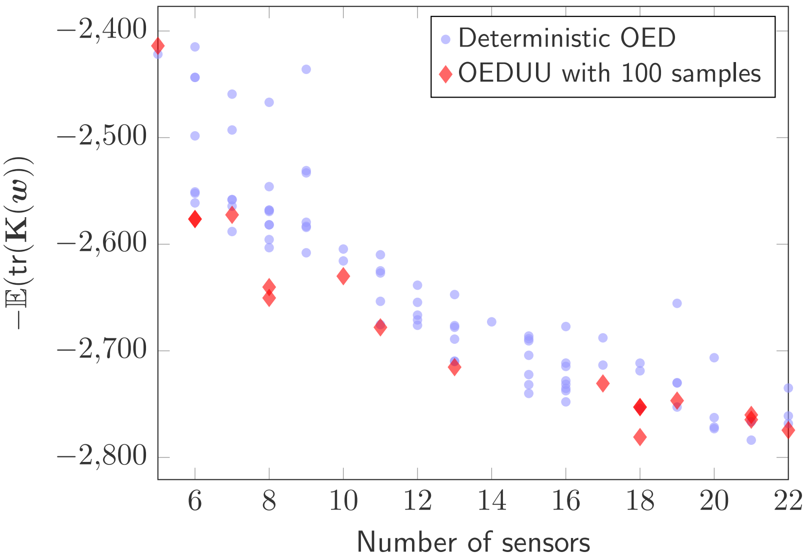
Design comparisons

How well do the designs do for initial condition inversion using the "true" v ?

Pointwise variance reduction



Deterministic vs. designs under uncertainty



Extension II: Nonlinear forward model

- Nonlinear model, Gaussian prior and noise $\not\Longrightarrow$ Gaussian posterior
- Use Gaussian approximation to posterior: for given $\boldsymbol{w}$ and $\boldsymbol{d}$
  - Compute the maximum a posteriori probability (MAP) estimate $m_{\text{MAP}}(\boldsymbol{w};\boldsymbol{d})$
  - Gaussian approximation to posterior at MAP point

$$\mathcal{N}\left(m_{\text{MAP}}(\boldsymbol{w};\boldsymbol{d}), \mathcal{H}^{-1}\left[m_{\text{MAP}}(\boldsymbol{w};\boldsymbol{d}), \boldsymbol{w};\boldsymbol{d}\right]\right)$$

- Data $\boldsymbol{d}$ not available *a priori*

Example: Log permeability m inversion from pressure measurements $u(x_i)$:

$$\begin{aligned} -\nabla\cdot(\exp(m)\nabla u) &= f && \text{in } \mathcal{D} \\ u &= 0 && \text{on } \Gamma_D \\ \exp(m)\nabla u\cdot\boldsymbol{n} &= g && \text{on } \Gamma_N \end{aligned}$$

Reference:

A. Alexanderian, N. Petra, G. Stadler and O. Ghattas: A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems, SISC, 2016.

Outline

Computation of posterior trace, other OED optimality criteria

OED with model errors

OED for nonlinear model

Extension II: Nonlinear forward model

- General formulation: minimize average posterior variance:

$$\underset{\boldsymbol{w}}{\text{minimize}} \qquad \operatorname{tr}\left(\mathcal{H}^{-1}\left[m_{\text{MAP}}(\boldsymbol{w};\boldsymbol{d}),\boldsymbol{w};\boldsymbol{d}\right]\right)$$

Extension II: Nonlinear forward model

- General formulation: minimize *expected* average posterior variance:

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \int \left\{ \text{tr}\left(\mathcal{H}^{-1}\left[m_{\text{MAP}}(\boldsymbol{w}; \boldsymbol{d}), \boldsymbol{w}; \boldsymbol{d} \right] \right) \right\} d\mu(\boldsymbol{d})$$

Extension II: Nonlinear forward model

- General formulation: minimize *expected* average posterior variance:

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \int \left\{ \text{tr}\left(\mathcal{H}^{-1}\left[m_{\text{MAP}}(\boldsymbol{w}; \boldsymbol{d}), \boldsymbol{w}; \boldsymbol{d} \right] \right) \right\} d\mu(\boldsymbol{d}) + \gamma P(\boldsymbol{w})$$

- $\gamma P(\boldsymbol{w})$: sparsifying penalty function

Extension II: Nonlinear forward model

- General formulation: minimize *expected* average posterior variance:

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \int \left\{ \text{tr}\left(\mathcal{H}^{-1}\left[m_{\text{MAP}}(\boldsymbol{w};\boldsymbol{d}), \boldsymbol{w}; \boldsymbol{d} \right] \right) \right\} d\mu(\boldsymbol{d}) + \gamma P(\boldsymbol{w})$$

- $\gamma P(\boldsymbol{w})$: sparsifying penalty function
- In practice: get data from a few training models $m_1, \ldots, m_n$
- m_i: draws from prior
- Training data:

$$\boldsymbol{d}_i = f(m_i) + \boldsymbol{\eta}_i, \quad i = 1, \ldots, n$$

Extension II: Nonlinear forward model

- General formulation: minimize *expected* average posterior variance:

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \int \left\{ \text{tr}\left(\mathcal{H}^{-1}\left[m_{\text{MAP}}(\boldsymbol{w}; \boldsymbol{d}), \boldsymbol{w}; \boldsymbol{d} \right] \right) \right\} d\mu(\boldsymbol{d}) + \gamma P(\boldsymbol{w})$$

- $\gamma P(\boldsymbol{w})$: sparsifying penalty function
- In practice: get data from a few training models $m_1, \dots, m_n$
- m_i: draws from prior
- Training data:

$$\boldsymbol{d}_i = f(m_i) + \boldsymbol{\eta}_i, \quad i = 1, \dots, n$$

- The problem to solve in practice:

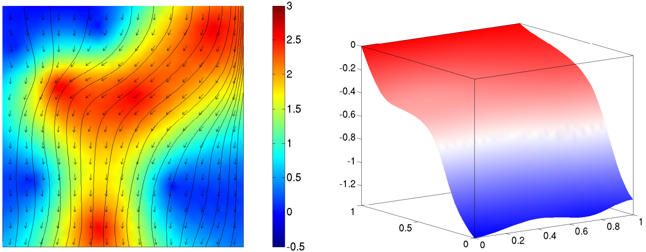
$$\underset{\boldsymbol{w}}{\text{minimize}} \; \frac{1}{n} \sum_{i=1}^{n} \text{tr}\left(\mathcal{H}^{-1}\left[m_{\text{MAP}}(\boldsymbol{w}; \boldsymbol{d}_i), \boldsymbol{w}; \boldsymbol{d}_i \right] \right) + \gamma P(\boldsymbol{w})$$

Application: subsurface flow

- Forward problem

$$
\begin{aligned}
-\nabla\cdot(e^m\nabla u) &= f && \text{in } \mathcal{D}\\
u &= 0 && \text{on } \Gamma_D\\
e^m\nabla u\cdot\boldsymbol{n} &= g && \text{on } \Gamma_N
\end{aligned}
$$

- u: pressure-field
- m: log-permeability (inversion parameter)



Left: true parameter, right: pressure-field

Bayesian inverse problem: Gaussian approximation

- Reduced cost functional (for given data $\boldsymbol{d}$)

$$
\begin{aligned}
\hat{\mathcal{J}}(\theta) &= \mathcal{J}(u(\theta), \theta) \\
&= \frac{1}{2\sigma^2_{noise}}(\mathcal{B}u - \boldsymbol{d})^T \mathbf{W}(\mathcal{B}u - \boldsymbol{d}) + \frac{1}{2}\langle \mathcal{C}_{\mathrm{pr}}^{-1}(m - m_{prior}), m - m_{prior} \rangle
\end{aligned}
$$

- $\mathcal{B}$ is observation operator
- MAP point is solution to

$$
\min_{m} \hat{\mathcal{J}}(m)
$$

where

$$
\begin{aligned}
-\nabla \cdot (e^m \nabla u) &= f && \text{in } \mathcal{D} \\
u &= 0 && \text{on } \Gamma_D \\
e^m \nabla u \cdot \boldsymbol{n} &= g && \text{on } \Gamma_N
\end{aligned}
$$

- Covariance of Gaussian approximation to posterior:

$$
\mathcal{H}(m)^{-1} = (\nabla^2 \hat{\mathcal{J}}(m))^{-1}
$$

Optimization problem for computing A-optimal design

- OED objective function:

$$\mathrm{tr}(\mathcal{H}^{-1}) \doteq \sum_i \big\langle z_i, \mathcal{H}^{-1} z_i \big\rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i$$

- Optimization problem:

$$\underset{\boldsymbol{w}}{\text{minimize}} \sum_i \langle z_i, y_i(\boldsymbol{w}) \rangle + \gamma P(\boldsymbol{w})$$

where

$$m_{\text{MAP}}(\boldsymbol{w}) = \arg\min_{\theta} \mathcal{J}\Big(u(\theta(\boldsymbol{w})), \theta(\boldsymbol{w})\Big)$$

$$\mathcal{H}\big(m_{\text{MAP}}(\boldsymbol{w})\big) y_i = z_i$$

Optimization problem for computing A-optimal design

OED objective function:

$$\mathrm{tr}(\mathcal{H}^{-1}) = \sum_i \langle z_i, \mathcal{H}^{-1} z_i \rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i$$

$$\underset{\boldsymbol{w}\in[0,1]^{n_s}}{\mathit{minimize}} \sum_i \langle z_i, y_i \rangle + \gamma P(\boldsymbol{w})$$

subject to

$$-\nabla \cdot (e^m \nabla u) = f \qquad \textit{(state)}$$

$$-\nabla \cdot (e^m \nabla p) = -\frac{1}{\sigma^2_{noise}} \mathcal{B}^* \mathbf{W} (\mathcal{B} u - \boldsymbol{d}) \qquad \textit{(adjoint)}$$

$$\mathcal{C}_{pr}^{-1}(m - m_{prior}) + e^m \nabla u \cdot \nabla p = 0 \qquad \textit{(gradient = 0)}$$

$$-\nabla \cdot (e^m \nabla v_i) = \nabla \cdot (y_i e^m \nabla u) \qquad \textit{(inc. state)}$$

$$-\nabla \cdot (e^m \nabla q_i) = \nabla \cdot (y_i e^m \nabla p) - \frac{1}{\sigma^2_{noise}} \mathcal{B}^* \mathbf{W} \mathcal{B} v_i \qquad \textit{(inc. adjoint)}$$

$$\mathcal{C}_{pr}^{-1} y_i + y_i e^m \nabla u \cdot \nabla p + e^m \left(\nabla v_i \cdot \nabla p + \nabla u \cdot \nabla q_i \right) = z_i \qquad \textit{(Hessian solve)}$$

Optimization problem for computing A-optimal design

OED objective function:

$$\mathrm{tr}(\mathcal{H}^{-1}) = \sum_i \langle z_i, \mathcal{H}^{-1} z_i \rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i$$

$$\underset{\boldsymbol{w} \in [0,1]^{n_s}}{minimize} \sum_i \langle z_i, y_i \rangle + \gamma P(\boldsymbol{w})$$

subject to

$$-\nabla \cdot (e^m \nabla u) = f \qquad \textit{(state)}$$

$$-\nabla \cdot (e^m \nabla p) = -\frac{1}{\sigma^2_{noise}} \mathcal{B}^* \mathbf{W} (\mathcal{B} u - \boldsymbol{d}) \qquad \textit{(adjoint)}$$

$$\mathcal{C}_{pr}^{-1}(m - m_{prior}) + e^m \nabla u \cdot \nabla p = 0 \qquad \textit{(gradient = 0)}$$

$$-\nabla \cdot (e^m \nabla v_i) = \nabla \cdot (y_i e^m \nabla u) \qquad \textit{(inc. state)}$$

$$-\nabla \cdot (e^m \nabla q_i) = \nabla \cdot (y_i e^m \nabla p) - \frac{1}{\sigma^2_{noise}} \mathcal{B}^* \mathbf{W} \mathcal{B} v_i \qquad \textit{(inc. adjoint)}$$

$$\mathcal{C}_{pr}^{-1} y_i + y_i e^m \nabla u \cdot \nabla p + e^m \left(\nabla v_i \cdot \nabla p + \nabla u \cdot \nabla q_i \right) = z_i \qquad \textit{(Hessian solve)}$$

Optimization problem for computing A-optimal design

OED objective function:

$$\text{tr}(\mathcal{H}^{-1}) = \sum_i \langle z_i, \mathcal{H}^{-1} z_i \rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i$$

$$\underset{\boldsymbol{w}\in[0,1]^{n_s}}{\textit{minimize}} \sum_i \langle z_i, y_i \rangle + \gamma P(\boldsymbol{w})$$

subject to

$$-\nabla \cdot (e^m \nabla u) = f \qquad \textit{(state)}$$

$$-\nabla \cdot (e^m \nabla p) = -\frac{1}{\sigma^2_{noise}} \mathcal{B}^* \mathbf{W} (\mathcal{B} u - \boldsymbol{d}) \qquad \textit{(adjoint)}$$

$$\mathcal{C}_{pr}^{-1}(m - m_{prior}) + e^m \nabla u \cdot \nabla p = 0 \qquad \textit{(gradient = 0)}$$

$$-\nabla \cdot (e^m \nabla v_i) = \nabla \cdot (y_i e^m \nabla u) \qquad \textit{(inc. state)}$$

$$-\nabla \cdot (e^m \nabla q_i) = \nabla \cdot (y_i e^m \nabla p) - \frac{1}{\sigma^2_{noise}} \mathcal{B}^* \mathbf{W} \mathcal{B} v_i \qquad \textit{(inc. adjoint)}$$

$$\mathcal{C}_{pr}^{-1} y_i + y_i e^m \nabla u \cdot \nabla p + e^m \left(\nabla v_i \cdot \nabla p + \nabla u \cdot \nabla q_i \right) = z_i \qquad \textit{(Hessian solve)}$$

Optimization problem for computing A-optimal design

OED objective function:

$$\mathrm{tr}(\mathcal{H}^{-1}) = \sum_i \langle z_i, \mathcal{H}^{-1} z_i \rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i$$

$$\underset{\boldsymbol{w}\in[0,1]^{n_s}}{minimize} \sum_i \langle z_i, y_i \rangle + \gamma P(\boldsymbol{w})$$

subject to

$$-\nabla\cdot(e^m \nabla u) = f \qquad (state)$$

$$-\nabla\cdot(e^m \nabla p) = -\frac{1}{\sigma^2_{noise}} \mathcal{B}^* \mathbf{W}(\mathcal{B}u - \boldsymbol{d}) \qquad (adjoint)$$

$$\mathcal{C}_{pr}^{-1}(m - m_{prior}) + e^m \nabla u \cdot \nabla p = 0 \qquad (gradient = 0)$$

$$-\nabla\cdot(e^m \nabla v_i) = \nabla\cdot(y_i e^m \nabla u) \qquad (inc.\ state)$$

$$-\nabla\cdot(e^m \nabla q_i) = \nabla\cdot(y_i e^m \nabla p) - \frac{1}{\sigma^2_{noise}} \mathcal{B}^* \mathbf{W} \mathcal{B} v_i \qquad (inc.\ adjoint)$$

$$\mathcal{C}_{pr}^{-1} y_i + y_i e^m \nabla u \cdot \nabla p + e^m \left(\nabla v_i \cdot \nabla p + \nabla u \cdot \nabla q_i \right) = z_i \qquad (Hessian\ solve)$$

Optimization problem for computing A-optimal design

OED objective function:

$$\mathrm{tr}(\mathcal{H}^{-1}) = \sum_i \langle z_i, \mathcal{H}^{-1} z_i \rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i$$

$$\underset{\boldsymbol{w}\in[0,1]^{n_s}}{\text{minimize}} \sum_i \langle z_i, y_i \rangle + \gamma P(\boldsymbol{w})$$

subject to

$$-\nabla\cdot(e^m \nabla u) = f \qquad \textit{(state)}$$

$$-\nabla\cdot(e^m \nabla p) = -\frac{1}{\sigma^2_{noise}} \mathcal{B}^* \mathbf{W}(\mathcal{B}u - \boldsymbol{d}) \qquad \textit{(adjoint)}$$

$$\mathcal{C}_{pr}^{-1}(m - m_{prior}) + e^m \nabla u \cdot \nabla p = 0 \qquad \textit{(gradient = 0)}$$

$$-\nabla\cdot(e^m \nabla v_i) = \nabla\cdot(y_i e^m \nabla u) \qquad \textit{(inc. state)}$$

$$-\nabla\cdot(e^m \nabla q_i) = \nabla\cdot(y_i e^m \nabla p) - \frac{1}{\sigma^2_{noise}} \mathcal{B}^* \mathbf{W} \mathcal{B} v_i \qquad \textit{(inc. adjoint)}$$

$$\mathcal{C}_{pr}^{-1} y_i + y_i e^m \nabla u \cdot \nabla p + e^m \left(\nabla v_i \cdot \nabla p + \nabla u \cdot \nabla q_i \right) = z_i \qquad \textit{(Hessian solve)}$$

The Lagrangian for the OED problem

- Consider the objective function $\Theta(\boldsymbol{w}) = \langle y, z \rangle, \quad \mathcal{H} y = z$

$$
\begin{aligned}
&\mathscr{L}^E(\boldsymbol{w}; u, m, p, v, q, y; u^*, m^*, p^*, v^*, q^*, y^*) \\
&= \langle z, y \rangle \\
&+ \langle e^m \nabla u, \nabla u^* \rangle - \langle f, u^* \rangle \\
&+ \langle e^m \nabla p, \nabla p^* \rangle + \sigma_{noise}^{-2} \langle \mathcal{B}^* \mathbf{W} (\mathcal{B} u - \boldsymbol{d}), p^* \rangle \\
&+ \left\langle \mathcal{C}_{pr}^{-1} (m - m_0), m^* \right\rangle + \langle m^* e^m \nabla u, \nabla p \rangle \\
&+ \langle e^m \nabla v, \nabla v^* \rangle + \langle y e^m \nabla u, \nabla v^* \rangle \\
&+ \langle e^m \nabla q, \nabla q^* \rangle + \langle y e^m \nabla p, \nabla q^* \rangle + \sigma_{noise}^{-2} \langle \mathcal{B}^* \mathbf{W} \mathcal{B} v, q^* \rangle \\
&+ \langle y^* e^m \nabla v, \nabla p \rangle + \left\langle y^*, \mathcal{C}_{pr}^{-1} y \right\rangle + \langle y^* e^m \nabla u, \nabla q \rangle + (y^* y e^m \nabla u, \nabla p) \\
&- \langle z, y^* \rangle
\end{aligned}
$$

The adjoint problem and the gradient

- The adjoint problem for $u^*, m^*, p^*, v^*, q^*, y^*$

$$
\begin{aligned}
\sigma_{noise}^{-2}\mathcal{B}^*\mathbf{W}\mathcal{B}q^* - \nabla\cdot(y^* e^m \nabla p) - \nabla\cdot(e^m \nabla v^*) &= 0\\
e^m \nabla q^* \cdot \nabla p + \mathcal{C}_{pr}^{-1} y^* + y^* e^m \nabla u \cdot \nabla p + e^m \nabla u \cdot \nabla v^* &= -z\\
-\nabla\cdot(e^m \nabla q^*) - \nabla\cdot(y^* e^m \nabla u) &= 0\\
\sigma_{noise}^{-2}\mathcal{B}^*\mathbf{W}\mathcal{B}p^* - \nabla\cdot(m^* e^m \nabla p) - \nabla\cdot(e^m \nabla u^*) &= b_1\\
e^m \nabla p^* \cdot \nabla p + \mathcal{C}_{pr}^{-1} m^* + m^* e^m \nabla u \cdot \nabla p + e^m \nabla u \cdot \nabla u^* &= b_2\\
-\nabla\cdot(e^m \nabla p^*) - \nabla\cdot(m^* e^m \nabla u) &= b_3
\end{aligned}
$$

- u, m, p, q, v, y, and q^*, v^*, y^* appear in expressions for b_1, b_2, b_3
- Gradient:

$$
g(\boldsymbol{w}) = \boldsymbol{\Gamma}_{\text{noise}}^{-1}(\mathcal{B}u - \boldsymbol{d}) \odot \mathcal{B}p^* + \boldsymbol{\Gamma}_{\text{noise}}^{-1}\mathcal{B}v \odot \mathcal{B}q^*
$$

- q^*, y^*, and v^* can be eliminated:

$$
q^* = -v, \quad y^* = -y, \quad v^* = -q
$$

The OED objective function and the gradient

- Solve inner optimization $\Longrightarrow$ $m(\boldsymbol{w})$, $u(m(\boldsymbol{w}))$ and $p(m(\boldsymbol{w}))$
- Solve for (v, y, q):

$$
\begin{aligned}
\sigma_{noise}^{-2}\mathcal{B}^*\mathbf{W}\mathcal{B}v - \nabla\cdot(ye^m\nabla p) - \nabla\cdot(e^m\nabla q) &= 0\\
e^m\nabla v\cdot\nabla p + \mathcal{C}_{\text{pr}}^{-1}y + ye^m\nabla u\cdot\nabla p + e^m\nabla u\cdot\nabla q &= z\\
-\nabla\cdot(e^m\nabla v) - \nabla\cdot(ye^m\nabla u) &= 0
\end{aligned}
$$

- Objective function evaluation: $\Theta(\boldsymbol{w}) = \langle y, z\rangle$
- Solve for (p^*, m^*, u^*):

$$
\begin{aligned}
\sigma_{noise}^{-2}\mathcal{B}^*\mathbf{W}\mathcal{B}p^* - \nabla\cdot(m^*e^m\nabla p) - \nabla\cdot(e^m\nabla u^*) &= b_1\\
e^m\nabla p^*\cdot\nabla p + \mathcal{C}_{\text{pr}}^{-1}m^* + m^*e^m\nabla u\cdot\nabla p + e^m\nabla u\cdot\nabla u^* &= b_2\\
-\nabla\cdot(e^m\nabla p^*) - \nabla\cdot(m^*e^m\nabla u) &= b_3
\end{aligned}
$$

- Gradient: $g(\boldsymbol{w}) = \boldsymbol{\Gamma}_{\text{noise}}^{-1}(\mathcal{B}u - \boldsymbol{d}) \odot \mathcal{B}p^* - \boldsymbol{\Gamma}_{\text{noise}}^{-1}\mathcal{B}v \odot \mathcal{B}v$

Computational cost: the number of PDE solves

r: rank of misfit Hessian

- Independent of mesh
- Weak dependence on sensor dimension

Cost of evaluating the OED objective function and gradient

1. Cost of solving the inner optimization $\sim 2 \times r \times \#\text{Newton iterations}$
2. Cost of computing OED objective $\sim 2 \times r \times n_{\text{tr}}$
3. Cost of computing OED gradient $\sim 2 \times n_{\text{tr}} + 2 \times r$

Low-rank approx to misfit Hessian $\Longrightarrow$ Cost in (2) and (3) $\sim 2 \times r$

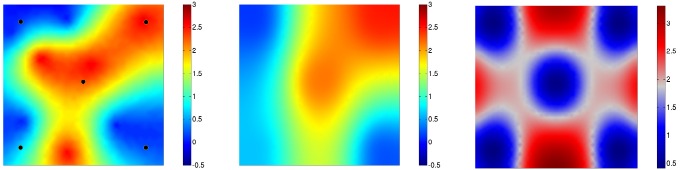
OED optimization problem:

- Solved via quasi-Newton interior point
- Number of interior point iterations insensitive to parameter/sensor dimension

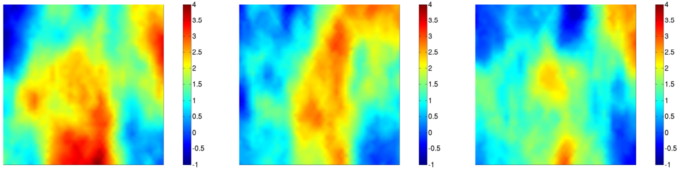
The prior and training models

- Prior knowledge: parameter value at few points and correlation information

truth   prior mean   prior variance



- Prior mean m_{prior} is a "smooth" least-squares fit to point measurements
- Prior covariance: $\mathcal{C}_{\text{pr}} = (\mathcal{A} + \alpha \sum_{i=1}^{N} \delta_i)^{-2}$, $\qquad \mathcal{A}m = -\nabla \cdot (\mathbf{D}\nabla m)$
- Draws from prior used to generate training data for OED



Computing an optimal design with sparsification

- Optimization problem with sparsity control

$$\underset{\boldsymbol{w}}{\text{minimize}} \sum_i \langle z_i, y_i(\boldsymbol{w})\rangle + \gamma P_\varepsilon(\boldsymbol{w})$$

where

$$\mathcal{H}\big(m_{\text{MAP}}(\boldsymbol{w})\big) y_i = z_i$$

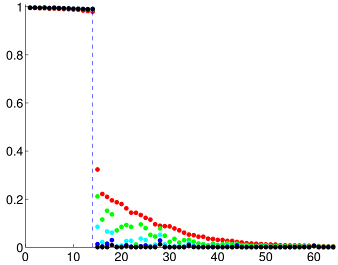
$$m_{\text{MAP}}(\boldsymbol{w}) = \arg\min_\theta \mathcal{J}\Big(u(\theta(\boldsymbol{w})), \theta(\boldsymbol{w})\Big)$$

- Continuation approach to approximate ℓ_0-penalty

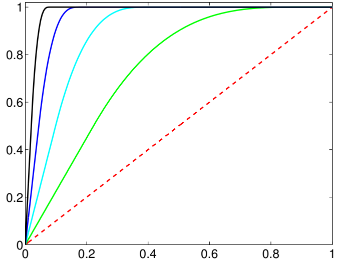
$$P_\varepsilon(\boldsymbol{w}) := \sum_{i=1}^{n_s} f_\varepsilon(w_i)$$

- Solve the problem with successively smaller values of ε

The optimal weight vector $\boldsymbol{w}$

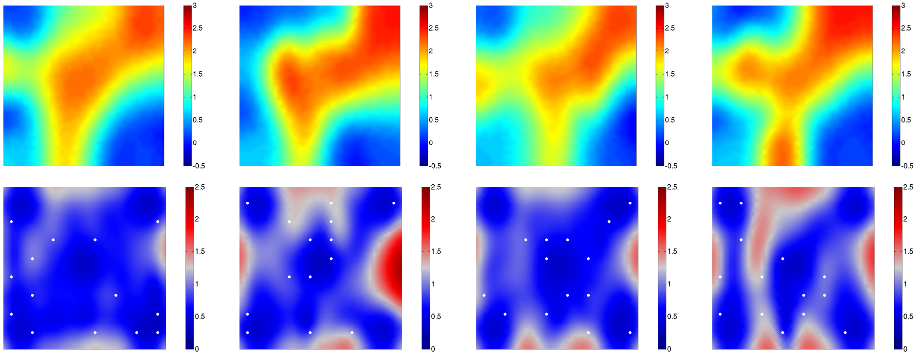


The functions f_ε

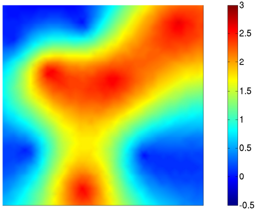


Effectiveness of the optimal design

Optimal | Random

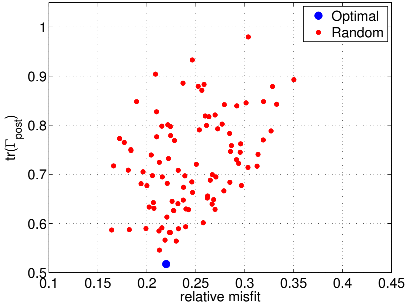


Truth



- Designs with 14 sensors
- Test quality of inferring the true field
- Comparing MAP point (top row) and posterior variance (Bottom row)

Optimal vs random designs

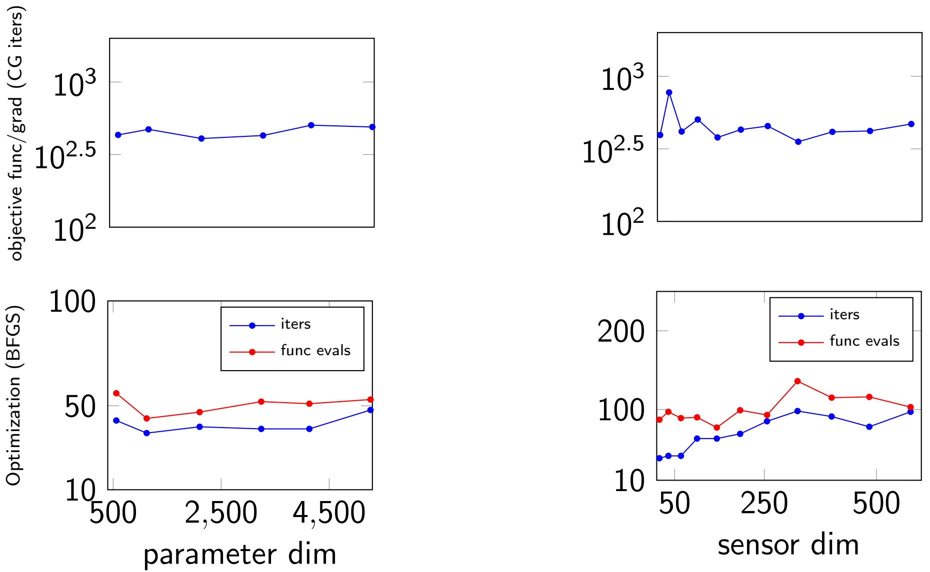


$$\text{relative misfit} = \frac{\| m_{\text{MAP}}(\boldsymbol{w}) - m_{\text{true}} \|}{\| m_{\text{true}} \|}$$

Sources of sub-optimality:

- Trace estimation
- Use of training data for OED
- Gaussian approximation

Scalability with respect to parameter/sensor dimension



Summary

- What is optimal experimental design in inverse problems?
- Design criteria?
- Linear and linearized problems
- Extensions to problems with model uncertainty and nonlinear problems
- Main challenges: Interpretation for nonlinear problems; computational cost,. . .