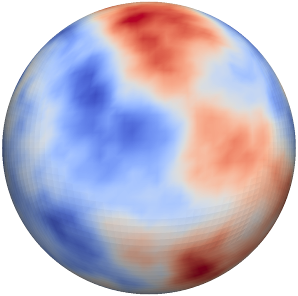
Scalable algorithms for optimal experimental design for infinite-dimensional nonlinear Bayesian inverse problems

Alen Alexanderian (Math/NC State), *Omar Ghattas* (ICES/UT-Austin),
Noémi Petra (Applied Math/UC Merced), Georg Stadler (Courant/NYU)

Data Assimilation and Inverse Problems
Mathematics Institute, University of Warwick
Coventry, United Kingdom

February 23, 2016

The conceptual optimal experimental design problem

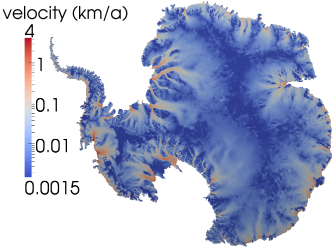


- Data: $\boldsymbol{d} \in \mathbb{R}^q$
- Model parameter field: $m = \{m(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathcal{D}}$
- Bayesian inverse problem:

  Use data $\mathbf{d}$ to infer the model parameter field m in the Bayesian sense, i.e., update the prior state of knowledge on m

- Optimal experimental design:

  How to design observation system for $\mathbf{d}$ to "optimally" infer m?



Some relevant references on OED for PDEs

- E. Haber, L. Horesh, L. Tenorio, Numerical methods for experimental design of large-scale linear ill-posed inverse problems, *Inverse Problems*, 24:125-137, 2008.
- E. Haber, L. Horesh, L. Tenorio, Numerical methods for the design of large-scale nonlinear discrete ill-posed inverse problems, *Inverse Problems*, 26, 025002, 2010.
- E. Haber, Z. Magnant, C. Lucero, L. Tenorio, Numerical methods for A-optimal designs with a sparsity constraint for ill-posed inverse problems, *Computational Optimization and Applications*, 1-22, 2012.
- Q. Long, M. Scavino, R. Tempone, S Wang, Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations, *Comp. Meth. in Appl. Mech. and Eng.*, 259:24–39, 2013.
- Q. Long, M. Scavino, R. Tempone, S. Wang, A Laplace method for under-determined Bayesian optimal experimental designs, *Comp. Meth. in Applied Mech. and Engineering*, 285:849–876, 2015.
- F. Bisetti, D. Kim, O. Knio, Q. Long, R. Tempone, Optimal Bayesian experimental design for priors of compact support with application to shocktube experiments for combustion kinetics, *Intl. Journal for Num. Methods in Engineering*, 2016.
- X. Huan and Y.M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, *Journal of Computational Physics*, 232:288-317, 2013.
- X. Huan, Y. Marzouk, Gradient-based stochastic optimization methods in Bayesian experimental design, *Intl. Journal for Uncertainty Quantification*, 4(6), 2014.

How we define an optimal design

- Given locations $\boldsymbol{x}_i$ of possible sensor sites, choose optimal weights w_i for a subset (using sparsity constraints):

$$\text{design} := \left\{ \begin{array}{c} \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_s} \\ w_1, \ldots, w_{n_s} \end{array} \right\}$$

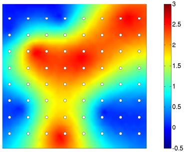
- Obtain data, solve Bayesian inverse problem:

$$\text{data} + \text{likelihood, prior} \Longrightarrow \text{posterior}$$

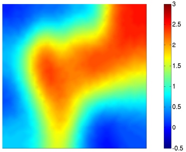
- We seek an A-optimal design:

$$\text{minimize average posterior variance}$$

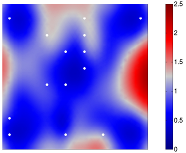
ground truth + sensor locations



MAP solution



variance



Bayesian inverse problem in Hilbert space (A. Stuart, 2010)

- $\mathcal{D}$: bounded domain $\qquad \mathcal{H} = L^2(\mathcal{D}) \qquad m \in \mathcal{H}$: parameter
- Parameter-to-observable map: $\boldsymbol{f} : \mathcal{H} \to \mathbb{R}^q$
- Additive Gaussian noise:

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

- Likelihood:

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

- Gaussian prior measure: $\mu_{\mathsf{prior}} = \mathcal{N}\left(m_{\mathsf{pr}}, \mathcal{C}_{\text{prior}}\right)$
- Bayes Theorem (in infinite dimensions):

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

A.M. Stuart, Inverse problems: A Bayesian perspective, *Acta Numerica*, 2010.

The experimental design and a weighted inference problem

$$\text{design} := \left\{ \begin{array}{c} \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_s} \\ w_1, \ldots, w_{n_s} \end{array} \right\}$$

- Want $w_i \in \{0,1\} \quad \overset{\text{relax}}{\Longrightarrow} \quad 0 \le w_i \le 1 \quad \overset{\text{threshold}}{\Longrightarrow} \quad w_i \in \{0,1\}$
- $\boldsymbol{w}$-weighted data-likelihood:

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

- $\boldsymbol{f}$: parameter-to-observable map
- $\boldsymbol{W} := \text{diag}(w_i)$
- $\boldsymbol{W}_\sigma := \frac{1}{\sigma^2_{\text{noise}}}\boldsymbol{W}$

A-optimal design in Hilbert space

- Covariance function: $c_{\text{post}}(\boldsymbol{x}, \boldsymbol{y}) = \text{Cov}\{m(\boldsymbol{x}), m(\boldsymbol{y})\}$

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

- Covariance operator:

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

- Mercer's Theorem:

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

- A-optimal design criterion:

Choose a sensor configuration to minimize $\text{tr}(\mathcal{C}_{\text{post}})$

Relationship to some other OED criteria

For Gaussian prior and noise, and linear parameter-to-observable map:

- Maximizing the *expected information gain* is equivalent to *D-optimal design*:

$$\mathsf{E}_{\mu_{\text{prior}}}\left\{\mathsf{E}_{\boldsymbol{d}|\boldsymbol{m}}\left\{\mathrm{D_{kl}}(\mu_{\text{post}},\mu_{\text{prior}})\right\}\right\} = -\frac{1}{2}\log\det\boldsymbol{\Gamma}_{\text{post}} + \frac{1}{2}\log\det\boldsymbol{\Gamma}_{\text{prior}}$$

- Minimizing *Bayes risk of posterior mean* is equivalent to *A-optimal design*:

$$\begin{aligned}\mathsf{E}_{\mu_{\text{prior}}}\left\{\mathsf{E}_{\boldsymbol{d}|\boldsymbol{m}}\left\{\|\boldsymbol{m}-\boldsymbol{m}_{\text{post}}(\boldsymbol{d})\|^2\right\}\right\} &= \int_{\mathcal{H}}\int_{\mathcal{Y}}\|\boldsymbol{m}-\boldsymbol{m}_{\text{post}}(\boldsymbol{d})\|^2\pi_{\text{like}}(\boldsymbol{d}|\boldsymbol{m})\,d\boldsymbol{d}\;\mu_{\text{prior}}(d\boldsymbol{m})\\ &= \mathrm{tr}(\boldsymbol{\Gamma}_{\text{post}})\end{aligned}$$

For infinite dimensions, see:

A. Alexanderian, P. Gloor, and O. Ghattas, On Bayesian A- and D-optimal experimental designs in infinite dimensions, *Bayesian Analysis*, to appear, 2016.
`http://dx.doi.org/10.1214/15-BA969`

Challenges

- The infinite-dimensional Bayesian inverse problem is “merely” an inner problem within OED
- Need trace of posterior covariance operator, which would be intractable in the large-scale setting
- We approximate the posterior covariance by the inverse of the Hessian at the maximum a posteriori (MAP) point, but explicit construction of the Hessian is still very expensive
- For nonlinear inverse problem, Hessian depends on data, which are not available *a priori*
- For efficient optimization, we need derivatives of the trace of the inverse Hessian with respect to the sensor weights
- We seek scalable algorithms, i.e., those whose cost (in terms of forward PDE solves) is independent of the data dimension and the parameter dimension

A-optimal design problem: Linear case

- Linear parameter-to-observable map: $\boldsymbol{f}(m) := \mathcal{F}m$
- Gaussian prior measure: $\mu_{\text{prior}} = \mathcal{N}\left(m_{\text{prior}}, \mathcal{C}_{\text{prior}}\right)$
- Gaussian prior and noise $\Longrightarrow$ Gaussian posterior

$$m_{\text{post}} = \arg\min_{m} \underbrace{\frac{1}{2}\left\|\boldsymbol{W}^{1/2}(\mathcal{F}m - \boldsymbol{d})\right\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} + \frac{1}{2}\left\langle \mathcal{C}^{-1}_{\text{prior}}(m - m_{\text{prior}}), m - m_{\text{prior}} \right\rangle}_{\mathcal{J}(m) := \text{negative log-posterior}}$$

$$\mathcal{C}_{\text{post}}(\boldsymbol{w}) = \mathcal{H}^{-1}(\boldsymbol{w}), \qquad \mathcal{H}(\boldsymbol{w}) = \mathcal{F}^*\boldsymbol{W}_\sigma\mathcal{F} + \mathcal{C}^{-1}_{\text{prior}}$$

- OED problem:

$$\underset{\boldsymbol{w}}{\text{minimize}}\ \text{tr}\left[\left(\mathcal{F}^*\boldsymbol{W}_\sigma\mathcal{F} + \mathcal{C}^{-1}_{\text{prior}}\right)^{-1}\right] + \text{penalty}$$

A. Alexanderian, N. Petra, G. Stadler, O. Ghattas, A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized ℓ_0-sparsification, *SIAM Journal on Scientific Computing*, 36(5):A2122–A2148, 2014.

A-optimal design problem: Nonlinear case

- Nonlinear case: 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 $m_{\text{MAP}}(\boldsymbol{w};\boldsymbol{d})$

$$m_{\text{MAP}}(\boldsymbol{w};\boldsymbol{d}) := \arg\min_{m} \mathcal{J}(m,\boldsymbol{w},\boldsymbol{d})$$

  - Gaussian approximation to posterior at MAP point

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

- Problem: data $\boldsymbol{d}$ not available *a priori*

A-optimal designs for nonlinear inverse problems

- General formulation: minimize average posterior variance:

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

A-optimal designs for nonlinear inverse problems

- General formulation: minimize expected average posterior variance:

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

A-optimal designs for nonlinear inverse problems

- General formulation: minimize expected average posterior variance:

$$\underset{\boldsymbol{w}}{\text{minimize}} \;\; \mathsf{E}_{\mu_{\text{prior}}} \mathsf{E}_{\boldsymbol{d}|m} \left\{ \text{tr}\left(\mathcal{H}^{-1} \left[m_{\text{MAP}}(\boldsymbol{w};\boldsymbol{d}), \boldsymbol{w}; \boldsymbol{d} \right] \right) \right\}$$

A-optimal designs for nonlinear inverse problems

- General formulation: minimize expected average posterior variance:

$$\underset{\boldsymbol{w}}{\text{minimize}} \;\; \mathsf{E}_{\mu_{\text{prior}}} \mathsf{E}_{\boldsymbol{d}|m} \left\{ \text{tr}\left(\mathcal{H}^{-1} \left[m_{\text{MAP}}(\boldsymbol{w};\boldsymbol{d}), \boldsymbol{w}; \boldsymbol{d} \right] \right) \right\} + \gamma P(\boldsymbol{w})$$

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

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

A-optimal designs for nonlinear inverse problems

- General formulation: minimize expected average posterior variance:

$$\underset{\boldsymbol{w}}{\text{minimize}} \ \mathsf{E}_{\mu_{\text{prior}}} \mathsf{E}_{\boldsymbol{d}|m} \left\{ \text{tr}\left(\mathcal{H}^{-1} \left[m_{\text{MAP}}(\boldsymbol{w}; \boldsymbol{d}), \boldsymbol{w}; \boldsymbol{d} \right] \right) \right\} + \gamma P(\boldsymbol{w})$$

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

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

- The problem to solve in practice:

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

Randomized trace estimation

- $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ — symmetric
- Trace estimator:

$$\mathrm{tr}(\boldsymbol{A}) \approx \frac{1}{n_{\mathrm{tr}}} \sum_{i=1}^{n_{\mathrm{tr}}} \boldsymbol{z}_i^T \boldsymbol{A} \boldsymbol{z}_i, \quad \boldsymbol{z}_i \text{ — random vectors}$$

- Gaussian trace estimator: $\boldsymbol{z}_i$ independent draws from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
- For $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$

$$\mathrm{E}\left\{\boldsymbol{z}^T \boldsymbol{A} \boldsymbol{z}\right\} = \mathrm{tr}(\boldsymbol{A}) \qquad \mathrm{Var}\left\{\boldsymbol{z}^T \boldsymbol{A} \boldsymbol{z}\right\} = 2 \left\|\boldsymbol{A}\right\|_F^2$$

- Clustered eigenvalues $\Longrightarrow$ Good approximation with small n_{tr}

Efficient means of approximating trace of inverse Hessian

M. Hutchinson, A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines (1990).

H. Avron and S. Toledo, Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix (2011).

Optimization problem for computing A-optimal design

- OED objective function:

$$\Psi(\boldsymbol{w}) = \frac{1}{n_{\mathsf{d}}} \sum_{i=1}^{n_{\mathsf{d}}} \mathsf{tr}\left[\mathcal{H}^{-1}\big(\boldsymbol{w}; \boldsymbol{d}_i\big)\right]$$

Optimization problem for computing A-optimal design

- OED objective function:

$$\Psi(\boldsymbol{w}) = \frac{1}{n_{\mathsf{d}}\, n_{\mathsf{tr}}} \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \left\langle z_k, \mathcal{H}^{-1}\big(\boldsymbol{w}; \boldsymbol{d}_i\big)\, z_k \right\rangle$$

Optimization problem for computing A-optimal design

- OED objective function:

$$\Psi(\boldsymbol{w}) = \frac{1}{n_{\mathsf{d}}\, n_{\mathsf{tr}}} \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \langle\, z_k, \underbrace{\mathcal{H}^{-1}\big(\boldsymbol{w}; \boldsymbol{d}_i\big)\, z_k}_{y_{ik}} \,\rangle$$

- Auxiliary variable: $y_{ik} = \mathcal{H}^{-1}\big(\boldsymbol{w}; \boldsymbol{d}_i\big) z_k$

Optimization problem for computing A-optimal design

- OED objective function:

$$\Psi(\boldsymbol{w}) = \frac{1}{n_{\mathsf{d}}\, n_{\mathsf{tr}}} \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \langle\, z_k, \underbrace{\mathcal{H}^{-1}\big(\boldsymbol{w}; \boldsymbol{d}_i\big)\, z_k}_{y_{ik}} \,\rangle$$

- Auxiliary variable: $y_{ik} = \mathcal{H}^{-1}\big(\boldsymbol{w}; \boldsymbol{d}_i\big) z_k$

OED optimization problem

$$\underset{\boldsymbol{w}}{\text{minimize}}\ \frac{1}{n_{\mathsf{d}}\, n_{\mathsf{tr}}} \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \langle z_k, y_{ik} \rangle + \gamma P(\boldsymbol{w})$$

where

$$m_{\mathsf{MAP}}(\boldsymbol{w}; \boldsymbol{d}_i) = \arg\min_{m} \mathcal{J}\Big(m, \boldsymbol{w}; \boldsymbol{d}_i\Big), \quad i = 1, \dots, n_{\mathsf{d}}$$

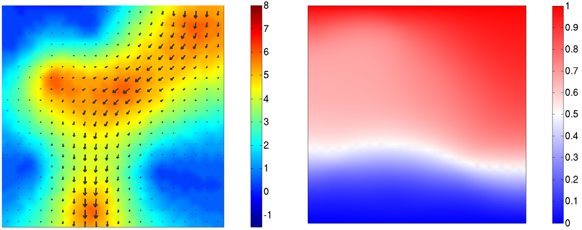
$$\mathcal{H}\Big(m_{\mathsf{MAP}}(\boldsymbol{w}; \boldsymbol{d}_i), \boldsymbol{w}; \boldsymbol{d}_i\Big)\, y_{ik} = z_k, \quad i = 1, \dots, n_{\mathsf{d}},\ k = 1, \dots, n_{\mathsf{tr}}$$

Application: subsurface flow

- Forward problem

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

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



Left: true parameter; Right: pressure-field

Bayesian inverse problem: Gaussian approximation

MAP point is solution to

$$\underset{m}{\text{minimize}}\ \mathcal{J}(m, \boldsymbol{w}; \boldsymbol{d}) := \frac{1}{2}(\mathcal{B}u - \boldsymbol{d})^T \boldsymbol{W}_\sigma (\mathcal{B}u - \boldsymbol{d}) + \frac{1}{2}\left\langle m - m_{prior}, m - m_{prior} \right\rangle_{\mathcal{E}}$$

where

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

- $\mathcal{B}$ is observation operator
- Cameron-Martin inner-product:

$$\langle m_1, m_2 \rangle_{\mathcal{E}} = \left\langle \mathcal{C}_{\text{prior}}^{-1/2} m_1, \mathcal{C}_{\text{prior}}^{-1/2} m_2 \right\rangle, \quad m_1, m_2 \in \text{range}(\mathcal{C}_{\text{prior}}^{1/2})$$

Covariance of Gaussian approximation to posterior:

$$\mathcal{C} = \mathcal{H}(m, \boldsymbol{w}; \boldsymbol{d})^{-1} = [\nabla_m^2 \mathcal{J}(m, \boldsymbol{w}; \boldsymbol{d})]^{-1}$$

with $m = m_{\text{MAP}}(\boldsymbol{w}; \boldsymbol{d})$

Optimization problem for A-optimal design

$$\underset{\boldsymbol{w}\in[0,1]^{n_s}}{\text{minimize}} \; \frac{1}{n_{\mathsf{d}} n_{\mathsf{tr}}} \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \langle z_k, y_{ik} \rangle + \gamma P(\boldsymbol{w})$$

where for $i = 1, \ldots, n_{\mathsf{d}}, \; k = 1, \ldots, n_{\mathsf{tr}}$

$$-\nabla\cdot(e^{m_i}\nabla u_i) = f$$

$$-\nabla\cdot(e^{m_i}\nabla p_i) = -\mathcal{B}^*\boldsymbol{W}_\sigma(\mathcal{B}u_i - \boldsymbol{d}_i)$$

$$\underbrace{\mathcal{C}_{\text{prior}}^{-1}(m_i - m_{\text{pr}}) + e^{m_i}\nabla u_i \cdot \nabla p_i}_{\nabla_m \mathcal{J}(m_i)} = 0$$

$$-\nabla\cdot(e^{m_i}\nabla v_{ik}) = \nabla\cdot(y_{ik}e^{m_i}\nabla u_i)$$

$$-\nabla\cdot(e^{m_i}\nabla q_{ik}) = \nabla\cdot(y_{ik}e^{m_i}\nabla p_i) - \mathcal{B}^*\boldsymbol{W}_\sigma\mathcal{B}v_{ik}$$

$$\underbrace{\mathcal{C}_{\text{prior}}^{-1}y_{ik} + y_{ik}e^{m_i}\nabla u_i \cdot \nabla p_i + e^{m_i}\left(\nabla v_{ik}\cdot\nabla p_i + \nabla u_i \cdot \nabla q_{ik}\right)}_{\mathcal{H}(m_i)y_{ik}} = z_k$$

The Lagrangian for the OED problem

$$
\begin{aligned}
&\mathcal{L}^{\mathsf{OED}}\left(\boldsymbol{w}, \{u_i\}, \{m_i\}, \{p_i\}, \{v_{ik}\}, \{q_{ik}\}, \{y_{ik}\}, \{u_i^*\}, \{m_i^*\}, \{p_i^*\}, \{v_{ik}^*\}, \{q_{ik}^*\}, \{y_{ik}^*\}\right) \\
&:= \frac{1}{n_{\mathsf{d}} n_{\mathsf{tr}}} \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \langle z_k, y_{ik} \rangle + \gamma P(\boldsymbol{w}) \\
&+ \sum_{i=1}^{n_{\mathsf{d}}} \left[\left\langle e^{m_i} \nabla u_i, \nabla u_i^* \right\rangle - \left\langle f, u_i^* \right\rangle - \left\langle h, u_i^* \right\rangle_{\Gamma_N} \right] \\
&+ \sum_{i=1}^{n_{\mathsf{d}}} \left[\left\langle e^{m_i} \nabla p_i, \nabla p_i^* \right\rangle + \left\langle \mathcal{B}^* \boldsymbol{W}_\sigma (\mathcal{B} u_i - \boldsymbol{d}_i), p_i^* \right\rangle \right] \\
&+ \sum_{i=1}^{n_{\mathsf{d}}} \left[\left\langle (m_i - m_{\mathsf{pr}}), m_i^* \right\rangle_{\mathcal{E}} + \left\langle m_i^* e^{m_i} \nabla u_i, \nabla p_i \right\rangle \right] \\
&+ \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \left[\left\langle e^{m_i} \nabla v_{ik}, \nabla v_{ik}^* \right\rangle + \left\langle y_{ik} e^{m_i} \nabla u_i, \nabla v_{ik}^* \right\rangle \right] \\
&+ \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \left[\left\langle e^{m_i} \nabla q_{ik}, \nabla q_{ik}^* \right\rangle + \left\langle y_{ik} e^{m_i} \nabla p_i, \nabla q_{ik}^* \right\rangle + \left\langle \mathcal{B}^* \boldsymbol{W}_\sigma \mathcal{B} v_{ik}, q_{ik}^* \right\rangle \right] \\
&+ \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \left[\left\langle y_{ik}^* e^{m_i} \nabla v_{ik}, \nabla p_i \right\rangle + \left\langle y_{ik}^*, y_{ik} \right\rangle_{\mathcal{E}} + \left\langle y_{ik}^* e^{m_i} \nabla u_i, \nabla q_{ik} \right\rangle + (y_{ik}^* y_{ik} e^{m_i} \nabla u_i, \nabla p_i) \right. \\
&\qquad\qquad \left. - \left\langle z_k, y_{ik}^* \right\rangle \right]
\end{aligned}
$$

The adjoint OED problem and the gradient

- The adjoint OED problem for $\{u_i^*\}, \{m_i^*\}, \{p_i^*\}, \{v_{ik}^*\}, \{q_{ik}^*\}, \{y_{ik}^*\}$

$$\mathcal{B}^* \boldsymbol{W}_\sigma \mathcal{B} q_{ik}^* - \nabla\cdot(y_{ik}^* e^{m_i} \nabla p_i) - \nabla\cdot(e^{m_i} \nabla v_{ik}^*) = 0$$

$$e^{m_i} \nabla q_{ik}^* \cdot \nabla p_i + \mathcal{C}_{\text{prior}}^{-1} y_{ik}^* + y_{ik}^* e^{m} \nabla u_i \cdot \nabla p_i + e^{m_i} \nabla u_i \cdot \nabla v_{ik}^* = -\frac{1}{n_{\mathsf{d}} n_{\mathsf{tr}}} z_k$$

$$-\nabla\cdot(e^{m_i} \nabla q_{ik}^*) - \nabla\cdot(y_{ik}^* e^{m} \nabla u_i) = 0$$

$$\mathcal{B}^* \boldsymbol{W}_\sigma \mathcal{B} p_i^* - \nabla\cdot(m_i^* e^{m_i} \nabla p_i) - \nabla\cdot(e^{m_i} \nabla u_i^*) = b_i^1$$

$$e^{m_i} \nabla p_i^* \cdot \nabla p_i + \mathcal{C}_{\text{prior}}^{-1} m_i^* + m_i^* e^{m_i} \nabla u_i \cdot \nabla p_i + e^{m_i} \nabla u_i \cdot \nabla u_i^* = b_i^2$$

$$-\nabla\cdot(e^{m_i} \nabla p_i^*) - \nabla\cdot(m_i^* e^{m_i} \nabla u_i) = b_i^3$$

- Gradient for the OED problem:

$$\boldsymbol{g}(\boldsymbol{w}) = \sum_{i=1}^{n_{\mathsf{d}}} \boldsymbol{\Gamma}_{\text{noise}}^{-1} (\mathcal{B} u_i - \boldsymbol{d}_i) \odot \mathcal{B} p_i^* + \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \boldsymbol{\Gamma}_{\text{noise}}^{-1} \mathcal{B} v_{ik} \odot \mathcal{B} q_{ik}^*,$$

- q_{ik}^*, y_{ik}^*, and v_{ik}^* can be eliminated:

$$q_{ik}^* = -v_{ik}, \quad y_{ik}^* = -y_{ik}, \quad v_{ik}^* = -q_{ik}$$

Computing the OED objective function and its gradient

- Solve inverse problems: $m_i(\boldsymbol{w})$, $u_i(m_i(\boldsymbol{w}))$ and $p_i(m_i(\boldsymbol{w}))$, $i = 1, \ldots, n_{\mathsf{d}}$
- Solve for $(v_{ik}, y_{ik}, q_{ik}), i = 1, \ldots, n_{\mathsf{d}}$, $k = 1, \ldots, n_{\mathsf{tr}}$:

$$
\begin{aligned}
-\nabla\cdot(e^{m_i}\nabla v_{ik}) &= \nabla\cdot(y_{ik}e^{m_i}\nabla u_i)\\
-\nabla\cdot(e^{m_i}\nabla q_{ik}) &= \nabla\cdot(y_{ik}e^{m_i}\nabla p_i) - \mathcal{B}^*\boldsymbol{W}_\sigma\mathcal{B}v_{ik}\\
\mathcal{C}_{\text{prior}}^{-1}y_{ik} + y_{ik}e^{m_i}\nabla u_i\cdot\nabla p_i + e^{m_i}\left(\nabla v_{ik}\cdot\nabla p_i + \nabla u_i\cdot\nabla q_{ik}\right) &= z_k
\end{aligned}
$$

- Objective function evaluation: $\Psi(\boldsymbol{w}) = \frac{1}{n_{\mathsf{d}}n_{\mathsf{tr}}}\sum_{i=1}^{n_{\mathsf{d}}}\sum_{k=1}^{n_{\mathsf{tr}}}\langle z_k, y_{ik}\rangle$
- Solve for (p_i^*, m_i^*, u_i^*), $i = 1, \ldots, n_{\mathsf{d}}$

$$
\begin{aligned}
\mathcal{B}^*\boldsymbol{W}_\sigma\mathcal{B}p_i^* - \nabla\cdot(m_i^*e^{m_i}\nabla p_i) - \nabla\cdot(e^{m_i}\nabla u_i^*) &= b_i^1\\
e^{m_i}\nabla p_i^*\cdot\nabla p_i + \mathcal{C}_{\text{prior}}^{-1}m_i^* + m_i^*e_i^m\nabla u_i\cdot\nabla p_i + e^{m_i}\nabla u_i\cdot\nabla u_i^* &= b_i^2\\
-\nabla\cdot(e^{m_i}\nabla p_i^*) - \nabla\cdot(m_i^*e_i^m\nabla u_i) &= b_i^3
\end{aligned}
$$

- Gradient:

$$
\boldsymbol{g}(\boldsymbol{w}) = \sum_{i=1}^{n_{\mathsf{d}}}\boldsymbol{\Gamma}_{\text{noise}}^{-1}(\mathcal{B}u_i - \boldsymbol{d}_i)\odot\mathcal{B}p_i^* - \frac{1}{n_{\mathsf{d}}n_{\mathsf{tr}}}\sum_{i=1}^{n_{\mathsf{d}}}\sum_{k=1}^{n_{\mathsf{tr}}}\boldsymbol{\Gamma}_{\text{noise}}^{-1}\mathcal{B}v_{ik}\odot\mathcal{B}v_{ik}
$$

Computational cost: the number of PDE solves

r: rank of Hessian of the data misfit

- Independent of mesh (beyond a certain resolution)
- Independent of sensor dimension (beyond a certain dimension)

Cost (in PDE solves) of evaluating OED objective function & gradient

1. Cost of solving inverse problems $\sim 2 \times r \times n_{\mathsf{d}} \times N_{\mathsf{newton}}$
2. Cost of computing OED objective $\sim 2 \times r \times n_{\mathsf{tr}} \times n_{\mathsf{d}}$
3. Cost of computing OED gradient $\sim 2 \times n_{\mathsf{tr}} \times n_{\mathsf{d}} + 2 \times r \times n_{\mathsf{d}}$

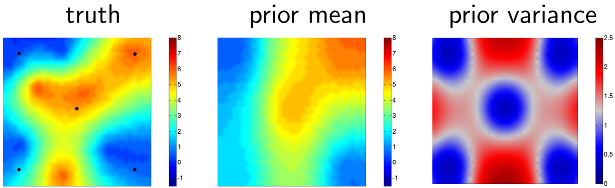
   Low-rank approx to misfit Hessian $\Longrightarrow$ Cost in (2) and (3) $\sim 2 \times r \times n_{\mathsf{d}}$

OED optimization problem:

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

The prior and training models

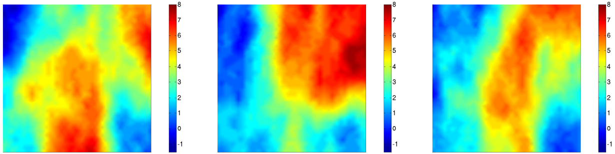
- Prior knowledge: parameter value at few points, correlation



- Prior mean: "smooth" least-squares fit to point measurements

$$m_{\mathsf{pr}} = \arg\min_{m} \frac{1}{2}\langle m, \mathcal{A}m \rangle + \frac{\alpha}{2} \sum_{i=1}^{N} \int_{\mathcal{D}} \delta_i(\boldsymbol{x}) \left[m(\boldsymbol{x}) - m_{\mathsf{true}}(\boldsymbol{x}) \right]^2 d\boldsymbol{x}$$

- Prior covariance: $\mathcal{C}_{\text{prior}} = (\mathcal{A} + \alpha \sum_{i=1}^{N} \delta_i)^{-2}, \qquad \mathcal{A}m = -\nabla \cdot (\boldsymbol{D} \nabla m)$
- Draws from prior used to generate training data for OED



- Optimization problem

$$\underset{\boldsymbol{w}}{\text{minimize}} \; \frac{1}{n_{\mathsf{d}}\, n_{\mathsf{tr}}} \sum_{i=1}^{n_{\mathsf{d}}} \sum_{k=1}^{n_{\mathsf{tr}}} \langle z_k, y_{ik} \rangle + \gamma P_\varepsilon(\boldsymbol{w})$$

where

$$m_{\mathsf{MAP}}(\boldsymbol{w}; \boldsymbol{d}_i) = \arg\min_m \mathcal{J}\Big(m, \boldsymbol{w}; \boldsymbol{d}_i\Big)$$

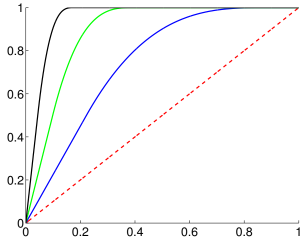
$$\mathcal{H}\big(m_{\mathsf{MAP}}(\boldsymbol{w}; \boldsymbol{d}_i), \boldsymbol{w}; \boldsymbol{d}_i\big) y_{ik} = z_k$$

- $n_{\mathsf{d}} = 5$, $n_{\mathsf{tr}} = 20$
- Continuation $\Rightarrow$ ℓ_0-penalty

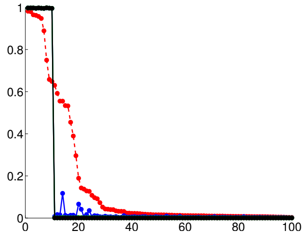
$$P_\varepsilon(\boldsymbol{w}) := \sum_{i=1}^{n_s} f_\varepsilon(w_i)$$

- Solve the problem with $\varepsilon \to 0$

The functions f_ε

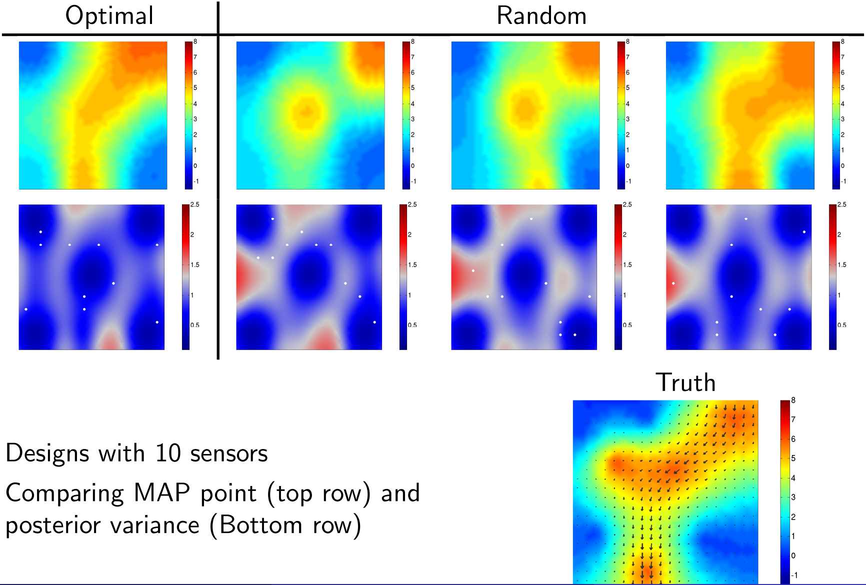


The optimal weight vector $\boldsymbol{w}$



Effectiveness of the optimal design

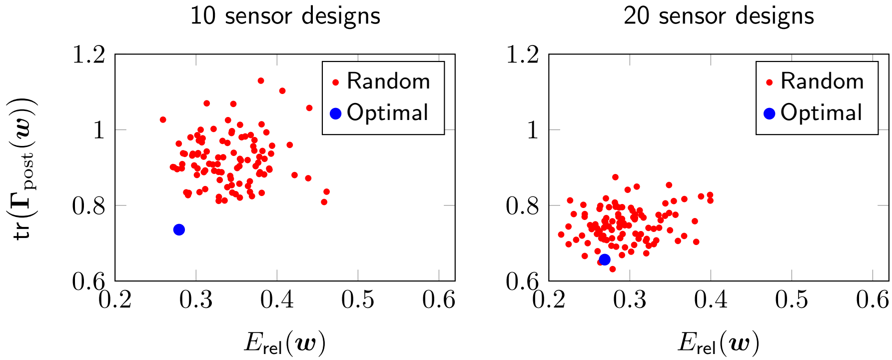
Numerical test: how well can we recover the “true model” from “true data”?



- Designs with 10 sensors
- Comparing MAP point (top row) and posterior variance (Bottom row)

Effectiveness of the optimal design

Numerical test: how well can we recover the "true model" from "true data"?

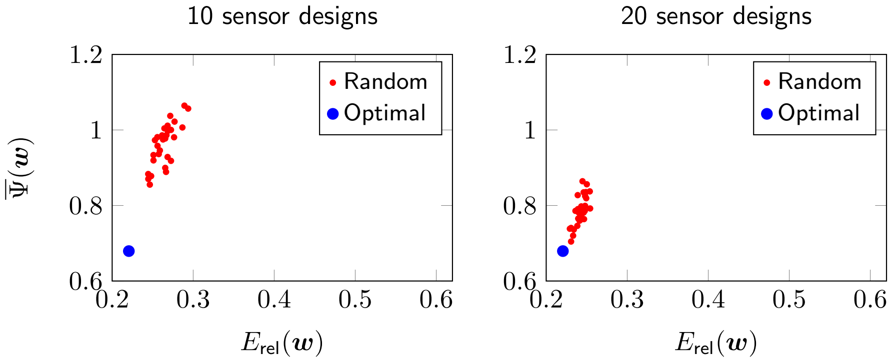


$$\boldsymbol{d} = f(m_{\mathsf{true}}) + \boldsymbol{\eta} \qquad \boldsymbol{\Gamma}_{\mathrm{post}}(\boldsymbol{w}) := \mathcal{H}^{-1}(m_{\mathsf{MAP}}(\boldsymbol{w};\boldsymbol{d}), \boldsymbol{w};\boldsymbol{d})$$

$$E_{\mathsf{rel}}(\boldsymbol{w}) := \frac{\|m_{\mathsf{MAP}}(\boldsymbol{w};\boldsymbol{d}) - m_{\mathsf{true}}\|}{\|m_{\mathsf{true}}\|}$$

Effectiveness of the optimal design

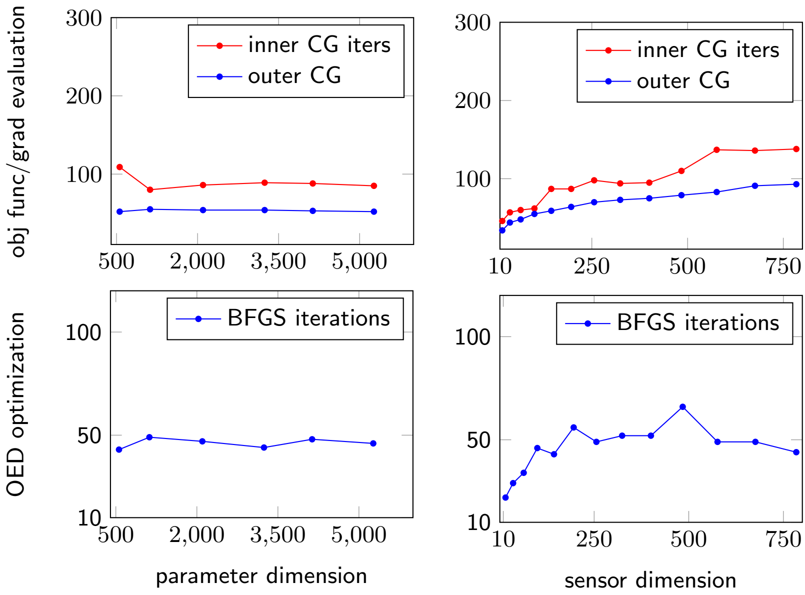
Numerical test: how well we do on average?



$$\overline{E}_{rel}(\boldsymbol{w}) = \frac{1}{n_d'} \sum_{i=1}^{n_d'} \frac{\|m_{\text{MAP}}(\boldsymbol{w}; \boldsymbol{d}_i) - m_i\|}{\|m_i\|} \qquad \overline{\Psi}(\boldsymbol{w}) = \frac{1}{n_d'} \sum_{i=1}^{n_d'} \text{tr}\big(\mathcal{H}^{-1}(m_{\text{MAP}}(\boldsymbol{w}; \boldsymbol{d}_i), \boldsymbol{w}; \boldsymbol{d}_i)\big)$$

$\boldsymbol{d}_i$ generated from $n_d' = 50$ draws of m_i from the prior
(different from the $n_{\text{d}} = 5$ used to solve the OED problem)

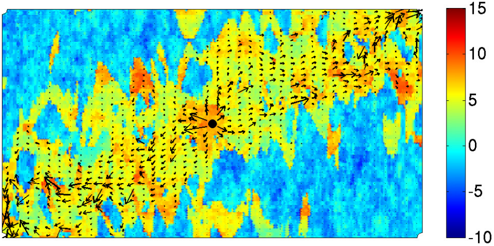
Scalability with respect to parameter/sensor dimension



Inversion of a realistic permeability field

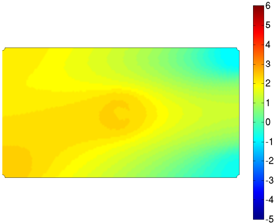
$$
\begin{aligned}
-\nabla \cdot (e^m \nabla u) &= f && \text{in } \mathcal{D} \\
u &= 0 && \text{on } \Gamma_D \\
e^m \nabla u \cdot \boldsymbol{n} &= 0 && \text{on } \Gamma_N
\end{aligned}
$$

- Γ_D: at the corners
- Γ_N: the rest of boundary
- Point source at center

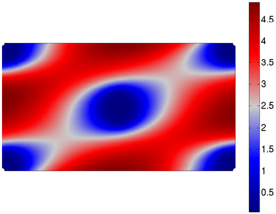


- Log-permeability field from SPE Comparative Solution Project
- Injection well at the center
- Production wells at four corners
- Parameter (and state) dimension: $n = 10,202$
- Potential sensor locations: $n_s = 128$

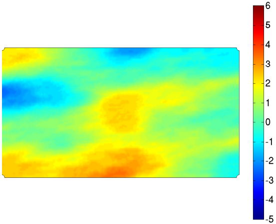
SPE model: the prior



Mean

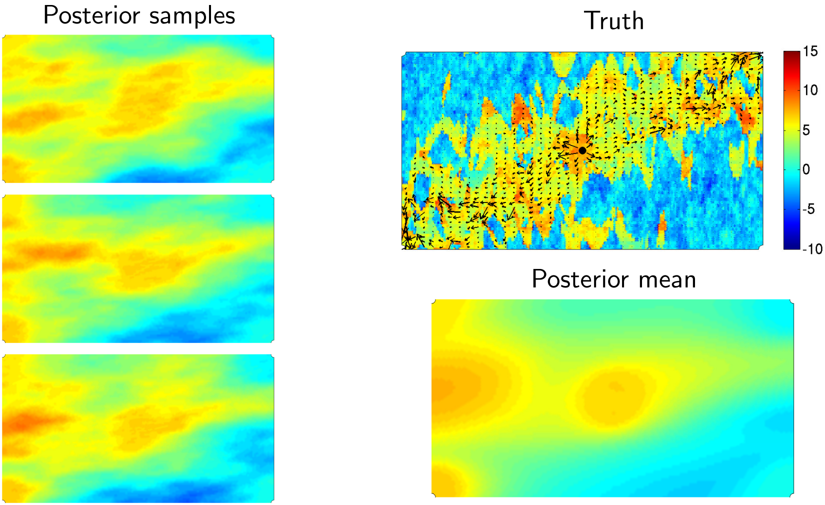


Standard deviation



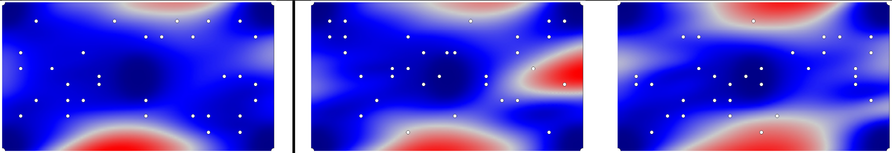
Random draw

SPE model: inference with the optimal design



SPE model: Compare optimal vs random design

Optimal	Random	



average var = 1.8824	average var = 2.4722	average var = 2.7206

- Designs with 32 sensors
- Observed data collected using true parameter on high resolution mesh $(n \sim 2.4 \times 10^5)$

Challenges revisited

- Infinite-dimensional inference
  Proper choice of prior, mass-weighted inner-product
- Need trace of posterior covariance (inverse of Hessian, large, dense, expensive matvecs)
  Randomized trace-estimator, Gaussian approximation in nonlinear case
- In nonlinear inverse problem, Hessian depends on data (not available *a priori*)
  Random draws from prior to generate training data
- Optimal design has combinatorial complexity in general
  $\boldsymbol{w}$*-weighted formulation, relax weights* $0 \leq w_j \leq 1$*, continuation to get 0–1*
- Conventional OED algorithms intractable for large-scale problems
  Infinite-dimensional formulation, gradient based optimization, scalable algorithms
- Gaussian assumption (on posterior of inverse problem) remains via linearized parameter-to-observable map; current worked aimed at higher-order approximation

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, *SIAM Journal on Scientific Computing*, 38(1):A243–A272, 2016. http://dx.doi.org/10.1137/140992564