Optimal Sensor Placement for Bayesian Linear Inverse Problems Governed by PDEs

Alen Alexanderian

Department of Mathematics
North Carolina State University

February 27, 2017
Motivating example: Diffusive transport of a contaminant with uncertain initial condition

- **Governing PDE** (forward model): advection-diffusion equation
- **Unknown/uncertain parameter**: initial concentration field
- **Inverse problem**: Use a vector $d$ of point (sensor) measurements of concentration at final time to reconstruct the initial state
Forward problem: time dependent advection-diffusion equation

\[ u_t - \kappa \Delta u + \mathbf{v} \cdot \nabla u = 0 \quad \text{in } \mathcal{D} \times [0, T] \]

\[ u(0, x) = m \quad \text{in } \mathcal{D} \]

\[ \kappa \nabla u \cdot \mathbf{n} = 0 \quad \text{on } \partial \mathcal{D} \times [0, T] \]

- \textit{m}: \textit{unknown} initial condition
- \textit{v}: velocity field
Solution of the forward problem

$t = 0$

$t = 1$

$t = 2$

$t = 3$
The inverse problem: reconstruct initial condition

The inverse problem of finding the unknown initial state based on sensor data

\[
\min_m \frac{1}{2} \|Bu(m) - d\|^2 + \frac{\alpha}{2} \langle Am, m \rangle
\]

where

\[
\begin{align*}
\frac{\partial u}{\partial t} - \kappa \Delta u + \mathbf{v} \cdot \nabla u &= 0 \quad \text{in } D \times [0, T] \\
\quad u(0, x) &= m \quad \text{in } D \\
\quad \kappa \nabla u \cdot \mathbf{n} &= 0 \quad \text{on } \partial D \times [0, T]
\end{align*}
\]

- \( B \): observation operator
- \( d = [d_1^T \; d_2^T \; \cdots \; d_{n_t}^T]^T \), \( d_i \in \mathbb{R}^{n_s}, n_s = \text{number of sensors} \)
- \( u \) linear in \( m \), \( u = Sm \implies \) linear parameter-to-observable map: \( \mathcal{F} = BS \)
- Can rewrite the optimization problem as

\[
\min_m \mathcal{J}(m) := \frac{1}{2} \| \mathcal{F}m - d \|^2 + \frac{\alpha}{2} \langle Am, m \rangle
\]
Solving the inverse problem

- Derivative of $\mathcal{J}$

\[
D\mathcal{J}(m)(\tilde{m}) = \frac{d}{d\varepsilon}\mathcal{J}(m + \varepsilon \tilde{m}) \big|_{\varepsilon=0} = \langle \mathcal{F}^*(\mathcal{F}m - d) + \alpha Am, \tilde{m} \rangle
\]

- Action of $\mathcal{F}^*$

$\mathcal{F}^* y = p(\cdot, 0)$, where $p$ is solution of the adjoint equation

\[
- p_t - \nabla \cdot (p v) - \kappa \Delta p = -\mathcal{B}^* y \\
p(T) = 0 \\
(v p + \kappa \nabla p) \cdot n = 0
\]

- Optimality condition

\[
(\mathcal{F}^* \mathcal{F} + \alpha A)m = \mathcal{F}^* d \quad \overset{\text{discretize}}{\Rightarrow} \quad (\mathcal{F}^* \mathcal{F} + \alpha A)m = \mathcal{F}^* d
\]

Solve the linear system using an iterative method, e.g. conjugate gradient
Solving the inverse problem: numerical results
Solving the inverse problem: numerical results

Truth

Sensor sites

Reconstruction
How to place sensors in an “optimal” way?

- Can formulate the optimal sensor placement problem as an optimal experimental design (OED) problem
- Can consider a statistical formulation of the inverse problem
- In addition to a reconstruction, we can also compute a statistical distribution of the parameters, conditioned on experimental data
- Find sensor locations so as to optimize the statistical quality of the reconstructed/inferred parameter
- In context of inverse problems a Bayesian formulation is natural
Bayesian inference: Bayes’ formula

\[ \pi_{\text{post}}(m|d) \propto \pi_{\text{like}}(d|m) \pi_{\text{prior}}(m) \]

- \( \pi_{\text{post}}(m|d) \) posterior pdf of \( m \)
- \( \pi_{\text{like}}(d|m) \) pdf of \( d \) given \( m \) (data likelihood)
- \( \pi_{\text{prior}}(m) \) prior pdf of \( m \)

pdf = probability density function

**Rev. Thomas Bayes**  
Bayes, T., An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, FRS Communicated by Mr. Price, in a Letter to John Canton, AMFRS. Philosophical Transactions, 1763.

**Pierre-Simon Laplace**  
Laplace, P.S., Théorie analytique des probabilités. 1820.
Modeling the initial state as a random function
Bayesian linear inverse problems

Assume linear parameter-to-observable map and additive Gaussian noise:

\[ d = Fm + \eta, \quad \eta \sim \mathcal{N}(0, \Gamma_{\text{noise}}) \]

Likelihood:

\[ \pi_{\text{like}}(d | m) \propto \exp \left\{ -\frac{1}{2} (Fm - d)^* \Gamma_{\text{noise}}^{-1} (Fm - d) \right\} \]

Gaussian prior:

\[ \pi_0(m) \propto \exp\left(-\frac{1}{2} m^T \Gamma_{\text{prior}}^{-1} m\right) \]
Bayesian linear inverse problems

For Bayesian linear inverse problem with Gaussian prior and noise, the posterior pdf is

\[ \pi_{\text{post}}(m|d) \propto \exp \left\{ -\frac{1}{2}(m - m_{\text{MAP}})^T (F^T \Gamma^{-1}_{\text{noise}} F + \Gamma^{-1}_{\text{prior}}) (m - m_{\text{MAP}}) \right\} \]

\[ \Rightarrow \mu_{\text{post}} = \mathcal{N}(m_{\text{MAP}}, \Gamma_{\text{post}}) \]

\[ \Gamma_{\text{post}}^{-1} = \frac{H_{\text{misfit}}}{D_m^2} = \begin{cases} F^T \Gamma^{-1}_{\text{noise}} F + \Gamma^{-1}_{\text{prior}} & (= D_m^2(\log \pi_{\text{post}})) \\ \mathcal{H}_{\text{misfit}} \end{cases} \]

\[ m_{\text{MAP}} = \arg\min_m \frac{1}{2} \| Fm - d \|_{\Gamma^{-1}_{\text{noise}}}^2 + \frac{1}{2} \langle \Gamma^{-1}_{\text{prior}} m, m \rangle \]
Bayesian inversion of the initial condition for 2D advection-diffusion equation

- Posterior mean, and posterior variance

<table>
<thead>
<tr>
<th>truth</th>
<th>posterior mean</th>
<th>posterior std deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="true-image" alt="Image" /></td>
<td><img src="post-mean" alt="Image" /></td>
<td><img src="post-std-deviation" alt="Image" /></td>
</tr>
</tbody>
</table>

- Posterior samples: $\nu = m_{\text{MAP}} + \Gamma_{\text{post}}^{1/2} M^{-1/2} n$, $n \sim \mathcal{N}(0, I)$

<table>
<thead>
<tr>
<th><img src="sample-1" alt="Image" /></th>
<th><img src="sample-2" alt="Image" /></th>
<th><img src="sample-3" alt="Image" /></th>
</tr>
</thead>
</table>
The optimal experimental design problem

A grid of candidate locations for observation points

- **Experimental design**: locations of observation points / sensors

  \[
  \text{design} := \left\{ x_1, \ldots, x_{N_s}, w_1, \ldots, w_{N_s} \right\}
  \]

- **Bayesian inversion**: data + likelihood, prior $\rightarrow$ posterior distribution of inversion parameter

- **Optimal experimental design (OED)**: Find sensor locations that result in minimized posterior uncertainty
Mathematical/computational challenges

- The inference problem is in an infinite-dimensional space
- Need to compute functionals of posterior covariance (inverse of Hessian, large, dense, expensive matvecs)
- With nonlinear inverse problems we are led to a bilevel optimization problem
- Optimal experimental design problem can have combinatorial complexity
- Conventional OED algorithms intractable for large-scale problem (due to high-dimensional parameters, expensive-to-evaluate PDE-based parameter-to-observable map, ...)

Alen Alexanderian (NCSU)
Optimal experimental design

- A-optimal design:
  
  **Minimize “average variance” of parameter function** $m$

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

- Covariance operator:

  $$[C_{\text{post}} u](x) = \int_D c(x, y) u(y) \, dy$$
A-optimal design:

Minimize “average variance” of parameter function $m$

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

Covariance operator:

$$[C_{\text{post}} u](x) = \int_{D} c(x, y) u(y) \, dy$$

Variance of $m$ at a given $x$:

$$\text{Var}\{m(x)\} = c(x, x)$$
Optimal experimental design

- A-optimal design:
  
  Minimize “average variance” of parameter function $m$

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

$$[C_{\text{post}} u](x) = \int_D c(x, y) u(y) \, dy$$

- Variance of $m$ at a given $x$:

$$\text{Var}\{m(x)\} = c(x, x)$$

- Average variance:

$$\int_D c(x, x) \, dx$$
Optimal experimental design

- A-optimal design:
  
  \textbf{Minimize “average variance” of parameter function } m

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

- Covariance operator:

  \[
  [C_{\text{post}} u](x) = \int_D c(x, y) u(y) \, dy
  \]

- Variance of \( m \) at a given \( x \):

  \[
  \text{Var}\{m(x)\} = c(x, x)
  \]

- Average variance:

  \[
  \int_D c(x, x) \, dx = \text{tr}(C_{\text{post}})
  \]
Optimal experimental design

- **A-optimal design:**
  
  Minimize “average variance” of parameter function \( m \)

- Covariance function: \( c(\mathbf{x}, \mathbf{y}) = \text{Cov} \{ m(\mathbf{x}), m(\mathbf{y}) \} \)

- Covariance operator:

  \[
  [C_{\text{post}} u](\mathbf{x}) = \int_{\mathcal{D}} c(\mathbf{x}, \mathbf{y}) u(\mathbf{y}) d\mathbf{y}
  \]

- Variance of \( m \) at a given \( \mathbf{x} \):

  \[
  \text{Var}\{m(\mathbf{x})\} = c(\mathbf{x}, \mathbf{x})
  \]

- Average variance:

  \[
  \int_{\mathcal{D}} c(\mathbf{x}, \mathbf{x}) d\mathbf{x} = \text{tr}(C_{\text{post}})
  \]

- Optimal design criterion:

  Choose a “design” to minimize \( \text{tr}(C_{\text{post}}) \)
A-optimal experimental design with sparsity control

\[
\begin{align*}
\text{minimize} \quad & \quad \operatorname{tr} \left[ \Gamma_{\text{post}}(w) \right] + \gamma P(w) \\
\text{subject to} \quad & \quad 0 \leq w \leq 1
\end{align*}
\]

\( \Gamma_{\text{post}}(w) = (\sigma_{\text{noise}}^{-2} F^* W F + \Gamma_{\text{prior}}^{-1})^{-1} \), \( W \): diagonal matrix with \( w_i \) on its diagonal

\( P(w) \): penalty term, \( \gamma > 0 \) (e.g., \( P(w) = \sum_j w_j \))

**Theorem**

*If the penalty function \( P \) is convex, then there exists a unique solution for the optimization problem \((\ast)\).*

- Need trace of inverse Hessian and its derivative
- Need many applications of the forward operator \( F \) \( \implies \) many PDE solves
Randomized trace estimation

- $A \in \mathbb{R}^{n \times n}$ — symmetric
- Trace estimator:
  \[
  \text{tr}(A) \approx \frac{1}{n_{\text{tr}}} \sum_{i=1}^{n_{\text{tr}}} z_i^T A z_i, \quad z_i \text{ — random vectors}
  \]
- Gaussian trace estimator: $z_i$ independent draws from $\mathcal{N}(0, I)$
- For $z \sim \mathcal{N}(0, I)$
  \[
  \mathbb{E}\{z^T A z\} = \text{tr}(A) \quad \text{Var}\{z^T A z\} = 2 \|A\|_F^2
  \]

Efficient means of approximating trace of posterior covariance

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).
A-optimal design: the objective function

- Randomized trace estimator:

\[
\text{tr}[\Gamma_{\text{post}}(w)] \approx \frac{1}{N} \sum_{i=1}^{N} \langle z_i, \Gamma_{\text{post}}(w) z_i \rangle =: \phi(w)
\]

- \(z_i\) random vectors (e.g. Gaussian)

- Note: \(\Gamma_{\text{post}} = H^{-1}\)

\[
H = \sigma_{\text{noise}}^{-2} F^* W F + \Gamma_{\text{prior}}^{-1} \\
= H_{\text{misfit}}(w) + \Gamma_{\text{prior}}^{-1}
\]

(for notational convenience, let \(\sigma_{\text{noise}} = 1\) from now on)
Inverse of the Hessian:

\[ H^{-1} = \left( H_{\text{misfit}} + \Gamma_{\text{prior}}^{-1} \right)^{-1} \]

\[ = \Gamma_{\text{prior}}^{1/2} \left( \Gamma_{\text{prior}}^{1/2} H_{\text{misfit}} \Gamma_{\text{prior}}^{1/2} + I \right)^{-1} \Gamma_{\text{prior}}^{1/2} \]
Application of inverse Hessian

- Inverse of the Hessian:

\[ H^{-1} = (H_{\text{misfit}} + \Gamma_{\text{prior}}^{-1})^{-1} \]

\[ = \Gamma_{\text{prior}}^{1/2} (\Gamma_{\text{prior}}^{1/2} H_{\text{misfit}} \Gamma_{\text{prior}}^{1/2} + I)^{-1} \Gamma_{\text{prior}}^{1/2} \]

- Low-rank \( H_{\text{misfit}} \) (prior preconditioned misfit Hessian):

\[ \tilde{H}_{\text{misfit}} \approx \sum_{i=1}^{r} \lambda_i \mathbf{v}_i \mathbf{v}_i^T \]
Application of inverse Hessian

- Inverse of the Hessian:

\[
H^{-1} = \left( H_{\text{misfit}} + \Gamma_{\text{prior}}^{-1} \right)^{-1} = \Gamma_{\text{prior}}^{1/2} \left( \Gamma_{\text{prior}}^{1/2} H_{\text{misfit}} \Gamma_{\text{prior}}^{1/2} + I \right)^{-1} \Gamma_{\text{prior}}^{1/2} \tilde{H}_{\text{misfit}}
\]

- Low-rank \( H_{\text{misfit}} \) (prior preconditioned misfit Hessian):

\[
\tilde{H}_{\text{misfit}} \approx \sum_{i=1}^{r} \lambda_i v_i v_i^T
\]

- Efficient \( H^{-1} \) apply:

\[
H^{-1} q \approx \Gamma_{\text{prior}}^{1/2} \left( V_r \Lambda_r V_r^* + I \right)^{-1} \Gamma_{\text{prior}}^{1/2} q = \Gamma_{\text{prior}}^{1/2} \left( I - V_r D_r V_r^* \right) \Gamma_{\text{prior}}^{1/2} q \quad \text{(Sherman-Morrison-Woodbury)}
\]

\[
D = \text{diag}\{\lambda_1/(1 + \lambda_1), \ldots, \lambda_r/(1 + \lambda_r)\}
\]
A-optimal design: the gradient

Objective function: \( \phi(w) = \frac{1}{N} \sum_{i=1}^{N} \langle z_i, q_i \rangle \)

For each \( i \), need one application of parameter-to-observable map \( F \):

\[ q_i \xrightarrow{F} (d_{T1}, d_{T2}, \ldots, d_{TN})^T, \]

where \( d_s = (d_{s1}, d_{s2}, \ldots, d_{sN})^T \). Then,

\[ \langle q_i, \partial_j H_{\text{misfit}} q_i \rangle = N^\tau \sum_{s=1}^{N^\tau} d_{js} d_{js}. \]
A-optimal design: the gradient

Objective function: \( \phi(\mathbf{w}) = \frac{1}{N} \sum_{i=1}^{N} \langle z_i, q_i \rangle \quad q_i = H^{-1}(\mathbf{w})z_i \)

- Gradient:
  \[
  \frac{\partial}{\partial w_j} H^{-1}(\mathbf{w}) = -H^{-1}(\mathbf{w}) \partial_j H(\mathbf{w}) H^{-1}(\mathbf{w}) = -H^{-1}(\mathbf{w}) \partial_j H_{\text{misfit}}(\mathbf{w}) H^{-1}(\mathbf{w})
  \]

  \[
  \frac{\partial \phi}{\partial w_j} = -\frac{1}{N} \sum_{i=1}^{N} \langle q_i, \partial_j H_{\text{misfit}} q_i \rangle \quad j = 1, \ldots, N_s
  \]
A-optimal design: the gradient

Objective function: \( \phi(w) = \frac{1}{N} \sum_{i=1}^{N} \langle z_i, q_i \rangle \quad q_i = H^{-1}(w)z_i \)

- Gradient:

\[
\frac{\partial}{\partial w_j} H^{-1}(w) = -H^{-1}(w) \partial_j H(w) H^{-1}(w) = -H^{-1}(w) \partial_j H_{\text{misfit}}(w) H^{-1}(w)
\]

\[
\frac{\partial \phi}{\partial w_j} = -\frac{1}{N} \sum_{i=1}^{N} \langle q_i, \partial_j H_{\text{misfit}} q_i \rangle \quad j = 1, \ldots, N_s
\]

- For each \( i \), need one application of parameter-to-observable map \( F \):

\[
q_i \xrightarrow{F} (d_{1s}^T, d_{2s}^T, \ldots, d_{Ns}^T)^T,
\]

where \( d_s = (d_{s1}, d_{s2}, \ldots, d_{sN_\tau})^T \). Then,

\[
\langle q_i, \partial_j H_{\text{misfit}} q_i \rangle = \sum_{s=1}^{N_\tau} d_{1s}^j d_{2s}^j.
\]
The forward operator

- Need many forward solves in the optimization process
- Idea: \( \mathbf{F} \) is low-rank (often)
- Note:
  \[
  \mathbf{F} = \underbrace{\mathbf{B}}_{\text{observation operator}} \underbrace{\mathbf{S}}_{\text{solution operator}}
  \]
- Idea: compute a low-rank SVD surrogate for \( \mathbf{F} \)
  \[
  \mathbf{F} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T
  \]
- Randomized SVD:
  - Independent matvecs (forward/adjoint) — can do in parallel
  - Simple but very robust algorithms
  - Backed by rigorous theory
  - Almost deterministic behavior

Randomized SVD for forward operator

- Idea: $F$ is low-rank (often); better idea: $\tilde{F} = F\Gamma_{\text{prior}}^{1/2}$ is even more so ...

$$\tilde{F} \approx \sum_{j=1}^{r} \sigma_j u_i v_i^T$$

SVD surrogate for $\tilde{F}$ $\Rightarrow$ no forward PDE solves in OED algorithm
At the cost of an upfront SVD for $F$:

- No PDE solves in the optimization process
- Efficient computation of cost/derivatives
- Independent of temporal/spatial mesh
A-optimal design: numerical results

- Sensor allocation

- Weight distribution
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
A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas. A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized $\ell^0$-sparsification. SISC. 2014.
Summary:
- A-Optimal sensor placement for PDE-based Bayesian linear inverse problems
- Scalable algorithms
- Efficient computation of OED objective/gradient (randomized methods in numerical linear algebra, low-rank approximations, ...)

Outlook:
- Optimal sensor placement for Bayesian nonlinear inverse problems: bilevel PDE-constrained optimization, adjoint based derivative computation, ...
  
- Goal-oriented OED (ongoing joint work with Ahmed Attia, Arvind Saibaba)
- D-optimal sensor placements for PDE-based inverse problems (joint work with Arvind Saibaba)