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).
- Particularly 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 $d$ of point measurements of concentration to infer distribution of $m$
- **Optimal experimental design problem**: Find sensor placements (to collect the data $d$) 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

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

- \( P(w) \): penalty term
- Discretized posterior covariance operator:
  \[
  \Gamma_{\text{post}}(w) = \left( \frac{1}{\sigma^2_{\text{noise}}} F^* W F + \Gamma_{\text{pr}}^{-1} \right)^{-1}
  \]

Main challenges:
- Choice of penalty function to achieve binary weights
- Re-computation of inverse covariance trace
- Gradient of posterior trace with respect to \( w \)
Randomized trace estimator:

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

- $z_i$ random vectors (e.g., Gaussian)

Computation of cost and gradient:

$$\phi(w) = \frac{1}{n_{\text{tr}}} \sum_{i=1}^{n_{\text{tr}}} \langle z_i, q_i \rangle \quad \frac{\partial \phi}{\partial w_j} = -\frac{1}{n_{\text{tr}}} \sum_{i=1}^{n_{\text{tr}}} \langle q_i, \partial_j \mathcal{H}_{\text{misfit}} q_i \rangle$$

- Need $q_i = \Gamma_{\text{post}}(w) z_i$, $i = 1, \ldots, n_{\text{tr}}$

- Need $F q_i$, $i = 1, \ldots, n_{\text{tr}}$ (for computing $\frac{\partial \phi}{\partial w_j}$)
SVD surrogate for the parameter-to-observable map

- Need many applications of $F$ in the optimization process
- Idea: $F$ is low-rank (often) $\implies$ compute low-rank SVD for $F$
- Better idea: Compute SVD surrogate for $\tilde{F} = F \Gamma_{pr}^{1/2}$
- Use randomized SVD*
- SVD surrogate for $\tilde{F}$:
  no forward/adjoint PDE solves in OED algorithm

Alternative: Rewrite posterior covariance 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

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 \( \mathbf{v} = \mathbf{v}(\xi) \)

\[
\begin{align*}
  u_t - \kappa \Delta u + \mathbf{v} \cdot \nabla u &= 0 \quad \text{in } D \times [0, T] \\
  u(0, x) &= m \quad \text{in } D \\
  \kappa \nabla u \cdot \mathbf{n} &= 0 \quad \text{on } \partial D \times [0, T]
\end{align*}
\]
Extension I: Forward model with uncertainty

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

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

Corresponding posterior covariance matrix:

\[
C_{\text{post}}(\mathbf{w}, \xi) = \left( \frac{1}{\sigma_{\text{noise}}^2} \mathcal{F}(\xi)^* \mathbf{W} \mathcal{F}(\xi) + C_{\text{pr}}^{-1} \right)^{-1}
\]
Extension I: Forward model with uncertainty

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

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

Corresponding posterior covariance matrix:

$$
C_{\text{post}}(w, \xi) = \left( \frac{1}{\sigma_{\text{noise}}^2} \mathcal{F}(\xi)^* W \mathcal{F}(\xi) + C_{\text{pr}}^{-1} \right)^{-1}
$$

OED formulation:

$$
\min_{w \in [0,1]^d} \text{tr} \left( C_{\text{post}}(w, \xi) \right) + \gamma P_\varepsilon(w)
$$

Extension I: Forward model with uncertainty

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

$$u_t - \kappa \Delta u + v \cdot \nabla u = 0 \quad \text{in } D \times [0, T]$$
$$u(0, x) = m \quad \text{in } D$$
$$\kappa \nabla u \cdot n = 0 \quad \text{on } \partial D \times [0, T]$$

Corresponding posterior covariance matrix:

$$C_{\text{post}}(w, \xi) = \left( \frac{1}{\sigma_{\text{noise}}^2} \mathcal{F}(\xi)^* W \mathcal{F}(\xi) + C_{\text{pr}}^{-1} \right)^{-1}$$

OED formulation:

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

Reference:

(1) Computation of trace, “measurement space approach”

\[ \phi(\xi, w) = \text{tr} \left[ \left( \frac{1}{\sigma^2} F(\xi)^* W F(\xi) + C_0^{-1} \right)^{-1} \right] \]

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

\[ \phi(\xi, w) = \text{tr} (C_0) - \text{tr} \left[ \frac{1}{\sigma^2} C_0 F^*(\xi) S^{-1}(\xi, w) W F(\xi) C_0 \right] \]

where \((I + \frac{1}{\sigma^2} W F(\xi) C_0 F^*(\xi)) = S(\xi, w)\)
(1) Computation of trace, “measurement space approach”

Rearranging terms in the trace, we have:

\[ \phi(\xi, w) = \text{tr}(C_0) - \text{tr}\left[ \frac{1}{\sigma^2} S^{-1}(\xi, w) W F(\xi) C_0^2 F^*(\xi) \right] \]

\[ = \text{tr}(C_0) - \text{tr}[K(\xi, w)] \]

First term independent of \( w \) \implies\ drop it

Optimal design now satisfies:

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

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

$$\int_{\Omega} \text{tr} [K(\xi, w)] P(d\xi) \approx \frac{1}{N} \sum_{i=1}^{N} \text{tr} [K(\xi_i, w)]$$
(3) Elimination of PDEs from the minimization

OEDUU objective (with discretized operators):

\[
\frac{1}{N} \sum_{i=1}^{N} \text{tr} [K(\xi_i, w)] = \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{j=1}^{d} \langle e_j, K(\xi_i, w)e_j \rangle \right)
\]

where

\[
K(\xi_i, w) = \frac{1}{\sigma^2} S^{-1}(\xi, w) W F_i C_0^2 F_i^*
\]

\[
S(\xi_i, w) = \left( I + \frac{1}{\sigma^2} W F_i C_0 F_i^* \right)
\]

Expensive to optimize:

- in each step of minimization need to compute trace for each sample
- evaluation of trace requires \(d\) applications of \(K(\xi_i, w)\)
- 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 \( C^\frac{1}{2} F(\xi_i) = \tilde{F}_i \)

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

Storing separate basis vectors for each \( \tilde{F}_i \) 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 \( Q \in \mathbb{R}^{d \times k} \) and \( \hat{Q} \in \mathbb{R}^{m \times k} \) (\( k \) small) such that \( \forall i \in [1, \cdots, N] \):

\[
\tilde{F}_i \approx QQ^* \tilde{F}_i \hat{Q}\hat{Q}^*
\]
Joint basis algorithm:

**Algorithm 1** Composite randomized range-finder algorithm

1. $N$ random matrices $\Omega_i \in \mathbb{R}^{d \times r}$, $\hat{\Omega}_i \in \mathbb{R}^{n \times r}$, $i = 1, \ldots, N$.
2. $Y_i = \tilde{F}_i \Omega_i$, $\hat{Y}_i = \tilde{F}_i^T \hat{\Omega}_i$.
3. Set $Y = [Y_1, \ldots, Y_N]$ and $\hat{Y} = [\hat{Y}_1, \ldots, \hat{Y}_N]$.
4. Compute SVDs of $Y = U \Sigma V^T$ and of $\hat{Y} = \hat{U} \hat{\Sigma} \hat{V}^T$.
5. Truncate SVDs up to desired tolerance.
6. $Q = U[:, 1 : k]$ and $\hat{Q} = \hat{U}[:, 1 : k]$.
7. **return** $Q$ and $\hat{Q}$.
(4) Reduced basis and clustering

Potential issues:

▶ Still need to apply $\tilde{F}_i$ to $k$ vectors to compute $Q^* \tilde{F}_i \hat{Q}$

▶ Requires minimal-residual solution to

\[
B_i \hat{Q}^T \Omega_i = Q^T Y_i, \tag{1}
\]

\[
B_i^T Q^T \hat{\Omega}_i = \hat{Q}^T \hat{Y}_i \tag{2}
\]

▶ Memory usage $\uparrow$ as $N \uparrow$

▶ Cluster “similar” uncertain parameters $\xi$ 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 $\tilde{F}_i$ to $k$ vectors to compute $Q^*\tilde{F}_i\hat{Q}$
- Find $B_i \approx Q^*\tilde{F}_i\hat{Q}$
- Requires minimal-residual solution to

\begin{align}
B_i\hat{Q}^T\Omega_i &= Q^TY_i, \\
B_i^TQ^T\hat{\Omega}_i &= \hat{Q}^T\hat{Y}_i
\end{align}

- Memory usage ↑ as $N$ ↑
Synthesizing a distribution for $\nu$

Solve Bayesian inverse problem for $\xi$ using synthetic data

$-\nabla \cdot (e^{\xi} \nabla p) = 0$ in $D$,

$p = 0$ on $\Gamma_L$,

$p = 1$ on $\Gamma_R$,

$e^{\xi} \nabla p \cdot n = 0$ on $\Gamma_N$

$\xrightarrow{\text{Sample}}$ $\xi_i \sim$ sample $\nu_i$

$\xrightarrow{\text{Leads to}}$ $\pi_\xi(\xi)$

$\pi_\xi(\xi)$: Gaussian approximation at MAP
Subsurface flow OEDUU

\[ w^* = \arg\min_{w \in [0,1]^d} - \frac{1}{N} \sum_{i=1}^{N} K(v_i, w) + \gamma \psi(w) \]

- Solve minimization problem many times with different \( \psi(w) \)
  - Change \( \psi(w) \) to approximate \( \ell_0 \) “norm” better in each iteration
- Each minimization solved with gradient-based method (projected BFGS)
- Number of minimization problems needed to solve \( \approx 10 \)
Design comparisons

How well do the designs do for initial condition inversion using the “true” $\nu$?

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

\[-E(\text{tr}(K(\omega)))\]

Number of sensors

- OED
- OEDUU with 100 samples
Extension II: Nonlinear forward model

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

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

- Data \(\mathbf{d}\) not available \textit{a priori}

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

\[
- \nabla \cdot (\exp(m) \nabla u) = f \quad \text{in} \ D
\]
\[
u = 0 \quad \text{on} \ \Gamma_D
\]
\[
\exp(m) \nabla u \cdot n = g \quad \text{on} \ \Gamma_N
\]

Reference:

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:

\[
\minimize_w \quad \operatorname{tr} \left( \mathcal{H}^{-1} \left[ m_{\text{MAP}}(w; d), w; d \right] \right)
\]
Extension II: Nonlinear forward model

- General formulation: minimize expected average posterior variance:

\[
\min_w \int \left\{ \text{tr} \left( H^{-1} [m_{MAP}(w; d), w; d] \right) \right\} d\mu(d)
\]
Extension II: Nonlinear forward model

- General formulation: minimize expected average posterior variance:

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

- \(\gamma P(w)\): sparsifying penalty function
Extension II: Nonlinear forward model

- General formulation: minimize expected average posterior variance:

\[
\min_w \int \left\{ \text{tr} \left[ H^{-1} [ m_{\text{MAP}}(w; d), w; d] \right] \right\} d\mu(d) + \gamma P(w)
\]

- \( \gamma P(w) \): sparsifying penalty function

- In practice: get data from a few training models \( m_1, \ldots, m_n \)

- \( m_i \): draws from prior

- Training data:

\[
d_i = f(m_i) + \eta_i, \quad i = 1, \ldots, n
\]
Extension II: Nonlinear forward model

- General formulation: minimize expected average posterior variance:

\[
\min_w \int \left\{ \text{tr} \left( \mathcal{H}^{-1} \left[ \text{MAP}(w; d), w; d \right] \right) \right\} \ d\mu(d) + \gamma P(w)
\]

- \( \gamma P(w) \): sparsifying penalty function

- In practice: get data from a few training models \( m_1, \ldots, m_n \)

- \( m_i \): draws from prior

- Training data:

\[
d_i = f(m_i) + \eta_i, \quad i = 1, \ldots, n
\]

- The problem to solve in practice:

\[
\min_w \frac{1}{n} \sum_{i=1}^{n} \text{tr} \left( \mathcal{H}^{-1} \left[ \text{MAP}(w; d_i), w; d_i \right] \right) + \gamma P(w)
\]
Application: subsurface flow

- **Forward problem**

\[- \nabla \cdot (e^m \nabla u) = f \quad \text{in } \mathcal{D}\]
\[u = 0 \quad \text{on } \Gamma_D\]
\[e^m \nabla u \cdot \mathbf{n} = g \quad \text{on } \Gamma_N\]

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

Left: true parameter, right: pressure-field
Bayesian inverse problem: Gaussian approximation

- Reduced cost functional (for given data \(d\))

\[
\hat{J}(\theta) = \mathcal{J}(u(\theta), \theta)
= \frac{1}{2\sigma_{\text{noise}}^2} (Bu - d)^T W (Bu - d) + \frac{1}{2} \langle C_{\text{pr}}^{-1}(m - m_{\text{prior}}), m - m_{\text{prior}} \rangle
\]

- \(B\) is observation operator
- MAP point is solution to

\[
\minimize_m \hat{J}(m)
\]

where

\[
- \nabla \cdot (e^m \nabla u) = f \quad \text{in } D
\]

\[
u = 0 \quad \text{on } \Gamma_D
\]

\[
e^m \nabla u \cdot n = g \quad \text{on } \Gamma_N
\]

- Covariance of Gaussian approximation to posterior:

\[
\mathcal{H}(m)^{-1} = (\nabla^2 \hat{J}(m))^{-1}
\]
Optimization problem for computing A-optimal design

- OED objective function:

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

- Optimization problem:

\[
\begin{align*}
\text{minimize} & \quad \sum_i \langle z_i, y_i(w) \rangle + \gamma P(w) \\
\text{where} & \\
\text{m}_{\text{MAP}}(w) & = \arg \min_{\theta} \mathcal{J}\left(u(\theta(w)), \theta(w)\right) \\
\mathcal{H}(m_{\text{MAP}}(w)) y_i & = z_i
\end{align*}
\]
Optimization problem for computing A-optimal design

OED objective function:
\[
\text{tr}(\mathcal{H}^{-1}) = \sum_i \langle z_i, \mathcal{H}^{-1} z_i \rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i
\]

\[
\begin{align*}
\text{minimize} & \quad \sum_i \langle z_i, y_i \rangle + \gamma P(w) \\
\text{subject to} & \quad \nabla \cdot (e^m \nabla u) = f \quad \text{(state)} \\
& \quad \nabla \cdot (e^m \nabla p) = -\frac{1}{\sigma_{\text{noise}}^2} \mathcal{B}^* \mathcal{W} (\mathcal{B} u - \mathcal{d}) \quad \text{(adjoint)} \\
& \quad C^{-1}_{pr} (m - m_{\text{prior}}) + e^m \nabla u \cdot \nabla p = 0 \quad \text{(gradient = 0)} \\
& \quad \nabla \cdot (e^m \nabla v_i) = \nabla \cdot (y_i e^m \nabla u) \quad \text{(inc. state)} \\
& \quad \nabla \cdot (e^m \nabla q_i) = \nabla \cdot (y_i e^m \nabla p) - \frac{1}{\sigma_{\text{noise}}^2} \mathcal{B}^* \mathcal{W} \mathcal{B} v_i \quad \text{(inc. adjoint)} \\
& \quad C^{-1}_{pr} y_i + y_i e^m \nabla u \cdot \nabla p + e^m (\nabla v_i \cdot \nabla p + \nabla u \cdot \nabla q_i) = z_i \quad \text{(Hessian solve)}
\end{align*}
\]
Optimization problem for computing A-optimal design

OED objective function:
\[
\text{tr}(\mathcal{H}^{-1}) = \sum_i \langle z_i, \mathcal{H}^{-1} z_i \rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i
\]

\[
\text{minimize} \sum_i \langle z_i, y_i \rangle + \gamma P(w) \quad \text{subject to}
\]

\[
-\nabla \cdot (e^m \nabla u) = f \quad \text{(state)}
\]

\[
-\nabla \cdot (e^m \nabla p) = -\frac{1}{\sigma_{\text{noise}}^2} B^* W (Bu - d) \quad \text{(adjoint)}
\]

\[
C_{\text{pr}}^{-1} (m - m_{\text{prior}}) + e^m \nabla u \cdot \nabla p = 0 \quad \text{(gradient = 0)}
\]

\[
-\nabla \cdot (e^m \nabla v_i) = \nabla \cdot (y_i e^m \nabla u) \quad \text{(inc. state)}
\]

\[
-\nabla \cdot (e^m \nabla q_i) = \nabla \cdot (y_i e^m \nabla p) - \frac{1}{\sigma_{\text{noise}}^2} B^* W B v_i \quad \text{(inc. adjoint)}
\]

\[
C_{\text{pr}}^{-1} y_i + y_i e^m \nabla u \cdot \nabla p + e^m (\nabla v_i \cdot \nabla p + \nabla u \cdot \nabla q_i) = z_i \quad \text{(Hessian solve)}
\]
Optimization problem for computing A-optimal design

OED objective function:
\[
\text{tr}(\mathcal{H}^{-1}) = \sum_i \langle z_i, \mathcal{H}^{-1} z_i \rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i
\]

\[
\minimize_{w \in [0,1]^{ns}} \sum_i \langle z_i, y_i \rangle + \gamma P(w)
\]

subject to
\[
-\nabla \cdot (e^m \nabla u) = f \quad \text{(state)}
\]
\[
-\nabla \cdot (e^m \nabla p) = -\frac{1}{\sigma_{\text{noise}}^2} B^* W (B u - d) \quad \text{(adjoint)}
\]
\[
C_{pr}^{-1} (m - m_{\text{prior}}) + e^m \nabla u \cdot \nabla p = 0 \quad \text{(gradient = 0)}
\]
\[
-\nabla \cdot (e^m \nabla v_i) = \nabla \cdot (y_i e^m \nabla u) \quad \text{(inc. state)}
\]
\[
-\nabla \cdot (e^m \nabla q_i) = \nabla \cdot (y_i e^m \nabla p) - \frac{1}{\sigma_{\text{noise}}^2} B^* W B v_i \quad \text{(inc. adjoint)}
\]
\[
C_{pr}^{-1} y_i + y_i e^m \nabla u \cdot \nabla p + e^m (\nabla v_i \cdot \nabla p + \nabla u \cdot \nabla q_i) = z_i \quad \text{(Hessian solve)}
\]
Optimization problem for computing A-optimal design

OED objective function:
\[ \text{tr}(\mathcal{H}^{-1}) = \sum_i \langle z_i, \mathcal{H}^{-1} z_i \rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i \]

\[
\text{minimize} \sum_i \langle z_i, y_i \rangle + \gamma P(w) \quad \text{subject to}
\]
\[ -\nabla \cdot (e^m \nabla u) = f \quad \text{(state)}
\]
\[ -\nabla \cdot (e^m \nabla p) = -\frac{1}{\sigma_{\text{noise}}^2} B^* W (B u - d) \quad \text{(adjoint)}
\]
\[ C_{pr}^{-1} (m - m_{\text{prior}}) + e^m \nabla u \cdot \nabla p = 0 \quad \text{(gradient = 0)}
\]
\[ -\nabla \cdot (e^m \nabla v_i) = \nabla \cdot (y_i e^m \nabla u) \quad \text{(inc. state)}
\]
\[ -\nabla \cdot (e^m \nabla q_i) = \nabla \cdot (y_i e^m \nabla p) - \frac{1}{\sigma_{\text{noise}}^2} B^* W B v_i \quad \text{(inc. adjoint)}
\]
\[ C_{pr}^{-1} y_i + y_i e^m \nabla u \cdot \nabla p + e^m (\nabla v_i \cdot \nabla p + \nabla u \cdot \nabla q_i) = z_i \quad \text{(Hessian solve)}
\]
Optimization problem for computing A-optimal design

OED objective function:
\[
\text{tr}(\mathcal{H}^{-1}) = \sum_i \langle z_i, \mathcal{H}^{-1} z_i \rangle = \sum_i \langle z_i, y_i \rangle, \quad \mathcal{H} y_i = z_i
\]

\[
\min_{\mathbf{w} \in [0,1]^{ns}} \sum_i \langle z_i, y_i \rangle + \gamma P(\mathbf{w})
\]

subject to
\[
-\nabla \cdot (e^m \nabla u) = f \quad \text{(state)}
\]
\[
-\nabla \cdot (e^m \nabla p) = -\frac{1}{\sigma_{\text{noise}}^2} B^* \mathbf{W} (B u - \mathbf{d}) \quad \text{(adjoint)}
\]
\[
C^{-1}_{pr} (m - m_{\text{prior}}) + e^m \nabla u \cdot \nabla p = 0 \quad \text{(gradient = 0)}
\]
\[
-\nabla \cdot (e^m \nabla v_i) = \nabla \cdot (y_i e^m \nabla u) \quad \text{(inc. state)}
\]
\[
-\nabla \cdot (e^m \nabla q_i) = \nabla \cdot (y_i e^m \nabla p) - \frac{1}{\sigma_{\text{noise}}^2} B^* \mathbf{W} B v_i \quad \text{(inc. adjoint)}
\]
\[
C^{-1}_{pr} y_i + y_i e^m \nabla u \cdot \nabla p + e^m (\nabla v_i \cdot \nabla p + \nabla u \cdot \nabla q_i) = z_i \quad \text{(Hessian solve)}
\]
The Lagrangian for the OED problem

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

$$
\mathcal{L}^E(w; u, m, p, v, q, y; u^*, m^*, p^*, v^*, q^*, y^*)
= \langle z, y \rangle
+ \langle e^m \nabla u, \nabla u^* \rangle - \langle f, u^* \rangle
+ \langle e^m \nabla p, \nabla p^* \rangle + \sigma_{noise}^{-2} \langle B^* W(Bu - d), p^* \rangle
+ \langle C_{pr}^{-1} (m - m_0), m^* \rangle + \langle m^* e^m \nabla u, \nabla p \rangle
+ \langle e^m \nabla v, \nabla v^* \rangle + \langle ye^m \nabla u, \nabla v^* \rangle
+ \langle e^m \nabla q, \nabla q^* \rangle + \langle ye^m \nabla p, \nabla q^* \rangle + \sigma_{noise}^{-2} \langle B^* WBv, q^* \rangle
+ \langle y^* e^m \nabla v, \nabla p \rangle + \langle y^*, C_{pr}^{-1} y \rangle + \langle y^* e^m \nabla u, \nabla q \rangle + (y^* ye^m \nabla u, \nabla p)
- \langle z, y^* \rangle
$$
The adjoint problem and the gradient

- The adjoint problem for \( u^*, m^*, p^*, v^*, q^*, y^* \)

\[
\sigma_{\text{noise}}^{-2} \mathcal{B}^* \mathcal{W} \mathcal{B} q^* - \nabla \cdot (y^* e^m \nabla p) - \nabla \cdot (e^m \nabla v^*) = 0
\]

\[
e^m \nabla q^* \cdot \nabla p + C_{pr}^{-1} y^* + y^* e^m \nabla u \cdot \nabla p + e^m \nabla u \cdot \nabla v^* = -z
\]

\[
- \nabla \cdot (e^m \nabla q^*) - \nabla \cdot (y^* e^m \nabla u) = 0
\]

\[
\sigma_{\text{noise}}^{-2} \mathcal{B}^* \mathcal{W} \mathcal{B} p^* - \nabla \cdot (m^* e^m \nabla p) - \nabla \cdot (e^m \nabla u^*) = b_1
\]

\[
e^m \nabla p^* \cdot \nabla p + C_{pr}^{-1} m^* + m^* e^m \nabla u \cdot \nabla p + e^m \nabla u \cdot \nabla u^* = b_2
\]

\[
- \nabla \cdot (e^m \nabla p^*) - \nabla \cdot (m^* e^m \nabla u) = b_3
\]

- \( u, m, p, q, v, y, \) and \( q^*, v^*, y^* \) appear in expressions for \( b_1, b_2, b_3 \)

- Gradient:

\[
g(w) = \Gamma_{\text{noise}}^{-1} (\mathcal{B} u - d) \odot \mathcal{B} p^* + \Gamma_{\text{noise}}^{-1} \mathcal{B} v \odot \mathcal{B} q^*
\]

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

\[
q^* = -v, \quad y^* = -y, \quad v^* = -q
\]
The OED objective function and the gradient

- Solve inner optimization \( \Rightarrow m(w), u(m(w)) \) and \( p(m(w)) \)
- Solve for \((v, y, q)\):
  \[
  \frac{\sigma_{\text{noise}}^{-2}}{B} W B v - \nabla \cdot (y e^m \nabla p) - \nabla \cdot (e^m \nabla q) = 0 \\
  e^m \nabla v \cdot \nabla p + C_{pr}^{-1} y + ye^m \nabla u \cdot \nabla p + e^m \nabla u \cdot \nabla q = z \\
  - \nabla \cdot (e^m \nabla v) - \nabla \cdot (ye^m \nabla u) = 0
  \]
- Objective function evaluation: \( \Theta(w) = \langle y, z \rangle \)
- Solve for \((p^*, m^*, u^*)\):
  \[
  \frac{\sigma_{\text{noise}}^{-2}}{B} W B p^* - \nabla \cdot (m^* e^m \nabla p) - \nabla \cdot (e^m \nabla u^*) = b_1 \\
  e^m \nabla p^* \cdot \nabla p + C_{pr}^{-1} m^* + m^* e^m \nabla u \cdot \nabla p + e^m \nabla u \cdot \nabla u^* = b_2 \\
  - \nabla \cdot (e^m \nabla p^*) - \nabla \cdot (m^* e^m \nabla u) = b_3
  \]
- Gradient: \( g(w) = \Gamma_{\text{noise}}^{-1} (B u - d) \odot B p^* - \Gamma_{\text{noise}}^{-1} B v \odot B v \)
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 $\sim 2 \times r \times n_{tr}$
3. Cost of computing OED gradient $\sim 2 \times n_{tr} + 2 \times r$

Low-rank approx to misfit Hessian $\implies$ 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

<table>
<thead>
<tr>
<th>truth</th>
<th>prior mean</th>
<th>prior variance</th>
</tr>
</thead>
</table>

- Prior mean $m_{\text{prior}}$ is a “smooth” least-squares fit to point measurements
- Prior covariance: $C_{pr} = (\mathcal{A} + \alpha \sum_{i=1}^{N} \delta_i)^{-2}$, \( \mathcal{A} m = -\nabla \cdot (D \nabla m) \)
- Draws from prior used to generate training data for OED
Computing an optimal design with sparsification

- Optimization problem with sparsity control
  
  \[
  \min_w \sum_i \langle z_i, y_i(w) \rangle + \gamma P_\varepsilon(w)
  \]

  where
  
  \[
  H(m_{\text{MAP}}(w)) y_i = z_i
  \]

  \[
  m_{\text{MAP}}(w) = \arg \min_\theta J\left(u(\theta(w)), \theta(w)\right)
  \]

- Continuation approach to approximate \(\ell_0\)-penalty

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

- Solve the problem with successively smaller values of \(\varepsilon\)
Effectiveness of the optimal design

- 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

Sources of sub-optimality:
- Trace estimation
- Use of training data for OED
- Gaussian approximation

\[
\text{relative misfit} = \frac{\| \hat{m}_{\text{MAP}}(\mathbf{w}) - m_{\text{true}} \|}{\| m_{\text{true}} \|}
\]
Scalability with respect to parameter/sensor dimension

- Objective func/grad (CG iters)
- Optimization (BFGS)

Parameter dim:
- iters
- func evals

Sensor dim:
- iters
- func evals
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, ...