Fast randomization methods for Bayesian optimal experimental design

Brad Marvin and Tan Bui-Thanh

Computational Engineering and Optimization (CEO) Group
Department of Aerospace Engineering and Engineering Mechanics
Institute for Computational Engineering and Sciences (ICES)
The University of Texas at Austin

AtMath conference, Nov 21, 2017
What is Inverse problem?

\[ u_1 w^2 + u_2 w + u_3 = 0 \]
What is Inverse problem?

\[ u_1 w^2 + u_2 w + u_3 = 0 \]

**Forward problem**

- **Given**: \( u_1, u_2, u_3 \)
- **Desire**: Compute solution \( w = \frac{-u_2 \pm \sqrt{u_2^2 - 4u_1u_3}}{2u_1} \)
What is Inverse problem?

Forward problem

- **Given**: \( u_1, u_2, u_3 \)
- **Desire**: Compute solution \( w = \frac{-u_2 \pm \sqrt{u_2^2 - 4u_1u_3}}{2u_1} \)

Inverse problem

- **Given**: \( w \)
- **Desire**: Compute \( u_1, u_2, u_3 = \)
### What is Inverse problem?

#### Forward problem
- **Given:** \( u_1, u_2, u_3 \)
- **Desire:** Compute solution \( w = \frac{-u_2 \pm \sqrt{u_2^2 - 4u_1u_3}}{2u_1} \)

#### Inverse problem
- **Given:** \( w \)
- **Desire:** Compute \( u_1, u_2, u_3 = ??? \)
What is Bayesian inversion method?

Maxwell Equations:

\[ \nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \quad \text{(Faraday)} \]
\[ \nabla \times \mathbf{H} = \varepsilon \frac{\partial \mathbf{E}}{\partial t}, \quad \text{(Ampere)} \]

\( \mathbf{E} \): Electric field, \( \mathbf{H} \): Magnetic field, \( \mu \): permeability, \( \varepsilon \): permittivity
What is Bayesian inversion method?

Maxwell Equations:

\[ \nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \quad \text{(Faraday)} \]

\[ \nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}, \quad \text{(Ampere)} \]

\( \mathbf{E} \): Electric field, \( \mathbf{H} \): Magnetic field, \( \mu \): permeability, \( \epsilon \): permittivity

Forward problem (discontinuous Galerkin discretization)

\[ y = F(u) \]

where \( G \) maps shape parameters \( u \) to electric/magnetic field \( y \) at the measurement points

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin
What is Bayesian inversion method?

Maxwell Equations:

\[ \nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \quad \text{(Faraday)} \]
\[ \nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}, \quad \text{(Ampere)} \]

**E**: Electric field, **H**: Magnetic field, **μ**: permeability, **ε**: permittivity

Forward problem (discontinuous Galerkin discretization)

\[ y = F(u) \]

where \( G \) maps shape parameters \( u \) to electric/magnetic field \( y \) at the measurement points

Inverse Problem

Given (possibly noise-corrupted) measurements on \( y \), infer \( u \)?
What is Bayesian inversion method?

**Bayes Theorem**

Solution to the inverse problem is given as a posterior PDF over parameter space:

\[ \pi_{\text{post}}(u|y_{\text{obs}}) \propto \pi_{\text{pr}}(u) \pi_{\text{like}}(y_{\text{obs}}|u) \]
What is Bayesian inversion method?

**Bayes Theorem**

Solution to the inverse problem is given as a posterior PDF over parameter space:

\[ \pi_{\text{post}}(u|y_{\text{obs}}) \propto \pi_{\text{pr}}(u)\pi_{\text{like}}(y_{\text{obs}}|u) \]

**Prior knowledge:** The obstacle is smooth:

\[ \pi_{\text{pr}}(u) \propto \exp \left( -\lambda \int_0^{2\pi} r''(u) d\theta \right) \]
What is Bayesian inversion method?

**Bayes Theorem**
Solution to the inverse problem is given as a posterior PDF over parameter space:

\[
\pi_{\text{post}}(u|y_{\text{obs}}) \propto \pi_{\text{pr}}(u) \pi_{\text{like}}(y_{\text{obs}}|u)
\]

**Prior knowledge:** The obstacle is smooth:

\[
\pi_{\text{pr}}(u) \propto \exp \left( -\lambda \int_0^{2\pi} r''(u) d\theta \right)
\]

**Likelihood:** Additive Gaussian noise, for example,

\[
\pi_{\text{like}}(y_{\text{obs}}|u) \propto \exp \left( -\frac{1}{2} \| G(u) - y_{\text{obs}} \|_{\Sigma_{\text{noise}}}^2 \right)
\]
Inverse wave propagation
Full wave form seismic inversion

Randomness

- Random errors in seismometer measurements are unavoidable
- Inadequacy of the mathematical model (elastodynamics)

Challenge

How to image the earth interior using forward computational model with \( \mathcal{O}(10^9) \) degree of freedoms?

(Marvin and Bui-Thanh)
Bayes Theorem (finite dimensions)

Solution to the inverse problem is given as a posterior PDF over parameter space:

$$\pi_{\text{post}}(u|d) \propto \pi_{\text{like}}(d|u) \times \pi_{\text{prior}}(u)$$
Bayes Theorem (finite dimensions)

Solution to the inverse problem is given as a posterior PDF over parameter space:

\[ \pi_{post}(u|d) \propto \pi_{like}(d|u) \times \pi_{prior}(u) \]

Bayes Theorem (infinite dimensions)

\[ \frac{\partial \nu}{\partial \mu} (u|d) \propto \exp(-J(u,d)) = \exp \left( -\frac{1}{2} \|d - F(u)\|_{R^K}^2 \right) \]

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin 6 / 47
FEM Discretization-invariant MCMC
2D elliptic inverse problems

Table: Average acceptance rate of the Metropolize-then-discretize MALA and the MALA as the mesh is refined.

<table>
<thead>
<tr>
<th></th>
<th>MTD MALA</th>
<th>MALA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$</td>
<td>0.63098</td>
<td>0.01000</td>
</tr>
<tr>
<td>$h/2$</td>
<td>0.63137</td>
<td>0.00235</td>
</tr>
<tr>
<td>$h/4$</td>
<td>0.63627</td>
<td>0.00549</td>
</tr>
</tbody>
</table>

Table: Average acceptance rate of the Metropolize-then-discretize (MTD) HMC, the discretize-then-Metropolize (DTM) HMC, and the prior-conditioned standard HMC as the mesh is refined.

<table>
<thead>
<tr>
<th></th>
<th>MTD HMC</th>
<th>HMC</th>
<th>prior-preconditioned standard HMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$</td>
<td>0.82</td>
<td>0.017</td>
<td>0.511</td>
</tr>
<tr>
<td>$h/2$</td>
<td>0.76</td>
<td>0.014</td>
<td>0.288</td>
</tr>
<tr>
<td>$h/4$</td>
<td>0.79</td>
<td>0.018</td>
<td>0.279</td>
</tr>
</tbody>
</table>
Discrete Monge-Kantorovich optimal transport

- $N$ prior particles: $\{u_j\}_{j=1}^{N} \sim \mu$
- Discrete Monge-Kantorovich transport problem

$$\min_{T} \sum_{i=1,j=1}^{N} \| u_i - u_j \|_2^2 M_{ij},$$

subject to

$$\sum_{i=1}^{N} T_{ij} = \frac{1}{N}, \quad \sum_{j=1}^{N} T_{ij} = w_i$$

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin 8 / 47
Gaussianization: An example of global seismic inversion

- **inversion field**: $c_p$ in acoustic wave equation
- **prior mean**: PREM (radially symmetric model)
- **“truth” model**: S20RTS (Ritsema et al.), (laterally heterogeneous)
- Piecewise-trilinear on same mesh as forward/adjoint 3rd order dG fields
- **dimensions**: 1.07 million parameters, 630 million field unknowns
- **Final time**: $T = 1000s$ with 2400 time steps
- A single forward solve takes 1 minute on 64K Jaguar cores

"truth", sources (black)  
MAP, receivers (white)  
Hessian eigenvalues

(Marvin and Bui-Thanh)  
OED for data assimilation  
ICES, UT Austin
Gaussianization: Uncertainty estimation

\[ C \approx C_0 - \frac{C_1}{2} V_r D_r V_r^* C_0^{1/2} \]
Randomized Misfit Approach for Big Data

\[ J = \frac{1}{2} \left\| \mathcal{L}^{-\frac{1}{2}} (d - F(u)) \right\|^2 + \frac{1}{2} \left\| u - u_0 \right\|_C^2, \quad \mathbb{E} [\varepsilon] = 0, \quad \mathbb{E} [\varepsilon \varepsilon^T] = \mathbf{I} \]

\[ = \frac{1}{2} \mathbb{E}_\varepsilon \left( \varepsilon^T \mathcal{L}^{-\frac{1}{2}} (d - F(u)) \right)^2 + \frac{1}{2} \left\| u - u_0 \right\|_C^2 \]

Monte Carlo

\[ \approx \frac{1}{2N} \sum_{j=1}^{N} \left( \varepsilon_j^T \mathcal{L}^{-\frac{1}{2}} (d - F(u)) \right)^2 + \frac{1}{2} \left\| u - u_0 \right\|_C^2 \]

\[ = \frac{1}{2} \left\| \Sigma^T \mathcal{L}^{-\frac{1}{2}} (d - F(u)) \right\|^2 + \frac{1}{2} \left\| u - u_0 \right\|_C^2 =: \tilde{J}_N \]

where

\[ \Sigma := \frac{1}{\sqrt{N}} \left[ \varepsilon_1, \ldots, \varepsilon_N \right] \in \mathbb{R}^{d \times N} \]

if \( N \ll d \) ⇒ substantially reducing the data

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin 11 / 47
Randomized Misfit Approach for Big Data

Numerical results for 2D elliptic inverse problem

\[ \tilde{u}^{MAP} = \arg \min_{\tilde{u}} \tilde{J}_N \]

\[ u^{MAP} = \arg \min_u J \]

\( N = 51, \ 67\% \) sparse

\( d = 1333 \)
Randomized Likelihood Method for Big Data

Numerical results for 2D elliptic inverse problem

\[ \tilde{u}^{MAP} = \arg\min_u \tilde{J}_N \]

\[ u^{MAP} = \arg\min_u J \]

\( N = 101, \, 67\% \text{ sparse} \)

\( d = 1333 \)
Randomized Likelihood Method for Big Data

Numerical results for 3D elliptic inverse problem

\[ \tilde{u}^{MAP} = \arg\min_u \tilde{J}_N \]

\[ u^{MAP} = \arg\min_u J \]

\( N = 50, \text{ 67\% sparse} \)

\( d = 2474 \)
Randomized Likelihood Method for Big Data
Numerical results for 3D elliptic inverse problem

\[ \tilde{u}^{MAP} = \arg\min_u \tilde{J}_N \]

\[ u^{MAP} = \arg\min_u J \]

\( N = 100, \ 67\% \ \text{sparse} \)

\( d = 2474 \)
Outline

1. A model inadequacy approach to data assimilation

2. OED for data assimilation
   - Exact OED
   - Upper bounds for OED

3. Scalable approximations for OED
   - Randomization + Gauss quadrature + Lanczos

4. Greedy with upper bounds

5. Conclusions
One step data assimilation

Model inadequacy as Bayesian inference (Extension of Nguyen et al. 2015)

\[ b(u, v) = f(v) + g(v), \quad \forall v \in V, \]
One step data assimilation

Model inadequacy as Bayesian inference (Extension of Nguyen et al. 2015)

\[ b(u, v) = f(v) + g(v), \quad \forall v \in V, \]

Infer \( g \) given noisy observations \( y \).
Postulate a Gaussian prior measure \( \mu := \mathcal{N}(g_0, C_0) \) on \( g \)
Postulate a Gaussian prior measure $\mu := \mathcal{N}(g_0, C_0)$ on $g$

Prior Covariance

$C_0 := \alpha^{-1} (I - \Delta)^{-s} =: \alpha^{-1} A^{-s}$

where

$D(A) := \left\{ u \in H^2(\Omega) : \frac{\partial u}{\partial n} = 0 \text{ on } \partial \Omega \right\}$. 
Postulate a Gaussian prior measure \( \mu := \mathcal{N}(g_0, C_0) \) on \( g \)

Prior Covariance

\[
C_0 := \alpha^{-1} (I - \Delta)^{-s} =: \alpha^{-1} A^{-s}
\]

where

\[
D(A) := \left\{ u \in H^2(\Omega) : \frac{\partial u}{\partial n} = 0 \text{ on } \partial \Omega \right\}.
\]

We choose \( s > n/2 \) (\( n \) is the spatial dimension) to ensure \( \nu \ll \mu \) via the Radon-Nikodym derivative

\[
\frac{d\nu}{d\mu}(g|y) \propto \exp \left( -\frac{1}{2\sigma^2} \| y - Fg \|^2 \right)
\]
Linear forward problem

posterior of $g$ and $u$

$$
\nu = \mathcal{N}(\bar{g}, C_g)
$$

$$
\bar{g} = C_g \left( \frac{1}{\sigma^2} F^* d + C_0^{-1} g_0 \right)
$$

$$
C_g = \left( \frac{1}{\sigma^2} F^* F + C_0^{-1} \right)^{-1}
$$
Linear forward problem

Posterior of $g$ and $u$

$$\nu = \mathcal{N}(\bar{g}, \mathcal{C}_g)$$

$$\bar{g} = \mathcal{C}_g \left( \frac{1}{\sigma^2} F^* d + \mathcal{C}_0^{-1} g_0 \right)$$

$$\mathcal{C}_g = \left( \frac{1}{\sigma^2} F^* F + \mathcal{C}_0^{-1} \right)^{-1}$$

Posterior of state $u \sim \mathcal{N}(\bar{u}, \mathcal{C}_u)$

$$\bar{u} = B^{-1} (\bar{g} + f)$$

$$\mathcal{C}_u = B^{-1} \mathcal{C}_g B^{-*} = \mathbf{B}^{-1} \left( \frac{1}{\sigma^2} F^* F + \mathcal{C}_0^{-1} \right)^{-1} B^{-*}$$
Linear forward model

\[-\nabla \cdot (e^w \nabla u) = 0 \quad \text{in } \Omega\]
\[-e^w \nabla u \cdot \mathbf{n} = Bi \, u \quad \text{on } \partial\Omega^{Bi}\]
\[-e^w \nabla u \cdot \mathbf{n} = -1 \quad \text{on } \partial\Omega^{RHS}\]
Posterior state versus the best-knowledge state

Noise: $\sigma_{obs} = 0.01$

Figure: $|\overline{u} - u^\dagger|$

Figure: $|u_{bk} - u^\dagger|$
Pointwise posterior standard deviation in state Noise: $\sigma_{\text{obs}} = 0.01$
Convection-diffusion equation

\[ \frac{\partial u}{\partial t} - \nabla \cdot (d \nabla u) + v \cdot \nabla u = 0 \quad \text{in} \ \Omega \]

\[ u = g \quad \text{on} \ \Gamma_{in} \]

\[ u(t = 0) = u_0 \quad \text{in} \ \Omega \]
Model Errors

Diffusion coefficient:

Velocity field:

Best Knowledge

Exact model

(Marvin and Bui-Thanh)

OED for data assimilation

ICES, UT Austin
Error introduced in diffusion coefficient

Error in best knowledge state:

Error in posterior State:

\[ t = 1.0 \]

(Marvin and Bui-Thanh)

\[ t = 1.8 \]

OED for data assimilation

\[ t = 2.5 \]

ICES, UT Austin
Error introduced in velocity field
Error in best knowledge state:

Error in posterior State:

\( t = 1.0 \)  \( t = 1.8 \)  \( t = 2.5 \)

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin
Error introduced in both terms

Error in best knowledge state:

Error in posterior State:

\[ t = 1.0 \]  \hspace{1cm}  \[ t = 1.8 \]  \hspace{1cm}  \[ t = 2.5 \]  

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin
Outline

1. A model inadequacy approach to data assimilation
2. OED for data assimilation
   - Exact OED
   - Upper bounds for OED
3. Scalable approximations for OED
   - Randomization + Gauss quadrature + Lanczos
4. Greedy with upper bounds
5. Conclusions

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin
Bayesian A-optimal design

- $0 \leq w \leq 1$: the weight vector at all the possible sensor locations
- $W$: diagonal matrix containing the weight vector $w$
- $B$: forward equation operator
- $C_u = B^{-1} \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} B^{-*}$
Bayesian A-optimal design

- $0 \leq \mathbf{w} \leq 1$: the weight vector at all the possible sensor locations
- $W$: diagonal matrix containing the weight vector $\mathbf{w}$
- $\mathbf{B}$: forward equation operator
- $\mathbf{C}_u = \mathbf{B}^{-1} \left( \frac{1}{\sigma^2} \mathbf{F}^* \mathbf{W} \mathbf{F} + \mathbf{C}_0^{-1} \right)^{-1} \mathbf{B}^{-*}$

Minimize the uncertainty in predicting $\mathbf{u}$

$$\min_{\mathbf{w}} \text{Tr} [\mathbf{C}_u] = \text{Tr} \left[ \mathbf{B}^{-1} \left( \frac{1}{\sigma^2} \mathbf{F}^* \mathbf{W} \mathbf{F} + \mathbf{C}_0^{-1} \right)^{-1} \mathbf{B}^{-*} \right]$$

- An action of $\mathbf{B}^{-1}$ or $\mathbf{F}$ is a forward PDE solve

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin 29 / 47
Bayesian A-optimal design

A numerical Result

\( \ell_1 \) regularization to promote sparsity

\[
\min_w \text{Tr} \left[ B^{-1} \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} B^{-*} \right] + \kappa \sum_{i=1}^{n_{\text{obs}}} w_i
\]

subject to \( 0 \leq w \leq 1 \)
Bayesian A-optimal design

A numerical Result

\[ \ell_1 \text{ regularization to promote sparsity} \]

\[
\min_w Tr \left[ B^{-1} \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} B^{-*} \right] + \kappa \sum_{i=1}^{n_{obs}} w_i \\
\text{subject to } 0 \leq w \leq 1
\]

- 961 possible observations and 961 states
- Pick 24 sensors
- Solve the optimization with trust region inexact reflective Newton CG
- \( \sigma^2 = 0.0001 \)
- \( \kappa = 1 \)
Figure: $|\bar{u} - u^\dagger|$
First upper bound for Bayesian A-optimal design

\[ C_u = B^{-1} \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} B^{-*} \]
First upper bound for Bayesian A-optimal design

\[ C_u = B^{-1} \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} B^{-1} \]

\[ C_g = \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} \]
First upper bound for Bayesian A-optimal design

\[ C_u = B^{-1} \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} B^{-*} \]

\[ C_g = \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} \]

We have, by cyclic rotation invariant + trace inequality for SPD matrices,

\[
\min_w Tr [C_u] \leq Tr [B^{-*} B^{-1}] Tr [C_g] \leq c \times Tr [C_g]
\]
First upper bound for Bayesian A-optimal design

\[ C_u = B^{-1} \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} B^{-1} \]

\[ C_g = \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} \]

We have, by cyclic rotation invariant + trace inequality for SPD matrices,

\[ \min_w Tr \left[ C_u \right] \leq Tr \left( B^{-1} B^{-1} \right) Tr \left[ C_g \right] \leq c \times Tr \left[ C_g \right] \]

Thus, minimize the uncertainty in \( g \) instead

\[ \min_w Tr \left[ C_g \right] = Tr \left[ \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} \right] \]
First upper bound for Bayesian A-optimal design

\[ \mathcal{C}_u = \mathbf{B}^{-1} \left( \frac{1}{\sigma^2} \mathbf{F}^* \mathbf{W} \mathbf{F} + \mathbf{C}_0^{-1} \right)^{-1} \mathbf{B}^{-*} \]

\[ \mathcal{C}_g = \left( \frac{1}{\sigma^2} \mathbf{F}^* \mathbf{W} \mathbf{F} + \mathbf{C}_0^{-1} \right)^{-1} \]

We have, by cyclic rotation invariant + trace inequality for SPD matrices,

\[
\min_{\mathbf{w}} \text{Tr} [\mathcal{C}_u] \leq \text{Tr} \left[ \mathbf{B}^{-*} \mathbf{B}^{-1} \right] \text{Tr} [\mathcal{C}_g] \leq c \times \text{Tr} [\mathcal{C}_g]
\]

Thus, minimize the uncertainty in \( g \) instead

\[
\min_{\mathbf{w}} \text{Tr} [\mathcal{C}_g] = \text{Tr} \left[ \left( \frac{1}{\sigma^2} \mathbf{F}^* \mathbf{W} \mathbf{F} + \mathbf{C}_0^{-1} \right)^{-1} \right]
\]

\[ \text{Avoid the forward solve } \mathbf{B}^{-1} \]
First upper bound for Bayesian A-optimal design

A numerical Result

Figure: $|\overline{u} - u^\dagger|$  

Figure: pointwise standard dev. in $u$  

(Marvin and Bui-Thanh)  

OED for data assimilation  

ICES, UT Austin
Second upper bound for Bayesian A-optimal design

\[ C_g = \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} = C_0^{\frac{1}{2}} \left( \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \right)^{-1} C_0^{\frac{1}{2}} \]
Second upper bound for Bayesian A-optimal design

\[ C_g = \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} = C_0^{1/2} \left( \frac{1}{\sigma^2} C_0^{1/2} F^* W F C_0^{1/2} + I \right)^{-1} C_0^{1/2} \]

Define \( C := \left( \frac{1}{\sigma^2} C_0^{1/2} F^* W F C_0^{1/2} + I \right)^{-1} \)

We have, by cyclic rotation invariant + trace inequality for SPD matrices,

\[
\min_w Tr [C_g] \leq Tr [C_0] \cdot Tr [C] \leq c \times Tr [C]
\]
Second upper bound for Bayesian A-optimal design

\[ C_g = \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} = C_0^{\frac{1}{2}} \left( \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \right)^{-1} C_0^{\frac{1}{2}} \]

Define \( C := \left( \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \right)^{-1} \)

We have, by cyclic rotation invariant + trace inequality for SPD matrices,

\[ \min_w Tr [C_g] \leq Tr [C_0] Tr [C] \leq c \times Tr [C] \]
Second upper bound for Bayesian A-optimal design

\[ C_g = \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} = C_0^{\frac{1}{2}} \left( \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \right)^{-1} C_0^{\frac{1}{2}} \]

Define \( C := \left( \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \right)^{-1} \)

We have, by cyclic rotation invariant + trace inequality for SPD matrices,

\[ \min_w Tr [C_g] \leq Tr [C_0] Tr [C] \leq c \times Tr [C] \]

Thus, minimize the trace of \( C \) instead

\[ \min_w Tr [C] = Tr \left[ \left( \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \right)^{-1} \right] \]
Second upper bound for Bayesian A-optimal design

\[ C_g = \left( \frac{1}{\sigma^2} F^* W F + C_0^{-1} \right)^{-1} = C_0^{1/2} \left( \frac{1}{\sigma^2} C_0^{1/2} F^* W F C_0^{1/2} + I \right)^{-1} C_0^{1/2} \]

Define \( C := \left( \frac{1}{\sigma^2} C_0^{1/2} F^* W F C_0^{1/2} + I \right)^{-1} \)

We have, by cyclic rotation invariant + trace inequality for SPD matrices,

\[ \min_w Tr[C_g] \leq Tr[C_0] Tr[C] \leq c \times Tr[C] \]

Thus, minimize the trace of \( C \) instead

\[ \min_w Tr[C] = Tr \left[ \left( \frac{1}{\sigma^2} C_0^{1/2} F^* W F C_0^{1/2} + I \right)^{-1} \right] \]

Additionally avoid the repeated application of the prior \( C_0 \)
First upper bound for Bayesian A-optimal design
A numerical Result

Figure: $|\overline{u} - u^\dagger|$  
Figure: pointwise standard dev. in $u$

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin
Third upper bounds for Bayesian A-optimal design

- $A \in \mathbb{R}^{n \times n}$: SPD
- $\mu_1 = \text{Tr}[A]$ 
- $\mu_2 = \|A\|_F^2$ 
- $\alpha$: a lower bound of the spectrum of $A$

Upper bound for trace of an inverse SPD matrix (Bai and Golub, 1996)

$$\text{Tr}[A^{-1}] \leq (\mu_1 \ n) \begin{pmatrix} \mu_2 & \mu_1 \\ \alpha^2 & \alpha \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix}$$
Third upper bounds for Bayesian A-optimal design

- \( A \in \mathbb{R}^{n \times n} \): SPD
- \( \mu_1 = Tr[A] \)
- \( \mu_2 = \|A\|_F^2 \)
- \( \alpha \): a lower bound of the spectrum of \( A \)

Upper bound for trace of an inverse SPD matrix (Bai and Golub, 1996)

\[
Tr[A^{-1}] \leq \left( \begin{array}{cc} \mu_1 & n \end{array} \right) \left( \begin{array}{cc} \mu_2 & \mu_1 \\ \alpha^2 & \alpha \end{array} \right)^{-1} \left( \begin{array}{c} n \\ 1 \end{array} \right)
\]

Can be applied: \( A^{-1} = \{C_u, C_g, C\} \) to avoid inverse of the posterior!

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin 36 / 47
Third upper bounds for Bayesian A-optimal design

An example

- \( A := C^{-1} = \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \)
- \( \mu_1 = Tr[A] \)
- \( \mu_2 = \|A\|^2_F \)
- \( \alpha = 1 \)

Upper bound for \( Tr[C] \)

\[
Tr[C] \leq \begin{pmatrix} \mu_1 & n \end{pmatrix} \begin{pmatrix} \mu_2 & \mu_1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix} =: J
\]
An example

- $A := C^{-1} = \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I$
- $\mu_1 = Tr [A]$
- $\mu_2 = \|A\|_F^2$
- $\alpha = 1$

Upper bound for $Tr [C]$

$$Tr [C] \leq \begin{pmatrix} \mu_1 & n \end{pmatrix} \begin{pmatrix} \mu_2 & \mu_1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix} =: J$$

Minimizing the upper bound $J$ instead!
Third upper bounds for Bayesian A-optimal design

An example

\[ A := c^{-1} = \frac{1}{\sigma^2} c_0^{1/2} F^* W F c_0^{1/2} + I \]

\[ \mu_1 = Tr[A] \]

\[ \mu_2 = \| A \|_F^2 \]

\[ \alpha = 1 \]

Upper bound for \( Tr[C] \)

\[ Tr[C] \leq (\mu_1 \ n) \begin{pmatrix} \mu_2 & \mu_1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix} =: J \]

Minimizing the upper bound \( J \) instead!

- Avoid the INVERSE completely!
Third upper bounds for Bayesian A-optimal design
A numerical Result

Figure: $| \mathbf{u} - \mathbf{u}^\dagger |$

Figure: pointwise standard dev. in $\mathbf{u}$
Outline

1. A model inadequacy approach to data assimilation

2. OED for data assimilation
   - Exact OED
   - Upper bounds for OED

3. Scalable approximations for OED
   - Randomization + Gauss quadrature + Lanczos

4. Greedy with upper bounds

5. Conclusions
Randomized trace estimators

Randomized trace (Hutchinson 1989)

- \( A \in \mathbb{R}^{n \times n} \): symmetric
- \( z \): random vector with zero mean and identity covariance
- Then

\[
Tr[A] = \mathbb{E}[z^T A z]
\]

Randomized trace + MC (Bai, Fahey, and Golub 1996 AND others...)

\[
Tr[A] \approx \frac{1}{N} \sum_{i=1}^{N} z_i^T A z_i
\]
Randomized trace estimators

Randomized trace (Hutchinson 1989)

- $A \in \mathbb{R}^{n \times n}$: symmetric
- $z$: random vector with zero mean and identity covariance
- Then
  \[
  Tr[A] = \mathbb{E}[z^T Az]
  \]

Randomized trace + MC (Bai, Fahey, and Golub 1996 AND others...)

- $Tr[A] \approx \frac{1}{N} \sum_{i=1}^{N} z_i^T A z_i$

Can be applied to: $Tr[C_u], Tr[C_g], Tr[C]$

(Marvin and Bui-Thanh)
Randomized trace estimators

Example: \( A = C = \left( \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \right)^{-1} \)

Challenges with the **INVERSE** for large-scale forward PDEs

- randomized SVD + Sherman-Woodbury (Alexanderian *et al.* 2014)
- **But** need this for every random vector and optimization iteration.
- **Furthermore**, expensive for Newton methods!

Our approach: Gaussian Quadrature + Lanczos (Bai *et al.* 1996)

- Let \( A = \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \) and define \( f(x) = x^{-1} \)
- Then

\[
z_i^T f(A) z_i = \int_a^b f(\lambda) \, d\mu(\lambda)
\]
Randomized trace estimators

Example: \[ A = C = \left( \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W FC_0^{\frac{1}{2}} + I \right)^{-1} \]

Challenges with the INVERSE for large-scale forward PDEs

- randomized SVD + Sherman-Woodbury (Alexanderian et al. 2014)
- But need this for every random vector and optimization iteration.
- Furthermore, expensive for Newton methods!

Our approach: Gaussian Quadrature + Lanczos (Bai et al. 1996)

- Let \[ A = \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W FC_0^{\frac{1}{2}} + I \] and define \[ f(x) = x^{-1} \]
- Then

\[
\begin{align*}
    z_i^T f(A) z_i &= \int_a^b f(\lambda) \, d\mu(\lambda)
\end{align*}
\]

- Gauss quadrature + Lanczos DOES NOT need to know \( d\mu \) nor \( a, b \)
Randomized trace estimators

Example: \( A = C = \left( \frac{1}{\sigma^2} C_0^{1/2} F^* W F C_0^{1/2} + I \right)^{-1} \)

Challenges with the INVERSE for large-scale forward PDEs

- randomized SVD + Sherman-Woodbury (Alexanderian et al. 2014)
- **But** need this for every random vector and optimization iteration.
- Furthermore, expensive for Newton methods!

Our approach: Gaussian Quadrature + Lanczos (Bai et al. 1996)

- Let \( A = \frac{1}{\sigma^2} C_0^{1/2} F^* W F C_0^{1/2} + I \) and define \( f(x) = x^{-1} \)
- Then

\[ z_i^T f(A) z_i = \int_a^b f(\lambda) \, d\mu(\lambda) \]

- Gauss quadrature + Lanczos **DOES NOT** need to know \( d\mu \) nor \( a, b \)
- Gauss quadrature + Lanczos **ONLY NEEDS** \( A * v \)
Randomized trace + Gauss quadrature + Lanczos
A numerical Result

Figure: $|\mathbf{u} - \mathbf{u}^\dagger|$  
Figure: pointwise standard dev. in $\mathbf{u}$

(Marvin and Bui-Thanh)  
OED for data assimilation  
ICES, UT Austin
Outline

1. A model inadequacy approach to data assimilation

2. OED for data assimilation
   - Exact OED
   - Upper bounds for OED

3. Scalable approximations for OED
   - Randomization + Gauss quadrature + Lanczos

4. Greedy with upper bounds

5. Conclusions
Greedy with upper bounds

Main idea

Recall the upper bound

\[
\text{Tr} \left[ A^{-1} \right] \leq J := \begin{pmatrix} \mu_1 & n \end{pmatrix} \begin{pmatrix} \mu_2 & \mu_1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix}
\]

where \( A = \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \), \( \mu_1 = \text{Tr} \left[ A \right] \) and \( \mu_2 = ||A||_F^2 \).
Greedy with upper bounds

Main idea

Recall the upper bound

\[
Tr[A^{-1}] \leq J := \begin{pmatrix} \mu_1 & n \\ \mu_2 & \mu_1 \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix}
\]

where \( A = \frac{1}{\sigma^2} C_0^{1/2} F^* W F C_0^{1/2} + I \), \( \mu_1 = Tr[A] \) and \( \mu_2 = ||A||_F^2 \)

Greedy sensor locations

- **Start**

\[
w_{opt} = [0, \ldots, 0], \quad \text{Optimal-set} = I = \emptyset
\]

(Marvin and Bui-Thanh) OED for data assimilation ICES, UT Austin 44 / 47
Greedy with upper bounds

Main idea

Recall the upper bound

\[
Tr[A^{-1}] \leq J := \begin{pmatrix} \mu_1 & n \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix}
\]

where \( A = \frac{1}{\sigma^2} C_0^{\frac{1}{2}} F^* W F C_0^{\frac{1}{2}} + I \), \( \mu_1 = Tr[A] \) and \( \mu_2 = \|A\|_F^2 \).

Greedy sensor locations

- **Start**
  \[ w_{opt} = [0, \ldots, 0], \quad \text{Optimal-set} = I = \emptyset \]

- **Step 1:** For \( i \notin I \), compute the the upper bound \( J_i \) for the sensor set \( I \cup \{i\} \)
Greedy with upper bounds

Main idea

Recall the upper bound

\[ \text{Tr} \left[ A^{-1} \right] \leq J := \begin{pmatrix} \mu_1 & n \end{pmatrix} \begin{pmatrix} \mu_2 & \mu_1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix} \]

where \( A = \frac{1}{\sigma^2} C_0^{1/2} F^* W F C_0^{1/2} + I \), \( \mu_1 = \text{Tr} [A] \) and \( \mu_2 = \|A\|_F^2 \).

Greedy sensor locations

- **Start**
  
  \( w_{opt} = [0, \ldots, 0] \), \( \text{Optimal-set} = I = \emptyset \)

- **Step 1:** For \( i \not\in I \), compute the the upper bound \( J_i \) for the sensor set \( I \cup \{i\} \)

- **Step 2:** Compute \( i^* = \min_i J_i \)
Greedy with upper bounds

Main idea

Recall the upper bound

\[ \text{Tr} \left[ A^{-1} \right] \leq J := \begin{pmatrix} \mu_1 & n \end{pmatrix} \begin{pmatrix} \mu_2 & \mu_1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} n \\ 1 \end{pmatrix} \]

where \( A = \frac{1}{\sigma^2} C_0^2 F^* W F C_0^2 + I \), \( \mu_1 = \text{Tr} [A] \) and \( \mu_2 = \|A\|_F^2 \).

Greedy sensor locations

- **Start**
  \[ w_{opt} = [0, \ldots, 0], \quad \text{Optimal-set} = I = \emptyset \]

- **Step 1**: For \( i \notin I \), compute the upper bound \( J_i \) for the sensor set \( I \cup \{i\} \)

- **Step 2**: Compute \( i^* = \min_i J_i \)

- **Step 3**: Enrich the optimal set \( I = I \cup \{i^*\} \). Then goto **Step 1**
Greedy with upper bounds
A numerical result

Figure: $|\bar{u} - u^\dagger|$
1. A model inadequacy approach to data assimilation

2. OED for data assimilation
   - Exact OED
   - Upper bounds for OED

3. Scalable approximations for OED
   - Randomization + Gauss quadrature + Lanczos

4. Greedy with upper bounds

5. Conclusions
## Summary and future work

### Summary

1. Pose model inadequacy as Bayesian inference
2. Use model inadequacy for data assimilation
3. Look at various approximations/bounds of the A-OED design
4. Use approximations/bounds as inexpensive proxy for optimization

### Future work

1. Thorough understanding and comparison of all approximations
2. Optimization solver
3. $\ell_0$ regularization
4. Large-scale PDE problems
5. OED for time dependent PDEs
6. Nonlinear problems