Robust Optimal Design of Large-Scale Nonlinear Bayesian Inverse Problems

Abhijit Chowdhary [1] Ahmed Attia [2] Alen Alexanderian [1]

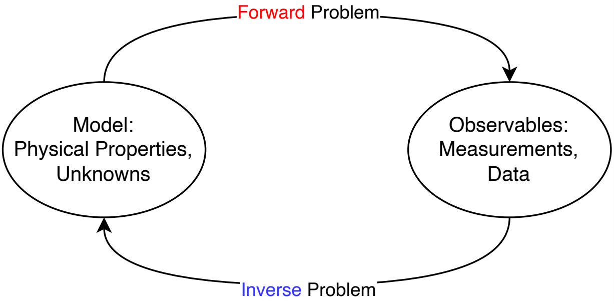
[1]North Carolina State University

[2]Argonne National Laboratory

November 16, 2024

Introduction

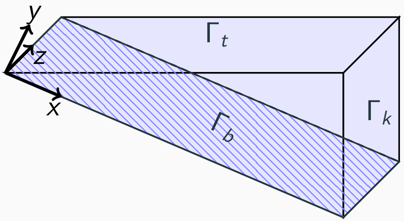
Introduction: Motivation

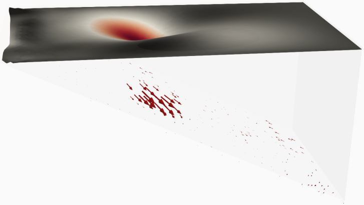


Applications

1. Fault-slip inference
2. Contaminant-source identification
3. Permeability field inversion in porous medium flow
4. Epidemic model calibration

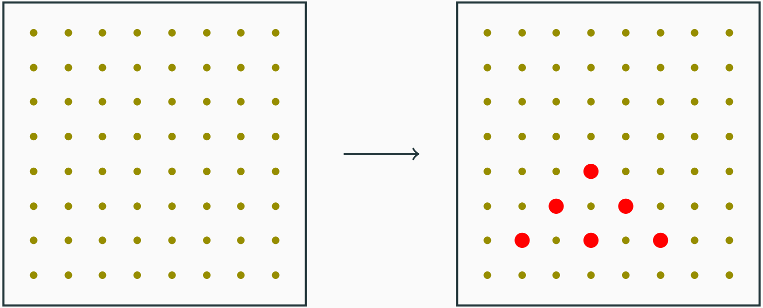
Example: 3D Linear Elasticity Modeling Fault Slip Inference





$$
\begin{aligned}
-\nabla \cdot [\mu(\nabla \mathbf{u} + (\nabla \mathbf{u})^T) + \lambda \nabla \cdot \mathbf{u}\mathbf{I}] &= \mathbf{0} && \text{in } \Omega, \\
\sigma(\mathbf{u})\mathbf{n} &= 0 && \text{on } \Gamma_t, \\
\mathbf{u} + \beta_k \sigma(\mathbf{u})\mathbf{n} &= \mathbf{0} && \text{on } \Gamma_k, \\
\mathbf{u} + \beta_s \sigma(\mathbf{u})\mathbf{n} &= \mathbf{h} && \text{on } \Gamma_s, \\
\mathbf{u} \cdot \mathbf{n} &= 0 && \text{on } \Gamma_b, \\
\delta \mathbf{T}(\sigma(\mathbf{u})\mathbf{n}) + \mathbf{T}\mathbf{u} &= \mathbf{m} && \text{on } \Gamma_b.
\end{aligned}
$$

Selecting the Optimal Experimental Design



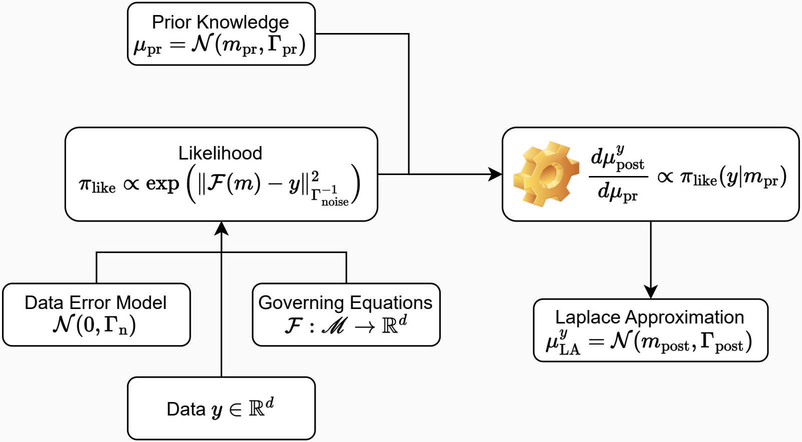
Optimal Experimental Design for Sensor Placement

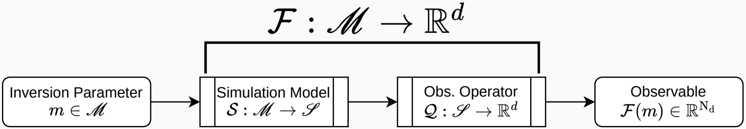
Q: Where should I allocate my sensor budget?

Goal: Enable robust (worst-case) optimal experimental design for infinite-dimensional nonlinear Bayesian inverse problems.

PDE-Constrained Bayesian Inversion

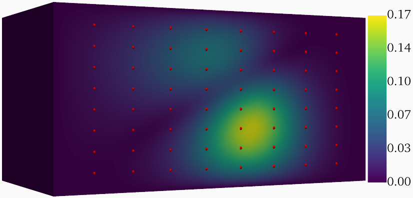
High Level Overview of Bayesian Inversion





Definition

The map $\mathcal{F} : \mathscr{M} \to \mathbb{R}^{N_d}$ is the parameter to observable map.



Governing Equations

For some appropriate space $\mathscr{U}$, consider the abstract weak PDE forumlation: Given $m \in \mathscr{M}$, find $u \in \mathscr{U}$ such that

$$a(u, p, m) = \big\langle p, \mathcal{A}(u, p) \big\rangle = 0, \quad \forall p \in \mathscr{V}, \tag{1}$$

- $u \in \mathscr{U}$ is the state variable,
- $m \in \mathscr{M}$ is the parameter to be inferred,
- $p \in \mathscr{V}$ is the test function,
- $a : \mathscr{U} \times \mathscr{V} \times \mathscr{M} \to \mathbb{R}$ is the weak form of the PDE

MAP Estimation and the Laplace Approximation

Definition (MAP Point)

Define the functional

$$\Phi(m) = \frac{1}{2}\left\|\mathcal{F}(m) - \mathbf{y}\right\|^2_{\Gamma_{\mathrm{n}}^{-1}} + \frac{1}{2}\left\|m - m_{\mathrm{pr}}\right\|^2_{\mathcal{C}_{\mathrm{pr}}^{-1}}$$

The **MAP point** m_{post} is the minimizer of Φ.

Definition (The Laplace Approximation)

The Laplace Approximation to the posterior measure is

$$\hat{\mu}^{\mathbf{y}}_{\mathrm{post}} = \mathcal{N}(m_{\mathrm{post}}, \mathcal{C}_{\mathrm{post}})$$

where $\mathcal{C}_{\mathrm{post}}^{-1} := \mathcal{H}_{\mathrm{m}}(m_{\mathrm{post}}) + \mathcal{C}_{\mathrm{pr}}^{-1}$ and

$$\mathcal{H}_{\mathrm{m}}(m_{\mathrm{post}}) := \mathcal{J}(m_{\mathrm{post}})^T \Gamma_{\mathrm{n}}^{-1} \mathcal{J}(m_{\mathrm{post}}).$$

Adjoint-Based Gradient Computation

Derivatives of Φ can be understood using the adjoint method.

$$\mathcal{L}(u, m, p) = \frac{1}{2}\|\mathbf{y} - \mathcal{Q}u\|^2_{\Gamma_{\mathrm{n}}^{-1}} + \frac{1}{2}\left\|m - m_{\mathrm{pr}}\right\|^2_{\mathcal{C}_{\mathrm{pr}}^{-1}} + a(u, m, p)\,.$$

The action of $\mathcal{H}_{\mathrm{m}}$ at $\hat{m}$ on m is

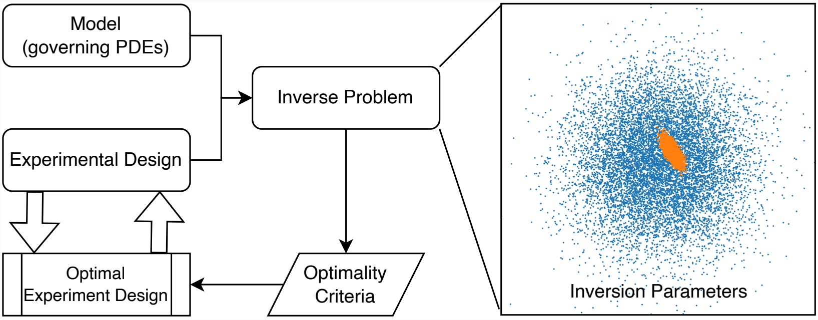
$$\mathcal{H}_{\mathrm{m}}(m)(\hat{m}, \tilde{m}) = \left\langle \tilde{m}, a_{mp}(u, m, p)\hat{p} \right\rangle,$$

where for all $\tilde{p} \in \mathscr{V}$ and $\tilde{u} \in \mathscr{V}$,

$$\begin{aligned}
a(u, m, \tilde{p}) &= 0\,, && \text{(State)}\\
\left\langle \tilde{u}, a_u(u, m, p) \right\rangle + \left\langle \tilde{u}, \mathcal{Q}^* \Gamma_{\mathrm{n}}^{-1}(\mathbf{y} - \mathcal{Q}u) \right\rangle &= 0\,, && \text{(Adjoint)}\\
\left\langle \tilde{p}, a_{pu}(u, m, p)\hat{u} \right\rangle + \left\langle \tilde{p}, a_{pm}(u, m, p)\hat{m} \right\rangle &= 0\,, && \text{(Incr. State)}\\
\left\langle \tilde{u}, a_{up}(u, m, p)\hat{p} \right\rangle + \left\langle \tilde{u}, \mathcal{Q}^* \Gamma_{\mathrm{n}}^{-1} \mathcal{Q}\hat{u} \right\rangle &= 0\,. && \text{(Incr. Adjoint)}
\end{aligned}$$

Robust Optimal Experimental Design

Optimal Experimental Design of Bayesian Inverse Problems



Definition (Optimal Experimental Design (OED))

Suppose

- $\boldsymbol{\xi} \in \{0,1\}^{\mathrm{N_d}}$ is a binary encoding of the observational configuration ($\xi_i = 1$ if sensor i is active, 0 otherwise),
- and $\mathrm{N_b} \ll \mathrm{N_d}$ is some budget constraint.

Then, the OED problem is defined as the optimization problem

$$\max_{\boldsymbol{\xi} \in \mathcal{S}(\mathrm{N_b})} \mathcal{U}(\boldsymbol{\xi}) \,,$$

where

$$\mathcal{S}(\mathrm{N_b}) = \left\{ \boldsymbol{\xi} \in \{0,1\}^{\mathrm{N_d}} : \sum_{i=1}^{\mathrm{N_d}} \xi_i = \mathrm{N_b} \right\} ,$$

and the utility (objective) $\mathcal{U}$ is chosen to quantify the quality of the design.

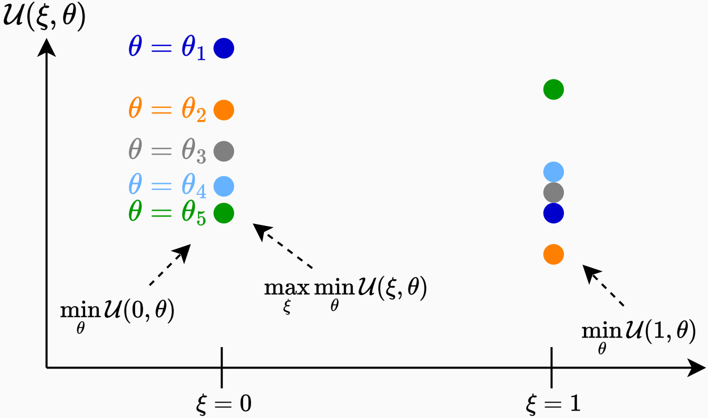
Definition (Robust Optimal Experimental Design (ROED))

Additionally, suppose there exists a set of parameters $\boldsymbol{\theta} \in \Theta$ that we are a-priori uncertain about. Then, the ROED problem is defined as the optimization problem

$$\max_{\boldsymbol{\xi} \in \mathcal{S}(\mathrm{N_b})} \min_{\boldsymbol{\theta} \in \Theta} \mathcal{U}(\boldsymbol{\xi}, \boldsymbol{\theta}) \,,$$

where

$$\mathcal{S}(\mathrm{N_b}) = \left\{ \boldsymbol{\xi} \in \{0, 1\}^{\mathrm{N_d}} : \sum_{i=1}^{\mathrm{N_d}} \xi_i = \mathrm{N_b} \right\},$$


$\mathcal{U}(\xi, \theta)$
$\theta = \theta_1$
$\theta = \theta_2$
$\theta = \theta_3$
$\theta = \theta_4$
$\theta = \theta_5$
$\min_\theta \mathcal{U}(0, \theta)$
$\max_\xi \min_\theta \mathcal{U}(\xi, \theta)$
$\min_\theta \mathcal{U}(1, \theta)$
$\xi = 0$
$\xi = 1$

Budget-Aware Probabilistic ROED

Idea: Model ξ as a random variable!

References

- **Probabilistic OED**: Ahmed Attia, Sven Leyffer, and Todd S. Munson. "Stochastic Learning Approach for Binary Optimization: Application to Bayesian Optimal Design of Experiments". In: *SIAM Journal on Scientific Computing* 44.2 (2022), B395–B427
- **Robust OED**: Ahmed Attia, Sven Leyffer, and Todd Munson. "Robust A-Optimal Experimental Design for Bayesian Inverse Problems". In: (2023). eprint: `arXiv:2305.03855`
- **Budget-Aware Probabilistic OED**: Ahmed Attia. "Probabilistic Approach to Black-Box Binary Optimization with Budget Constraints: Application to Sensor Placement". In: *arXiv preprint arXiv:2406.05830* (2024)

Budget-Aware Probabilistic ROED

Definition (Conditional Bernoulli Model)

Let $\boldsymbol{\xi} \in \{0,1\}^{\mathrm{N_d}}$ be a multivariate Bernoulli random variable parameterized by the policy $\mathbf{p} \in [0,1]^{\mathrm{N_d}}$. Let $|\boldsymbol{\xi}| \equiv |\boldsymbol{\xi}| = \sum_{i=1}^{\mathrm{N_d}} \xi_i$ be the total number of active (equal to 1) entries in $\boldsymbol{\xi}$, and define

$$S = \{1,\dots,\mathrm{N_d}\};\ O = \{i \in S : p_i = 0\};\ I = \{i \in S : p_i = 1\};\ T = S \setminus \{O \cup I\}\,.$$

Then, the probability mass function (PMF) of the conditional Bernoulli model is

$$\mathbb{P}(\boldsymbol{\xi}|\mathbf{p}, |\boldsymbol{\xi}| = z) = \begin{cases} \frac{\prod\limits_{i\in T} w^{\xi_i}}{R(z-|I|,T)}, & \text{if } \xi_j = p_j, \forall j \in \{I \cup O\} \text{ and } \sum\limits_{j\in T} \xi_j = z - |I| \\ 0, & \text{otherwise} \end{cases}$$

where

$$R(k,A) = \sum_{\substack{B\subseteq A \\ |B|=k}} \prod_{i\in B} w_i\,; \qquad w_i = \frac{p_i}{1-p_i}, \forall i \in \{1,\dots,\mathrm{N_d}\}\,.$$

Budget-Aware Probabilistic ROED

Assume that $\boldsymbol{\xi}$ is a random variable endowed with the **conditional Bernoulli distribution** $\mathbb{P}(\boldsymbol{\xi}|\mathbf{p}, |\boldsymbol{\xi}| = \mathrm{N_b})$.

$$\max_{\mathbf{p}\in[0,1]^{\mathrm{N_d}}} \mathfrak{U}(\mathbf{p}) := \mathbb{E}_{\boldsymbol{\xi}\sim\mathbb{P}(\boldsymbol{\xi}|\mathbf{p},|\boldsymbol{\xi}|=\mathrm{N_b})} \left[\min_{\boldsymbol{\theta}\in\Theta} \mathcal{U}(\boldsymbol{\xi}, \boldsymbol{\theta}) \right] .$$

Budget-Aware Probabilistic ROED

Assume that $\boldsymbol{\xi}$ is a random variable endowed with the **conditional Bernoulli distribution** $\mathbb{P}(\boldsymbol{\xi}|\mathbf{p}, |\boldsymbol{\xi}| = \mathrm{N_b})$.

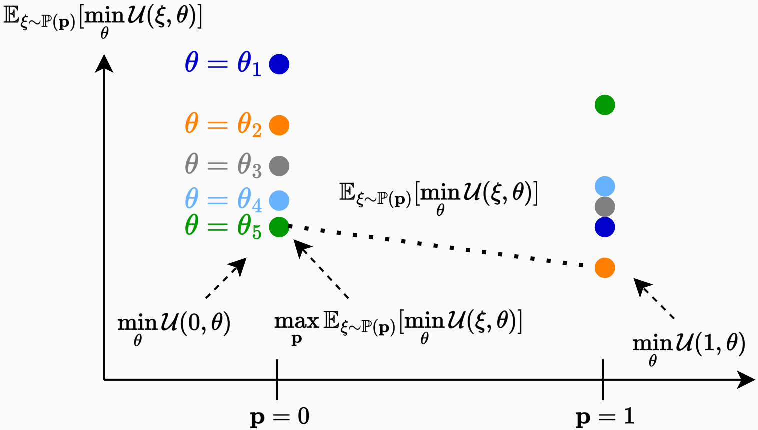
$$\max_{\mathbf{p}\in[0,1]^{\mathrm{N_d}}} \mathfrak{U}(\mathbf{p}) := \mathbb{E}_{\boldsymbol{\xi}\sim\mathbb{P}(\boldsymbol{\xi}|\mathbf{p},|\boldsymbol{\xi}|=\mathrm{N_b})} \left[\min_{\boldsymbol{\theta}\in\Theta} \mathcal{U}(\boldsymbol{\xi}, \boldsymbol{\theta}) \right] .$$

Furthermore,

$$\boldsymbol{\nabla}_{\mathbf{p}}\mathfrak{U}(\mathbf{p}) = \mathbb{E}_{\boldsymbol{\xi}\sim\mathbb{P}(\boldsymbol{\xi}|\mathbf{p},|\boldsymbol{\xi}|=\mathrm{N_b})} \left[\min_{\boldsymbol{\theta}\in\Theta} \mathcal{U}(\boldsymbol{\xi}, \boldsymbol{\theta}) \, \boldsymbol{\nabla}_{\mathbf{p}} \log \mathbb{P}(\boldsymbol{\xi}|\mathbf{p}, |\boldsymbol{\xi}| = \mathrm{N_b}) \right] .$$

Policy Optimization

Can use gradient based optimization for **p** without requiring gradients of $\mathcal{U}$!


$\mathbb{E}_{\xi\sim\mathbb{P}(\mathbf{p})}[\min_\theta \mathcal{U}(\xi,\theta)]$
$\theta = \theta_1$
$\theta = \theta_2$
$\theta = \theta_3$
$\theta = \theta_4$
$\theta = \theta_5$
$\mathbb{E}_{\xi\sim\mathbb{P}(\mathbf{p})}[\min_\theta \mathcal{U}(\xi,\theta)]$
$\min_\theta \mathcal{U}(0,\theta)$
$\max_{\mathbf{p}} \mathbb{E}_{\xi\sim\mathbb{P}(\mathbf{p})}[\min_\theta \mathcal{U}(\xi,\theta)]$
$\min_\theta \mathcal{U}(1,\theta)$
$\mathbf{p} = 0$
$\mathbf{p} = 1$

Polyak's Algorithm

$$\max_{\mathbf{p}\in[0,1]^{\mathrm{N_d}}} \mathbb{E}_{\boldsymbol{\xi}\sim\mathbb{P}(\boldsymbol{\xi}|\mathbf{p},|\boldsymbol{\xi}|=\mathrm{N_b})} \left[\min_{\boldsymbol{\theta}\in\overline{\Theta}^{(k)}} \mathcal{U}(\boldsymbol{\xi},\boldsymbol{\theta}) \right]$$

$$\overline{\Theta}^{(k+1)} = \overline{\Theta}^{(k)} \cup \arg\min_{\boldsymbol{\theta}\in\Theta} \mathcal{U}(\boldsymbol{\xi}^{\mathrm{opt}},\boldsymbol{\theta})$$

$\overline{\Theta}^{(k)}$ is a sample of k uncertain parameters.

E.S. Levitin and B.T. Polyak. "Constrained minimization methods". In: *USSR Computational Mathematics and Mathematical Physics* 6.5 (1966)

The Expected Information Gain

References

- **Expected Information Gain for Large-Scale OED**: Keyi Wu, Peng Chen, and Omar Ghattas. "A Fast and Scalable Computational Framework for Large-Scale High-Dimensional Bayesian Optimal Experimental Design". In: *SIAM/ASA Journal on Uncertainty Quantification* 11.1 (2023), pp. 235–261
- **Adjoint-Based Eigenvalue Sensitivities**: Abhijit Chowdhary et al. "Sensitivity analysis of the information gain in infinite-dimensional Bayesian linear inverse problems". In: *International Journal for Uncertainty Quantification* 14.6 (2024)

EIG as a Utility

Definition (Information Gain)

$$D_{\mathrm{KL}}(\mu_{\mathrm{post}}^{\mathbf{y}} \| \mu_{\mathrm{pr}}) = \int \log\left(\frac{\mathrm{d}\mu_{\mathrm{post}}^{\mathbf{y}}}{\mathrm{d}\mu_{\mathrm{pr}}}\right) \mathrm{d}\mu_{\mathrm{post}}^{\mathbf{y}}$$

Definition (Expected Information Gain)

$$\begin{aligned} \overline{D_{\mathrm{KL}}} &= \mathbb{E}_{\mathbf{y}}\Big[D_{\mathrm{KL}}(\mu_{\mathrm{post}}^{\mathbf{y}} \,\|\, \mu_{\mathrm{pr}})\Big] \\ &= \int_{\mathscr{M}} \int_{\mathscr{Y}} D_{\mathrm{KL}}(\mu_{\mathrm{post}}^{\mathbf{y}} \,\|\, \mu_{\mathrm{pr}}) \pi_{\mathrm{like}}(\mathbf{y}|m) \, \mathrm{d}\mathbf{y} \, \mathrm{d}\mu_{\mathrm{pr}}(m) \end{aligned}$$

Definition (EIG using the Laplace Approximation)

Under the Laplace approximation, it can be shown that

$$D_{\mathrm{KL}}(\hat{\mu}_{\mathrm{post}}^{\mathbf{y}} \,\|\, \mu_{\mathrm{pr}}) =$$
$$\frac{1}{2}\Big[\log\det\Big(\mathcal{I} + \widetilde{\mathcal{H}}_{\mathrm{m}}\Big) - \mathrm{tr}\Big(\widetilde{\mathcal{H}}_{\mathrm{m}}\big[\mathcal{I} + \widetilde{\mathcal{H}}_{\mathrm{m}}\big]^{-1}\Big) + \big\|m_{\mathrm{post}} - m_{\mathrm{pr}}\big\|^2_{\mathcal{C}_{\mathrm{pr}}^{-1}}\Big]$$

with $\widetilde{\mathcal{H}}_{\mathrm{m}} = \mathcal{C}_{\mathrm{pr}}^{1/2}\mathcal{H}_{\mathrm{m}}\mathcal{C}_{\mathrm{pr}}^{1/2}$. Hence,

$$\overline{D_{\mathrm{KL}}} \approx \frac{1}{\mathrm{N_{SAA}}} \sum_{i=1}^{\mathrm{N_{SAA}}} D_{\mathrm{KL}}(\hat{\mu}_{\mathrm{post}}^{\mathbf{y}_i} \,\|\, \mu_{\mathrm{pr}}) \,,$$

where for every $i \in \{1, \dots, \mathrm{N_{SAA}}\}$, the data $\mathbf{y}_i$ is

$$\mathbf{y}_i = \boldsymbol{\mathcal{F}}(m_i) + \boldsymbol{\eta}_i \,,$$

where $m_i \sim \mu_{\mathrm{pr}}$ and $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \Gamma_{\mathrm{n}})$.

$\widetilde{\mathcal{H}}_{\mathrm{m}}$ is often low-rank, hence

$$\widetilde{\mathcal{H}}_{\mathrm{m}}\phi = \sum_{n=1}^{\infty} \lambda_n \langle \phi, \omega_n \rangle \omega_n \approx \sum_{n=1}^{r} \lambda_n \langle \phi, \omega_n \rangle \omega_n, \quad \phi \in \mathscr{M},$$

so, we can define

$$\overline{D_{\mathrm{KL}}} \approx \frac{1}{\mathrm{N_{SAA}}} \sum_{i=1}^{\mathrm{N_{SAA}}} D_{\mathrm{KL}}^{(r)}(\hat{\mu}_{\mathrm{post}}^{\mathbf{y}_i} \,\|\, \mu_{\mathrm{pr}}),$$

with

$$D_{\mathrm{KL}}^{(r)}(\hat{\mu}_{\mathrm{post}}^{\mathbf{y}_i} || \mu_{\mathrm{pr}})(\boldsymbol{\xi}, \boldsymbol{\theta}) = \sum_{n=1}^{r} \left[\log\Big(1 + \lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})\Big) - \frac{\lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})}{1 + \lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})} \right] + \left\| m_{\mathrm{post}}^i(\boldsymbol{\xi}, \boldsymbol{\theta}) - m_{\mathrm{pr}} \right\|_{\mathcal{C}_{\mathrm{pr}}^{-1}}^2.$$

Low-Rank Approximation of the EIG with fixed MAP estimate

$\widetilde{\mathcal{H}}_{\mathrm{m}}$ is often low-rank, hence

$$\widetilde{\mathcal{H}}_{\mathrm{m}}\phi = \sum_{n=1}^{\infty} \lambda_n \langle \phi, \omega_n \rangle \omega_n \approx \sum_{n=1}^{r} \lambda_n \langle \phi, \omega_n \rangle \omega_n, \quad \phi \in \mathscr{M} ,$$

so, we can define

$$\overline{D_{\mathrm{KL}}} \approx \frac{1}{\mathrm{N_{SAA}}} \sum_{i=1}^{\mathrm{N_{SAA}}} D_{\mathrm{KL}}^{(r)}(\hat{\mu}_{\mathrm{post}}^{\mathbf{y}_i} \,\|\, \mu_{\mathrm{pr}}) ,$$

with

$$D_{\mathrm{KL}}^{(r)}(\hat{\mu}_{\mathrm{post}}^{\mathbf{y}_i} || \mu_{\mathrm{pr}})(\boldsymbol{\xi}, \boldsymbol{\theta}) =$$

$$\sum_{n=1}^{r} \left[\log\Big(1 + \lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})\Big) - \frac{\lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})}{1 + \lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})} \right] + ||m_{\mathrm{post}}^i(\cancelto{\boldsymbol{\xi}^{\mathrm{all}}}{\boldsymbol{\xi}}, \cancelto{\overline{\boldsymbol{\theta}}}{\boldsymbol{\theta}}) - m_{\mathrm{pr}}||_{\mathcal{C}_{\mathrm{pr}}^{-1}}^2 .$$

Utility for ROED

For the ROED problem, we will use the utility

$$\mathcal{U}(\boldsymbol{\xi}, \boldsymbol{\theta}) = \frac{1}{\mathrm{N_{SAA}}} \sum_{i=1}^{\mathrm{N_{SAA}}} \hat{\mathcal{U}}(\mathbf{y}_i, \boldsymbol{\xi}, \boldsymbol{\theta})$$

with

$$\hat{\mathcal{U}}(\mathbf{y}_i, \boldsymbol{\xi}, \boldsymbol{\theta}) = \sum_{n=1}^{r} \left[\log\Big(1 + \lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})\Big) - \frac{\lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})}{1 + \lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})} \right] + C_i \,.$$

Evaluating $\nabla_\theta \mathcal{U}$

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} \mathcal{U}(\boldsymbol{\xi}, \boldsymbol{\theta}) = \frac{1}{\mathrm{N}_{\mathrm{SAA}}} \sum_{i=1}^{\mathrm{N}_{\mathrm{SAA}}} \boldsymbol{\nabla}_{\boldsymbol{\theta}} \hat{\mathcal{U}}(\mathbf{y}_i, \boldsymbol{\xi}, \boldsymbol{\theta}) \,,$$

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} \hat{\mathcal{U}}(\mathbf{y}_i, \boldsymbol{\xi}, \boldsymbol{\theta}) = \boldsymbol{\nabla}_{\boldsymbol{\theta}} \left(\frac{1}{2} \sum_{n=1}^{r} \left[\log\Big(1 + \lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})\Big) - \frac{\lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})}{1 + \lambda_n^i(\boldsymbol{\xi}, \boldsymbol{\theta})} \right] \right) \,.$$

Q: How do we compute $\nabla_\theta \lambda_n^i(\xi, \theta)$?

$$\frac{1}{2}\sum_{n=1}^{r}\left[\log(1+\lambda_n)-\frac{\lambda_n}{1+\lambda_n}\right],$$

The above expression has the following constraints

$$\begin{aligned}\langle\phi,\mathcal{H}_{\mathrm{m}}\psi_n\rangle &= \lambda_n\langle\phi,\psi_n\rangle_{\mathcal{C}_{\mathrm{pr}}^{-1}}, && \forall\phi\in\mathscr{V},\forall n=1,\ldots,r,\\ \langle\psi_n,\psi_n\rangle_{\mathcal{C}_{\mathrm{pr}}^{-1}} &= 1, && \forall n=1,\ldots,r,\end{aligned}$$

and

$$\mathcal{H}_{\mathrm{m}}(m)(\psi_n,\phi)=\left\langle\phi,a_{mp}(u,m,p)\hat{p}\right\rangle,$$

such that for all $\tilde{p}\in\mathscr{V}$ and $\tilde{u}\in\mathscr{V}$,

$$\begin{aligned}\left\langle\tilde{p},a_p(u,m,p)\right\rangle &= 0, && \text{(State)}\\ \left\langle\tilde{u},a_u(u,m,p)\right\rangle+\left\langle\tilde{u},\mathcal{Q}^*\widehat{\Gamma}_{\mathrm{n}}^{\dagger}(\boldsymbol{\xi},\boldsymbol{\theta})(\mathbf{y}-\mathcal{Q}u)\right\rangle &= 0, && \text{(Adjoint)}\\ \left\langle\tilde{p},a_{pu}(u,m,p)\hat{u}_n\right\rangle+\left\langle\tilde{p},a_{pm}(u,m,p)\psi_n\right\rangle &= 0, && \text{(Incr. State)}\\ \left\langle\tilde{u},a_{up}(u,m,p)\hat{p}_n\right\rangle+\left\langle\tilde{u},\mathcal{Q}^*\widehat{\Gamma}_{\mathrm{n}}^{\dagger}(\boldsymbol{\xi},\boldsymbol{\theta})\mathcal{Q}\hat{u}_n\right\rangle &= 0, && \text{(Incr. Adjoint)}\end{aligned}$$

Lagrangian for $\nabla_\theta \mathcal{U}$

$$
\begin{aligned}
&\mathcal{L}^{\mathrm{IG}}\big(u, m, p, \{\psi_n\}_{n=1}^r, \{\hat{u}_n\}_{n=1}^r, \{\hat{p}_n\}_{n=1}^r, \\
&\quad u^*, p^*, \{\lambda_n^*\}_{n=1}^r, \{\hat{u}_n^*\}_{n=1}^r, \{\hat{p}_n^*\}_{n=1}^r; \boldsymbol{\theta}\big) \\
&\quad = \frac{1}{2}\sum_{n=1}^r \left[\log\Big(1 + \big\langle \psi_n, a_{mp}\hat{p}_n \big\rangle\Big) - \frac{\big\langle \psi_n, a_{mp}\hat{p}_n \big\rangle}{1 + \big\langle \psi_n, a_{mp}\hat{p}_n \big\rangle}\right] \\
&\qquad + \big\langle p^*, a_p \big\rangle + \big\langle u^*, a_u \big\rangle + \Big\langle u^*, \mathcal{Q}^*\widehat{\Gamma}_{\mathrm{n}}^{\dagger}(\mathbf{y} - \mathcal{Q}u) \Big\rangle \\
&\qquad + \sum_{n=1}^r \left[\big\langle \hat{p}_n^*, a_{pu}\hat{u}_n + a_{pm}\psi_n \big\rangle + \big\langle \hat{u}_n^*, a_{up}\hat{p}_n \big\rangle + \Big\langle \hat{u}_n^*, \mathcal{Q}^*\widehat{\Gamma}_{\mathrm{n}}^{\dagger}\mathcal{Q}\hat{u}_n \Big\rangle\right] \\
&\qquad + \sum_{n=1}^r \lambda_n^* \left[\langle \psi_n, \psi_n \rangle_{\mathcal{C}_{\mathrm{pr}}^{-1}} - 1\right].
\end{aligned}
$$

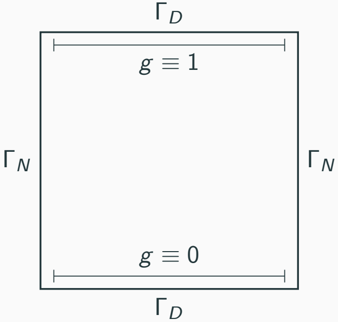
Computational Complexity Summary for $\mathcal{U}$

Procedure	Cost (in PDE solves)
Evaluation	$\mathcal{O}(4\mathrm{N_b} \cdot \mathrm{N_{SAA}})$
Gradient	$\mathcal{O}((3 + 5\mathrm{N_b}) \cdot \mathrm{N_{SAA}})$
Simultaneous Value/Gradient	$\mathcal{O}((3 + 5\mathrm{N_b}) \cdot \mathrm{N_{SAA}})$

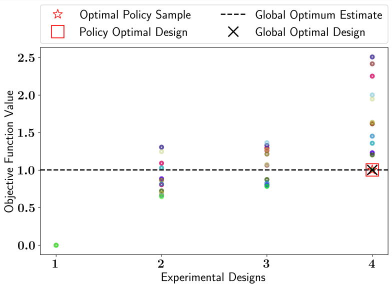
Numerical Results

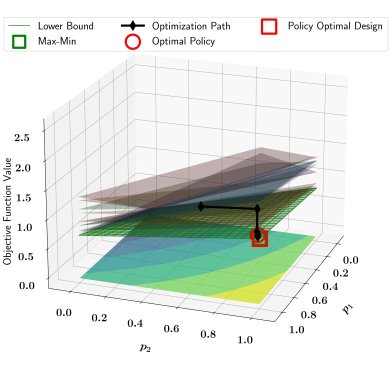
Example: Permeability Identification in Poisson's Equation

$$
\begin{aligned}
-\nabla \cdot (\exp(m)\nabla u) &= 0 && \text{in } \Omega := (0,1)^2\,, \\
\exp(m)\nabla u \cdot \mathbf{n} &= 0 && \text{on } \Gamma_N := \{0,1\} \times (0,1)\,, \\
u &= g && \text{on } \Gamma_D := (0,1) \times \{0,1\}\,.
\end{aligned}
$$



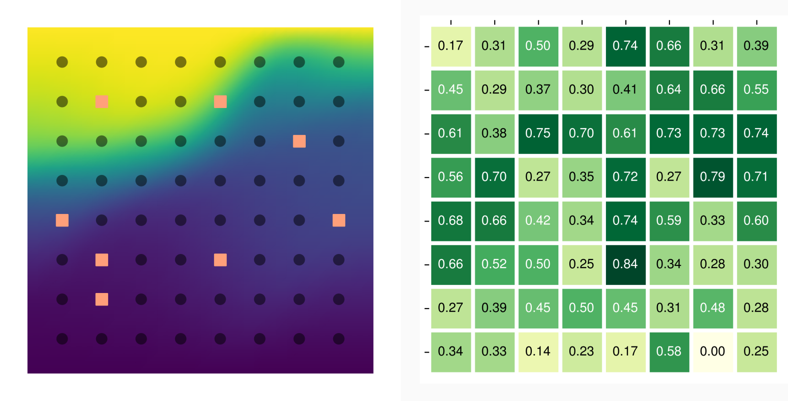
Two Sensor Experiment





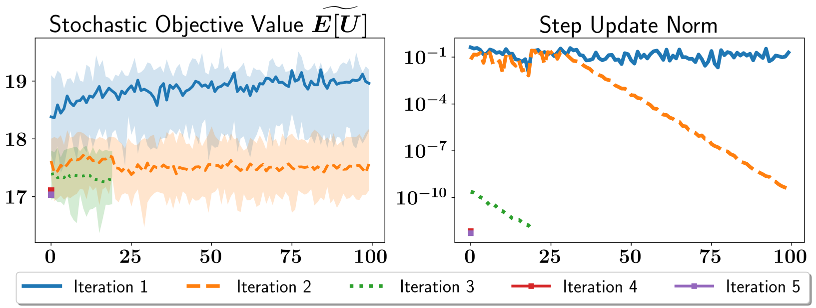
$$\Gamma_{\mathrm{n}}(\boldsymbol{\theta} := [\sigma_1, \sigma_2, \rho]) = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

64 Sensor, Budget 8 Experiment



Results of the 64 sensor, budget 8, experiment. Left: Optimal design discovered by sampling from policy and selecting the design with the highest utility. Right: Optimal policy $\mathbf{p}_{\boldsymbol{\theta}}^{\mathrm{opt}}$ discovered by the stochastic optimization algorithm, visualized across the sensor grid.

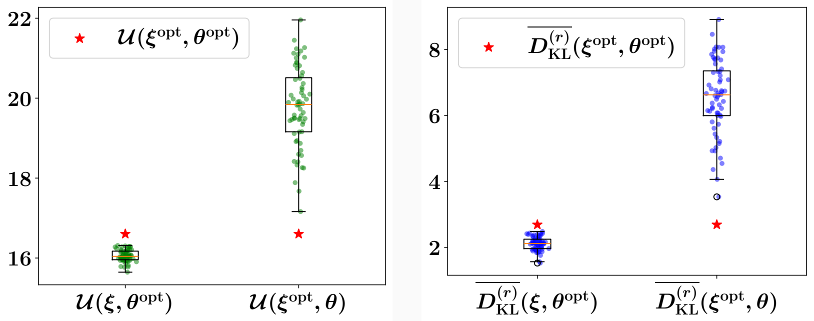
Progression of the ROED algorithm



Optimization trajectory of the 64 sensor, budget 8, experiment.

- Left: Progress of an estimate to the expectation of the utility $\mathcal{U}$ over designs sampled from the policy. The line represents the mean of the expectation whereas the top and bottom of the shaded region represent the maximum and minimum respectively.
- Right: Norm of the update in the policy $\mathbf{p}$.

Sanity check for $\mathcal{U}$



A visualization of the quality of $(\xi^{\text{opt}}, \theta^{\text{opt}})$ for the 64 sensor, budget 8, ROED experiment.

Conclusion

Conclusion

Future Directions

- Moving beyond the Laplace approximation
- Further Surrogate Modeling

To be presented in a minisymposia at **SIAM CSE 2025**!

Abhijit Chowdhary, Ahmed Attia, and Alen Alexanderian. “Robust optimal design of large-scale Bayesian nonlinear inverse problems”. In: *arXiv preprint arXiv:2409.09137* (2024)