GOal-Oriented Optimal Experimental Design

GO-OED for PDE-based Bayesian Linear Inverse Problems

Ahmed Attia

Statistical and Applied Mathematical Science Institute (SAMSI)
19 TW Alexander Dr, Durham, NC 27703
attia@ {samsi.info || vt.edu}

Alen Alexanderian, and Arvind Krishna Saibaba

Department of Mathematics
North Carolina State University

SAMSI
January 26, 2017

Outline

Motivation

Inverse problems

Data collection, and sensor placement

Optimal Experimental Design (OED)

Optimal experimental design (OED)

Sensor placement as OED problem

Goal-Oriented OED

GO-OED preliminary results

Conclusions and Future Work

Motivation

- Inverse problems, and data assimilation, enable making efficient and accurate predictions about the state of large-scale systems such as the atmosphere.

Motivation

- Inverse problems, and data assimilation, enable making efficient and accurate predictions about the state of large-scale systems such as the atmosphere.
- The solution of the inverse problem is sensitive to the quality of the collected measurements.

Motivation

- Inverse problems, and data assimilation, enable making efficient and accurate predictions about the state of large-scale systems such as the atmosphere.
- The solution of the inverse problem is sensitive to the quality of the collected measurements.
- It is intuitive that we should collect as much data as possible, however measurement sensors are not always cheap!

Motivation

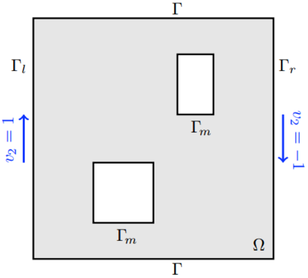
- Inverse problems, and data assimilation, enable making efficient and accurate predictions about the state of large-scale systems such as the atmosphere.
- The solution of the inverse problem is sensitive to the quality of the collected measurements.
- It is intuitive that we should collect as much data as possible, however measurement sensors are not always cheap!
- Generally, observational grids are designed such that the collected measurements lead to predictions with minimum uncertainty (OED).

Motivation

- Inverse problems, and data assimilation, enable making efficient and accurate predictions about the state of large-scale systems such as the atmosphere.
- The solution of the inverse problem is sensitive to the quality of the collected measurements.
- It is intuitive that we should collect as much data as possible, however measurement sensors are not always cheap!
- Generally, observational grids are designed such that the collected measurements lead to predictions with minimum uncertainty (OED).
- We might be interested in designing an observational grid that minimizes a QoI dependent on the model state, rather than the state itself (GO-OED).

Example: Advection-Diffusive transport

- Consider the concentration of a contaminant u in the domain $\Omega \in \mathbb{R}^2$.
- Assume, we are interested in inferring the initial distribution of the contaminant, from measurements b taken after the contaminant has been subjected to diffusive transport. For example, consider measuring u on some parts Γ_m of (or all) the domain boundary.



Example: Initial Condition in Advection-Diffusive transport

- Inverse Problem: given measurements b over time interval $[T_1, T]$, find the initial condition u_0 such that the evolution of the contaminant distribution is governed by the PDE, and is consistent with the collected measurements.

Example: Initial Condition in Advection-Diffusive transport

- Inverse Problem: given measurements b over time interval $[T_1, T]$, find the initial condition u_0 such that the evolution of the contaminant distribution is governed by the PDE, and is consistent with the collected measurements.

- A Continuous Formulation: solve the optimization problem:

$$\min_{u_0} J(u_0) := \frac{1}{2}\int_{T_1}^{T}\int_{\Gamma_m} (\mathcal{B}(u) - b)^2 dx dt + \frac{\gamma}{2}\int_{\Omega} u_0^2 dx$$

  where u is constrained by the advection diffusion PDEs.

- A Discrete Formulation: solve the optimization problem:

$$\min_{\mathbf{u}_0} J(\mathbf{u}_0) := \frac{1}{2}\sum_{i=1}^{m} (\mathcal{B}(\mathbf{u}_i) - \mathbf{b}_i)^T \Gamma_{\rm noise}^{-1} (\mathcal{B}(\mathbf{u}_i) - \mathbf{b}_i) + \frac{1}{2}\mathbf{u}_0^T \mathbf{R}^{-1}\mathbf{u}_0$$

  where $\mathbf{u}$ is a spatial discretization of u, and is constrained by the discretized advection diffusion PDEs, and $\{\mathbf{b}_i\}_{i=1,2,\ldots,m}$ are discrete-time observations.

Example: Initial Condition in Advection-Diffusive transport
A Discrete Bayesian Formulation:

- Let $\mathbf{u}_0$ be a random vector with prior distribution $\mathcal{N}(0, \mathbf{C}_{\rm prior})$.
- Assume a likelihood function:

$$\mathcal{P}(\mathbf{b}_1, \ldots, \mathbf{b}_m | \mathbf{u}_0) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^{m} (\mathbf{b}_i - \mathcal{B}(\mathbf{u}_i))^T \Gamma_{\rm noise}^{-1} (\mathbf{b}_i - \mathcal{B}(\mathbf{u}_i)) \right)$$

- The posterior (Bayes' theorem):

$$\mathcal{P}(\mathbf{u}_0 | \mathbf{b}_1, \ldots, \mathbf{b}_m) \propto \exp\left(-\tfrac{1}{2} \textstyle\sum_{i=1}^{m} (\mathcal{B}(\mathbf{u}_i) - \mathbf{b}_i)^T \Gamma_{\rm noise}^{-1} (\mathbf{b}_i - \mathcal{B}(\mathbf{u}_i)) \right) + \tfrac{1}{2} \mathbf{u}_0^T \mathbf{C}_{\rm prior}^{-1} \mathbf{u}_0$$

  where $\Gamma_{\rm noise}$ is an observation error covariance matrix, and $\mathbf{C}_{\rm prior}$ is a parameter prior covariance matrix.

- **Solution strategy**: Given observation(s), and a prior, describe the posterior PDF, and use it to make predictions.

Data collection, and sensor placement

- Design of experiments (DOE) involves applying engineering principles and techniques for data collection so as to ensure the conclusions made based on the experiment are valid and reliable.
- Suppose we have N_{obs} candidate observation locations. Where do we put sensors, to actually collect measurements?
- **The main idea**: Find the measurement locations such that the uncertainty in the QoI is minimized.

Problem Formulation:

- Consider an additive noise model:

$$\mathbf{b} = \mathcal{F}(m) + \eta\,.$$

where $\mathcal{F}$ is the "Forward Operator" e.g. $(\mathcal{B} \circ \mathcal{S})$.

- In a Bayesian formulation, we obtain the posterior law for m through,

$$\frac{d\mu_{\text{post}}^{\mathbf{b}}}{d\mu_{\text{pr}}} \propto \pi_{\text{like}}(\mathbf{b}|m)\,,$$

where the data likelihood is given by, $\pi_{\text{like}}(\mathbf{b}|m) = \rho_{\text{noise}}(d - \mathcal{F}(m))$.

- An experimental design, ξ, will specify the way data is collected (e.g. sensor placement). In OED problem, we seek ξ_{opt} that minimizes the uncertainty in the inferred parameter m, i.e. we consider:

$$\frac{d\mu_{\text{post}}^{\mathbf{b}|\xi}}{d\mu_{\text{pr}}} \propto \pi_{\text{like}}(\mathbf{b}|m;\xi)\,.$$

Optimal Experimental Design (OED)

- The way one chooses to quantify "posterior uncertainty" leads to the choice of the design criterion.
- Alphabetical criteria:
  1. minimize the trace of the posterior covariance (A-Optimality)
  2. minimize the determinant of the posterior covariance (D-Optimality)
  3. minimize the largest eigenvalue of the posterior covariance (E-Optimality)
  4. etc.

Optimal Experimental Design (OED)

- In the context of optimal sensor placement, the design ξ is a binary vector $\boldsymbol{w}$, corresponding to candidate sensor locations.

Optimal Experimental Design (OED)

- In the context of optimal sensor placement, the design ξ is a binary vector $\boldsymbol{w}$, corresponding to candidate sensor locations.

- We solve a relaxed version, where the design ξ is a vector $\boldsymbol{w}$ of weights $\in [0, 1]$ corresponding to a set of candidate locations for sensors.

OED for Sensor Placement

- $\boldsymbol{w}$ enters the Bayesian inverse problem through the data likelihood, amounting to a weighted data likelihood:

$$\pi_{\text{like}}(\mathbf{b}|m;\boldsymbol{w}) \propto \exp\left\{-\frac{1}{2}\left(\mathcal{F}(m)-\mathbf{b}\right)^T \mathbf{W}^{1/2}\boldsymbol{\Gamma}_{\text{noise}}^{-1}\mathbf{W}^{1/2}\left(\mathcal{F}(m)-\mathbf{b}\right)\right\},$$

  where $\mathbf{W} = diag(w_1, \ldots, w_{N_s})$.

- Posterior (Gaussian Linear case) $\mathcal{N}(0, \mathbf{C}_{\rm post})$ with

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

  where $\mathbf{W}_\sigma = \mathbf{W}^{1/2}\boldsymbol{\Gamma}_{\text{noise}}^{-1}\mathbf{W}^{1/2}$.

- OED: find $\boldsymbol{w}$ that minimizes ($\mathbf{C}_{\rm post}$ trace, determinant, etc.)

- what if we are interested in a prediction quantity $p = \mathcal{P}(m)$ rather than the parameter itself? e.g. the average contaminant concentration within a specific distance from the buildings' walls. (**GO-OED**)

GO-OED: problem formulation

- We consider a linear Θ_{pred} of the form:

$$\Theta_{\text{pred}}(m) = \mathcal{P} m \,,$$

where $\mathcal{P}$ is a linear prediction operator, e.g. a forward solve followed by a restriction operator.

- In the standard Linear-Gaussian settings:

$$\nu_{\text{prior}} = \mathcal{N}\left(\widetilde{m}_{\rm pr}, \widetilde{\mathbf{C}}_{\rm pr}\right) \,, \qquad \nu_{\text{post}}^{\boldsymbol{y}|\boldsymbol{w}} = \mathcal{N}\left(\widetilde{m}_{\rm post}, \widetilde{\mathbf{C}}_{\rm post}\right) \,,$$

with

$$\widetilde{m}_{\rm pr} = \mathcal{P} m_{\rm pr} \,, \qquad \widetilde{m}_{\rm post} = \mathcal{P} m_{\rm MAP} \,, \text{and,}$$
$$\widetilde{\mathbf{C}}_{\rm pr} = \mathcal{P}\mathcal{C}_{\text{pr}}\mathcal{P}^* \,, \qquad \widetilde{\mathbf{C}}_{\rm post} = \mathcal{P}\mathcal{C}_{\text{post}}\mathcal{P}^* \,,$$

where, the posterior MAP, and the posterior covariance, of the inverse problem, are given by:

$$m_{\rm MAP} = \mathcal{C}_{\text{post}} \left(\mathcal{F}^* \mathbf{W}_\sigma \boldsymbol{y} + \mathcal{C}_{\text{pr}}^{-1} m_{\rm pr}\right) ;$$
$$\mathcal{C}_{\text{post}}(\boldsymbol{w}) = \left(\mathcal{F}^* \mathbf{W}_\sigma \mathcal{F} + \mathcal{C}_{\text{pr}}^{-1}\right)^{-1} = \left(\mathbf{H}_{\text{misfit}}(w) + \mathcal{C}_{\text{pr}}^{-1}\right)^{-1} .$$

GO-OED: A-Optimality

- The optimal design ($w^{\mathrm{A}}_{\mathrm{optimal}}$) is given by:

$$w^{\mathrm{A}}_{\mathrm{optimal}} = \operatorname*{arg\,min}_{\boldsymbol{w} \in \mathbb{R}^{N_s}} \operatorname{tr}\left(\widetilde{\mathbf{C}}_{\mathrm{post}} \right) := \operatorname{tr}\left(\mathcal{P} \, [\mathbf{H}(w)]^{-1} \, \mathcal{P}^* \right),$$

where $\mathbf{H}(w)$ is the Hessian of the posterior negative-log.

GO-OED: A-Optimality

- The optimal design ($w^{\mathrm{A}}_{\mathrm{optimal}}$) is given by:

$$w^{\mathrm{A}}_{\mathrm{optimal}} = \underset{\boldsymbol{w} \in \mathbb{R}^{N_s}}{\arg\min}\, \mathrm{tr}\Big(\widetilde{\mathbf{C}}_{\mathrm{post}}\Big) := \mathrm{tr}\Big(\mathcal{P}\ [\mathbf{H}(w)]^{-1}\ \mathcal{P}^*\Big)\,,$$

  where $\mathbf{H}(w)$ is the Hessian of the posterior negative-log.

- With temporally-independent observation errors, the Hessian reads:

$$\mathbf{H}(w) = \sum_{k=1}^{N_\tau} \mathcal{F}^*_{0,k} \mathbf{W}_\sigma \mathcal{F}_{0,k} + \mathcal{C}_{\mathrm{pr}}^{-1} = \mathbf{H}_{\mathrm{misfit}}(w) + \mathcal{C}_{\mathrm{pr}}^{-1}\,,$$

  where $\mathcal{F}_{0,k}$ is the parameter-to-observable map from (discrete) time 0 to time t_k, and $\mathcal{F}^*_{0,k}$ is it's adjoint.

GO-OED: A-Optimality

- The gradient of A-optimality objective, with respect to the design, is given by:

$$\nabla_w \mathrm{tr}\left(\widetilde{\mathbf{C}}_{\rm post}\right) = -\sum_{k=1}^{N_\tau} \sum_{j} \left[\left(\Gamma_N^{-1/2} \mathcal{F}_{0,k} \, [\mathbf{H}(w)]^{-1} \, \mathcal{P}^* \, \boldsymbol{e}_j\right) \odot \left(\Gamma_N^{-1/2} \mathcal{F}_{0,k} \, [\mathbf{H}(w)]^{-1} \, \mathcal{P}^* \, \boldsymbol{e}_j\right) \right],$$

where $\boldsymbol{e}_i$ is the i^{th} coordinate vector in $\mathbb{R}^{N_{\rm pred}}$.

GO-OED: D-Optimality

- The optimal design ($w^{\mathrm{D}}_{\mathrm{optimal}}$) is given by:

$$w^{\mathrm{D}}_{\mathrm{optimal}} = \underset{\boldsymbol{w} \in \mathbb{R}^{N_s}}{\arg\min} \log\det\left(\widetilde{\mathbf{C}}_{\mathrm{post}}\right) := \log\det\left(\mathcal{P}\ [\mathbf{H}(w)]^{-1}\ \mathcal{P}^*\right).$$

GO-OED: D-Optimality

- The optimal design ($w^{\mathrm{D}}_{\mathrm{optimal}}$) is given by:

$$w^{\mathrm{D}}_{\mathrm{optimal}} = \operatorname*{arg\,min}_{\boldsymbol{w} \in \mathbb{R}^{N_s}} \log\det\left(\widetilde{\mathbf{C}}_{\mathrm{post}}\right) := \log\det\left(\mathcal{P}\ [\mathbf{H}(w)]^{-1}\ \mathcal{P}^*\right).$$

- The gradient of D-optimality objective reads:

$$\nabla_w \log\det\left(\widetilde{\mathbf{C}}_{\mathrm{post}}\right) = -\sum_{k=1}^{N_\tau}\sum_j \left[\begin{array}{c} \left(\Gamma_N^{-1/2}\mathcal{F}_{0,k}\,[\mathbf{H}(w)]^{-1}\,\mathcal{P}^*\Sigma_{\mathrm{pred}}^{-1/2}(\boldsymbol{w})\boldsymbol{e}_j\right) \\ \odot\left(\Gamma_N^{-1/2}\mathcal{F}_{0,k}\,[\mathbf{H}(w)]^{-1}\,\mathcal{P}^*\Sigma_{\mathrm{pred}}^{-1/2}(\boldsymbol{w})\boldsymbol{e}_j\right)\end{array}\right],$$

where $\boldsymbol{e}_i$ is the i^{th} coordinate vector in $\mathbb{R}^{N_{\mathrm{pred}}}$, and $\left(\widetilde{\mathbf{C}}_{\mathrm{post}}\right)^{-1} = \left(\Sigma_{\mathrm{pred}}(\boldsymbol{w})\right)^{-1/2}\left(\Sigma_{\mathrm{pred}}^T(\boldsymbol{w})\right)^{-T/2}$

GO-OED: D-Optimality

- Alternatively, the gradient of D-optimality objective can be written on the form:

$$\frac{\partial}{\partial w_i}\left(\log\det\left(\widetilde{\mathbf{C}}_{\rm post}\right)\right) = -\sum_{k=1}^{N_\tau}\left(\boldsymbol{I}_{k,i}\left(\mathcal{P}\left[\mathbf{H}(w)\right]^{-1}\mathcal{P}^*\right)^{-1}\boldsymbol{I}_{k,i}^T\right),$$

where:

$$\boldsymbol{I}_{k,i}^T = \left(\mathcal{P}\ \left[\mathbf{H}(w)\right]^{-1}\mathcal{F}_{0,k}^*\Gamma_N^{-T/2}\right)\boldsymbol{e}_i\,; \quad \forall i = 1,2,\ldots,N_{\rm obs}\,,$$

with $\boldsymbol{e}_i$ being the i^{th} coordinate vector in $\mathbb{R}^{N_{\rm obs}}$.

GO-OED: Bayesian D-Optimality (KL-Divergence)

- Consider the expected Kullback-Leibler divergence from the prior and posterior predictive distributions of Θ_{pred}:

$$\mathrm{D}_{\mathrm{KL}}(\nu_{\mathrm{post}}^{\boldsymbol{y}|\boldsymbol{w}} \| \nu_{\mathrm{prior}}) \,,$$

GO-OED: Bayesian D-Optimality (KL-Divergence)

- Consider the expected Kullback-Leibler divergence from the prior and posterior predictive distributions of Θ_{pred}:

$$
\mathsf{D}_{\mathsf{KL}}(\nu_{\text{post}}^{\boldsymbol{y}|\boldsymbol{w}} \| \nu_{\text{prior}}) \,,
$$

- The expected information gain can be defined as:

$$
\begin{aligned}
\Psi(\boldsymbol{w}) &:= \int \int \mathsf{D}_{\mathsf{KL}}(\nu_{\text{post}}^{\boldsymbol{y}|\boldsymbol{w}} \| \nu_{\text{prior}}) \, \pi_{\text{like}}(\boldsymbol{y}|m) d\boldsymbol{y} \, \mu_{\text{pr}}(dm) \\
&= \mathrm{E}_m \{ \mathrm{E}_{\boldsymbol{y}|m} \mathsf{D}_{\mathsf{KL}}(\nu_{\text{post}}^{\boldsymbol{y}|\boldsymbol{w}} \| \nu_{\text{prior}}) \}
\end{aligned}
$$

where the KL divergence measure given by:

$$
\begin{aligned}
\mathsf{D}_{\mathsf{KL}}(\nu_{\text{post}}^{\boldsymbol{y}|\boldsymbol{w}} \| \nu_{\text{prior}}) = \frac{1}{2} & \left(\mathsf{tr} \left(\widetilde{\mathbf{C}}_{\mathrm{pr}}^{-1} \widetilde{\mathbf{C}}_{\mathrm{post}} \right) \right) \\
& + \frac{1}{2} \left(\widetilde{\mu}_{\mathrm{post}} - \widetilde{\mu}_{\mathrm{pr}} \right)^T \widetilde{\mathbf{C}}_{\mathrm{pr}}^{-1} \left(\widetilde{\mu}_{\mathrm{post}} - \widetilde{\mu}_{\mathrm{pr}} \right) \\
& - \frac{N_{\mathrm{pred}}}{2} + \frac{1}{2} \log \left(\frac{\det \widetilde{\mathbf{C}}_{\mathrm{pr}}}{\det \widetilde{\mathbf{C}}_{\mathrm{post}}} \right) .
\end{aligned}
$$

GO-OED: Bayesian D-Optimality (KL-Divergence)

- The expected information gain reduces to:

$$
\begin{aligned}
\Psi(\boldsymbol{w}) = & \frac{1}{2}\mathsf{tr}\left(\left(\mathcal{P}\mathcal{C}_{\mathsf{pr}}^{1/2}\right)^{\dagger} \left(\mathcal{P}\mathcal{C}_{\mathsf{pr}}^{1/2}\right)\right) - \frac{N_{\mathrm{pred}}}{2} \\
& + \frac{1}{2} \log\det\left(\mathcal{P}\mathcal{C}_{\mathsf{pr}}\mathcal{P}^{T}\right) - \frac{1}{2}\log\det\left(\mathcal{P}\mathcal{C}_{\mathsf{post}}\mathcal{P}^{T}\right).
\end{aligned}
$$

GO-OED: Bayesian D-Optimality (KL-Divergence)

- The expected information gain reduces to:

$$\Psi(\boldsymbol{w}) = \frac{1}{2}\mathrm{tr}\Big(\left(\mathcal{P}\mathcal{C}_{\mathrm{pr}}^{1/2}\right)^{\dagger}\left(\mathcal{P}\mathcal{C}_{\mathrm{pr}}^{1/2}\right)\Big) - \frac{N_{\mathrm{pred}}}{2} + \frac{1}{2}\log\det\left(\mathcal{P}\mathcal{C}_{\mathrm{pr}}\mathcal{P}^{T}\right) - \frac{1}{2}\log\det\left(\mathcal{P}\mathcal{C}_{\mathrm{post}}\mathcal{P}^{T}\right).$$

- Maximizing the expected information gain is equivalent to minimizing the following quantity:

$$\widetilde{\Psi}(\boldsymbol{w}) = \log\det\left(\mathcal{P}\mathcal{C}_{\mathrm{post}}\mathcal{P}^{T}\right).$$

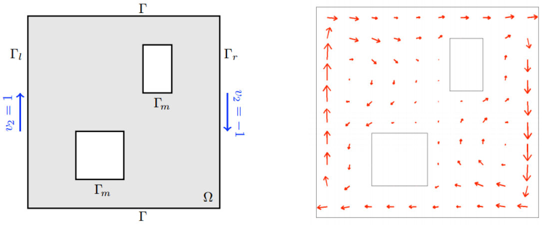
This is exactly the standard D-optimality criterion we addressed before.

Example: Advection-Diffusive transport

- Let u be the solution of:

$$
\begin{aligned}
u_t - \kappa \Delta u + \mathbf{v} \cdot \nabla u = 0 \quad & \text{in } \Omega \times [0, T] \\
u(0, x) = u_0 \quad & \text{in } \Omega \\
\kappa \nabla u \cdot \mathbf{n} = 0 \quad & \text{on } \partial\Omega \times [0, T]
\end{aligned}
$$

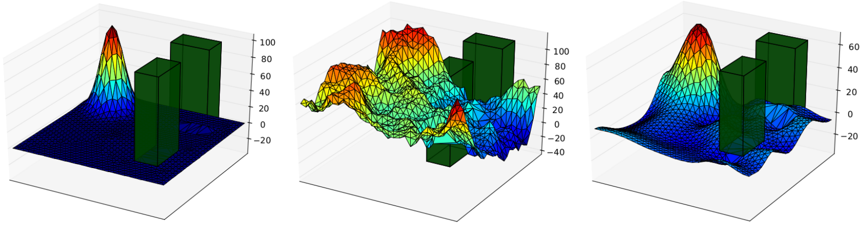
where κ is the diffusivity, and $\mathbf{v}$ is the velocity field:



- Here, $\mathrm{Dof} = 7836$,
- $T_0 = 0,\ T_1 = 1,\ T_{\mathrm{final}} = 3,\ dt = 0.2$,
- $N_{\mathrm{obs}} = 20$ observation points are randomly selected in the domain,
- the observation noise level is 0.05 of the magnitude of the initial true solution.

Preliminary results: the AD problem

The solution of the inverse problem (using hIPPYlib):



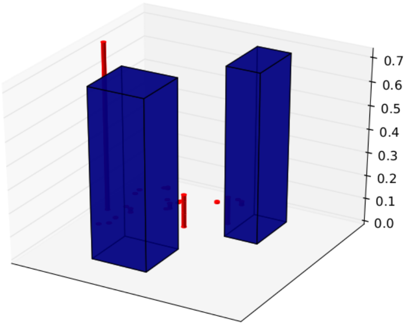
(c) True m (d) A prior sample (e) The MAP

Figure: Inverse problem components and solution (MAP)

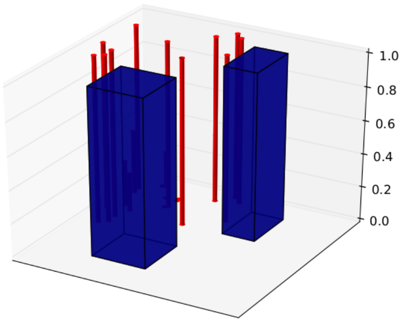
GO-OED Results: Advection-Diffusion model

Let $\mathcal{P}$ be a solution operator from the initial time $T_0 = 0$ to a future time $T = 4$, followed by a restriction operator that observes the concentration of the contaminant within a distance of $\epsilon = 0.009$ around the two buildings.

Assuming an ℓ_1 norm with penalty $\alpha = 10^{-3}$:



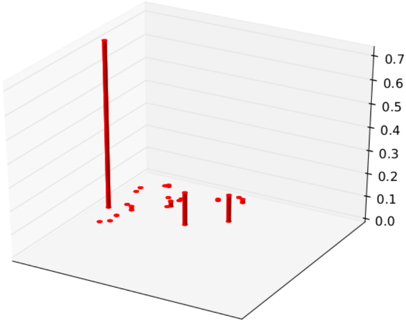
(a) A-Optimal Design



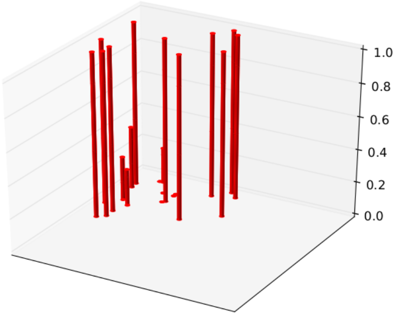
(b) D-Optimal Design

Figure: Goal-Oriented OED A, and D-Optimality results

GO-OED Results: Advection-Diffusion model



(a) A-Optimal Design



(b) D-Optimal Design

Figure: Goal-Oriented OED A, and D-Optimality results

GO-OED Results

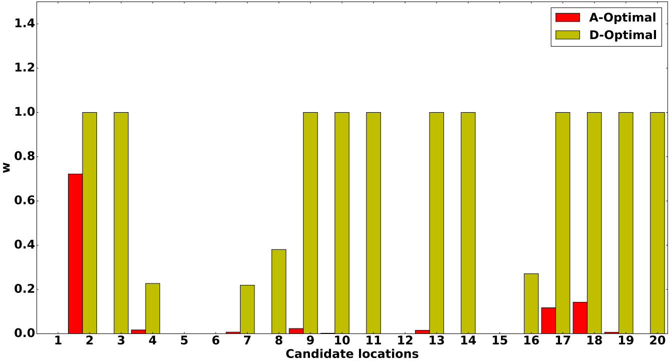


Figure: Goal-Oriented OED A, and D-Optimality results

Conclusions and Future Work

- We have extended the standard OED framework to the case where the QoI is a linear transformation of the inverse problem solution rather than the solution itself.
- Developed the A-optimality, the D-optimality criteria and the associated gradients,
- Showed the equivalence between standard D-optimality, and Bayesian D-optimality (expected information gain).

Conclusions and Future Work

- We have extended the standard OED framework to the case where the QoI is a linear transformation of the inverse problem solution rather than the solution itself.
- Developed the A-optimality, the D-optimality criteria and the associated gradients,
- Showed the equivalence between standard D-optimality, and Bayesian D-optimality (expected information gain).

- We are currently comparing results to the standard approach, i.e. with $\mathcal{P} = \mathbf{I}$, to understand the impact of incorporating Goals, on the optimal design.
- We will consider other regularization norms in addition to the ℓ_1 norm.

References

- Alexanderian, Alen, et al. "A-Optimal Design of Experiments for Infinite-Dimensional Bayesian Linear Inverse Problems with Regularized ℓ_0-Sparsification." SIAM Journal on Scientific Computing 36.5 (2014): A2122-A2148.
- Chaloner, Kathryn, and Isabella Verdinelli. "Bayesian experimental design: A review." Statistical Science (1995): 273-304.
- Haber, Eldad, Lior Horesh, and Luis Tenorio. "Numerical methods for experimental design of large-scale linear ill-posed inverse problems." Inverse Problems 24.5 (2008): 055012.
- Villa, Umberto, Petra, Noemi, and Ghattas, Omar. "hIPPYlib: An Extensible Software Framework for Large-Scale Deterministic and Linearized Bayesian Inversion" (2016): url = http://hippylib.github.io

Thank You!