New Efficient Approaches for Model-Constrained Optimal Design of Experiments

Ahmed Attia

Mathematics and Computer Science Division (MCS),
Argonne National Laboratory (ANL)
attia@mcs.anl.gov

LANS Seminar @ MCS . ANL
March 17, 2021

Thanks to:

Emil Constantinescu, Sven Leyffer, and Todd Munson
Mathematics and Computer Science Division,
Argonne National Laboratory

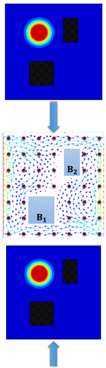
Motivation: parameter identification

Consider the contaminant concentration u in domain $\Omega \in \mathbb{R}^2$

Advection-Diffusion

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

- **Forward Problem:** given model state/parameter predict the expected model observations
- **Inverse Problem:** given spatiotemporal measurements, and a prior infer the **true QoI, e.g., function of the model state/parameter**
- **OED (sensor placement):** given a set of n_s candidate locations, determine the *optimal* positions, possibly under budget or sparsity constraints



Forward Problem and Bayesian Inversion

- **Forward problem:**

$$\mathbf{y} = \mathcal{F}(\theta) + \delta\,; \qquad \rho = \mathbf{P}\theta$$

  - $\theta \in \mathbb{R}^{\mathrm{N}_{\mathrm{state}}}$: discretized model parameter, e.g., IC
  - $\mathbf{y} \in \mathbb{R}^{\mathrm{N}_{\mathrm{obs}}}$: spatiotemporal sensor observations
  - $\rho \in \mathbb{R}^{\mathrm{N}_{\mathrm{pred}}}$: goal/prediction/QoI

- **Bayesian inverse problem (goal-oriented):**
  - **The prior:** knowledge about the QoI ρ prior to obtaining new observations

$$\rho := \mathbf{P}\theta \sim \mathcal{N}(\mathbf{P}\theta_{\mathrm{pr}}, \mathbf{P}\mathbf{\Gamma}_{\mathrm{pr}}\mathbf{P}^*)$$

    where $\mathbf{P}$ here is a linear goal/prediction operator
  - **The likelihood:** Gaussian observation noise;

$$\mathcal{L}(\mathbf{y}|\theta) \propto \exp\left(-\frac{1}{2}\left\|\mathcal{F}(\theta) - \mathbf{y}\right\|^2_{\mathbf{\Gamma}_{\mathrm{noise}}^{-1}}\right); \qquad \|\mathbf{x}\|^2_{\mathbf{A}} = \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}$$

  - **The posterior**: distribution of the QoI ρ conditioned on observations

Bayes' theorem: *Posterior* $\propto$ *Likelihood* $\times$ *Prior*

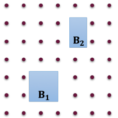
For a linear operator $\mathbf{F}$, the posterior is Gaussian $\mathcal{N}(\rho_{\text{post}}, \mathbf{\Sigma}_{\mathrm{post}})$ with

$$\rho_{\text{post}} = \mathbf{P}\mathbf{\Gamma}_{\mathrm{post}}\left(\mathbf{\Gamma}_{\mathrm{pr}}^{-1}\theta_{\mathrm{pr}} + \mathbf{F}^*\mathbf{\Gamma}_{\mathrm{noise}}^{-1}\mathbf{y}\right), \quad \mathbf{\Sigma}_{\mathrm{post}} = \underbrace{\mathbf{P}\left(\mathbf{F}^*\mathbf{\Gamma}_{\mathrm{noise}}^{-1}\mathbf{F} + \mathbf{\Gamma}_{\mathrm{pr}}^{-1}\right)^{-1}\mathbf{P}^*}_{\text{data-independent; only uncertainties}}$$

where $\mathbf{F}^*$, $\mathbf{P}^*$ is the adjoints of $\mathbf{F}$, $\mathbf{P}$

OED: sensor placement

Find the best subset (i.e., λ) of sensor location such as to optimize some utility function (e.g. identification accuracy, total uncertainties, information gain, etc.)



$$\xi := \left\{ \begin{array}{c} \mathbf{x}_1, \ldots, \mathbf{x}_{n_s} \\ \zeta_1, \ldots, \zeta_{n_s} \end{array} \right\} \rightarrow \mathbf{W}(\zeta) \rightarrow$$

$\mathbf{\Gamma}_{\rm noise}^{-1} \longrightarrow \mathbf{W}_{\Gamma}(\zeta)$ E.g.,

$$\mathbf{W}_{\Gamma}(\zeta) := \begin{cases} \mathbf{W}^{\frac{1}{2}}(\zeta)\mathbf{\Gamma}_{\rm noise}^{-1}\mathbf{W}^{\frac{1}{2}}(\zeta), \text{ or,} \\ \mathbf{\Gamma}_{\rm noise}^{-\frac{1}{2}}\mathbf{W}\mathbf{\Gamma}_{\rm noise}^{-\frac{1}{2}} \end{cases}$$

- **Weighted-likelihood:**

$$\mathcal{L}(\mathbf{y}|\theta;\zeta) \propto \exp\left(-\frac{1}{2}\left\|\mathbf{F}(\theta)-\mathbf{y}\right\|^2_{\mathbf{W}_{\Gamma}(\zeta)}\right)$$

- **Weighted posterior covariance:**

$$\mathbf{\Sigma}_{\rm post}(\zeta) = \mathbf{P}\left(\mathbf{F}^*\mathbf{W}_{\Gamma}(\zeta)\mathbf{F} + \mathbf{\Gamma}_{\rm pr}^{-1}\right)^{-1}\mathbf{P}^*$$

- **OED optimization problem:**

Binary OED

$$\zeta^{\rm opt} = \mathop{\arg\min}_{\zeta \in \{0,1\}^{n_s}} \mathcal{J}(\zeta) := \Psi(\zeta) + \alpha\,\Phi(\zeta)$$

Relaxed OED + Rounding

$$\zeta^{\rm opt} = \mathop{\arg\min}_{\zeta \in [0,1]^{n_s}} \mathcal{J}(\zeta) := \Psi(\zeta) + \alpha\,\Phi(\zeta)$$

- Ψ: utility function; $\mathrm{Tr}\left(\mathbf{\Sigma}_{\rm post}\right) \rightarrow$ A-optimality, $\det\left(\mathbf{\Sigma}_{\rm post}\right) \rightarrow$ D-optimality, etc.
- Φ: penalty function; ℓ_0, ℓ_1 etc.

OED (relaxed): sensor placement

Solving OED optimization problem: gradient-based approach

- **A-optimality:** $\Psi := \Psi^{\mathsf{GA}}$

$$\frac{\partial \Psi^{\mathsf{GA}}}{\partial \zeta_i} = -\operatorname{Tr}\left(\mathbf{P} \left(\mathbf{H}(\zeta) \right)^{-1} \mathbf{F}^* \frac{\partial \mathbf{W}_{\Gamma}(\zeta)}{\partial \zeta_i} \mathbf{F} \left(\mathbf{H}(\zeta) \right)^{-1} \mathbf{P}^* \right)$$

- **D-optimality:** $\Psi := \Psi^{\mathsf{GD}}$

$$\frac{\partial \Psi^{\mathsf{GD}}}{\partial \zeta_i} = \operatorname{Tr}\left(\mathbf{\Sigma}_{\rm post}^{-1}(\zeta) \mathbf{P} \left(\mathbf{H}(\zeta) \right)^{-1} \mathbf{F}^* \frac{\partial \mathbf{W}_{\Gamma}(\zeta)}{\partial \zeta_i} \mathbf{F} \left(\mathbf{H}(\zeta) \right)^{-1} \mathbf{P}^* \right)$$

$$\mathbf{H}(\zeta) = \mathbf{\Gamma}_{\rm pr}^{-1} + \mathbf{F}^* \mathbf{W}_{\Gamma}(\zeta) \mathbf{F}$$

OED Challenges

- **Linear vs. Nonlinear forward operator $\mathbf{F}$**
  1. Laplace/Gaussian approximation: iterate (OED $\rightarrow$ MAP $\rightarrow$ OED ...)
  2. Use information criteria; e.g., KL divergence between posterior and prior
- **Computational cost of the OED criterion and the gradient**
  1. Randomized approximation: A-optimality is easy, D-optimality is challenging!
  2. Reduced order model $\mathbf{F} \rightarrow \widetilde{\mathbf{F}}$ (POD, DNN, ...)
- **Rounding (SUR, continuation, ...) is challenging**
- **Observation correlations:**
  - No observation correlation:
    - $\mathbf{\Gamma}_{\rm noise}$ and $\mathbf{W}$ are diagonal
    - $\mathbf{W}_{\Gamma}(\zeta) = \text{diag}\,(\zeta) \rightarrow \mathbf{W}^{\frac{1}{2}}(\zeta)\mathbf{\Gamma}_{\rm noise}^{-1}\mathbf{W}^{\frac{1}{2}}(\zeta) \equiv \mathbf{\Gamma}_{\rm noise}^{-\frac{1}{2}}\mathbf{W}\mathbf{\Gamma}_{\rm noise}^{-\frac{1}{2}}$
  - **Spatiotemporal observation correlations** (generally overlooked)
    - A general/flexible approach to weight observation variances/covariances $\longleftarrow$ **Discussed Next**

Stochastic learning for binary OED
Second part of this talk

Schur-product OED formulation

We formulate the weighted likelihood as:

$$\mathcal{L}(\mathbf{y}|\theta;\zeta) \propto \exp\left(-\frac{1}{2}\left\|\mathbf{F}(\theta)-\mathbf{y}\right\|^2_{\mathbf{W}_\Gamma(\zeta)}\right),$$

$$\mathbf{W}_\Gamma(\zeta) := \mathbf{L}^\mathsf{T}\widetilde{\mathbf{\Gamma}}^{-1}_{\text{noise}}(\zeta)\mathbf{L}\,; \qquad \widetilde{\mathbf{\Gamma}}_{\text{noise}}(\zeta) := \mathbf{L}\Big(\mathbf{\Gamma}_{\text{noise}} \odot \mathbf{W}(\zeta)\Big)\mathbf{L}^\mathsf{T}\,.$$

- $\odot$ is the Hadamard (Schur) product
- Entries of the localization/weighting matrix $\mathbf{W}(\zeta)$ influence the observation correlations
- $\mathbf{W}(\zeta)$ is a symmetric and doubly nonnegative weighting kernel, with entries:

$$\varpi\left(t_k, t_l;\, \zeta_i, \zeta_j\right); \quad k,l = 1,2,\ldots,n_t\,;\; i,j = 1,2,\ldots,n_\text{s}$$

  * ϖ is a symmetric weighting (localization) function, such that:

$$\varpi\left(t_k, t_l;\, \zeta_i, \zeta_j\right) = \varpi\left(t_l, t_k;\, \zeta_i, \zeta_j\right) = \varpi\left(t_k, t_l;\, \zeta_j, \zeta_i\right) = \varpi\left(t_l, t_k;\, \zeta_j, \zeta_i\right)$$

Let $n_t = 2,\, n_\text{s} = 2$

$$\mathbf{\Gamma}_{\text{noise}} = \begin{bmatrix} \mathbf{R}_{1,1} & \mathbf{R}_{1,2} \\ \mathbf{R}_{2,1} & \mathbf{R}_{2,2} \end{bmatrix} \rightarrow \mathbf{W}(\zeta) := \begin{bmatrix} \varpi(t_1,t_1;\,\zeta_1,\zeta_1) & \varpi(t_1,t_1;\,\zeta_1,\zeta_2) & \varpi(t_1,t_2;\,\zeta_1,\zeta_1) & \varpi(t_1,t_2;\,\zeta_1,\zeta_2) \\ \varpi(t_1,t_1;\,\zeta_2,\zeta_1) & \varpi(t_1,t_1;\,\zeta_2,\zeta_2) & \varpi(t_1,t_2;\,\zeta_2,\zeta_1) & \varpi(t_1,t_2;\,\zeta_2,\zeta_2) \\ \varpi(t_2,t_1;\,\zeta_1,\zeta_1) & \varpi(t_2,t_1;\,\zeta_1,\zeta_2) & \varpi(t_2,t_2;\,\zeta_1,\zeta_1) & \varpi(t_2,t_2;\,\zeta_1,\zeta_2) \\ \varpi(t_2,t_1;\,\zeta_2,\zeta_1) & \varpi(t_2,t_1;\,\zeta_2,\zeta_2) & \varpi(t_2,t_2;\,\zeta_2,\zeta_1) & \varpi(t_2,t_2;\,\zeta_2,\zeta_2) \end{bmatrix}$$

Schur product OED formulation: space correlations

$$\mathbf{\Gamma}_{\rm noise} = \bigoplus_{m=1}^{n_t} (\mathbf{R}_m) := \text{diag}\left(\mathbf{R}_1, \ldots, \mathbf{R}_{n_t}\right) \,, \quad \mathbf{L} = \bigoplus_{m=1}^{n_t} (\mathbf{L}_m) \,, \quad \mathbf{W}_{\Gamma}(\zeta) = \bigoplus_{m=1}^{n_t} \left(\mathbf{V}_m^{\dagger}(\zeta)\right) \,,$$

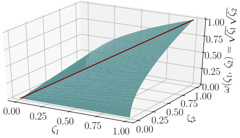
$$\mathbf{V}_m^{\dagger}(\zeta) = \mathbf{L}_m^{\mathsf{T}} \mathbf{V}_m^{-1}(\zeta) \mathbf{L}_m \,, \qquad \mathbf{V}_m(\zeta) = \mathbf{L}_m \left(\mathbf{R}_m \odot \left(\sum_{i,j=1}^{n_{\rm s}} \omega(\zeta_i,\, \zeta_j) \mathbf{e}_i \mathbf{e}_j^{\mathsf{T}} \right) \right) \mathbf{L}_m^{\mathsf{T}} \,,$$

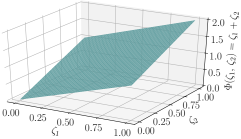
$$\Phi(\zeta) = \left\| \left(\omega(\zeta_1, \zeta_1),\, \ldots,\, \omega(\zeta_{n_{\rm s}}, \zeta_{n_{\rm s}}) \right)^{\mathsf{T}} \right\|_p$$

SQRT kernel:

$$\omega(\zeta_i, \zeta_j) = \sqrt{\zeta_i}\sqrt{\zeta_j};$$

$$\frac{\partial \omega(\zeta_i, \zeta_j)}{\partial \zeta_k} = \frac{1}{2} \left(\frac{\sqrt{\zeta_j}}{\sqrt{\zeta_i}} \delta_{i,k} + \frac{\sqrt{\zeta_i}}{\sqrt{\zeta_j}} \delta_{j,k} \right)$$





- $\zeta_i, \zeta_j \in [0, 1]$; $i, j = 1, 2, \ldots, n_{\rm s}$,
- $\delta_{i,k}$: standard delta function
- $\zeta \in [0, 1] \rightarrow$ constrained-optimization (box-constraints)
- *Relates to traditional OED formulation*

Schur product OED formulation: space correlations

Solving Shur OED optimization problem: gradient-based approach

$$\frac{\partial \mathbf{W}_{\Gamma}(\zeta)}{\partial \zeta_i} = \frac{\partial \left(\mathbf{L}^{\mathsf{T}} \widetilde{\mathbf{\Gamma}}_{\rm noise}^{-1}(\zeta) \mathbf{L} \right)}{\partial \zeta_i} = - \mathbf{L}^{\mathsf{T}} \widetilde{\mathbf{\Gamma}}_{\rm noise}^{-1}(\zeta) \mathbf{L} \left(\mathbf{\Gamma}_{\rm noise} \odot \frac{\partial \mathbf{W}(\zeta)}{\partial \zeta_i} \right) \mathbf{L}^{\mathsf{T}} \widetilde{\mathbf{\Gamma}}_{\rm noise}^{-1}(\zeta) \mathbf{L}$$

$$= - \mathbf{W}_{\Gamma}(\zeta) \left(\mathbf{\Gamma}_{\rm noise} \odot \frac{\partial \mathbf{W}(\zeta)}{\partial \zeta_i} \right) \mathbf{W}_{\Gamma}(\zeta)$$

$$\mathbf{W}' = [\eta_1, \, \eta_2, \dots, \, \eta_{n_{\rm s}}] \, , \quad \eta_j = \left(\frac{1}{1+\delta_{1,j}} \frac{\partial \omega(\zeta_1, \zeta_j)}{\partial \zeta_j}, \dots, \frac{1}{1+\delta_{n_{\rm s},j}} \frac{\partial \omega(\zeta_{n_{\rm s}}, \zeta_j)}{\partial \zeta_j} \right)^{\mathsf{T}}$$

- **A-optimality:** $\zeta^{\rm opt} = \underset{\zeta \in [0,\,1]^{n_{\rm s}}}{\arg\min} \, \mathcal{J}(\zeta) := \mathrm{Tr}\left(\mathbf{\Sigma}_{\rm post}(\zeta) \right) + \alpha \, \Phi(\zeta)$

$$\nabla_{\zeta} \Psi^{\mathsf{GA}}(\zeta) = 2 \sum_{m=1}^{n_t} \mathrm{diag} \left(\mathbf{V}_m^{\dagger}(\zeta) \mathbf{F}_{0,m} \mathbf{H}^{-1}(\zeta) \mathbf{P}^* \mathbf{P} \mathbf{H}^{-1}(\zeta) \mathbf{F}_{m,0}^* \mathbf{V}_m^{\dagger}(\zeta) \left(\mathbf{R}_m \odot \mathbf{W}' \right) \right)$$

- **D-optimality:** $\zeta^{\rm opt} = \underset{\zeta \in [0,\,1]^{n_{\rm s}}}{\arg\min} \, \mathcal{J}(\zeta) := \log\det\left(\mathbf{\Sigma}_{\rm post}(\zeta) \right) + \alpha \, \Phi(\zeta)$

$$\nabla_{\zeta} \Psi^{\mathsf{GD}}(\zeta) = 2 \sum_{m=1}^{n_t} \mathrm{diag} \left(\mathbf{V}_m^{\dagger}(\zeta) \mathbf{F}_{0,m} \mathbf{H}^{-1}(\zeta) \mathbf{P}^* \mathbf{\Sigma}_{\rm post}^{-1}(\zeta) \mathbf{P} \mathbf{H}^{-1}(\zeta) \mathbf{F}_{m,0}^* \mathbf{V}_m^{\dagger}(\zeta) \left(\mathbf{R}_m \odot \mathbf{W}' \right) \right)$$

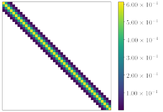
Experimental Settings

- Numerical model (AD): u solves:

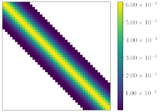
$$
\begin{aligned}
u_t - \kappa \Delta u + \mathbf{v} \cdot \nabla u &= 0 \quad \text{in } \mathcal{D} \times [0, T], \\
u(x,\, 0) &= \theta \quad \text{in } \mathcal{D}, \\
\kappa \nabla u \cdot \mathbf{n} &= 0 \quad \text{on } \partial\mathcal{D} \times [0, T]
\end{aligned}
$$

  - * $\Omega \in \mathbb{R}^2$ is an open and bounded domain
  - * u the concentration of a contaminant in the domain Ω
  - * κ is the diffusivity,
  - * $\mathbf{v}$ is the velocity field

- Observational setup: $n_s = 43$ candidate sensor locations, with
  - * observation times $t_k := t_1 + s\Delta t$
  - * $\Delta t = 0.2$ is the model simulation time step;
  - * $t_1 = 1,\ s = 0, 1, \ldots, 5$
  - * Observation correlations; synthetic, created with Gaspari-Cohn, and 5% noise level

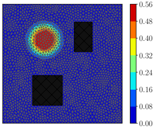


$\mathbf{R}_k$ with $\ell = 1$

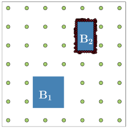


$\mathbf{R}_k$ with $\ell = 3$

- QoI: ρ is the contaminant concentration predicted around the second building at time $t_p = 2.2$

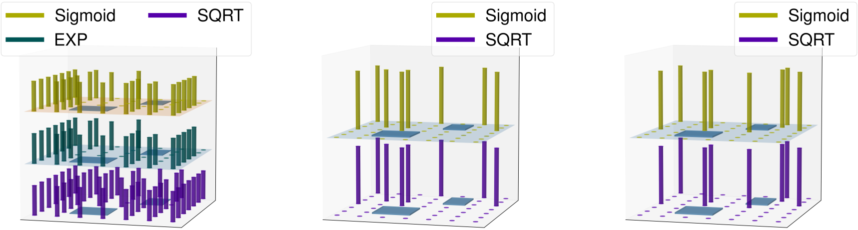


Ground truth (initial parameter) θ



QoI (prediction) gridpoints.

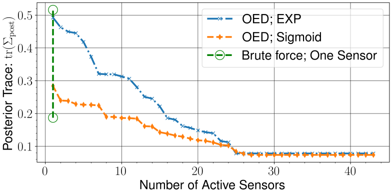
Numerical Results: space-correlation I



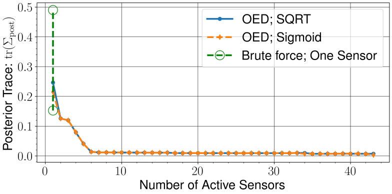
$\ell = 0$ (No correlation) $\quad\quad$ $\ell = 1$ $\quad\quad$ $\ell = 3$

Optimal design weights resulting from solving the OED problem, with $\alpha = 0$.
Sensors with weights above 0.5 are activated.

Numerical Results: space-correlation II



$\ell = 0.$



$\ell = 3.$

Objective value (post-trace) evaluated for increasing number of sensors, corresponding to highest optimal weights. Here, $\alpha = 0$.

Schur product OED formulation: space-time correlations

Entries of $\mathbf{\Gamma}_{\rm noise}^{-1}$ are weighted by

$$\varpi(\zeta_i, \zeta_j;\, t_m, t_n) := \rho(t_m,\, t_n)\, \omega(\zeta_i, \zeta_j)\,; \qquad \begin{array}{c} i, j = 1, 2, \ldots, n_{\rm s}\,, \\ m, n = 1, 2, \ldots, n_t\,. \end{array}$$

$\rho(t_m,\, t_n)$ is a symmetric temporal decorrelation function:

- $\rho(t_m,\, t_m) = 1$
- $\rho(t_m,\, t_m)$ is conversely related with $d(t_m,\, t_n)$, the distance between t_m and t_n

---

- Gauss:

$$\rho(t_m,\, t_n) := \exp\left(\frac{-d(t_m,\, t_n)}{2\ell^2}\right)$$

- Gaspari-Cohn

$$\rho(t_m, t_n) = \begin{cases} -\frac{\upsilon^5}{4} + \frac{\upsilon^4}{2} + \frac{5\upsilon^3}{8} - \frac{5\upsilon^2}{3} + 1\,, & 0 \leq \upsilon \leq 1 \\ \frac{\upsilon^5}{12} - \frac{\upsilon^4}{2} + \frac{5\upsilon^3}{8} + \frac{5\upsilon^2}{3} - 5\upsilon + 4 - \frac{2}{3\upsilon}\,, & 1 \leq \upsilon \leq 2 \\ 0\,, & \upsilon \geq 2\,, \end{cases}$$

$\upsilon := \frac{d(t_m, t_n)}{\ell}$ and, ℓ is a predefined correlation length scale

Schur product OED formulation: space-time correlations

Solving Shur-OED optimization problem: gradient-based approach

$$\mathbf{W} := \left[\varpi \left(\zeta_{(k-1)\%n_s+1}, \zeta_{(h-1)\%n_s+1};\, t_{\left\lfloor \frac{k-1}{n_t} \right\rfloor +1}, t_{\left\lfloor \frac{h-1}{n_t} \right\rfloor +1} \right) \right]_{k,h=1,2,\ldots,\mathrm{N_{obs}}}$$

$$\vartheta_{i,m}[k] := \frac{1}{1+\delta_{i,\,(k-1)\%n_s+1}} \frac{\partial}{\partial \zeta_i} \varpi \left(\zeta_i, \zeta_{(k-1)\%n_s+1};\, t_m, t_{\left\lfloor \frac{k-1}{n_t} \right\rfloor +1} \right) ; \qquad \begin{array}{l} i = 1, \ldots, n_s\,, \\ m = 1, \ldots, n_t\,, \\ k = 1, \ldots, \mathrm{N_{obs}} \end{array}$$

$$\frac{\partial \mathbf{W}_\Gamma}{\partial \zeta_i} = -\mathbf{W}_\Gamma(\zeta) \left[\sum_{m=1}^{n_t} \mathbf{e}_q \left((\mathbf{\Gamma}_{\rm noise}^{-1} \mathbf{e}_q) \odot \vartheta_{i,m} \right)^{\mathsf{T}} + \sum_{m=1}^{n_t} \left((\mathbf{\Gamma}_{\rm noise}^{-1} \mathbf{e}_q) \odot \vartheta_{i,m} \right) \mathbf{e}_q^{\mathsf{T}} \right] \mathbf{W}_\Gamma(\zeta)\,, \quad \begin{array}{c} q = i + (m-1) n_s\,, \\ i = 1, \ldots, n_s \end{array}$$

- A-optimality: $\Psi := \Psi^{\mathsf{GA}}$

$$\nabla_\zeta \Psi^{\mathsf{GA}}(\zeta) = 2 \sum_{i=1}^{n_s} \sum_{m=1}^{n_t} \mathbf{e}_i \mathbf{e}_q^{\mathsf{T}} \mathbf{W}_\Gamma(\zeta) \mathbf{F} \mathbf{H}^{-1}(\zeta) \mathbf{P}^* \mathbf{P} \mathbf{H}^{-1}(\zeta) \mathbf{F}^* \mathbf{W}_\Gamma(\zeta) \left((\mathbf{\Gamma}_{\rm noise}^{-1} \mathbf{e}_q) \odot \vartheta_{i,m} \right)$$

- D-optimality: $\Psi := \Psi^{\mathsf{GD}}$

$$\nabla_\zeta \Psi^{\mathsf{GD}}(\zeta) = 2 \sum_{i=1}^{n_s} \sum_{m=1}^{n_t} \mathbf{e}_i \left(\mathbf{\Gamma}_{\rm noise}^{-1} \mathbf{e}_q \odot \vartheta_{i,m} \right)^{\mathsf{T}} \mathbf{W}_\Gamma(\zeta) \mathbf{F} \mathbf{H}^{-1}(\zeta) \mathbf{P}^* \mathbf{\Sigma}_{\rm post}^{-1}(\zeta) \mathbf{P} \mathbf{H}^{-1}(\zeta) \mathbf{F}^* \mathbf{W}_\Gamma(\zeta) \mathbf{e}_q$$

Schur product OED formulation

Efficient evaluation of the OED objective Ψ:

- reduced-order approximation of $\mathbf{F}$
- randomized approximation of Ψ

A-optimality: randomized-trace approximation: recall: $\mathbf{W}_\Gamma(\zeta) = \mathbf{L}^\mathsf{T} \left(\mathbf{L} \Big(\mathbf{\Gamma}_{\text{noise}} \odot \mathbf{W}(\zeta) \Big) \mathbf{L}^\mathsf{T} \right)^{-1} \mathbf{L}$

$$\Psi^{\mathsf{GA}}(\zeta) \approx \widetilde{\Psi}^{\mathsf{GA}}(\zeta) = \frac{1}{n_r} \sum_{r=1}^{n_r} \mathbf{z}_r^\mathsf{T} \mathbf{P} \big(\mathbf{F}\, \mathbf{W}_\Gamma(\zeta)\, \mathbf{F}^* + \mathbf{\Gamma}_{\text{pr}}^{-1} \big)^{-1} \mathbf{P}^* \mathbf{z}_r \,, \quad \mathbf{z}_r \in \mathbb{R}^{\mathrm{N}_{\text{pred}}}$$

$$\frac{\partial \widetilde{\Psi}^{\mathsf{GA}}(\zeta)}{\partial \zeta_i} = \frac{1}{n_r} \sum_{r=1}^{n_r} \mathbf{z}_r^\mathsf{T} \mathbf{P} \mathbf{H}^{-1}(\zeta) \mathbf{F}^* \mathbf{W}_\Gamma(\zeta) \left(\mathbf{\Gamma}_{\text{noise}} \odot \frac{\partial \mathbf{W}(\zeta)}{\partial \zeta_i} \right) \mathbf{W}_\Gamma(\zeta) \mathbf{F} \mathbf{H}^{-1}(\zeta) \mathbf{P}^* \mathbf{z}_r$$

- space correlations:

$$\nabla_\zeta \widetilde{\Psi}^{\mathsf{GA}}(\zeta) = \frac{2}{n_r} \sum_{r=1}^{n_r} \sum_{m=1}^{n_t} \psi_{r,m}^* \odot \big((\mathbf{R}_m \odot \mathbf{W}')\, \psi_{r,m} \big) \,; \qquad \begin{aligned} \psi_{r,m} &:= \mathbf{V}_m^\dagger(\zeta) \mathbf{F}_{0,m} \mathbf{H}^{-1}(\zeta) \mathbf{P}^* \mathbf{z}_r \,, \\ \psi_{r,m}^* &:= \big(\mathbf{z}_r^\mathsf{T} \mathbf{P} \mathbf{H}^{-1}(\zeta) \mathbf{F}_{m,0}^* \mathbf{V}_m^\dagger(\zeta) \big)^\mathsf{T} \end{aligned}$$

- space-time correlations:

$$\nabla_\zeta \widetilde{\Psi}^{\mathsf{GA}}(\zeta) = 2 \sum_{r=1}^{n_r} \sum_{i=1}^{n_s} \sum_{m=1}^{n_t} \mathbf{e}_i \psi_r^* \mathbf{e}_q \left((\mathbf{\Gamma}_{\text{noise}} \mathbf{e}_q) \odot \vartheta_{i,m} \right)^\mathsf{T} \psi_r \,; \qquad \begin{aligned} \psi_r &:= \mathbf{W}_\Gamma(\zeta) \mathbf{F} \mathbf{H}^{-1}(\zeta) \mathbf{P}^* \mathbf{z}_r \,, \\ \psi_r^* &:= \mathbf{z}_r^\mathsf{T} \mathbf{P} \mathbf{H}^{-1}(\zeta) \mathbf{F}^* \mathbf{W}_\Gamma(\zeta) \end{aligned}$$

OED Challenges (relaxed formulation)

- Computational cost (objective/criterion and gradient)
- Differentiability (objective and regularizer)
- Enforcing sparsity
- Rounding/thresholding

Can we solve binary OED problems efficiently without relaxation+rounding?

Solving the Binary OED problem (without relaxation+rounding) I

The Binary OED Optimization Problem

$$\zeta^{\text{opt}} = \underset{\zeta \in \{0,1\}^{n_s}}{\arg\min}\, \mathcal{J}(\zeta) := \Psi(\zeta) + \alpha\, \Phi(\zeta)\,,$$

* the function $\Phi(\zeta)$ promotes regularization or sparsity on the design, e.g., $\Phi(\zeta) := \|\zeta\|_0$, or $\Phi(\zeta) := \|\zeta\|_0 = \lambda \quad \leftarrow$ **nondifferentiable!**
* α is a user-defined regularization parameter

IDEA

- Associate with each candidate sensor x_i, $i = 1, \ldots, n_s$: a probability of activation $\theta_i \in [0, 1]$
- Let $\zeta_i \in \{0, 1\}$ be the status of the ith sensor: $0 \rightarrow$ OFF , $1 \rightarrow$ ON
- ζ_i are independent Bernoulli random variables:

$$\mathbb{P}\left(\zeta_i = \nu_i | \theta_i\right) = \theta_i^{\nu_i} (1 - \theta_i)^{\nu_i} = \begin{cases} \theta_i\,, & \nu_i = 1 \\ (1 - \theta_i)\,, & \nu_i = 0 \end{cases}$$

- Probability associated with any observational configuration (described by the joint PMF (policy)): $\mathbb{P}\left(\zeta|\theta\right) := \prod_{i=1}^{n_s} \theta_i^{\zeta_i} \left(1 - \theta_i\right)^{1-\zeta_i}\,, \quad \zeta_i \in \{0, 1\}, \quad \theta_i \in [0, 1]\,.$

Solving the Binary OED problem (without relaxation+rounding) II

We replace the original problem with:

Stochastic OED Optimization Problem

$$\theta^{\text{opt}} = \underset{\theta\in[0,1]^{n_s}}{\arg\min}\, \Upsilon(\theta) := \mathbb{E}_{\zeta\sim\mathbb{P}(\zeta|\theta)}\Big[\mathcal{J}(\zeta)\Big] = \sum_{k=1}^{2^{n_s}} \mathcal{J}(\zeta[k])\mathbb{P}\left(\zeta[k]|\theta\right) ,$$

where $k = 1 + \sum_{i=1}^{n_s} \zeta_i\, 2^{i-1}$

- θ^{opt} : parameter of an optimal policy that describes probability of sensors' activation
- Solution: (exact) gradient approach:

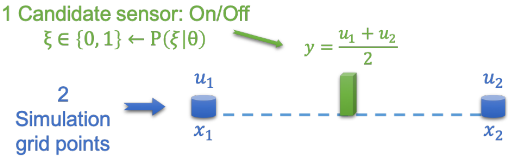
$$\nabla_\theta\, \Upsilon(\theta) = \nabla_\theta\, \mathbb{E}_{\zeta\sim\mathbb{P}(\zeta|\theta)}\Big[\mathcal{J}(\zeta)\Big] = \nabla_\theta \sum_{\zeta} \mathcal{J}(\zeta)\mathbb{P}\left(\zeta|\theta\right) = \sum_{k=1}^{2^{n_s}} \mathcal{J}(\zeta[k])\nabla_\theta\, \mathbb{P}\left(\zeta[k]|\theta\right) ,$$

$$\nabla_\theta\, \mathbb{P}\left(\zeta[k]|\theta\right) = \sum_{j=1}^{n_s} \left.\frac{\partial\mathbb{P}\left(\zeta[k]|\theta\right)}{\partial\theta_j}\right|_{\zeta=\zeta[k]} = \sum_{j=1}^{n_s} (-1)^{1-\zeta_j[k]} \prod_{\substack{i=1\\ i\neq j}}^{n_s} \theta_i^{\zeta_i[k]}\left(1-\theta_i\right)^{1-\zeta_i[k]}\mathbf{e}_j$$

$\nabla_\zeta \mathcal{J}(\zeta)$ is no longer required

Solving the Binary OED problem (without relaxation+rounding) III

One-dimensional Example; one candidate sensor:



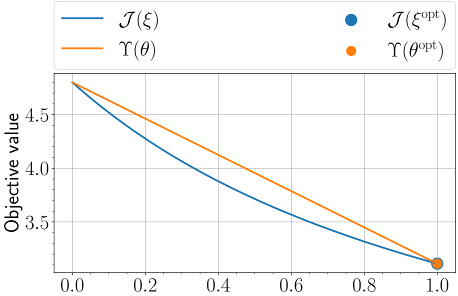
$$\mathbf{F} := \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}; \quad \mathbf{\Gamma}_{\rm pr} := \mathrm{diag}\,(4,1); \quad \mathbf{\Gamma}_{\rm noise} := 1 \quad \Rightarrow$$

$$\mathcal{J}(\zeta) := \mathrm{Tr}\left(\left(\mathbf{F}^\intercal \frac{\zeta}{\mathbf{\Gamma}_{\rm noise}} \mathbf{F} + \mathbf{\Gamma}_{\rm pr}^{-1} \right)^{-1} \right) = \mathrm{Tr}\left(\begin{bmatrix} 0.25\zeta + 0.25 & 0.25\zeta_1 \\ 0.25\zeta & 0.25\zeta + 1 \end{bmatrix}^{-1} \right)$$

$$\Upsilon(\theta) := \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\theta)} \left[\mathcal{J}(\zeta) \right] = (1-\theta)\mathcal{J}(0) + \theta \mathcal{J}(1)$$

$$\nabla_\theta \Upsilon(\theta) = \mathcal{J}(1) - \mathcal{J}(0)$$

Solving the Binary OED problem (**without relaxation+rounding**) IV



The objective function $\mathcal{J}$ of the relaxed OED problem and the objective function Υ of the corresponding stochastic OED problem for a one-dimensional problem.

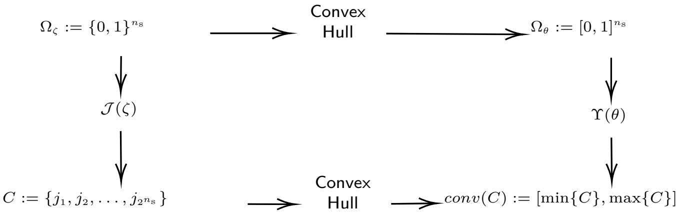
Relation between $\mathcal{J}$, Υ and their respective domains and codomains

Lemma

The optimal solutions of the binary (deterministic) and the stochastic OED problem are such that

$$\underset{\zeta\in\Omega_\zeta}{\arg\min}\,\mathcal{J}(\zeta) \subseteq \underset{\theta\in\Omega_\theta}{\arg\min}\,\Upsilon(\theta)\,.$$

Moreover, if the optimal solution $\zeta^{\rm opt}$ is unique, then $\theta^{\rm opt}=\zeta^{\rm opt}$, where $\theta^{\rm opt}$ is the unique optimal solution of the stochastic OED optimization problem



Solving the stochastic OED optimization problem I

- Evaluating the objective Υ and the gradient $\nabla_\theta \Upsilon$ require enumerating all possible designs
- *"The kernel trick"*: $\nabla_\theta \log(\mathbb{P}(\zeta|\theta)) = \frac{1}{\mathbb{P}(\zeta|\theta)} \nabla_\theta \mathbb{P}(\zeta|\theta) \rightarrow \nabla_\theta \mathbb{P}(\zeta|\theta) = \mathbb{P}(\zeta|\theta)\ \nabla_\theta \log(\mathbb{P}(\zeta|\theta))$

$$
\begin{aligned}
\nabla_\theta\, \Upsilon(\theta) &= \sum_{k=1}^{2^{n_s}} \mathcal{J}(\zeta[k])\, \nabla_\theta \mathbb{P}(\zeta[k]|\theta) \\
&= \sum_{k=1}^{2^{n_s}} \Big(\mathcal{J}(\zeta[k])\, \nabla_\theta\, \log \mathbb{P}(\zeta[k]|\theta) \Big) \mathbb{P}(\zeta[k]|\theta) = \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\theta)} \Big[\mathcal{J}(\zeta)\, \nabla_\theta\, \log \mathbb{P}(\zeta|\theta) \Big]
\end{aligned}
$$

- The gradient can be written as an expectation of gradients; this enables us to approximate the gradient using MC sampling by following a stochastic optimization approach

$$
\begin{aligned}
\mathbf{g} \approx \widehat{\mathbf{g}} &:= \frac{1}{\mathrm{N_{ens}}} \sum_{j=1}^{\mathrm{N_{ens}}} \mathcal{J}(\zeta[j]) \nabla_\theta\, \log \mathbb{P}(\zeta[j]|\theta) \\
&= \frac{1}{\mathrm{N_{ens}}} \sum_{j=1}^{\mathrm{N_{ens}}} \mathcal{J}(\zeta[j]) \sum_{i=1}^{n_s} \left(\frac{\zeta_i[j]}{\theta_i} + \frac{\zeta_i[j]-1}{1-\theta_i} \right) \mathbf{e}_i
\end{aligned}
$$

where $\zeta[j] \sim \mathbb{P}(\zeta|\theta) := \prod_{i=1}^{n_s} \theta_i^{\zeta_i}\ (1-\theta_i)^{1-\zeta_i}\ , \quad , j = 1, 2, \ldots, \mathrm{N_{ens}}$

Solving the stochastic OED optimization problem II

- A stochastic steepest-descent algorithm:

$$\theta^{(n+1)} = \mathbf{L}\left(\theta^{(n)} - \eta^{(n)}\widehat{\mathbf{g}}^{(n)}\right); \qquad \mathbf{L}(\theta_i) := \min\{1, \max\{0, \theta_i\}\} \equiv \begin{cases} 0 & \text{if} \quad \theta_i < 0\,, \\ \theta_i & \text{if} \quad \theta_i \in [0,1]\,, \\ 1 & \text{if} \quad \theta_i > 1\,, \end{cases}$$

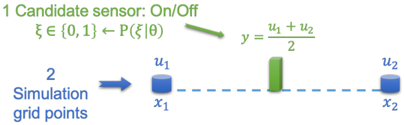
Algorithm 1 Stochastic optimization algorithm for binary OED

Input: Initial distribution parameter $\theta^{(0)}$, step size schedule $\eta^{(n)}$; sample sizes $\mathrm{N}_{\mathrm{ens}}$, m

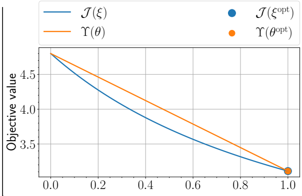
Output: ζ^{opt}

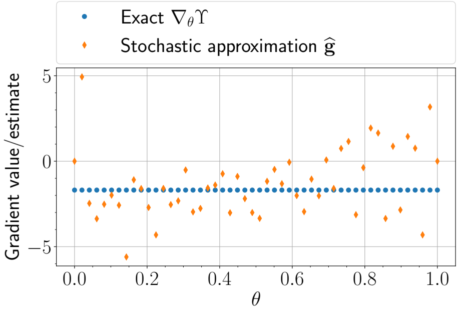
1: Initialize $n = 0$
2: **while** Not Converged **do**
3: Update $n \leftarrow n + 1$
4: Sample $\{\zeta[i]; i = 1, 2, \ldots, \mathrm{N}_{\mathrm{ens}}\} \sim \mathbb{P}\left(\zeta|\theta^{(n)}\right)$
5: Calculate $\widehat{g}^{(n)} = \frac{1}{\mathrm{N}_{\mathrm{ens}}} \sum_{j=1}^{\mathrm{N}_{\mathrm{ens}}} \left(\mathcal{J}(\zeta[j])\right) \sum_{i=1}^{n_s} \left(\frac{\zeta_i[j]}{\theta_i} + \frac{\zeta[j]_i - 1}{1-\theta_i}\right) \mathbf{e}_i$
6: Update $\theta^{(n+1)} = \mathbf{L}\left(\theta^{(n)} - \eta^{(n)}\widehat{g}^{(n)}\right)$
7: **end while**
8: Set $\theta^{\mathrm{opt}} = \theta^{(n)}$
9: Sample $\{\zeta[j]; j = 1, 2, \ldots, m\} \sim \mathbb{P}\left(\zeta|\theta^{\mathrm{opt}}\right)$, and calculate $\mathcal{J}(\zeta[j])$
10: **return** ζ^{opt}: the design ζ with smallest value of $\mathcal{J}$ in the sample.

One dimensional example: revisited



$$\mathcal{J}(\zeta) = \mathrm{Tr}\left(\begin{bmatrix} 0.25\zeta + 0.25 & 0.25\zeta_1 \\ 0.25\zeta & 0.25\zeta + 1 \end{bmatrix}^{-1}\right)$$





The estimate is unbiased, but it exhibits high variability!

Improving the Stochastic Optimization Procedure I

Lemma 4.1: Convergence guarantees, and statistics of the gradient estimator

The stochastic estimator $\widehat{\mathrm{g}}$ is **unbiased**, with sampling total variance $\mathrm{var}\left(\widehat{\mathbf{g}}\right)$, such that

$$\mathbb{E}\left[\widehat{\mathbf{g}}\right] = \mathbf{g} = \nabla_\theta \Upsilon(\theta)$$

$$\mathrm{var}\left(\widehat{\mathbf{g}}\right) = \frac{1}{\mathrm{N_{ens}}} \mathrm{var}\left(\mathcal{J}(\zeta) \nabla_\theta \log \mathbb{P}\left(\zeta|\theta\right)\right) \leq C < \infty$$

$$\mathbb{E}\left[\widehat{\mathbf{g}}^{\mathsf{T}}\widehat{\mathbf{g}}\right] = \mathbb{E}\left\|\widehat{\mathbf{g}}\right\|^2 \leq K_1 + K_2 \left\|\mathbf{g}\right\|^2 = K_1 + K_2 \mathbf{g}^{\mathsf{T}}\mathbf{g}\,, K_1, K_2 > 0$$

- **Variance reduction:**
  * Increase $\mathrm{N_{ens}}$
  * Use antithetic variates and/or importance sampling
  * **Introduce baseline to the objective**

Improving the Stochastic Optimization Procedure II

- **Policy gradient with baseline:**

  the objective function Υ can be replaced with the following baseline version:

$$\Upsilon^{\mathrm{b}}(\theta) := \mathbb{E}_{\zeta\sim\mathbb{P}(\zeta|\theta)}\Big[\mathcal{J}(\zeta) - b\Big] \,,$$

- By applying the kernel trick again

$$\nabla_\theta \Upsilon^{\mathrm{b}}(\theta) := \mathbf{g} = \mathbb{E}_{\zeta\sim\mathbb{P}(\zeta|\theta)}\Big[\left(\mathcal{J}(\zeta) - b\right) \nabla_\theta \log \mathbb{P}\left(\zeta|\theta\right)\Big]$$

$$\approx \widehat{\mathbf{g}}^{\mathrm{b}} := \frac{1}{\mathrm{N_{ens}}} \sum_{j=1}^{\mathrm{N_{ens}}} \left(\mathcal{J}(\zeta[j]) - b\right) \nabla_\theta \log \mathbb{P}\left(\zeta[j]|\theta\right)$$

Note that: $\Upsilon^{\mathrm{b}}(\theta) = \mathbb{E}_{\zeta\sim\mathbb{P}(\zeta|\theta)}\Big[\mathcal{J}(\zeta)\Big] - b \quad \Rightarrow \quad \begin{cases} \arg\min_\theta \Upsilon^{\mathrm{b}} = \arg\min_\theta \Upsilon \,, \\ \nabla_\theta \Upsilon^{\mathrm{b}} = \nabla_\theta \Upsilon - \nabla_\theta b = \nabla_\theta \Upsilon \end{cases}$

How do we choose b?

Improving the Stochastic Optimization Procedure III

Lemma 4.2: Statistics of the gradient estimator (with baseline)

Let $\mathbf{d} = \frac{1}{\mathrm{N_{ens}}} \sum_{j=1}^{\mathrm{N_{ens}}} \nabla_\theta \log \mathbb{P}\left(\zeta[j] | \theta\right)$, then

$$\mathbb{E}\left[\widehat{\mathbf{g}}^{\mathrm{b}}\right] = \mathbf{g} = \nabla_\theta \Upsilon(\theta)\,;$$

$$\mathsf{var}\left(\widehat{\mathbf{g}}^{\mathrm{b}}\right) = \mathsf{var}\left(\widehat{\mathbf{g}}\right) - 2b\mathbb{E}\left[\widehat{\mathbf{g}}^{\mathsf{T}}\mathbf{d}\right] + \frac{b^2}{\mathrm{N_{ens}}} \sum_{i=1}^{n_{\mathrm{s}}} \frac{1}{\theta_i - \theta_i^2}$$

The expression of $\mathsf{var}\left(\widehat{\mathbf{g}}^{\mathrm{b}}\right)$ is quadratic in b and can be minimized over b:

$$\begin{aligned} b^{\mathrm{opt}} &= \frac{\mathrm{N_{ens}}}{\sum_{i=1}^{n_{\mathrm{s}}} \frac{1}{\theta_i - \theta_i^2}} \mathbb{E}\left[\widehat{\mathbf{g}}^{\mathsf{T}}\mathbf{d}\right] \\ &\approx \frac{\mathrm{N_{ens}}}{\sum_{i=1}^{n_{\mathrm{s}}} \frac{1}{\theta_i - \theta_i^2}} \frac{1}{b_m} \sum_{e=1}^{b_m} \widehat{\mathbf{g}}[e]^{\mathsf{T}}\mathbf{d}[e]\,, \end{aligned}$$

where $\mathbf{g}[e]$ and $\mathbf{d}[e]$ are realizations of $\mathbf{g}$ and $\mathbf{d}$, respectively.

Improving the Stochastic Optimization Procedure IV

The following function is then used to estimate the optimal baseline b^{opt}

$$b^{\mathrm{opt}} \approx \widehat{b}^{\mathrm{opt}} := \frac{\sum_{e=1}^{b_m} \left(\sum_{j=1}^{\mathrm{N_{ens}}} \mathcal{J}(\zeta[e,j]) \nabla_{\theta} \log \mathbb{P}\left(\zeta[e,j]|\theta\right) \right)^{\mathsf{T}} \left(\sum_{j=1}^{\mathrm{N_{ens}}} \nabla_{\theta} \log \mathbb{P}\left(\zeta[e,j]|\theta\right) \right)}{\mathrm{N_{ens}}\, b_m \sum_{i=1}^{n_{\mathrm{s}}} \frac{1}{\theta_i - \theta_i^2}}$$

Algorithm 2 Stochastic optimization for binary OED with the optimal baseline.

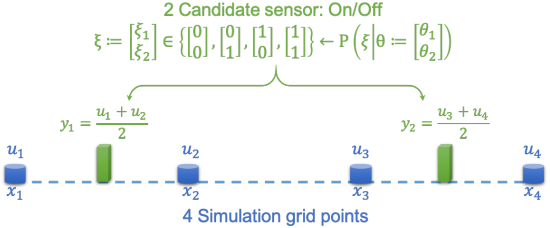
Input: Initial distribution parameter $\theta^{(0)}$, step size schedule $\eta^{(n)}$, sample sizes $\mathrm{N}_{\rm ens}$, m, baseline batch size b_m

Output: $\zeta^{\rm opt}$

initialize $n = 0$
while Not Converged **do**
    Update $n \leftarrow n + 1$
    Sample $\{\zeta[j]; j = 1, 2, \ldots, \mathrm{N}_{\rm ens}\} \sim \mathbb{P}\left(\zeta|\theta^{(n)}\right)$
    Calculate $b = \textsc{OptimalBaseline}(\theta^{(n)},\, \mathrm{N}_{\rm ens},\, b_m)$
    Calculate $\mathbf{g}^{(n)} = \frac{1}{\mathrm{N}_{\rm ens}} \sum_{j=1}^{\mathrm{N}_{\rm ens}} \left(\mathcal{J}(\zeta[j] - b)\right) \sum_{i=1}^{n_s} \left(\frac{\zeta_i[j]}{\theta_i} + \frac{\zeta[j]_i - 1}{1-\theta_i} \right) \mathbf{e}_i$
    Update $\theta^{(n+1)} = \mathbf{L}\left(\theta^{(n)} - \eta^{(n)} g^{(n)}\right)$
end while
Set $\theta^{\rm opt} = \theta^{(n)}$
Sample $\{\zeta[j]; j = 1, 2, \ldots, m\} \sim \mathbb{P}\left(\zeta|\theta^{\rm opt}\right)$, and calculate $\mathcal{J}(\zeta[j])$
return $\zeta^{\rm opt}$: the design ζ with smallest value of $\mathcal{J}$ in the sample.

function $\textsc{OptimalBaseline}(\theta,\, \mathrm{N}_{\rm ens},\, b_m)$
    Initialize $b \leftarrow 0$
    for $e \leftarrow 1$ to b_m **do**
        for $j \leftarrow 1$ to $\mathrm{N}_{\rm ens}$ **do**
            Sample $\zeta[j] \sim \mathbb{P}\left(\zeta|\theta\right)$
            Calculate $\mathbf{r}[j] = \sum_{i=1}^{n_s} \left(\frac{\zeta_i[j]}{\theta_i} + \frac{\zeta[j]_i - 1}{1-\theta_i} \right) \mathbf{e}_i$
        end for
        Calculate $\mathbf{d}[e] = \frac{1}{\mathrm{N}_{\rm ens}} \sum_{j=1}^{\mathrm{N}_{\rm ens}} \mathbf{r}[j]$
        Calculate $\mathbf{g}[e] = \frac{1}{\mathrm{N}_{\rm ens}} \sum_{j=1}^{\mathrm{N}_{\rm ens}} \mathcal{J}(\zeta[j])\, \mathbf{r}[j]$
        Update $b \leftarrow b + \left(\mathbf{g}[e]\right)^{\mathsf{T}} \mathbf{d}[e]$
    end for
    Update $b \leftarrow b \times \frac{\mathrm{N}_{\rm ens}}{b_m \sum_{i=1}^{n_s} \frac{1}{\theta_i - \theta_i^2}}$
    return b
end function

Numerical results: 2D (2 candidate sensors) example I



$$\mathbf{F} := \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

$$\mathbf{\Gamma}_{\rm pr} := \text{diag}\,(4, 1, 0.25, 1)$$

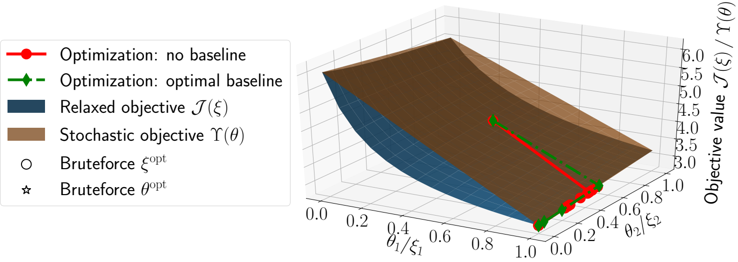
$$\mathbf{\Gamma}_{\rm noise} := \text{diag}\,(0.25, 1)$$

$$\Rightarrow \mathcal{J}(\zeta) = \mathrm{Tr}\left(\begin{bmatrix} \zeta_1 + 0.25 & \zeta_1 & 0 & 0 \\ \zeta_1 & \zeta_1 + 1 & 0 & 0 \\ 0 & 0 & 0.25\zeta_2 + 4 & 0.25\zeta_2 \\ 0 & 0 & 0.25\zeta_2 & 0.25\zeta_2 + 1 \end{bmatrix}^{-1} \right)$$

$$\Upsilon(\theta) = \Upsilon(\theta) = \sum_{k=1}^{4} \mathcal{J}(\zeta[k])\mathbb{P}\left(\zeta[k]|\theta\right); \quad \nabla_\theta \Upsilon(\theta) = \sum_{k=1}^{4} \mathcal{J}(\zeta[k])\nabla_\theta \mathbb{P}\left(\zeta[k]|\theta\right),$$

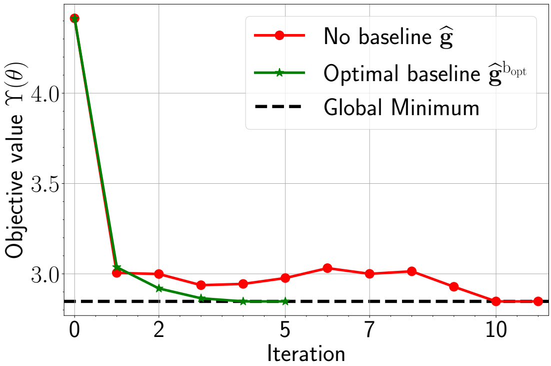
$$\nabla_\theta \mathbb{P}\left(\zeta|\theta\right) = \begin{bmatrix} (-1)^{1-\zeta_1}\theta_2(1-\theta_2) \\ (-1)^{1-\zeta_2}\theta_1(1-\theta_1) \end{bmatrix}, \quad \zeta \in \left\{ \zeta[1] = \begin{bmatrix}0\\0\end{bmatrix}, \zeta[2] = \begin{bmatrix}0\\1\end{bmatrix}, \zeta[3] = \begin{bmatrix}1\\0\end{bmatrix}, \zeta[4] = \begin{bmatrix}1\\1\end{bmatrix} \right\}$$

Numerical results: 2D (2 candidate sensors) example II



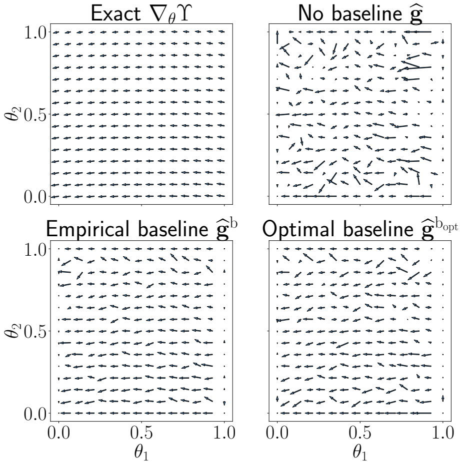
Surface plot of the objective function $\mathcal{J}$ of the relaxed OED problem and the objective function Υ of the corresponding stochastic OED problem.

Numerical results: 2D (2 candidate sensors) example III



Value of the objective function Υ evaluated at each iteration of the algorithm until convergence. The initial parameter $\theta^{(0)}$ of the optimizer is set to $(0.5, 0.5)^{\mathsf{T}}$, and the algorithm terminates when the magnitude of the projected gradient (pgtol) is lower than 10^{-8}.

Numerical results: 2D (2 candidate sensors) example IV



Exact and stochastic approximation of the gradient at multiple realizations of θ. In addition to the optimal estimate, shown are results of an empirical baseline $b := \frac{\mathcal{J}(0)+\mathcal{J}(1)}{2}$ (without guarantees!).

Numerical results: Advection-Diffusion I

- Numerical model (AD): u solves:

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

  - * $\Omega \in \mathbb{R}^2$ is an open and bounded domain
  - * u the concentration of a contaminant in Ω
  - * κ is the diffusivity,
  - * $\mathbf{v}$ is the velocity field

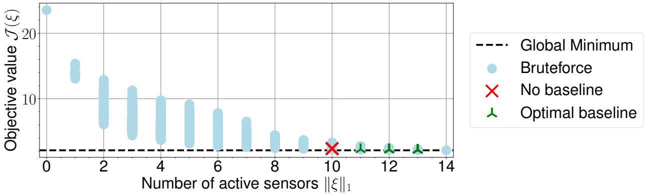
- Observational setup: $n_s = 14$ candidate sensor locations; $2^{14} = 16384$ possible designs,
  - * observation times $t_k := t_1 + s\Delta t$; $\Delta t = 0.2$ is the model simulation time step;
  - * $t_1 = 1,\ s = 0, 1, \ldots, 20$
  - * Observation correlations; synthetic, created with Gaspari-Cohn, and 5% noise level

- The stochastic optimization problem:

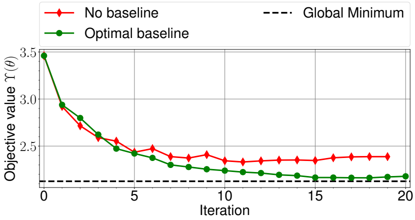
$$
\theta^{\text{opt}} = \underset{\theta \in [0,1]^{n_s}}{\arg\min}\, \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\theta)} \Big[\underbrace{\mathrm{Tr}\left(\left(\mathbf{M}^{-1} \mathbf{F}^{\mathsf{T}} \mathbf{\Gamma}_{\text{noise}}^{-1/2} \mathrm{diag}\,(\zeta)\, \mathbf{\Gamma}_{\text{noise}}^{-1/2} \mathbf{F} + \mathbf{\Gamma}_{\text{pr}}^{-1} \right)^{-1} \right)}_{\mathcal{J}(\zeta)} + \alpha\, \Phi(\zeta) \Big]
$$

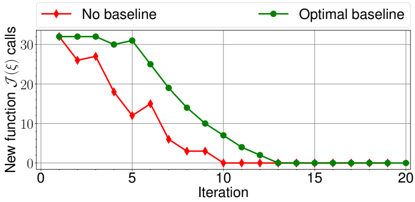
Numerical results: Advection-Diffusion II

Dropping the penalty term: $\alpha = 0$



Results of the policy gradient procedures compared with the brute-force search of all candidate binary designs.



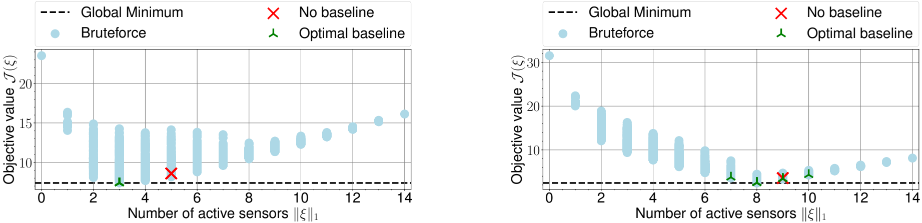


Behavior of the policy gradient procedures over consecutive iterations. Left: the value of the objective Υ at each iteration of the optimization procedures. Right: number of new function $\mathcal{J}$ evaluations carried out.

Numerical results: Advection-Diffusion III

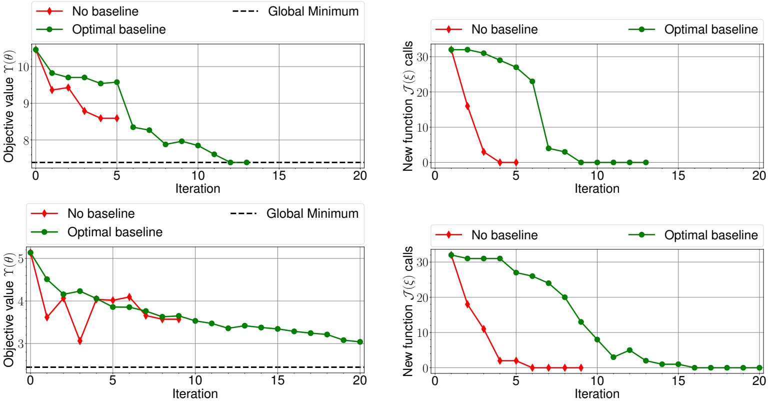
Results with sparsity and fixed-budget constraints:

- $\alpha = 1,\ \Phi(\zeta) := \|\zeta\|_0$
- $\alpha = 1,\ \Phi(\zeta) := \left|\|\zeta\|_0 - \lambda\right|$ for a budget λ



Left: $\Phi(\zeta) := \|\zeta\|_0$. Right: $\Phi(\zeta) := \|\zeta - \lambda\|_0$, where $\lambda = 8$. In both cases, we set the sparsity penalty parameter to $\alpha = 1.0$.

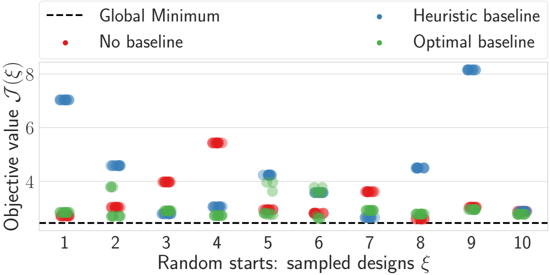
Numerical results: Advection-Diffusion IV

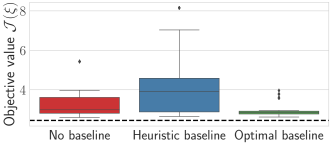


Top: $\Phi(\zeta) := \|\zeta\|_0$. Bottom: $\Phi(\zeta) := \|\zeta - \lambda\|_0$, where $\lambda = 8$. In both cases, we set the sparsity penalty parameter to $\alpha = 1.0$.

Numerical results: Advection-Diffusion V

Initial policy and effect of the baseline





The optimization procedures are run 10 times from different (random) initial policies, and 10 designs are sampled from the policy returned by each algorithm in each experiment.

Concluding Remarks and References

1. Ahmed Attia, and Emil Constantinescu. "Optimal Experimental Design for Inverse Problems in the Presence of Observation Correlations." arXiv preprint arXiv:2007.14476 (2020).
   - → Properly accommodate observation correlations in **relaxed OED problems**
   - → D-optimality extension is straight forward, however it is computationally expensive

2. Ahmed Attia, Sven Leyffer, and Todd Munson. "Stochastic Learning Approach to Binary Optimization for Optimal Design of Experiments." arXiv preprint arXiv:2101.05958 (2021).
   - → Solve the **binary OED optimization without relaxation**
   - → Convert a binary design domain into a bounded continuous domain, where the optimal solution of the two problems coincide
   - → The stochastic formulation enables utilizing efficient stochastic optimization algorithms to solve binary optimization problems, e.g., SAA, etc.
   - → The solution of the stochastic OED problem is an optimal parameter $\theta^{\rm opt}$ that can be used for sampling binary designs ζ by sampling $\mathbb{P}\left(\zeta|\theta^{\rm opt}\right)$, even if only a suboptimal solution is found
   - → Nonsmooth penalty functions Φ can be utilized in defining $\mathcal{J}$, for example to enforce sparsity or budget constraints;

- These slides are available
  https://www.mcs.anl.gov/~attia/conferences.html
- Questions are welcome
  attia@mcs.anl.gov

U.S. DEPARTMENT OF ENERGY | Office of Science