Bayesian Optimal Experimental Design

Luis Espath

Joint with J Beck, AG Carlon, BM Dia & R Tempone

—Department of Mathematics—RWTH Aachen University—

July 16, 2020

Outline—Section §1

Optimal Experimental Design (OED) I

This presentation is based on [1, 2, 3].

- Find the design that maximizes the success of an experiment for some desired experimental goal
- Predict the outcome **without** performing the real experiment
- Model-based approach: Important when resources are limited

Data model assumption:

$$\boldsymbol{Y}(\boldsymbol{\xi}) = \boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{\xi}) + \boldsymbol{\epsilon},$$

- $\boldsymbol{Y} = (y_1, y_2, \ldots, y_q) \in \mathbb{R}^q$, measurements
- $\boldsymbol{\epsilon} \in \mathbb{R}^q$, measurement error
- $\boldsymbol{g} \in \mathbb{R}^q$, model predictions (usually, the solution of a math. model)
- $\boldsymbol{\theta} \in \mathbb{R}^d$, quantities of interest
- $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_s)$, design parameters

Measurement error assumption: The measurement errors $\boldsymbol{\epsilon}$ are zero-mean Gaussian errors with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$

Optimal Experimental Design (OED) II

Data model with repetitive experiments:

$$\boldsymbol{y}_i(\boldsymbol{\xi}) = \boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{\xi}) + \boldsymbol{\epsilon}_i, \ i = 1, \ldots, N_e$$

where N_e is the number of repetitive experiments with design $\boldsymbol{\xi}$.

The full set of data: $\boldsymbol{Y} = \{\boldsymbol{y}_i\}_{i=1}^{N_e}$.

Simple OED example: a linear regression model

Linear regression model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

where

- $\boldsymbol{y}$, vector of observation values
- $\boldsymbol{X}$ is the design matrix
- $\boldsymbol{\epsilon}$ is the error (zero-mean) vector with known covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$
- $\boldsymbol{\theta}$ is an unknown parameter vector of interest.

Inverse problem: Estimate $\boldsymbol{\theta}$

Classical approach:

- Least square estimate ($\arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\epsilon}\|^2$): $\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$
- $\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-\top}$

OED goal: Find design $\boldsymbol{X}$ that minimize $\mathrm{Cov}(\hat{\boldsymbol{\theta}})$

Alphabetic optimalities

What does it mean to minimize $\mathrm{Cov}(\hat{\boldsymbol{\theta}}) := \boldsymbol{\Lambda}$?

- $\boldsymbol{A}$-optimality: minimize the trace of the covariance matrix, $\mathrm{tr}\,(\boldsymbol{\Lambda})$
- $\boldsymbol{C}$-optimality: minimize the variance of a predefined linear combination of quantities of interest: $\boldsymbol{c} \cdot \boldsymbol{\Lambda} \boldsymbol{c}$
- $\boldsymbol{D}$-optimality: minimize the determinant of the covariance matrix, $\det\,(\boldsymbol{\Lambda})$
- $\boldsymbol{E}$-optimality: minimize the maximum eigenvalue of the covariance matrix $\boldsymbol{\Lambda}$

For nonlinear models the problem is much more challenging!

Entropy based expected information gain in a Bayesian setting:

Lindley, D. V. (1956). On a measure of the information provided by an experiment. The Annals of Mathematical Statistics, 27(4), 986-1005.

Entropy-based OED criterion

1948, Information Entropy Definition: Claude Shannon introduced the concept of information entropy:

$$\text{Entropy} = \sum_i p_i \log \frac{1}{p_i},$$

for probabilities p_i. The differential entropy for a continuous random variable with *probability density function* (pdf) $p(\boldsymbol{x})$ with support $\mathcal{X}$,

$$\text{Differential Entropy} = \int_{\mathcal{X}} \log \frac{1}{p(\boldsymbol{x})} p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}.$$

1956, Shannon Entropy-based OED: Dennis V. Lindley proposed an entropy-based measure of the information provided in an experiment: the relative entropy of the posterior pdf $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ with respect to the prior pdf $\pi(\boldsymbol{\theta})$:

$$\text{Relative Entropy} = \underbrace{\int \log \frac{1}{\pi(\boldsymbol{\theta})} \pi(\boldsymbol{\theta}|\boldsymbol{Y})\, \mathrm{d}\boldsymbol{\theta}}_{=\text{Cross Entropy}} - \underbrace{\int \log \frac{1}{\pi(\boldsymbol{\theta}|\boldsymbol{Y})} \pi(\boldsymbol{\theta}|\boldsymbol{Y})\, \mathrm{d}\boldsymbol{\theta}}_{=\text{Differential Entropy}},$$

in expectation.

Information gain

The **information gain** (Lindley, 1956) is here defined by the Kullback-Leibler (KL) divergence between posterior and prior pdfs

$$D_{\mathrm{KL}}(\pi(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi}) \parallel \pi(\boldsymbol{\theta})) := \int_{\Theta} \log\left(\frac{\pi(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi})}{\pi(\boldsymbol{\theta})}\right) \pi(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\theta},$$

where

- $\pi(\boldsymbol{\theta})$, the prior pdf for $\boldsymbol{\theta}$
- $\pi(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi})$, the posterior pdf for $\boldsymbol{\theta}$ given $\boldsymbol{Y}$ for design $\boldsymbol{\xi}$

How to obtain the posterior pdf? Use *Bayes' rule*:

$$\pi(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi}) = \frac{p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})}{p(\boldsymbol{Y}|\boldsymbol{\xi})}\pi(\boldsymbol{\theta})$$

where

- $p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})$, the likelihood function for $\boldsymbol{Y}$
- $p(\boldsymbol{Y}|\boldsymbol{\xi})$, the evidence, i.e., $p(\boldsymbol{Y}|\boldsymbol{\xi}) = \int_{\Theta} p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}$

Expected Information Gain (EIG) criterion

Information gain:

$$D_{\mathrm{KL}}(\pi(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi}) \parallel \pi(\boldsymbol{\theta})) = \int_{\Theta} \log\left(\frac{p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})}{p(\boldsymbol{Y}|\boldsymbol{\xi})}\right) \frac{p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})\pi(\boldsymbol{\theta})}{p(\boldsymbol{Y}|\boldsymbol{\xi})} \,\mathrm{d}\boldsymbol{\theta}.$$

The information about $\boldsymbol{Y}$ is given by the data model.

The **expected information gain (EIG)** is given by

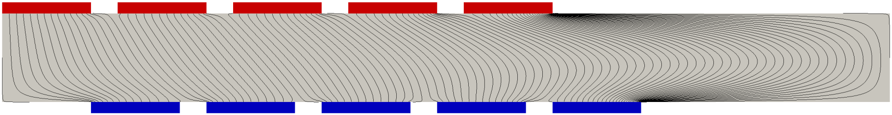
$$\begin{aligned} I(\boldsymbol{\xi}) &:= \int D_{\mathrm{KL}}(\pi(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi}) \parallel \pi(\boldsymbol{\theta}))\, p(\boldsymbol{Y})\, d\boldsymbol{Y} \\ &= \int\int \log\left(\frac{p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})}{\int_{\Theta} p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}}\right) p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})\, d\boldsymbol{Y}\, \pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}, \\ &= \mathbb{E}\left[\mathbb{E}\left[\log\left(\frac{p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})}{\mathbb{E}\left[p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})\right]}\right)\right]\right], \end{aligned}$$

where, because of the assumed data model, the likelihood follows a multivariate Normal distribution

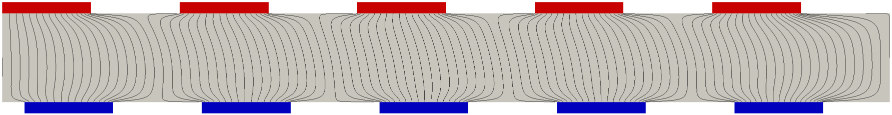
$$p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi}) = \det\left(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\right)^{-\frac{N_e}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{N_e} \|\boldsymbol{y}_i - \boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{\xi})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1}}\right).$$

Electrical Impedance Tomography Example I

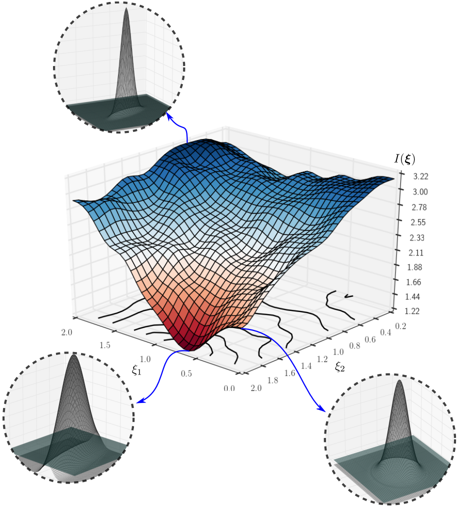
Best experiment



Worst experiment



Electrical Impedance Tomography Example II



Estimation EIG

The expected information can be computed separately for each $\boldsymbol{\xi}$.

EIG criterion:

$$I = \mathbb{E}\left[\mathbb{E}\left[\log\left(\frac{p(\boldsymbol{Y}|\boldsymbol{\theta})}{\mathbb{E}\left[p(\boldsymbol{Y}|\boldsymbol{\theta})\right]}\right)\right]\right],$$

where the likelihood is given by the *data model*:

$$p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi}) = \det\left(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\right)^{-\frac{N_e}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{N_e}\|\boldsymbol{y}_i - \boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{\xi})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1}}\right).$$

Conventional choice: Double-loop Monte Carlo sampling

Double-loop Monte Carlo (DLMC) for estimating EIG

The DLMC estimator for EIG:

$$\mathcal{I}_{\mathrm{dl}} := \frac{1}{N}\sum_{n=1}^{N}\log\left(\frac{p(\boldsymbol{Y}_n|\boldsymbol{\theta}_n,\boldsymbol{\xi})}{\frac{1}{M}\sum_{m=1}^{M}p(\boldsymbol{Y}_n|\boldsymbol{\theta}_{n,m},\boldsymbol{\xi})}\right)$$

where $\boldsymbol{\theta}_n, \boldsymbol{\theta}_{n,m} \overset{\text{i.i.d.}}{\sim} \pi(\boldsymbol{\theta})$ and $\boldsymbol{Y}_n \overset{\text{i.i.d.}}{\sim} p(\boldsymbol{Y}|\boldsymbol{\theta}_n,\boldsymbol{\xi})$.

Average work model: $\mathcal{W} = N \times M$ likelihood evaluations!

Bias and variance can be modeled (Ryan, 2003):

Bias: $|\mathbb{E}\left[\mathcal{I}_{\mathrm{dl}}\right] - I| \approx \frac{C_1}{M}$, for some constant $C_1 > 0$ (Not unbiased!)

Variance: $\mathbb{V}[\mathcal{I}_{\mathrm{dl}}] \approx \frac{C_2}{N} + \frac{C_3}{NM}$, for some constants $C_2, C_3 > 0$.

How to control the estimator error?

Control the accuracy: Choose N and M to control the estimator's accuracy

Goal: For a specified error tolerance, $\mathrm{TOL} > 0$, at a confidence level given by $0 < \alpha \ll 1$, we select the values of N and M minimizing the computational work, while satisfying

$$P(|\mathcal{I}_{\mathrm{dl}}(N, M) - I| \leq \mathrm{TOL}) \geq 1 - \alpha.$$

Total Error = Bias + Statistical Error

Split the total error of a random estimator $\mathcal{I}$ into a bias component and a statistical error component

$$|I - \mathcal{I}| \leq |I - \mathbb{E}[\mathcal{I}]| + |\mathbb{E}[\mathcal{I}] - \mathcal{I}|.$$

Optimization problem: minimize the estimator's work, $\mathcal{W}$, subject to

$$|I - \mathbb{E}[\mathcal{I}]| \leq (1-\kappa)\mathrm{TOL}$$

$$|\mathbb{E}[\mathcal{I}] - \mathcal{I}| \leq \kappa\mathrm{TOL}$$

for $\kappa \in (0,1)$, and the latter should hold with probability $1-\alpha$.
By using the **CLT** theorem, the statistical error constraint is replaced by

$$C_\alpha \sqrt{\mathbb{V}[\mathcal{I}]} \leq \kappa\mathrm{TOL}$$

for $C_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$ where Φ is the cumulative probability distribution (CDF) of a standard, normal random variable.

Optimal choice of DLMC parameters

Optimization problem for DLMC:

$$(N^*, M^*, \kappa^*) = \underset{(N,M,\kappa)}{\arg\min}\, N \times M \quad \text{subject to} \quad \begin{cases} \frac{C_{dl,3}}{M} \leq (1-\kappa)\mathrm{TOL} \\ \frac{C_{dl,1}}{N} + \frac{C_{dl,2}}{NM} \leq (\kappa/C_\alpha)^2\, \mathrm{TOL}^2 \end{cases}$$

Optimal choice of number of samples (Beck et al. [1]):

$$N^* = \left\lceil \frac{C_\alpha^2}{2\kappa^*} \frac{C_{dl,1}}{1-\kappa^*} \mathrm{TOL}^{-2} \right\rceil$$

$$M^* = \left\lceil 0.5 \frac{C_{dl,2}}{1-\kappa^*} \mathrm{TOL}^{-1} \right\rceil$$

Optimal average work: For a specified TOL with the specified probability of success,

$$\text{Optimal average work} = N^* M^* \propto \mathrm{TOL}^{-3},$$

as $\mathrm{TOL} \to 0$.

Improving the constant, Long et at. [4]: Laplace-based importance sampling I

Assumption: The posterior pdf of $\boldsymbol{\theta}$ can be well approximated by a Gaussian.

Laplace approximation: Approximate the posterior pdf, $\pi(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi})$, by a multivariate normal pdf, $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi})$, following $\mathcal{N}\left(\hat{\boldsymbol{\theta}}(\boldsymbol{Y},\boldsymbol{\xi}),\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}(\boldsymbol{Y},\boldsymbol{\xi}))\right)$, i.e.,

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi}) = (2\pi)^{-\frac{d}{2}}\det(\hat{\boldsymbol{\Sigma}})^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\|^2_{\hat{\boldsymbol{\Sigma}}^{-1}}\right),$$

with *maximum a posteriori* (MAP) estimate,

$$\hat{\boldsymbol{\theta}}(\boldsymbol{Y},\boldsymbol{\xi}) \coloneqq \arg\max_{\boldsymbol{\theta}\in\Theta}\log\pi(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\xi}) = \arg\max_{\boldsymbol{\theta}\in\Theta}\log(p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\xi})p(\boldsymbol{\theta}))$$

$$= \arg\max_{\boldsymbol{\theta}\in\Theta}\left[-\frac{1}{2}\sum_{i=1}^{N_e}\|\boldsymbol{y}_i-\boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{\xi})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1}}+\log(\pi(\boldsymbol{\theta}))\right],$$

This optimization problem can typically be solved sufficiently with $\sim$ 20-40 extra solves per outer sample (indexed by n).

Improving the constant, Long et at. [4]: Laplace-based importance sampling II

and with a second-order Taylor expansion, we have that $\hat{\boldsymbol{\Sigma}}^{-1} = -\nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}} \log \pi(\boldsymbol{\theta}|\boldsymbol{Y}\boldsymbol{\xi})$,

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}) &= \nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}\boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{\xi}) \cdot \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}{}^{-1} \sum_{i=1}^{N_e} (\boldsymbol{y}_i - \boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{\xi})) \\ &+ N_e \nabla_{\boldsymbol{\theta}}\boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{\xi}) \cdot \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}{}^{-1} \nabla_{\boldsymbol{\theta}}\boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{\xi}) - \nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}} \log(\pi(\boldsymbol{\theta})) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{N_e}}\right).\end{aligned}$$

Where the leading terms can be obatined by dropping the first term $\mathcal{O}_{\mathbb{P}}\left(\frac{1}{N_e}\right)$, thus, arriving at

$$\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\theta},\boldsymbol{\xi}) = N_e \nabla_{\boldsymbol{\theta}}\boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{\xi}) \cdot \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}{}^{-1} \nabla_{\boldsymbol{\theta}}\boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{\xi}) - \nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}} \log(\pi(\boldsymbol{\theta})) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{N_e}}\right).$$

DLMC with Laplace-based Importance Sampling (DLMCIS)

Laplace-based importance sampling for the inner expectation:

$$p(\boldsymbol{Y}) = \int p(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = \int \frac{p(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{Y})}\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{Y})\,\mathrm{d}\boldsymbol{\theta}$$

DLMCIS estimator *(Ryan et al., 2015; Beck et al. [1])*:

$$\mathcal{I}_{\mathrm{dlis}} := \frac{1}{N}\sum_{n=1}^{N}\log\left(\frac{p(\boldsymbol{Y}|\boldsymbol{\theta}_n)}{\frac{1}{M}\sum_{m=1}^{M}\frac{p(\boldsymbol{Y}_n|\tilde{\boldsymbol{\theta}}_{n,m})\pi(\tilde{\boldsymbol{\theta}}_{n,m})}{\tilde{\pi}(\tilde{\boldsymbol{\theta}}_{n,m}|\boldsymbol{Y}_n)}}\right),$$

where $\tilde{\boldsymbol{\theta}}_{n,m} \overset{\text{i.i.d.}}{\sim} \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{Y}_n)$. (**no additional bias!**)

Optimal average work: $\propto \mathrm{TOL}^{-3}$ as $\mathrm{TOL} \to 0$.

The *proportionality constant* is substantially smaller for DLMCIS than for DLMC ! (**variance reduction**)

Numerical demonstration

Nonlinear scalar model:

$$y_i(\xi) = \theta^3\xi^2 + \theta\exp(-|0.2-\xi|) + \epsilon_i, \quad \text{for } i = 1, 2, \ldots, N_e,$$

with

- $\theta \sim \mathcal{U}(0,1)$
- $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 10^{-3})$
- $N_e = 1$ <u>and</u> $N_e = 10$

Expected information gain

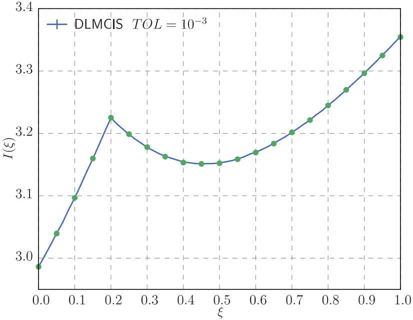


Figure: Expected information gain

Computational work for a single design

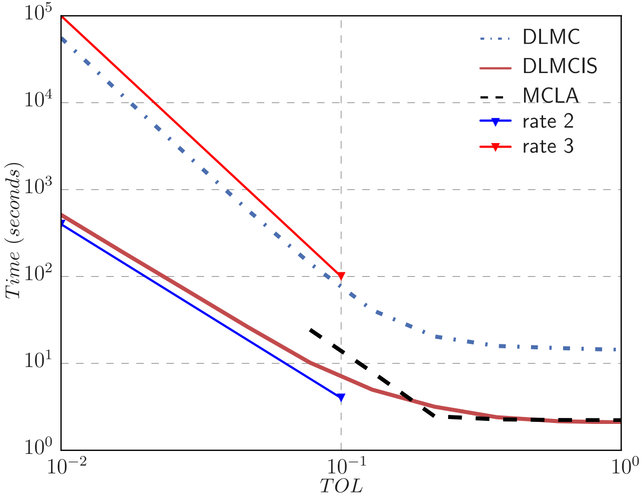


Figure: Average running time vs. TOL

Optimal estimator parameter choices

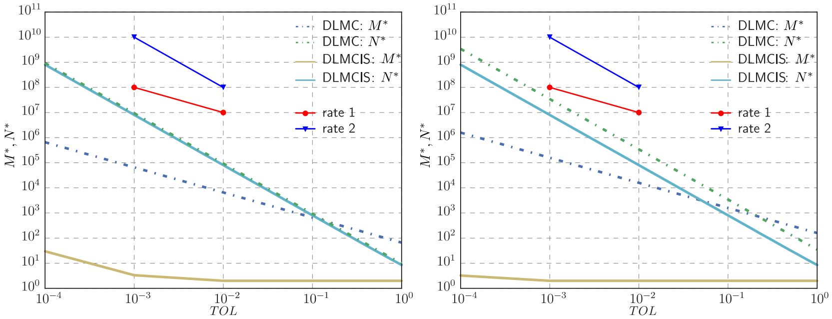


Figure: Left: $N_e = 1$, Right: $N_e = 10$

Consistency between TOL and Error

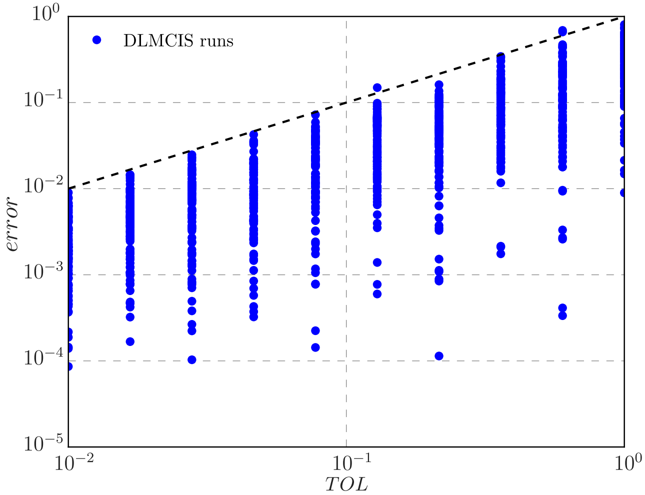


Figure: Error vs. TOL

Outline—Section §2

Stochastic optimization—Introduction

Find $\boldsymbol{\theta}^*$ such that

$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi} \in \Theta}{\arg\min}\, \mathbb{E}[f(\boldsymbol{\xi}, \boldsymbol{\theta})].$$

Deterministic algorithm:

$$\boldsymbol{\xi}_{k+1} = \boldsymbol{\xi}_k + \alpha_k \nabla_{\boldsymbol{\xi}} \mathbb{E}[f(\boldsymbol{\xi}_k, \boldsymbol{\theta})], \quad k \geq 0.$$

Stochastic gradient algorithm:

$$\boldsymbol{\xi}_{k+1} = \boldsymbol{\xi}_k + \alpha_k \nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k, \boldsymbol{\theta}_i), \quad k \geq 0,$$

where $\boldsymbol{\theta}_i$ is a single random realization.
Through what follows, we want to derive an inexpensive and accurate estimator

$$\nabla_{\boldsymbol{\xi}} F(\boldsymbol{\xi}) = \nabla_{\boldsymbol{\xi}} \mathbb{E}[f(\boldsymbol{\xi}, \boldsymbol{\theta})].$$

Optimal Bayesian experimental design—Gradient DLMC and MCLA I

Gradient of the inner loop of DLMC

$$\mathcal{G}_{DLMC}(\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{Y}) \coloneqq \nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k, \boldsymbol{\theta}_i) = \nabla_{\boldsymbol{\xi}} \left(\log \left(\frac{p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\xi})}{\sum_{m=1}^{M} p(\boldsymbol{Y}|\boldsymbol{\theta}_m^*, \boldsymbol{\xi})} \right) \right).$$

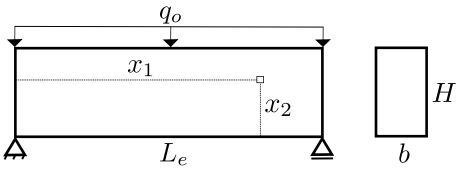
Gradient of the inner loop of MCLA

$$\mathcal{G}_{LA}(\boldsymbol{\xi}, \boldsymbol{\theta}) \coloneqq \nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k, \boldsymbol{\theta}_i) = -\frac{1}{2} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}) \colon \nabla_{\boldsymbol{\xi}} \boldsymbol{\Sigma}(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}) = -\sum_{k=1}^{d} \hat{\sigma}_k^{-1} \nabla_{\boldsymbol{\xi}} \hat{\sigma}_k.$$

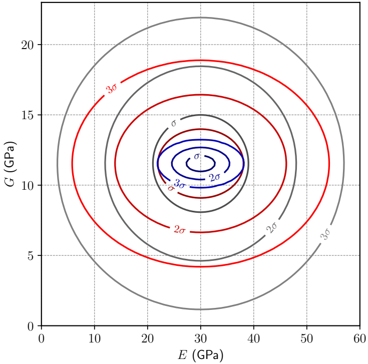
Carlon et al. [3]

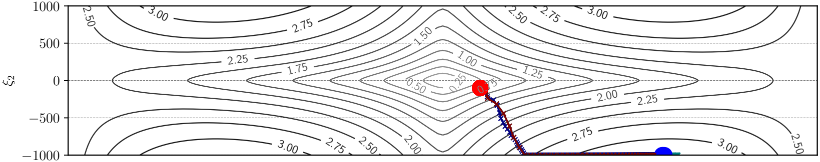
Stochastic optimization—Application to OED [3] I

Goal: Find the optimal position to attach a strain gauge in a Timoshenko beam (measuring strains ε_{11} and ε_{12}) to maximize the information about the Young and shear modulus, (E, G).

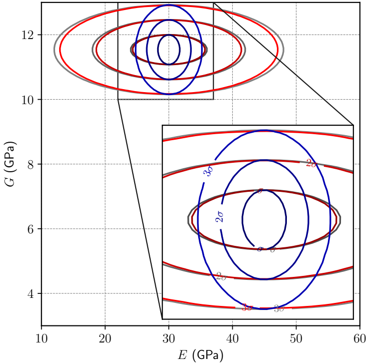


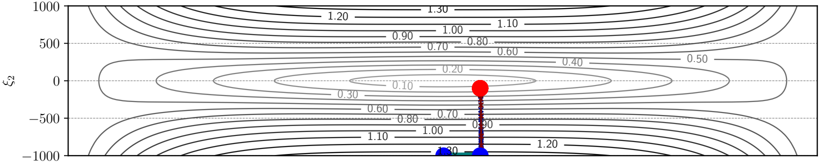
Stochastic optimization—Application to OED [3] II



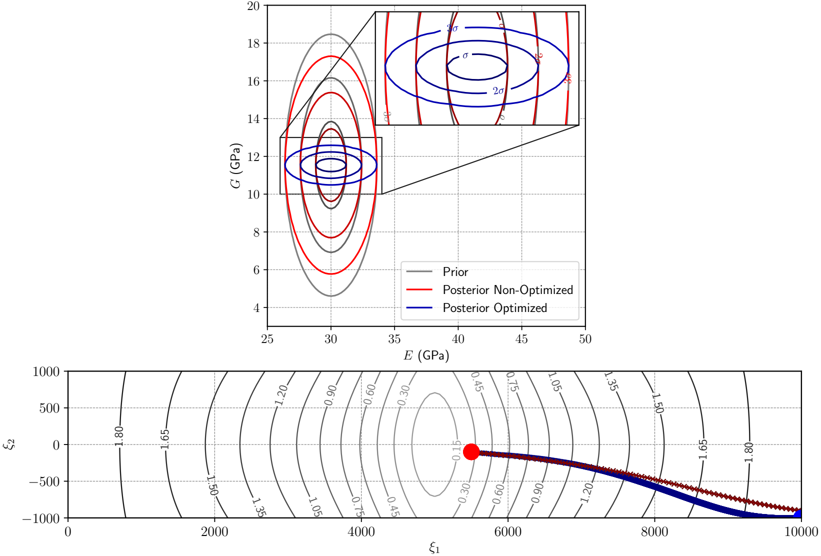


Stochastic optimization—Application to OED [3] III





Stochastic optimization—Application to OED [3] IV

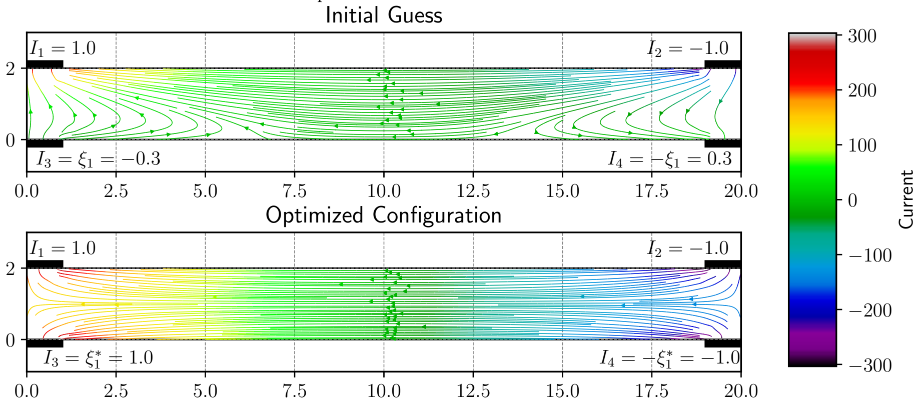


Stochastic optimization—Application II to OED I

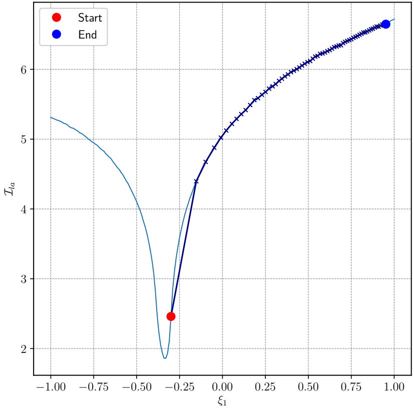
Goal: Find the optimal current pattern to be injected in a composite plate in order to recover the angle orientation of each ply.

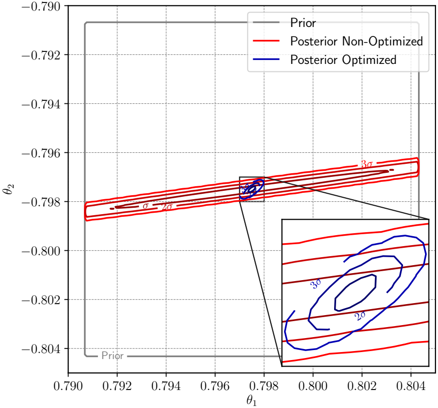
Stochastic optimization—Application II to OED II

Example 1: Four electrodes and one variable



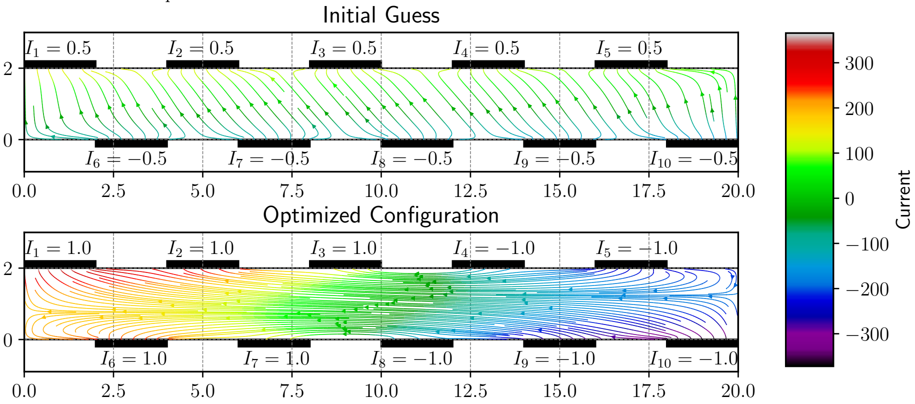
Stochastic optimization—Application II to OED III



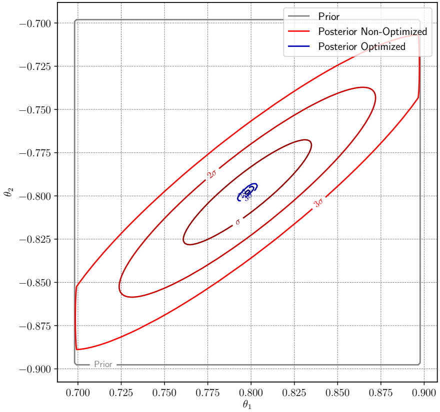


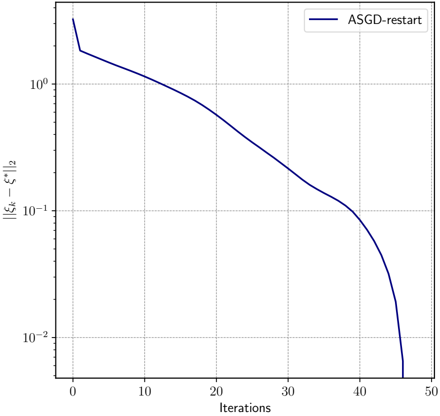
Stochastic optimization—Application II to OED IV

Example 3: Ten electrodes and nine variables Ten electrodes and nine variables



Stochastic optimization—Application II to OED V





Outline—Section §3

Joakim Beck, Ben Mansour Dia, Luis FR Espath, Quan Long, and Raul Tempone.
Fast bayesian experimental design: Laplace-based importance sampling for the expected information gain.
Computer Methods in Applied Mechanics and Engineering, 334:523–553, 2018.

Joakim Beck, Ben Mansour Dia, Luis FR Espath, and Raul Tempone.
Multilevel double loop monte carlo and stochastic collocation methods with importance sampling for bayesian optimal experimental design.
arXiv preprint arXiv:1811.11469, 2018.

Andre Gustavo Carlon, Ben Mansour Dia, Luis FR Espath, Rafael Holdorf Lopez, and Raul Tempone.
Nesterov-aided stochastic gradient methods using laplace approximation for bayesian design optimization.
arXiv preprint arXiv:1807.00653, 2019.

Quan Long, Marco Scavino, Raúl Tempone, and Suojin Wang.
Fast estimation of expected information gains for bayesian experimental designs based on laplace approximations.
Computer Methods in Applied Mechanics and Engineering, 259:24–39, 2013.