Efficient Bayesian Optimal Experimental Design for Physical Models

Quan Long

United Technologies Research Center
East Hartford, CT, USA

April 19, 2016
UTC Institute for Advanced Systems Engineering, University of Connecticut, Storrs, CT

Introduction

Introduction




- Experimental design is important when resources are limited.

Introduction

We first consider a linear regression model:

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

- The simple least square estimation: $\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$
- $\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Sigma} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}$

We want $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ to be as "small" as possible.

Introduction

Some alphabetic optimalities:

- **A**–optimality: minimize the trace of the covariance matrix $\mathbf{tr}(\boldsymbol{\Sigma})$
- **C**–optimality: minimize the variance of a predefined linear combination of parameters $(\beta^T \boldsymbol{\Sigma}^{-1} \beta)^{-1}$
- **D**–optimality: minimize the determinant of the covariance matrix $|\boldsymbol{\Sigma}|$
- **E**–optimality: minimize the maximum eigenvalue of the covariance matrix $\mathbf{max}(\sigma_{ii})$

Entropy based expected information gain in a Bayesian setting.

Computational Challenges in OED for nonlinear systems

- The sampler
- The optimizer
- The forward problem solver

Major Notations

- $p(\cdot)$: probability density function
- $\boldsymbol{\theta}$: unknown parameter vector
- $\boldsymbol{\theta}_0$: the d dimensional vector of the "true" parameters used to generate the synthetic data
- $\boldsymbol{\xi}$: the vector of control parameters, also known as the experimental setup
- $\boldsymbol{g}$: the deterministic model
- $\boldsymbol{y}_i$: the i^{th} observation vector
- $\bar{\boldsymbol{y}} = \{\boldsymbol{y}_i\}_{i=1}^{M}$: a set of observation vectors
- ϵ_i: the additive independent and identically distributed (i.i.d.) measurement noise

Bayesian framework for experimental design and expected information gain

- Prior of parameters: $p(\boldsymbol{\theta})$.
- Posterior (post experimental) of parameters by Bayes' theorem:

$$p(\boldsymbol{\theta}|\bar{\mathbf{y}}, \xi) = \frac{p(\bar{\mathbf{y}}|\boldsymbol{\theta}, \xi)p(\boldsymbol{\theta})}{p(\bar{\mathbf{y}})}\,.$$

- Kullback-Leibler divergence (information gain) between prior and posterior to measure the usefulness of an experiment

$$D_{KL} := \int_{\boldsymbol{\Theta}} \log\left(\frac{p(\boldsymbol{\theta}|\bar{\mathbf{y}}, \xi)}{p(\boldsymbol{\theta})}\right) p(\boldsymbol{\theta}|\bar{\mathbf{y}}, \xi) d\boldsymbol{\theta}\,.$$

(if $p(\boldsymbol{\theta}|\bar{\mathbf{y}}) = p(\boldsymbol{\theta})$, then $D_{KL} = 0.$)

- **Expected information gain** :

$$I(\xi) = \int D_{KL}\, p(\bar{\mathbf{y}}|\xi) d\bar{\mathbf{y}}\,.$$

Double–loop Monte Carlo

- The expected information gain can be rearranged as follows

$$I = \int_{\boldsymbol{\Theta}} \int_{\mathcal{Y}} \log\left(\frac{p(\bar{\mathbf{y}}|\boldsymbol{\theta})}{p(\bar{\mathbf{y}})}\right) p(\bar{\mathbf{y}}|\boldsymbol{\theta}) d\bar{\mathbf{y}} p(\boldsymbol{\theta}) d\boldsymbol{\theta} \,.$$

- This integral can be evaluated using Monte Carlo sampling.

$$I_{DLMC} = \frac{1}{N_o} \sum_{I=1}^{N_o} \log\left(\frac{p(\bar{\mathbf{y}}_I|\boldsymbol{\theta}_I)}{p(\bar{\mathbf{y}}_I)}\right) ,$$

where $\boldsymbol{\theta}_I$ is drawn from $p(\boldsymbol{\theta})$, $\bar{\mathbf{y}}_I$ is drawn from $p(\bar{\mathbf{y}}|\boldsymbol{\theta}_I)$. The so-called "double–loop" comes from the nested Monte Carlo to evaluate the marginal density

$$p(\bar{\mathbf{y}}_I) = \int_{\boldsymbol{\Theta}} p(\bar{\mathbf{y}}_I|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \approx \frac{1}{N_i} \sum_{J=1}^{N_i} p(\bar{\mathbf{y}}_I|\boldsymbol{\theta}_J) \,.$$

Double–loop Monte Carlo

We have the following estimates:

- $\mathbf{Bias}(I_{DLMC}) = \mathbf{E}(I_{DLMC} - I) = \mathcal{O}\left(\frac{1}{N_i}\right)$
- $\mathbf{Var}(I_{DLMC}) = \mathcal{O}\left(\frac{1}{N_o}\right)$
- To control the MSE, enforcing $\mathbf{Var}(I_{DLMC}) + \mathbf{Bias}(I_{DLMC})^2 = tol^2$
- To achieve tolerance *tol*, the total work is $N_o \times N_i = \mathcal{O}\left(tol^{-3}\right)$

Laplace method and generalized Laplace method

Laplace approximation of $I(\xi)$ for determined models

Idea: use an asymptotic (with respect to the number of experiments) to approximate the integration
Laplace Approximation:
Assuming nonzero second derivative and bounded third derivative of f:

$$\int exp\left[Mf(x)\right]dx = \sqrt{\frac{2\pi}{M|f''(x_0)|}}exp\left[Mf(x_0)\right] + \mathcal{O}\left(\frac{1}{M}\right).$$

Hint:

$$f(x) = f(x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \mathcal{O}\left(|x - x_0|^3\right).$$

Laplace approximation of $I(\xi)$ for determined models

Synthetic data model:

$$\boldsymbol{y}_i = \boldsymbol{g}(\theta_0, \xi) + \epsilon_i \,, i = 1, \ldots, M \,,$$

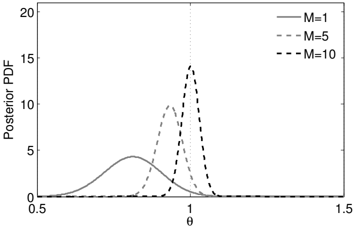


Figure 1: Posterior pdfs as M increases.

Laplace approximation of $I(\xi)$ for determined models

- Truncated Taylor expansion of $log(p(\boldsymbol{\theta}|\{\boldsymbol{y}_i\}))$ leads to a normal distribution $\mathcal{N}(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$.

Theorem 1

$$I = \int_{\Theta} \int_{\mathcal{Y}} \underbrace{\left[-\frac{1}{2} \log((2\pi)^d |\boldsymbol{\Sigma}|) - \frac{d}{2} - h(\hat{\boldsymbol{\theta}}) - \frac{tr(\boldsymbol{\Sigma}\, \boldsymbol{H}_p(\hat{\boldsymbol{\theta}}))}{2} \right]}_{D_{KL}}$$

$$p(\bar{\boldsymbol{y}}|\boldsymbol{\theta}_0) d\bar{\boldsymbol{y}} p(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0 + \mathcal{O}\left(\frac{1}{M^2}\right)$$

Q. Long, M. Scavino, R. Tempone, S. Wang: Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations, *Computer Methods in Applied Mechanics and Engineering* 259 (2013) 24-39.

Under-determined models

So far, the results are useful when the Laplace approximation can be applied: a dominant mode (or multiple equivalently dominant modes) exists.

Question: How about the cases, where an non-informative manifold exists?

Example 1: $g = (\theta_1^2 + \theta_2^2)^3 \xi^2 + (\theta_1^2 + \theta_2^2) \exp[-|0.2 - \xi|]$

Example 2:

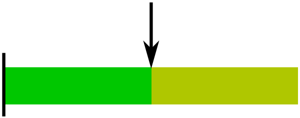
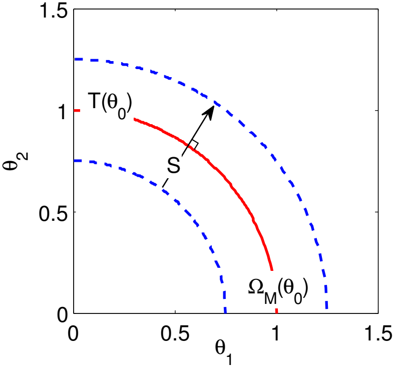


Figure 2: A cantilever beam.

The non-informative manifold



The definition of non-informative manifold

The definition of the manifold and a small region containing this manifold [2] :

$$T(\boldsymbol{\theta}_0) := \{\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d \ : \ p(\bar{\boldsymbol{y}}|\boldsymbol{\theta}) - p(\bar{\boldsymbol{y}}|\boldsymbol{\theta}_0) = \mathbf{0}\},$$

$$\Omega_M(\boldsymbol{\theta}_0) := \{\boldsymbol{\theta} \in \mathbb{R}^d : dist(\boldsymbol{\theta}, T(\boldsymbol{\theta}_0)) \leq \ell_0 M^{-\alpha}\}$$

[2] The volume of $\Omega_M(\boldsymbol{\theta}_0)$ contracts to zero in a slower rate than the square root of the number of replicate experiments M, i.e., $\alpha \in (0, 0.5)$.

Local reparameterization

- The diffeomorphism mapping: $\boldsymbol{f} : \Omega_{Ms,t} \to \Omega_M$
- Cost function: $F(\boldsymbol{\theta}) := \frac{1}{2}(\boldsymbol{g}(\boldsymbol{\theta}) - \boldsymbol{g}(\boldsymbol{\theta}_0))^T \boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{g}(\boldsymbol{\theta}) - \boldsymbol{g}(\boldsymbol{\theta}_0))$
- Hessian of F: $\boldsymbol{H}(\boldsymbol{f}(0, \boldsymbol{t})) = [\boldsymbol{U}\ \boldsymbol{V}]\, \boldsymbol{\Lambda}\, [\boldsymbol{U}\ \boldsymbol{V}]^T$
- Local coordinate $\boldsymbol{s}$: $\boldsymbol{s} = \boldsymbol{U}^T(\theta - \boldsymbol{f}(0, \boldsymbol{t}))$
- Prior weight function: $p(\boldsymbol{s}, \boldsymbol{t}) := p_{\boldsymbol{\Theta}}(\boldsymbol{f}(\boldsymbol{s}, \boldsymbol{t}))|\boldsymbol{J}|$
- Posterior weight function: $p(\boldsymbol{s}, \boldsymbol{t}|\bar{\boldsymbol{y}}) := p_{\boldsymbol{\Theta}}(\boldsymbol{f}(\boldsymbol{s}, \boldsymbol{t})|\bar{\boldsymbol{y}})|\boldsymbol{J}|$
- Due to Bayes' theorem, we have

$$p(\boldsymbol{s}, \boldsymbol{t}|\bar{\boldsymbol{y}}) = \frac{p(\bar{\boldsymbol{y}}|\boldsymbol{s},\boldsymbol{t})p(\boldsymbol{s},\boldsymbol{t})}{p(\bar{\boldsymbol{y}})} \quad \text{for} \quad (\boldsymbol{s}, \boldsymbol{t}) \in \Omega_{Ms,t}$$

Change of coordinates for the K–L divergence (D_{KL})

Approximated K–L divergence using the local coordinates $\boldsymbol{t}$ and $\boldsymbol{s}$:

$$D_{KL}(\bar{\boldsymbol{y}}) = \int_{T_{\boldsymbol{t}}} \int_{[-\ell_0 M^{-\alpha},\ \ell_0 M^{-\alpha}]} \log\left(\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{p(\boldsymbol{s},\boldsymbol{t})}\right) p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}}) p(\boldsymbol{t}|\bar{\boldsymbol{y}}) d\boldsymbol{s} d\boldsymbol{t} + \mathcal{O}_P\left(e^{-M^{\ell_0 \delta}}\right)$$

Laplace approximation for the conditional information gain

Gaussian approximations:

$$\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}}) = \frac{1}{(\sqrt{2\pi})^r|\boldsymbol{\Sigma}_{s|t}|^{1/2}} \exp\left[-\frac{(\boldsymbol{s}-\hat{\boldsymbol{s}})^T\boldsymbol{\Sigma}_{s|t}^{-1}(\boldsymbol{s}-\hat{\boldsymbol{s}})}{2}\right]$$

$$\tilde{p}(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}}) = p(\hat{\boldsymbol{s}},\boldsymbol{t}|\bar{\boldsymbol{y}}) \exp\left[-\frac{(\boldsymbol{s}-\hat{\boldsymbol{s}})^T\boldsymbol{\Sigma}_{s|t}^{-1}(\boldsymbol{s}-\hat{\boldsymbol{s}})}{2}\right]$$

$$\tilde{p}(\boldsymbol{s},\boldsymbol{t}) = p(\hat{\boldsymbol{s}},\boldsymbol{t}) \exp\left[\nabla \log p(\hat{\boldsymbol{s}},\boldsymbol{t})(\boldsymbol{s}-\hat{\boldsymbol{s}}) + \frac{(\boldsymbol{s}-\hat{\boldsymbol{s}})^T H_p(\hat{\boldsymbol{s}},\boldsymbol{t})(\boldsymbol{s}-\hat{\boldsymbol{s}})}{2}\right]$$

The information gain D_{KL} can be approximated by

$$D_{KL} = \int_{T_{\boldsymbol{t}}} \underbrace{\int_{[-\ell_{\mathbf{0}}M^{-\alpha},\ell_{\mathbf{0}}M^{-\alpha}]} \log\left(\frac{\tilde{p}(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{\tilde{p}(\boldsymbol{s},\boldsymbol{t})}\right) \tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s}}_{D_{\boldsymbol{s}|\boldsymbol{t}}}\, p(\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{t}$$

$$+\,\mathcal{O}_P\left(\frac{1}{M}\right),$$

with

$$D_{\boldsymbol{s}|\boldsymbol{t}} = -\log\left(\int_{T_{\boldsymbol{t}}} p(\hat{\boldsymbol{s}},\boldsymbol{t})|\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2}d\boldsymbol{t}\right) - \frac{r}{2}\log(2\pi) - \frac{r}{2} + \mathcal{O}_P(\frac{1}{M}).$$

Laplace approximation for the expected information gain for under determined models

Theorem 2

The expected information gain can be expressed as

$$I = \int_{\boldsymbol{\Theta}} \int_{\mathcal{Y}} \mathbf{1}_{\Omega_M} \left[-\log \left(\int_{T_{\boldsymbol{t}}} p(\hat{\boldsymbol{s}}, \boldsymbol{t}) |\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2} d\boldsymbol{t} \right) - \frac{r}{2} \log(2\pi) - \frac{r}{2} \right]$$
$$p(\bar{\boldsymbol{y}}|\boldsymbol{\theta}_0) p(\boldsymbol{\theta}_0) d\bar{\boldsymbol{y}} d\boldsymbol{\theta}_0 + \mathcal{O}\left(\frac{1}{M}\right),$$

where the error $\mathcal{O}\left(\frac{1}{M}\right)$ is dominated by the standard Laplace approximation in $\boldsymbol{s}$ direction.

Q. Long, M. Scavino, R. Tempone, S. Wang: A Laplace Method for Under-Determined Bayesian Optimal Experimental Designs. *Computer Methods in Applied Mechanics and Engineering* 285 (2015) 849-876.

Simplification of the integration over the manifold $T_{\boldsymbol{t}}$

Approximation of the conditional covariance matrix (by Woodbury's formula)

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}} =& \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{s}|\boldsymbol{t}} + O_P\left(\frac{1}{M\sqrt{M}}\right) \\
\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{s}|\boldsymbol{t}} =& \frac{1}{M}\left\{\boldsymbol{U}^T\left[\boldsymbol{J}_g(\boldsymbol{f}(\hat{\boldsymbol{s}},\boldsymbol{t}))^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{J}_g(\boldsymbol{f}(\hat{\boldsymbol{s}},\boldsymbol{t}))\right]\boldsymbol{U}\right\}^{-1}.
\end{aligned}
$$

Note that $|\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{s}|\boldsymbol{t}}|$ is independent to $\boldsymbol{t}$ for a given value of $\boldsymbol{s}$.

Simplification of the integration over the manifold $T_{\boldsymbol{t}}$

Theorem 3

The expected information gain can be expressed as

$$I = \int_{\boldsymbol{\Theta}} \int_{\mathcal{Y}} \mathbf{1}_{\Omega_M} \left[-\log \left(\int_{T_{\boldsymbol{t}}} p(\hat{\boldsymbol{s}}, \boldsymbol{t}) d\boldsymbol{t} \right) - \frac{1}{2} \log |\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{s}|\boldsymbol{t}}| - \frac{r}{2} \log(2\pi) - \frac{r}{2} \right] p(\bar{\boldsymbol{y}}|\boldsymbol{\theta}_0) p(\boldsymbol{\theta}_0) d\bar{\boldsymbol{y}} d\boldsymbol{\theta}_0 + \mathcal{O}\left(\frac{1}{M}\right),$$

- $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{s}|\boldsymbol{t}}$ is independent to $\boldsymbol{t}$.

Q. Long, M. Scavino, R. Tempone, S. Wang: A Laplace Method for Under-Determined Bayesian Optimal Experimental Designs. *Computer Methods in Applied Mechanics and Engineering* 285 (2015) 849-876.

Simplification of the integration over the manifold $T_{\boldsymbol{t}}$

We can furthermore approximate the maximum posterior solution of $\boldsymbol{s}$ for a given value of $\boldsymbol{t}$, i.e., $\hat{\boldsymbol{s}}$, by $\mathbf{0}$. The result 3 can be simplified to the following result 4.

Theorem 4

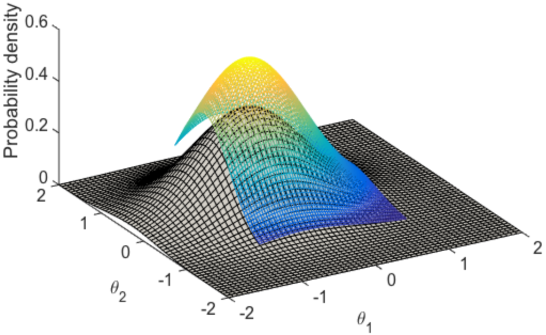
The expected information gain can be approximated by

$$I = \int_{\boldsymbol{\Theta}} \left[-\log \left(\int_{T_{\boldsymbol{t}}} p(\mathbf{0}, \boldsymbol{t}) d\boldsymbol{t} \right) - \frac{1}{2} \log |\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{s}|\boldsymbol{t}}| \right] p(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0 - \frac{r}{2} \log(2\pi) - \frac{r}{2} + \mathcal{O}\left(\frac{1}{M}\right) .$$

Q. Long, M. Scavino, R. Tempone, S. Wang: A Laplace Method for Under-Determined Bayesian Optimal Experimental Designs. *Computer Methods in Applied Mechanics and Engineering* 285 (2015) 849-876.

Truncated Gaussian approximation

Truncated Gaussian approximation



Truncated Gaussian approximation

Theorem 5

The expected information gain can be approximated by

$$I(\xi) = \int_{\Theta} \tilde{D}_{KL}(\boldsymbol{\theta}_0, \xi) p(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0 + \mathcal{O}\left(\frac{1}{N_e}\right),$$

with

$$\tilde{D}_{KL}(\boldsymbol{\theta}_0, \xi) = \int_{\Theta} \frac{\phi(\boldsymbol{\theta}|\mathbf{y})}{p(\boldsymbol{\theta})} \phi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \quad \textit{and} \quad \phi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\tilde{p}(\boldsymbol{\theta}|\mathbf{y}, \xi)}{\int_{\Theta} \tilde{p}(\boldsymbol{\theta}|\mathbf{y}, \xi) d\boldsymbol{\theta}}.$$

F. Bisetti, D. Kim, O. Knio, Q. Long, R. Tempone: Optimal Bayesian Experimental Design for Priors of Compact Support with Application to Shock-Tube Experiments for Combustion Kinetics. *International Journal for Numerical Methods in Engineering* (2016) DOI: 10.1002/nme.5211.

Multi level monte carlo for OED

General theory of multi level monte carlo

- Telescopic sum of expectations:

$$\mathbb{E}[P_L] = \sum_{l=0}^{L} \mathbb{E}[P_l - P_{l-1}],$$

where $P_{-1} = 0$.

- The MLMC estimator of $\mathbb{E}(P_L)$ reads

$$Y = \sum_{l=0}^{L} Y_l = \sum_{l=0}^{L} \frac{1}{N_l} \sum_{n=1}^{N_l} \left(P_l(\omega_n) - P_{l-1}(\omega_n)\right).$$

General theory of multi level monte carlo

Theorem 6

Let P denote a RV and P_l its numerical approximation on level l. If there exist independent estimators Y_l based on N_l MC samples, each with expected cost C_l and variance V_l, and positive constants α, β, γ, c_1, c_2, c_3, such that $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and

i. $|\mathbb{E}[P_l - P]| \leq c_1 2^{-\alpha l}$,

ii. $\mathbb{E}[Y_l] = \begin{cases} \mathbb{E}[P_0] & \text{if } l = 0 \\ \mathbb{E}[P_l - P_{l-1}] & \text{if } l > 0 \end{cases}$,

iii. $V_l \leq c_2 2^{-\beta l}$,

iv. $C_l \leq c_3 2^{\gamma l}$,

then there exists a positive constant c_4, such that for any $TOL < e^{-1}$ there are values L and N_l for which the multilevel estimator $Y = \sum_{l=0}^{L} Y_l$ has a mean-square-error with bound:

$$MSE := \mathbb{E}[(Y - \mathbb{E}[P])^2] < TOL^2$$

with a computational complexity C with bound:

$$\mathbb{E}[C] \leq \begin{cases} c_4 TOL^{-2} & \text{if } \beta > \gamma \\ c_4 TOL^{-2}(\log TOL)^2 & \text{if } \beta = \gamma \\ c_4 TOL^{-2-(\gamma-\beta)/\alpha} & \text{if } \beta < \gamma. \end{cases}$$

Monte Carlo Complexity:

$$\mathcal{O}\left(TOL^{-2-\frac{\gamma}{\alpha}}\right)$$

MLMC for nested integration

- Recap:

$$I_{DLMC} = \frac{1}{N_o} \sum_{I=1}^{N_o} \log \left(\frac{p(\bar{\mathbf{y}}_I | \boldsymbol{\theta}_I)}{p(\bar{\mathbf{y}}_I)} \right),$$

- This integral of expected information gain can be evaluated using the multi level estimator:

$$I_{MLMC} = \sum_{l=0}^{\infty} Y_l ,$$

$$Y_l = \frac{1}{N_{ol}} \sum_{i=1}^{N_{ol}} \left[\log \left(\frac{p(\bar{\mathbf{y}}_i | \boldsymbol{\theta}_i)}{p_l(\bar{\mathbf{y}}_i)} \right) - \frac{1}{2} \log \left(\frac{p(\bar{\mathbf{y}}_i | \boldsymbol{\theta}_i)}{p_{l-1}(\bar{\mathbf{y}}_i)} \right) - \frac{1}{2} \log \left(\frac{p(\bar{\mathbf{y}}_i | \boldsymbol{\theta}_i)}{p_{l-1}(\bar{\mathbf{y}}_i)} \right) \right]$$

- This estimator has a complexity of $\mathcal{O}(TOL^{-2})$ according to the theorem of MLMC.

MLMC for Laplace method

- Using discretization of the physical model to define level

$$\mathbb{E}[P_L] = \sum_{l=0}^{L} \mathbb{E}[P_l - P_{l-1}], \quad \text{with} \quad P_{-1} = 0.$$

where

$$P_l(\boldsymbol{\theta}) = \frac{1}{2}\log((2\pi)^d |\boldsymbol{\Sigma}_l(\hat{\boldsymbol{\theta}})|) - \frac{d}{2} - h(\boldsymbol{\theta}), \text{ and}$$
$$\boldsymbol{\Sigma}_l(\hat{\boldsymbol{\theta}}) \approx \left(N_e \boldsymbol{J}_l(\hat{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{J}_l(\hat{\boldsymbol{\theta}}) - \nabla\nabla h(\hat{\boldsymbol{\theta}})\right)^{-1}.$$

Applications

Illustrative example

$y = (\theta_1 + \theta_2)^3 \xi^2 + (\theta_1 + \theta_2)\exp[-|0.2 - \xi|] + \epsilon$, with $\epsilon \sim \mathcal{N}(0, 10^{-3})$.

Gaussian mixture prior

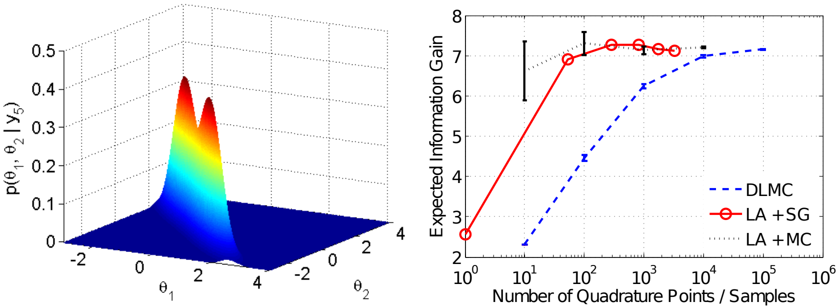
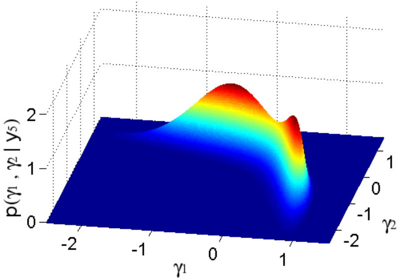


Figure 3: Left: the posterior pdf ($M = 5$); right: convergence. $\xi = 1$.

Illustrative example

Log Gaussian mixture prior $\gamma = \log \boldsymbol{\theta}$



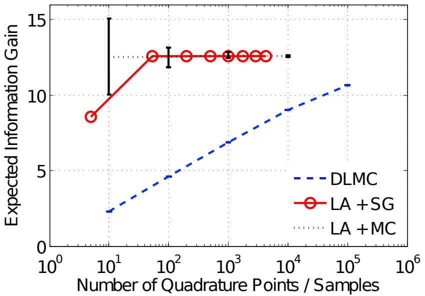
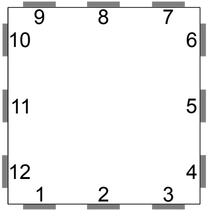
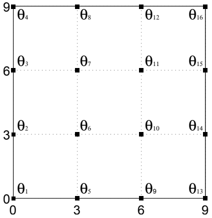
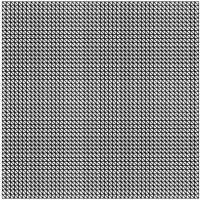


Figure 4: Left: the posterior pdf ($M = 5$); right: convergence. $\xi = 1$.

Impedance tomography







The parameters: piecewise linear conductivity field $\theta(\boldsymbol{x})$ controlled by the random vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{16})^T$.

Laplace equation: $\nabla \cdot \boldsymbol{q}(\boldsymbol{x}) = 0, \qquad \boldsymbol{q}(\boldsymbol{x}) = -\theta(\boldsymbol{x})\nabla u(\boldsymbol{x})$

Boundary conditions:

$$\boxed{\begin{array}{c} \int_{a_j} \boldsymbol{q} \cdot \boldsymbol{n} \, d\boldsymbol{x} = I_j \, , \quad j = 1, ..., I, \\ \boldsymbol{q} \cdot \boldsymbol{n} = 0 \quad \text{on} \quad \delta\Omega_N / \bigcup_{j=1}^{I} a_j \\ \sum_{j=1}^{I} U_j = 0 \, , \quad \sum_{j=1}^{I} I_j = 0 \end{array}}$$

Measurement: $y_j = \frac{1}{|a_j|} \int_{a_j} u_h(\boldsymbol{x}) d\boldsymbol{x} + \epsilon \, , \quad j = 1, ..., I$

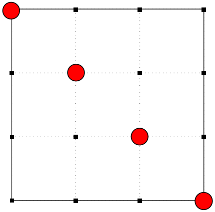
Impedance tomography

Similar to what we have done in the first example, we set the prior as a mixture log Gaussian ($\gamma = \log \boldsymbol{\theta}$) which adopts the following form:

$$p(\gamma) = 0.5 \times p_1(\gamma) + 0.5 \times p_2(\gamma)\,, \qquad (1)$$

where $p_1(\gamma)$ is the pdf which has mean $\mathbf{0}$, and $p_2(\gamma)$ is the pdf of a multivariate Gaussian with mean vector and covariance matrix as follows

$\gamma_0(4) = \gamma_0(7) = \gamma_0(10) = \gamma_0(13) = 2$
$\gamma_0(i) = 0, i \neq 4, 7, 10, 13$
$\Sigma_p(4,4) =$
$\Sigma_p(7,7) = \Sigma_p(10,10) = \Sigma_p(13,13) = 1$
$\Sigma_p(i,i) = 0.01, i \neq 4, 7, 10, 13$
$\Sigma_p(i,j) = 0, i \neq j$



Impedance tomography

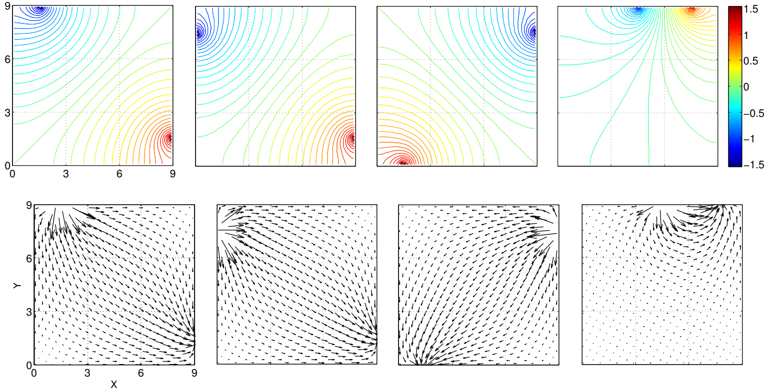


Figure 5: Voltage iso–contours and current patterns generated by the best and worst sensor placements.

Seismic source inversion

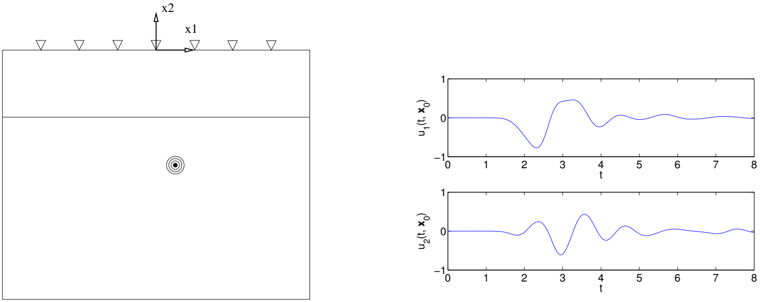


Figure 6: The two-layered spatial domain $D = [-10000, 10000] \times [-15000, 0]$ with stress-free and non-reflecting boundary conditions. An array of N_R receivers are located on the ground surface in equidistant recording points.

Q. Long, M. Motamed, R. Tempone: Fast Bayesian optimal experimental design for seismic source inversion. *Computer Methods in Applied Mechanics and Engineering*. 291 (2015) 123-145.

Seismic source inversion

The parameters: the source location, moment tensor components, and start time and frequency in the time function.

The forward problem: elastodynamic wave equations.

$$\rho(\mathbf{x})\,\mathbf{u}_{tt}(t,\mathbf{x}) - \nabla\cdot\boldsymbol{\sigma}(\mathbf{u}(t,\mathbf{x})) = \mathbf{f}(t,\mathbf{x};\boldsymbol{\theta}) \qquad \text{in } [0,T]\times D,$$
$$\boldsymbol{\sigma}(\mathbf{u}) = \lambda(\mathbf{x})\,\nabla\cdot\mathbf{u}\,\mathbf{I} + \mu(\mathbf{x})\left(\nabla\mathbf{u} + (\nabla\mathbf{u})^{\top}\right)$$

Initial and boundary conditions:

$$\begin{aligned}
&\mathbf{u}(0,\mathbf{x}) = \mathbf{0}, \quad \mathbf{u}_t(0,\mathbf{x}) = \mathbf{0} && \text{on } \{t=0\}\times D,\\
&\boldsymbol{\sigma}(\mathbf{u}(t,\mathbf{x}))\cdot\hat{\mathbf{n}} = \mathbf{0} && \text{on } [0,T]\times\partial D_0,\\
&\mathbf{u}_t(t,\mathbf{x}) = \boldsymbol{B}(\mathbf{x})\,\boldsymbol{\sigma}(\mathbf{u}(t,\mathbf{x}))\cdot\hat{\mathbf{n}} && \text{on } [0,T]\times\partial D_1.
\end{aligned}$$

Measurements: $\boldsymbol{y} = \boldsymbol{u} + \boldsymbol{\epsilon} = (u_1,\ldots,u_d)^{\top} + \boldsymbol{\epsilon}$.

Source term: $\mathbf{f}(t,\mathbf{x};\boldsymbol{\theta}) = S(t)\,\mathbf{M}\,\nabla\delta(\mathbf{x}-\mathbf{x}_s)$.

Priors:

$\theta_1 \sim \mathcal{U}(-1000,1000)$, $\theta_2 \sim \mathcal{U}(-3000,-1000)$, $\theta_3 \sim \mathcal{U}(0.5,1.5)$,
$\theta_4 \sim \mathcal{U}(3,5)$, $\theta_5,\theta_6,\theta_7 \sim \mathcal{U}(10^{13},10^{15})$.

Seismic source inversion

The experiment with $d_R = 1000$ gives the maximum information. Both lumping and sparsifying the seismograms give suboptimal designs.

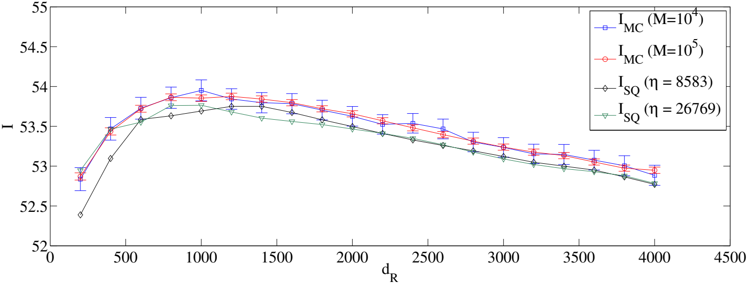


Figure 7: The expected information gain, computed both by Monte Carlo sampling (together with 99.7% confidence interval) and by sparse quadrature, for 20 different design scenarios.

Design of shock-tube experiments for combustion kinetics

- Forward model: A set of ordinary differential equations (ODE) describing the ignition of a reactive mixture.
- $H + O_2 \rightleftharpoons OH + O$
- Observable: maximum slope of the time history of water concentration
- Reaction constant/rate:

$$k_j^f = A_j T^{b_j} \exp\left(-E_j/\mathcal{R}T\right).$$

F. Bisetti, D. Kim, O. Knio, Q. Long, R. Tempone: Optimal Bayesian Experimental Design for Priors of Compact Support with Application to Shock-Tube Experiments for Combustion Kinetics. *International Journal for Numerical Methods in Engineering* (2016) DOI: 10.1002/nme.5211.

Convergence test

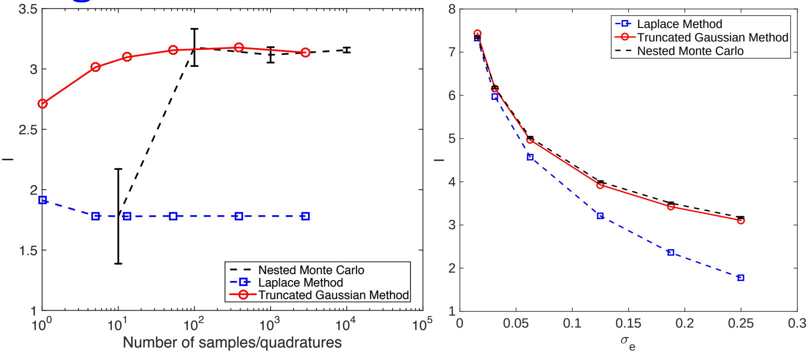


Figure 8: Convergence of the expected information gain of three experiments with $\xi_1 = [1500, 5\%]^\top$, $\xi_2 = [1100, 0.5\%]^\top$, $\xi_3 = [1500, 0.5\%]^\top$ and $\sigma_e = 0.25$. The statistical error bars represent 95% confidence intervals.

Truncated Gaussian approximation reduces significantly the error of direct Laplace method.

Convergence and CPU time

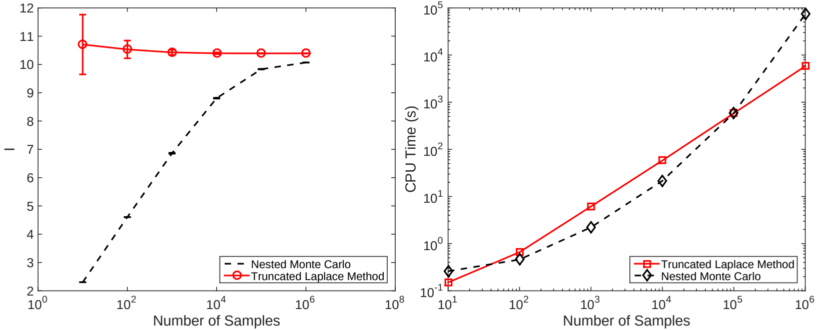


Figure 9: Left: convergence of the expected information gain; Right: CPU time.

Design of a single experiment

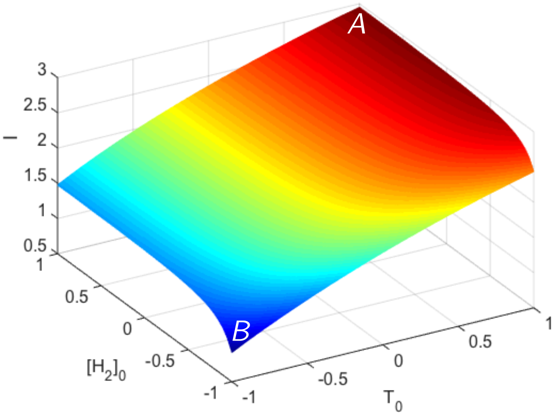


Figure 10: The expected information gain of a single experiment with $\sigma_e = 0.25$. Note that the ranges of T_0 and $[\mathrm{H}_2]_0$ are normalized to $[-1, 1]$.

Validation using legacy data

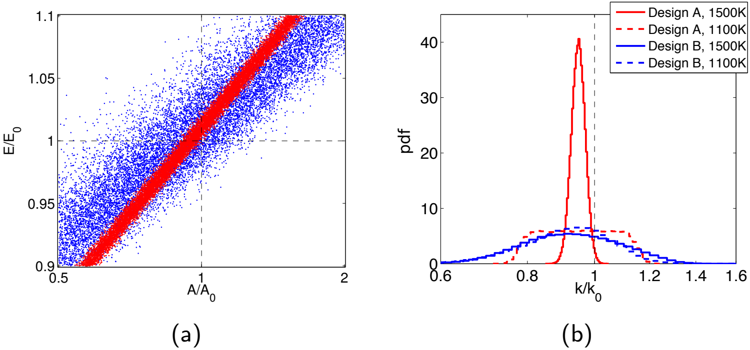
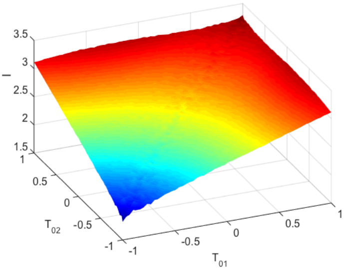


Figure 11: (a) Posterior samples of A and E based on real data from a single experiment: low (blue) and high (red) temperature designs. (b) The probability densities of k at 1100 K and 1500 K. Data extracted from Hong et al. 2011.

Higher temperature leads to higher concentration of pdf.

Design of two experiments under different temperatures



(a) (b)

Figure 12: Expected information gain for the two-run experimental design problem. In both experiments, $[\mathrm{H}_2]_0 = 5\%$. (a): $\sigma_e = 0.25$ and (b): $\sigma_e = 0.025$. The ranges of T_{01} and T_{02} are normalized to $[-1, 1]$.

Level of measurement noise changes the optimal design.

Comparison of DLMC, MLMC, LA+MC, LA+MLMC

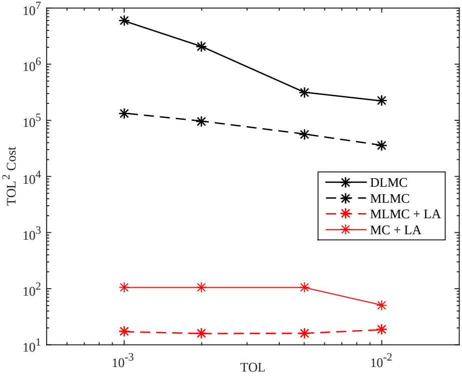


Figure 13: Cost comparison between the different methods

Conclusions

- Extend Bayesian experimental design methodology based on the Laplace approximation from classical scenario to under determined models.
- (Generalized) Laplace method has huge computational advantage over the nested integration.
- Approximating the posterior by a truncated Gaussian distribution in the case of priors with compact supports.
- Multi level approach should be used to accelerate computation when there is a lack of measure concentration.

Reference

- Q. Long, M. Scavino, R. Tempone, S. Wang: Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations, *Computer Methods in Applied Mechanics and Engineering* 259 (2013) 24-39.
- Q. Long, M. Scavino, R. Tempone, S. Wang: A Laplace method for under-determined Bayesian optimal experimental designs. *Computer Methods in Applied Mechanics and Engineering* 285 (2015) 849-876.
- Q. Long, M. Motamed, R. Tempone: Fast Bayesian optimal design for seismic source inversion, *Computer Methods in Applied Mechanics and Engineering* 291 (2015) 123-145.
- F. Bisetti, D. Kim, O. Knio, Q. Long, R. Tempone: Optimal Bayesian experimental design for priors of compact support with application to shock-tube experiments for combustion kinetics. *International Journal for Numerical Methods in Engineering* (2016) DOI: 10.1002/nme.5211.
- B.M. Dia, L. Espath, C. B. Issaid, Q. Long, R. Tempone: A multi level computational approach for Bayesian Experimental Design, *In preparation* (2016).
- M. Giles: Multilevel Monte Carlo path simulation. *Operations Research* 56 (2008) 607-617.
- K. J. Ryan: Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *Journal of Computational and Graphical Statistics* 12 (2003) 585-603.