Sequential design of experiments on a stochastic multi-fidelity simulator

STROH Rémi\(^{(a,b)}\)
BECT Julien\(^{(a)}\), VAZQUEZ Emmanuel\(^{(a)}\)
DEMEYER Séverine\(^{(b)}\), FISCHER Nicolas\(^{(b)}\)

(a) Laboratoire des Signaux et Systèmes (L2S)
(b) Laboratoire National de métrologie et d’Essais (LNE)
1. Introduction

2. Sequential design of experiments

3. Academic example
Outline

1. Introduction

2. Sequential design of experiments

3. Academic example
• Fire safety: conformity of a smoke extraction system
  • Expensive experiments ➔ use of numerical models

Real Experiment
• Images from [Kerber, 2005]

Fire Dynamics Simulator (FDS)
Main properties of the considered simulators:

1. Multi-fidelity
2. Tunable cost
3. Stochastic outputs
Multi-fidelity: same physical phenomenon ➔ several models with various accuracy.
Tunable cost

- Cost of observation: function of the fidelity $C(t)$
  - Cheap simulation, but low fidelity
  - High fidelity simulation, but time-consuming
Stochastic outputs

\[(x, t) \xrightarrow{} \text{Simulator} \xrightarrow{} Z \sim \mathbb{P}_{x,t}^\text{sim}\]

- Stochastic: same input \(\rightarrow\) different outputs

![Graph showing stochastic outputs distribution](image)
• Probability of exceeding a critical threshold $z^{crit}$

$$p(x, t^{HF}) = \mathbb{P}^{sim}_{x,t^{HF}}(Z > z^{crit})$$

• $t^{HF}$: the highest-fidelity level
• **Goal**: selecting $(x_1, t_1), \ldots, (x_n, t_n)$ to estimate the function $p$ with a minimal cost $C(t_1) + \cdots + C(t_n)$
  - Observations $(x_i, t_i; z_i)_{1 \leq i \leq n} \Rightarrow$ Estimation $\hat{p}_n$ of $p$

---

The 22\textsuperscript{nd} march 2017  Sequential design of experiments on a stochastic multi-fidelity simulator
• **Goal**: selecting \((x_1, t_1), \ldots, (x_n, t_n)\) to estimate the function \(p\) with a minimal cost \(C(t_1) + \cdots + C(t_n)\)

• **Sequential design**
  - use the \(n\) first observations to select the \((n + 1)^{th}\) observation
1. Introduction

2. Sequential design of experiments

3. Academic example
Model of the simulator

- **Prior distribution:**
  - Output $Z$ at $x, t$ follows a normal distribution
    \[ Z|\xi \sim \mathcal{N}(\xi(x, t), \lambda(x, t)) \]
  - Mean function $\xi$: Gaussian process
    \[ \xi \sim \mathcal{GP}(m, k) \]

- **Posterior distribution: kriging**
  - Mean function $\xi|\chi_n$
    \[ \xi|\chi_n \sim \mathcal{GP}(m_n, k_n) \]
Auto-regressive multi-fidelity covariance

- **Mean function**

\[
\xi(x, t) = \begin{cases} 
\xi_{LF}(x) & \text{if } t = 1 \\
\rho \xi_{LF}(x) + \delta(x) & \text{if } t = 2 
\end{cases}
\]

- \(\xi_{LF}\): low-fidelity simulator
- \(\xi_{HF} = \rho \xi_{LF} + \delta\): high-fidelity simulator, linked to the low-fidelity by a linear relationship

- **Covariance function**

\[
k((x, t), (x', t')) = \begin{cases} 
k_{LF}(x, x') & \text{if } t = t' = 1 \\
\rho k_{LF}(x, x') & \text{if } t \neq t' \\
\rho^2 k_{LF}(x, x') + k_{\delta}(x, x') & \text{if } t = t' = 2 
\end{cases}
\]

- \(k_{LF}\): covariance of the low-fidelity simulator
- \(k_{\delta}\): covariance of the difference between high- and low-fidelity levels
- \(\rho\): correlation between the low- and high-fidelity levels

Non-stationary multi-fidelity covariance

- **Mean function**
  \[ \xi(x, t) = \xi_0(x) + \epsilon(x, t) \]
  - \( \xi_0 \): ideal simulator (Ex: mesh size = 0)
  - \( \epsilon \): system error between ideal and real simulators at \( t \)

- **Covariance function**
  \[ k((x, t), (x', t')) = k_0(x, x') + r(t, t') \cdot k_\epsilon(x, x') \]
  - \( k_0 \): covariance of \( \xi_0 \)
  - \( k_\epsilon \): covariance of \( \epsilon \) according to \( x \)
  - \( r \): rules the decrease of the error
  - [Picheny and Ginsbourger, 2013], [Tuo et al., 2014]
Probability of exceeding the critical threshold

- Probability of exceeding the critical threshold:

\[ p(x, t) = \mathbb{P}_{x,t}^{sim}(Z > z^{\text{crit}} | \chi_n) = \Phi \left( \frac{\xi(x, t) - z^{\text{crit}}}{\sqrt{\lambda(x, t)}} \right) \]

- First and second moments

  - Expectation: \( \mathbb{E}_n[p(x, t)] = \Phi(u_n(x, t)) = \hat{p}_n(x, t) \)
  - Variance: \( \text{Var}_n[p(x, t)] = \Phi_2(u_n(x, t), u_n(x, t); r_n(x, t)) - \Phi^2(u_n(x, t)) \)

\[ u_n(x, t) = \frac{m_n(x,t)-z^{\text{crit}}}{\sqrt{\sigma_n^2(x,t)+\lambda(x,t)}} \quad r_n(x, t) = \frac{\sigma_n^2(x,t)}{\sigma_n^2(x,t)+\lambda(x,t)} \]

\[ \sigma_n^2(x, t) = k_n((x, t), (x, t)) \]

- \( \Phi \): cumulative distribution function (cdf) of the normal distribution
- \( \Phi_2 \): cdf of the bivariate normal distribution
Measure of uncertainty

\[ H_n = \mathbb{E}_n [\| \hat{p}_n(\cdot, t^{HF}) - p(\cdot, t^{HF}) \|^2] = \int_X \text{Var}_n [p(x, t^{HF})] \, dx \]

- $L^2$-norm of the error of the estimator at the highest level of fidelity
Stepwise Uncertainty Reduction

- Measure of uncertainty
  \[ H_n = \mathbb{E}_n[\|\hat{p}_n(\cdot, t^{HF}) - p(\cdot, t^{HF})\|^2] = \int_X \text{Var}_n[p(x, t^{HF})] dx \]

- Stepwise uncertainty reduction algorithm
  \[ (x_{n+1}, t_{n+1}) = \arg\min_{x,t} \{ \mathbb{E}_n[H_{n+1}|X_{n+1} = x, T_{n+1} = t] \} \]

- [Vazquez and Bect, 2009]
Measure of uncertainty

\[ H_n = \mathbb{E}_n [\| \hat{\rho}_n(\cdot, t^{HF}) - p(\cdot, t^{HF}) \|^2] = \int_X \text{Var}_n [p(x, t^{HF})] dx \]

Stepwise uncertainty reduction algorithm

\[ (x_{n+1}, t_{n+1}) = \arg\min_{x,t} \{ \mathbb{E}_n [H_{n+1} | X_{n+1} = x, T_{n+1} = t] \} \]

Analytical expression

\[ \mathbb{E}_n [H_{n+1} | X_{n+1} = x, T_{n+1} = t] \]

\[ = \int_X [\Phi_2(u_n(y, t^{HF}), u_n(y, t^{HF}); r_n(y, t^{HF}))-\Phi_2(u_n(y, t^{HF}), u_n(y, t^{HF}); \tilde{r}_n((x, t), (y, t^{HF})))]dy \]

\[ \tilde{r}_n((x, t), (y, t^{HF})) = \frac{k((x, t),(y, t^{HF}))^2}{(\sigma^2_n(x, t)+\lambda(x, t)) \cdot (\sigma^2_n(y, t^{HF})+\lambda(y, t^{HF}))} \]
• Different costs $C(x, t)$ of observations
  ➔ Trade-off between $H_n$ reduction and cost $C(x, t)$

• [Huang et al. 2006], [Le Gratiet and Cannamela, 2015]: comparison between benefit and cost
Maximum Speed of Uncertainty Reduction (MSUR)

\[(x_{n+1}, t_{n+1}) = \arg\max_{x,t} \left\{ \frac{H_n - \mathbb{E}_n[H_{n+1}|X_{n+1} = x, T_{n+1} = t]}{C(x, t)} \right\} \]

- MSUR = Benefit/Cost
- Adaptable for any measure of uncertainty \(H_n\)
- If \(C\) is constant \(\Rightarrow\) equivalent to SUR algorithm
If the cost depends only on the level $C(x, t) = C(t)$

Algorithm: separate optimization of the point $x$ and the level $t$

1. $x^*(t) = \arg\min_x \{\mathbb{E}_n[H_{n+1}|X_{n+1} = x, T_{n+1} = t]\}$

2. $t_{n+1} = \arg\max_t \left\{ \frac{H_n - \mathbb{E}_n[H_{n+1}|X_{n+1} = x^*(t), T_{n+1} = t]}{C(t)} \right\}$

3. $x_{n+1} = x^*(t_{n+1})$
1. Introduction

2. Sequential design of experiments

3. Academic example
1. Introduction

2. Sequential design of experiments

3. Academic example
   a. Presentation of the example
   b. Comparison of sequential designs of experiments
Consider a damped harmonic simulator with random drive
\[ \ddot{X}(t) + 2\zeta \omega_0 \dot{X}(t) + \omega_0^2 X(t) = W(t) \]

- \( \omega_0 \): the undamped angular frequency
- \( \zeta \): the damping ratio
- \( W \): a Brownian motion, with spectral density \( S = 1 \)
- Initial conditions: \( X(t = 0) = 0, \dot{X}(t = 0) = 0 \)

[Au and Beck, 2001]
• Consider a damped harmonic simulator with random drive
\[ \ddot{X}(t) + 2\zeta\omega_0\dot{X}(t) + \omega_0^2 X(t) = W(t) \]

• Ideal simulator
\[ F: (\omega_0, \zeta) \mapsto \max_{0 \leq t \leq t_{\text{end}} = 30} \{ \log|X(t)| \} \]

\( \omega_0 = 15.708 \text{ rad/s}; \zeta = 0.2 \)

\( \log(|\cdot|) \)
• Approximation by an explicit Exponential Euler Scheme

\[ X(n \cdot dt) \approx \tilde{X}_n \]

• Multi-fidelity simulator

\[ f : (\omega_0, \zeta, dt) \mapsto \max_{0 \leq n \leq \left\lfloor \frac{t_{\text{end}}}{dt} \right\rfloor} \{ \log |\tilde{X}_n| \} \]

\[ dt = 1 \text{ s} \quad dt = 0.2 \text{ s} \quad dt = 0.05 \text{ s} \quad dt = 0.01 \text{ s} \]

• [Jentzen and Kloeden, 2009]
The output distribution at \((\omega_0, \zeta, dt)\) can be approximated by a normal distribution.

10^5 simulations at \(\omega_0 = 15.708\) rad/s and \(\zeta = 0.2\).

- \(dt = 1\) s
- \(dt = 0.2\) s
- \(dt = 0.05\) s
- \(dt = 0.01\) s
Mean function $\xi$

- $10^5$ simulations
- $0 \leq \omega_0 \leq 30 \text{ rad/s}$, $0 \leq \zeta \leq 1$ - Grid: $100 \times 100$

$dt = 1 \text{ s}$  
$dt = 0.2 \text{ s}$  
$dt = 0.05 \text{ s}$  
$dt = 0.01 \text{ s}$
• Critical threshold $z^{crit} = -3$

\begin{align*}
dt &= 1 \text{ s} \\
dt &= 0.2 \text{ s} \\
dt &= 0.05 \text{ s} \\
dt &= 0.01 \text{ s}
\end{align*}
• True probability of exceeding the threshold $p(\omega_0, \zeta; dt)$

\[ dt = 1 \text{ s} \quad dt = 0.2 \text{ s} \quad dt = 0.05 \text{ s} \quad dt = 0.01 \text{ s} \]
• Computation time $C(dt)$: linear in $1/dt$.

<table>
<thead>
<tr>
<th>Time step $dt$</th>
<th>1 s</th>
<th>0.2 s</th>
<th>0.05 s</th>
<th>0.01 s</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU Time (ms)</td>
<td>0.799 ms</td>
<td>1.85 ms</td>
<td>5.78 ms</td>
<td>26.7 ms</td>
</tr>
<tr>
<td>Cost function $C(dt)$</td>
<td>0.030</td>
<td>0.069</td>
<td>0.217</td>
<td>1.00</td>
</tr>
</tbody>
</table>
1. Introduction

2. Sequential design of experiments

3. Academic example
   a. Presentation of the example
   b. Comparison of sequential designs of experiments
• Target: probability of exceeding the threshold at the highest level of fidelity $dt = 0.01$ s

• Initial design: Nested LHS on 5 levels

<table>
<thead>
<tr>
<th>$dt$ (s)</th>
<th>1.00</th>
<th>0.50</th>
<th>0.33</th>
<th>0.25</th>
<th>0.20</th>
<th>0.17</th>
<th>0.10</th>
<th>0.05</th>
<th>0.02</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nb. points</td>
<td>180</td>
<td>60</td>
<td>20</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

• [Qian, 2009]

• Initial budget: $9.87 \times$
  
  $1 \times = \text{cost for 1 observation at the level } dt = 0.01 \text{ s}$
  
  $1.96 \times = \text{cost for 1.96 observations at the level } dt = 0.02 \text{ s}$
  
  $33.4 \times = \text{cost for 33.4 observations at the level } dt = 1 \text{ s}$
  
  $\ldots$
Comparison between designs of experiments

- Initial budget: 9.87 €
  Supplementary budget: 10 €
- 6 designs of experiments (DoE)
  - 5 Single level DoE
  - Multi-level DoE

<table>
<thead>
<tr>
<th>Sequential design</th>
<th>Criterion</th>
<th>Nb. Points Initial design</th>
<th>Nb. Points Final design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single level ($dt = 0.17 s$)</td>
<td>SUR</td>
<td>275</td>
<td>275+145 = 420</td>
</tr>
<tr>
<td>Single level ($dt = 0.10 s$)</td>
<td>SUR</td>
<td>275</td>
<td>275 + 85 = 360</td>
</tr>
<tr>
<td>Single level ($dt = 0.05 s$)</td>
<td>SUR</td>
<td>275</td>
<td>275 + 46 = 321</td>
</tr>
<tr>
<td>Single level ($dt = 0.02 s$)</td>
<td>SUR</td>
<td>275</td>
<td>275 + 19 = 294</td>
</tr>
<tr>
<td>Single level ($dt = 0.01 s$)</td>
<td>SUR</td>
<td>275</td>
<td>275 + 10 = 285</td>
</tr>
<tr>
<td>Multi-level</td>
<td>MSUR</td>
<td>275</td>
<td>275 + ? = ?</td>
</tr>
</tbody>
</table>
Comparison between designs of experiments

- Initial budget: 9.87 ¤
  Supplementary budget: 10 ¤
- 6 designs of experiments (DoE)
  - 5 Single level DoE
  - Multi-level DoE
- Same model:
  - Same covariance function
  - Hyper-parameters estimated on a large design
  - Fixed hyper-parameters during the sequential designs
- Each DoE: 12 repetitions
Criteria of comparison

\[ \mathbb{L}^2 \text{-error on the probability function}\]

\[ \sqrt{\int_{[0;30] \times [0;1]} (\hat{p}_n(x, t^{HF}) - p(x, t^{HF}))^2 dx} \]

\[ t^{HF} = 0.01 \text{ s} \]
Multi-fidelity is better

$L^2$-error on the probability function

- Low-fidelity levels are biased
  High-fidelity levels are slow
- In this example, multi-fidelity finds the best trade-off
Goal: sequential design of experiments to estimate probability on stochastic multi-fidelity numerical models

New SUR criteria to estimate probability of exceeding a threshold on stochastic simulator

Adaptation to multi-fidelity model → Maximum Speed of Uncertainty Reduction (MSUR)
  - **MSUR = (Uncertainty Reduction)/Cost**

Results on an academic example → automatic trade-off between cost and fidelity
Do you have any questions?

Thank you for your attention!


References (1/5): SUR-algorithm
References (2/5): auto-regressive multi-fidelity


http://rspa.royalsocietypublishing.org/content/465/2102/649

Mean and variance functions

- Mean function $\xi$ and variance function $\lambda$

![Mean function $\xi$](image1)

![Mean function $\xi$](image2)

![Mean function $\xi$](image3)

![Mean function $\xi$](image4)

$dt = 1 \text{ s}$

$dt = 0.2 \text{ s}$

$dt = 0.05 \text{ s}$

$dt = 0.01 \text{ s}$