

Sequential design of multi-fidelity computer experiments: maximizing the rate of stepwise uncertainty reduction

Rémi STROH^{*†} Julien BECT^{*} Séverine DEMEYER[†]
Nicolas FISCHER[†] Damien MARQUIS[‡] Emmanuel VAZQUEZ^{*}

^{*} Université Paris-Saclay, CNRS, CentraleSupélec,
Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France
Email: firstname.lastname@l2s.centralesupelec.fr

[†] Département sciences des données et incertitudes
[‡] Département comportement au feu et sécurité incendie
Laboratoire national de métrologie et d'essais, Trappes, France
Email: firstname.lastname@lne.fr

May 10, 2021

SUPPLEMENTARY MATERIAL

Contents

A	Introduction	3
B	Non-sequential designs in multi-fidelity	3
C	A new SUR criterion	4
C.1	Criterion definition and result statement	4
C.2	A useful identity	5
C.3	Proof of Proposition 1	6
D	Additional information regarding the examples	7
D.1	One-dimensional example	7
D.2	Random damped harmonic oscillator	7

A Introduction

This document contains supplementary material for the article “Sequential design of multi-fidelity computer experiments: maximizing the rate of stepwise uncertainty reduction”. It is organized as follows. Section B provides a short literature review on non-sequential designs for multi-fidelity, which complements the literature review on sequential designs given in Section 3.1 of the article. Section C provides (with proof) a new SUR criterion, which covers as a special case the criterion provided (without proof) in Section 3.2 of the article. Finally, Section D provides additional information regarding the examples presented in Sections 4.2 and 4.3 of the article.

B Non-sequential designs in multi-fidelity

In this section, we provide a very brief literature review on non-sequential designs for multi-fidelity.

A common recommendation for multi-level designs is nesting. A multi-level design is nested when any observed point at a level $\delta^{(s)}$ is also observed at every lower-fidelity level $\delta^{(s')}$, $s' < s$. Furthermore, space-filling designs are also expected to ensure observations in the whole input domain, as usual in Gaussian process regression.

A simple method to create a nested design is proposed by Forrester et al. (2007). It draws a maximin Latin Hypercube Sampling (LHS) at the lowest fidelity design, and then selects subsets of this LHS at the next levels. Le Gratiet and Garnier (2014) suggest to start with the highest-fidelity level and add points on the lower-fidelity levels to ensure better space-filling properties. The reader is also referred to Rennen et al. (2010) for a method which applies when there is only two levels of fidelity.

In our work, we use the method proposed by Qian (2009), which construct Nested Latin Hypercube Sampling (NLHS). An NLHS is a nested design with the property of being an LHS at each level of fidelity. We add a maximin optimization step to ensure that the design is space-filling at each level.

Note that this method was extended in several directions. He and Qian (2011), Yang et al. (2014), Guo et al. (2017), and Xu et al. (2017) propose methods to build NLHS with particular structures, such as orthogonality.

C A new SUR criterion

C.1 Criterion definition and result statement

Let $\xi = (\xi(\mathbf{x}))_{\mathbf{x} \in \mathbb{X}}$ denote a Gaussian process prior for the mean response of a stochastic simulator with Gaussian responses—i.e., a simulator which produces random responses

$$Z \mid \xi \sim \mathcal{N}(\xi(\mathbf{x}), \lambda(\mathbf{x})),$$

where $\mathbf{x} \in \mathbb{X}$ denotes the vector of inputs of the simulator and $\lambda : \mathbb{X} \rightarrow [0, +\infty)$ is a known variance function. Assume that the quantity of interest is the probability function $\alpha : \mathbb{X} \rightarrow [0, 1]$ defined by

$$\alpha(\mathbf{x}) = \mathbb{P}(Z_{\mathbf{x}} > z^{\text{crit}}),$$

where $z^{\text{crit}} \in \mathbb{R}$ is a given threshold and $Z_{\mathbf{x}}$ denotes a (future) response of the simulator with \mathbf{x} as the input. More explicitly, we have

$$\alpha(\mathbf{x}) = \Phi\left(\frac{\xi(\mathbf{x}) - z^{\text{crit}}}{\sqrt{\lambda(\mathbf{x})}}\right). \quad (1)$$

Let $\hat{\alpha}_n(\mathbf{x})$ denote the posterior mean of $\alpha(\mathbf{x})$ given n responses Z_1, \dots, Z_n of the simulator at design points $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{X}$ (possibly selected in a sequential manner). Let μ denote a positive, bounded measure on \mathbb{X} , and consider the measure of uncertainty H_n defined by

$$H_n = \mathbb{E}_n \left(\int (\alpha(\mathbf{x}) - \hat{\alpha}_n(\mathbf{x}))^2 \mu(d\mathbf{x}) \right) = \int \text{var}_n(\alpha(\mathbf{x})) \mu(d\mathbf{x}).$$

The following result provides a tractable expression for the corresponding SUR criterion for a batch $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_l)_{1 \leq l \leq q} \in \mathbb{X}^q$ of candidate points. The criterion presented in the article corresponds to the fully sequential case ($q = 1$) and to a particular choice of the measure μ .

Proposition 1. *Let m_n (resp. k_n) denote the posterior mean (resp. the posterior covariance) of ξ given n observations. Let*

$$J_n(\tilde{\mathbf{x}}) = \mathbb{E} \left(H_{n+q} \mid \mathbf{X}_{n+1} = \tilde{\mathbf{x}}_1, \dots, \mathbf{X}_{n+q} = \tilde{\mathbf{x}}_q \right)$$

and $G_n(\tilde{\mathbf{x}}) = H_n - J_n(\tilde{\mathbf{x}})$. Then,

$$J_n(\tilde{\mathbf{x}}) = \int \left[\Phi_2(u_n(\mathbf{x}), a_n(\mathbf{x}); r_n(\mathbf{x})) - \Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); \tilde{r}_n(\mathbf{x}; \tilde{\mathbf{x}})) \right] \mu(d\mathbf{x})$$

and

$$G_n(\tilde{\mathbf{x}}) = \int \left[\Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); \tilde{r}_n(\mathbf{x}; \tilde{\mathbf{x}})) - \Phi(a_n(\mathbf{x}))^2 \right] \mu(d\mathbf{x}),$$

with Φ the cdf of the standard normal distribution, $\Phi_2(\cdot, \cdot; \rho)$ the cdf of the standard bi-variate normal distribution with correlation ρ , and

$$\begin{aligned} a_n(\mathbf{x}) &= (m_n(\mathbf{x}) - z^{\text{crit}}) / \sqrt{v_n(\mathbf{x})}, & v_n(\mathbf{x}) &= k_n(\mathbf{x}, \mathbf{x}) + \lambda(\mathbf{x}), \\ r_n(\mathbf{x}) &= k_n(\mathbf{x}, \mathbf{x}) / v_n(\mathbf{x}), & \tilde{r}_n(\mathbf{x}; \tilde{\mathbf{x}}) &= \nu_n(\mathbf{x}, \mathbf{x}; \tilde{\mathbf{x}}) / v_n(\mathbf{x}), \\ \nu_n(\mathbf{x}, \mathbf{x}'; \tilde{\mathbf{x}}) &= k_n(\mathbf{x}, \mathbf{x}') - k_{n+q}(\mathbf{x}, \mathbf{x}'; \tilde{\mathbf{x}}) = k_n(\tilde{\mathbf{x}}, \mathbf{x})^T K_n(\tilde{\mathbf{x}}, \tilde{\mathbf{x}})^{-1} k_n(\tilde{\mathbf{x}}, \mathbf{x}'), \\ k_n(\tilde{\mathbf{x}}, \mathbf{x}) &= (k_n(\tilde{\mathbf{x}}_l, \mathbf{x}))_{1 \leq l \leq q}, & K_n(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}) &= (k_n(\tilde{\mathbf{x}}_l, \mathbf{x}_{l'}) + \lambda(\tilde{\mathbf{x}}_l) \cdot \delta_{l=l'})_{1 \leq l, l' \leq q}. \end{aligned}$$

Remark 1. In the expressions of $J_n(\tilde{\mathbf{x}})$ and $G_n(\tilde{\mathbf{x}})$, the only part which depends on the future design $\tilde{\mathbf{x}}$ is $\int \Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); \tilde{r}_n(\mathbf{x}; \tilde{\mathbf{x}})) \mu(d\mathbf{x})$, which must be maximized.

C.2 A useful identity

Let $\tilde{\Phi}_d(\cdot; m, K)$ denote the cumulative distribution function of the d -variate normal distribution with mean m and covariance matrix K . The following identity is used by Chevalier et al. (2014) for the computation of SUR criteria similar to ours.

Lemma 2. *Let $W \sim \mathcal{N}(m, K)$ be a d -dimensional Gaussian vector. Then, for any mean vector m' and covariance matrix K' ,*

$$\mathbb{E}(\tilde{\Phi}_d(W; m', K')) = \tilde{\Phi}_d(m; m', K + K').$$

Proof. Let $W' \sim \mathcal{N}(m', K')$ be independent from W . Then

$$\mathbb{E}(\tilde{\Phi}_d(W, m', K')) = \mathbb{E}(\mathbb{P}(W' \leq W \mid W)) = \mathbb{P}(W' \leq W),$$

and, using that $W' - W = W'' - m$ with $W'' \sim \mathcal{N}(m', K + K')$,

$$\mathbb{P}(W' \leq W) = \mathbb{P}(W'' \leq m) = \tilde{\Phi}_d(m; m', K + K'). \quad \square$$

Corollary 3. *Let $m, m' \in \mathbb{R}$, $v, v' \in [0, +\infty)$ and $W \sim \mathcal{N}(m, v)$. Then*

$$\mathbb{E}\left(\Phi\left(\frac{W - m'}{\sqrt{v'}}\right)\right) = \Phi\left(\frac{m - m'}{\sqrt{v + v'}}\right)$$

and

$$\mathbb{E}\left(\Phi\left(\frac{W - m'}{\sqrt{v'}}\right)^2\right) = \Phi_2\left(\frac{m - m'}{\sqrt{v + v'}}, \frac{m - m'}{\sqrt{v + v'}}; \frac{v}{v + v'}\right).$$

C.3 Proof of Proposition 1

Proof. Recall from (1) that

$$\alpha(\mathbf{x}) = \Phi \left(\frac{\xi(\mathbf{x}) - z^{\text{crit}}}{\sqrt{\lambda(\mathbf{x})}} \right).$$

It follows from Corollary 3 that

$$\mathbb{E}_{n+q}(\alpha(\mathbf{x})) = \Phi \left(\frac{m_{n+q}(\mathbf{x}) - z^{\text{crit}}}{\sqrt{\lambda(\mathbf{x}) + k_{n+q}(\mathbf{x}, \mathbf{x})}} \right) = \Phi \left(\frac{m_{n+q}(\mathbf{x}) - z^{\text{crit}}}{\sqrt{v_{n+q}(\mathbf{x})}} \right), \quad (2)$$

and thus

$$\text{var}_{n+q}(\alpha(\mathbf{x})) = \mathbb{E}_{n+q}(\alpha(\mathbf{x})^2) - \Phi \left(\frac{m_{n+q}(\mathbf{x}) - z^{\text{crit}}}{\sqrt{v_{n+q}(\mathbf{x})}} \right)^2. \quad (3)$$

Let us now compute separately the expectation with respect to \mathbb{P}_n of the two terms in the right-hand side of (3). For the first term we have

$$\begin{aligned} \mathbb{E}_n(\mathbb{E}_{n+q}(\alpha(\mathbf{x})^2)) &= \mathbb{E}_n(\alpha(\mathbf{x})^2) = \mathbb{E}_n \left(\Phi \left(\frac{\xi(\mathbf{x}) - z^{\text{crit}}}{\sqrt{\lambda(\mathbf{x})}} \right)^2 \right) \\ &= \Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); r_n(\mathbf{x})), \end{aligned} \quad (4)$$

where we have applied the second part of Corollary 3 with $m = m_n(\mathbf{x})$, $v = k_n(\mathbf{x}, \mathbf{x})$, $m' = z^{\text{crit}}$, and $v' = \lambda(\mathbf{x})$. For the second term we observe that, under \mathbb{P}_n , m_{n+q} is a Gaussian process with mean m_n and covariance function $\nu_n(\cdot, \cdot; \tilde{\mathbf{x}})$. Therefore $m_{n+q}(\mathbf{x}) \sim \mathcal{N}(m_n(\mathbf{x}), \nu(\mathbf{x}, \mathbf{x}; \tilde{\mathbf{x}}))$, and it follows that

$$\mathbb{E}_n \left(\Phi \left(\frac{m_{n+q}(\mathbf{x}) - z^{\text{crit}}}{\sqrt{v_{n+q}(\mathbf{x})}} \right)^2 \right) = \Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); \tilde{r}_n(\mathbf{x})), \quad (5)$$

where we have used again the second part of Corollary 3 with $m = m_n(\mathbf{x})$, $v = \nu_n(\mathbf{x}, \mathbf{x}; \tilde{\mathbf{x}})$, $m' = z^{\text{crit}}$ and $v' = v_{n+q}(\mathbf{x})$. Indeed,

$$\begin{aligned} v + v' &= \nu_n(\mathbf{x}, \mathbf{x}; \tilde{\mathbf{x}}) + v_{n+q}(\mathbf{x}) \\ &= (k_n(\mathbf{x}, \mathbf{x}) - k_{n+q}(\mathbf{x}, \mathbf{x})) + (k_{n+q}(\mathbf{x}, \mathbf{x}) + \lambda(\mathbf{x})) = v_n(\mathbf{x}), \end{aligned}$$

therefore

$$\begin{aligned} \frac{m - m'}{\sqrt{v + v'}} &= \frac{m_n(\mathbf{x}) - z^{\text{crit}}}{\sqrt{v_n}} = a_n(\mathbf{x}), \\ \frac{v}{v + v'} &= \frac{\nu_n(\mathbf{x}, \mathbf{x}; \tilde{\mathbf{x}})}{v_n(\mathbf{x})} = \tilde{r}_n(\mathbf{x}). \end{aligned}$$

Combining (3)–(5) and integrating on \mathbb{X} with respect to μ yields the desired expression for $J_n(\mathbf{x})$. Similarly, combining (2) with $q = 0$ and (4) we have

$$H_n = \int \left(\Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); r_n(\mathbf{x})) - \Phi(a_n(\mathbf{x}))^2 \right) \mu(d\mathbf{x})$$

and the expression of $G_n(\tilde{\mathbf{x}})$ follows. □

D Additional information regarding the examples

D.1 One-dimensional example

This section provides additional information regarding the Bayesian model used in the “One-dimensional example” presented in Section 4.2 of the article.

The model used in this example is the one proposed by Kennedy and O’Hagan (2000) and reviewed in Section 2.1 of the article, with $S = 2$ levels. The two independent Gaussian processes η_1 and η_2 are stationary processes with unknown constant means and Matérn covariance functions with regularity $5/2$:

$$\eta_s \sim \text{GP} \left(m_s, \sigma_s^2 \mathcal{M}_{5/2}(a_s(\cdot - \cdot)) \right), \quad s \in \{1, 2\}.$$

Independent priors are used for all the remaining hyper-parameters of the model:

- Improper uniform prior distributions on \mathbb{R} are used for the means m_s .
- The parameters of the covariance function follow log-normal distributions:

$$\begin{aligned} \log(\sigma_s^2) &\sim \mathcal{N}(2 \log(0.2), \log(100)^2), \\ \log(a_s) &\sim \mathcal{N}(\log(2), \log(10)^2). \end{aligned}$$

- Finally, the regression term between the two levels follows a normal prior distribution:

$$\rho \sim \mathcal{N}(1, 2^2).$$

D.2 Random damped harmonic oscillator

This section provides additional information regarding the explicit exponential Euler scheme used in the “Random damped harmonic oscillator” example (Section 4.3 of the article).

Consider a stochastic differential equation

$$d\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_t dt + \mathbf{b} dB_t,$$

where (\mathbf{Y}_t) is a vector-valued stochastic process, (B_t) a Wiener process, \mathbf{b} a vector and \mathbf{A} a matrix. Let δ a time step, we would like to approximate (\mathbf{Y}_t) by a finite difference method: $\mathbf{Y}_k^{(\delta)} \approx \mathbf{Y}_{k\delta}$. The explicit exponential Euler scheme is a finite difference method proposed by Jentzen and Kloeden (2009) to ensure stability when approximating a stochastic equation. The method is to apply recursively the formula

$$\mathbf{Y}_{k+1}^{(\delta)} = \exp(\mathbf{A}\delta) \left[\mathbf{Y}_k^{(\delta)} + \sqrt{2\pi S\delta} \cdot \mathbf{b} \cdot U \right]$$

with S the spectral intensity of the Brownian motion, and U a normal random vector.

In particular, for the application of Section 4.2, we have

$$d \begin{pmatrix} Y_t \\ \dot{Y}_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -2\zeta\omega_0 \end{pmatrix} \begin{pmatrix} Y_t \\ \dot{Y}_t \end{pmatrix} dt + \begin{pmatrix} 0 \\ 1 \end{pmatrix} dB_t.$$

Consequently, the approximation with a finite time-step δ is

$$\begin{pmatrix} Y_{k+1}^{(\delta)} \\ \dot{Y}_{k+1}^{(\delta)} \end{pmatrix} = \exp \begin{pmatrix} 0 & \delta \\ -\omega_0^2\delta & -2\zeta\omega_0\delta \end{pmatrix} \left[\begin{pmatrix} Y_k^{(\delta)} \\ \dot{Y}_k^{(\delta)} \end{pmatrix} + \begin{pmatrix} 0 \\ \sqrt{2\pi S\delta}u \end{pmatrix} \right], \quad u \sim \mathcal{N}(0, 1).$$

References

- Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.
- Forrester, A. I. J., Sóbester, A., and Keane, A. J. (2007). Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 463(2088):3251–3269.
- Guo, B., Chen, X.-P., and Liu, M.-Q. (2017). Construction of latin hypercube designs with nested and sliced structures. *Statistical Papers*, pages 1–14.
- He, X. and Qian, P. Z. G. (2011). Nested orthogonal array-based latin hypercube designs. *Biometrika*, 98(3):721–731.

- Jentzen, A. and Kloeden, P. E. (2009). Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space–time noise. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 465(2102):649–667.
- Kennedy, M. C. and O’Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13.
- Le Gratiet, L. and Garnier, J. (2014). Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5):365–386.
- Qian, P. Z. G. (2009). Nested latin hypercube designs. *Biometrika*, 96(4):957–970.
- Rennen, G., Husslage, B., Van Dam, E. R., and Den Hertog, D. (2010). Nested maximin latin hypercube designs. *Structural and Multidisciplinary Optimization*, 41(3):371–395.
- Xu, J., Duan, X., Wang, Z., and Yan, L. (2017). A general construction for nested latin hypercube designs. *Statistics & Probability Letters*.
- Yang, J., Liu, M.-Q., and Lin, D. K. J. (2014). Construction of nested orthogonal latin hypercube designs. *Statistica Sinica*, 24(1):211–219.