

Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set

Clément Chevalier, IMSV, University of Bern

Julien Bect, Supelec

David Ginsbourger, IMSV, University of Bern

Emmanuel Vazquez, Supelec

Victor Picheny, INRA

Yann Richet, IRSN

Abstract

Stepwise Uncertainty Reduction (SUR) strategies aim at constructing a sequence of points for evaluating a function f in such a way that the residual uncertainty about a quantity of interest progressively decreases to zero. Using such strategies in the framework of Gaussian process modeling has been shown to be efficient for estimating the volume of excursion of f above a fixed threshold. However, SUR strategies remain cumbersome to use in practice because of their high computational complexity, and the fact that they deliver a single point at each iteration. In this paper we introduce several multi-point sampling criteria, allowing the selection of batches of points at which f can be evaluated in parallel. Such criteria are of particular interest when f is costly to evaluate and several CPUs are simultaneously available. We also manage to drastically reduce the computational cost of these strategies through the use of closed form formulas. We illustrate their performances in various numerical experiments, including a nuclear safety test case. Basic notions about kriging, auxiliary problems, complexity calculations, R code and datas are available online as supplementary materials.

Keywords: Design of computer experiments, Gaussian process models, Batch-sequential strategies, Probability of failure, Active learning, Inversion

1 Introduction

Whether in natural sciences, engineering, or economics, the study of complex phenomena is increasingly relying on numerical simulations. From an end user's perspective, a numerical simulator can often be considered as a black box taking a number of real-valued parameters as inputs and returning one or several quantities of interest after a post-processing stage. Formally, the space of inputs is a set $\mathbb{X} \subset \mathbb{R}^d$ and the simulator can be viewed as a function $f : \mathbb{X} \rightarrow \mathbb{R}$ that maps the inputs to a cost or a performance indicator. In many practical applications, the objective is to obtain information about the simulator from a number of runs, or, in other words, to infer a quantity of interest from a number of evaluations of f . A problem that is often at stake is the estimation of the probability that a cost exceeds a given threshold. This problem corresponds to the estimation of the volume α^* of the excursion set $\Gamma^* = \{x \in \mathbb{X} : f(x) \geq T\}$, with T a given threshold, under a measure $\mathbb{P}_{\mathbb{X}}$ on \mathbb{X} . In safety analysis, $\mathbb{P}_{\mathbb{X}}$ typically models the uncertainty on input parameters. If f is expensive to evaluate, the estimation of α^* must be performed with a limited number of evaluations of f , which naturally excludes brute-force approaches like crude Monte Carlo sampling.

A popular approach is to construct a response surface (also known as surrogate or meta-model) based on available evaluations of f , together with an uncertainty measure about this surface. Using this uncertainty measure is one of the key concepts in the design and analysis of computer experiments (see, e.g., Santner et al., 2003; Fang et al., 2006; Bayarri et al., 2007; Forrester et al., 2008; Gramacy and Lee, 2009, and references therein). It has been found to be a convenient and powerful tool, providing efficient answers to the issues of designing experiments (Sacks et al., 1989b) or global optimization (Jones et al., 1998) for instance.

For the problem of estimating a probability of failure, several sampling strategies based on a kriging metamodel have already been proposed (see Bect et al., 2012, for a review). Note that some of these strategies were initially designed to estimate the boundary of the excursion set (and not its volume) but, as these problems are quite close, we expect these criteria to have fairly good performances for the problem of estimating a probability of failure. The sampling criteria proposed by Ranjan et al. (2008), Bichon et al. (2008) and Echard et al. (2010) consist of heuristic modifications of the famous Expected Improvement criterion of Jones et al.

(1998). They compute a *pointwise* trade-off between predicted closeness to the threshold T , and high prediction uncertainty. In contrast, Stepwise Uncertainty Reduction (SUR) strategies (Vazquez and Bect, 2009; Bect et al., 2012) rely on global measures of uncertainty about the excursion set Γ^* and take into account the important fact that sampling at a point x also brings useful information on the behaviour of f in a neighbourhood of x . Numerical experiments (reported by Bect et al., 2012) showed that SUR criteria widely outperform pointwise criteria in terms of quickly estimating the true volume of excursion α^* .

Perhaps the most natural SUR sampling criterion, for the problem of estimating a probability of failure, is the expected posterior variance of the volume of the *random* excursion set $\Gamma = \{x \in \mathbb{X} : \xi(x) \geq T\}$, where ξ is a Gaussian process modeling our current (prior) knowledge about f . This criterion has been considered impractical in previous publications (Vazquez and Bect, 2009; Bect et al., 2012), since its computation *seems* to require conditional simulations of the Gaussian process ξ , which are very expensive. Alternative SUR strategies were proposed instead: in short, they consist of defining a measure of uncertainty dedicated to the problem at hand, and then sampling sequentially at the location that will reduce the most, in expectation, this uncertainty.

An example of application of a SUR strategy is shown on Figure 1. Here a simulator f calculates whether a storage facility of plutonium powder presents risks of nuclear chain reactions or not, as a function of two variables, the mass and the concentration of Plutonium (see Section 4.2). A sequential sampling of this 2-dimensional “expensive” function, using a SUR strategy, manages to identify with few evaluations the set of “dangerous” configurations.

Despite their very good performance in applications, SUR strategies still have important drawbacks. Computing the value of a SUR criterion at a single point $x_{n+1} \in \mathbb{X}$ is indeed computationally demanding since it relies on numerical integration. Besides, these strategies were designed to sample one point at a time while practitioners often have the capacity to run $r > 1$ simulations in parallel. The high numerical complexity of simply computing the value of a sampling criterion at one point suggests why, despite their very good performances on numerical experiments, SUR strategies based on kriging are not yet widely used by practitioners for the problem of estimating a probability of failure.

In this paper, we bring new solutions to the issues mentioned above. We first introduce

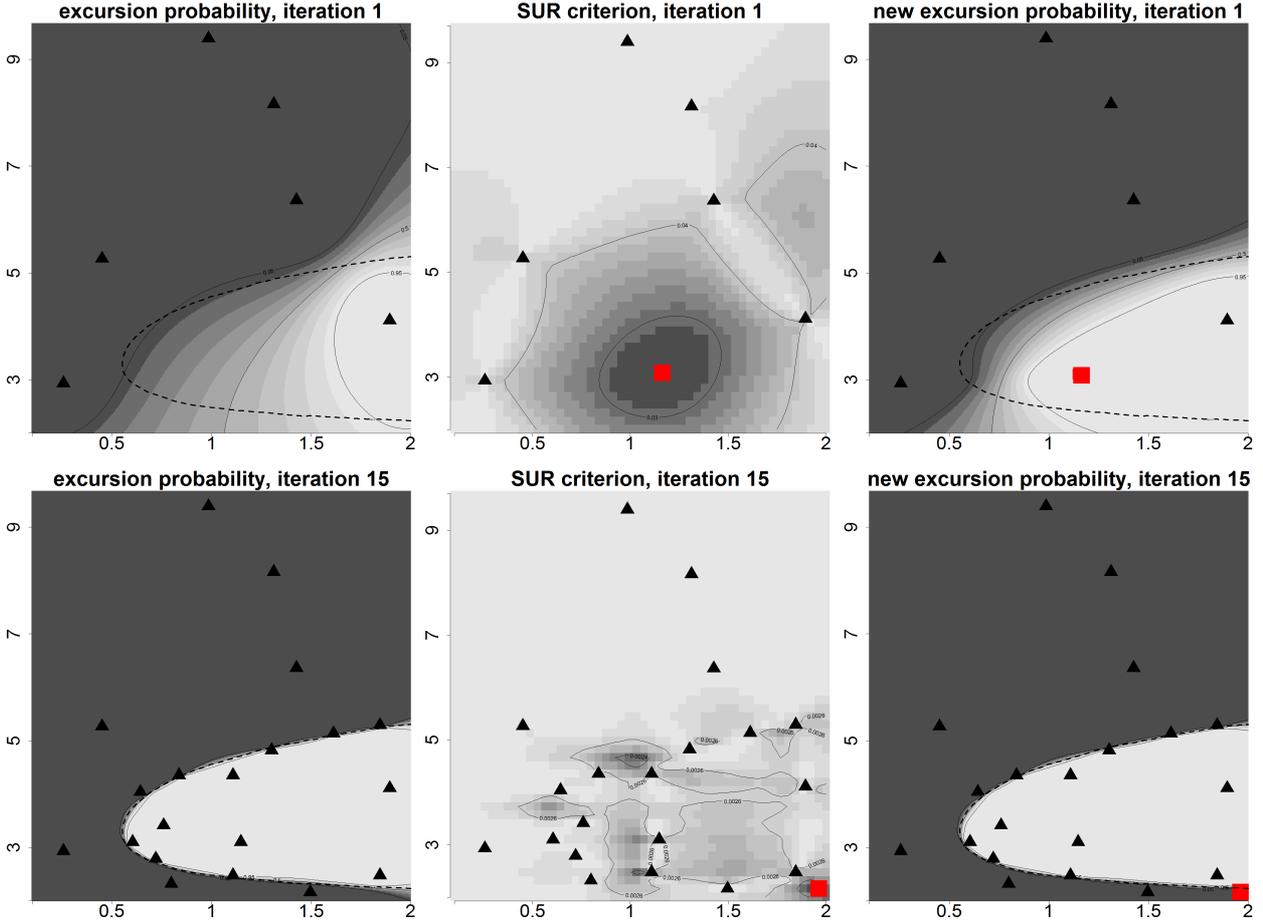


Figure 1: SUR strategy (first and last iteration) applied to a nuclear criticality safety simulator. The black triangles stand for the design points at the current iteration. The square is the point sampled using the SUR criterion. Areas in black (resp. white) correspond to excursion probabilities near 0 (resp. 1). The dotted line indicates a fine approximation of the true excursion set’s boundary.

new *multi-point* SUR sampling criteria and provide methods and algorithms to run them in a reasonable time. In particular, we show that the (one step look-ahead) expected posterior variance criterion, presented as unaffordable in Bect et al. (2012) can be computed quickly, without simulating any Gaussian process realizations.

The paper is organised as follows: Section 2 introduces notations and gives two examples of SUR criteria with their new multi-point versions. The theoretical basis of our methods to quickly compute the criteria are detailed in Section 3 and our new algorithms are tested in Section 4 on different test cases, including a nuclear safety application. For the sake of brevity, basic notions about kriging, details about the choice of integrations points and

detailed complexity calculations are provided as Supplementary Material.

2 Kriging-based Stepwise Uncertainty Reduction

A Stepwise Uncertainty Reduction (SUR) strategy seeks to construct a sequence X_1, X_2, \dots of evaluation points of f so as to quickly reduce the residual uncertainty about a quantity of interest given the information provided by the evaluation results. More precisely, SUR strategies are based on three main ideas. The first (Bayesian) idea is to consider f as a sample path of a random process ξ , assumed to be Gaussian for the sake of tractability. The second idea is to introduce a measure of the uncertainty about the quantity of interest conditioned on the observations at time n , or, more precisely, on the σ -algebra \mathcal{A}_n generated by $\{(X_i, \xi(X_i)), 1 \leq i \leq n\}$. We will denote by H_n such a measure of uncertainty, which is an \mathcal{A}_n -measurable random variable. The third idea is to choose evaluation points sequentially in order to minimize, at each step n , the *expected* value of the future uncertainty measure H_{n+1} with respect to the random outcomes of the new evaluation of ξ :

$$X_{n+1} \in \operatorname{argmin}_{x_{n+1} \in \mathbb{X}} J_n(x_{n+1}), \quad (1)$$

where

$$J_n(x_{n+1}) := \mathbf{E}_n \left(H_{n+1} \mid X_{n+1} = x_{n+1} \right), \quad (2)$$

and $\mathbf{E}_n(\cdot)$ stands for the conditional expectation $\mathbf{E}(\cdot \mid \mathcal{A}_n)$.

Depending on the definition given to the measure of uncertainty, many SUR strategies can be designed in order to infer any quantity of interest. For the question of estimating a probability of failure, two SUR strategies are presented in this section.

Example 1: criterion $J_n^{(\alpha)}$. Recall that we denote by Γ the random excursion set $\{x \in \mathbb{X} : \xi(x) \geq T\}$ and α its volume, $\alpha = \mathbf{P}_{\mathbb{X}}(\Gamma)$. The conditional variance $\operatorname{Var}_n(\alpha)$ of α is a natural choice for H_n to quantify the (residual) uncertainty about α^* given \mathcal{A}_n . In the rest of the paper, we denote this uncertainty measure by $H_n^{(\alpha)}$. A possible SUR strategy to estimate α^* would consist, at step n , of choosing as next evaluation point a minimizer of the criterion:

$$J_n^{(\alpha)}(x_{n+1}) := \mathbf{E}_n \left(\operatorname{Var}_{n+1}(\alpha) \mid X_{n+1} = x_{n+1} \right) \quad (3)$$

A quite natural multi-point extension of this criterion is now introduced. The following criterion depends on $r > 0$ points $(x_{n+1}, \dots, x_{n+r}) \in \mathbb{X}^r$:

$$J_n^{(\alpha)}(x_{n+1}, \dots, x_{n+r}) := \mathbf{E}_n \left(\text{Var}_{n+r}(\alpha) \mid X_{n+1} = x_{n+1}, \dots, X_{n+r} = x_{n+r} \right). \quad (4)$$

Note that this criterion is considered intractable in Bect et al. (2012) for $r = 1$ because its computation has a very high numerical complexity (it apparently requires the simulation of a large number of Gaussian process realizations). We will see in the next sections that both one-point and multi-point versions of this criterion can be computed quickly and used in applications.

Example 2: criterion $J_n^{(\Gamma)}$. The excursion set can be characterized by the random function $\mathbb{1}_{\{\xi > T\}}$. For $x \in \mathbb{X}$, the random variable $\mathbb{1}_{\{\xi(x) > T\}}$ has conditional expectation

$$p_n(x) := \mathbf{E}_n \mathbb{1}_{\{\xi(x) > T\}} = \mathbf{P}(\xi(x) > T \mid \mathcal{A}_n) = \Phi \left(\frac{m_n(x) - T}{s_n(x)} \right),$$

where $m_n(x)$ and $s_n(x)$ are the kriging mean and variance at point x at time n and Φ denotes the cumulative distribution function (c.d.f.) of the standard Gaussian distribution. See the Supplementary Material for a brief reminder about kriging and the notation used throughout the paper. The random variable $\mathbb{1}_{\{\xi(x) > T\}}$ has conditional variance $p_n(x)(1 - p_n(x))$, so that $\int_{\mathbb{X}} p_n(1 - p_n) d\mathbf{P}_{\mathbb{X}}$ can serve as a measure of global uncertainty about α^* . We denote this uncertainty measure by $H_n^{(\Gamma)}$, and the corresponding SUR sampling criterion is

$$J_n^{(\Gamma)}(x_{n+1}) := \mathbf{E}_n \left(\int_{\mathbb{X}} p_{n+1}(1 - p_{n+1}) d\mathbf{P}_{\mathbb{X}} \mid X_{n+1} = x_{n+1} \right). \quad (5)$$

This criterion was first introduced by Bect et al. (2012). Again, a natural extension is the following new multi-point criterion:

$$J_n^{(\Gamma)}(x_{n+1}, \dots, x_{n+r}) = \mathbf{E}_n \left(\int_{\mathbb{X}} p_{n+r}(1 - p_{n+r}) d\mathbf{P}_{\mathbb{X}} \mid X_{n+1} = x_{n+1}, \dots, X_{n+r} = x_{n+r} \right). \quad (6)$$

In Bect et al. (2012), the numerical computation of $J_n^{(\Gamma)}$ in (6) is considered only for $r = 1$ and is based on quadrature formulas written as

$$J_n^{(\Gamma)}(x_{n+1}) \approx \frac{1}{M} \sum_{q=1}^Q \sum_{m=1}^M w^{(q)} v_{n+1}(x^{(m)}; x_{n+1}, y_{n+1}^{(q)}), \quad (7)$$

where Q is the number of points used to approximate the conditional expectation with respect to the random outcome of the evaluation at x_{n+1} , which has a $\mathcal{N}(m_n(x_{n+1}), s_n^2(x_{n+1}))$

distribution, M is the number of points used to obtain a Monte-Carlo approximation of $H_{n+1}^{(\Gamma)}$, $v_{n+1} := p_{n+1}(1 - p_{n+1})$. The $x^{(m)}$'s are i.i.d. according to $\mathbf{P}_{\mathbb{X}}$, $(y_{n+1}^{(1)}, \dots, y_{n+1}^{(Q)})$ and $(w^{(1)}, \dots, w^{(Q)})$ stand for the quadrature points and quadrature weights of the Gauss-Hermite quadrature. Here the computation of $v_{n+1}(x^{(m)}; x_{n+1}, y_{n+1}^{(q)})$ in (7) involves the calculation of the kriging mean and the kriging variance at $x^{(m)}$ from the evaluations of ξ at X_1, \dots, X_n and x_{n+1} . It follows that the computation of $J_n^{(\Gamma)}$ at one point has a $O(n^3 + Mn^2 + MQ)$ complexity. Additional details about algorithmic complexities are provided in the Supplementary Material. Since we need to evaluate $J_n^{(\Gamma)}$ several times to carry out the minimization in (1), the computational cost of this SUR sampling strategy with an implementation relying on (7) can be very large.

The problem becomes even more difficult for $r > 1$, which requires a higher value for Q . Indeed, when $r > 1$, we have to approximate a conditional expectation with respect to the random outcome of the Gaussian vector $(\xi(x_{n+1}), \dots, \xi(x_{n+r}))^\top$, which requires a discretization of an integral over \mathbb{R}^r . As a consequence, the complexity to compute the multi-point SUR criterion presented above is expected to rise quickly with r , which makes it impractical even for moderate values of r .

The next section develops useful properties allowing to circumvent these issues. In particular, new analytical formulas allow us to eliminate of the cumbersome integral over \mathbb{R}^r and make it possible to compute efficiently both one-point and multi-point criteria.

3 Efficient calculation of multi-point SUR criteria

In this section, we provide new expressions allowing efficient computation of the two multi-point SUR criteria introduced in the previous section.

3.1 Criterion $J_n^{(\Gamma)}$

As explained in the previous sections, the proposed multi-point criterion $J_n^{(\Gamma)}$ is the conditional expectation given \mathcal{A}_n of the future uncertainty measure $H_{n+r}^{(\Gamma)}$, assuming that the objective function will be evaluated at r new points. Such a future uncertainty is an \mathcal{A}_{n+r} -measurable random variable, meaning that the computation of its conditional expectation given \mathcal{A}_n requires the discretization of an integral over \mathbb{R}^r . It turns out that the complexity

for computing $J_n^{(\Gamma)}$ can be drastically reduced, using the new analytical expressions given below.

Proposition 1.

$$J_n^{(\Gamma)}(x_{n+1}, \dots, x_{n+r}) = \int_{\mathbb{X}} \Phi_2 \left(\begin{pmatrix} a(x) \\ -a(x) \end{pmatrix}, \begin{pmatrix} c(x) & 1 - c(x) \\ 1 - c(x) & c(x) \end{pmatrix} \right) \mathbf{P}_{\mathbb{X}}(dx), \quad (8)$$

where:

- $\Phi_2(\cdot, M)$ is the c.d.f. of the centered bivariate Gaussian with covariance matrix M
- $a(x) := (m_n(x) - T)/s_{n+r}(x)$,
- $\mathbf{b}(x) := \frac{1}{s_{n+r}(x)} \boldsymbol{\Sigma}^{-1} (k_n(x, x_{n+1}), \dots, k_n(x, x_{n+r}))^\top$,
- $c(x) := 1 + \mathbf{b}(x)^\top \boldsymbol{\Sigma} \mathbf{b}(x) = s_n^2(x)/s_{n+r}^2(x)$,
- $\boldsymbol{\Sigma}$ is the $r \times r$ covariance matrix of $(\xi(x_{n+1}), \dots, \xi(x_{n+r}))^\top$ conditional on \mathcal{A}_n .

Proof. First, an interchange of integral and expectation (Fubini-Tonelli theorem) yields

$$J_n^{(\Gamma)}(x_{n+1}, \dots, x_{n+r}) = \int_{\mathbb{X}} \mathbf{E}_n(p_{n+r}(x)(1 - p_{n+r}(x))) \mathbf{P}_{\mathbb{X}}(dx), \quad (9)$$

where the conditioning on $X_{n+i} = x_{n+i}$'s is not explicitly reproduced, to simplify notation. Now, using the kriging update formula (see, e.g., Barnes and Watson (1992); Gao et al. (1996); Emery (2009), as well as Chevalier et al. (2013a)), we obtain:

$$m_{n+r}(x) = m_n(x) + (k_n(x, x_{n+1}), \dots, k_n(x, x_{n+r})) \boldsymbol{\Sigma}^{-1} \mathbf{y}_{\text{centered}}, \quad (10)$$

where $\mathbf{y}_{\text{centered}} := (\xi(x_{n+1}) - m_n(x_{n+1}), \dots, \xi(x_{n+r}) - m_n(x_{n+r}))^\top$, so that

$$p_{n+r}(x) = \Phi \left(a(x) + \mathbf{b}(x)^\top \mathbf{y}_{\text{centered}} \right). \quad (11)$$

A plug-in of (10) in the integrand of (6) gives:

$$\mathbf{E}_n(p_{n+r}(x)(1 - p_{n+r}(x))) = \int_{\mathbb{R}^r} \Phi(a(x) + \mathbf{b}(x)^\top \mathbf{u}) \Phi(-a(x) - \mathbf{b}(x)^\top \mathbf{u}) \Psi(\mathbf{u}) d\mathbf{u}, \quad (12)$$

where Ψ is the $\mathcal{N}(\mathbf{0}, \Sigma)$ density of $\mathbf{y}_{\text{centered}}$ knowing \mathcal{A}_n . By definition of Φ , we then get

$$\begin{aligned} \mathbb{E}_n(p_{n+r}(x)(1 - p_{n+r}(x))) &= P_n(N_1 < a(x) + \mathbf{b}(x)^\top \mathbf{y}_{\text{centered}}, N_2 < -a(x) - \mathbf{b}(x)^\top \mathbf{y}_{\text{centered}}) \\ &= P_n(N_1 - \mathbf{b}(x)^\top \mathbf{y}_{\text{centered}} < a(x), N_2 + \mathbf{b}(x)^\top \mathbf{y}_{\text{centered}} < -a(x)), \end{aligned}$$

where $(N_1, N_2)^\top \sim \mathcal{N}(\mathbf{0}, I_2)$ independently of $\mathbf{y}_{\text{centered}}$ and $P_n(\cdot) := P(\cdot | \mathcal{A}_n)$. Finally, $N_1 - \mathbf{b}(x)^\top \mathbf{y}_{\text{centered}}$ and $N_2 + \mathbf{b}(x)^\top \mathbf{y}_{\text{centered}}$ form a Gaussian couple with componentwise variances equal to $c(x)$ and covariance $1 - c(x)$, so that stated result directly follows by integration over \mathbb{X} . \square

Remark 1. *In Proposition 1, we managed to eliminate an integral over \mathbb{R}^r . Moreover, the given formula is “exact” in the sense that we no longer have to compute an estimate (relying on quadrature points) of an integral over \mathbb{R}^r . Besides, the computation of $J_n^{(\Gamma)}$ is now available for $r > 1$ at a cost that does not quickly increase with r . For n observations and M discretization points for the integral over \mathbb{X} , the complexity to compute $J_n^{(\Gamma)}$ for one batch of r points is mainly of $O(rMn)$ if we assume that $r \ll n \ll M$ (which is often the case in practice) and that some quantities have been pre-computed (see algorithms in the Supplementary Material for more details). This means that the complexity is roughly linear in r , which ensures that batches with large r values can be used in applications. The criterion is now easier to compute than multi-point criteria dedicated to optimization, like the multi-point Expected Improvement (Ginsbourger et al., 2010; Chevalier and Ginsbourger, 2013).*

Remark 2. *When the integral over \mathbb{X} is discretized based on M integration points, the computation of the $J_n^{(\Gamma)}$ criterion requires the calculation of the updated kriging variance $s_{n+r}^2(x)$ for each of the M points. The updated kriging variance can be efficiently calculated using a kriging variance update formula given in Chevalier et al. (2013a).*

Remark 3. *By reducing (6) to (8), we achieved to reduce the integral over \mathbb{R}^r to an integral over \mathbb{R}^2 (Φ_2). Moreover, although calculating Φ_2 is not trivial, this bivariate integral is standard, and very efficient numerical procedures to compute it are available. For instance, Genz (1992) wrote routines in Fortran77 which have been included in many R packages (e.g., in `mnormt`, `pbivnorm`, `mvtnorm`, all available on CRAN).*

3.2 Criterion $J_n^{(\alpha)}$

In the kriging framework and conditional on \mathcal{A}_n , the expectation of the volume of excursion α is given by $\hat{\alpha} := \int_{\mathbb{X}} p_n d\mathbb{P}_{\mathbb{X}}$. As explained in Section 2, the conditional variance $\text{Var}_n(\alpha)$ of α given \mathcal{A}_n is a very natural choice to quantify the uncertainty about α but, even for $r = 1$, it has been considered intractable. In fact, with the help of the kriging update formulas (See (10)) and the calculation schemes introduced in the proof of Proposition 1, we will now show that this criterion can be expressed in a numerically tractable form, for both one-point and multi-point versions.

Proposition 2.

$$J_n^{(\alpha)}(x_{n+1}, \dots, x_{n+r}) = \gamma_n - \int_{\mathbb{X} \times \mathbb{X}} \Phi_2 \left(\begin{pmatrix} a(z_1) \\ a(z_2) \end{pmatrix}, \begin{pmatrix} c(z_1) & d(z_1, z_2) \\ d(z_1, z_2) & c(z_2) \end{pmatrix} \right) \mathbb{P}_{\mathbb{X}}(dz_1) \mathbb{P}_{\mathbb{X}}(dz_2), \quad (13)$$

where

- Φ_2, a, \mathbf{b}, c and Σ are defined as in Proposition 1,
- $d(z_1, z_2) := \mathbf{b}(z_1)^\top \Sigma \mathbf{b}(z_2)$,
- γ_n is a constant, in the sense that it does not depend on $(x_{n+1}, \dots, x_{n+r})$.

Proof. Omitting again the conditioning on the $X_{n+i} = x_{n+i}$'s in the notations, we have:

$$\begin{aligned} J_n^{(\alpha)}(x_{n+1}, \dots, x_{n+r}) &:= \mathbb{E}_n \left(\text{Var}_{n+r}(\alpha) \mid X_{n+1} = x_{n+1}, \dots, X_{n+r} = x_{n+r} \right) \\ &= \mathbb{E}_n \left(\mathbb{E}_{n+r} \left(\int_{\mathbb{X}} (\mathbb{1}_{\{\xi(z) > T\}} - p_{n+r}(z)) \mathbb{P}_{\mathbb{X}}(dz) \right)^2 \right) \\ &= \mathbb{E}_n \left(\mathbb{E}_{n+r} \iint_{\mathbb{X} \times \mathbb{X}} (\mathbb{1}_{\{\xi(z_1) > T\}} - p_{n+r}(z_1)) (\mathbb{1}_{\{\xi(z_2) > T\}} - p_{n+r}(z_2)) \mathbb{P}_{\mathbb{X}}(dz_1) \mathbb{P}_{\mathbb{X}}(dz_2) \right) \\ &= \mathbb{E}_n \left(\iint_{\mathbb{X} \times \mathbb{X}} (\mathbb{E}_{n+r}(\mathbb{1}_{\{\xi(z_1) > T\}} \mathbb{1}_{\{\xi(z_2) > T\}}) - p_{n+r}(z_1) p_{n+r}(z_2)) \mathbb{P}_{\mathbb{X}}(dz_1) \mathbb{P}_{\mathbb{X}}(dz_2) \right) \end{aligned}$$

By applying the law of total expectation, we see that, for any $(z_1, z_2) \in \mathbb{X}^2$:

$$\mathbb{E}_n(\mathbb{E}_{n+r}(\mathbb{1}_{\{\xi(z_1) > T\}} \mathbb{1}_{\{\xi(z_2) > T\}})) = \mathbb{E}_n(\mathbb{1}_{\{\xi(z_1) > T\}} \mathbb{1}_{\{\xi(z_2) > T\}}) = P(\xi(z_1) > T, \xi(z_2) > T \mid \mathcal{A}_n)$$

Thus, this quantity does not depend on the choice of the r points $(x_{n+1}, \dots, x_{n+r})$. Writing $\gamma_n := \iint_{\mathbb{X} \times \mathbb{X}} P(\xi(z_1) > T, \xi(z_2) > T | \mathcal{A}_n) P_{\mathbb{X}}(dz_1) P_{\mathbb{X}}(dz_2)$, $J_n^{(\alpha)}$ simplifies to

$$J_n^{(\alpha)}(x_{n+1}, \dots, x_{n+r}) = \gamma_n - \iint_{\mathbb{X} \times \mathbb{X}} \mathbf{E}_n(p_{n+r}(z_1)p_{n+r}(z_2)) P_{\mathbb{X}}(dz_1) P_{\mathbb{X}}(dz_2).$$

The end result is obtained using similar calculations as in the proof of Proposition 1. □

Remark 4. *This new expression is very similar to the expression found in Proposition 1 and can be computed with the same complexity. However, in practice, the number of integration points M has to be higher because the domain to be discretized is $\mathbb{X} \times \mathbb{X}$. In the examples of Section 4, we use importance sampling techniques to choose these M integration points. In Section 4.1, we empirically demonstrate that using M^2 integration points to compute the $J_n^{(\alpha)}$ criterion and using M points to compute $J_n^{(\Gamma)}$ yields comparable performances for estimating the true volume of excursion in the case where the unknown function is actually a Gaussian process realization.*

The calculation of the $J_n^{(\Gamma)}$ and $J_n^{(\alpha)}$ criteria based on Proposition 1 and 2 requires knowledge of the covariance function k . When it is only assumed that $k = k_\theta$ belongs to some parametric family of covariance functions, a common approach consists of plugging in the maximum likelihood (or maximum a posteriori) estimate of the parameter vector θ , and then re-estimating this parameter at each iteration. This plug-in approach has been used previously for contour estimation problems (see, e.g., Ranjan et al. (2008); Picheny et al. (2010)).

The formula proposed in Propositions 1 and 2 are also perfectly compatible with a fully Bayesian approach. Indeed, (2) can be rewritten as

$$J_n(x_{n+1}) = \int \mathbf{E}_n(H_{n+1} \mid X_{n+1} = x_{n+1}, \theta = t) \pi_n(dt),$$

where π_n denotes the posterior distribution of θ . The value of the inner expectation can be obtained, as in the plug-in approach, using either (8) or (13). The outer integral, however, usually admits no closed-form expression, but can be approximated using for instance a Sequential Monte Carlo (SMC) method (as in, e.g., Gramacy and Polson (2011); Benassi et al. (2011)). Fully Bayesian versions of the SUR criteria thus come with an increased

computational cost, with respect to plug-in approaches, but are expected to provide a better assessment of the actual uncertainties when only a small number of evaluations is available. In particular, the uncertainty H_n and the expectation of the future uncertainty $J_n(\cdot)$ might be underestimated with the plug-in method. The fully Bayesian approach has not been implemented in the context of this research, and will be the topic of future work; a maximum-likelihood-based plug-in approach will be used in Section 4.

4 Applications

In this section, we illustrate our sequential and batch-sequential strategies on three test cases: (i) simulated realizations of a two-dimensional Gaussian process; (ii) a two-dimensional nuclear safety case study; and (iii) two test functions for which the excursion set is not easy to identify. All tests are realized using the KrigInv R package (Chevalier et al., 2013b), offering a simple access to the previously described one-point and multi-point criteria. Throughout this section, we use a separable Matérn covariance with parameter $\nu = 3/2$. In Section 4.1, the parameters of this covariance kernel are assumed to be known, while in Sections 4.2 and 4.3 we plug in maximum likelihood estimates (renewed at each iteration) of the covariance parameters.

4.1 Benchmark on simulated Gaussian process realizations

The first goal of this section is to compare the *one-point* versions of the $J_n^{(\Gamma)}$ and $J_n^{(\alpha)}$ criteria. The test functions are 200 independent realizations of a two-dimensional Gaussian process indexed by $[0, 1]^2$. The covariance parameters for the kriging models are fixed equal to the actual ones of the Gaussian process. Besides comparing the two criteria, we want to estimate the effect of numerical integration errors on the global performance of the SUR strategies. The criterion $J_n^{(\alpha)}$ requires the computation of an integral over $\mathbb{X} \times \mathbb{X}$, so it is expected that the error will be higher than for the criterion $J_n^{(\Gamma)}$, which requires an integration over \mathbb{X} only. Therefore, as a rule of thumb, we use M integration points for $J_n^{(\Gamma)}$ and M^2 for $J_n^{(\alpha)}$.

For each Gaussian process realization, we adjust the threshold T in order to have a constant volume of excursion $\alpha^* = 0.2$. The volumes are calculated using 1000 reference points (obtained with a Sobol sequence), and thus exactly 200 points are in the excursion

set for each realization. The initial design consists of $n_0 = 12$ points using maximin Latin Hypercube Sampling (LHS), and a total of $n_1 = 40$ points are added to the design, using either $J_n^{(\Gamma)}$ or $J_n^{(\alpha)}$. For all realizations, the performance of both criteria are measured in term of the relative squared volume error $\text{SE} := (\hat{\alpha} - \alpha^*)^2/\alpha^{*2}$, where $\hat{\alpha}$ is the estimated volume, equal to the average probability of excursion of the reference points.

Two strategies are considered for numerical integration: first, we use $M = 50$ and 100 integration points to compute $J_n^{(\Gamma)}$ obtained using a Sobol sequence. In that case, the M^2 points used to compute $J_n^{(\alpha)}$ correspond to a $M \times M$ grid. We also test the use of $M = 50$ points chosen using a specific instrumental distribution (and renewed at each iteration) versus $M^2 = 2500$ points over $\mathbb{X} \times \mathbb{X}$ chosen using some other distribution on $\mathbb{X} \times \mathbb{X}$. In this last case, the M^2 points are not on a grid. Further details about the choice of the integration points are given in the Supplementary Material, Section B.

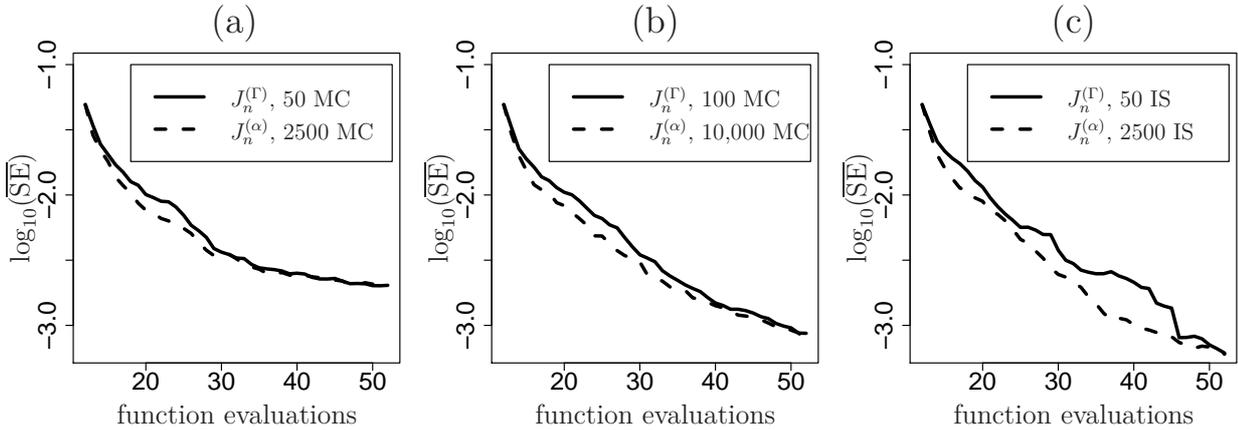


Figure 2: Benchmark simulation: performance (measured in term of mean squared relative error) of two SUR strategies based on the $J_n^{(\Gamma)}$ and $J_n^{(\alpha)}$ criteria, in function of the number of integration points and the method for choosing them (importance sampling or quasi Monte Carlo).

Figure 2 shows the evolution of $\overline{\text{SE}}$, the average of the SE values over the 200 realizations, as a function of the number of observations. The number of integration points has a direct impact on the performance of both SUR strategies (see, Figures 2(a)-2(b)), since the experiments corresponding to $M = 50$ with quasi-Monte Carlo sampling provide the worst results.

The closeness of pairs of lines in Figures 2(a)-2(c) suggests that the $J_n^{(\Gamma)}$ criterion with M integration points has roughly the same performance as the $J_n^{(\alpha)}$ criterion with M^2 integration

points. This suggests that, in high dimension, the criterion $J_n^{(\Gamma)}$ should be chosen since it requires a significantly lower computational effort.

Figure 2(c) indicates that the use of importance sampling (with a well chosen instrumental distribution) significantly improves the performance of these strategies, especially after a high number of iterations. Indeed, as the algorithm progresses, the criterion becomes more difficult to calculate with good accuracy as explained in the Supplementary Material, Section B. In that case, a clever choice of the integration points has a crucial impact on the global performance of the strategy.

The average computation time per run for $J_n^{(\Gamma)}$ in Figures 2(a)-2(c) are respectively 50, 65 and 72 seconds. For $J_n^{(\alpha)}$ computation times are 151, 393 and 211 seconds. These tests were performed on a computer with a 2.53GHz CPU and 3Gb of RAM.

From these numerical experiments, we can conclude that the criterion $J_n^{(\Gamma)}$ roughly achieves the same performances as $J_n^{(\alpha)}$ at a lower computational cost. Thus, in the next applications, we will mostly focus our attention on the $J_n^{(\Gamma)}$ criterion and its multi-point extension.

4.2 Nuclear safety test case

In this section, we illustrate a batch-sequential SUR strategy on an engineering problem, and provide a heuristic strategy for minimizing $J_n^{(\Gamma)}$ when the batch size r is large.

A system involving fissile materials may produce a chain reaction based on neutrons, which are both a product and an initiator of fission reactions. Nuclear criticality safety assessment aims at avoiding “criticality accidents” (overproduction of neutrons) within the range of operational conditions. In order to check subcriticality of a system, the neutrons multiplication factor, k_{eff} , is estimated using a costly simulator. In our case, the system is a storage facility of plutonium powder, whose k_{eff} depends on two input parameters: the mass of plutonium (MassePu) and the concentration of plutonium (logConcPu). We aim at finding the set of “dangerous” configurations $\{(\text{MassePu}, \text{logConcPu}) : k_{\text{eff}}(\text{MassePu}, \text{logConcPu}) > T\}$, where T is threshold fixed at 0.95. The main issue lies in the high cost to evaluate k_{eff} at a single configuration. Many CPU are simultaneously available, which means that our sampling strategy has to provide us, at each iteration, with a batch of $r > 1$ points at which to evaluate the simulator in parallel.

In this section, we run our algorithms on this two-dimensional problem. The $J_n^{(\Gamma)}$ sampling criterion is used with an initial design of experiment (a maximim LHS) of $n = 6$ points. The criterion is computed using $M = 600$ integration points renewed at each iteration, sampled from a specific instrumental distribution. See Supplementary Material, Section B, for additional details.

At each iteration, batches of $r = 4$ points are evaluated in parallel, and so a minimization problem of dimension $r \times d = 8$ needs to be solved. Direct optimization can be performed using classical optimal design tools, such as simulated annealing (Sacks and Schiller, 1988; Morris and Mitchell, 1995), genetic algorithms (Hamada et al., 2001) or subset selection (Gramacy and Lee, 2009). However, when $r \times d$ is large, finding a good trade-off between exploration and local convergence is challenging. In addition here, the region with high criterion values is likely to be narrow, in particular when the number of observations n becomes large. Instead of performing the optimization of J_n^Γ directly on the \mathbb{X}^r space, we propose, similar to Sacks et al. (1989a), the following sequential heuristic:

- find the point x_{n+1} minimizing the criterion for $r = 1$;
- while $k < r$, consider the points x_{n+1}, \dots, x_{n+k} as fixed, find x_{n+k+1} such that the set of points $x_{n+1}, \dots, x_{n+k}, x_{n+k+1}$ minimizes the criterion for $r = k + 1$, and set $k \leftarrow k + 1$.

Hence, r consecutive optimizations in dimension d are performed instead of the more difficult $r \times d$ dimension optimization. Note that this approach makes it possible to use larger values of r .

The evolution of the algorithm is shown in Figure 3. The excursion set is accurately identified in a few (three) iterations of the parallel SUR strategy. After 18 evaluations (i.e. six initial evaluations plus three iterations, each providing a batch of $r = 4$ points), the excursion probability $p_n(x)$ does not depart much from the true function $\mathbb{1}_{x \in \Gamma^*}$.

A key question is to compare performances between the one-point and the multi-point criteria. If the total number of evaluations of f is identical, we generally expect the multi-point criterion to have a worse performance than the one-point criterion, in terms of reducing the uncertainty measure $H_n^{(\Gamma)}$. Indeed, for n a multiple of r , say, the n^{th} evaluation point is chosen based on $n - 1$ past evaluations in the one-point case, while it is only chosen based on $n - r$ evaluations in the multi-point case. In an ideal case, however, the uncertainty

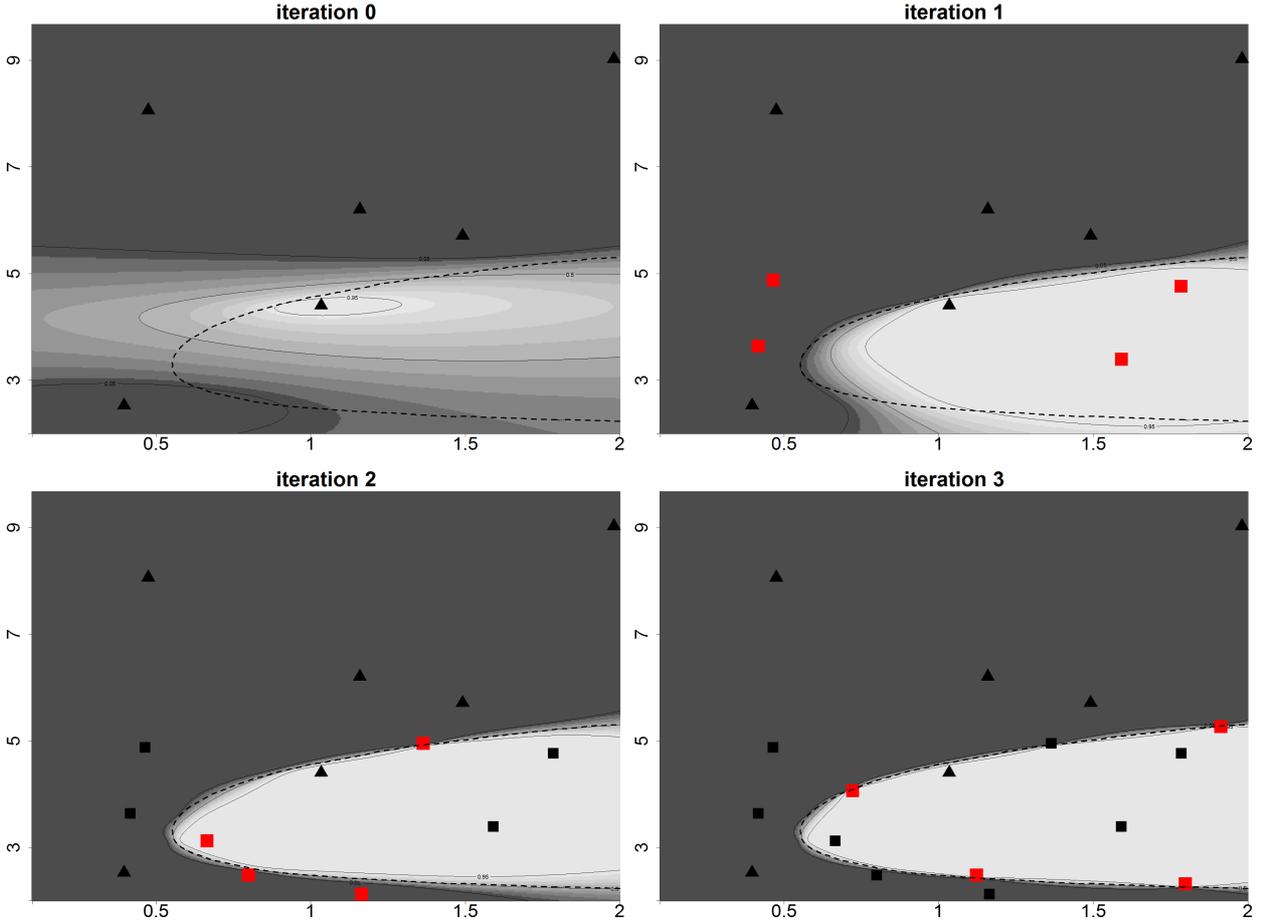


Figure 3: Plot of the function $p_n(x) = P_n(x \in \Gamma) = \Phi\left(\frac{m_n(x) - T}{s_n(x)}\right)$ after n evaluations of the simulator. The triangles are the six points of the initial design. The squares are the points sampled using the $J_n^{(\Gamma)}$ criterion. The points sampled at the considered iteration are highlighted in red. Areas in black correspond to $p_n(x) \approx 0$ and areas in white correspond to $p_n(x) \approx 1$. The dotted line indicates the true excursion set. The contour lines indicate the three level sets $p_n(x) = 0.05, 0.5$ and 0.95 .

would decrease at (almost) the same rate anyway, meaning that n/r iterations of the multi-point criterion would give the same remaining uncertainty as n iterations of the one-point one. Thus, if f is very expensive to evaluate, the time saving for the practitioner might be considerable.

Figure 4 shows the evolution of the uncertainty measure $H_n^{(\Gamma)}$ obtained during the uncertainty reduction with the one-point and the multi-point criteria. It also shows $J_n^{(\Gamma)}(\mathbf{x}_n^*)$, which is the values of the $J_n^{(\Gamma)}$ criterion (with $r = 4$) at its current minimizer \mathbf{x}_n^* . Note that,

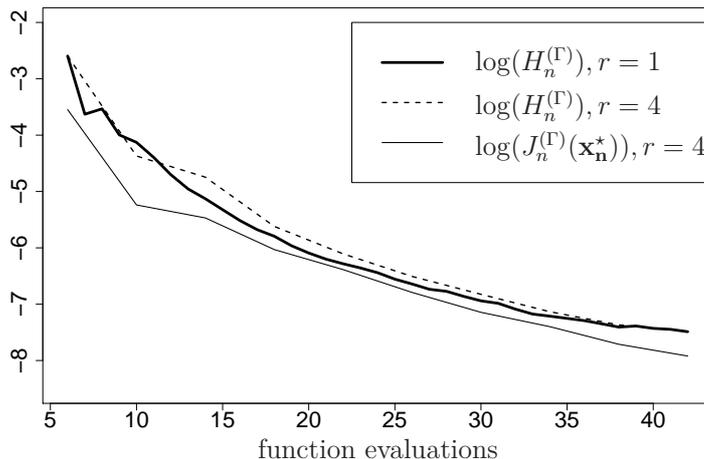


Figure 4: Evolution of $H_n^{(\Gamma)}$ during the sequential sampling strategy on the nuclear safety case study.

here, \mathbf{x}_n^* is a batch of r points. One can see on Figure 4 that at each iteration, $J_n^{(\Gamma)}(\mathbf{x}_n^*)$ is lower than $H_n^{(\Gamma)}$. This was to be expected, since $J_n^{(\Gamma)}(\mathbf{x}_n^*)$ is precisely the expectation of the future uncertainty measure $H_{n+r}^{(\Gamma)}$ if the r points \mathbf{x}_n^* are added to the design of experiments.

A striking conclusion of this section is that, here, the multi-point criterion almost has the *same performance* as the one-point one, in term of reducing the uncertainty measure $H_n^{(\Gamma)}$, which corresponds to the ideal case mentioned before. Figure 5 represents the points actually sampled with the one-point criterion. We see that, once the Kriging metamodel has correctly located the contour $\{x \in \mathbb{X} : f(x) = T\}$, both one-point and multi-point criteria tend to spread points in a neighbourhood of this contour. This intuitively explains the similar performances: since the chosen points are, typically, “far” from each other, the response at one point does not bring much information about the responses at the other points.

4.3 Experiments with the Hartman6 and Rastrigrin functions

The Hartman6 function is a well known 6-dimensional function used in box-constrained global optimisation (Torn and Zilinskas, 1989). We test our SUR strategies on this function for two main reasons. First, we want to illustrate that the sampling strategy works (i.e., is able to recover the true volume of excursion) on higher dimensional functions, and provides

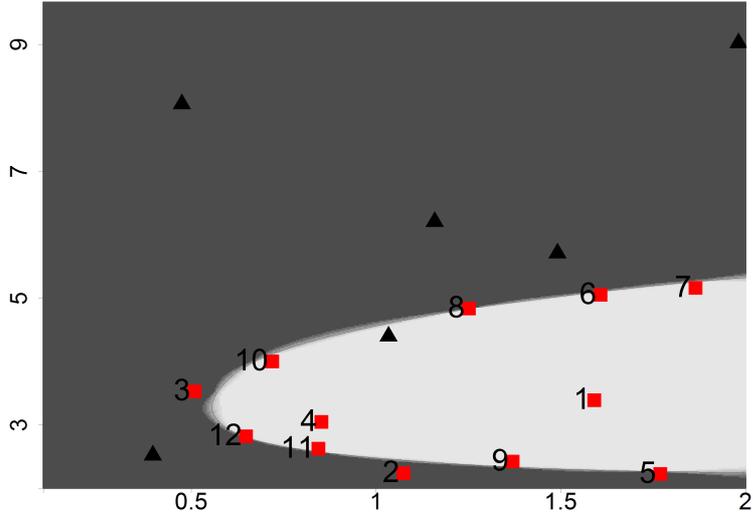


Figure 5: Plot of the $p_n(x)$ function after 12 iterations of the one-point $J_n^{(\Gamma)}$ criterion. The triangles are the points of the initial design. The squares are the points sampled at the different iterations.

better performances than a “basic” random sampling strategy. Second, we want again to compare the performances of the multi-point $J_n^{(\Gamma)}$ criterion with the one-point criterion. We follow Jones et al. (1998) and define the objective function y_H by

$$y_H(x) = -\log(-\text{Hartman6}(x)), \quad x \in \mathbb{R}^6. \quad (14)$$

A threshold of $T = 4$ is fixed for y_H . As a second test-case, we consider the Rastrigin function, restricted to the domain $[0, 2.5]^2$, with threshold $T = 438$. This function has several unconnected excursion sets (see Figure 6) and is one of the 24 noiseless test functions of the Black-Box Optimization Benchmark (BBOB) (Finck et al., 2010; Hansen et al., 2010). In the sequel, this function will be denoted by y_R .

The true excursion set of y_R is easy to visualize on a plot (see Figure 6). However, visualizing the excursion set of y_H is not trivial. Note that each function is considered as “difficult” here for different reasons. y_H is rather smooth and is expected to be correctly approximated using a kriging model with stationary covariance. However, the boundary of the excursion set is a manifold in dimension 5, which is hard to learn, as the excursion set does not have a simple shape. For y_R the difficulties are mainly linked to the fact that the function varies more in some regions (e.g. in the small excursion set on the bottom left). Thus the approximation of y_R using a model with stationary covariance is expected to be

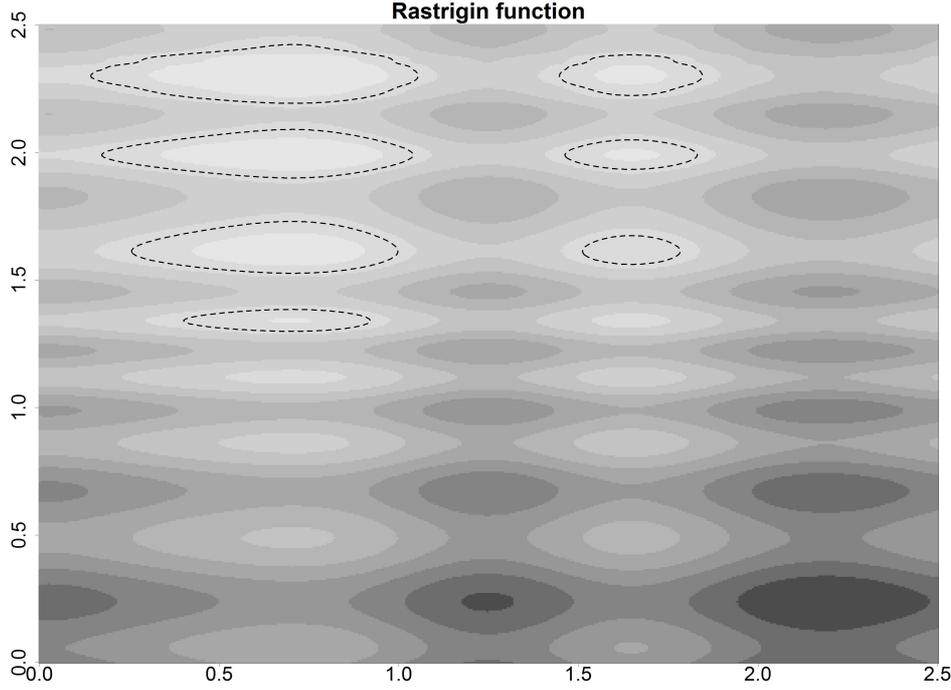


Figure 6: Plot of the Rastrigin function y_R , along with the contours (black dashed lines) of its excursion set above the threshold $T = 438$.

problematic.

In our tests, we use two measures of performance:

- the uncertainty measure $H_n^{(\Gamma)}$,
- the relative squared volume error SE, defined in Section 4.1.

We are interested in the empirical quantiles q_{50} and q_{90} of $H_n^{(\Gamma)}$ and SE obtained from 100 runs, each corresponding to a different initial design of experiments. Each initial design is again a maximin LHS. The quantile q_{50} helps to identify the performance of a “typical” run while q_{90} provides information on the worst runs and thus on the robustness of the method. The size of the initial design is 10 times the dimension of the input domain, i.e., 60 points for y_H and 20 points for y_R . For both functions, the performance calculations are based on 10,000 reference points. In this example, for y_H , $\alpha^* = 0.2127$ which means that exactly 2127 of the 10,000 reference points are in the excursion set. For y_R we have $\alpha^* = 0.0815$.

In Figure 7, the multi-point $J_n^{(\Gamma)}$ criterion, with $r = 4$, and the one-point criterion are compared in terms of $H_n^{(\Gamma)}$ and SE. As a benchmark, we also test a sampling strategy which

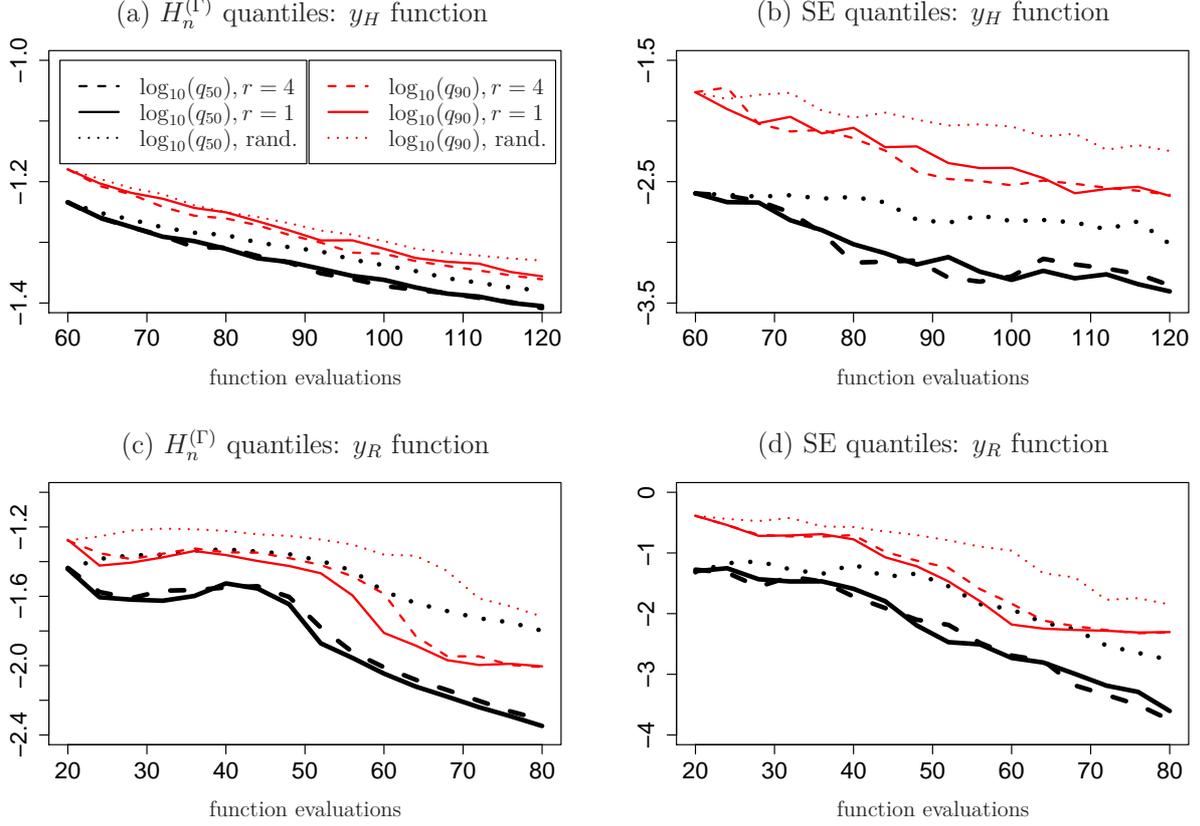


Figure 7: Evolution of the quantiles of the uncertainty measure $H_n^{(\Gamma)}$ and of the squared relative volume error SE during both the considered sequential and batch-sequential algorithms, on the y_H and y_R functions.

selects the evaluation points randomly, with a uniform distribution over \mathbb{X} . The one-point and multi-point $J_n^{(\Gamma)}$ criteria are calculated based on 250 integration points renewed at each iteration, using an instrumental distribution (see the Supplementary Material, Section B, for more detail). A total of 60 new points are evaluated for each instance of any of the three considered strategies. The heuristic optimization strategy presented in Section 4.2 is used for the multi-point criterion.

In Figures 7(b) and 7(d), the SUR strategies are sequentially reducing the relative squared volume error faster than the basic random sampling strategy. If we look at the relative error (the square-root of SE) we see that, on y_H , from an initial median relative error of approximately 5% we end with a median relative error of approximately 2% for the SUR strategies, and 3.1% for the random sampling. For y_R the median of the relative error

is initially 22% and is reduced to 1.6%, 1.3% and 4.1% when using the sequential SUR, batch-sequential and random sampling strategies respectively.

We also see that the 90% quantile of the relative error (i.e. the 10% worst runs) of both SUR strategies is higher (for both y_H and y_R) than the median relative error obtained with a random sampling. For y_R , if we look at the uncertainty measure $H_n^{(\Gamma)}$ instead of SE, the worst runs of the SUR strategy still outperform the random sampling (see Figure 7(c)). The discrepancy between these two indicators ($H_n^{(\Gamma)}$ and SE) can be explained by the difficulty to accurately estimate the covariance parameters from a small number of observations. This is particularly true for the function y_R where one of the seven excursion set connected components is sometimes “missed” by both the one-point and multi-point SUR criteria, due to non-stationary behaviour of the function which may lead to wrong uncertainty quantifications. A fully Bayesian approach on the covariance parameters may mitigate – but not solve – this problem and improve the performance of the worst runs. However, as mentioned before, the cost to compute (and minimize) the $J_n^{(\Gamma)}$ criterion in a fully Bayesian framework is expected to be substantially higher. We verified empirically that when more iterations were run, all connected components of the excursion sets were found. However, in general, we may recommend to users of these methods to “invest” a substantial part of the evaluation budget in an appropriate initial design in cases where the covariance parameters are difficult to learn.

Finally, we observed again that the multi-point criterion has a global performance which is usually close to the one-point criterion.

5 Conclusion and future work

In this paper, we presented algorithms for the computation of one-point and multi-point kriging-based infill sampling criteria. We showed how the proposed analytical formulas enable a practically sound implementation of the Stepwise Uncertainty Reduction (SUR) criteria proposed in Bect et al. (2012) and of batch-sequential versions of them. In particular, the complexity for computing a SUR criterion giving r points for simultaneous evaluation is “only” linear in r . Sampling criteria that were previously unaffordable in practical applications can now be used for sequential or batch-sequential inversion. In addition, we showed

that the proposed multi-point SUR criteria do perform extremely well in terms of estimating a probability of failure in competitive wall-clock time. For low values of r , computing one iteration of the multi-point criterion improves the accuracy of the estimation at almost the same pace than r sequential iterations of the one-point criterion. In applications on expensive-to-evaluate simulators, this allows considerable time savings for the practitioners. Finally, a new version of the R package KrigInv (Chevalier et al., 2013b) is now available online, and allows using the presented sequential and batch-sequential strategies.

Further enhancements of the presented methods remain possible, and a selection of them were mentioned in this paper. Sequential Monte Carlo methods might be an interesting alternative to compute a set of integration points that “evolves” from one iteration to another (see, e.g, Li et al. (2012)). The necessary optimization of these criteria could also be improved with gradient calculations, possibly in the whole space of dimension rd . Finally, from a more theoretical perspective, approaches directly based on random set notions (considering a “variance” of the excursion set itself, rather than the variance of the excursion volume) may provide elegant alternative sampling criteria for inversion and related problems.

Supplementary Materials

This paper has supplementary materials, available online. It consists of a zip file which contains a help file (readme.txt) and the following files:

fastSUR SupplementaryMaterial: This is a pdf file providing basis notions on kriging, details on the choice of the integration points and detailed complexity calculation of the algorithms presented in this paper.

R code: The code used to provide the results of Sections 4.1 and 4.3 is available in five R files: two files for each section and an additional file, `utils.R`, containing some useful functions.

R data: The results we obtained with this code are stored in two RData files: `Technometrics_section41.RData` and `Technometrics_section43.RData`.

Acknowledgments: Clément Chevalier gratefully acknowledges support from the French Nuclear Safety Institute (IRSN) and from the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (Ecole des Mines de Saint-Etienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments.

References

- R. J. Barnes and A. Watson. Efficient updating of kriging estimates and variances. *Mathematical geology*, 24(1):129–133, 1992.
- M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu. A framework for validation of computer models. *Technometrics*, 49(2):138–154, 2007.
- J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, 2012.
- R. Benassi, J. Bect, and E. Vazquez. Robust gaussian process-based global optimization using a fully bayesian expected improvement criterion. In Springer, editor, *Learning and Intelligent Optimization, Lecture Notes in Computer Science*, volume 6683, pages 176–190, 2011.
- B. J. Bichon, M. S. Eldred, L. P. Swiler, S. Mahadevan, and J. M. McFarland. Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA Journal*, 46(10):2459–2468, 2008.
- C. Chevalier and D. Ginsbourger. Fast computation of the multi-points expected improvement with applications in batch selection. *Proceedings of the LION7 conference, Lecture Notes in Computer Science*, 2013.
- C. Chevalier, D. Ginsbourger, and X. Emery. Corrected kriging update formulae for batch-sequential data assimilation. *Proceedings of the IAMG2013 conference, Madrid*, 2013a. URL <http://arxiv.org/abs/1203.6452>.

- C. Chevalier, V. Picheny, and D. Ginsbourger. Kriging: An efficient and user-friendly implementation of batch-sequential inversion strategies based on kriging. *Computational Statistics & Data Analysis*, 2013b. doi: <http://dx.doi.org/10.1016/j.csda.2013.03.008>.
- B. Echard, N. Gayton, and M. Lemaire. Kriging-based Monte Carlo simulation to compute the probability of failure efficiently: AK-MCS method. In *6èmes Journées Nationales de Fiabilité, 24–26 mars, Toulouse, France, 2010*.
- X. Emery. The kriging update equations and their application to the selection of neighboring data. *Computational Geosciences*, 13(1):211–219, 2009.
- K.-T. Fang, R. Li, and A. Sudjianto. *Design and modeling for computer experiments*. Chapman & Hall / CRC Press, 2006.
- S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical report, Research Center PPE, 2009, 2010.
- A. I. J. Forrester, A. Sóbester, and A. J. Keane. *Engineering design via surrogate modelling: a practical guide*. Wiley, 2008.
- H. Gao, J. Wang, and P. Zhao. The updated kriging variance and optimal sample design. *Mathematical Geology*, 28:295–313, 1996.
- A. Genz. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1:141–149, 1992.
- D. Ginsbourger, R. Le Riche, and L. Carraro. Kriging is well-suited to parallelize optimization. In L. M. Hiot, Y. S. Ong, Y. Tenne, and C.-K. Goh, editors, *Computational Intelligence in Expensive Optimization Problems*, volume 2 of *Adaptation Learning and Optimization*, pages 131–162. Springer, 2010.
- R. Gramacy and H. Lee. Adaptive design and analysis of supercomputer experiments. *Technometrics*, 51(2):130–145, 2009.

- R. B. Gramacy and N. G. Polson. Particle learning of gaussian process models for sequential design and optimization. *Journal of Computational and Graphical Statistics*, 20(1):pp. 102–118, 2011.
- M. Hamada, H. Martz, C. Reese, and A. Wilson. Finding near-optimal bayesian experimental designs via genetic algorithms. *The American Statistician*, 55(3):175–181, 2001.
- N. Hansen, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical report, INRIA 2009, 2010.
- D. R. Jones, M. Schonlau, and J. William. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.
- L. Li, J. Bect, and E. Vazquez. Bayesian Subset Simulation : a kriging-based subset simulation algorithm for the estimation of small probabilities of failure. Technical report, 2012. URL <http://publicationslist.org/jbect>. Submitted to PSAM 11 and ESREL 2012.
- M. Morris and T. Mitchell. Exploratory designs for computational experiments. *Journal of statistical planning and inference*, 43(3):381–402, 1995.
- V. Picheny, D. Ginsbourger, O. Roustant, R. T. Haftka, and N.-H. Kim. Adaptive designs of experiments for accurate approximation of target regions. *Journal of Mechanical Design*, 132(7), 2010.
- P. Ranjan, D. Bingham, and G. Michailidis. Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541, 2008.
- J. Sacks and S. Schiller. Spatial designs. *Statistical decision theory and related topics IV*, 2 (S S):385–399, 1988.
- J. Sacks, S. Schiller, and W. Welch. Designs for computer experiments. *Technometrics*, 31 (1):41–47, 1989a.
- J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435, 1989b.

T. J. Santner, B. J. Williams, and W. Notz. *The Design and Analysis of Computer Experiments*. Springer Verlag, 2003.

A. Torn and A. Zilinskas. *Global Optimization*. Springer Verlag, New York, 1989.

E. Vazquez and J. Bect. A sequential Bayesian algorithm to estimate a probability of failure. In *Proceedings of the 15th IFAC Symposium on System Identification, SYSID 15th IFAC Symposium on System Identification*, Saint-Malo France, 2009.