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

Clément Chevalier, Julien Bect, David Ginsbourger, Emmanuel Vazquez, Victor Picheny, Yann Richet

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, CERFACS
Yann Richet, IRSN

April 2, 2012

Abstract

Stepwise Uncertainty Reduction (SUR) strategies aim at constructing a sequence of sampling points for a function $f : \mathbb{R}^d \to \mathbb{R}$, in such a way that the residual uncertainty about a quantity of interest becomes small. In the context of Gaussian Process-based approximation of computer experiments, these strategies have been shown to be particularly efficient for the problem of estimating the volume of excursion of a function f above a threshold. However, these strategies remain difficult to use in practice because of their high computational complexity, and they only deliver at each iteration a single point to evaluate. In this paper we introduce parallel sampling criteria, which allow selecting several sampling points simultaneously. Such criteria are of particular interest when the function f is expensive to evaluate and many CPUs are available. We also manage to drastically reduce the computational cost of these strategies using closed form expressions. We illustrate their performances in various numerical experiments, including a nuclear safety test case.

Keywords: Computer experiments, Gaussian processes, Sequential design, 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 $f : \mathbb{X} \to \mathbb{R}$ 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 $\alpha^\star$ of the excursion set $\Gamma^\star = \{x \in \mathbb{X} : f(x) \geq T\}$, with T a given threshold, under a measure $\mathsf{P}_\mathbb{X}$ on $\mathbb{X}$. In safety analysis, $\mathsf{P}_\mathbb{X}$ typically models the uncertainty on input parameters. If f is expensive to evaluate, the estimation of $\alpha^\star$ must be performed with a limited number of evaluations of f, which naturally excludes brute-force approaches like Monte Carlo sampling.

A popular approach consists in constructing 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, 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. [1989]) 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., 2011, 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., 2011] rely on global measures of uncertainty about the excursion set $\Gamma^\star$ and take into account the important fact that sampling at a point x also brings useful information on the neighbourhood of x. Numerical experiments [reported by Bect et al., 2011] showed that SUR criteria widely outperform pointwise criteria in terms of quickly estimating the true volume of excursion $\alpha^\star$.

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., 2011], 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 in 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, on a real test case. 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. A sequential sampling of this 2-dimensional "expensive" function, using a SUR strategy, manages to identify with very few evaluations the set of "dangerous" configurations.

Despite their very good performances 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 very computer demanding since it relies on numerical integration. Besides, these strategies where designed to sample one point at a time while practionners often the have the capacity to run $r > 1$ simulations in parallel. This very high numerical complexity to simply compute the value of a sampling criterion at one point mainly explains 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 new *parallel* SUR sampling criteria and provide methods and algorithms allowing to run

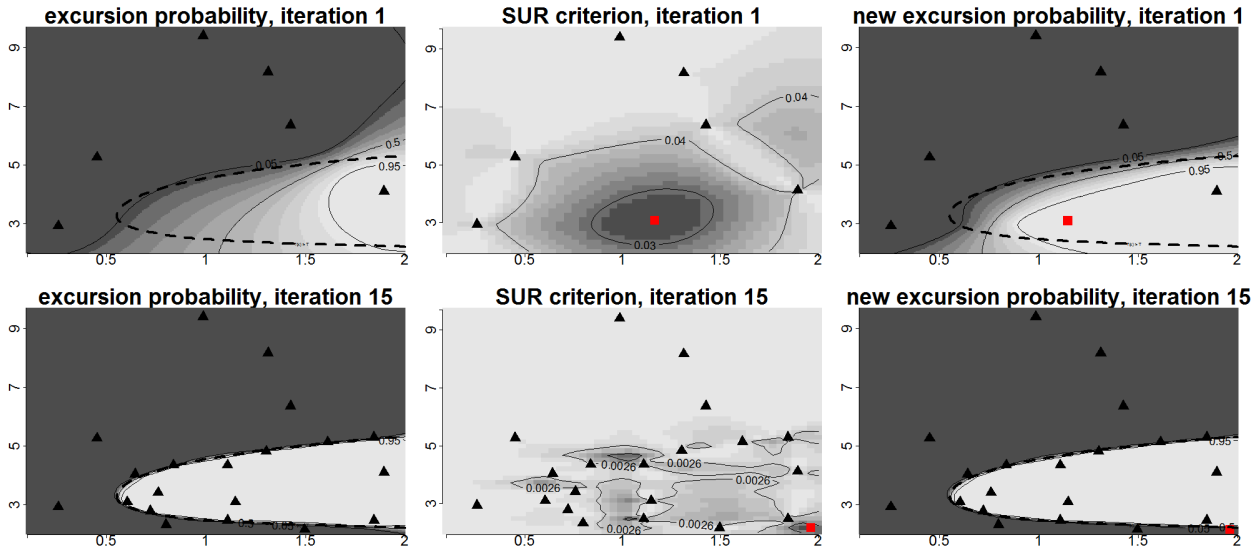

Figure 1: SUR strategy (first and last iteration) applied to a nuclear criticality safety simulator. The black triangles stand for the design points at the current iteration. The red square is the point sampled using the SUR criterion. Areas in black (resp. white) correspond to excursion probabilities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

them in a very reasonable time. In particular, we show that the unaffordable (one step look-ahead) optimal criterion presented in Bect et al. [2011] can be computed quickly, without simulating any Gaussian Process realization. Furthermore, we illustrate the use of parallel criteria in real-life applications, and investigate their performances on several test cases.

The paper is organised as follows: Section 2 introduces notations and gives two examples of SUR criteria (including the optimal onse-step-lookahead criterion) with their new parallel 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 and details about the choice of the integrations points are presented in appendix. In addition, detailed computational complexity calculations are provided as Supplementary Material.

2 Kriging-based Stepwise Uncertainty Reduction

A Stepwise Uncertainty Reduction (SUR) strategy aims at constructing a sequence $X_1, X_2, \ldots$ of evaluation points of f in such a way that the residual uncertainty about a quantity of interest given the information provided by the evaluation results becomes small. 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, which is assumed 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 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:

$$X_{n+1} \ = \ \underset{x_{n+1} \in \mathbb{X}}{\operatorname{argmin}} \, J_n(x_{n+1}) \tag{1}$$

where

$$J_n(x_{n+1}) := \mathsf{E}_n \left(H_{n+1} \,\middle|\, X_{n+1} = x_{n+1} \right) , \tag{2}$$

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

Depending of the definition given to the measure of uncertainty, many sequential 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, $\alpha = \mathsf{P}_{\mathbb{X}}(\Gamma)$. The conditional variance $\mathrm{Var}_n\left(\alpha\right)$ of α is a natural choice for H_n to quantify the (residual) uncertainty about $\alpha^\star$ given $\mathcal{A}_n$. In the rest of the paper, we denote this uncertainty by $H_n^{(\alpha)}$. A possible SUR strategy to estimate $\alpha^\star$ would consist, at step n, in choosing as next evaluation point an optimizer of the criterion:

$$J_n^{(\alpha)}(x_{n+1}) := \mathsf{E}_n \left(Var_{n+1}(\alpha) \,\middle|\, X_{n+1} = x_{n+1} \right) \tag{3}$$

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

$$J_n^{(\alpha)}(x_{n+1}, \ldots, x_{n+r}) := \mathsf{E}_n \left(Var_{n+r}(\alpha) \,\middle|\, X_{n+1} = x_{n+1}, \ldots, X_{n+r} = x_{n+r} \right) \tag{4}$$

Note that the latter criterion, is considered intractable in Bect et al. [2011] for $r = 1$ because its computation has a very high numerical complexity (it requires the simulation of a large number of Gaussian Process realizations). We will see in the next sections that both parallel and non parallel versions of this criterion can be computed quickly and used in applications.

Example 2: criterion $J_n^{(\Gamma)}$. The excursion volume can be characterized by the random variable $_{\{\xi(x)>T\}}$. This random variable has conditional expectation:

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

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

$$J_n^{(\Gamma)}(x_{n+1}) := \mathsf{E}_n\left(\int_{\mathbb{X}} p_{n+1}(1 - p_{n+1})\mathrm{d}\mathsf{P}_{\mathbb{X}} \;\middle|\; X_{n+1} = x_{n+1}\right). \tag{5}$$

This criterion was first introduced by Bect et al. [2011]. Again, a natural extension is the following new parallel criterion:

$$J_n^{(\Gamma)}(x_{n+1}, \ldots, x_{n+r}) = \mathsf{E}_n\left(\int_{\mathbb{X}} p_{n+r}(1 - p_{n+r})\mathrm{d}\mathsf{P}_{\mathbb{X}} \;\middle|\; X_{n+1} = x_{n+1}, \ldots, X_{n+r} = x_{n+r}\right). \tag{6}$$

In Bect et al. [2011], the numerical computation of $J_n^{(\Gamma)}$ in (6) is considered only for $r = 1$ and is based on quadrature formulas written as

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

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 $H_{n+1}^{(\Gamma)}$. The $x^{(m)}$'s are i.i.d. according to $\mathsf{P}_{\mathbb{X}}$; $(y_{n+1}^{(1)}, \ldots, y_{n+1}^{(Q)})$ and $(w^{(1)}, \ldots, w^{(Q)})$ 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 $X_1, \ldots, X_n$ and x_{n+1}. It follows (See supplementary materialfor more detail about algorithmic complexities) that the computation of $J_n^{(\Gamma)}$ at one point has a $O(n^3 + Mn^2 + MQ)$ complexity. 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 implemented using (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 $(\xi(x_{n+1}), \ldots, \xi(x_{n+r}))^\top$, which requires a discretization of an integral over $\mathbb{R}^r$. As a consequence, the complexity to compute the parallel SUR criterion presented above is expected to rise quickly with r, which makes it impractical even for small r.

The next section brings useful properties allowing to circumvent these issues. In particular, new analytical formulas allow us to get rid of the cumbersome integral over $\mathbb{R}^r$ and make it possible to compute efficiently both parallel and non-parallel criteria.

3 Efficient calculation of parallel SUR criteria

In this section, we provide new expressions allowing to efficiently compute the two parallel SUR strategies introduced in the previous section.

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

As explained in the previous sections, the proposed parallel criterion $J_n^{(\Gamma)}$ is the conditional expectation given $\mathcal{A}_n$ of the future uncertainty $H_{n+r}^{(\Gamma)}$, assuming that r new points will be evaluated. 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 to discretize 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.

$$J_n^{(\Gamma)}(x_{n+1}, \ldots, x_{n+r}) = \int_{\mathbb{X}} \Phi_2\left(\begin{pmatrix} a(x) \\ -a(x) \end{pmatrix}, \begin{pmatrix} c(x) & 1-c(x) \\ 1-c(x) & c(x) \end{pmatrix}\right) \mathsf{P}_{\mathbb{X}}(\mathrm{d}x), \qquad (8)$$

where:

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

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

$$J_n^{(\Gamma)}(x_{n+1}, \ldots, x_{n+r}) = \int_{\mathbb{X}} \mathsf{E}_n\left(p_{n+r}(x)(1 - p_{n+r}(x))\right) \mathsf{P}_{\mathbb{X}}(\mathrm{d}x), \tag{9}$$

where the conditioning on $X_{n+i} = x_{n+i}$'s is not explicitely reproduced, to alleviate notations. Now, using the kriging update formula (see, e.g., Barnes and Watson [1992], Gao et al. [1996], Emery [2009], as well as Chevalier and Ginsbourger [2012]), we obtain:

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

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

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

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

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

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

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

where $(N_1, N_2)^T \sim \mathcal{N}(\mathbf{0}, I_2)$ independently of $\mathbf{y}_{\text{centered}}$. 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 the anounced result directly follows by integration over $\mathbb{X}$. □

Remark 1. *For the latter Proposition 1, we managed to get rid of 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 such integral over* $\mathbb{R}^r$*. Besides, the computation of* $J_n^{(\Gamma)}$ *is now available for* $r > 1$ *at a cost that is not quickly increasing 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* $r << n << M$ *(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 values for* r *can be used in applications.*

Remark 2. *When the integral over* $\mathbb{X}$ *is discretized based on* M *integration points, the computation of the* $J_n^{(\Gamma)}$ *criterion requires to calculate the updated kriging variance* $s^2_{n+r}(x)$ *for each of the* M *points. The updated kriging variance can be efficiently calculated using a kriging variance update formula given and proven in Chevalier and Ginsbourger [2012].*

Remark 3. *By reducing equation* (6) *to equation* (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 there exist very efficient numerical procedures to compute it. For instance, Genz [1992] wrote routines in Fortran77 which have been wrapped in many R Packages (e.g., mnormt, pbivnorm, mvtnorm, available on CRAN).*

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

In the kriging framework and conditionally on $\mathcal{A}_n$, the conditional expectation of the volume of excursion α is given by $\widehat{\alpha} := \int_{\mathbb{X}} p_n d\mathsf{P}_{\mathbb{X}}$. As explained before, the conditional variance $\mathrm{Var}_n\left(\alpha\right)$ of α given $\mathcal{A}_n$ is a very natural choice to quantify the uncertainty about α but, even for $r = 1$, it was considered intractable so far. In fact, with the help of the kriging update formulas (See Eq. 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 parallel and non-parallel versions.

Proposition 2.

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

where

- $\Phi_2, a, \mathbf{b}$, *and* $\boldsymbol{\Sigma}$ *are defined as in Proposition 1,*
- $d(z_1,z_2) := \mathbf{b}(z_1)^\top \boldsymbol{\Sigma} \mathbf{b}(z_2)$
- γ_n *is a constant, in the sense that it does not depend on* $(x_{n+1},\ldots,x_{n+r})$.

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

$$\begin{aligned}
J_n^{(\alpha)}(x_{n+1},\ldots,x_{n+r}) &:= \mathsf{E}_n\left(Var_{n+r}(\alpha) \;\middle|\; X_{n+1}=x_{n+1},\ldots,X_{n+r}=x_{n+r}\right) \\
&= \mathsf{E}_n\left(\mathsf{E}_{n+r}\left(\int_{\mathbb{X}} (\;_{\{\xi(x)>T\}} - p_{n+r}(z))\mathsf{P}_{\mathbb{X}}(\mathrm{d}z)\right)^2\right) \\
&= \mathsf{E}_n\left(\mathsf{E}_{n+r}\iint_{\mathbb{X}\times\mathbb{X}} (\;_{\{\xi(z_1)>T\}} - p_{n+r}(z_1))(\;_{\{\xi(z_2)>T\}} - p_{n+r}(z_2))\mathsf{P}_{\mathbb{X}}(\mathrm{d}z_1)\mathsf{P}_{\mathbb{X}}(\mathrm{d}z_2)\right) \\
&= \mathsf{E}_n\left(\iint_{\mathbb{X}\times\mathbb{X}} \left(\mathsf{E}_{n+r}(\;_{\{\xi(z_1)>T\}}\;_{\{\xi(z_2)>T\}}) - p_{n+r}(z_1)p_{n+r}(z_2)\right)\mathsf{P}_{\mathbb{X}}(\mathrm{d}z_1)\mathsf{P}_{\mathbb{X}}(\mathrm{d}z_2)\right)
\end{aligned}$$

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

$$\mathsf{E}_n(\mathsf{E}_{n+r}(\;_{\{\xi(z_1)>T\}}\;_{\{\xi(z_2)>T\}})) = \mathsf{E}_n(\;_{\{\xi(z_1)>T\}}\;_{\{\xi(z_2)>T\}}) = P(\xi(z_1)>T,\xi(z_2)>T|\mathcal{A}_n)$$

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

$$J_n^{(\alpha)}(x_{n+1},\ldots,x_{n+r}) = \gamma_n - \iint_{\mathbb{X}\times\mathbb{X}} \mathsf{E}_n(p_{n+r}(z_1)p_{n+r}(z_2))\mathsf{P}_{\mathbb{X}}(\mathrm{d}z_1)\mathsf{P}_{\mathbb{X}}(\mathrm{d}z_2).$$

The end result is obtained using similar calculations as in the proof of Property 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 integrations 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.*

4 Applications

In this section, we illustrate our sequential sampling strategies on several test cases. The examples include simulated realizations of two-dimensional Gaussian Processes, a two-dimensional nuclear safety case study and a six-dimensional test function.

4.1 Benchmark on simulated Gaussian Process realizations

The first objective of this section is to compare the *non parallel* versions of the $J_n^{(\Gamma)}$ and $J_n^{(\alpha)}$ criteria. The test functions are 200 independent realizations of a two-dimensional Gaussian Process (GP) indexed by $[0,1]^2$. The covariance parameters for the kriging models are fixed equal to the actual ones of the GP. Besides comparing the two criteria, we want to estimate the effect of numerical integration erros on the global performance of the SUR strategies. The criterion $J_n^{(\alpha)}$ requires to compute an integral over $\mathbb{X} \times \mathbb{X}$, so is it 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 GP realization, we fix the threshold T in order to have a constant volume of excursion $\alpha^\star = 0.2$. The volumes are calculated using 1 000 reference points, so for each Gaussian Process realization, exactly 200 points are in the excursion set. 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 the $J_n^{(\Gamma)}$ criterion or the $J_n^{(\alpha)}$ criterion. For all realizations, the performance of both criteria are measured in term of the relative squared volume error $\mathrm{SE} := (\widehat{\alpha} - \alpha^\star)^2/\alpha^{\star 2}$, where $\widehat{\alpha}$ 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; *note that the integration points are not bound to lie among the* 1 000 *reference points.* 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 Appendix, section B.

Figure 2 draws the evolution of $\overline{SE}$, the average of the SE values over the 200 realizations, as a function of the number of observations. First, one can see that the number of integration points has a direct impact on the performance of both SUR strategy, since the experiments corresponding to $M = 50$ with quasi-Monte Carlo sampling –based here on a Sobol sequence– provide the worst results (bold curves with the MC legend on Figure 2).

Besides, 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.

A third conclusion is that the use of importance sampling (with a well chosen instrumental distribution) has a significant impact on 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 a good accuracy as explained in Appendix B. In that case, a clever choice of the integration points has a crucial impact on the global performance of the strategy.

From this application we can conclude that the criterion $J_n^{(\Gamma)}$ roughly achieves the same performances as $J_n^{(\alpha)}$ at a lower computational cost. This is why, in the next applications, we will mostly focus our attention on the $J_n^{(\Gamma)}$ criterion and its parallel extension.

4.2 Nuclear safety test case

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

A system involving fissile materials may produce a chain reaction based on neutrons,

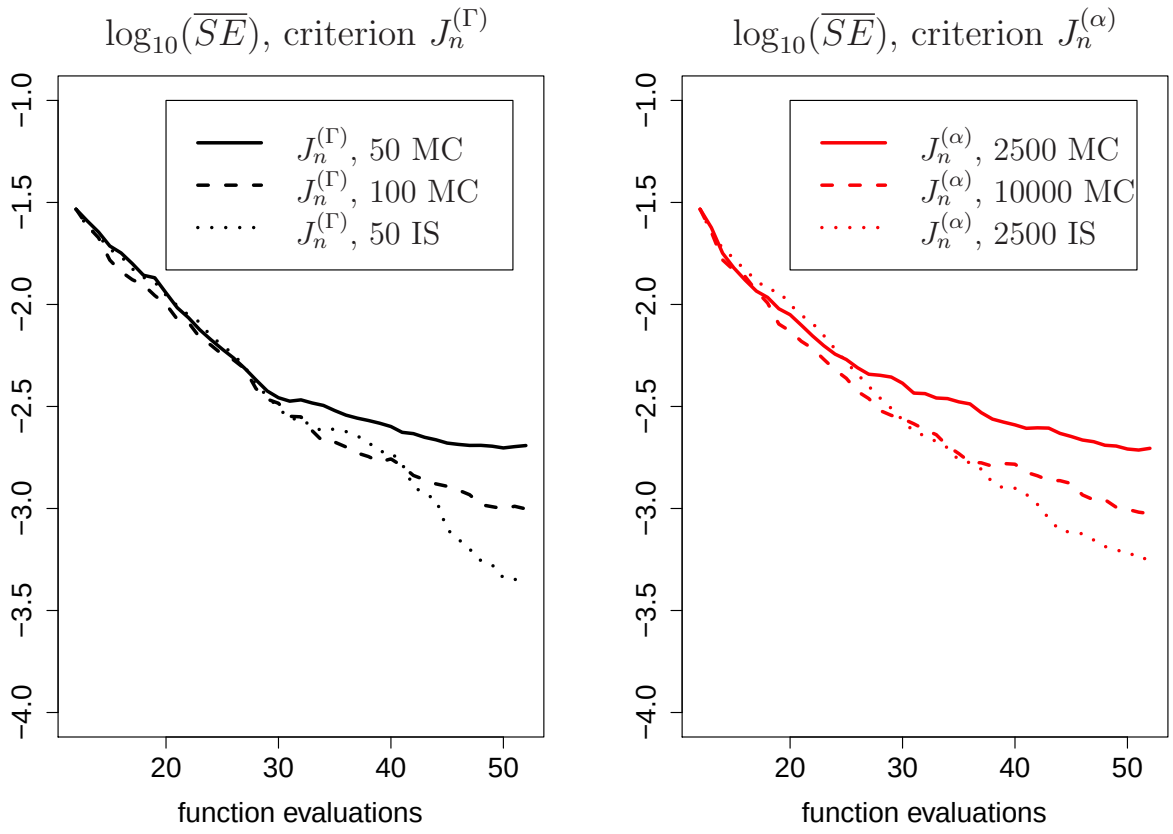

Figure 2: Performance (measured in term of mean squared relative error) of two SUR strategies based on the $J_n^{(\Gamma)}$ and $J_n^{(\alpha)}$ criteria, in function of the number of integration points and the method for choosing them (importance sampling or quasi Monte Carlo).

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 one single configuration. Many CPU are available to evaluate points in parallel, which means that our sampling strategy has to provide us, at each iteration, a number of points $r > 1$ at which to evaluate the simulator simultaneously.

In this section, we run our stepwise algorithms on this two-dimensional problem. The $J_n^{(\Gamma)}$ sampling criterion is used with an initial design of experiment of $n = 6$ points. The criterion is computed using $M = 600$ integration points renewed at each iteration, sampled from a specific instrumental distribution (See Appendix B for more detail). At each iteration, batches of $r = 4$ points are evaluated in parallel. Instead of performing the optimization of $J_n^{(\Gamma)}$ directly on the $\mathbb{X}^r$ space, we propose the following heuristic:

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

This heuristic is of course sub-optimal but it allows us to replace the difficult optimization in $r \times d$ dimensions into r consecutive optimizations in dimension d. Note that this option allows using high values of r.

The evolution of the algorithm is shown on Figure 3. One can see that the excursion set is accurately identified in 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 $_{x\in\Gamma^\star}$.

A key question here is to compare performances between the parallel criterion and the non-parallel one ($r = 1$). If the total number of evaluation of f is strictly identical, we generally expect the parallel criterion to have a worse performance than the non parallel one, in term of reducing the uncertainty $H_n^{(\Gamma)}$, because in the non parallel case the n^{th} evaluation point is chosen based on $n-1$ past evaluations, while in the parallel case it is chosen based on $n-r$ evaluations. In an ideal case, the uncertainty would decrease at the same rate, meaning that n/r iterations of the parallel criterion gives the same remaining uncertainty as n iterations of the non-parallel one (for n a multiple of r, say). Thus, if f is very expensive to evaluate, the time saving for the practitioner might be considerable.

Figure 4 gives the evolution of the uncertainty $H_n^{(\Gamma)}$ obtained during the uncertainty reduction with the parallel and the non-parallel criteria. It also shows $J_n^{(\Gamma)}(\mathbf{x}_\mathbf{n}^\star)$, which is the values of the $J_n^{(\Gamma)}$ criterion (with $r = 4$) at its current minimizer $\mathbf{x}_\mathbf{n}^\star$. Note that, here, $\mathbf{x}_\mathbf{n}^\star$ is a batch of r points. One can see on Figure 4 that at each iteration, $J_n^{(\Gamma)}(\mathbf{x}_\mathbf{n}^\star)$ is lower

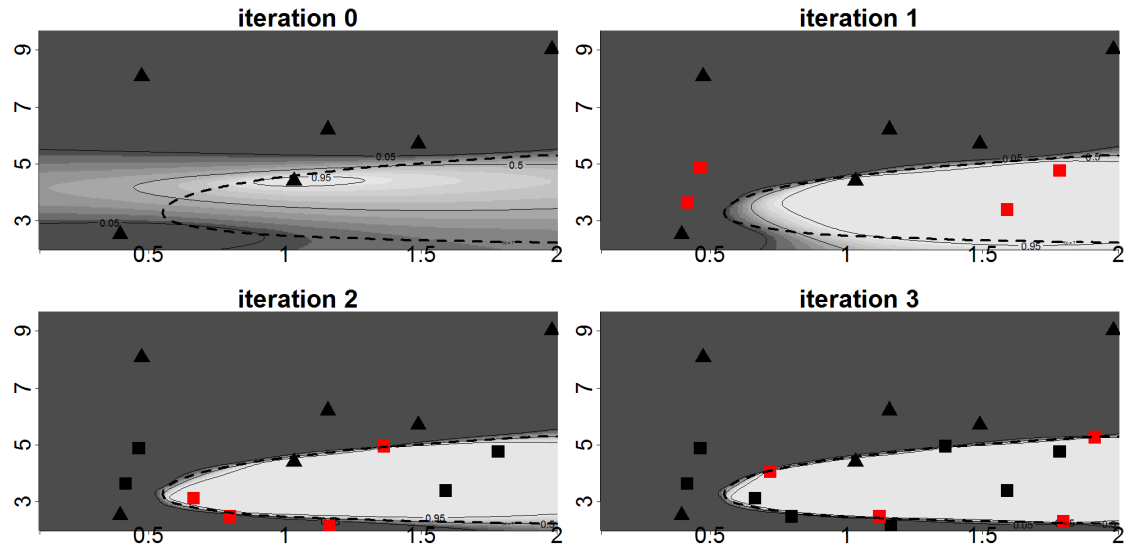

Figure 3: Plot of the function $p_n(x) = P_n(x \in \Gamma) = \Phi\left(\frac{m_n(x)-T}{s_n(x)}\right)$ after n evaluations of the simulator. The triangles are the six points of the initial DOE. The squares are the points sampled using the $J_n^{(\Gamma)}$ criterion. Areas in black correspond to $p_n(x) \approx 0$ 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.

than $H_n^{(\Gamma)}$. This was to be expected, since $J_n^{(\Gamma)}(\mathbf{x_n^\star})$ is precisely the expectation of the future uncertainty $H_{n+r}^{(\Gamma)}$ if the r points $\mathbf{x_n^\star}$ are added to the design of experiments.

A striking conclusion to this section is that, here, the parallel criterion has the *same performance* as the non-parallel one, in term of reducing the uncertainty $H_n^{(\Gamma)}$, which corresponds to the ideal case mentioned before.

4.3 Six dimensional example

The Hartman6 function is a well known 6-dimensional function used in unconstrained global optimisation (Torn and Zilinskas [1989]). We test our SUR strategies on this function for two reasons. First we want to prove that the sampling strategy works (i.e., is able to recover the true volume of excursion) on higher dimensional functions, and provides better performances than a "basic" random sampling strategy. Second, we want to confirm the compared performances of the parallel $J_n^{(\Gamma)}$ criterion with the non parallel one observed

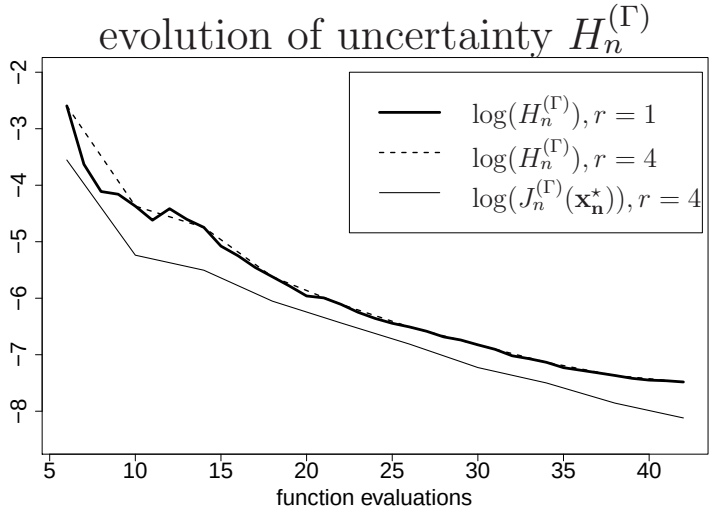

Figure 4: Evolution of $H_n^{(\Gamma)}$ during the sequential sampling strategy on the nuclear safety case study. The optimum $J_n^{(\Gamma)}(\mathbf{x_n^\star})$ of the $J_n^{(\Gamma)}$ criterion is usually lower than $H_n^{(\Gamma)}$ and corresponds to the expectation of $H_n^{(\Gamma)}$ at the next iteration.

earlier. We follow Jones et al. [1998] and perform the following change of variables:

$$y_H : \mathbf{x} \in \mathbb{R}^6 \mapsto -\log(-\text{Hartman6}(\mathbf{x})) \ .$$

We work with a threshold $T = 4$ and use two measures of performance:

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

All the performance calculations are done using 10 000 reference points (with a sobol sequence). In this example, $\alpha^\star = 0.2127$ which means that exactly 2127 of the 10 000 reference points are in the excursion set.

The results are averaged over 100 random initial design of experiments of 36 points (all of them being maximin Latin Hypercube Designs, generated with the R package *lhs* Carnell [2009]). The average uncertainty and the squared error are denoted by $\overline{H_n^{(\Gamma)}}$ and $\overline{SE}$ respectively.

In Figure 5, the parallel $J_n^{(\Gamma)}$ criterion, with $r = 4$, and the non parallel one are tested and compared. The criteria are calculated based on 250 integration points renewed at each iteration, using an instrumental distribution. A total of 80 new points are evaluated for

each instance of any of the two considered strategies. Two main conclusions can be obtained from Figure 5. First, as anticipated, the SUR strategies are sequentially reducing the relative real volume error faster than the basic random sampling strategy. From a relative error, in absolute value, of approximately 15% (with the initial design), we end up with a relative error (after having added 80 new observations) of approximately 3.3% on average. Second, the parallel strategy has again almost the same performance as the non parallel one. This means that we are, again, very close to the "ideal case" mentioned in the previous application.

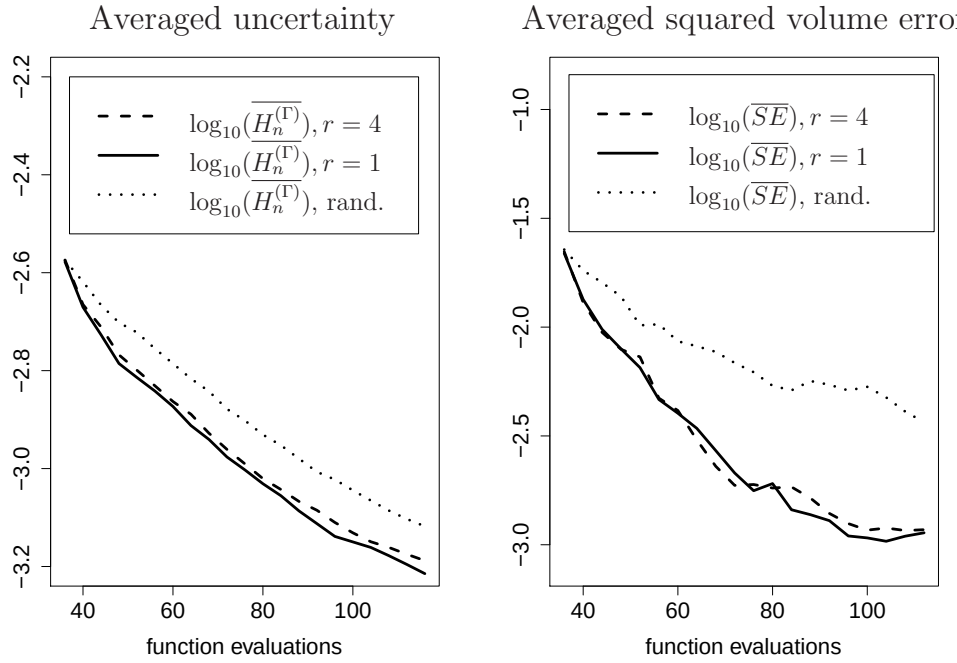

Figure 5: Evolution of $\overline{H_n^{(\Gamma)}}$ and of the averaged squared relative volume error during both the considered sequential and batch sequential algorithms, on the y_H function.

5 Conclusion and future work

In this paper, we presented algorithms for the computation of parallel and non-parallel Kriging-based infill sampling criteria. We showed that the use of the formulas introduced in this paper enables a practically sound implementation of the Stepwise Uncertainty Reduction (SUR) criteria proposed in Bect et al. [2011] and of batch-sequential versions of them. In particular, the complexity for computing a SUR criterion giving r points to evaluate si-

multaneously is "only" linear in r. Sampling criteria that were previously unaffordable in practical applications can now be used for parallel or non-parallel inversion. In addition, we showed that the proposed parallel SUR criteria do perform extremely well, in terms of quickly estimating a probability of failure. For low values of r, computing one iteration of the parallel criterion improves the accuracy of the estimation at almost the same pace than r sequential iterations of the non-parallel criterion. In applications on expensive-to-evaluate simulators, this allows a considerable time saving for the practitioners. Finally, a new version of the R package KrigInv (Chevalier et al. [2012]) is now available online, and allows using the presented sequential and batch-sequential strategies.

Further improvement are possible in this work and 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.). 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.

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*, pages 1–21, 2011. DOI: 10.1007/s11222-011-9241-4.

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.

R. Carnell. *lhs: Latin Hypercube Samples*, 2009. R package version 0.5.

C. Chevalier and D. Ginsbourger. Corrected kriging update formulae for batch-sequential data assimilation. March 2012. URL `http://arxiv.org/abs/1203.6452`.

C. Chevalier, V. Picheny, and D. Ginsbourger. *KrigInv: Kriging-based Inversion for Deterministic and Noisy Computer Experiments*, 2012. URL `http://CRAN.R-project.org/package=KrigInv`. R package version 1.2.

V. Dubourg. *Métamodèles adaptatifs pour l'analyse de fiabilité et l'optimisation sous contrainte fiabiliste*. PhD thesis, IFMA, 2011.

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.

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. 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. Submitted to PSAM 11 and ESREL 2012. URL `http://publicationslist.org/jbect`.

P. Ranjan, D. Bingham, and G. Michailidis. Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541, 2008.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

R. Rubinstein and D. Kroese. *Simulation and the Monte Carlo method*. Wiley, 2008.

J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435, 1989.

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 2009 15th IFAC Symposium on System Identification, SYSID 2009*, Saint-Malo France, 2009.

APPENDIX

A Kriging mean and variance

In this section, we shall recall how to obtain the kriging mean and variance. Let $\xi \sim \mathrm{GP}(m, k)$ be a Gaussian random process with mean function $m(\cdot) = \mathsf{E}(\xi(\cdot))$ and covariance function $k(\cdot,\cdot) = \mathrm{cov}(\xi(\cdot), \xi(\cdot))$. We assume that the mean can be written as a linear combination

$$m(\cdot) \;=\; \sum_{i=1}^{l} \beta_i \, p_i(\cdot) \tag{14}$$

of basis functions $p_1, \ldots, p_l$ (very often, these are monomials), where $\beta_1, \ldots, \beta_l$ are *unknown* parameters. The covariance k is assumed to be a given symmetric *strictly* positive function.

The kriging predictor of ξ at a point $x \in \mathbb{X}$ from n observations $\xi(x_1), \ldots, \xi(x_n)$ is the best linear unbiased predictor (BLUP) of $\xi(x)$ from the observations, that we shall denote by

$$m_n(x) \;=\; \lambda(x; \underline{x}_n)^{\mathsf{T}} \begin{pmatrix} \xi(x_1) \\ \vdots \\ \xi(x_n) \end{pmatrix}. \tag{15}$$

The vector of *kriging weights* $\lambda(x; \underline{x}_n) \in \mathbb{R}^n$ can be obtained by solving the linear system

$$\underbrace{\begin{pmatrix} K(\underline{x}_n) & p(\underline{x}_n)^{\mathsf{T}} \\ p(\underline{x}_n) & \mathbf{0} \end{pmatrix}}_{:=\widetilde{K}(\underline{x}_n)} \cdot \underbrace{\begin{pmatrix} \lambda(x; \underline{x}_n) \\ \mu(x; \underline{x}_n) \end{pmatrix}}_{:=\widetilde{\lambda}(x; \underline{x}_n)} \;=\; \underbrace{\begin{pmatrix} k(x, \underline{x}_n) \\ p(x) \end{pmatrix}}_{:=\widetilde{k}(x, \underline{x}_n)} \tag{16}$$

where $K(\underline{x}_n)$ is the $n \times n$ covariance matrix of the random vector $(\xi(x_1), \ldots, \xi(x_n))^{\mathsf{T}}$, $p(\underline{x}_n)$ is the $l \times n$ matrix with general term $p_i(x_j)$, $k(x, \underline{x}_n)$ is the column vector with general term $k(x, x_i)$ and $\mu(x; \underline{x}_n)$ is a vector of l Lagrange multipliers associated to the unbiasedness constraint.

The covariance function of the prediction error

$$k_n(x, y) \;:=\; \mathsf{E}_n\left(\left(\xi(x) - m_n(x)\right) \left(\xi(y) - m_n(y)\right) \right), \tag{17}$$

also called *kriging covariance*, can be written using the notations of equation (16) as

$$k_n(x, y) \;=\; k(x, y) - \widetilde{k}(x, \underline{x}_n)^{\mathsf{T}} \, \widetilde{\lambda}(y; \underline{x}_n). \tag{18}$$

The conditional variance of the prediction error at a point $x \in \mathbb{X}$, also called *kriging variance*, will be denoted by $s_n^2(x) := k_n(x, x)$.

Remark 5. *To ensure that the conditional process $f|\mathcal{A}_n$ is still Gaussian when the mean function is of the form (14), with an unknown vector of parameters β, it is necessary to adopt a Bayesian approach and to use an (improper) uniform distribution over $\mathbb{R}^l$ as a prior distribution for β (see Bect et al. [2011], Section 2.3, Proposition 2, and the references therein for more detail).*

B Modelling choices

B.1 Choice of the integration points

The criteria studied in this paper involve the numerical computation of integrals over the domains $\mathbb{X}$ or $\mathbb{X} \times \mathbb{X}$. We showed in section 4 that the number of integration points has an important impact on the performance of the corresponding strategies. In this section we deal with the question of how these integration points are chosen.

The $J_n^{(\Gamma)}$ and $J_n^{(\alpha)}$ criteria may be written under the following general form:

$$J_n(x_{n+1}, \ldots, x_{n+r}) = \text{Const} \pm \int_D \mathsf{E}_n \left(v_{n+r}(\mathbf{u}) \;\middle|\; X_{n+1} = x_{n+1}, \ldots, X_{n+r} = x_{n+r} \right) Q(d\mathbf{u}) \quad (19)$$

More specifically, for the $J_n^{(\Gamma)}$ criterion, we have Const $= 0$, $D := \mathbb{X}$, $v_{n+r}(x) = p_{n+r}(x)(1 - p_{n+r}(x))$ and $Q := \mathsf{P}_{\mathbb{X}}$. For the $J_n^{(\alpha)}$ criterion, Const $= \gamma_n$, $D := \mathbb{X} \times \mathbb{X}$, $v_{n+r}(z_1, z_2) = p_{n+r}(z_1)p_{n+r}(z_2)$ and $Q := \mathsf{P}_{\mathbb{X}} \otimes \mathsf{P}_{\mathbb{X}}$. Note that in the remainder of this section, we omit the conditioning on $(X_{n+1} = x_{n+1}, \ldots, X_{n+r} = x_{n+r})$ in order to simplify the notations.

For both criteria, a straightforward option to for calculating the integral over D would be to use Monte Carlo sampling with distribution Q. However, importance sampling techniques (see, e.g., Rubinstein and Kroese [2008], Robert and Casella [2004]) are a good choice for reducing the variance of the Monte Carlo error in case a suitable instrumental density is chosen. For the integral in equation 19, a natural choice for such an instrumental density is:

$$h(\mathbf{u}) \propto v_n(\mathbf{u})Q(\mathbf{u}) \; . \quad (20)$$

Indeed, if a point $\mathbf{u}$ has, in expectation, a high "future uncertainty" $v_{n+r}(\mathbf{u})$ it generally means that the current uncertainty at point $\mathbf{u}$, $v_n(\mathbf{u})$ is already high. For the $J_n^{(\Gamma)}$ criterion, we thus propose the following instrumental density:

$$h(\mathbf{u}) \propto p_n(\mathbf{u})(1 - p_n(\mathbf{u}))\mathsf{P}_{\mathbb{X}}(\mathbf{u}) \tag{21}$$

Similarly, for the $J_n^{(\alpha)}$ criterion, the proposed instrumental density is:

$$h(z_1, z_2) \propto p_n(z_1)p_n(z_2)\mathsf{P}_{\mathbb{X}} \otimes \mathsf{P}_{\mathbb{X}}(z_1, z_2) \tag{22}$$

Remark 6. *Using importance sampling techniques for the problem of estimating a probability of failure with a kriging metamodel has already been proposed and applied in Dubourg [2011].*

B.2 Sampling from our instrumental distribution

Sampling from the densities h defined above in Equations 21 and 22 is a difficult task. Figure 6 shows the value of the density $p_n(x)(1-p_n(x))\mathsf{P}_{\mathbb{X}}(x)$, which is (up to a multiplicative factor) the instrumental density proposed to compute the criterion $J_n^{(\Gamma)}$. When $n = 20$ evaluations of the simulator are available, one may remark (right graph) that the support of the instrumental density becomes very narrow. This issue complicates the use of standard MCMC algorithms to obtain a sample distributed according to the instrumental density. Sequential Monte Carlo methods may provide a nice solution to this issue but they have not been investigated and implemented yet in the KrigInv package (Chevalier et al. [2012]). Instead, we decided to use a simpler approximation. The idea consists in replacing the integral in 19 by the following estimator:

$$\int_D \mathsf{E}_n\left(v_{n+r}(\mathbf{u})\right) Q(d\mathbf{u}) \approx \frac{1}{N}\sum_{j=1}^{N} \mathsf{E}_n(v_{n+r}(\mathbf{u}_j)) \ , \tag{23}$$

where N is a large number and $\mathbf{u}_1, \ldots, \mathbf{u}_N$ is a i.i.d sample of N points with distribution Q. Rather than aiming at computing the integral of origin, we try to approximate this finite sum using a discrete instrumental density proportional to:

$$\sum_{j=1}^{N} v_n(\mathbf{u}_j)\delta_{\mathbf{u}_j}$$

Such new discrete instrumental density is easy to sample from.

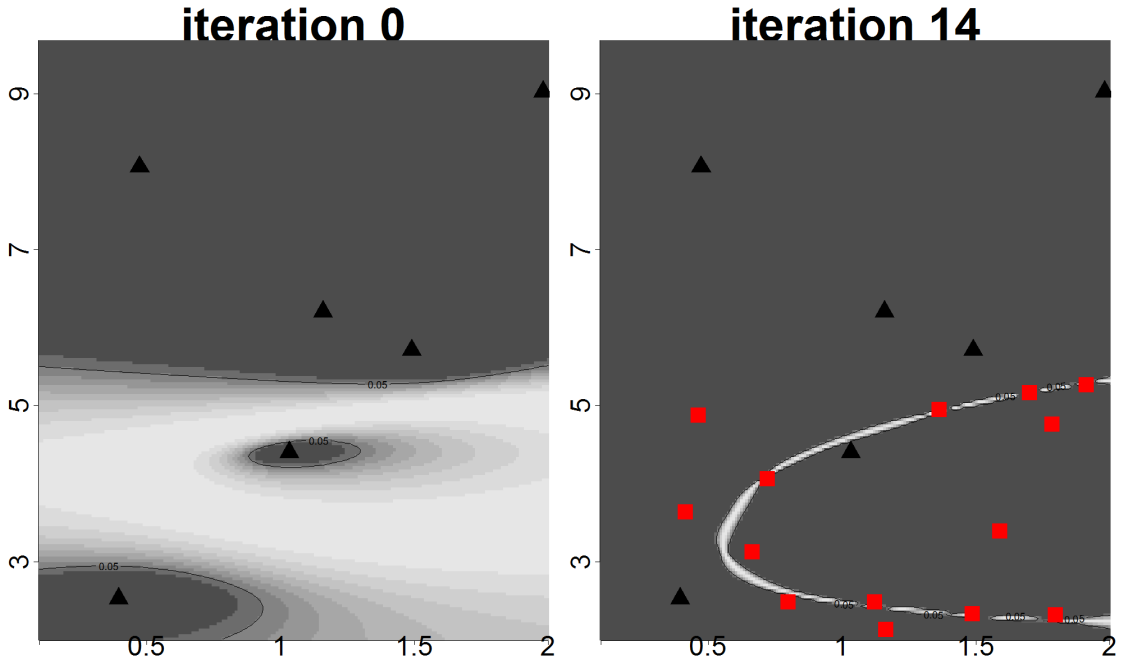

Figure 6: Plot of the function $p_n(x)(1-p_n(x))$ after n evaluations of the MORET code (left). The triangles are the six points of the initial DOE. Right: the squares are the points sampled using the $J_n^{(\Gamma)}$ criterion. Areas in black correspond to low uncertainty zones.

Of course, this method has important limitations. In particular the new "objective" quantity $\frac{1}{N}\sum_{j=1}^{N} \mathsf{E}_n(v_{n+r}(\mathbf{u}_j))$ can be completely different from $\int_D \mathsf{E}_n\left(v_{n+r}(\mathbf{u})\right) Q(d\mathbf{u})$ if all the N points from the initial large sample have an uncertainty v_n close to zero, or if N is not large enough. In essence, both N and the number of draws should tend to infinity in order for the estimator to converge to the true value of the integral. However, even if adapted MCMC approaches are likely to perform better in future implementations, this simple and easy option proposed here already provided a significantly improved calculation of the proposed SUR criteria compared to a standard quasi-Monte Carlo approach, as presented in section 4.