Statistica Sinica **28** (2018), 853-877
doi:https://doi.org/10.5705/ss.202016.0160

# SEQUENTIAL DESIGN OF EXPERIMENTS FOR ESTIMATING QUANTILES OF BLACK-BOX FUNCTIONS

T. Labopin-Richard and V. Picheny

*Université Paul Sabatier and INRA*

*Abstract:* Estimating quantiles of black-box deterministic functions with random inputs is a challenging task when the number of function evaluations is severely restricted, which is typical for computer experiments. This article proposes two new sequential Bayesian methods for quantile estimation based on the Gaussian process metamodel. Both rely on the Stepwise Uncertainty Reduction paradigm, hence aim at providing a sequence of function evaluations that reduces an uncertainty measure associated with the quantile estimator. The proposed strategies are tested on several numerical examples, showing that accurate estimators can be obtained using only a small number of function evaluations.

*Key words and phrases:* Gaussian processes, risk assessment, stepwise uncertainty reduction.

## 1. Introduction

In the last decades, the question of designing experiments for the efficient exploration and analysis of numerical black-box models has received wide interest, and metamodel-based strategies have been shown to offer efficient alternatives in many contexts, such as optimization or uncertainty quantification. We consider here the question of estimating quantiles of the output of a black-box model. More precisely, let $g : \mathbb{X} \subset \mathbb{R}^d \to \mathbb{R}$ denote the output of interest of the model, the inputs of which can vary within $\mathbb{X}$. We assume here that the multivariate input $X$ is modelled as a random vector; then, our objective is to estimate a quantile of $g(X)$:

$$q^{\alpha}(g(X)) = q^{\alpha}(Y) = F_Y^{-1}(\alpha), \tag{1.1}$$

for a fixed level $\alpha \in (0,1)$, where $F_U^{-1} := \inf\{\mathbf{x} : F_U(\mathbf{x}) \geq u\}$ denotes the generalized inverse of the cumulative distribution function of a random variable $U$. We consider here only random vectors $X$ and functions $g$ regular enough to have $F_Y\left(F_Y^{-1}(\alpha)\right) = \alpha$ (that is, $F_Y$ is continuous). Since the level $\alpha$ is fixed, we omit the index in the sequel.

A natural idea to estimate a quantile consists in using its empirical estimator: having at hand a sample $(X_i)_{i=1,\dots,n}$ of the input law $X$, we run it through the computer model to obtain a sample $(Y_i)_{i=1,\dots,n}$ of the output $Y$. Then, denoting $Y_{(k)}$ the $k$-th order statistic of the previous sample, the estimator

$$q_n := Y_{(\lfloor n\alpha \rfloor + 1)} \tag{1.2}$$

is consistent and asymptotically Gaussian under weak assumptions on the model (see David and Nagaraja (2003) for more details). However, for computationally expensive models, the sample size is drastically limited, which makes the estimator (1.2) impractical. In that case, one may replace the sample $(X_i)_i$ by a sequence of well-chosen points that provide a useful information for the quantile estimation. This is the basis of the large field of importance sampling, for which many solutions have been proposed, using either parametric (see e.g. Egloff et al. (2010); Cornuet et al. (2012)) or non-parametric approaches (see e.g. Zhang (1996); Morio (2012)).

When the available data are scarce, an interesting alternative is to rely on metamodels (a.k.a. surrogate models). The observation set is used to build a fast-to-evaluate approximation of $g$, and use this approximation (metamodel) to estimate $q_n$. Such an approach is often combined with importance sampling strategies; see e.g. Bucher and Bourgund (1990); Bourinet, Deheeger and Lemaire (2011); Cannamela, Garnier and Iooss (2008); Morio (2012) for works based on support vector machines, neural networks, linear regression, or kriging.

In this article, we focus on the Gaussian process (GP) metamodel, which has the advantage of being particularly well-suited for sequential sampling, adding observations one at a time, using the metamodel to guide the process. Many algorithms, following Močkus (1975); Jones, Schonlau and Welch (1998), have been proposed for optimization, or for the estimation of a probability of exceedance (see for instance Bect et al. (2012) for a review), which is the dual problem of quantile estimation.

GP-based algorithms specifically dedicated to quantile estimation are scarcer in the literature. Oakley (2004) proposed a two-step strategy: first, generate an initial set of observations to obtain a first estimator of the quantile, then increase the set of observations by a second set likely to improve the estimator. Jala et al. (2016) proposed two sequential methods (called GPQE and GPQE+), based on the GP-UCB optimization algorithm of de Freitas, Zoghi and Smola (2012), that is, making use of the confidence bounds provided by the Gaussian Process model.

In this paper we propose two new algorithms based on *Stepwise Uncertainty*

*Reduction* (SUR), a framework that has been successfully applied to closely related problems such as optimization (Picheny (2013)), or the estimation of a probability of exceedance (Bect et al. (2012); Chevalier et al. (2014)). A first SUR strategy has been proposed for the quantile case in Arnaud et al. (2010) and Jala et al. (2012) that relies on expensive simulation procedures. However, finding a statistically sound algorithm with a reasonable cost of computation, in particular when the problem dimension increases, is still an open problem.

The rest of the paper is organized as follow. In Section 2, we introduce the basics of GP modelling, our quantile estimator and the SUR framework. Section 3 describes our two algorithms. Some numerical simulations to test the two methods are presented in Section 4, followed by concluding comments in Section 5. Most of the proofs are deferred to the Appendix.

## 2. Gaussian Process Modelling and Sequential Experiments

### 2.1. Model definition

We consider here the classical GP framework in computer experiments (Sacks et al. (1989); Rasmussen and Williams (2006)): we suppose that $g$ is the realization of a GP denoted by $G(\cdot)$ with known mean $\mu$ and covariance function $c$.

Given an observed sample contained in the event $\mathcal{A}_n = \{(\mathbf{x}_1, g_1), (\mathbf{x}_2, g_2), \dots (\mathbf{x}_n, g_n)\}$ with all $\mathbf{x}_i \in \mathbb{X}$ and $g_i = g(x_i)$, the distribution of $G|\mathcal{A}_n$ is entirely known:

$$\mathcal{L}\left(G|\mathcal{A}_n\right) = GP(m_n(\cdot), k_n(\cdot,\cdot))\,,$$

where $\mathcal{L}$ refers to the law and with, $\forall \mathbf{x} \in \mathbb{X}$,

$$m_n(\mathbf{x}) = \mathbb{E}(G(\mathbf{x})|\mathcal{A}_n) = c_n(\mathbf{x})^T C_n^{-1}\mathbf{g_n}, \tag{2.1}$$

$$k_n(\mathbf{x},\mathbf{x}') = \mathrm{Cov}\left(G(\mathbf{x}), G(\mathbf{x}')|\mathcal{A}_n\right) = c(\mathbf{x},\mathbf{x}') - c_n(\mathbf{x})^T C_n^{-1} c_n(\mathbf{x}'), \tag{2.2}$$

where $c_n(\mathbf{x}) = [c(\mathbf{x}_1,\mathbf{x}),\dots,c(\mathbf{x}_n,\mathbf{x})]^T$, $C_n = [c(\mathbf{x}_i,\mathbf{x}_j)]_{1\le i,j\le n}$ and $\mathbf{g_n} = [g_1,\dots, g_n]$. In the sequel, we also write $s_n^2(\mathbf{x}) = k_n(\mathbf{x},\mathbf{x})$.

We use here the standard Kriging framework (Stein (2012)), where the covariance function depends on unknown parameters that are inferred from $\mathcal{A}_n$, using maximum likelihood estimates for instance. Usually, the estimates are used as face value, but updated when new observations are added to the model.

### 2.2 Quantile estimation

Since each call to the code $g$ is expensive, the sequence of inputs to evaluate,

$\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, must be chosen carefully to make our estimator as accurate as possible. The general scheme based on GP modelling proceeds as follows:

- For an initial budget $N_0$, build an initialisation sample $(\mathbf{x}_0^i, g(\mathbf{x}_0^i))_{i=1\ldots N_0}$, typically using a space-filling strategy, and compute the estimator of the quantile $q_{N_0}$.
- At each step $n+1 > N_0$ and until the budget $N$ of evaluations is reached: knowing the current set of observations $\mathcal{A}_n$ and estimator $q_n$, choose the next point to evaluate $\mathbf{x}_{n+1}$, based on a so-called *infill criterion*. Evaluate $g(\mathbf{x}_{n+1})$ and update the observations $\mathcal{A}_{n+1}$ and the estimator $q_{n+1}$.
- $q_N$ is the estimator of the quantile to return.

**Quantile estimator.** Considering that, conditionally on $\mathcal{A}_n$, the best approximation of $G(\mathbf{x})$ is $m_n(\mathbf{x})$, an intuitive estimator of the quantile (as chosen in Oakley (2004) for instance) is simply the quantile of the GP mean:

$$q_n := q_X(m_n(X)) = q_X\left(\mathbb{E}_G\left[G(X)|\mathcal{A}_n\right]\right), \tag{2.3}$$

where $q_X$ is the quantile with regard to the measure on $X$ and $\mathbb{E}_G$ the expectation with regard to $G$. In the following, the subscripts are dropped when there is no ambiguity.

Another natural idea is to consider the estimator that minimizes the mean square error $\mathbb{E}\left((q-q_n)^2\right)$ among all $\mathcal{A}_n$-measurable estimators:

$$q_n = \mathbb{E}_G\left(q_X(G(X))|\mathcal{A}_n\right). \tag{2.4}$$

This estimator is used for instance in Jala et al. (2016). Despite its theoretical qualities, it cannot be expressed in a computationally tractable form. Hence, in the sequel, we focus on the estimator (2.3).

**Sequential Sampling and Stepwise Uncertainty Reduction.** Consider methods based on the sequential maximization of an infill criterion of the form

$$\mathbf{x}_{n+1}^* = \underset{\mathbf{x}_{n+1}\in\mathbb{X}}{\operatorname{argmax}}\, J_n(\mathbf{x}_{n+1}), \tag{2.5}$$

where $J_n$ is a function that depends on $\mathcal{A}_n$ (through the GP conditional distribution) and on $\mathbf{x}_{n+1}$, a candidate observation location.

Intuitively, an efficient strategy explores $\mathbb{X}$ enough to obtain a GP model reasonably accurate everywhere, and exploits previous results to identify the area with response values close to the quantile and sample more densely there.

To this end, the concept of Stepwise Uncertainty Reduction (SUR) has been proposed originally in Geman and Jedynak (1996) as a trade-off between exploitation and exploration, and has been successfully adapted to optimization (Villemonteix, Vazquez and Walter (2009); Picheny (2013)) or probability of failure estimation frameworks (Bect et al. (2012); Chevalier et al. (2014)). The general principle of SUR strategies is to define an uncertainty measure related to the objective pursued, and add sequentially the observation that reduces the most uncertainty. The main difficulty of such an approach is to evaluate the potential impact of a candidate point $\mathbf{x}_{n+1}$ without having access to $g(\mathbf{x}_{n+1}) = g_{n+1}$, that would require running the computer code.

In the quantile estimation context, Jala et al. (2012) and Arnaud et al. (2010) proposed to choose the next point to evaluate as the minimizer of the conditional variance of the quantile estimator (2.4). This strategy showed promising results, as it substantially outperformed more classical strategies, and, with a small number of input variables, managed to identify the quantile area (that is, where $g$ is close to its quantile) and choose the majority of the points in it. However, computing their criterion is very costly, as it requires drawing many GP realizations, which hinders its use in practice for dimensions larger than two.

## 3. Two Sequential Strategies for Quantile Estimation

In this section, we propose two new infill criteria dedicated to quantile estimation. Both are based on a closed-form expression of the updated value of the quantile estimator when an observation is added to $\mathcal{A}_n$. This update formula is first given in Section 3.1, then the two criteria are derived in Sections 3.2 and 3.3.

### 3.1. Update formula for the quantile estimator

We focus on the estimator (2.3), which is, at step $n$, the quantile of the random vector $m_n(X)$. Since no closed-form expression is available, we approach it by using the empirical quantile. Let $\mathbf{X}_{\mathrm{MC}} = (\mathbf{x}^1_{\mathrm{MC}}, \ldots, \mathbf{x}^l_{\mathrm{MC}})$ be an independent sample of size $l$, distributed as $X$. We compute $m_n(\mathbf{X}_{\mathrm{MC}})$ and order this vector by denoting $m_n(\mathbf{X}_{\mathrm{MC}})_{(i)}$ the $i$-th coordinate. Then we choose

$$q_n = m_n(\mathbf{X}_{\mathrm{MC}})_{(\lfloor l\alpha \rfloor + 1)}. \tag{3.1}$$

**Remark** 1. Since the observation points $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ do not follow the distribution of $X$, they cannot be used to estimate the quantile. Hence, a different set $(\mathbf{X}_{\mathrm{MC}})$ must be used.

Consider that a new observation $g_{n+1} = g(\mathbf{x}_{n+1})$ is added to $\mathcal{A}_n$. The key to building a SUR strategy is to measure the impact of this observation on our estimator. To do so, we introduce the notion of *quantile point*, denoted by $\mathbf{x}_n^q$, as the point of $\mathbf{X}_{\mathrm{MC}}$ such that

$$q_n = m_n(\mathbf{x}_n^q).$$

This formulation allows us to provide a closed-form expression of the value of the estimator $q_{n+1} = m_{n+1}(\mathbf{x}_{n+1}^q)$ as a function of the past observations $\mathcal{A}_n$, the past quantile estimator $q_n$, a candidate point $\mathbf{x}_{n+1}$ and its corresponding (deterministic) evaluation $g_{n+1}$.

A classical property of the GP model, linking its means at steps $n$ and $n+1$, is

$$m_{n+1}(\mathbf{x}) = m_n(\mathbf{x}) + \frac{k_n(\mathbf{x}_{n+1}, \mathbf{x})}{s_n^2(\mathbf{x}_{n+1})} \left(g_{n+1} - m_n(\mathbf{x}_{n+1})\right), \tag{3.2}$$

where $(\mathbf{x}_{n+1}, g_{n+1})$ is a new observational event. For detailed calculations, discussion and complexity analysis, see Chevalier, Ginsbourger and Emery (2014). This allows us to compute $m_{n+1}$ very efficiently, as it does not require inverting the updated covariance matrix $C_{n+1}$, and shows explicitely its linear dependency with respect to $g_{n+1}$. Computing $k_n(\mathbf{x}_{n+1}, \mathbf{x})$ and $s_n^2(\mathbf{x}_{n+1})$ has a $O(n^2)$ complexity (see (2.2)), provided that the inverse of $C_n$ has been computed beforehand.

By (3.2), we have

$$m_{n+1}(\mathbf{X}_{\mathrm{MC}}) = m_n(\mathbf{X}_{\mathrm{MC}}) + \frac{k_n(\mathbf{X}_{\mathrm{MC}}, \mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1})^2} \left(g_{n+1} - m_n(\mathbf{x}_{n+1})\right). \tag{3.3}$$

We see directly that once $\mathbf{x}_{n+1}$ is fixed, the vector $m_{n+1}(\mathbf{X}_{\mathrm{MC}})$ is determined by the value of $g_{n+1}$. Our objective is to derive, for all $g_{n+1} \in \mathbb{R}$, which point of $\mathbf{X}_{\mathrm{MC}}$ is the quantile point, the point satisfying

$$m_{n+1}(\mathbf{X}_{\mathrm{MC}})_{\lfloor l\alpha \rfloor + 1} = m_{n+1}\left(\mathbf{x}_{n+1}^q\right). \tag{3.4}$$

Write $\mathbf{b} = m_n(\mathbf{X}_{\mathrm{MC}})$ and $\mathbf{a} = k_n(\mathbf{X}_{\mathrm{MC}}, \mathbf{x}_{n+1})$, vectors of $\mathbb{R}^l$, and $z = g_{n+1} - m_n(\mathbf{x}_{n+1})/s_n^2(\mathbf{x}_{n+1})$, so that the updated mean is simply displayed as a linear function of $z$, $\mathbf{b} + \mathbf{a}z$. Our problem can then be interpreted graphically: each coordinate of $m_{n+1}(\mathbf{X}_{\mathrm{MC}})$ is represented by a straight line

$$b_i + a_i z,\ \ i \in \{1, \dots, l\}, \tag{3.5}$$

and the task of finding $\mathbf{x}_{n+1}^q$ for any value of $g_{n+1}$ amounts to finding the $\lfloor l\alpha \rfloor + 1$ lowest line for any value of $z$. A similar graphical interpretation can be found in Scott, Frazier and Powell (2011) in an optimization context.

We have that the lines' order changes only when two lines intersect each other. There are $(l(l-1))/2$ intersection points, with values given by $(b_r - b_s)/(a_s - a_r)$ $(1 \le s, r \le l,\ s \ne r)$.

Denote by $I_1, \ldots, I_L$, in increasing order, the intersection points at which the index of the $\lfloor l\alpha \rfloor + 1$ lowest line changes ($L \le l(l-1)/2$). We set $I_0 = -\infty$ and $I_{L+1} = +\infty$, and introduce $(B_i)_{0 \le i \le L}$, the sequence of intervals between intersection points:

$$B_i = [I_i, I_{i+1}] \text{ for } i \in [0, L] \tag{3.6}$$

For any $z \in B_i$, the order of $(b_i + a_i z)$ is fixed.

Denoting by $j_i$ the index of the $\lfloor l\alpha \rfloor + 1$ lowest line, we have

$$\mathbf{x}_{n+1}^q = \mathbf{x}_{\mathrm{MC}}^{j_i}, \quad z \in B_i, \tag{3.7}$$

the quantile point when $z \in B_i$, which we henceforth write $\mathbf{x}_{n+1}^q(B_i)$.

**Proposition 1.** *In the previous notation, at step $n$ for the candidate point $\mathbf{x}_{n+1}$, we have*

$$q_{n+1}(\mathbf{x}_{n+1}, g_{n+1}) = \sum_{i=0}^{L} m_{n+1}(\mathbf{x}_{n+1}^q(B_i)) \mathbf{1}_{z \in B_i}.$$

Intuitively, the updated quantile is the updated GP mean at the $\mathbf{X}_{\mathrm{MC}}$ point that depends on which interval $g_{n+1}$ (or equivalently, $z$) falls.

Figure 1 provides an example for $l = 5$, and $\alpha = 40\%$. The values of $a$ and $b$ are given by a GP model, which allows us to draw the straight lines as a function of $z$. Each line corresponds to a point $\mathbf{x}_{\mathrm{MC}}^i$. The intersections for which the quantile point changes are shown by the vertical lines. For each interval, the segment corresponding to the quantile point (second lowest line) is shown in bold. We see that depending on the value of $z$ (that is, the value of $g_{n+1}$), the quantile point changes. In the example, $j_i$ takes successively as values 2, 3, 1, 4, 3, and 5.

**Remark** 2. Although the number of intersections grows quadratically with the MC sample size, finding the set of quantile points can be done very efficiently, based on two important elements: the number of distinct quantile points is much smaller than the number of intersections; there are at most two changes in the order of the straight lines moving from an interval to another adjacent one (nothing changes except the positions of the intersected lines which are inverted). This latter feature allows us to avoid numerous calls to sorting functions. An efficient algorithm to extract the quantile points indices and effective intervals is given in Appendix 6.4.

### 3.2. Infill criterion based on probability of exceedance

Proposition 1 allows us to express the quantile estimator at step $n + 1$ as

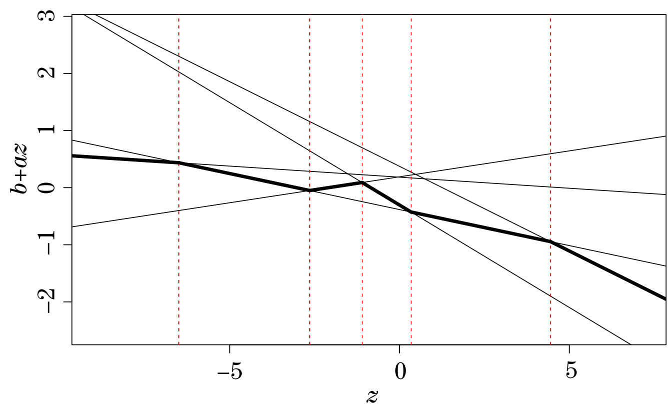


Figure 1. Evolution of the quantile point as a function of the value of $z$. Each plain line represents a point of $\mathbf{X}_{\mathrm{MC}}$, and the vertical lines the relevant intersections $I_i$. The second lowest line is shown in bold.

a function of the candidate point $\mathbf{x}_{n+1}$ and corresponding value $g_{n+1}$. In this section, we use this formulation to define a SUR criterion, an uncertainty measure related to our estimator that can be minimized by a proper choice of $\mathbf{x}_{n+1}$.

This criterion is inspired by related work in probability of failure estimation (Bect et al. (2012)) and multi-objective optimization (Picheny (2013)), that take advantage of the closed-form expressions of probabilities of exceeding thresholds in the GP framework. Our idea is to express the quantile estimation problem in terms of probability of exceedance in order to obtain a criterion in closed form.

By definition, the quantile is related to the probability by

$$\mathbb{P}(G(X) \geq q(G(X))) = 1 - \alpha. \tag{3.8}$$

The probability $\mathbb{P}(G(\mathbf{x}) \geq q_n | \mathcal{A}_n)$, available for any $\mathbf{x} \in \mathbb{X}$, is in the ideal case ($G$ is exactly known) either zero or one and, if $q_n = q(G(X))$, the proportion of ones is exactly $1 - \alpha$. At step $n$, a measure of error is then

$$H_n^{\mathrm{prob}} = \left| \int_{\mathbb{X}} \mathbb{P}(G(\mathbf{x}) \geq q_n | \mathcal{A}_n) d\mathbf{x} - (1 - \alpha) \right| = \left| \Gamma_n - (1 - \alpha) \right|, \tag{3.9}$$

with $\Gamma_n = \int_{\mathbb{X}} \mathbb{P}\left(G(\mathbf{x}) \geq q_n | \mathcal{A}_n\right) d\mathbf{x}$.

Following the SUR paradigm, we want to add at step $n+1$ an observation $(\mathbf{x}_{n+1}, g_{n+1})$ such that $H_{n+1}^{\mathrm{prob}}$ is minimal. However, computing $H_{n+1}^{\mathrm{prob}}$ requires evaluating $g_{n+1}$ (to obtain the updated distribution of $G(\mathbf{x})$ and updated value of $q_{n+1}$), which prevents us from searching for the optimal $\mathbf{x}_{n+1}$.

To circumvent this problem, we replace $g_{n+1}$ by its distribution conditional on $\mathcal{A}_n$ and then take the expectation of $H_{n+1}^{\mathrm{prob}}$ on this law. Writing $G_{n+1} =$

$G(\mathbf{x}_{n+1})$ as a random variable following this conditional distribution, we define $A_{n+1} = \mathcal{A}_n \cup (\mathbf{x}_{n+1}, G_{n+1})$, that is random through $G_{n+1}$. We can then choose the criterion to minimize (indexed by $\mathbf{x}_{n+1}$ to make the dependency explicit):

$$J_n^{\mathrm{prob}}(\mathbf{x}_{n+1}) = \left|\mathbb{E}\left(\Gamma_{n+1}(\mathbf{x}_{n+1})\right) - (1-\alpha)\right|, \tag{3.10}$$

where now,

$$\Gamma_{n+1}(\mathbf{x}_{n+1}) = \int_{\mathbb{X}} \mathbb{P}\left(G(\mathbf{x}) \geq Q_{n+1} | A_{n+1}\right) d\mathbf{x}. \tag{3.11}$$

**Proposition 2.** *Under our first strategy,*

$$\begin{aligned}
J_n^{prob}(\mathbf{x}_{n+1}) = \Bigg| \int_{\mathbb{X}} \sum_{i=1}^{L-1} \Bigg[ & \Phi_{r_i^n}\left(e_n^i(\mathbf{x}_{n+1};\mathbf{x}), f_n^i(\mathbf{x}_{n+1}, I_{i+1})\right) \\
& - \Phi_{r_i^n(\mathbf{x}_{n+1},\mathbf{x})}\left(e_n^i(\mathbf{x}_{n+1};\mathbf{x}), f_n(\mathbf{x}_{n+1}, I_i)\right) + \Phi_{r_i^n}\left((e_n^i(\mathbf{x}_{n+1};\mathbf{x}), f_n^i(\mathbf{x}_{n+1}, I_1)\right) \\
& + \Phi_{-r_i^n}\left(e_n^i(\mathbf{x}_{n+1};\mathbf{x}), -f_n^i\left((\mathbf{x}_{n+1}, I_L\right)\right) \Bigg] d\mathbf{x} - (1-\alpha) \Bigg|,
\end{aligned}$$

*where*

$$e_n^i(\mathbf{x}_{n+1};\mathbf{x};\mathbf{x}_{n+1}^q(B_i)) = \frac{m_n(\mathbf{x}) - m_n(\mathbf{x}_{n+1}^q(B_i))}{\sigma_W}, \;\; f_n^i(\mathbf{x}_{n+1}; I_i) = I_i s_n(\mathbf{x}_{n+1}),$$

$$\sigma_W = s_n(\mathbf{x})^2 + \frac{k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1})^2}{s_n(\mathbf{x}_{n+1})^2} - 2\frac{k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) k_n(\mathbf{x}, \mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1})^2}$$

*and* $\Phi_{r_n^i}$ *is the cumulative distribution function (CDF) of the centered Gaussian law of covariance matrix* $\begin{pmatrix} 1 & r_n^i \\ r_n^i & 1 \end{pmatrix}$, *with*

$$r_n^i = \frac{k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) - k_n(\mathbf{x}, \mathbf{x}_{n+1})}{\sqrt{\begin{matrix} s_n(\mathbf{x})^2 + (k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1})^2)/(s_n(\mathbf{x}_{n+1})^2) \\ -2(k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) k_n(\mathbf{x}, \mathbf{x}_{n+1}))/(s_n(\mathbf{x}_{n+1})^2) \end{matrix}}\, s_n(\mathbf{x}_{n+1})}.$$

The proof is deferred to the Appendix.

This criterion has a favourable form since it writes as a function of GP quantities at step $n$, $m_n$, $s_n$ and $k_n$, that can be computed quickly once the model is established. It does not require conditional simulations, as does the criterion in Jala et al. (2016), with advantages in computational cost and evaluation precision.

Evaluating this criterion does require a substantial computational effort, as it takes the form of an integral over $\mathbb{X}$, which must be done numerically. An obvious choice here is to use the set $\mathbf{X}_{\mathrm{MC}}$ as integration points. It also relies on the bivariate Gaussian CDF, which must be computed numerically. Efficient programs can be found, such as the R package `pbivnorm` (Kenkel (2012)), which make this task relatively inexpensive.

### 3.3. Infill criterion based on the quantile variance

Accounting for the fact that $J^{\text{prob}}$ is still expensive to compute, we propose an alternative that does not require numerical integration over $\mathbb{X}$.

It is based on choosing the point that has a *maximal effect* on the posterior value of the estimator. The variance of the updated estimator, $\text{Var}\,(q_{n+1}|A_{n+1})$ with $G_{n+1}$ random, is a good indicator of this potential effect, as it measures the sensitivity of $q_{n+1}$ to the possible values of $g(\mathbf{x}_{n+1})$.

Our second strategy is then

$$J_n^{\text{Var}}(\mathbf{x}_{n+1}) = \text{Var}_{G_{n+1}}(q_{n+1}|A_{n+1}), \tag{3.12}$$

where once again $A_{n+1}$ denotes the conditioning on $\mathcal{A}_n \cup (\mathbf{x}_{n+1}, G_{n+1})$, with $G_{n+1}$ random.

It is straightforward to see that choosing $\mathbf{x}_{n+1} \in \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ results in $\text{Var}(q_{n+1}|A_{n+1}) = \text{Var}(q_n|\mathcal{A}_n) = 0$.

**Proposition 3.** *Conditionally on $\mathcal{A}_n$ and on the choice of $\mathbf{x}_{n+1}$,*

$$\begin{aligned}
J_n^{Var}(\mathbf{x}_{n+1}) &= \sum_{i=1}^{L} \left[k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1})\right]^2 V(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i) P_i \\
&+ \sum_{i=1}^{L} \left[m_n(\mathbf{x}_{n+1}^q(B_i) - k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) E(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i)\right]^2 (1 - P_i)\, P_i \\
&- 2 \sum_{i=2}^{L} \sum_{j=1}^{i-1} \left[m_n(\mathbf{x}_{n+1}^q(B_i) - k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) E(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i)\right] P_i \\
&\quad \left[m_n(\mathbf{x}_{n+1}^q(B_i) - k_n(\mathbf{x}_{n+1}^q(B_i), \mathbf{x}_{n+1}) E(s_n(\mathbf{x}_{n+1}), I_{j+1}, I_j)\right] P_j \\
&\quad \textit{if } s_n(\mathbf{x}_{n+1}) \neq 0 \textit{ and 0 otherwise,}
\end{aligned}$$

*where*

$$\begin{aligned}
&P_i = \Phi(s_n(\mathbf{x}_{n+1}) I_{i+1}) - \Phi(s_n(\mathbf{x}_{n+1}) I_i), \\
&E(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i) = \frac{1}{s_n(\mathbf{x}_{n+1})} \left( \frac{\phi(s_n(\mathbf{x}_{n+1}) I_{i+1}) - \phi(s_n(\mathbf{x}_{n+1}) I_i)}{\Phi(s_n(\mathbf{x}_{n+1}) I_{i+1}) - \Phi(s_n(\mathbf{x}_{n+1}) I_i)} \right), \\
&V(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i) = \\
&\frac{1}{s_n(\mathbf{x}_{n+1})^2} \left[ 1 + \frac{s_n(\mathbf{x}_{n+1}) \phi(I_{i+1}) - s_n(\mathbf{x}_{n+1}) \phi(I_i)}{\Phi(I_{i+1}) - \Phi(I_i)} - \left( \frac{\phi(I_{i+1}) - \phi(I_i)}{\Phi(I_{i+1}) - \Phi(I_i)} \right)^2 \right],
\end{aligned}$$

*for $\Phi$ and $\phi$ respectively the CDF and density function of the standard Gaussian law.*

The proof is deferred to the Appendix.

This criterion writes only as a function of GP quantities at step $n$, $m_n$, $s_n$, $k_n$, and $\mathbf{x}_{n+1}$. It does not require numerical integration nor the bivariate CDF.

Intuitively, a point $\mathbf{x}_{n+1}$ has a large $J_n^{\mathrm{Var}}$ if it changes significantly the GP mean in regions that are critical to defining the quantile. In contrast, an observation $\mathbf{x}_{n+1}$ in either well-known regions (low GP variance) or non-critical ones (GP mean very different from the quantile estimator) would not change the estimator, regardless of the value of $g_{n+1}$. In that sense, it realizes a trade-off between exploitation and exploration.

Figure 2 is an illustration that shows how different values of $\mathbf{x}_{n+1}$ and $g_{n+1}$ affect the estimator. Here, an initial model with five observations is updated with either $\mathbf{x}_{n+1} = 0.97$ (left) or $\mathbf{x}_{n+1} = 0.5$ (right), and the 15% quantile is considered. Different updated values of the GP mean ($m_{n+1}$) and quantile ($q_{n+1}$) are shown, depending on the value taken by $g_{n+1}$. Here, we show the values corresponding to the middle of each interval $B_i$. We see that for $\mathbf{x}_{n+1} = 0.97$, even if $g_{n+1}$ takes extreme values, the quantile does not change significantly (low $J_n^{\mathrm{Var}}$). In constrast, for $\mathbf{x}_{n+1} = 0.5$, different values of $g_{n+1}$ lead to different shapes of the GP mean and consequently different values of the quantile (high $J_n^{\mathrm{Var}}$). Hence, the point $\mathbf{x}_{n+1} = 0.5$ can be considered as highly informative for our estimator, while $\mathbf{x}_{n+1} = 0.97$ is not.

### 3.4. Practical recommendations

**Finding the new observation** Finding $\mathbf{x}_{n+1}$ (Equation (2.5)) requires solving an optimization problem, that may not be straightforward. To ease this step, we propose that a (large) set of candidates be generated from the distribution of $\mathbf{X}$, from which a shorter set of "promising" points is extracted. Those points are drawn randomly from the large set with weights equal to $\phi\left(q_n - m_n(\mathbf{x})/s_n(\mathbf{x})\right)$, so that higher weights are given to points either close to the current quantile estimate and/or with high uncertainty. The criterion is evaluated on this subset of points and the best is chosen as the next infill point. In addition, a local optimization can be performed (for instance the BFGS algorithm, see Liu and Nocedal (1989)), starting from the best point of the subset.

**Choosing $\mathbf{X}_{\mathbf{MC}}$** The size of $\mathbf{X}_{\mathrm{MC}}$, which affects the precision of the criteria, is limited in practice by computational costs, in particular with $J^{\mathrm{prob}}$ (in our

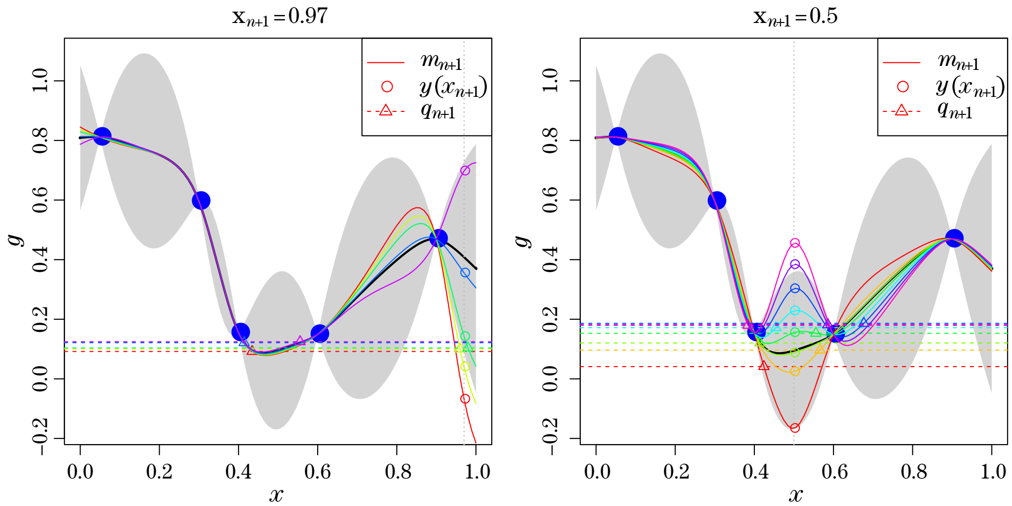


Figure 2. Illustration of $J^{\text{Var}}$. The GP model is shown in bold line and grey area. Updated GP means (plain lines) are shown depending on the value of $g_{new}$ (circles) for either $\mathbf{x}_{n+1} = 0.97$ (left) or $\mathbf{x}_{n+1} = 0.5$ (right). The corresponding 15% quantiles $q_{n+1}$ are shown with dotted horizontal lines, along with the quantile points $\mathbf{x}^q_{n+1}$ (triangles).

implementation, the maximum size is of the order of $10^4$). However, we found that renewing $\mathbf{X}_{\text{MC}}$ at each iteration sufficiently mitigates this issue.

**Budget and stopping criteria** Choosing the size of the initial observation set and the number of iterations is a classical issue with GP-based algorithms. A common rule-of-thumb is to use $n_0 = 5 \times d$ for the initial set. For $J^{\text{Var}}$, the iterative procedure can stop when the maximum of the criterion is below a small threshold. For $J^{\text{prob}}$, one can consider the difference $J_n^{\text{prob}} - J_{n+1}^{\text{prob}}$. In the following section, predetermined numbers of observations are used, and we use one third of the observations for the initial set. In an optimization context, the choice of this proportion was found as not significant compared to other factors (Picheny, Wagner and Ginsbourger (2013)).

## 4. Experiments

### 4.1. Two-dimensional example

As an illustrating example, we use here the classical Branin test function (Dixon and Szegö (1978), see Equation (C.1) in Appendix). On $[0,1]^2$, the range of this function is approximately $[0, 305]$.

We took $X_1, X_2 \sim \mathcal{U}[0,1]$, and searched for the 85% quantile. The initial set

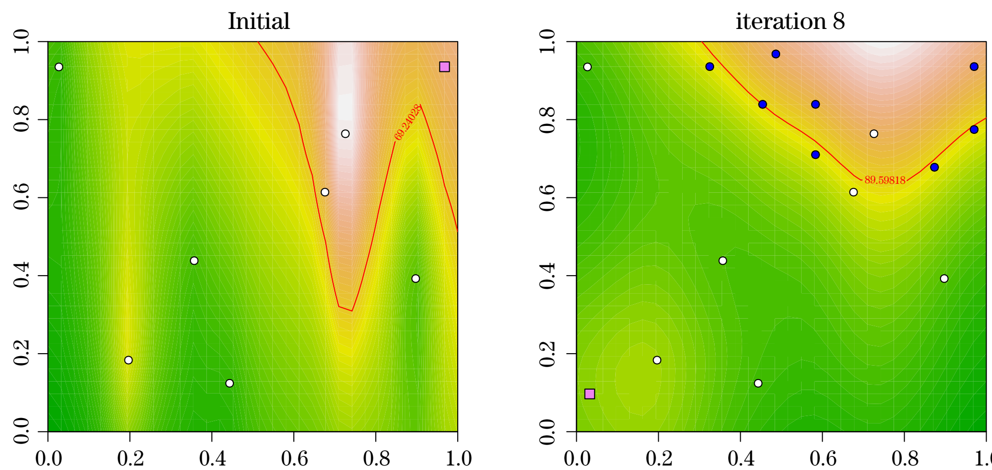


Figure 3. Contour lines of the GP mean and experimental set at $n = 7$ (left) and $n = 15$ (right) with $J^{\mathrm{Var}}$. The initial observations are shown with white circles, the observations added by the sequential strategy with plain circles, and the next point to evaluate with squares. The line shows the contour corresponding to the quantile estimate.

of experiments consisted of seven observations generated using Latin Hypercube Sampling (LHS), and 15 observations were added sequentially using both SUR strategies. The GP models learning, prediction, and update was performed using the `R` package `DiceKriging` (Roustant, Ginsbourger, and Deville (2012)). The covariance was chosen as Matérn 3/2 and the mean as a linear trend.

For $\mathbf{X}_{\mathrm{MC}}$, we used a 1000-point uniform sample on $[0, 1]^2$. For simplicity purpose, the search of $\mathbf{x}_{n+1}$ was performed on $\mathbf{X}_{\mathrm{MC}}$, although a continuous optimizer algorithm could have been used here. The actual quantile was computed using a $10^5$-point sample.

Figure 3 shows the set of experiments, along with contour lines of the GP model mean, for two intermediate stages of the $J^{\mathrm{Var}}$ run, to reveal the dynamics of our strategy. From the initial design of experiments, the top right corner of the domain was identified as the region containing the highest 15% values (Figure 3 left). Several observations were added in that region until the kriging approximation became accurate (blue circles, Figure 3 right), then a new region (bottom left corner) was explored (square point, Figure 3 right).

Figure 4 reports the final set of experiments and GP models obtained by both criteria, and Figure 5 the evolution of the estimators. The two strategies lead to relatively similar observation sets, that mostly consist of values close to the contour line corresponding to the $85^{th}$ quantile (exploitation points), and a few

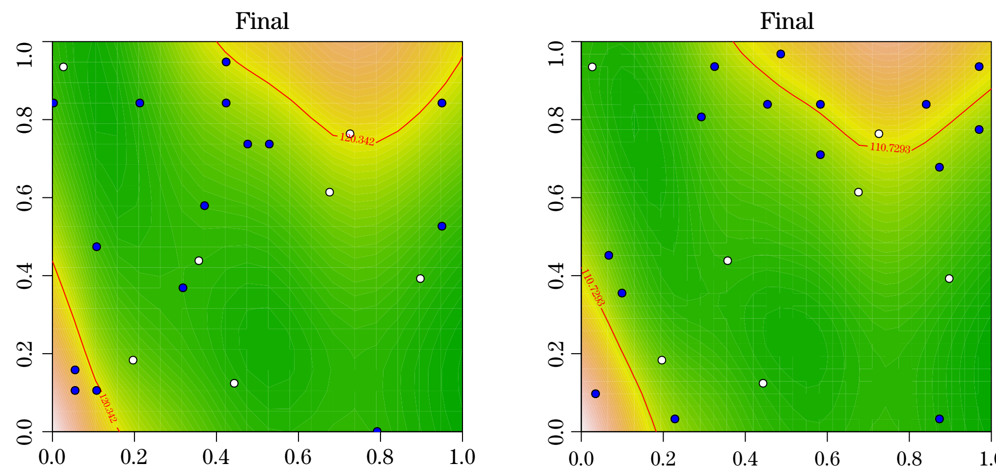


Figure 4. Comparison of observation sets obtained using $J^{\text{prob}}$ (left) and $J^{\text{Var}}$ (right).

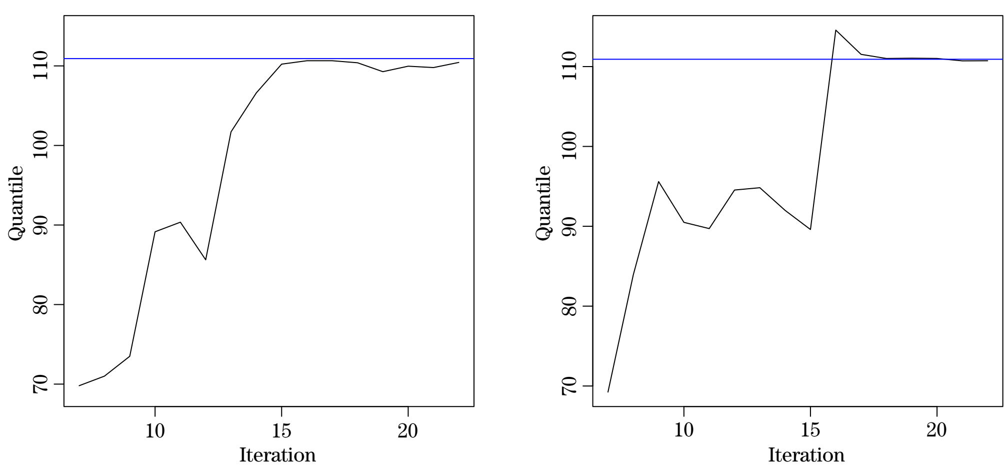


Figure 5. Evolution of the quantile estimates using $J^{\text{prob}}$ (left) and $J^{\text{Var}}$ (right) for the 2D problem. The horizontal line shows the actual $85^{th}$ quantile.

space-filling points (exploration points). With 18 observations, both estimators are close to the actual value (in particular with respect to the range of the function), yet additional observations might be required to achieve convergence (Figure 5).

### 4.2. Four- and six-dimensional examples

We consider now two more difficult test functions, with four and six dimensions, respectively (*hartman* and *ackley* functions, see (C.2) and (C.3) in Appendix). Both are widely used to test optimization strategies (Dixon and Szegö (1978)), and are bowl-shaped, multi-modal functions.

We took $\mathbf{X} \sim \mathcal{N}(1/2, \Sigma)$, with $\Sigma$ a symmetric matrix with diagonal elements equal to 0.1 and other elements equal to 0.05. The initial set of observations was taken as a 30-point LHS generated from the density of $\mathbf{X}$ (Helton and Davis (2003)), and 60 observations were added sequentially. A 3,000-point sample from the distribution of $\mathbf{X}$ was used for $\mathbf{X}_{\mathrm{MC}}$ (renewed at each iteration), and the actual quantile was computed using a $10^5$-point sample. The GP covariance was chosen as Matérn 3/2 and the mean as a linear trend.

The criteria were optimized as follow: a set of $10^5$ candidates was generated from the distribution of $\mathbf{X}$, out of which a subset of 300 promising points was extracted to evaluate the criterion, as described in Section 3.4. For $J^{\mathrm{Var}}$ a local optimization was performed, starting from the best point of the subset (using the BFGS algorithm, see Liu and Nocedal (1989)). Due to computational constraints, this step was not applied to $J^{\mathrm{prob}}$. Preliminary experiments have shown that only a limited gain is achieved by this step.

As an baseline strategy for comparison purpose, we included a "random search": $\mathbf{x}_{n+1}$'s were sampled randomly from the distribution of $X$. We also included the two-step approach, 30 initial observations and 60 additions, as proposed in Oakley (2004).

Quantile levels were considered to cover a variety of situations: 5% and 97% for the 4D problem and 15% and 97% for the 6D problem. Due to the bowl-shape of the functions, low levels are defined by small regions close to the center of the support of $X$, while high levels correspond to the edges of the support of $X$. Besides, it is reasonable to assume that levels farther away from 50% are more difficult to estimate.

As an error metric $\varepsilon$, we took the absolute difference between the quantile estimator and its actual value. We show this error as a percentage of the variation range of the test function. Since $X$ is not bounded, the range is defined as the difference between the 0.05 and 0.95 quantiles of $g(X)$.

To assess the robustness of our approach, the experiments were run ten times for each case, starting with a different initial set of observations. The evolution of the estimators (average, lowest and highest error metric values over the ten runs) is given in Figure 6. Table 1 shows the final precision metric of all alternatives.

We see that, except for 4D, $\alpha = 0.97$, and $J^{\mathrm{prob}}$, on average both strategies provide estimates with less than 2% error after approximately 30 iterations (for

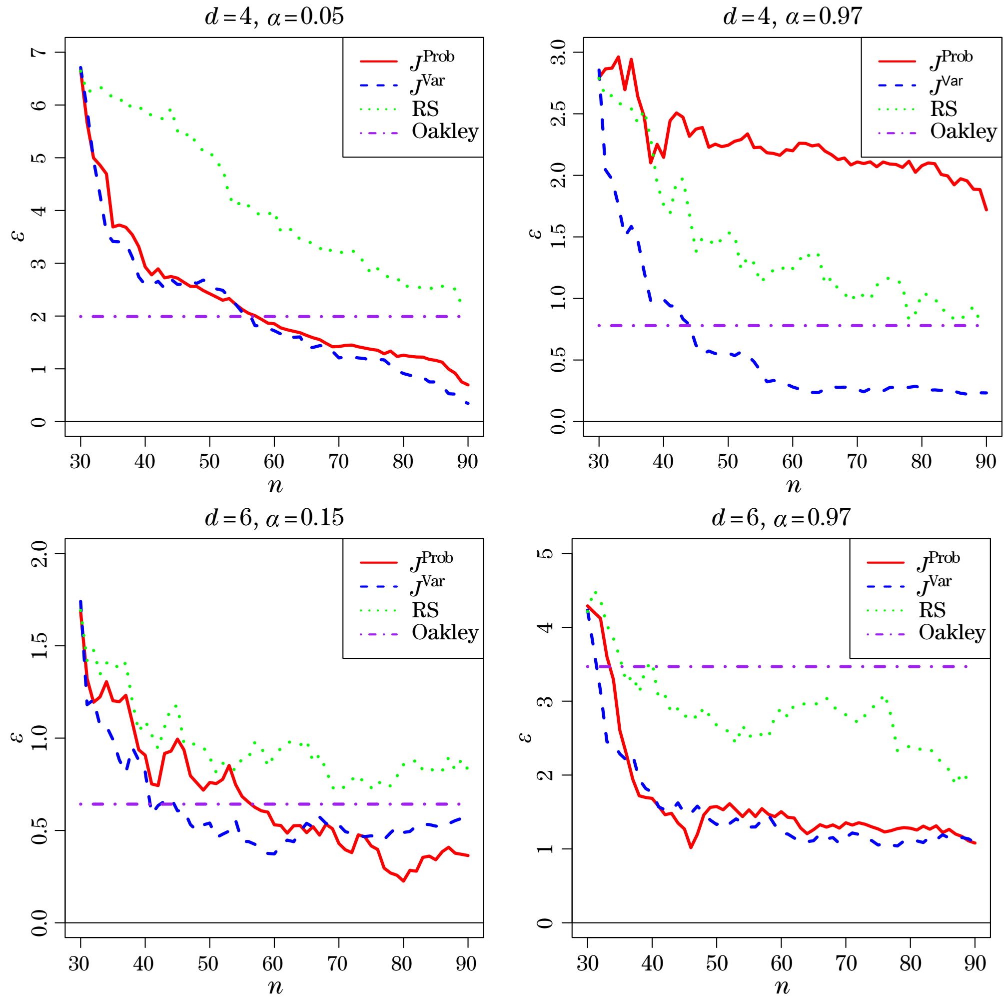


Figure 6. Evolution of the quantile estimates using $J^{\text{prob}}$, $J^{\text{Var}}$, random search (RS) or Oakley's two-step approach, for the 4D and 6D problems and several quantile levels. The lines show the average error. Note that since Oakley's approach is not sequential, the corresponding lines represent the estimates based on 90 observations.

a total of 60 function evaluations), which plainly justifies the use of GP models and sequential strategies in a constrained budget context.

For $d = 4$, $\alpha = 0.05$, both methods seem to converge to the actual quantile. For $d = 4$, $\alpha = 0.97$, $J^{\text{prob}}$ performs surprisingly poorly; we conjecture that a more exploratory behavior (compared to $J^{\text{Var}}$) hinders its performance here. $J^{\text{Var}}$ reaches a good estimate quickly with less that 1% error, yet seems to then converge slowly to the exact solution. This might be explained by the relative

Table 1. Average error metrics based on 90 observations.

| Pb | Random search | Oakley | $J^{\text{prob}}$ | $J^{\text{Var}}$ |
|---|---|---|---|---|
| $d = 4, \alpha = 0.05$ | 2.23 | 1.99 | 0.69 | 0.35 |
| $d = 4, \alpha = 0.97$ | 0.85 | 0.78 | 1.72 | 0.23 |
| $d = 6, \alpha = 0.15$ | 0.83 | 0.64 | 0.36 | 0.56 |
| $d = 6, \alpha = 0.97$ | 1.86 | 3.47 | 1.08 | 1.08 |

mismatch between the GP model and the test function.

For $d = 6$, $\alpha = 0.15$, both approaches reach consistently less than 1% error. However, they only moderately outperform the random search strategy here. This might indicate that for central quantile values, less gain can be achieved by sequential strategies, as a large region of the design space needs to be learned to characterize the quantile, making space-filling strategies, for instance, competitive.

For $d = 6$, $\alpha = 0.97$, both approaches largely outperform random search, yet after a first few very efficient steps seem to converge only slowly to the actual quantile.

From Table 1, we see that both SUR approaches substantially outperform random search and Oakley's two-step approach, except $J^{\text{prob}}$ for $d = 4, \alpha = 0.97$. $J^{\text{prob}}$ is the best approach for $d = 6, \alpha = 0.15$, both SUR strategies perform silimiary for $d = 6, \alpha = 0.97$, and $J^{\text{Var}}$ is best for the two other problems. Interestingly, Oakley's approach is outperformed by random search for $d = 6, \alpha = 0.97$. This can be explained by the difficulty of the approach, acknowledged by the authors, when the quantile is defined by several distinct regions.

In general, those experiments show the ability of our approach to handle multi-modal black-box functions, with input space dimensions typical of GP-based approaches.

## 5. Concluding Comments

We have proposed two sequential Bayesian strategies for quantile estimation. They rely on the analytical update formula for the GP-based estimator, obtained thanks to the particular form of the GP equations and the introduction of the *quantile point* concept. Two criteria have then been proposed for which closed-form expression have been derived, hence avoiding the use of computationally intensive conditional simulations. Numerical experiments in dimensions two to six have demonstrated the potential of both approaches.

Some limitations of the proposed method call for future improvements. The strategies rely on the set $\mathbf{X}_{\mathrm{MC}}$, whose size is in practice limited by the computational resources to a couple of thousands at most. This may hinder the use of our method for extreme quantile estimation, or for highly multi-modal functions. Combining adaptive sampling strategies or subset selection methods with our approaches may prove useful in this context.

Accounting for the GP model error (due to an inaccurate estimation of its hyper-parameters or a poor choice of kernel) is also an important task which could greatly improve the efficiency and robustness of the approach. A fully Bayesian approach as in Kennedy and O'Hagan (2001); Gramacy and Lee (2008) could address this issue, yet at the price of additional computational expense.

## Supplementary Materials

The `R` code implementing the methods described in this article is provided as supplementary material.

## Acknowledgment

The authors would like to thank Damien Leroux (MIAT, Université de Toulouse, INRA) for his help on the efficient implementation of the algorithms on `R`.

## Appendix

### A. Proof of proposition 2

We denote by $\mathbb{E}_n$ and $\mathbb{P}_n$ the expectation and the probability conditionally on the event $\mathcal{A}_n$. Starting from (3.11), we have

$$\begin{aligned}\mathbb{E}\left(\Gamma_{n+1}(\mathbf{x}_{n+1})\right) &= \mathbb{E}\left[\int_{\mathbb{X}} \mathbb{P}(G(x) \geq q_{n+1})|A_{n+1})dx\right] \\ &= \int_{\mathbb{X}} \mathbb{E}\left[\mathbb{E}_n\left[\mathbf{1}_{G(x)\geq q_{n+1}(x_{n+1})}|G_{n+1}\right]\right]dx \\ &= \int_{\mathbb{X}} \mathbb{E}_n\left[\mathbf{1}_{G(x)\geq q_{n+1}(\mathbf{x}_{n+1})}\right]dx \\ &= \int_{\mathbb{X}} \mathbb{P}_n(G(x) \geq q_{n+1}(\mathbf{x}_{n+1}))dx.\end{aligned}$$

We get then that

$$J_n^{\mathrm{prob}}(\mathbf{x}_{n+1}) = \left|\int_{\mathbb{X}} \mathbb{P}_n(G(x) \geq q_{n+1}(\mathbf{x}_{n+1}))dx - (1-\alpha)\right|.$$

To get a closed form of our criterion, we have to develop $\mathbb{P}_n(G(x) \geq q_{n+1}(x_{n+1}))$. Writing $Z = G_{n+1} - m_n(\mathbf{x}_{n+1})/s_n(\mathbf{x}_{n+1})^2$, we have

$$
\begin{aligned}
\mathbb{E}_n\left(\mathbf{1}_{G(x)\geq q_{n+1}(\mathbf{x}_{n+1})}\right) &= \sum_{i=0}^{L} \mathbb{E}_n\left[\mathbf{1}_{G(x)\geq m_{n+1}(\mathbf{x}^q_{n+1}(B_i)}\mathbf{1}_{Z\in B_i}\right] \\
&= \sum_{i=1}^{L-1}\left(\mathbb{P}_n\left[G(x) \geq m_{n+1}(\mathbf{x}^q_{n+1}(B_i)) \cap Z \leq I_{i+1}\right]\right. \\
&\quad \left. - \mathbb{P}_n\left[G(x) \geq m_{n+1}(\mathbf{x}^q_{n+1}(B_i)) \cap Z \leq I_i\right)\right]) \\
&\quad + \mathbb{P}_n(G(x) \geq m_{n+1}(\mathbf{x}^q_{n+1}(B_1)) \cap Z \leq I_1) \\
&\quad + \mathbb{P}_n(G(x) \geq m_{n+1}(\mathbf{x}^q_{n+1}(B_L)) \cap Z \geq I_L)\ .
\end{aligned}
$$

Now,

$$
\begin{aligned}
T_n : &= \mathbb{P}_n\left(G(x) \geq m_{n+1}(\mathbf{x}^q_{n+1}(B_i)) \cap Z \leq I_i\right) \\
&= \mathbb{P}_n\left(m_{n+1}(\mathbf{x}^q_{n+1}(B_i)) - G(x) \leq 0 \cap Z \leq I_i\right)\ ,
\end{aligned}
$$

is the CDF of the couple $(m_{n+1}(\mathbf{x}^q_{n+1}(B_i) - G(x)), Z) := (W, Z)$ , at point $(0, I_i)$. This random vector, conditionally on $\mathcal{A}_n$ is Gaussian. We denote by $M$ and $R$ its mean vector and covariance matrix, respectively.

Thanks to (3.2), we have

$$
m_{n+1}(\mathbf{x}^q_{n+1}(B_i)) = m_n(\mathbf{x}^q_{n+1}(B_i)) - k_n(\mathbf{x}^q_{n+1}(B_i), \mathbf{x}_{n+1})Z\ , \tag{A.1}
$$

which gives

$$
M = \begin{pmatrix} m_n(\mathbf{x}^q_{n+1}(B_i)) - m_n(x) \\ 0 \end{pmatrix},\ R = \begin{pmatrix} \mathrm{Var}(W) & \mathrm{Cov}(W,Z) \\ \mathrm{Cov}(W,Z) & \mathrm{Var}(Z) \end{pmatrix}\ ,
$$

$$
\begin{aligned}
\mathrm{Var}(W) := \sigma_W = s_n(x)^2 &+ \frac{k_n(\mathbf{x}^q_{n+1}(B_i), \mathbf{x}_{n+1})^2}{s_n(\mathbf{x}_{n+1})^2} \\
&- 2\frac{k_n(\mathbf{x}^q_{n+1}(B_i), \mathbf{x}_{n+1})k_n(x, \mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1})^2},
\end{aligned}
$$

$$
\mathrm{Cov}(W,Z) = \frac{k_n(\mathbf{x}^q_{n+1}(B_i), \mathbf{x}_{n+1}) - k_n(x, \mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1})^2} \text{ and } \mathrm{Var}(Z) = \frac{1}{s_n(\mathbf{x}_{n+1})^2}\ .
$$

We can conclude by centering and normalizing that

$$
\begin{aligned}
T_n &= \mathbb{P}_n\left(W \leq 0 \cap Z \leq I_i\right) \\
&= \mathbb{P}\left(\frac{W - (m_n(\mathbf{x}^q_{n+1}(B_i)) - m_n(x))}{\sqrt{Var(W)}}\right.
\end{aligned}
$$

$$\leq \frac{m_n(x) - m_n(\mathbf{x}^q_{n+1}(B_i))}{\sqrt{Var(W)}} \cap s_n(\mathbf{x}_{n+1})Z \leq I_i s_n(\mathbf{x}_{n+1})\Bigg)$$

$$:= \mathbb{P}\left(S \leq e^i_n(\mathbf{x}_{n+1}; x; \mathbf{x}^q_{n+1}(B_i)) \cap T \leq f^i_n(\mathbf{x}_{n+1}; I_i)\right) \ ,$$

where $(S,T)$ is a Gaussian random vector of law $\mathcal{N}\left(0, \begin{pmatrix} 1 & r^i_n \\ r^i_n & 1 \end{pmatrix}\right)$ with

$$r^i_n := r_n(\mathbf{x}_{n+1}; x; \mathbf{x}^q_{n+1}(B_i)) = \frac{k_n(\mathbf{x}^q_{n+1}(B_i), \mathbf{x}_{n+1}) - k_n(x, \mathbf{x}_{n+1})}{\sqrt{\mathrm{Var}(W)} s_n(\mathbf{x}_{n+1})} \ .$$

$$e^i_n(\mathbf{x}_{n+1}; x; \mathbf{x}^q_{n+1}(B_i)) = \frac{m_n(x) - m_n(\mathbf{x}^q_{n+1}(B_i))}{\sqrt{\mathrm{Var}(W)}} \ ,$$

and

$$f^i_n(\mathbf{x}_{n+1}; I_i) = I_i s_n(\mathbf{x}_{n+1}).$$

Finally, we get for $1 \leq i \leq L$,

$$\mathbb{P}_n(G(x) \geq m_{n+1}(\mathbf{x}^q_{n+1}(B_i)) \cap Z \leq I_i)$$
$$= \Phi_{r^i_n}\left(e^i_n(\mathbf{x}_{n+1}; x; \mathbf{x}^q_{n+1}(B_i)), f^i_n(\mathbf{x}_{n+1}; I_i)\right) \ ,$$

where we denote by $\Phi_r$ the cumulative distribution function of the centered Gaussian random vector of covariance matrix $\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$ .

Similarly, for $0 \leq i \leq L$,

$$\mathbb{P}_n(G(x) \geq m_{n+1}(\mathbf{x}^q_{n+1}(B_i)) \cap Z \leq I_{i+1})$$
$$= \Phi_{r^i_n}\left(e^i_n(\mathbf{x}_{n+1}; x; \mathbf{x}^q_{n+1}(B_i)), f^i_n(\mathbf{x}_{n+1}; I_{i+1})\right) \ ,$$
$$\mathbb{P}_n(G(x) \geq m_{n+1}(\mathbf{x}^q_{n+1}(B_i)) \cap Z \geq I_L)$$
$$= \Phi_{-r^i_n}\left(e^i_n(\mathbf{x}_{n+1}; x; \mathbf{x}^q_{n+1}(B_i)), -f^i_n(\mathbf{x}_{n+1}; I_L)\right) \ .$$

## B. Proof of proposition 3

**Lemma 1.** *Let $E_1, \ldots E_n$ be mutually exclusive and exhaustive events. Then, for a random variable $U$,*

$$Var(U) = \sum_{i=1}^n \mathrm{Var}(U \mid E_i)\mathbb{P}(E_i) + \sum_{i=1}^n \mathbb{E}(U \mid E_i)^2(1 - \mathbb{P}(E_i))\mathbb{P}(E_i)$$
$$- 2\sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E}(U \mid E_i)\mathbb{P}(E_i)\mathbb{E}(U \mid E_j)\mathbb{P}(E_j) \ .$$

In our case, we want to compute $\mathrm{Var}(q_{n+1}(\mathbf{x}_{n+1})|\mathcal{A}_n) := \mathrm{Var}_n(q_{n+1}(\mathbf{x}_{n+1}))$. Since the events $\{Z \in B_i\}_{1 \leq i \leq L}$ are mutually exclusive and exhaustive, we can apply Lemma 1,

$$
\begin{aligned}
\mathrm{Var}_n(q_{n+1}(\mathbf{x}_{n+1})) = &\sum_{i=1}^{L} \mathrm{Var}_n(m_{n+1}(\mathbf{x}^q_{n+1}(B_i))|Z \in B_i)\mathbb{P}_n(Z \in B_i) \\
&+ \sum_{i=1}^{L} \mathbb{E}_n\left(m_{n+1}(\mathbf{x}^q_{n+1}(B_i))|Z \in B_i\right)^2 \\
&\quad (1 - \mathbb{P}_n(Z \in B_i))\,\mathbb{P}_n(Z \in B_i) \\
&- 2\sum_{i=2}^{L}\sum_{j=1}^{i-1} \mathbb{E}_n\left(m_{n+1}(\mathbf{x}^q_{n+1}(B_i))|Z \in B_i\right)\mathbb{P}_n\left(Z \in B_i\right) \\
&\times \mathbb{E}_n\left(m_{n+1}(\mathbf{x}^q_{n+1}(B_j))|Z \in B_j\right)\mathbb{P}_n\left(Z \in B_j\right)\ .
\end{aligned}
$$

Thanks to (A.1), we get

$$
m_{n+1}(\mathbf{x}^q_{n+1}(B_i)) = m_n(\mathbf{x}^q_{n+1}(B_i)) - k_n(\mathbf{x}^q_{n+1}(B_i), \mathbf{x}_{n+1})Z\ .
$$

Then,

$$
\begin{aligned}
\mathrm{Var}_n(q_{n+1}(\mathbf{x}_{n+1})) = &\sum_{i=1}^{n} k_n(\mathbf{x}^q_{n+1}(B_i), \mathbf{x}_{n+1})^2\, \mathrm{Var}_n(Z|Z \in B_i)\mathbb{P}_n(Z \in B_i) \\
&+ \sum_{i=1}^{L} \left(m_n(\mathbf{x}^q_{n+1}(B_i)) - k_n(\mathbf{x}^q_{n+1}(B_i), \mathbf{x}_{n+1})\mathbb{E}_n\left(Z|Z \in B_i\right)\right)^2 \\
&\times (1 - \mathbb{P}_n(Z \in B_i))\,\mathbb{P}_n(Z \in B_i) \\
&- 2\sum_{i=2}^{L}\sum_{j=1}^{i-1} \left(m_n(\mathbf{x}^q_{n+1}(B_i)) - k_n(\mathbf{x}^q_{n+1}(B_j), \mathbf{x}_{n+1})\right. \\
&\quad \left.\mathbb{E}_n\left(Z|Z \in B_j\right)\right)\mathbb{P}_n\left(Z \in B_i\right) \\
&\times \left(m_n(\mathbf{x}^q_{n+1}(B_j)) - k_n(\mathbf{x}^q_{n+1}(B_j), \mathbf{x}_{n+1})\mathbb{E}_n\left(Z|Z \in B_j\right)\right) \\
&\quad \mathbb{P}_n\left(Z \in B_j\right).
\end{aligned}
$$

Since $Z$ is a centered Gaussian random variable of variance $s_n(\mathbf{x}_{n+1})^{-2}$ we have

$$
P_i := \mathbb{P}_n(Z \in B_i) = \Phi(s_n(\mathbf{x}_{n+1})I_{i+1}) - \Phi(s_n(\mathbf{x}_{n+1})I_i)\ .
$$

To conclude, we have to find analytical forms for the quantities Var $(Z \mid I_i < Z < I_{i+1})$ and $\mathbb{E}\left(Z \mid I_i < Z < I_{i+1}\right)$. To do so, we use a result on truncated Gaussian random variable, see Tallis (1961).

**Lemma 2.** *If $U$ is a real random variable such that $U \sim \mathcal{N}(\mu, \sigma)$, and $u$ and $v$*

*are two real numbers, then*

$$\mathbb{E}(U|u < U < v) = \mu + \frac{\phi\left((v-\mu)/\sigma\right) - \phi\left((u-\mu)/\sigma\right)}{\Phi\left((v-\mu)/\sigma\right) - \Phi\left((w-\mu)/\sigma\right)}\sigma \ ,$$

$$\begin{aligned}\mathrm{Var}(U|u < U < v) = \sigma^2 \Bigg[ 1 + &\frac{(u-\mu)/\sigma\phi\left((u-\mu)/\sigma\right) - (v-\mu)/\sigma\phi\left((v-\mu)/\sigma\right)}{\Phi\left((v-\mu)/\sigma\right) - \Phi\left((u-\mu)/\sigma\right)} \\ &- \left(\frac{\phi\left((u-\mu)/\sigma\right) - \phi\left((v-\mu)/\sigma\right)}{\Phi\left((v-\mu)/\sigma\right) - \Phi\left((u-\mu)/\sigma\right)}\right)^2 \Bigg] \ .\end{aligned}$$

We apply Lemma 2 with $U = Z$, $u = I_i$ and $v = I_{i+1}$, and conclude that

$$\begin{aligned}&E(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i) \\ &:= E_n(Z|Z \in B_i) = \frac{\phi(s_n(\mathbf{x}_{n+1})I_i) - \phi(s_n(\mathbf{x}_{n+1})I_{i+1})}{\Phi(s_n(\mathbf{x}_{n+1})I_{i+1}) - \Phi(s_n(\mathbf{x}_{n+1})I_i)} \frac{1}{s_n(\mathbf{x}_{n+1})} \ , \\ &V(s_n(\mathbf{x}_{n+1}), I_{i+1}, I_i) := \mathrm{Var}_n(Z|Z \in B_i) \\ &= \frac{1}{s_n(\mathbf{x}_{n+1})^2} \Bigg[ 1 + \frac{I_i s_n(\mathbf{x}_{n+1})\phi(s_n(\mathbf{x}_{n+1})I_i) - s_n(\mathbf{x}_{n+1})I_{i+1}\phi(s_n(\mathbf{x}_{n+1})I_i)}{\Phi(s_n(\mathbf{x}_{n+1})I_{i+1}) - \Phi(s_n(\mathbf{x}_{n+1})I_i)} \\ &\quad - \left(\frac{\phi(s_n(\mathbf{x}_{n+1})I_i) - \phi(s_n(\mathbf{x}_{n+1})I_{i+1})}{\Phi(s_n(\mathbf{x}_{n+1})I_{i+1}) - \Phi(s_n(\mathbf{x}_{n+1})I_i)}\right)^2 \Bigg] \ .\end{aligned}$$

## C. Test functions

The two-dimensional Branin function is

$$g(\mathbf{x}) = \left(\bar{x}_2 - \frac{5.1\bar{x}_1^2}{4\pi^2} + \frac{5\bar{x}_1}{\pi} - 6\right)^2 + \left(10 - \frac{10}{8\pi}\right)\cos(\bar{x}_1) + 10 \qquad \text{(C.1)}$$

with: $\bar{x}_1 = 15 \times x_1 - 5$, $\bar{x}_2 = 15 \times x_2$.

The four-dimensional Hartman function is

$$g(\mathbf{x}) = \frac{-1}{1.94}\left[2.58 + \sum_{i=1}^{4} C_i \exp\left(-\sum_{j=1}^{4} a_{ji}\left(x_j - p_{ji}\right)^2\right)\right], \qquad \text{(C.2)}$$

with

$$\mathbf{C} = \begin{bmatrix} 1.0 \\ 1.2 \\ 3.0 \\ 3.2 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 10.00 & 0.05 & 3.00 & 17.00 \\ 3.00 & 10.00 & 3.50 & 8.00 \\ 17.00 & 17.00 & 1.70 & 0.05 \\ 3.50 & 0.10 & 10.00 & 10.00 \end{bmatrix},$$

$$\mathbf{p} = \begin{bmatrix} 0.1312 & 0.2329 & 0.2348 & 0.4047 \\ 0.1696 & 0.4135 & 0.1451 & 0.8828 \\ 0.5569 & 0.8307 & 0.3522 & 0.8732 \\ 0.0124 & 0.3736 & 0.2883 & 0.5743 \end{bmatrix}.$$

The six-dimensional Ackley function is

$$g(\mathbf{x}) = 20 + \exp(1) - 20\exp\left(-0.2\sqrt{\frac{1}{4}\sum_{i=1}^{4} x_i^2}\right) - \exp\left[\frac{1}{4}\sum_{i=1}^{4}\cos(2\pi x_i)\right] \quad \text{(C.3)}$$

### D. Enumerating quantile points

We provide an efficient algorithm for finding the quantile points, as described in Section 3.1. This amounts to finding all the indices of the empirical quantiles of $\mathbf{b}+\mathbf{a}z$, when $\mathbf{b}$ and $\mathbf{a}$ are fixed vectors of size $l$ and $z$ is a scalar that takes all values in $\mathbb{R}$.

An intuitive algorithm computes all the intersection points defined by all the combinations of $b_u - b_v/a_v - a_u$, then evaluates $\mathbf{b}+\mathbf{a}z$ with $z$ taking the value at the middle of the interval defined by two consecutive intersection points, then extracts the index of the quantile. However, this requires ordering $(l(l-1))/2+1$ times vectors of size $l$, which is computationally intensive when $l$ is large.

We propose the following algorithm, that avoids considering all the intersection points and does not require extracting vector quantiles, but only their minimal values. Its principle is, given the line index corresponding to the quantile for a value of $z$, to search which line intersects it first as $z$ increases. The algorithm starts at $z=-\infty$ and the initial quantile line corresponds to the quantile of $\mathbf{a}$ (the values of $\mathbf{b}$ being then negligible). The algorithm main loop stops when there are no more intersections ($z=+\infty$). The algorithm is given in pseudo-code in Algorithm 1.

**Algorithm 1** Pseudo-code for finding quantile points.

Set: $z=-\infty$, $j=$ index of the $\alpha-$quantile of $\mathbf{a}$, $\mathbf{J}=j$, $\mathbf{Z}=z$
**while** There exists $k$ such that $b_k - b_j/a_j - a_k > z$ **do**
  Find $k = \arg\min_{1\le r\le l, r\ne j} b_r - b_j/a_j - a_r$ such that $b_k - b_j/a_j - a_k > z$
  Update: $z = b_k - b_j/a_j - a_k$, $j=k$
  Save: $\mathbf{Z}=[\mathbf{Z},z]$, $\mathbf{J}=[\mathbf{J},j]$
**end while**
Return: $\mathbf{Z}$ (critical intersection points), $\mathbf{J}$ (indices of all quantile points).

## References

Arnaud, A., Bect, J., Couplet, M., Pasanisi, A. and Vazquez, E. (2010). Évaluation d'un risque d'inondation fluviale par planification séquentielle d'expériences. In *42èmes Journées de Statistique*. Marseille, France, France.

Bect, J., Ginsbourger, D., Li, L., Picheny, V. and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing* **22**, 773–793.

Bourinet, J.-M., Deheeger, F. and Lemaire, M. (2011). Assessing small failure probabilities by combined subset simulation and support vector machines. *Structural Safety* **33**, 343–353.

Bucher, C. G. and Bourgund, U. (1990). A fast and efficient response surface approach for structural reliability problems. *Structural Safety* **7**, 57–66.

Cannamela, C., Garnier, J. and Iooss, B. (2008). Controlled stratification for quantile estimation. *The Annals of Applied Statistics* , 1554–1580.

Chevalier, C., Ginsbourger, D., Bect, J., Vazquez, E., Picheny, V. and Richet, Y. (2014). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics* **56**.

Chevalier, C., Ginsbourger, D. and Emery, X. (2014). Corrected kriging update formulae for batch-sequential data assimilation. In *Mathematics of Planet Earth*. Springer, pp. 119–122.

Cornuet, J., MARIN, J.-M., Mira, A. and Robert, C. P. (2012). Adaptive multiple importance sampling. *Scandinavian Journal of Statistics* **39**, 798–812.

David and Nagaraja (2003). *Order Statistics*. Wiley.

de Freitas, N., Zoghi, M. and Smola, A. J. (2012). Exponential regret bounds for Gaussian process bandits with deterministic observations. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*.

Dixon, L. and Szegö, G. (1978). *Towards Global Optimisation 2*, vol. 2. North Holland.

Egloff, D., Leippold, M. et al. (2010). Quantile estimation with adaptive importance sampling. *The Annals of Statistics* **38**, 1244–1278.

Geman, D. and Jedynak, B. (1996). An active testing model for tracking roads in satellite images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **18**, 1–14.

Gramacy, R. B. and Lee, H. K. (2008). Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association* **103**.

Helton, J. C. and Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety* **81**, 23–69.

Jala, M., Lévy-Leduc, C., Moulines, E., Conil, E. and Wiart, J. (2012). Sequential design of computer experiments for parameter estimation with application to numerical dosimetry. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE.

Jala, M., Lévy-Leduc, C., Moulines, É., Conil, E. and Wiart, J. (2016). Sequential design of computer experiments for the assessment of fetus exposure to electromagnetic fields. *Technometrics* **58**, 30–42.

Jones, D. R., Schonlau, M. and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* **13**, 455–492.

Kenkel, B. (2012). *pbivnorm: Vectorized Bivariate Normal CDF*. R package version 0.5-1.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**, 425–464.

Liu, D. C. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical Programming* **45**, 503–528.

Močkus, J. (1975). On bayesian methods for seeking the extremum. In *Optimization Techniques IFIP Technical Conference*. Springer.

Morio, J. (2012). Extreme quantile estimation with nonparametric adaptive importance sampling. *Simulation Modelling Practice and Theory* **27**, 76–89.

Oakley, J. (2004). Estimating percentiles of uncertain computer code outputs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **53**, 83–93.

Picheny, V. (2013). Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction. *Statistics and Computing*, 1–16.

Picheny, V., Wagner, T. and Ginsbourger, D. (2013). A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization* **48**, 607–626.

Rasmussen, C. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Roustant, O., Ginsbourger, D. and Deville, Y. (2012). Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software* **51**, 1–55.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 409–423.

Scott, W., Frazier, P. and Powell, W. (2011). The correlated knowledge gradient for simulation optimization of continuous parameters using Gaussian process regression. *SIAM Journal on Optimization* **21**, 996–1026.

Stein, M. L. (2012). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media.

Tallis, G. M. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society. Series B (Methodological)*, 223–229.

Villemonteix, J., Vazquez, E. and Walter, E. (2009). An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization* **44**, 509–534.

Zhang, P. (1996). Nonparametric importance sampling. *Journal of the American Statistical Association* **91**, 1245–1253.

Institut de Mathématiques de Toulouse (CNRS UMR 5219), Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France.

E-mail: tatiana.labopin@math.univ-toulouse.fr

MIAT, Université de Toulouse, INRA, Castanet-Tolosan, France.

E-mail: victor.picheny@inra.fr

(Received March 2016; accepted February 2017)