Adaptive Designs of Experiments for Accurate Approximation of a Target Region

Victor Picheny, David Ginsbourger, Olivier Roustant, Raphael T. Haftka, Nam-Ho Kim

HAL Id: hal-00319385
https://hal.science/hal-00319385v2

Submitted on 22 Jun 2010

Adaptive Designs of Experiments for Accurate Approximation of a Target Region

Victor Picheny
Ecole Centrale Paris
victor.picheny@ecp.fr
David Ginsbourger
University of Bern
david.ginsbourger@stat.unibe.ch
Olivier Roustant
Ecole des Mines de St Etienne
roustant@emse.fr
Raphael T. Haftka
University of Florida
Email: haftka@ufl.edu
Nam-Ho Kim
University of Florida
nkim@ufl.edu

This paper addresses the issue of designing experiments for a metamodel that needs to be accurate for a certain level of the response value. Such a situation is common in constrained optimization and reliability analysis. Here, we propose an adaptive strategy to build designs of experiments that is based on an explicit trade-off between reduction of global uncertainty and exploration of regions of interest. A modified version of the classical integrated mean square error criterion is used that weights the prediction variance with the expected proximity to the target level of response. The method is illustrated by two simple examples. It is shown that a substantial reduction of error can be achieved in the target regions, with reasonable loss of global accuracy. The method is finally applied to a reliability analysis problem; it is found that the adaptive designs significantly outperform classical space-filling designs.

1   Introduction

In the past decades, the use of metamodeling techniques has been recognized to efficiently address the issues of prediction and optimization of expensive-to-compute numerical simulators or black-box functions [1, 2]. A metamodel (or surrogate model) is an approximation to system response constructed from its value at a limited number of selected input values, the design of experiments (DoE). In many engineering problems, the total number of function evaluations is drastically limited by computational cost; hence, it is of crucial interest to develop methods for efficiently selecting the experiments.

In this paper, we focus on a particular application where metamodels are used in a way that their accuracy is crucial for certain level-sets. This situation is common in two popular frameworks:

> In constrained optimization, the constraint function often relies on expensive calculations. For instance, a typical structural optimization formulation is to minimize a weight such that the maximum stress, computed by finite element analysis, does not exceed a certain value. When using a metamodel to approximate the constraint, it is of utmost importance that the approximation error is minimal on the boundary that separates the feasible designs from infeasible ones. Substantial errors for values far from the boundary, on the other hand, are not detrimental.
> In reliability analysis, a metamodel is often used to propagate the uncertainty of random input variables to the performance function of a system [3, 4]. In particular, the probability of failure of the system can be computed using sampling techniques (i.e. Monte-Carlo Simulations, MCS), by counting the number of responses that are above a certain threshold. The contour line of the response equal to the threshold must be known accurately to discriminate between samples.

The objective of the present work is to provide a methodology to construct a design of experiments such that the metamodel accurately approximates the vicinity of a boundary in design space defined by a target value of the function of interest. Mourelatos et al. [5] used a combination of global and local metamodels to first detect the critical regions and then obtain a locally accurate approximation. Ranjan et al. [6] proposed a modified version of the famous EGO algorithm (Efficient Global Optimization, [7]) to sequentially explore the domain region along a contour line. Tu et al. used a modified D-optimal strategy for boundary-focused polynomial regression [8]. Vazquez and Bect [9] proposed an iterative strategy for accurate computation of a probability of failure based on Kriging. In this paper, we present an alternative criterion to choose sequentially the experiments, based on an explicit trade-off between the exploration of the target region (on the vicinity of the contour line) and reduction of the global uncertainty (prediction variance) in the metamodel.

The paper is organized as follows: in Section 2, the Kriging model and the framework of design of experiments are described. In Section 3, the original criterion of selecting experiments is presented, followed by its associated sequen-

tial strategy to derive designs of experiments in Section 4. Results are presented for two analytical examples in Section 5. Finally, the criterion is applied to a probability of failure estimation problem.

2 Kriging Metamodel and Design of Experiments

Let us first introduce some notation. We denote by y the response of a numerical simulator or function that is to be studied:

$$y : D \subset \mathbf{R}^d \longrightarrow \mathbf{R}$$
$$\mathbf{x} \longmapsto y(\mathbf{x}) \tag{1}$$

where $\mathbf{x} = \{x_1, ..., x_d\}^T$ is a vector of input variables, and D is the design space. In order to build a metamodel, the response y is observed at n distinct locations $\mathbf{X}$:

$$\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n]$$
$$\mathbf{Y} = [y(\mathbf{x}_1), ..., y(\mathbf{x}_n)]^T = y(\mathbf{X}) \tag{2}$$

In Eqn. 2, choosing $\mathbf{X}$ is called the *design of experiments* (DoE), and $\mathbf{Y}$ is the vector of observations. Since the response y is expensive to evaluate, we approximate it by a simple model M, called the *metamodel* or *surrogate model*, based on assumptions on the nature of y and on its observations $\mathbf{Y}$ at the points of the DoE. In this paper, we present a particular metamodel, Universal Kriging (UK), and we discuss some important issues about the choice of the design of experiments.

2.1 Universal Kriging Model

The main hypothesis behind the Kriging model is to assume that the true function y is one realization of a Gaussian stochastic process Y, $y(\mathbf{x}) = Y(\mathbf{x}, \omega)$, where ω belongs to the underlying probability space Ω. In the following we use the notation $Y(\mathbf{x})$ for the process and $Y(\mathbf{x}, \omega)$ for one realization. For Universal Kriging [10], Y is typically of the form:

$$Y(\mathbf{x}) = \sum_{j=1}^{p} \beta_j f_j(\mathbf{x}) + Z(\mathbf{x}) \tag{3}$$

where f_j are linearly independent known functions, and Z is a Gaussian process [11] with zero mean and stationary covariance kernel k with known correlation structure and parameters.

Under such hypothesis, the best linear unbiased predictor (BLUP) for $Y(\mathbf{x})$ (for any $\mathbf{x}$ in D), knowing the observations $\mathbf{Y}$, is given by the following equation [10, 11]:

$$m_K(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \hat{\beta} + \mathbf{c}(\mathbf{x})^T \mathbf{C}^{-1} \left(\mathbf{Y} - \mathbf{F} \hat{\beta} \right) \tag{4}$$

where $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), ..., f_p(\mathbf{x})]^T$ is $p \times 1$ vector of basis functions, $\hat{\beta} = \left[\hat{\beta}_1, ..., \hat{\beta}_p \right]^T$ is $p \times 1$ vector of estimates of

β, $c(\mathbf{x}) = [k(\mathbf{x}, \mathbf{x}_1), ..., k(\mathbf{x}, \mathbf{x}_n)]^T$ is $n \times 1$ vector of covariance, $\mathbf{C} = [k(\mathbf{x}_i, \mathbf{x}_j)]_{1 \leq i, j \leq n}$ is $n \times n$ covariance matrix, and $\mathbf{F} = [\mathbf{f}(\mathbf{x}_1), ..., \mathbf{f}(\mathbf{x}_n)]^T$ is $n \times p$ experimental matrix. In Eqn. 4, $\hat{\beta}$ is the vector of generalized least square estimates of β:

$$\hat{\beta} = \left(\mathbf{F}^T \mathbf{C}^{-1} \mathbf{F} \right)^{-1} \mathbf{F}^T \mathbf{C}^{-1} \mathbf{Y} \tag{5}$$

In addition, the Universal Kriging model provides an estimate of the accuracy of the mean predictor, the Kriging prediction variance:

$$s_K^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{c}(\mathbf{x})^T \mathbf{C}^{-1} \mathbf{c}(\mathbf{x})$$
$$+ \left(\mathbf{f}(\mathbf{x})^T - \mathbf{c}(\mathbf{x})^T \mathbf{C}^{-1} \mathbf{F} \right) \left(\mathbf{F}^T \mathbf{C}^{-1} \mathbf{F} \right)^{-1} \left(\mathbf{f}(\mathbf{x})^T - \mathbf{c}(\mathbf{x})^T \mathbf{C}^{-1} \mathbf{F} \right)^T \tag{6}$$

where σ^2 is the process variance. Note that if the prediction variance is written in terms of correlations (instead of covariance here), Eqn. 6 can be factorized by σ^2. For details of derivations, see for instance [10, 11]. It is important to notice here that the Kriging variance in Eqn. 6, assuming that the covariance parameters are known, does not depend on the observations $\mathbf{Y}$, but only on the Kriging model and on the design of experiments.

We denote by $M(\mathbf{x})$ the Gaussian process conditional on the observations $\mathbf{Y}$:

$$M := (M(\mathbf{x}))_{\mathbf{x} \in D} = (Y(\mathbf{x})|Y(\mathbf{X}) = \mathbf{Y})_{\mathbf{x} \in D} = (Y(\mathbf{x})|obs)_{\mathbf{x} \in D} \tag{7}$$

The Kriging model provides the marginal distribution of M at a prediction point $\mathbf{x}$:

$$M(\mathbf{x}) \sim \mathcal{N} \left(m_K(\mathbf{x}), s_K^2(\mathbf{x}) \right) \tag{8}$$

The Kriging mean m_K interpolates the function $y(\mathbf{x})$ at the design of experiment points:

$$m_K(\mathbf{x}_i) = y(\mathbf{x}_i), \ 1 \leq i \leq n \tag{9}$$

The Kriging variance is null at the observation points $\mathbf{x}_i$, and greater than zero elsewhere:

$$s_K^2(\mathbf{x}_i) = 0, \quad 1 \leq i \leq n \quad and \quad s_K^2(\mathbf{x}) \geq 0, \quad \mathbf{x} \neq \mathbf{x}_i \tag{10}$$

Besides, the Kriging variance increases with the low values of the covariance between $Y(\mathbf{x})$ and $Y(\mathbf{x}_i)$ ($1 \leq i \leq n$). Some parameters of the covariance kernel are often unknown and must be estimated based on the observations, using maximum likelihood, cross-validation or variogram techniques for instance (see [10, 11]). However, in the Kriging model they are considered as known. To account for additional variability due to the parameter estimation, one may use Bayesian Kriging models (see [12, 13]), which will not be detailed here. With such models, Eqn. 8 does not stand in general. However, the methodology proposed here also applies to Bayesian Kriging, with the appropriate modifications of the calculations shown in Section 3.

2.2 Design of experiments

Choosing the set of experiments (sampling points) **X** plays a critical role in the accuracy of the metamodel and the subsequent use of the metamodel for prediction. DoEs are often based on geometric considerations, such as Latin Hypercube sampling (LHS) [14], or Full-factorial designs [15]. In this section, we introduce two important notions: model-oriented and adaptive designs.

2.2.1 Model-oriented designs

Model-oriented designs aim at maximizing the quality of statistical inference of a given metamodel. In linear regression, [16, 17], A- and D- optimal designs minimize the uncertainty in the coefficients, when uncertainty is due to noisy observations. Formally, the A- and D-optimality criteria are, respectively, the trace and determinant of Fisher's information matrix.

These criteria are particularly relevant in regression since minimizing the uncertainty in the coefficients also minimizes the uncertainty in the prediction (Kiefer, [16]). For Kriging, uncertainties in covariance parameters and prediction are not simply related. Instead, a natural alternative is to take advantage of the prediction variance associated with the metamodel, assuming that the covariance structure and parameters are accurately estimated. The prediction variance allows us to build measures that reflect the overall accuracy of Kriging. Two different criteria are available: the integrated mean square error (IMSE) and maximum mean square error (MMSE) [18]:

$$IMSE = \int_D MSE(\mathbf{x})d\mu(\mathbf{x}) \qquad (11)$$

$$MMSE = max_{\mathbf{x}\in D}[MSE(\mathbf{x})] \qquad (12)$$

μ is a positive measure on D and

$$MSE(\mathbf{x}) = E\left[(m_K(\mathbf{x}) - M(\mathbf{x}))^2\right] = s_K^2(\mathbf{x}) \qquad (13)$$

Note that the above criteria are often called I-criterion and G-criterion, respectively, in the regression framework. The IMSE is a measure of the average accuracy of the metamodel, while the MMSE measures the risk of large error in prediction.

Optimal designs are *model-dependent*, in the sense that the optimality criterion is determined by the choice of the metamodel. In regression, A- and D-criteria depend on the choice of the basis functions, while in Kriging, the prediction variance s_K^2 depends on the linear trend, the covariance structure, and parameter values. However, one may notice that, assuming that the trend and covariance structures are known, none of the criteria depends on the response values at the design points.

2.2.2 Adaptive designs

The previous DoE strategies choose all the points of the design before computing any observation. It is also possible to build the DoE sequentially, by choosing a new point as a function of the other points and their corresponding response values. Such approach has received considerable attention from the engineering and mathematical statistic communities, for its advantages of flexibility and adaptability over other methods [19, 20].

Typically, the new point achieves a maximum on some criterion. For instance, a sequential DoE can be built by making at each step a new observation where the prediction variance is maximal. Sacks et al. [18] use this strategy as a heuristic to build IMSE-optimal designs for Kriging. The advantage of sequential strategy here is twofold. Firstly, it is computationally efficient because it transforms an optimization problem of dimension $n \times d$ (for the IMSE minimization) into k optimizations of dimension d. Secondly, it allows us to reevaluate the covariance parameters after each observation. In the same fashion, Williams et al. [21], Currin et al. [22], and Santner [2] use a Bayesian approach to derive sequential IMSE designs. Osio and Amon [23] proposed a multistage approach to enhance first space-filling in order to accurately estimate the Kriging covariance parameters and then refine the DoE by reducing the model uncertainty. Some reviews of adaptive sampling in engineering design can be found in Jin et al. [24].

In general, a particular advantage of sequential strategies over other DoEs is that they can integrate the information given by the first k observation values to choose the $(k+1)^{th}$ training point, for instance by reevaluating the Kriging covariance parameters. It is also possible to define response-dependent criteria, with naturally leading to surrogate-based optimization. One of the most famous adaptive strategy is the EGO algorithm Jones et al. [7], used to derive sequential designs for the optimization of deterministic simulation models, by choosing at each step the point that maximizes the expected improvement, a functional that represents a compromise between exploration of unknown regions and local search. Jones [25] also proposes maximum probability of improvement as an alternative criterion.

In this paper, the objective is not optimization, but to accurately fit a function when it is close to a given threshold. It is then obvious that the DoE needs to be built according to the observation values, hence sequentially. Shan and Wang [26] proposed a rough set based approach to identify sub-regions of the design space that are expected to have performance values equal to a given level. Ranjan et al. [6] proposed a sequential DoE method for contour estimation, which consists of a modified version of the EGO algorithm. The functional minimized at each step is a trade-off between uncertainty and proximity to the actual contour. Tu et al. [8] used a weighted D-optimal strategy for polynomial regression, the acceptable sampling region at each step being limited by approximate bounds around the target contour. Oakley [27] uses Kriging and sequential strategies for uncertainty propagation and estimation of percentiles of the output of computer codes. Vazquez and Bect [9] proposed an iterative strategy for probability of failure estimation by minimizing the classification error when using Kriging. All these papers aim at constructing DoEs for accurate approximation of sub-
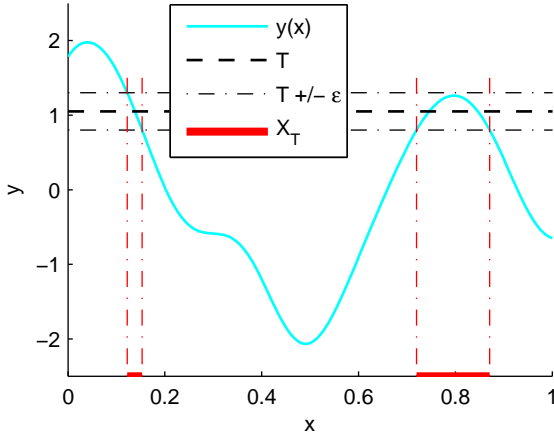
Fig. 1. One-dimensional illustration of the target region. Here, $T = 1$ and $\varepsilon = 0.2$. The target region consists of two distinct intervals.

regions of the design space. Our work proposes an alternative criterion which focuses on the integral of the prediction variance (rather than punctual criterion).

3 Weighted IMSE Criterion

In this section, we present a variation of the IMSE criterion, adapted to the problem of fitting a function accurately for a certain level-set. The controlling idea of this work is that the surrogate does not need to be globally accurate, but only in some critical regions, which are the vicinity of the target boundary.

3.1 Target region defined by an indicator function

The IMSE criterion is convenient because it sums up the uncertainty associated with the Kriging model over the entire domain D. However, we are interested in predicting Y accurately in the vicinity of a level-set $y^{-1}(T) = \{\mathbf{x} \in D : y(\mathbf{x}) = T\}$ (T a constant). Then, such a criterion is not suitable since it weights all points in D according to their Kriging variance, which does not depend on the observations $\mathbf{Y}$, and hence does not favor zones with respect to properties concerning their y values but only on the basis of their position with respect to the DoE.

We propose to change the integration domain from D to a neighborhood of $y^{-1}(T)$ in order to learn y accurately near the contour line. We define a *region of interest* $\mathbf{X}_{T,\varepsilon}$ (parameterized by $\varepsilon > 0$) as the subset in D whose image is within the bounds $T - \varepsilon$ and $T + \varepsilon$:

$$\mathbf{X}_{T,\varepsilon} = y^{-1}\left([T-\varepsilon, T+\varepsilon]\right) = \{\mathbf{x} \in D \,|\, y(\mathbf{x}) \in [T-\varepsilon, T+\varepsilon]\} \tag{14}$$

Figure 1 illustrates a one-dimensional function with the region of interest being at $T = 1$ and $\varepsilon = 0.2$. Note that the target region consists of two distinct intervals.

With the region of interest, the targeted IMSE criterion

is defined as follows:

$$imse_T = \int_{X_{T,\varepsilon}} s_K^2(\mathbf{x})d\mathbf{x} = \int_D s_K^2(\mathbf{x}) \mathbf{1}_{[T-\varepsilon, T+\varepsilon]}\left[y(\mathbf{x})\right] d\mathbf{x} \tag{15}$$

where $\mathbf{1}_{[T-\varepsilon, T+\varepsilon]}\left[y(\mathbf{x})\right]$ is the indicator function, equal to 1 when $y(\mathbf{x}) \in [T-\varepsilon, T+\varepsilon]$ and 0 elsewhere. Finding a design that minimizes $imse_T$ would make the metamodel accurate in the subset $\mathbf{X}_{T,\varepsilon}$, which is exactly what we want. Weighting the IMSE criterion over a region of interest is classical and proposed for instance by [15], pp.433-434. However, the notable difference here is that this region is unknown *a priori*.

Now, we can adapt the criterion in the context of Kriging modeling, where y is a realization of a random process Y (see Section 2.1).

Thus, $imse_T$ is defined with respect to the event ω:

$$imse_T = \int_D s_K^2(\mathbf{x}) \mathbf{1}_{[T-\varepsilon, T+\varepsilon]}\left[Y(\mathbf{x}, \omega)\right] d\mathbf{x} = I(\omega) \tag{16}$$

To come back to a deterministic criterion, we consider the expectation of $I(\omega)$, conditional on the observations:

$$
\begin{aligned}
IMSE_T &= E\left[I(\omega)\Big| obs\right] \\
&= E\left[\int_D s_K^2(\mathbf{x}) \mathbf{1}_{[T-\varepsilon, T+\varepsilon]}\left[Y(\mathbf{x}, \omega)\right] d\mathbf{x}\Big| obs\right] \tag{17}
\end{aligned}
$$

Since the quantity inside the integral is positive, we can commute the expectation and the integral:

$$
\begin{aligned}
IMSE_T &= \int_D s_K^2(\mathbf{x}) E\left[\mathbf{1}_{[T-\varepsilon, T+\varepsilon]}\left[Y(\mathbf{x}, \omega)\right]\Big| obs\right] d\mathbf{x} \\
&= \int_D s_K^2(\mathbf{x}) E\left[\mathbf{1}_{[T-\varepsilon, T+\varepsilon]}\left[M(\mathbf{x})\right]\right] d\mathbf{x} \\
&= \int_D s_K^2(\mathbf{x}) W(\mathbf{x}) d\mathbf{x} \tag{18}
\end{aligned}
$$

According to Eqn. 18, the reduced criterion is the average of the prediction variance weighted by the function $W(\mathbf{x})$. Besides, $W(\mathbf{x})$ is simply the probability that the response is inside the interval $[T-\varepsilon, T+\varepsilon]$:

$$
\begin{aligned}
W(\mathbf{x}) &= E\left[\mathbf{1}_{[T-\varepsilon, T+\varepsilon]}\left[M(\mathbf{x})\right]\right] \\
&= P\left(M(\mathbf{x}) \in [T-\varepsilon, T+\varepsilon]\right) \tag{19}
\end{aligned}
$$

Using Eqn. 8), we obtain a simple analytical form for $W(\mathbf{x})$:

$$W(\mathbf{x}) = \int_{T-\varepsilon}^{T+\varepsilon} g_{N\left(m_K(\mathbf{x}), s_K^2(\mathbf{x})\right)}(u)\,du \tag{20}$$

where $g_{N(m_K(\mathbf{x}),\sigma_K^2(\mathbf{x}))}(u)$ is the probability density function (PDF) of $M(\mathbf{x})$. By integrating the PDF we obtain:

$$W(\mathbf{x}) = \Phi\left(\frac{T+\varepsilon-m_K(\mathbf{x})}{s_K(\mathbf{x})}\right) - \Phi\left(\frac{T-\varepsilon-m_K(\mathbf{x})}{s_K(\mathbf{x})}\right) \quad (21)$$

where Φ is the CDF of the standard normal distribution. Note that by dividing $W(\mathbf{x})$ by the constant 2ε, it is possible to define the weight function with $\varepsilon \to 0$:

$$\lim_{\varepsilon \to 0} \frac{W(\mathbf{x})}{2\varepsilon} = g_{N(m_K(\mathbf{x}),s_K^2(\mathbf{x}))}(T) \quad (22)$$

which is the PDF of the Kriging distribution evaluated at threshold.

3.2 Target region defined by a Gaussian density

Defining the region of interest as $\mathbf{X}_{T,\varepsilon}$ is intuitive and makes it easy to derive the weight function. However, one might prefer a criterion that continuously increases the importance of the location when the response approaches the threshold. For instance, we can choose a triangular function (with a maximum at T) or a sigmoid function. Here, we choose to use the probability density function of a normal distribution which leads to a simple analytical form of the weight function. In the spirit of Eqn. 19, the Gaussian-based weight function is therefore defined as follows:

$$W(\mathbf{x}) = E\left[g_\varepsilon(M(\mathbf{x})-T)\right] \quad (23)$$

where $g_\varepsilon(u)$ is the PDF of $N(0,\sigma_\varepsilon^2)$.
When $M(\mathbf{x})$ stands for the Kriging model, we can obtain a simple form for the weight function:

$$W(\mathbf{x}) = \int_{-\infty}^{+\infty} g_\varepsilon(u-T) g_{N(m_K(\mathbf{x}),s_K^2(\mathbf{x}))}(u)\, du \quad (24)$$

This integral is the convolution of the two Gaussian densities, which is well-known to be the density of a sum of independent Gaussian variables. Hence, we obtain:

$$W(\mathbf{x}) = \frac{1}{\sqrt{2\pi\left(\sigma_\varepsilon^2 + s_K^2(\mathbf{x})\right)}} e^{\left(-\frac{1}{2}\frac{(m_K(\mathbf{x})-T)^2}{\sigma_\varepsilon^2 + s_K^2(\mathbf{x})}\right)} \quad (25)$$

This new weight function depends on a single parameter σ_ε that allows us to select the size the domain of interest around the target level of the function. A large value of σ_ε would enhance space-filling, since the weight function would tend to a constant and the weighted IMSE to a uniform IMSE criterion. On the contrary, a small value would enhance the accuracy of the surrogate on a narrow region around the contour line of interest. In particular when epsilon tends to zero,

the weight function tends to the density $g_{N(m_K(\mathbf{x}),s_K^2(\mathbf{x}))}(T)$, which is purely local.

In practice, it has been found that the choice of σ_ε -excepting very large and very small values- has little impact on the criterion and its use in sequential strategies. It only becomes important when the number of observations is very large (thus, the target region is well-known). In the numerical examples of Section 5, we chose σ_ε equal to approximately five percent of the output range.

3.3 Illustration

We consider a one-dimensional case, where the function y to approximate is a realization of a Gaussian process (so the Kriging is an accurate model for y) with isotropic Gaussian covariance structure:

$$k(\mathbf{u},\mathbf{v}) = \sigma^2 \exp\left[-\left(\frac{\|\mathbf{u}-\mathbf{v}\|}{\theta}\right)^2\right] \quad (26)$$

y is defined on $D = [0,1]$; the design of experiments consists of five observations equally spaced in this interval. The level-set of interest T is chosen as 1.3, and both ε and σ_ε are taken as 0.2. Figure 2 represents the true function, the Kriging metamodel and corresponding weights. The weight function in Eqn. 21 is shown as "interval", while that in Eqn. 25 is called "Gaussian".

Among the five observations, one is substantially closer to T than the others. As a consequence, the weight functions are large around this observation point. For the indicator-based weight function, the weights are null at the observation points, since on this example no observation is inside the target value interval. For the Gaussian-based weight, we can observe a smoothing effect compared to the interval. For both functions, high weights are given to regions for which the actual function is inside the target interval. Both weight functions are also non-zero where the uncertainty is high, even if the Kriging mean is far from T (around $x = 0.65$ and 0.85).

3.4 Application To Probability Of Failure Estimation
3.4.1 Probability of failure using metamodel

Failure of a system can usually be determined through a criterion, called a limit-state, g. The limit-state is defined such that the system is considered safe if $g \leq 0$, and failed otherwise. For instance, the limit-state of a structure can be defined as the difference between response, r, (e.g., maximum stress or strain) and capacity, c, (e.g., maximum allowable stress or strain), $g = r - c$.
The limit-state depends on a set of factors $\mathbf{U}$ (for instance in structural analysis, material properties and loadings), which are often uncertain, and the limit-state shows random distribution. Then the safety of the system is evaluated in terms of reliability or probability of failure. The probability of failure is defined as:

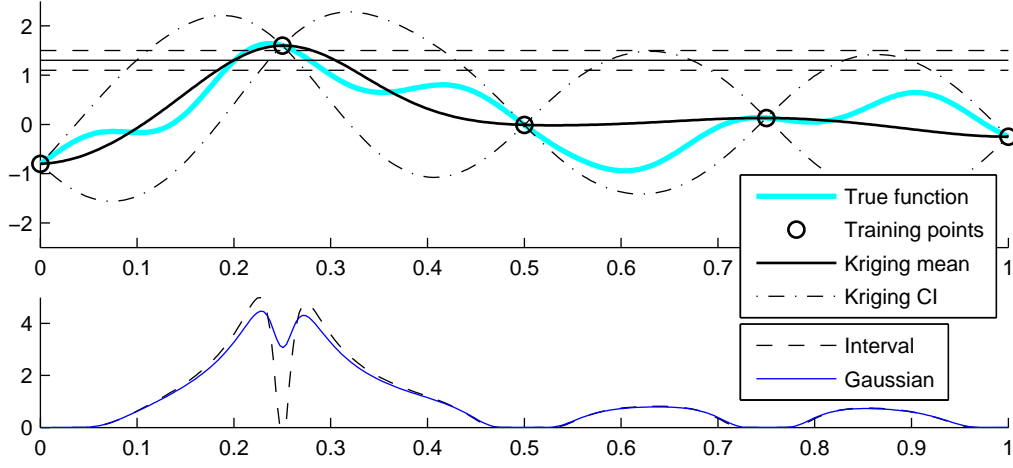$$P_f = Prob\left(g(\mathbf{U}) \geq 0\right) \quad (27)$$

Fig. 2. Illustration of the weights functions. Upper graph: true function, observations, Kriging mean and confidence intervals; the target region is represented by the horizontal lines at $T - \varepsilon$, T and $T + \varepsilon$. Lower graph: weight functions. Both weights are large where the true function is inside the target region, but also signaling regions of high uncertainties (around $x = 0.65$ and 0.85).

where $\mathbf{U}$ is a (multivariate) random variable.

There are many methods for calculating the failure probability of a system [3, 28, 29]. Some of them use the relation between input random variables and the limit-state (e.g., first-order reliability method) and some consider the limit-state as a black-box (e.g., Monte-Carlo Simulations, MCS). MCS generates samples of the limit-state and calculates the number of failed runs [3]. The ratio between the numbers of failures and the total sample size approximates the probability of failure of the system:

$$\hat{P}_f = \frac{1}{N} \sum_{i=1}^{N} 1_{[0,+\infty[} [g(\mathbf{u}_i)] \qquad (28)$$

where the $\mathbf{u}_i$'s are i.i.d. (independent and identically distributed) replicates of the random variable $\mathbf{U}$ ($1 \leq i \leq N$). The accuracy of MCS strongly depends on the number of runs used, especially when the probability of failure is low. When the cost of simulation is high, engineers can afford to have only a small number of runs, which is not good enough to estimate the reliability with acceptable accuracy [30]. Hence, using a metamodel to approximate the limit-state g is a natural solution to the lack of data; MCS is then performed on the metamodel that is inexpensive to evaluate. Instead of using the indicator function on the Kriging mean, we use the full Kriging information by computing, at each sampling point, the probability that the response exceeds the threshold:

$$\hat{P}_f = \frac{1}{N} \sum_{i=1}^{N} 1 - \Phi_k^{(i)}(0) \qquad (29)$$

where $\Phi_k^{(i)}$ denotes the cumulative distribution function (CDF) of the Kriging model at $\mathbf{x}_i$ ($N(m_k(\mathbf{u}_i), s_k^2(\mathbf{u}_i))$). If the Kriging variance is small, the CDF becomes equivalent

to the indicator function, being 1 if the Kriging mean exceeds the threshold zero and 0 otherwise. On the other hand, when the variance is high or the predicted response close to the threshold, using the Kriging distribution offers a smoothing effect by giving a number between zero and one instead of a Boolean number.

3.4.2 Adaptation of the weighted IMSE criterion

When approximating the limit-state, it is clear that accuracy is critical in the regions where it is close to zero, since error in that region is likely to affect the probability estimate. The region of interest can be further refined by taking into account the distribution of the input variables. Indeed, let us consider the case of two distinct failure regions, with the probability that the input falls onto the first region being much larger than the probability that it falls onto the other). Instead of focusing equally on the two critical regions, it will be more efficient to spend more computational effort on the one that will affect most the probability estimate. In the same sense, when refining the surrogate in a single critical region, it is efficient to refine only where the input probability is high. To address this probability distribution of input variables, we modify the weighted IMSE criterion by integrating the weighted MSE not with a uniform measure, but with the law μ of the input variables. In the usual case that μ admits a PDF $f(x)$ with respect to the Lebesgue measure, we then have:

$$\mathrm{IMSE}_T = \int_D s_K^2(\mathbf{x}) W(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) = \int_D s_K^2(\mathbf{x}) W(\mathbf{x}) f(\mathbf{x}) \, \mathrm{d}\mathbf{x} \qquad (30)$$

In practice, the criterion becomes the integral of the product of three quantities: the prediction variance, the weight function and probability density function of the input variables.

Table 1. Procedure of the $IMSE_T$-based sequential DoE strategy.

Create an initial DoE, $\mathbf{X}_k$, and generate observations

$\mathbf{Y}_k = y(\mathbf{X}_k)$

For i going from one to the total number of additional

observations n:

    Fit the Kriging model to the data $\{\mathbf{X}_{k+i-1}, \mathbf{Y}_{k+i-1}\}$

    Find a new training point $\mathbf{x}_{new}$ that minimizes the

criterion $\mathrm{IMSE_T}(\{\mathbf{X}_{k+i-1}, \mathbf{x}_{new}\})$

    Compute the new observation $y_{new} = y(\mathbf{x}_{new})$

    Update the DoE and observations:

        $\mathbf{X}_{k+i} = \{\mathbf{X}_{k+i-1}, \mathbf{x}_{new}\}$

        $\mathbf{Y}_{k+i} = \{\mathbf{Y}_{k+i-1}, y_{new}\}$

End of loop

4 Sequential Strategies For Selecting Experiments

4.1 Building DoEs using the targeted IMSE criterion

Without any observation, the weight function $W(\mathbf{x})$ is, assuming stationarity, a constant (the probability is the same everywhere). Every time a new observation is performed, the weight function will more precisely discriminate the regions of interest from the others. Hence, the procedure to build an optimal DoE is necessarily iterative. If we add one observation at a time we can use the procedure shown in Table 1.

A good evaluation of the covariance parameters is critical to obtain a good Kriging model. Besides, those parameters directly affect the weight function: for instance, underestimation of the range (θ in Eqn. 26) makes the weight function flat (constant), which enhances space-filling; on the contrary, overestimation of the range leads to a very discriminating (over-confident) weight function. The Kriging parameters can be reevaluated after every new observation, or only from the initial DoE before the iterative procedure. However, re-evaluating the parameters at each iteration is computationally intensive, which can harm the efficiency of the method. Hence, one would consider estimating the parameters only when necessary, as proposed in Gano et al. [31]. In the numerical examples used in this work, we found that after a first few iterations, the parameter re-evaluation had a negligible impact on the efficiency of the method.

Defining a stopping criterion for this problem is an open and complex question. We consider here that in most cases the number of observations is very limited so the iterative process stops at early stage. Ideally, the adaptive process should be stopped when the identified target region does not change significantly over several adaptations, which can be detected by looking at changes in the probability to be inside target regions.

Finding the new training point requires an inner optimization procedure. When the classical IMSE criterion is considered,

the optimization can be expressed as:

$$\min_{\mathbf{x}_{new} \in D} \mathrm{IMSE}(\mathbf{X}_{k+1}) = \mathrm{IMSE}(\{\mathbf{X}_k, \mathbf{x}_{new}\}) \quad (31)$$

where $\mathrm{IMSE}(\{\mathbf{X}_k, \mathbf{x}_{new}\}) = \int_D s_K^2(\mathbf{x}|\{\mathbf{X}_k, \mathbf{x}_{new}\})\,d\mathbf{x}$.

$s_K^2(\mathbf{x}|\{\mathbf{X}_k, \mathbf{x}_{new}\})$ is the variance at $\mathbf{x}$ of the Kriging model based on the design of experiments $\mathbf{X}$ augmented with the training point $\mathbf{x}_{new}$. Since the Kriging variance does not depend on the observation, there is no need to have $y(\mathbf{x}_{new})$ to compute the IMSE.

In contrast, the weighted IMSE depends on the observations through the weight function $W(\mathbf{x})$. The weight function cannot take into account the new observation, since the response is not available. Hence, when expressing the weighted IMSE as a function of $\mathbf{x}_{new}$, we update only the variance part under the integral:

$$\mathrm{IMSE_T}(\mathbf{X}_k, \mathbf{Y}_k, \mathbf{x}_{new}) = \int_D s_K^2(\mathbf{x}|\{\mathbf{X}_k, \mathbf{x}_{new}\})\,W(\mathbf{x}|\mathbf{X}_k, \mathbf{Y}_k)\,d\mathbf{x}$$

$$(32)$$

where $s_K^2(\mathbf{x}|\{\mathbf{X}_k, \mathbf{x}_{new}\})$ is the same as in Eqn. 31 and $W(\mathbf{x}|\mathbf{X}_k, \mathbf{Y}_k)$ is the weight function based on the existing DoE. Using this expression, we have the simple formulation for the inner optimization problem:

$$\min_{\mathbf{x}_{new} \in D} \mathrm{IMSE_T}(\mathbf{X}_k, \mathbf{Y}_k, \mathbf{x}_{new}) \quad (33)$$

4.2 Solving the optimization problem

Finding the new observation $\mathbf{x}_{new}$ by solving the optimization problem of Eqn. 33 is, in practice, challenging. Indeed, the $IMSE_T$ criterion in Eqn. 32 must be evaluated by numerical integration, which is computationally intensive. Besides, for any candidate $\mathbf{x}_{new}$, the Kriging model must be reevaluated with this new observation to obtain $s_K^2(\mathbf{x}|\{\mathbf{X}_k, \mathbf{x}_{new}\})$). Therefore we propose here some alternatives that may be used to reduce the cost.

A popular heuristic to minimize sequentially the IMSE is to find the point where the prediction variance is maximum [[18, 21]], which can be used here with the weighted prediction variance. This strategy has the advantage of saving both the numerical integration and the inversion of a new covariance matrix. However, the prediction variance is likely to have many (local or global) maximizers, which are not equivalent in terms of the IMSE. In particular, many optima are located on the boundaries, which is very inefficient for the IMSE minimization. To compensate for this issue, one may in a first time get a large number of local optima using adapted optimization strategies (multi-start, etc.), and in a second time evaluate those optima in terms of the weighted IMSE criterion, and perform a local optimization on the best point. It is to be noted that the gradients of the weighted MSE can be calculated analytically (in the fashion of Ginsbourger [32] (Chapter 4) for the Expected Improvement criterion).
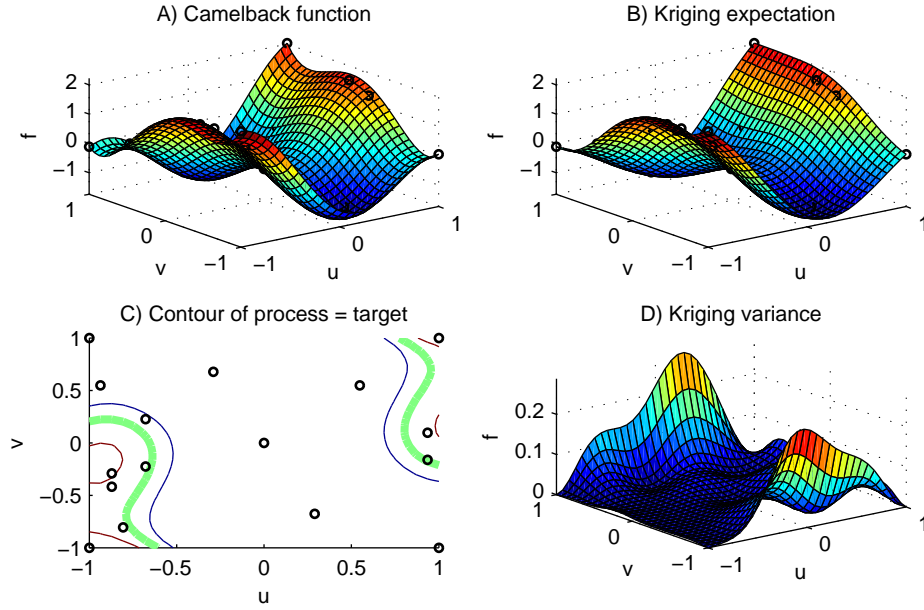
Fig. 3. Optimal design after 11 iterations. The contour lines correspond to the true function at levels T (bold line) and $[T - \sigma_\varepsilon, T + \sigma_\varepsilon]$, which delimit the actual target regions. Most of the training points are chosen close to the target region. The Kriging variance is very small in these regions and large in non-critical regions.

A valuable computational shortcut can be achieved in the update of the inverse of the covariance matrix when adding an observation. Let us call $\mathbf{C}_k$ the covariance matrix corresponding to a DoE with k observations. Then, the covariance matrix of the DoE augmented with the $k+1$th observation can be written:

$$\mathbf{C}_{k+1} = \begin{bmatrix} \sigma^2 & \mathbf{c}_{new}^T \\ \mathbf{c}_{new} & \mathbf{C}_k \end{bmatrix} \qquad (34)$$

with $\mathbf{c}_{new}^T = [k(\mathbf{x}_{new}, \mathbf{x}_1), \dots, k(\mathbf{x}_{new}, \mathbf{x}_k)]$ a $1 \times k$ vector. Using Schur's complement formula [33], we get:

$$\mathbf{C}_{k+1}^{-1} = \begin{bmatrix} 1 & 0 \\ -\mathbf{C}_k^{-1}\mathbf{c}_{new} & \mathbf{I}_k \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2 - \mathbf{c}_{new}^T \mathbf{C}_k^{-1} \mathbf{c}_{new}} & 0 \\ 0 & \mathbf{C}_k^{-1} \end{bmatrix} \begin{bmatrix} 1 & -\mathbf{c}_{new}^T \mathbf{C}_k^{-1} \\ 0 & \mathbf{I}_k \end{bmatrix}$$
$$(35)$$

This formula allows to compute $\mathbf{C}_{k+1}^{-1}$ from $\mathbf{C}_k^{-1}$ without doing any matrix inversion, and compute $s_K^2(\mathbf{x}|\{\mathbf{X}_k, \mathbf{x}_{new}\})$ at reasonable cost.

Another typical problem of sequential strategies for Kriging is the ill-conditioning of the covariance matrix, which happens in particular when two (or more) observations are very close to each other. Since the IMSE criterion enhances exploration, this risk is limited here. Therefore, when the number of iterations is large, the observations can concentrate on the target region and the covariance matrix becomes difficult to invert. In that case, it is possible to add a small diagonal matrix (nugget effect) to the covariance function in order to facilitate the inversion (Neal [34]).

In general, the criterion has several local minimizers. Then, it is necessary to use global optimization methods, such as population-based methods, multi-start strategies, etc. In the test problems presented in this chapter, we optimize the criterion on a fine grid for low dimensions, and using the population-based CMA-ES algorithm [Covariance Matrix Adaptation Evolution Strategies, [35]] for higher dimensions. Experimentation showed that due to the numerical integration precision, the targeted IMSE strategy becomes inefficient for dimensions higher than ten.

5 Numerical Examples

In this section, we evaluate the accuracy and efficiency of the methods presented in the sections 3 and 4 through numerical examples. We consider three examples: the first is the fitting of an analytical function in two dimensions with estimated covariance parameters. The second is the fitting of realizations of random processes in six dimensions with known covariance parameters, which allows us to decompose the problem and evaluate the relevance of our criterion since in this case there is no modeling error. Finally, the method is applied to probability of failure estimation.

5.1 Two-dimensional example

The first example is the approximation of a two-dimensional parametric function from the optimization literature (*Camelback function*, [36]). The original function is modified (bounds are different and a negative constant is added) and the target is chosen in order to have two failure regions, one dominating the other. The two-dimensional design space is given as $[1,1]^2$. The performance function is

defined as

$$f(u,v) = \left(4 - 2.1\bar{u}^2 + \frac{1}{3}\bar{u}^4\right)\bar{u}^2 + \frac{2}{3}\bar{u}\bar{v} + \frac{16}{9}\left(-4 + \frac{16}{9}\bar{v}^2\right)\bar{v}^2 - \Big($$

(36)

where $\bar{u} = 1.2u - 0.1$ and $\bar{v} = 0.9v$.

For both numerical integration and optimization, the design space is discretized in a 32×32 grid. We present the results for the following configuration:

> ==Target value T is chosen as 1.3,==
> Gaussian-based weight function is used, with parameter $\sigma_\varepsilon = 0.2$,
> Initial DoE consists of the four corners and the center of the domain,
> 11 points are added iteratively to the DoE as described in the previous section.

An isotropic Gaussian covariance function (Eqn. 26) is chosen for the Kriging model. The covariance parameters (process variance σ^2 and range θ) are estimated from the initial 5-point DoE, and re-estimated after each new observation, using the MatLab toolbox GPML [11]. The final results are presented in Figure 3.

Figure 3 A) is the plot of the true function, and Figure 3 B) is that of the Kriging mean. In the contour plot in Figure 3 C), it is shown that there are two critical regions. After 11 iterations, the sequential strategy used four points to explore the first critical region, three points to explore the second region, and four points for space-filling. As shown in Figure 3 D), the Kriging variance becomes small near the critical regions, while it is relatively large in the non-critical region.

Figure 4 shows the evolution of the target contour line for the kriging expectation, which is a good indicator of the quality of the surrogate. We see that because the first four iterations (Figure 4 B)) are used for space-filling, the Kriging contour line is very different from the actual one. After eight iterations (Figure 4 C)), the two target regions are found and additional sampling points are chosen close to the actual contour line. Final state (Figure 4 D)) shows that the kriging contour line is close to the actual one.

5.2 Six-Dimensional Example

In the second example, we consider a realization of a six-dimensional isotropic Gaussian process with Gaussian covariance function. The design space is $[-11]^6$. In order to limit the complexity (number of non-connected target regions) of the target region, we add a linear trend to the Gaussian process. We take $\sigma^2 = 1$, $\theta = 0.1$ and $\beta = [1\ldots1]$. The weighted IMSE criterion is computed by Quasi Monte-Carlo integration. The integration points are chosen from a Sobol sequence [37] to ensure a good space-filling, and are changed at each step to limit the risk of keeping a hole in the integration region over the iterations. At each step, the optimization is performed using the population-based optimizer CMA-ES [35]. The number of integration points is chosen equal to 5,000, and the number of function evaluations for CMA-ES is limited to 1,000. With this set-up, one
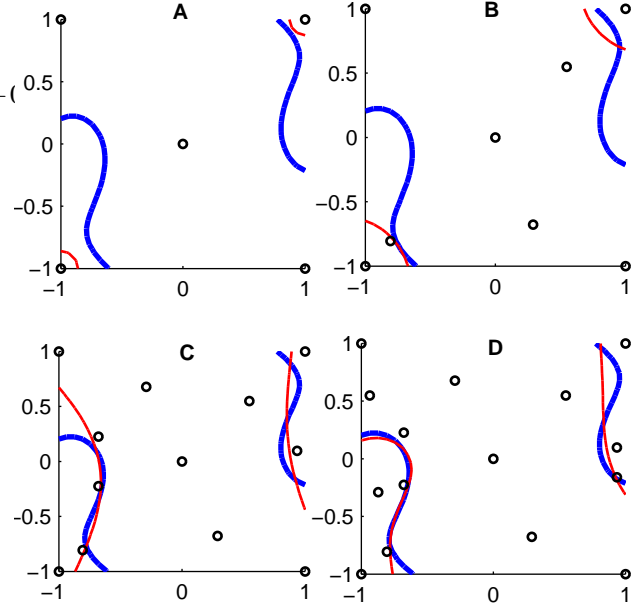

Fig. 4. Evolution of Kriging target contour line (thin line) compared to actual (bold line) during the sequential process: A) Initial, B) after four iterations, C) after eight iterations, D) final.

optimization (which is the computational bottleneck) takes of the order of two minutes on a PC with a 1.8GHz processor and 1Go RAM. For comparison, the two-dimensional problem described earlier requires about 1 second to perform the optimization.

We present the results for the following configurations:

> Target value is chosen as 2
> Gaussian-based weight function is used, with $\sigma_\varepsilon = 0.05$.
> Initial DoE consists of 20 points chosen from Latin-hypercube sampling (LHS)
> 70 points are added iteratively to the DoE.

The Kriging parameters are not estimated here, but taken equal to the covariance parameters of the true function. Hence, no modeling error is involved since the function to approximate corresponds exactly to the assumptions of the Kriging model, and the error of Kriging is only due to the lack of sampling. The advantage of using such test case is to decompose the problem: here, we evaluate only the relevance and efficiency of our criterion, regardless the difficulty of estimating the covariance parameters from a small number of observations.

For comparison purpose, we generate a classical space-filling DoE that consists of 90 LHS points with maximin criterion. First, we represent the error at 10,000 (uniformly distributed) data points (Figure 5). The classical space-filling DoE leads to a uniform error behavior, while the optimal DoE lead to large errors when the response is far from the target value, while small errors when it is close to the target.

In order to analyze the error in the target region, we draw the boxplots of the errors for the test points where responses are inside the domain $[T - 2\sigma_\varepsilon, T + 2\sigma_\varepsilon]$ (Figure 6). Compared to the space-filling strategy, the optimal design reduces
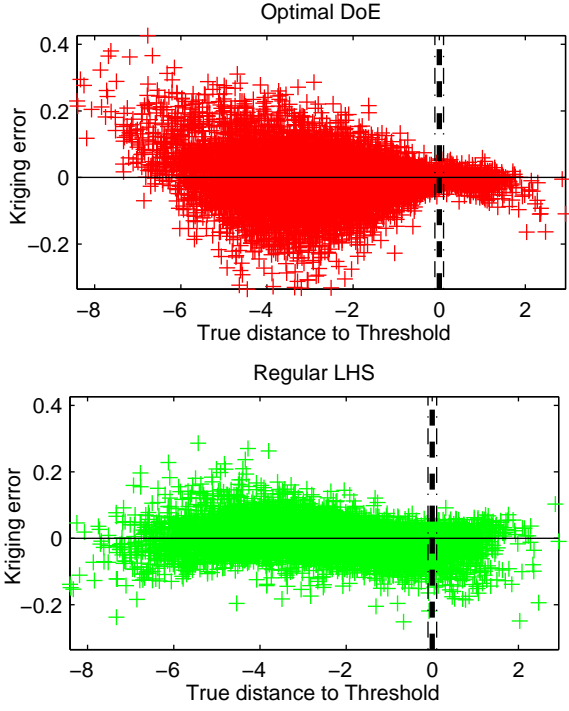
Fig. 5. Comparison of error distribution for two 90-point DoEs: optimal DoE (top) and classical LHS (bottom). The x-axis is the difference between the true function and the threshold, the y-axis is the error. Three vertical bars are drawn at $-2\sigma_\varepsilon$, 0 and $+2\sigma_\varepsilon$ for the target region. The error is on average smaller for the LHS design, but the optimal DoE reduces substantially the error in the target region.
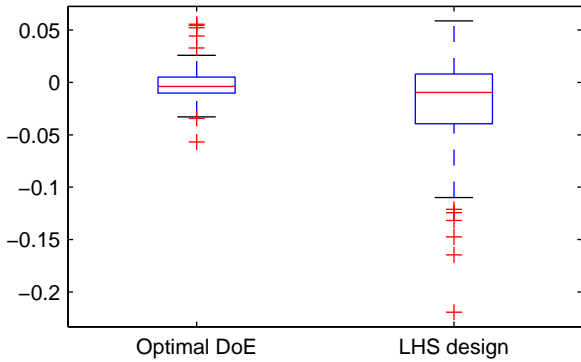

Fig. 6. Boxplots of errors for the 90-point LHS and optimal designs for the test points where responses are inside the domain $[T - 2\sigma_\varepsilon, T + 2\sigma_\varepsilon]$. Error at these points is about 2.5 times smaller for the optimal designs for both intervals.

significantly the error. In particular, the interquartiles interval is 2.5 times smaller for the optimal DoE.

5.3 Reliability Example

The limit state function is taken as the Camelback function used in the previous section. Let U and V be independent Gaussian variables with zero mean and standard deviation taken at 0.28; i.e., $U, V \sim N\left(0, 0.28^2\right)$. Then, the failure
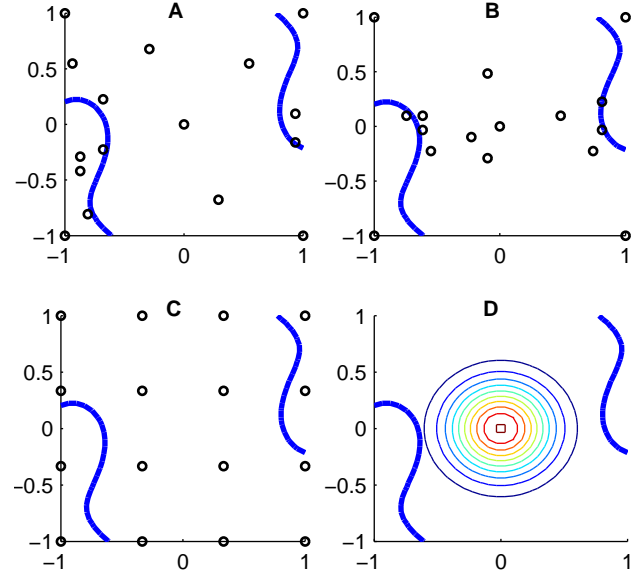

Fig. 7. Optimal design with A) uniform integration measure, B) input distribution integration measure; C) Full factorial designs with 16 points. Plain green line shows the limit of the failure region; Input distribution is shown in D).

is defined when f becomes greater than 1.3. Thus, the limit state is defined as

$$G = f(U,V) - 1.3 \qquad (37)$$

For this example, we generate two adaptive designs: the first is generated sequentially as described previously, with uniform integration measure (Eqn. 18); the second is generated using the input distribution as integration measure (Eqn. 30). Both use the four corners and the center of the domain as starting DoE and 11 iterations are performed. For comparison purpose, a 16-point full factorial design is also used. It is found that an Ordinary Kriging model (UK without linear trend) with isotropic Gaussian covariance function approximates well the function. The covariance parameters are computed using the toolbox GPML for all the DoEs. For the sequential DoEs the parameters are re-evaluated at each new observation.

Figure 7 draws the two optimal designs obtained and the full factorial designs. Both optimal designs concentrate the computational effort on the failure regions and the center of the domain. With uniform measure integration in Figure 7 A), the DoE is more space-filling than the one based on the distribution (shown in Figure 7 D)). By taking the input distribution into account in Figure 7 B), we see that all the observations are located relatively close to the center of the domain. Part of each target regions is not explored, since it is far from the center.

Finally, we perform 10^7 MCS on the three metamodels to compute the probability of failure estimates. 10^7 MCS are also performed directly on the test function to obtain the true probability of failure. Results are reported in Table 2. The full-factorial design leads to 77% error, while both optimal

Table 2. Probability of failure estimates for the three DoEs and the actual function based on 10^7 MCS. The standard deviation of all estimates is of the order of 2×10^{-5}.

DoE	P_f (%)	Relative error
Full Factorial	0.17	77 %
Optimal without input distribution	0.70	7 %
Optimal with input distribution	0.77	3 %
Probability estimate based on 10^7 MCS	0.75	

designs lead to a small error. Substantial improvement is obtained by taking the input distribution into account.

6 Conclusions

In this paper, we have addressed the issue of choosing a design of experiments when the Kriging metamodel was used to approximate a function accurately around a particular level-set. This situation frequently occurs in constrained optimization and reliability analysis. We proposed a modified version of the classical IMSE criterion, obtained by weighting the prediction variance using a Kriging-based measure of the expected proximity to target values. The choice of a new observation based on such criterion is a trade-off between exploration of the target region (on the vicinity of the contour line) and reduction of the global uncertainty (prediction variance) in the metamodel.

We applied our strategy to examples in two and six dimensions. In two dimensions, we showed that the adaptive sampling efficiently explored the target regions while ensuring space-filling. In six dimensions, we showed that compared to a classical space-filling design, the error reduction in the target region was by a factor of 2.5.

Finally, the method was tested for reliability estimation on an analytical example. An additional criterion was adapted to integrate the distribution of input random variables. It was found that both criterion-based strategies significantly outperformed space-filling designs, and taking into account the input distribution provides additional improvement in the accuracy of the probability of failure.

However, it has been found some limitations to the method, which were not solved here and requires future work to apply the method to a wide range of problems:

Since it relies on numerical integration, the method can become computationally expensive if a large number of integration points are needed to compute the criterion. We found that for dimensions higher than ten, the criterion minimization becomes critical without the use of complex and problem-dependant numerical procedures, such as dimension reduction or adapted numerical integration.

Secondly, it is important to recall that it is a model-believer strategy, since the criterion is entirely based on the Kriging model. Although sequential strategies allow some correction of the model during the process (through re-estimation of the parameters for instance), the success of the method will strongly depend on the capability of the Kriging model to fit the actual response.

Future research may compare the results obtained with this method to alternative methods, in particular in the frameworks of reliability analysis and constrained optimization.

References

[1] Queipo, N., Haftka, R., Shyy, W., Goel, T., Vaidyanathan, R., and Kevin Tucker, P., 2005. "Surrogate-based analysis and optimization". *Progress in Aerospace Sciences,* **41**(1), pp. 1–28.

[2] Santner, T., Williams, B., and Notz, W., 2003. *The design and analysis of computer experiments*. Springer.

[3] Melchers, R., 1987. *Structural reliability: analysis and prediction*. Ellis Horwood.

[4] Rajashekhar, R., and Ellingwood, B., 1993. "A new look at the response surface approach for reliability analysis". *Structural Safety*, pp. 205–220.

[5] Mourelatos, Z., Kuczera, R., and Latcha, M., 2006. "An Efficient Monte Carlo Reliability Analysis using Global and Local Metamodels". In Proceedings of 11th AIAA/ISSMO multidisciplinary analysis and optimization conference, September, Portsmouth, VA.

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

[7] Jones, D., Schonlau, M., and Welch, W., 1998. "Efficient global optimization of expensive black-box functions". *Journal of Global Optimization,* **13**(4), pp. 455–492.

[8] Tu, C., and Barton, R., 1997. "Production yield estimation by the metamodel method with a boundary-focused experiment design". In ASME Design Engineering Technical Conference, American Society Of Mechanical Engineers.

[9] Vazquez, E., and Bect, J., 2009. "A sequential bayesian algorithm to estimate a probability of failure". In Proceedings of 15th IFAC Symposium on System Identification (SYSID), Saint-Malo, France.

[10] Matheron, G., 1969. "Le krigeage universel". *Cahiers du centre de morphologie mathematique,* **1**.

[11] Rasmussen, C., and Williams, C., 2006. *Gaussian processes for machine learning*. Springer.

[12] Martin, J., and Simpson, T., 2004. "A Monte Carlo simulation of the Kriging model". In 10th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, pp. 2004–4483.

[13] Oakley, J., and O'Hagan, A., 2004. "Probabilistic sensitivity analysis of complex models: a Bayesian approach". *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, pp. 751–769.

[14] McKay, M., Beckman, R., and Conover, W., 2000. "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code". *Technometrics*, pp. 55–61.

[15] Box, G., and Draper, N., 1986. *Empirical model-building and response surface*. Wiley Series in Probability and Mathematical Statistics.

[16] Kiefer, J., 1961. "Optimum designs in regression problems, II". *The Annals of Mathematical Statistics*, pp. 298–325.

[17] St John, R., and Draper, N., 1975. "D-optimality for regression designs: a review". *Technometrics*, pp. 15–23.

[18] Sacks, J., Schiller, S., and Welch, W., 1989. "Designs for computer experiments". *Technometrics*, pp. 41–47.

[19] Lin, Y., Luo, D., Bailey, T., Khire, R., Wang, J., and Simpson, T. W., 2008. "Model Validation and Error Modeling to Support Sequential Sampling". In Proceedings of the ASME Design Engineering Technical Conferences - Design Automation Conference, New York, NY, DETC2008/DAC-49336.

[20] Turner, C., Campbell, M., and Crawford, R., 2003. "Generic sequential sampling for meta-model approximations". In ASME 2003 Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 2–6.

[21] Williams, B., Santner, T., and Notz, W., 2000. "Sequential design of computer experiments to minimize integrated response functions". *Statistica Sinica,* **10**(4), pp. 1133–1152.

[22] Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D., 1991. "Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments". *Journal of the American Statistical Association*, pp. 953–963.

[23] Osio, I., and Amon, C., 1996. "An engineering design methodology with multistage Bayesian surrogates and optimal sampling". *Research in Engineering Design,* **8**(4), pp. 189–206.

[24] Jin, R., Chen, W., and Sudjianto, A., 2002. "On sequential sampling for global metamodeling in engineering design". In Proceedings of DETC, Vol. 2, Citeseer.

[25] Jones, D., 2001. "A taxonomy of global optimization methods based on response surfaces". *Journal of Global Optimization,* **21**(4), pp. 345–383.

[26] Shan, S., and Wang, G., 2004. "Space exploration and global optimization for computationally intensive design problems: a rough set based approach". *Structural and Multidisciplinary Optimization,* **28**(6), pp. 427–441.

[27] Oakley, J., 2004. "Estimating percentiles of uncertain computer code outputs". *Applied statistics*, pp. 83–93.

[28] Enevoldsen, I., and Sørensen, J., 1994. "Reliability-based optimization in structural engineering". *Structural Safety,* **15**(3), pp. 169–196.

[29] Haldar, A., and Mahadevan, S., 2000. *Probability, reliability and statistical methods in engineering design*. Wiley New York.

[30] Ben-Haim, Y., and Elishakoff, I., 1990. *Convex models of uncertainty in applied mechanics*. North-Holland.

[31] Gano, S., Renaud, J., Martin, J., and Simpson, T., 2006. "Update strategies for kriging models used in variable fidelity optimization". *Structural and Multidisciplinary Optimization,* **32**(4), pp. 287–298.

[32] Ginsbourger, D., 2009. "Metamodeles multiples pour l'approximation et l'optimisation de fonctions numeriques multivariables". PhD thesis, Ecole Nationale Superieure des Mines de Saint Etienne.

[33] Zhang, F., 2005. *The Schur complement and its applications*. Springer Verlag.

[34] Neal, R., 1997. "Monte Carlo implementation of Gaussian process models for Bayesian regression and classification". *Arxiv preprint physics/9701026*.

[35] Hansen, N., and Kern, S., 2004. "Evaluating the CMA evolution strategy on multimodal test functions". *Lecture Notes in Computer Science*, pp. 282–291.

[36] Chipperfield, A., Fleming, P., Pohlheim, H., and Fonseca, C., 1994. "Genetic algorithm toolbox users guide". *ACSE Research Report,* **512**.

[37] Sobol, I., 1976. "Uniformly distributed sequences with an additional uniform property". *Zh. Vychisl. Mat. Mat. Fiz,* **16**, pp. 1332–1337.