XL CILAMCE
IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING
NOVEMBER 11-14, 2019
Praiamar Natal Hotel & Convention
Natal, RN-BRAZIL

# SEQUENTIAL APPROXIMATE OPTIMIZATION USING KRIGING AND RADIAL BASIS FUNCTIONS

**Marina A. Maia**
**Leonardo G. Ribeiro**
*marinaalvesmaia@gmail.com*
*leo14330@gmail.com*
**Evandro Parente Jr.**
*evandro@ufc.br*
**Antônio M. C. Melo**
*macario@ufc.br*
*Laboratório de Mecânica Computacional e Visualização (LMCV), Departamento de Engenharia Estrutural e Construção Civil, Universidade Federal do Ceará, Campus do Pici, Centro de Tecnologia, Bloco 728, 60455-760, Fortaleza, Ceará, Brasil*

**Abstract.** Despite steady advance in computing power, the number of function evaluations in global optimization problems is often limited due to time-consuming analyses. In structural optimization problems, for instance, these analyses are typically carried out using the Finite Elements Method (FEM). This issue is especially critical when dealing with bio-inspired algorithms, where a high number of trial designs are usually required. Therefore, surrogate models are a valuable alternative to help reduce computational cost. With that in mind, present work proposes three Sequential Approximate Optimization (SAO) techniques. For that purpose, two surrogate models were chosen: the Radial Basis Functions (RBF) and Kriging. As for the infill criteria, three methodologies were investigated: the Expected Improvement, the Density Function and the addition of the global best. Two bio-inspired meta-heuristics were used in different stages of the optimization, namely Particle Swarm Optimization and Genetic Algorithm. To validate the proposed methodologies, a set of benchmarks functions were selected from the literature. Results showed a significant reduction in the number of high-fidelity evaluations. In terms of accuracy, efficiency, and robustness, Kriging excelled in most categories for all problems. Finally, these techniques were applied to the solution of a laminated composite plate, which demands a more complex analysis using FEM.

**Keywords:** Optimization, Sequential Approximate Optimization, RBF, Kriging.

## 1 Introduction

Simulation-based analysis has increasingly been used to explore design alternatives at preliminary stages of engineering design problems. Often these analyses require computationally intensive numerical simulations such as Finite Element Analysis (FEA) and Isogeometric Analysis (IGA). However, despite steady advances in computing power, these high fidelity simulations may pose a serious limitation to engineering design optimization, thereby undermining its potential. This is especially critical when population-based meta-heuristics algorithms are employed (e.g., Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO)), as it is the case explored in this work, due to the high number of evaluations involved. On the other hand, these algorithms do not demand information on the function gradient, which can be especially useful when discrete variables are considered (a common consideration in engineering problems as a result of manufacturing constraints). In addition to that, meta-heuristics algorithms are less prone to be trapped in local minima [1].

The basic approach to overcome these issues is to use surrogate models (also known as *meta-models*) in lieu of high fidelity models. In brief, surrogate models may be understood as approximate models of the true response surface of objective functions and/or constraints built upon a limited number of the true responses. These observations are usually determined by a Design of Experiments (DoE), which consists of a group of stochastic and deterministic methodologies used to define the sample points distribution on the design space. Thus, one can use this cheaper model to predict points on the design space that were not yet evaluated at a lower computational cost when compared to that of running the true function.

Several methodologies were proposed for building the surrogate model, such as the Radial Basis Functions (RBF), Kriging, the Artificial Neural Networks (ANN) and the Support Vector Regression (SVR). In this paper, the RBF and the Kriging Model are the meta-models used to approximate the HF responses. These are two of the most intensively studied meta-models and they were both originated from the geostatistics community [2, 3]. The Kriging Model is very flexible due to the wide range of correlation functions that may be chosen and to the high number of model parameters that should be adjusted in the approximation. This model is well-suited for deterministic applications and, more importantly, it has the ability to estimate errors in its predictions. [4], while RBF deals with a much simpler formulation with lower computational cost (yet robust) with respect to Kriging. RBF is also well adapted for numerical experiments with deterministic errors and was credited as the most dependable by Jin et al. [5] in a comparative study including 5 metamodelling techniques.

Another important feature to define a surrogate model is the sampling plan size. When kept constant throughout the optimization, the surrogate model is said to be static. This type of modeling usually demands a large number of samples to produce reasonable approximations, which may come across with the idea of reducing the computational cost itself. On the other hand, when the initial sampling plan is updated, we have the so-called Sequential Approximate Optimization (SAO). In this approach, the sampling plan is updated with points on the design space capable of improving the surrogate model and help the algorithm find the global optimum.

To select a new infill point, one must choose the infill criteria that best fits the goals of the surrogate model application. Some infill criteria aim the global improvement of the surrogate model accuracy (exploration), some aim the improvement in the local vicinity of the optimum found so far (exploitation) and others seek the balance between exploration and exploitation with the addition of a single point. Alternatively, one may use parallelization techniques to reduce execution time. Several authors have reported speed gains and improved algorithm performance in laminated composite optimization (the structural optimization type of problem explored later on) using these techniques [6–8]. However, this approach usually demands high-performance clusters computers and is not addressed in this work.

The present work uses different SAO approaches to find the optimum solution of benchmark problems, as well as for validation purposes. A numerical comparison between the two surrogate models (RBF and Kriging) is carried out in terms of accuracy, computational efficiency and simplicity of implementation. In addition to that, the influence of each infill criterion on the SAO performance is considered. After that, a structural optimization problem involving composite structures is presented.

## 2 Surrogate models

Surrogate models are an attempt of mapping a function of $m$ variables $y = f(\mathbf{x})$, which is treated as a black box, that converts the design variables vector $\mathbf{x}$ into a scalar $y$. This black-box function may represent experimental results or deterministic outputs provided by numerical simulations. For that matter, a limited number of samples of the design space and their corresponding High-Fidelity Model (HFM) analyses results are used to build a model capable of finding the best guess of $\hat{f}(\mathbf{x})$ using cheaper-to-evaluate functions. This way, the computational cost may be drastically reduced.

According to Forrester et al. [3], surrogate modeling consists of three main parts: (i) preparation of the data, which includes the screening of the design variables, (ii) sampling plan and (iii) choice of the modeling approach. The most common sampling plans used are the Latin Hypercube Sampling (LHS), as well as its improved version Optimized Latin Hypercube Sampling (OLHS), and Hammersley Sequence Sampling (HSS), discussed in the following Section. After that, the model parameters are chosen to best fit the data using closed-form equations, estimation criteria such as the Maximum Likelihood Estimator (MLE) or Cross-Validation (CV) techniques.

There is a wide variety of surrogate modeling techniques and several levels of complexity: single-fidelity models, hybrid surrogate models, adaptive sampling-based and multi-fidelity surrogate models. Typical single-fidelity surrogate models include the classical Polynomial Regression, Artificial Neural Networks, Support Vector Regression, Radial Basis Functions and Kriging Model, also referred to as Design and Analysis of Computer Experiments (DACE).

Of particular interest to this work is the adaptive sampling-based approach, also referred to as SAO. This approach uses auxiliary criteria called *infill strategies* to choose new sample point(s) in the design space with high potential to improve the model accuracy, providing better generalization properties to the surrogate model. Sections 2.2 and 2.3 deal with the mathematical formulation of the last two mentioned surrogate models, respectively.

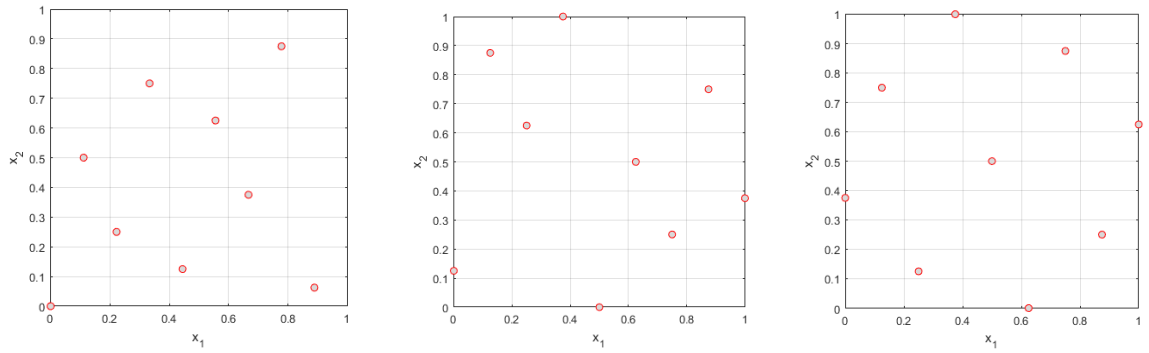
### 2.1 Sampling plan

In general, it is desirable to create a uniform distribution of the sampling points in order to achieve a certain level of uniformity in the accuracy throughout the design space. A sampling plan with such feature is said to be *space-filling* [9]. The full-factorial sampling technique is the most straightforward approach, although it is easily compromised by the Curse of Dimensionality in high-dimensional problems [3].

Three DoE techniques stand out among the numerous formulations in the literature: the Hammersley Sequence Sampling (HSS), the Latin Hypercube Sampling (LHS) and the Optimized Latin Hypercube Sampling (OLHS). The HSS is a deterministic low-discrepancy experimental design proposed by Kalagnanam and Diwekar [10]) for placing $n$ points in a $m$-dimensional hypercube based on the Hammersley's points [11]. Fig. 1(a) depicts the sampling plan obtained for $m$ = 2 and $n$ = 9.

As for the Latin Hypercube Sampling, this is a stochastic sampling plan obtained by splitting the design space into equally sized hypercubes (also called *bins*), placing one point in each and ensuring that from each occupied bin it is possible to exit the design space along any direction parallel with any of the axes without encountering any other occupied bin. However, this approach can lead to a poor sampling plan in terms of "space-fillingness" [3], as shown in Fig. 1(b). To overcome this issue, a metric known as *maximin* can be used to quantify the uniformity of a LHS. Based on this metric, Morris and Mitchell [12] proposed a criterion to find the best LHS arrangement, also known as Optmized Latin Hypercube Sampling, as illustrated in Fig. 1(c). The present work uses the MATLAB code provided by Forrester et al. [3] to generate the OLHSs. The authors use a GA to solve the optimization problem.

Note that the design variables are shown in the domain $[0, 1]^m$. This is a standard proceeding that aims to eliminate the effect of scale discrepancy on the surrogate model performance, particularly on the determination of the widths of the RBF (discussed in the following sections). Forrester et al. [3] states that the scaling of the observed data does not affect the values of the Kriging model hyper-parameters, but the scaling of the design space does. Thus, the normalized design variables can be described by:



(a) Hammersley Sequence (b) Latin Hypercube (c) Optimized Latin Hypercube

Figure 1. Different sampling plan techniques for 2D problem

$$x_{i_{sc}} = \frac{x_i - x_{low}}{x_{upp} - x_{low}}, \tag{1}$$

where $x_{low}$ and $x_{upp}$ are the lower and the upper bounds of the design variable $x_i$.

Finally, there are no formal guidelines regarding the number of sampling points required, which is also a motivation to use SAO, where the points is incrementally increased until a certain level of accuracy is reached. In practice, there are a few recommendations in literature such as the rule of thumb considered by Jin and Jung [13] and Jones et al. [14], where $n \approx 10\ m$.

## 2.2 Radial Basis Functions

The Radial Basis Function surrogate model uses a linear combination of radially symmetric functions centered around a set of points (basis function centres) scattered around the design space according to a given data set [3]. This type of surrogate model was first proposed by Hardy [2] aiming the interpolation of topographic scattered data.

Considering a scalar function $f(\mathbf{x})$ and the vector of sample points in the design space $\mathbf{x} = \{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\}^T$ which provides the vector of responses $\mathbf{y} = \{y^{(1)}, y^{(2)}, \ldots, y^{(n)}\}^T$, a typical form of RBF can be expressed as:

$$\hat{f}(x) = \mathbf{w}^T \psi = \sum_{i=1}^{N_c} w_i \ \psi \ (\underbrace{||\mathbf{x} - \mathbf{c}^{(i)}||}_{r}), \tag{2}$$

where $N_c$ is the number of basis functions centres, $\hat{f}(x)$ is the prediction of the true response function, $\mathbf{c}^{(i)}$ refers to the $i$th-centre among the $N_c$ basis function centres and $\boldsymbol{\psi}$ is the vector of size $N_c$ containing the values of the basis functions $\psi$ themselves evaluated at the Euclidean distances ($r$) between the testing point $\mathbf{x}$ and the centers of the basis functions $\mathbf{c}^{(i)}$. In this work, boldface indicates a vector or a matrix.

The Gaussian function is chosen as the basis function of RBF models. This type of function provide more freedom to improve the generalization when compared to other basis function that depend exclusively on the distance between points at the cost of a more complex parameter estimation process:

$$\psi = \mathrm{e}^{-r^2/(2\sigma^2)}, \tag{3}$$

where $\sigma$ is known as *spread* or *shape parameter*. In addition to that, under certain assumptions, Gaussian functions always lead to symmetric positive definite Gram matrices [3]. Ensuring the interpolation condition $f(\mathbf{x}) = \hat{f}(\mathbf{x})$, a system that is linear in terms of the basis functions weights $\mathbf{w}$ despite any possible non-linearity of the $f(\mathbf{x})$ response is obtained, which is an important characteristic of RBFs [15].

One of the conditions for the system to have one single solution is that it must be square (i.e. $N_c = n$, where $n$ is the number of samples) and that the samples points must be sufficiently distant from each other [3]. This is because very closely spaced points in $\mathbf{x}$ can cause ill-conditioning and as consequence, failure of the Cholesky factorization may occur. Usually, this is not a problem in space-filling sampling plans but may become an issue in the SAO approach if infill points are added subsequently in a very small area of interest in the design space. Also, calculations are simplified if the basis coincide with the data points, that is $\mathbf{c}^{(i)} = \mathbf{x}^{(i)}$ for any $i = 1 \dots n$, leading to the following equation:

$$\mathbf{\Psi}\ \mathbf{w} = \mathbf{y} \implies \mathbf{w} = \mathbf{\Psi^{-1}}\ \mathbf{y}, \tag{4}$$

where $\mathbf{\Psi}$ is the so-called *Gram matrix* defined as $\Psi_{i,j} = \psi_j\ (||\mathbf{x}^{(i)} - \mathbf{x}^{(j)}||)$, where $i, j = 1, \dots, n$. Thus, the basis function weights are obtained simply by solving the linear system in Eq. (4).

Furthermore, to avoid overfitting, a phenomenon that occurs when the built model closely explain the training data set but fails to generalize well when applied to unseen data, one may split the initial data set into separate *training* and *testing* subsets. If the model performs much better on the training subset than on the testing subset, that is an indicator of overfitting. Particularly in RBF models, these concerns can be addressed by determining the basis functions weights by minimizing the least square error considering a regularization parameter ($\lambda_i$) [16]:

$$E = \sum_{i=1}^{n} \left(f(x_i) - \hat{f}(x_i)\right)^2 + \sum_{i=1}^{n} \lambda_i\ w_i^2, \tag{5}$$

where $\lambda_i$ is recommended to be sufficiently small (e.g.: $1.0\ 10^{-3}$). This way, the data training is equivalent to finding $\mathbf{w}$ according to:

$$\mathbf{w} = \left(\mathbf{\Psi^T}\ \mathbf{\Psi} + \boldsymbol{\lambda}\right)^{-1}\ \Psi\ \mathbf{y}, \tag{6}$$

where $\boldsymbol{\lambda}$ is a diagonal matrix. It is worth mentioning that although the correct choice of the weights alone *does* ensure that the approximation can replicate the training data, this is not enough to guarantee a model with a minimum estimated generalization error. That is a role played by the choice of the model parameters (i.e., $\boldsymbol{\sigma}$ in the Gaussian function). In fact, the surrogate model parameters definition is typically evaluated prior to the weights' evaluation.

Several approaches were proposed to evaluate the shape parameter $\sigma$. Fig. 2 depicts the behaviour of a Gaussian function with varying $\sigma$. It can be seen that lower values of $\sigma$ may lead to non-smooth response surfaces. Moreover, too large values of spread may cause to a Runge's phenomenon interpolation [17]. Therefore, the estimation of such a parameter is a key-stage to achieving a good surrogate model.

Many researchers have proposed equations that simplify its calculation, skipping an arduous task that could be itself another topic for optimization, as proposed by Wu et al. [17]. Usually, the proposed closed-form expressions depend on the maximum distance between the sample points ($d_{max}$). In this work the formulation proposed by Kitayama and Yamazaki [18] is used. The authors considered the sparseness and density of the sampling points to propose a different spread value for each sampling point given by:

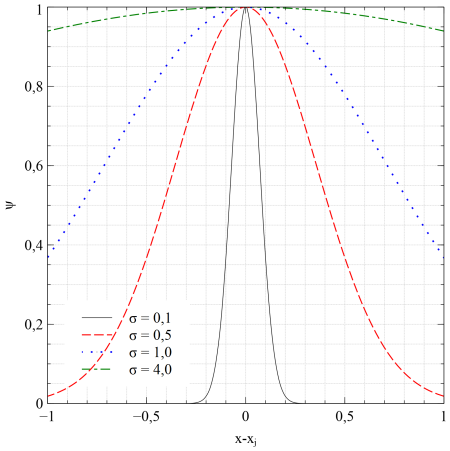



Figure 2. Varying $\sigma$ for Gaussian function

$$\sigma_i = \frac{d_{i,max}}{\sqrt{m}\,\sqrt[m]{n-1}}, \tag{7}$$

where $d_{i,max}$ is the maximum distance between sample $i$ and any other sample point in the design space. Kitayama and Yamazaki [18] used these widths associated with an adaptive scaling technique. In brief, the adaptive scaling technique aims to make all widths to be bigger than unity. This way, a smooth regression of the response surface is achieved as the influence area of each basis on the design domain is higher. However, in the authors' experience, this methodology may lead to convergence problems in the solution of optimization problems. Thus, the bases widths are simply evaluated using Eq. (7) and the conventional scaling technique (see Eq. (7)). Another alternative to get better predictions of the basis widths can be obtained (at a higher computational cost) using cross-validation [3, 19].

### 2.3 Kriging

This model was developed by the mining engineer Daniel Krige and made its way into engineering design through Sacks et al. [20] when the technique was applied to the approximation of computer experiments. Kriging, as RBF, is a nonparametric interpolating model, which means that the training points are involved in the determination of the unknown parameters and that the model exactly interpolates the responses at the sample points. As a matter of fact, Kriging has multiple variants and its general form may be described as

$$\hat{y}(\mathbf{x}) = g(\mathbf{x}) + Z(\mathbf{x}), \tag{8}$$

where the first term concerns the global trend of the model and the second term refers to the localized deviations that are autocorrelated. Often the global trend is taken as unknown, but constant ($\mu$), which results in the definition of the Ordinary Kriging, the version explored here.

When using Kriging, the responses should be understood as the result of a stochastic process, even if they come from a deterministic computer code (e.g., FEA). More specifically, in Eq. (8), $z(\mathbf{x})$ is assumed to be a realization of a stochastic process with mean zero and covariance given by [21]:

$$cov[\mathbf{Y}, \mathbf{Y}] = \sigma^2\, \mathbf{\Psi}, \tag{9}$$

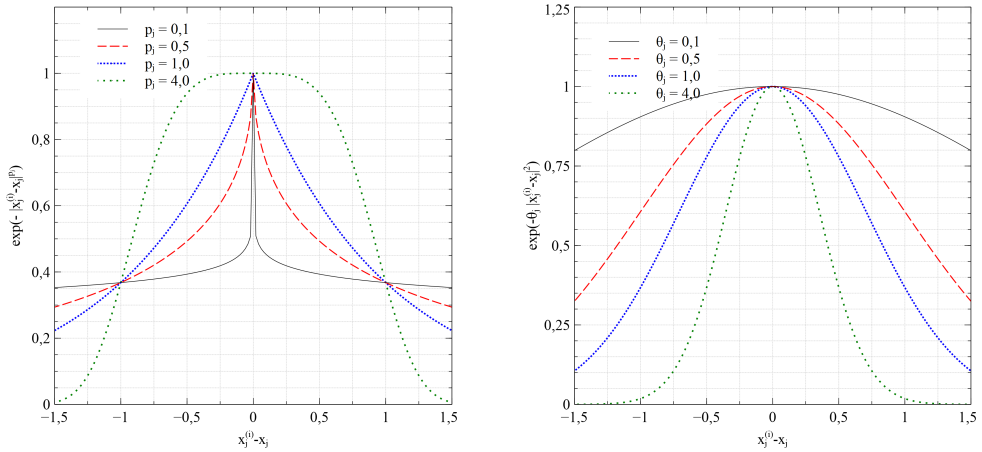
where $\sigma^2$ is the process variance and $\mathbf{\Psi}$ is the correlation matrix of all samples. To build this matrix, the sample data $\mathbf{x} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}\}^T$ with observed responses $\mathbf{y} = \{y^{(1)}y^{(2)}, \ldots, y^{(n)}\}^T$ should actually be seen as the realization of a stochastic process represented by $\mathbf{Y} = \{y^{(1)}y^{(2)}, \ldots, y^{(n)}\}^T$ with mean $\mathbf{1}\ \mu$, where $\mathbf{1}$ is an $n$ x 1 vector of ones. Moreover, the random variables must be spatially correlated with each other using a basis function. In this work, the correlation function is given by:

$$cor[Y(\mathbf{x}^{(i)}), Y(\mathbf{x}^{(l)})] = \exp(\sum_{j=1}^{m} \theta_j \, |x_j^{(i)} - x_j^{(l)}|^{p_j}). \tag{10}$$

Thus, the correlation ($\Psi$) matrix of *all* samples can be evaluated as:

$$\mathbf{\Psi} = \begin{bmatrix} cor[Y(\mathbf{x}^{(1)}), Y(\mathbf{x}^{(1)})] & \ldots & cor[Y(\mathbf{x}^{(1)}), Y(\mathbf{x}^{(n)})] \\ \vdots & \ddots & \vdots \\ cor[Y(\mathbf{x}^{(n)}), Y(\mathbf{x}^{(1)})] & \ldots & cor[Y(\mathbf{x}^{(n)}), Y(\mathbf{x}^{(n)})] \end{bmatrix}. \tag{11}$$

This matrix is positive semi-definite and symmetric with diagonals of ones as a result of the correlation function given in Eq. (10). To compute $\Psi$, the hyper-parameters $\mathbf{p}$ and $\boldsymbol{\theta}$ must be determined, or "tuned". This is a vital step to the model fitting since these parameters have a huge influence on the response surface and on the model accuracy. In Fig. 3(a), the behaviour of the correlation function is observed considering $\theta_j = 1$ fixed and varying $p_j$. The increase of this parameter leads to a smoother variation of the spatial correlation. In particular, for the lowest value considered, a near discontinuity occurs between near points. On Fig. 3(b), $p_j = 2.0$ is fixed and $\theta_j$ varies from 0.1 to 4.0, similarly to the spread on the RBF model. In this case, a low value of $\theta_j$ means that all sample points are highly correlated. In other words, $Y(x_j)$ is similar across the sample, while a high value indicates that there is a significant difference between the $Y(x_j)$'s $\theta_j$. Therefore, $\theta_j$ may be understood as a measure for how "active" is the approximated function regarding the design variable $x_j$.




(a) Varying $p_j$ (b) Varying $\theta_j$

Figure 3. Variation on the hyper-parameter of the correlation function

To tune the hyper-parameters, the Maximum Likelihoood Estimate is chosen. In this approach, the Likelihood Function (LF) described below should be maximized:

$$L(\mathbf{Y}|\ \mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}}\ \exp(-\frac{\sum(\mathbf{Y}^{(i)} - \mu)^2}{2\sigma^2}). \tag{12}$$

Expressing this equation in terms of the sample data and taking the natural logarithm to simplify the maximization:

$$\ln(L) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2}\ln(|\mathbf{\Psi}|) - \frac{(\mathbf{y} - \mathbf{1}\mu)\mathbf{\Psi}^{-1}(\mathbf{y} - \mathbf{1}\mu)}{2\pi\sigma^2}. \tag{13}$$

This way, to obtain the estimates for $\mu$ and $\sigma^2$, one must differentiate $L$ in respect to each of the variables and equate them to 0, which gives

$$\hat{\mu} = \frac{\mathbf{1^T\Psi^{-1}y}}{\mathbf{1^T\Psi 1}}, \tag{14}$$

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{1}\hat{\mu})^T\mathbf{\Psi^{-1}}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{15}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the estimates for the mean and variance, respectively. Replacing these estimates in Eq. (13) and removing constant terms to give what is known as *concentrated ln-likelihood function*:

$$\ln(L) \approx -\frac{n}{2}\ln(\hat{\sigma}^2) - \frac{1}{2}\ln|\mathbf{\Psi}|. \tag{16}$$

This function cannot be differentiated to obtain an analytic expression to describe the hyper-parameters. Thus, an iterative numerical optimization technique must be used. In this work, bioinspired algorithms are used for both levels of the optimization (the first being the surrogate model parameter tuning and the second being the structural optimization itself).

In short, the MLE evaluates the probability of a given data set having resulted from $\hat{f}(\mathbf{x})$ is evaluated according to a chosen distribution (usually, the normal distribution) and maximized. Alternatively, if $n$ and $m$ are not too large, direct search through the search space may be a simple and feasible option. Typically, $\mathbf{p}$ is fixed and only $\boldsymbol{\theta}$ is tuned, reducing the complexity of the Kriging optimization problem. The most commonly used $p$ value is 2, which corresponds to a Gaussian function [22].

Finally, to predict the target response at a new point $\mathbf{x}$, the following equation can be applied:

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \boldsymbol{\psi}^{\boldsymbol{T}}\ \mathbf{\Psi^{-1}}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{17}$$

where $\boldsymbol{\psi}$ is the correlation vector between $\mathbf{x}$ and all training points. According to Forrester et al. [3], this prediction maximizes the likelihood of the sample data and the prediction, given our correlation parameters. Last, but equally important, is the estimated error measure provided by the Gaussian process, also referred to as the Mean Squared Error (MSE) ($\hat{s}^2$):

$$\hat{s}^2(\mathbf{x}) = \hat{\sigma}^2\Big[1 - \boldsymbol{\psi}^{\boldsymbol{T}}\mathbf{\Psi^{-1}}\boldsymbol{\psi} + \frac{1 - \mathbf{1^T\Psi^{-1}}\boldsymbol{\psi}}{\mathbf{1^T\Psi^{-1}1}}\Big]. \tag{18}$$

The third term accounts for the uncertainty in the estimate of $\mu$ ($\hat{\mu}$). The MSE provides a measure of the quality of the prediction, it is always non-negative and values closer to zero are better. Another important observation is that the MSE is equal to zero at any sampled point since there is no uncertainty about points evaluated using the HFM. The correlation between errors also affects the prediction accuracy in the sense that the farther we get from the sample points, the less confidence in the prediction we get. The complete derivation of the equations given in this Section can be found in Sacks et al. [20].

### 2.4 Surrogate model performance assessment

At this stage, the quality or fidelity of the surrogate model is assessed using performance measures such as the NRMSE. In addition to accuracy measures, efficiency and robustness are also important features in a surrogate model quality assessment. These three aspects can be formally defined as:

a) accuracy: measures how close the surrogate model is to the true function over the design space. In this work, the metric selected to quantify this feature is the NRMSE;
b) efficiency: measures the amount of resources employed to produce reasonable results (e.g. 1% accuracy). To do so, the number of HFM evaluations is considered;
c) robustness: measures the ability of the model to consistently present good results in different test problems. To that end, the standard deviation of NRMSE is used to a set of benchmark problems;

The accuracy metric is given by

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{n_v}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_v}(y_i)^2}}, \tag{19}$$

where $n_v$ is the number of validation points considered. Lower values of NRMSE indicate better performances. Note that the non-normalized version of the RMSE typically provides values at the same order of the true function, which could result in an unfair comparison between different functions. This metric is argued to be among the best overall measures of model performance [23].

Finally, the standard deviation of the NRMSE ($STD_{NRMSE}$) is given by:

$$STD_{NRMSE} = \sqrt{\frac{\sum_{i=1}^{n_r} NRMSE_i - \overline{NRMSE}}{n_r - 1}}, \tag{20}$$

where $n_r$ is the number of runs for the same problem set (same sampling plan, optimization parameters, etc.) and $\overline{NRMSE}$ is the mean of the NRMSEs values obtained by the $n_r$ runs. Smaller $STD_{NRMSE}$ values suggest a more robust surrogate model or to put it another way, the less problem-dependent the surrogate model is.

## 3 Sequential Approximate Optimization

Historically, Sequential Approximate Optimization began with the work of Schmit and Farshi [24]. The authors used mathematical programming methods and approximation concepts to improve structural synthesis efficiency by alleviating excesses such as the consideration of all constraints rather than working on critical or "near" critical constraints at each stage of the iterative design process. Another important contribution was provided by Jones et al. [14] with the widely known Efficient Global Optimizatiion (EGO) algorithm.

Bottom line, the core of the SAO approach lays on the criteria used to update the surrogate model. To optimization applications, using solely exploration-based strategies may reduce the convergence efficiency as many non-profitable areas are investigated, whilst using solely exploitation-based strategies

may lead to a deceptive optimum if the initial sampling plan does not provide good coverage of the design space and the global accuracy is compromised. This trade-off is often addressed with the use of *hybrid criteria* (e.g., Probability of Improvement (PI) and Expected Improvement (EI)). These methods strive the balancing between exploitation and exploration.

In this work, three frameworks are considered:

1. **KRG EI**: This method is based on the implementation of the EGO algorithm proposed by Jones et al. [14]. Kriging is used to build the surrogate model and the new infill point is determined using the EI criterion (see Section 3.2). To benchmarking purposes, the actual difference between the best sample found so far and the optimal point is used as the stopping criterion. However, in real-life problems, the EI value should be adopted as stopping criterion;
2. **RBF EI**: This method is basically the same applied in the KRG EI, but using the RBF as the surrogate model [25]. The basis function widths are evaluated using Eq. (7);
3. **RBF Best+DF**: This method is based on the formulation proposed by Kitayama et al. [16]. However, in this work, only two new points are added at each iteration of the optimization algorithm, instead of $m/2$ new sample points. One point is simply taken as the best non-repeated solution at each generation and the other infill point is determined by the Density Function criterion (see Section 3.1).

In all three approaches, two stopping criteria are considered: the maximum number of generations and a minimum accuracy level. The latter is only applied when dealing with benchmark problems. The infill points determined by the Expected Improvement criterion using the RBF as surrogate model were implemented following the formulation proposed by Sóbester et al. [25], without the cycling weight feature. The SAO-RBFs were originally implemented by Balreira [26] on an in-house program named Bio-Inspired Optimization System (BIOS) and modified for this work with the addition of the EI criterion, while the SAO-Kriging approach was fully implemented by the authors of this paper. BIOS is written in C++ language and uses the OOP philosophy [27].

When dealing with the EI criterion, a variation of the EGO algorithm is employed, as depicted in Fig. 4. In this version, the algorithm starts with the general proceeding of a SAO: generation of the initial sampling plan using a DoE and evaluation of sample points using the HFM. Next, the Leave-One-Out Cross-Validation (LOOCV) is applied to determine the Kriging model that provides the lowest error value, which is given by:

$$\frac{y(\mathbf{x}^{(i)}) - \hat{y}(\mathbf{x}^{(i)})}{\hat{s}(\mathbf{x}^{(i)})}, \tag{21}$$

where $\hat{s}$ is calculated as shown in Eq. (18). This error is referred as "standardized cross-validated residual" and should be roughly in the interval [-3, 3] in order to validate the Kriging model. Jones et al. [14] proposes a more "visual" and qualitative diagnostic of the model fitting by plotting error measures and suggesting, in case of fail, the transformation of the dependent variable, typically using the log transformation ($\log(y)$) or the inverse transformation ($-1/y$). Note that if the RBF is used as the surrogate model, the selection and training of the initial surrogate model is much simpler: the basis functions widths are evaluate as proposed by Kitayama et al. [16].

To define the hyper-parameters of the Kriging model, the Maximum Likelihood Estimator is used. Once the LOOCV is done and the initial Kriging model is fit and selected, the algorithms proceeds to the maximization of the EI, which can be used as both the infill criterion and stopping criterion. Jones et al. [14] suggest that if the expected improvement is less than 1% of the best current function value (on the untransformed scale), the algorithm should be terminated. Otherwise, the design with highest EI value is sampled and Kriging hyper-parameters are re-estimated, and iterate. However, for benchmarking purposes, the actual error between the best sample point found so far and the optimal point will be considered.

Finally, the last method (RBF Best+DF) is depicted in Figure 5. The algorithm starts with the definition of the initial sampling plan and evaluation of the points using the HFM. Then, the initial

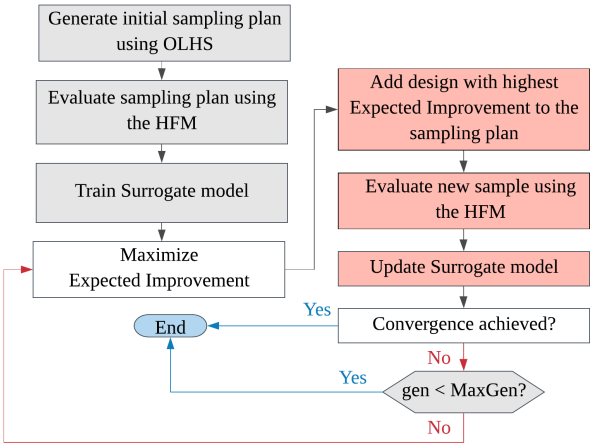



Figure 4. KRG EI and RBF EI algorithm

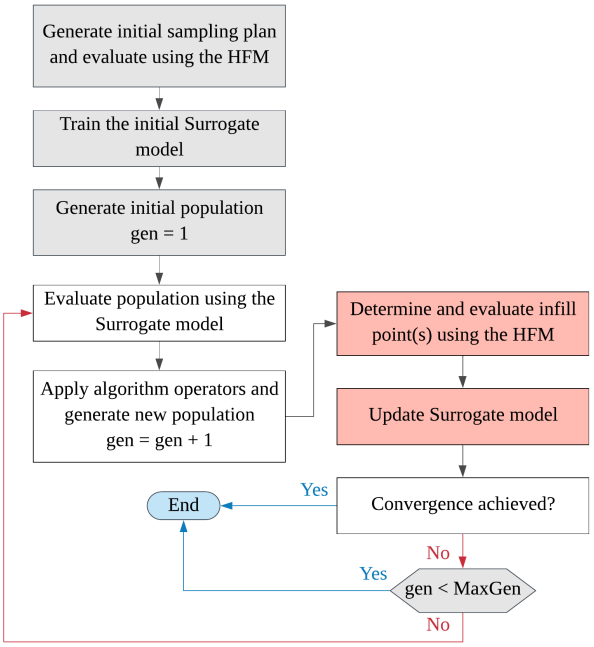



Figure 5. RBF Best+DF algorithm

surrogate model is built and the initial population of the algorithm is generated randomly. After that, the main optimization loop is initiated. In this loop, the trial designs are evaluated using the surrogate model and the algorithm operators (in this case, crossover, mutation, etc.) are applied to create the new population. After that, an infill criterion is used to determine two new infill points to update the sampling plan. As a consequence, the surrogate model should be updated and the hyper-parameters re-evaluated. When the stopping criterion is achieved, the best trial design found by the optimization algorithm is compared to the best sampling point. If there is a sampling point with better objective function than the current best individual, then the best solution will be taken as the sampling point.

### 3.1 Density function

The Density Function (DF) strategy was introduced by Kitayama et al. [16] and successfully applied by the authors [16, 28] and Pan et al. [29]. This function is used to find sparse regions in the design space and determine an infill point. It is expected that this point will then lead to a better global approximation.

The basic procedure to build the density function is described as follows:

1. At the sampling points, the output is replaced with +1, the maximum normalized value allowed for a design variable as in $\mathbf{y}^D = \{1, 1, \ldots, 1\}^T$.
2. The weight vector of the DF is given by:

$$\mathbf{w}^D = (\boldsymbol{\Psi}^{\boldsymbol{T}}\boldsymbol{\Psi} + \boldsymbol{\lambda})\boldsymbol{\Psi}^{\boldsymbol{T}}\mathbf{y}^D. \tag{22}$$

3. Solve the minimization problem where the objective function is given by

$$D(\mathbf{x}) = \sum_{i=1}^{n} w_i^D \, \psi_i((x)). \qquad \text{for } i = 1, 2, \ldots, m. \tag{23}$$

Note that any optimization algorithm could be applied. In this work, the PSO algorithm is used.

4. Add best point obtained in Step 3 as update point.

In the RBF Best+DF, the best solution found so far by the GA algorithm is added to the initial sampling plan and the surrogate model (and the DF) is updated. Then, the optimal point of the minimization of the DF is added to the sampling plan and the surrogate model is, again, updated. This is illustrated in Figure 6 through a simple 1D problem. In Fig. 6(a), the initial surrogate model and the initial sampling plan are shown. Next, the DF is built (see Fig. 6(b)) and minimized. Then, the optimal point of this optimization problem is used to updated the surrogate model, as shown in 6(c).

### 3.2 Expected Improvement

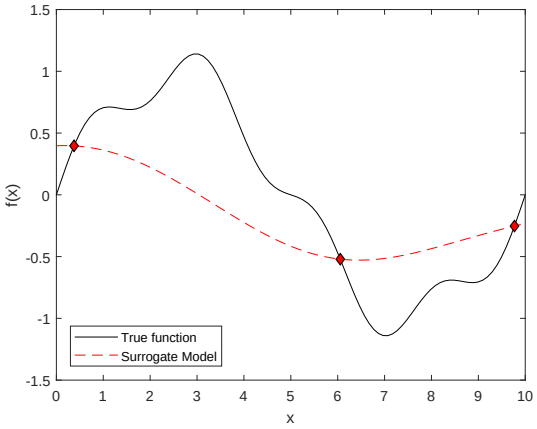
This approach was proposed by Mockus et al. [30] and takes into account the magnitude of possible improvement on the best current best value $(I(\mathbf{x}))$. The EI aims to balance the desire to search at locations with good predicted values (exploitation) with the desire to check where the uncertainty of prediction is large (exploration). This concept gained popularity when Jones et al. [14] applied it as an infill criterion in their EGO algorithm. If the current best value is $y_{min}$, then $I(\mathbf{x})$ may be written as $I(x) = y_{min} - y(\mathbf{x})$, or more precisely:

$$E(I(\mathbf{x})) = \begin{cases} (y_{min} - \hat{y}(\mathbf{x})) \; \Phi\left(\frac{y_{min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) + s \; \phi\left(\frac{y_{min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right), & \text{if } \hat{s} > 0 \\ 0, & \text{if } \hat{s} = 0, \end{cases} \tag{24}$$

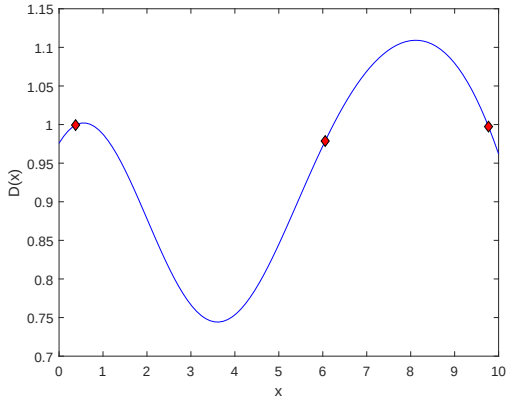
where $\Phi$ and $\phi$ are the normal cumulative distribution function and probability density function, respectively. Evaluating Eq. (24) using the error function (erf), we arrive at:

$$E(I(\mathbf{x})) = \begin{cases} (y_{min} - \hat{y}(\mathbf{x})) \left[\frac{1}{2} + \frac{1}{2} \, \text{erf}\left(\frac{y_{min} - \hat{y}(\mathbf{x})}{\hat{s}\sqrt{2}}\right)\right] \; + \; \hat{s} \, \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-(y_{min} - \hat{y}(\mathbf{x}))^2}{2\hat{s}^2}\right], & \text{if } \hat{s} > 0. \\ 0, & \text{if } \hat{s} = 0. \end{cases} \tag{25}$$

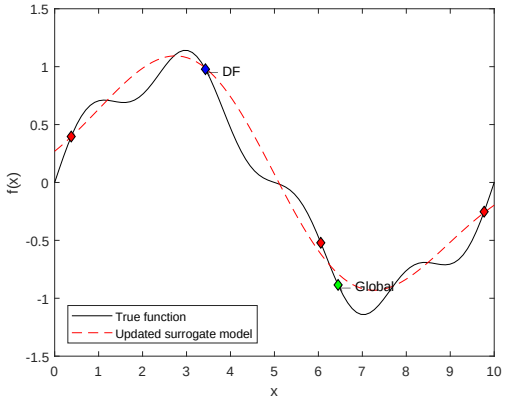
The first term of Eq. (25) corresponds to the *exploitation* contribution, where the predicted difference between the current minimum and the prediction $\hat{y}$ in $\mathbf{x}$, penalized by the probability of improvement, is computed. Whilst the second term corresponds to the *exploration* contribution, which provides larger values in areas where uncertainties are high (i.e., unsampled areas). Typically, $y_{min}$ is taken as the minimum over previous observations [14, 21, 31], but it may also be set as the minimum of the expected value of the target [32]. For further details on the development of Eq. (24), the reader is encouraged to look for the reference [21].




(a) Initial surrogate model

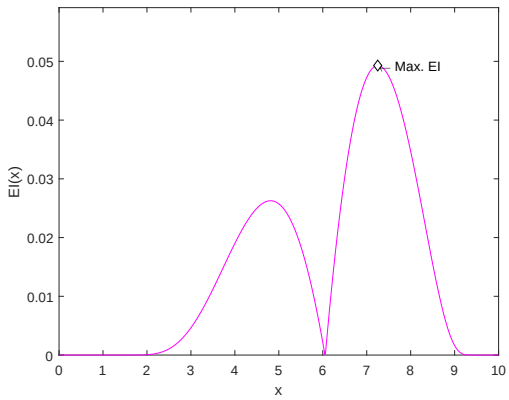



(b) Density Function

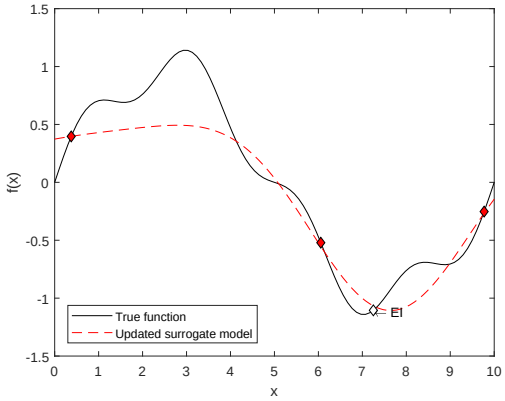



(c) Updated surrogate model

Figure 6. Addition of two new samples (Global best + DF)




(a) Expected Improvement




(b) Updated surrogate model

Figure 7. Addition of one new sample point (EI)

In this work, the GA algorithm is used to maximize the EI due to its good performance in previous works at the LMCV. This criterion has been proved to find the global optimum [33]. However, Sóbester et al. [25] stresses that, although the criterion offers a simple and combined form of fusing exploration and exploitation, if the problem in hand is likely to yield a simple, unimodal surface, searching the predictor will probably work better. Conversely, if the target landscape is extremely multimodal, biasing

the search towards sampling in thus far unexplored areas could lead to faster convergence than the EI criterion. Figure 7(a) illustrates the EI for the 1D problem illustrated in Fig. 6(a) and Fig. 7(b) shows the updated surrogate model with the addition of the point that maximizes the EI function.

When dealing with constraints, one must define whether they should be evaluated using a HFM or using a surrogate model. In the first case, the EI is simply set to 0 if the trial design violates any constraint. However, if the constraint is also evaluated using a surrogate model, Forrester et al. [3] suggest a few approaches to deal with this uncertainty, including a Penalty approach.

## 4 Results

This section presents the results for 4 benchmark problems of different dimensionalities, as well as the application of the SAO for a structural problem, namely a laminated composite plate. The benchmark problems are unconstrained mathematical functions and were carefully selected to illustrate the use of the different surrogate models discussed in this work. Note that these functions are not expensive-to-evaluate in a sense that the optimization with the true-function would be prohibitive, but they were chosen due to the abundance of historical search performance data that could be used for comparison.

### 4.1 Mathematical functions

Table 1 presents the main features of the mathematical functions selected and Table 2 presents a more detailed description of the number of design variables, the design domain and the global minimum of each function. The coefficients $A_{ij}$ and $P_{ij}$ of the Hartmann 6 function are shown in Dixon and Szegi [34].

Table 1. Main features of test functions

| Function name | Code | $m$ | Design domain | Local minima | Global minima |
|---|---|---|---|---|---|
| Branin | BN | 2 | $x_1 \in [-5, 10]\ \ x_2 \in [0, 15]$ | 3 | 3 |
| Goldstein Price | GP | 2 | $x_i \in [-2, 2]$ | 4 | 1 |
| Hartmann | H3 | 3 | $x_i \in [0, 1]$ | 4 | 1 |
| | H6 | 6 | $x_i \in [0, 1]$ | 6 | 1 |

To minimize the Density Function and maximize the Likelihood Function, the PSO algorithm was applied using the swarm size equal to 100, 250 iterations, probability of mutation ($p_{mut}$) equal to 10%. When dealing with KRG EI and RBF EI, the Genetic Algorithm was employed to maximize the Expected Improvement using 100 individuals, 250 generations, $p_{mut}$ = 15% and crossover rate ($r_{cross}$) equal to 80% at each cycle (i.e. each search for the design with highest EI). In both cases, the maximum number of HFM evaluations is set to 50 for the BN, GP and H3 problems and 100 for the H6 function. This stopping criterion was based on the values found by the original EGO algorithm [14].

The RBF Best+DF employs the GA using 100 individuals, 50 generations and $p_{mut}$ = 15% for BN, GP and H3. For the H6 function, the maximum number of generations was increased to 100. Note that, in this case, at each generation, 2 new points are added, which results in the maximum number of 100 HFM evaluations for the low-dimensional problems and 200 for the H6 problem.

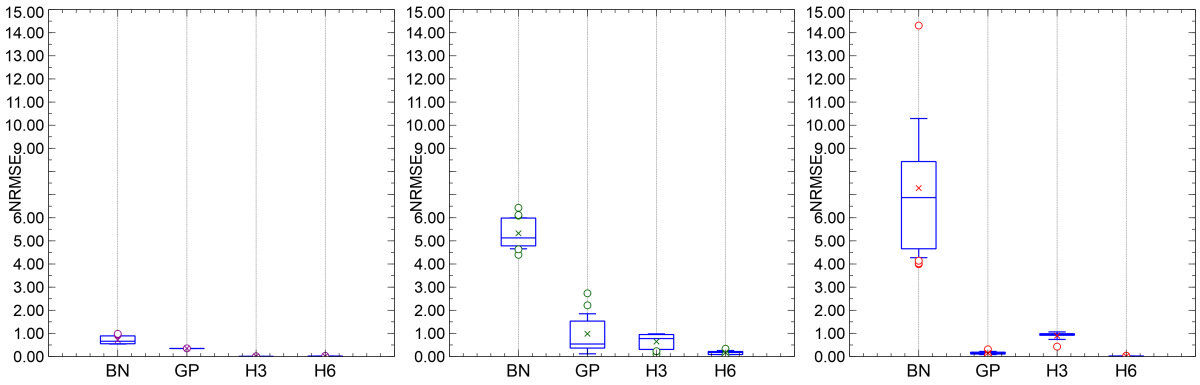
Finally, to create the initial sampling plan, the OLHS was employed. For the low-dimensional problems, $n = 10$ and for the H6 problem, $n = 20$. All results were averaged over 10 runs. The performance of each SAO technique regarding the number of HFM evaluations needed to reach 1% of difference to the actual optimal response is compared to the results found in the literature [14, 25, 35, 36] and shown in Table 3. The robustness metrics are shown in Table 4. In both cases, boldface indicates the best performance.

Table 2. Mathematical test functions

| Code | Function | Global minimum |
|---|---|---|
| BN | $f(\mathbf{x}) = (x_2 - \frac{5.1}{4\pi^2}\, x_1^2 + \frac{5}{\pi}\, x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})\cos(x_1) + 10)$ | $\mathbf{x} = \{-\pi, 12.28\}$<br>$\mathbf{x} = \{\pi, 2.275\}$<br>$\mathbf{x} = \{9.425, 2.475\}$<br>$f_{opt} = 0.398$ |
| GP | $f(\mathbf{x}) = [1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2)]$<br>$[30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)]$ | $\mathbf{x} = \{0.000, -1.000\}$<br>$f_{opt} = 3.000$ |
| H3<br>H6 | $f(\mathbf{x}) = -\sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{m} A_{ij}(x_j - P_{ij})^2\right)$ | $\mathbf{x} = \{0.202, 0.150,$<br>$0.477, 0.275,$<br>$0.312, 0.657\}$<br>$f_{opt} = -3.322$ |

Table 3. Number of HFM evaluations needed for at least 1% accuracy

| | | | | | Present work | | |
|---|---|---|---|---|---|---|---|
| Function | WEIF [25] | Gutmann [36] | DIRECT [35] | EGO [14] | KRG EI | RBF EI | RBF BEST+DF |
| Branin | 34 | 44 | 63 | **28** | **28** | - | 79 |
| Goldstein Pr. | **32** | 63 | 101 | **32** | 42 | - | 60 |
| Hartmann 3 | 28 | 25 | 83 | 35 | **17** | 45 | 93 |
| Hartmann 6 | **33** | 112 | 213 | 121 | 56 | - | 131 |




(a) KRG EI (b) RBF EI (c) RBF BEST+DF

Figure 8. NRMSE for different SAO techniques

Note that, considering the maximum number of HFM evaluations (or generations) stopping criterion, the only framework not to reach the 1% accuracy is the RBF EI. The same pattern is observed when we look at the H6 problem, where the maximum number of generations was doubled. The average error for the Branin, Goldstein Price and Hartmann 6 functions using this method are 36%, 79% and 14%, respectively. In general, KRG EI is in-line with the results found in the literature and, in some cases, performed better than the algorithm by Jones et al. [14] and Sóbester et al. [25]. The latter uses a weighted version of the EI expression and determines the basis function widths using the LOOCV

technique. Furthermore, the addition of two new points also provides good results, but at the cost of a much higher number of HFM evaluations.

When it comes to the robustness, KRG EI is the method with the lowest NRMSE standard deviations, which means that this methodology provides the most consistent (i.e., less variable) surrogate models. Figure 8 illustrates the accuracy of the different SAO techniques. One may observe the following order of accuracy: KRG EI $>$ RBF EI $>$ RBF Best+DF, which is not reflected in the algorithm capacity of finding the global optimum.

Table 4. Robustness $STD_{NRMSE}$

| Function | KRG EI | RBF EI | RBF BEST+DF |
|---|---|---|---|
| Branin | **0.185** | 0.706 | 3.171 |
| Goldstein Pr. | **0.006** | 0.396 | 0.073 |
| Hartmann 3 | **0.004** | 0.352 | 0.171 |
| Hartmann 6 | **0.009** | 0.092 | 0.012 |

### 4.2 Laminated composite plate

This problem was originally proposed by Balreira [26] and deals with the maximization of the strength of a simply supported square laminated plate. The side length is worth 0.720 m and a circular hole ($r = 0.072\ m$) is placed in the middle of it. The laminated plate is made of 40 plies, each with 0.3 mm, and fixed material (see Table 5).

Table 5. Material properties

| $E_1$ (GPa) | $E_2$ (GPa) | $G_{12}$ (GPa) | $\nu_{12}$ | $\varepsilon_1^u$ | $\varepsilon_2^u$ | $\gamma_{12}^u$ |
|---|---|---|---|---|---|---|
| 127.59 | 13.03 | 6.41 | 0.30 | 0.008 | 0.029 | 0.015 |

The design variables are the fiber orientation of each ply and only balanced symmetrical layups are allowed, which reduces the number of design variables to 10. In addition to that, due to manufacturing constraints, the fiber orientations are considered to be discrete and can assume the following values: $0^\circ$ , $45^\circ$ or $90^\circ$. In short, the optimization problem may be described as:

$$\begin{aligned} &\text{Find} && \mathbf{x} = [\theta_1, \theta_2, ..., \theta_{10}]. \\ &\text{that maximize} && \min(\lambda_s, \lambda_b) \\ &\text{s. t.} && \text{Max contiguous plies} \leq 4, \end{aligned} \tag{26}$$

where $\lambda_s$ and $\lambda_b$ are the safety factor for buckling and material failure, respectively.

The strain failure load must be maximized considering the Maximum Strain Criterion and a minimum safety factor ($S_f$) of 1.50, as shown in Eq. (27), where $k$ refers to the strains of the $k$-th ply. The HFM analysis is carried out by an in-house program named Finite AnalysiS Tool (FAST). FAST is a program developed by LMCV collaborators based on the same premises as BIOS. Fig. 9 illustrates the mesh of quadratic shell elements considered in the analyses.

$$\lambda_s = \min_k \left( \min \left( \frac{\varepsilon_1^u}{S_f \varepsilon_1^k}, \frac{\varepsilon_2^u}{S_f \varepsilon_2^k}, \frac{\gamma_{12}^u}{S_f \gamma_{12}^k} \right) \right). \tag{27}$$

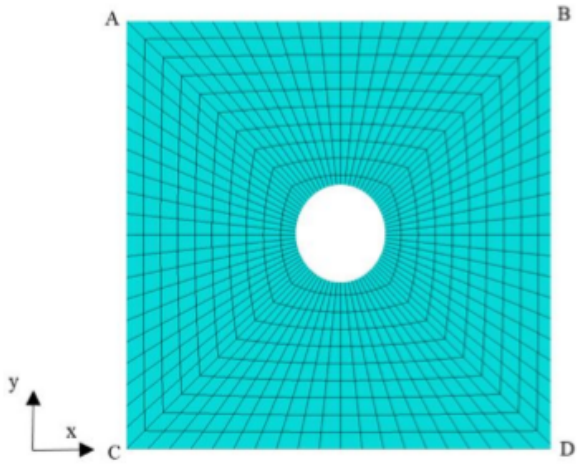



Figure 9. Laminated plate with circular hole

The initial sampling plan was generated using the HSS for $n = 185$. However, some of these points were discarded due to the contiguity constraint violation. In this case, before checking the constraint, each variable was rounded to the nearest discrete value allowed. For example, if $0.00° < x_i < 33.33°$, the algorithm will set $x_i$ as $0°$. If $33.33° \leq x_i < 66.66°$, $x_i$ will be set as $45°$ and so on. By the end of this process of rounding and constraint verification, only 99 sampling points were feasible.

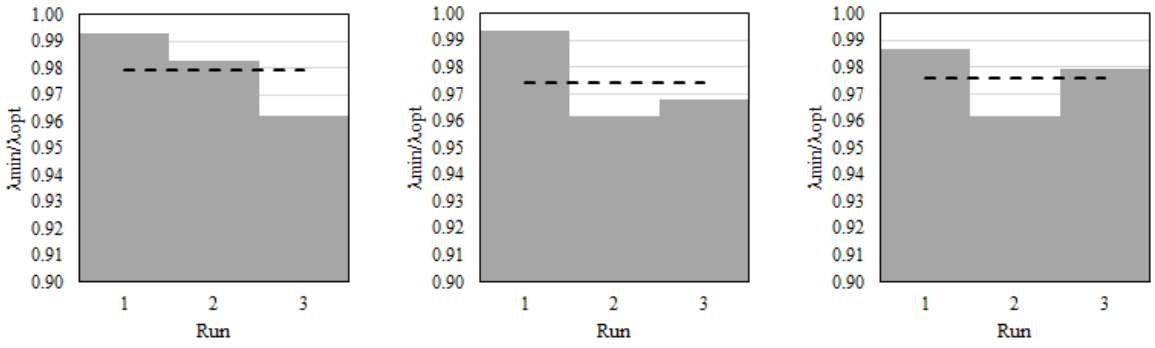
The reference results were obtained using the GA with the following optimization parameters: 3 runs, 50 individuals, 50 generations, $p_{mut} = 0.10$ and $r_{cross} = 0.80$. These parameters were also replicated for the SAO techniques. However, on the KRG EI and RBF EI methods, these are only the optimization parameters for the EI maximization. In these methods, only one infill point is added at each search, resulting in 50 HFM evaluations. On the RBF Best+DF, two infill points are added per generation, resulting in 100 infill points. In all cases, the maximization is treated as the negative of the objective function and the accuracy stopping criterion is not considered. Only the best performances are shown in Table 6. Note that all values were obtained using the HFM.

Table 6. Best layups for laminated plate problem

| Method | Layup (°) | $\lambda_b$ | $\lambda_s$ | $\lambda_{min}$ | Diff (%) |
|---|---|---|---|---|---|
| HFM | $[\pm 45_3\ 90_2\ 0_4\ (90_2\ 0_2)_2]_s$ | 1218.02 | 1237.87 | 1218.02 | - |
| KRG EI | $[\pm 45_2\ 90_2\ 0_2 \pm 45\ 0_2\ 90_2\ 0_4 \pm 45]_s$ | 1209.63 | 1275.17 | 1209.63 | 0.69 |
| RBF EI | $[\pm 45_2\ 0_2\ 90_2\ \pm 45\ 90_2\ 0_4\ 90_2\ 0_2]_s$ | 1210.18 | 1237.87 | 1210.18 | 0.64 |
| RBF BEST+DF | $[\pm 45\ 90_2\ \pm 45\ 0_4\ \pm 45_2\ 0_4\ 90_2]_s$ | 1201.62 | 1275.17 | 1201.62 | 1.35 |

In this problem, the RBF EI method provided the highest strength load factor, being only 0.64% lower than the best result obtained using the HFM. This result is closely followed by the Kriging EI and RBF Best+DF approaches. However. Figure 10 shows that this good performance should be treated carefully. The average $\lambda_{min}$ of the 3 approaches is represented by the dashed black lines. Again, the results obtained by KRG EI are more consistent, followed by the RBF Best+DF and, lastly, by the RBF EI. However, one must be cautious since the number of runs considered is low.

As for the number of HFM models, it is clear that the KRG EI and RBF EI are the methods with the lowest number of evaluations since only one point is added per iteration. Both methods use 149 evaluations (being 99 of them part of the initial sampling plan), while the RBF Best+DF uses 199 evaluations. This represents a reduction of 94% and 92%, respectively, on the 2500 evaluations used when dealing with the direct optimization of the HFM. It is worth mentioning that the suitability of each method is




(a) KRG EI (b) RBF EI (c) RBF BEST+DF

Figure 10. Normalized $\lambda_{min}$ for different SAO techniques

strongly influenced by the cost of the problem analysis and the computational budget available.

## 5 Final remarks

This work has compared the performance of three SAO techniques using multiple infill criteria, providing insightful observations on the performance of the different frameworks considered. For that end, benchmark problems of different dimensionalities and orders of nonlinearity were selected. As shown, Kriging excels in most categories.

Another interesting observation comes from the fact that the surrogate accuracy does not necessarily imply on the capacity of finding the global optimum, although a misleading response surface may be a difficulty for the optimization algorithm. When dealing with multiple global optimal solutions, namely the Branin function, all methods presented worse accuracy metrics (i.e., not all optimal points are close to their true function values) when compared to the ones with a single global optimum. However, this did not seem to be a hindrance for the optimization algorithm. This is especially important for the laminated problem, where multiple layups can lead to the same objective function value.

The maximization of the strength of the laminated plate was also well-succeed. All three methods provided layups that were only 2% far from the load factor obtained using the HFM for all evaluations. In all cases, the laminated composite plate fails first due to instability. In addition to that, the reduction in the number of HFM evaluations is a valuable gain when dealing with time-consuming structural analyses.

Finally, it is worth noting that efficiency is only considered by the number of HFM evaluations and not by the CPU time, for example. This is an important topic to be considered in future works since the time employed to define the hyper-parameters of Kriging is significantly higher than the one employed to define the width of the RBFs using closed-form expressions.

## Acknowledgements

The financial support by FUNCAP (Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) are gratefully acknowledged.

## References

[1] Arora, J. S., 2012. *Introduction to Optimum Design*. Elsevier Academic Press, Iowa, second edition.

[2] Hardy, R. L., 1971. Multiquadric Equations of Topography and Other Irregular Surfaces. *Journal of Geophysical Research*, vol. 76, n. 8, pp. 1905–1915.

[3] Forrester, A. I. J., Sbester, A., & Keane, A. J., 2008. *Engineering Design via Surrogate Modelling*.

[4] Wang, G. G. & Shan, S., 2006. Review of Metamodeling Techniques in Support of Engineering Design Optimization. *Journal of Mechanical Design*, pp. 1–42.

[5] Jin, R., Chen, W., & Simpson, T. W., 2001. Comparative studies of metamodelling techniques under multiple modelling criteria. *Structural and Multidisciplinary Optimization*, vol. 23, n. 1, pp. 1–13.

[6] Rocha, I. B. C. M., Parente Jr, E., & Melo, A. M. C., 2014. A hybrid shared/distributed memory parallel genetic algorithm for optimization of laminate composites. *Composite Structures*, vol. 107, n. 1, pp. 288–297.

[7] Omkar, S. & Senthilnath, J., 2011. Artificial Bee Colony (ABC) for multi-objective design optimization of composite structures. *Applied Soft Computing*, vol. 11, n. 1, pp. 489–499.

[8] Punch, W., Averill, R., Goodman, E., Lin, S.-C., Ding, Y., & Yip, Y., 1994. Optimal design of laminated composite structures using coarse-grain parallel genetic algorithms. *Computing Systems in Engineering*, vol. 5, n. 4, pp. 415 – 423. 3rd National Symposium on Large-Scale Structural Analysis for High-Performance Computers and Workstations.

[9] Simpson, T. W., Peplinski, J. D., Koch, P. N., & Allen, J. K., 2001. Metamodels for Computer-based Engineering Design : Survey and recommendations. *Engineering with Computers*, pp. 129–150.

[10] Kalagnanam, J. R. & Diwekar, U. M., 1997. An efficient sampling technique for off-line quality control. *Technometrics*, vol. 39, n. 3, pp. 308–319.

[11] Hammersley, J. M., 1960. Monte carlo methods for solving multivariable problems. *Annals of the New York Academy of Sciences*, vol. 86, n. 3, pp. 844–874.

[12] Morris, M. D. & Mitchell, T. J., 1995. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, vol. 43, n. 3, pp. 381 – 402.

[13] Jin, S.-s. & Jung, H.-j., 2016. Sequential surrogate modeling for efficient finite element model updating. *Computers and Structures*, vol. 168, pp. 30–45.

[14] Jones, D. R., Schonlau, M., & W. J. Welch, 1998. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, vol. 13, pp. 455–492.

[15] Song, X., Lv, L., Sun, W., & Zhang, J., 2019. A radial basis function-based multi-fidelity surrogate model: exploring correlation between high-fidelity and low-fidelity models. *Structural and Multidisciplinary Optimization*.

[16] Kitayama, S., Arakawa, M., & Yamazaki, K., 2011. Sequential Approximate Optimization using Radial Basis Function network for engineering optimization. *Optim. Eng. Journal*, pp. 535–557.

[17] Wu, Z., Wang, D., Okolo N., P., Jiang, Z., & Zhang, W., 2016. Unified estimate of Gaussian kernel width for surrogate models. *Neurocomputing*, vol. 203, pp. 41–51.

[18] Kitayama, S. & Yamazaki, K., 2011. Simple estimate of the width in Gaussian kernel with adaptive scaling technique. *Applied Soft Computing Journal*, vol. 11, n. 8, pp. 4726–4737.

[19] Ferreira, A., Fasshauer, G., Batra, R., & Rodrigues, J., 2008. Static deformations and vibration analysis of composite and sandwich plates using a layerwise theory and rbf-ps discretizations with optimal shape parameter. *Composite Structures*, vol. 86, pp. 328–343.

[20] Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P., 1989. Design and Analysis of Computer Experiments. *Statistical Science*, vol. 4, n. 4, pp. 409–435.

[21] Jones, D. R., 2001. A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization*, vol. 21, n. 4, pp. 39.

[22] Wang, J. T., Wang, C. J., & Zhao, J. P., 2017. Frequency response function-based model updating using Kriging model. *Mechanical Systems and Signal Processing*, vol. 87, n. August 2016, pp. 218–228.

[23] Li, J. & Heap, A. H., 2008. A Review of Spatial Interpolation Methods for Environmental Scientists. Technical report, Australian Government, Canberra, Australia.

[24] Schmit, L. & Farshi, B., 1974. Some approximation concepts for structural synthesis. *AIAA Journal*, vol. 12, n. 5, pp. 692–699.

[25] Sóbester, A., Leary, S. J., & Keane, A. J., 2005. On the design of optimization strategies based on global response surface approximation models. *Journal of Global Optimization*, vol. 33, n. 1, pp. 31–59.

[26] Balreira, D. S., 2018. *Otimização Sequencial Aproximada de Estruturas Laminadas de Material Compósito*. Master of civil engineering, Universidade Federal do Ceará.

[27] Rocha, I. B. C. M., 2013. *Análise e otimização de cascas laminadas considerando não-linearidade geométrica e falha progressiva*. Masters of civil engineering, Universidade Federal do Ceará.

[28] Kitayama, S., Srirat, J., Arakawa, M., & Yamazaki, K., 2013. Sequential approximate multiobjective optimization using radial basis function network. *Structural and Multidisciplinary Optimization*, vol. 48, n. 3, pp. 501–515.

[29] Pan, G., Ye, P., Wang, P., & Yang, Z., 2014. A sequential optimization sampling method for metamodels with radial basis functions. *Scientific World Journal*, vol. 2014.

[30] Mockus, J., Tiesis, V., & Zilinskas, A., 1978. *The application of Bayesian methods for seeking the extremum*, volume 2, pp. 117–129.

[31] Gan, N. & Gu, J., 2018. Hybrid meta-model-based design space exploration method for expensive problems. *Structural and Multidisciplinary Optimization*, pp. 907–917.

[32] Gelbart, M. A., Snoek, J., & Adams, R. P., 2014. Bayesian Optimization with Unknown Constraints. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pp. 1–14, Quebec City, Quebec, Canada. AUAI Press.

[33] Locatelli, M., 1997. Bayesian Algorithms for One-Dimensional Global Optimization. *Journal of Global Optimization*, vol. 10, n. 1, pp. 57–76.

[34] Dixon, L. C. W. & Szegi, G. P., 1978. *The Global optimization problem: an introduction*.

[35] Bjorkman, M. & Holmström, K., 1999. Global optimization using the direct algorithm in matlab. *Advanced Modeling and Optimization*, vol. 1, pp. 17–28.

[36] Gutmann, H.-M., 2001. A radial basis function method for global optimization. *Journal of Global Optimization*, vol. 19, pp. 201–227.