Felipe Carraro

A stochastic Kriging approach for the minimization of integrals

Florianópolis

2017

Felipe Carraro

A stochastic Kriging approach for the minimization of integrals

Federal University of Santa Catarina

Department of Civil Engineering

Graduation Program in Civil Engineering

Advisor: Prof. Rafael Holdorf Lopez

Co-advisor: Prof. Leandro Fleck Fadel Miguel

Florianópolis

2017

Felipe Carraro

A stochastic Kriging approach for the minimization of integrals

Esta Dissertação foi julgada adequada para a obtenção do Título de MESTRE em Engenharia Civil e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Civil - PPGEC da Universidade Federal de Santa Catarina.

Florianópolis, 05 de outubro de 2017:

Glicério Trichês, Dr.
Coordenador PPGEC

Comissão Examinadora:

Prof. Rafael Holdorf Lopez, Dr. (Orientador)
Universidade Federal de Santa Catarina

Prof. Breno Pinheiro Jacob, Dr. (Videoconferência)
Universidade Federal do Rio de Janeiro

Prof. Marcelo Krajnc Alves, Ph.D.
Universidade Federal de Santa Catarina

Prof. Wellison José de Santana Gomes, Dr.
Universidade Federal de Santa Catarina

Resumo

Este estudo tem como objetivo propor um método eficiente baseado na metodologia *Stochastic Kriging (SK)* para a solução de problemas de minimização de integrais. O modelo *SK* é usado para criar uma aproximação rápida para a função a ser minimizada. Além disso, estimativas de variância da simulação de Monte Carlo são usadas para auxiliar a otimização. O procedimento de otimização aplica o algoritmo *Efficient Global Optimization* com o critério de preenchimento *Augmented Expected Improvement*. Observa-se que o alvo de variância influencia os resultados da otimização. Um alvo muito baixo faz com a otimização torne-se demasiadamente custosa, enquanto que um alvo muito alto pode estagnar a otimização. Desta maneira, uma seleção adaptativa do alvo de variância é proposta. De modo a verificar o desempenho do método proposto vários testes comparativos são conduzidos. O método é aplicado para diversas funções de referência da literatura submetidas a um ruído estocástico. Além disso, comparações são realizadas em relação a um algoritmo eficiente conhecido e ao uso de quadraturas para a avaliação da integral. Por fim, a abordagem proposta é também aplicada a um problema de engenharia estrutural. Os resultados destacam o desempenho eficiente do método bem como a sua consistência ao longo de diversas execuções independentes.

Palavras-chave: stochastic kriging. otimização global. minimização de integrais.

Resumo expandido

Introdução

Atualmente, muitas áreas da engenharia como civil, mecânica, naval e aeroespacial usam técnicas de otimização como uma ferramenta para resolver problemas reais. Apesar do aumento da capacidade computacional disponível, os ambientes de engenharia continuam a evoluir a analisar sistemas mais complexos, que podem incluir mais detalhes, por exemplo, modelos mais refinados. A análise desse tipo de problema, em termos de otimização, pode tornar-se inviável uma vez que uma única análise do modelo pode ser bastante custosa computacionalmente. Pode ser necessário a avaliação das incertezas inerentes ao problema estudado. Neste caso, há ainda um custo adicional relacionado ao fato do modelo analisado ser estocástico.

Objetivos

Este estudo tem como objetivo propor um método eficiente baseado na metodologia *Stochastic Kriging (SK)* para a solução de problemas de minimização de integrais. Mais especificamente busca-se estudar técnicas de metamodelagem associadas a simulações de Monte Carlo. A partir destas técnicas, objetiva-se a implementação de algoritmos computacionais bem como a realização de testes para a verificação do desempenho obtido. Uma das metas deste estudo também é propor uma maneira de incluir o erro cometido pela simulação de Monte Carlo na construção do metamodelo. Mais ainda, aplicar o método proposto a um problema de engenharia, como por exemplo o de controle ótimo da resposta dinâmica de estruturas sujeitas a carregamentos transientes. Por fim, busca-se verificar a eficiência dos métodos estudados.

Metodologia

O modelo *SK* é usado para criar uma aproximação rápida para a função a ser minimizada. Além disso, estimativas de variância da simulação de Monte Carlo são usadas para auxiliar a otimização. O procedimento de otimização aplica o algoritmo *Efficient Global Optimization* com o critério de preenchimento *Augmented Expected Improvement*. No entanto, este critério é modificado para que seja possível a inserção de pontos no modelo utilizando-se alvos de variância diversos. Mais ainda, de forma a balancear as características de exploração e refinamento da resposta, o alvo é alterado pelo uso de uma função de decaimento exponencial. Esta função realiza a adaptação do alvo conforme as características do ponto que deve ser adicionado ao metamodelo.

De modo a verificar o desempenho do método proposto, vários testes comparativos são conduzidos. Inicialmente, o método é aplicado para quatro funções de referência da literatura submetidas a um ruído estocástico. As funções possuem 1, 2, 6 e 10 dimensões. O ruído estocástico é aplicado multiplicativamente a função determinística por meio de uma variável aleatória normal com média unitária e desvio padrão arbitrado para cada problema.

Além disso, comparações são realizadas entre o método proposto e um algoritmo eficiente conhecido, chamado *Globalized Bounded Nelder–Mead (GBNM)*. Ainda, verifica-se existe vantagem em realizar a avaliação da integral que descreve o problema de otimização utilizando-se quadraturas ao invés da técnica de simulação de Monte Carlo.

Por fim, a abordagem proposta é também aplicada a um problema de engenharia estrutural. Neste problema o objetivo é minimizar o valor esperado da probabilidade de falha de um pórtico plano de 10 pavimentos sujeito a ação sísmica. Uma forma de aleatoriedade do problema advém do sismo, que é simulado utilizando-se um processo estocástico filtrado pelo espectro Kanai-Tajimi. A outra, vem incorporada nos parâmetros estruturais de massa, amortecimento e rigidez. Estes são descritos por variáveis aleatórias com uma média fixa e um certo coeficiente de

variação. A solução do problema se dá com uso do método proposto sendo que a parte dinâmica do problema é solucionada a partir da equação de Lyapunov. Esta equação descreve o problema em sua formulação de espaço de estado.

Resultados e discussão

Observa-se que o alvo de variância influencia os resultados da otimização. Um alvo muito baixo faz com a otimização torne-se demasiadamente custosa, enquanto que um alvo muito alto pode estagnar a otimização. É esta situação que motiva o uso de um esquema adaptativo para a seleção do alvo de variância. A partir dos resultados da otimização dos problemas referência da literatura, observam-se reduções no custo computacional e na variabilidade dos resultados quando da utilização do alvo adaptativo em relação ao constante. O uso de quadraturas, em sua forma mais simples com a utilização do produto de regras unidimensionais, acresce demasiadamente o número de avaliações para se chegar a uma resposta similar ao obtido pelo método proposto. Com relação ao algoritmo *GBNM* utilizado para comparação, o método proposto se mostrou mais eficiente e mesmo alterando-se os parâmetros internos daquele algoritimo não foi possível obter um número de avaliações equivalente para a partir do problema de duas dimensões. Os resultados destacam o desempenho eficiente do método bem como a sua consistência ao longo de diversas execuções independentes.

Considerações finais

De maneira geral pode-se dizer que os objetivos do estudo foram atingidos. Conseguiu-se propor um método que incorpora o erro de Monte Carlo na criação do metamodelo. Além disso, utilizando-se o algoritmo EGO foi possível realizar o refino iterativo do metamodelo objetivando a localização do ótimo global. Questões como a parada inesperada do algoritmo ou o alto número de avalições que estavam relacionadas ao

valor do alvo de variância foram solucionados por meio do esquema adaptativo. O método proposto obteve uma maior eficiência com relação ao número de chamadas da função objetivo além de consistentemente chegar mais próximo ao ótimo global.

Palavras-chave: stochastic kriging. otimização global. minimização de integrais.

Abstract

This study aims at proposing an efficient method based on the Stochastic Kriging (SK) methodology for the solution of integral minimization problems. The SK metamodel is used to create a fast approximation for the function being minimized. Moreover, variance estimates from Monte Carlo simulation are used to aid the optimization. The minimization procedure employed the Efficient Global Optimization algorithm with the Augmented Expected Improvement infill criterion. It can be observed that the target variance influences the optimization results. Setting it too low causes the optimization to becomes too costly, while setting it too high might stall the optimization. Therefore, an adaptive target setting is proposed. In order to verify the performance of the proposed method multiple benchmarks are conducted. The method is applied to a number of noisy benchmark functions from literature. Moreover, comparisons are made against a known efficient optimization algorithm as well as an integral evaluation approach using quadratures. At last, the proposed approach is also applied to a structural engineering problem. The results highlight the efficient performance of the method as well as its consistency over multiple independent runs.

Keywords: stochastic kriging. global optimization. integral minimization.

Contents

List of Figures

List of Tables

1 Introduction

Optimization can be seen as the search for the best outcome of a given process while satisfying certain restrictions. In engineering, its use for design and planning is extensive since long back. However, with the increasingly availability of computing power it has gained significant popularity. In recent times, many engineering fields such as civil, mechanical, naval and aerospace use it as a tool to solve real-life problems.

Although the increase in available computing power, the current engineering environment continues to evolve and the analysis of complex systems may now include more detail, *i.e.*, model refinements. When dealing with large or complex models such as Finite Element Analysis (FEA), illustrated in Figure 1, or Computational Fluid Dynamics (CFD), each analysis can become computationally costly.
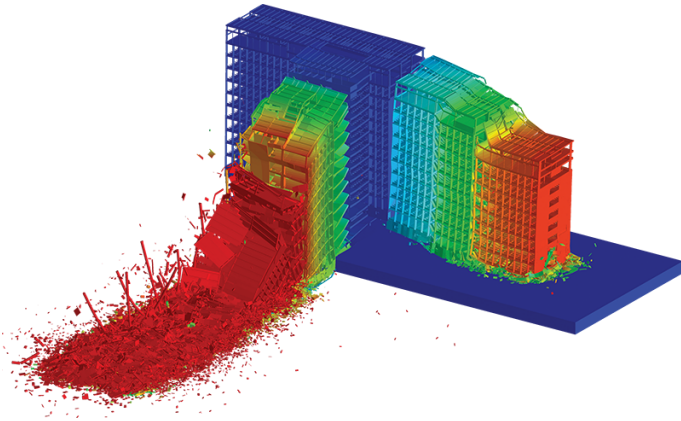
Figure 1 – FEA of a structural collapse simulation

Source – Applied Science International, LLC (2016)

Thus, refining the mathematical/mechanical model leads to a more expensive problem to optimize. However, this is not the only

source of added complexity. For a long time in engineering, researchers have focused on improving structural models. One could analyze, for example, a beam considering a higher order beam model, plasticity, damage theories, and other sources of non- linearity, and approximate the solution using a state-of-the-art finite element model. All of this procedure would still be a rough representation of reality if the intrinsic randomness of materials (rock, soil, concrete) and loads (wind, earthquake motion) were disregarded and a deterministic average was used (SUDRET; KIUREGHIAN, 2000).

Impact Risk
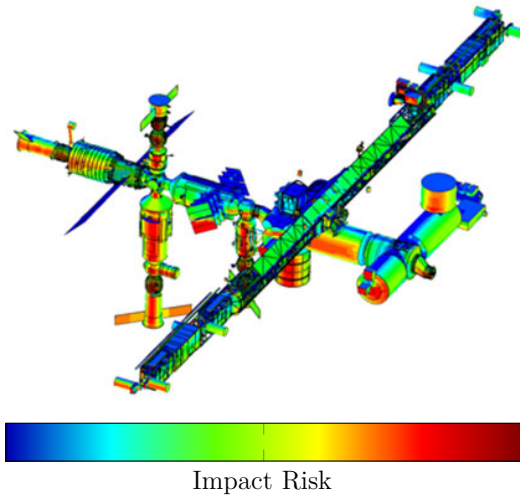
Figure 2 – International Space Station Risk of Impact from Orbital Debris

Source – NASA (2014b)

Accounting for the randomness present in the model makes it possible to achieve a reliable design. For example, Figure 2 shows a model of the International Space Station (ISS) with the associated risk of impact by orbital debris. It cannot be known for certain how particles free in space will behave, but by using probabilistic models it is possible to determine which parts of the spacecraft are more vulnerable to this source of impact. With this information designers may use more

protective shielding in those regions with larger risk and reduce the cost of the structure with less protection in non-critical or improbable to be damaged parts. This type of optimization aims to effectively reduce the probability of failure of the structure and also generate economical savings. This is especially important considering a high investment like the ISS, where estimated construction and operation costs reach almost 75 billion dollars (NASA, 2014a).

Orbital debris may seem a bit distant from an engineer's daily life but the same concept may be applied for other sources of unpredictable loads such as: wind, earthquakes, action of waves, ice, *etc.* In Figure 3 two examples of failure events associated with natural uncertain phenomena are shown. Thus, cost and reliability are two important features in any project and the trade-off between them should be considered.

When trying to optimize large or complex systems including also the system's reliability, the amount of computational resources required may render the task intractable. An efficient optimization algorithm must be able to find the best or at least a reasonably good result under limited time or computational budget.

Classic optimization methods make use of gradients to try to find extreme values of an objective function. However, considering the FEA example, there is usually no explicit function to be used, as the output comes from a computer simulation. Thus, obtaining the derivatives may be either impossible or very costly to approximate. Nevertheless, derivative-free algorithms exist, *e.g.* the Nelder-Mead (NELDER; MEAD, 1965; NHAMAGE *et al.*, 2014). Even so, both types of algorithms may not be successful in finding the global optimum of the usually highly non-linear, multimodal and non-convex problems that arise in engineering. As a result of such algorithms becoming trapped at local optimum, the outcome becomes dependent on the starting point.

Metaheuristics are another class of algorithms, widely used in engineering optimization. They can be used to solve black-box functions as they make very few assumptions about function characteristics and usually develop their search procedure based solely on evaluated responses.

(a) Wind turbine collapse failure under
    extreme wind

(b) Building collapse under earthquake

Figure 3 – Structural failure

Source – Brome (2010) and Takats (2005)

On the other hand, their usefulness is diminished when considering
expensive functions. Usually a large number of function calls are needed
in order to obtain convergence, leading to a prohibitive execution time
(HUANG *et al.*, 2006).

       Beyond both of these approaches, there is an alternative method
that can cope with the limitations discussed: metamodels. Metamodels
are simpler approximate models, created by adjusting a response surface
based on a sample of simulated points from the original model. By being
simple, the metamodel can be used as a replacement of the original
expensive function, reducing the computational burden of performing
numerous simulations. A widely applied metamodel to deal with the
type of problem discussed above is called Kriging. A prediction example

using Kriging is shown in Figure 4.
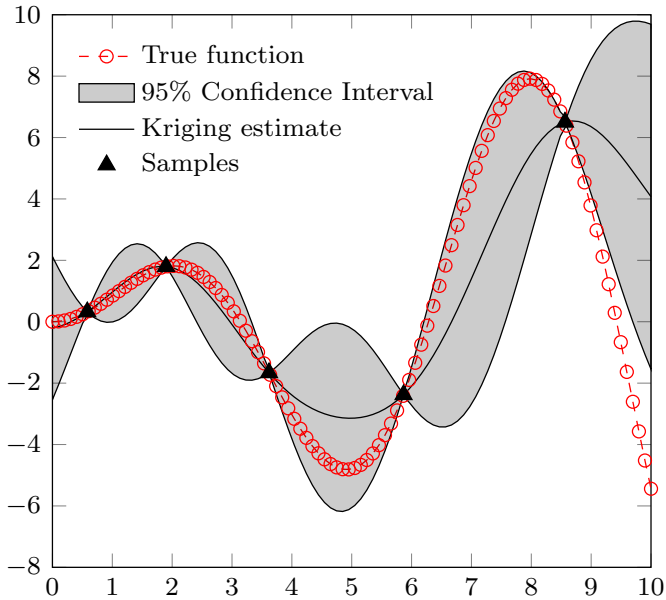
Figure 4 – Example of Kriging metamodel prediction

Kriging in its usual formulation considers the approximation of deterministic functions. But what if the objective function possesses randomness? This randomness could result from, for example, uncertainty in the input parameters or noise on the function response. In such cases, a more recent extension called Stochastic Kriging may be employed. Under this approach, the minimization problem is written as the minimization of an integral. This integral form defines a class of problems which are the focus of this study. This form arises naturally in some engineering cases. Yet, some problems may be rewritten with this integral form. Stochastic Kriging formulation, its characteristics and how it can be employed in order to solve the class of problems discussed, are addressed in this study.

1.1   Objectives

1.1.1   General Objective

Develop a method based on Stochastic Kriging for the minimization of expensive to evaluate functions that depend on an integral.

1.1.2   Specific Objectives

- Study the use of kriging metamodeling techniques coupled with Monte Carlo simulation;

- Implement computational algorithms and test solution approaches found in literature;

- Propose a way to include the error committed with Monte Carlo simulation in the metamodel construction;

- Apply the metamodeling framework to a practical engineering problem, *e.g.*, optimal control of dynamic response under transient loads;

- Verify the overall efficiency of the studied methods.

1.2   Dissertation structure

This dissertation is structured in chapters. The initial chapters aim to constructively gather the knowledge for the understanding and development of the proposed method as well as its application. With the theoretical basis well established, the final chapters present numerical experiments with the proposed method, results are discussed and conclusion are drawn. A brief overview of each chapter is presented below.

In Chapter 2 the theory and formulation of general optimization problems are presented. In addition, problem categories and optimization algorithms are also discussed.

Chapter 3 concisely describes the probabilistic background needed for the characterization of random parameters. Moreover, it details how uncertainties are considered in optimization as well as the formulation of problems involving random variables.

Chapter 4 explains the general metamodeling concept and gives examples of existent methods. It then, focuses in discussing the approach used in this study: Kriging. The formulation of the Kriging predictor and the process of parameter estimation is presented. The coupling of optimization and metamodeling is reviewed under the Efficient Global Optimization method.

In Chapter 5 resides the main contribution of this study. It introduces the concepts for dealing with stochastic problems. Moreover, it employs the Stochastic Kriging framework with proposed modifications to solve a class of integral minimization problems.

In Chapter 6 a few benchmark problems are solved in order to show the metamodeling technique efficiency as well as to demonstrate the proper functioning of the implemented algorithms.

In Chapter 7 conclusions are drawn in respect of the results obtained in the benchmarks as well as the overall performance of the proposed method. Additionally, it presents suggestions regarding further studies related to the studied topic.

2 Optimization

Optimization might be defined as the science of determining the "best" solutions to certain mathematically defined problems, which are often models of physical reality (FLETCHER, 1987). Its practical use begins with the definition of at least one objective function, which represents a measure of performance. This function depends on certain characteristics of the system, called design variables. Thus, the goal of optimization is to find the design variables that return the best value for the objective function. In an engineering context, an example could be that of optimizing the sections of a planar steel frame (CARRARO *et al.*, 2016). A simple example of a frame that could be optimized is shown in Figure 5.
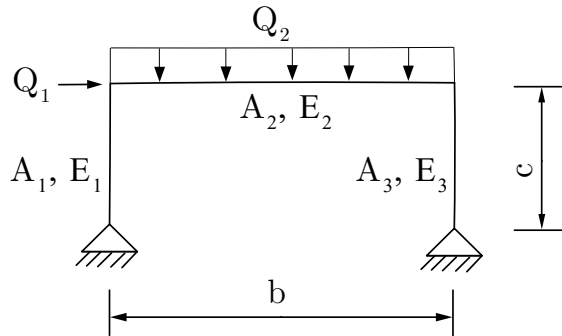
Figure 5 – Example of a simple plane frame

Three elements characterize an optimization problem:

1. Objective function: this function associates the system parameters and measures a certain performance value. In the steel frame example, an objective function could be the total volume or weight of the structure or the monetary cost to build it. In Figure 5, the objective could be to find A_1, A_2 and A_3 which minimize the

volume of the structure. Thus, the objective function would be written as $f_{obj} = Vol = A_1c + A_2b + A_3c$, where b and c are the lengths of beam and columns, respectively.

2. Design variables: these are the input parameters which modify the system response and can be selected in order to improve the objective function value. In the example, the design variables could be the steel profiles associated with groups of beams or columns. These profiles would be chosen from a catalog, *i.e.* from a list of commercially available profiles. When design variables have to be chosen based on discrete set, the problem is classified as a Discrete Optimization (PAPADIMITRIOU; STEIGLITZ, 1982; LEE, 2004). On the other hand, the design variables could be the cross-sectional area of the structural elements. In such situation, the variable may be represented by an infinite amount of positive values. When design variables are real or defined over a real range, they are called continuous and the problem is a Continuous Optimization (LUENBERGER, 1969). From Figure 5, A_1, A_2 and A_3 could be continuous design variables to be optimized. Additionally, there are problems posed using both continuous and discrete variables. Such problems are called Mixed Variable Optimization.

3. Search space: this is the space that contains all the possible inputs to the objective function. In a steel frame optimization, the search space could be the W-shaped profiles from a standard specification.

2.1   Formulation of an optimization problem

Transcription of an optimization problem into a mathematical formulation is a critical step. If the problem formulation is improper, the solution for the problem is most likely going to be unacceptable (ARORA, 2007). The usual formulation of an optimization problem is as follows:

Find a vector of design variables $\mathbf{d} \in \mathbb{R}^n$:

$$\mathbf{d} = \{d_1, d_2, \ldots, d_n\}, \tag{2.1}$$

which minimize the objective function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$:

$$f_{obj} = f(\mathbf{d}, \mathbf{x}), \tag{2.2}$$

with parameters:

$$\mathbf{x} = \{x_1, x_2, \ldots, x_m\}, \tag{2.3}$$

subject to the constraints:

$$g_i(\mathbf{d}, \mathbf{x}) \leq 0; \quad i = 1, \ldots, n_{ic}, \tag{2.4}$$

$$h_j(\mathbf{d}, \mathbf{x}) = 0; \quad j = 1, \ldots, n_{ec}, \tag{2.5}$$

where $g_i(\mathbf{d}, \mathbf{x})$ and $h_j(\mathbf{d}, \mathbf{x})$ represent the functions that establish inequality and equality constraints, respectively, while n_{ic} and n_{ec} represent the number of such functions.

In the frame example parameters could be, for instance, the steel elastic modulus (E_1, E_2, E_3) or the external loads applied (Q_1, Q_2). Constraints on the objective function could come from stress restrictions on each member, based on a design code, or a limited lateral displacement. They could also be imposed on the design variables, for example, by limiting the cross-sectional area to a certain range, *e.g.* $1 \leq A_i \leq 5$ cm^2 for i={1,2,3} in Figure 5.

When seeking to minimize a function it may occur that there are no imposed restrictions. In this case, the problem is classified as an Unconstrained Optimization problem. Otherwise, if restrictions exist on the objective function or on the design variables, the problem is from the Constrained Optimization class.

It is assumed that the objective function is to be minimized, but that entails no loss of generality since the minimum of $-f(\mathbf{d}, \mathbf{x})$ occurs where the maximum of $f(\mathbf{d}, \mathbf{x})$ takes place.

Further considerations on the terminology and aspects of this formulation will be addressed in section 2.2.

2.2   Aspects and classification of optimization problems

This section details aspects of the optimization problem that lead to different classification and usually different solution procedures.

2.2.1   Convexity

Convexity in an important characteristic of an optimization problem. Knowing beforehand that a function is convex enables the use of specialized methods that exploit this property. Thus, the search becomes faster if compared to more general methods and the solution is guaranteed to be a global optimum as shall be discussed in subsection 2.2.2.

In order to define a convex function, the definition of a convex set is required. A set $\mathbb{S}$ is said to be convex if given any two points p_1 and p_2 in $\mathbb{S}$, the line segment $\overline{p_1 p_2}$ is also in $\mathbb{S}$. This concept can be extended for the n-dimensional space. Mathematically, a parametric representation of a line segment between points $\mathbf{d}^{(1)}$ and $\mathbf{d}^{(2)}$ can be formulated as follows:

$$\mathbf{d} = \alpha \mathbf{d}^{(2)} + (1 - \alpha)\mathbf{d}^{(1)};\ 0 \le \alpha \le 1. \tag{2.6}$$

If the entire line segment is in $\mathbb{S}$, then it is a convex set (ARORA, 2004).

Considering now a function $f(\mathbf{d})$ defined on a convex set $\mathbb{S}$, this function is said convex if it satisfies:

$$f(\alpha \mathbf{d}^{(2)} + (1 - \alpha)\mathbf{d}^{(1)}) \le \alpha f(\mathbf{d}^{(2)}) + (1 - \alpha)f(\mathbf{d}^{(1)});\ 0 \le \alpha \le 1. \tag{2.7}$$

This condition is necessary and sufficient and applies to n-dimensional functions. An illustrative example of a single variable convex function is shown in Figure 6.

In practice, applying Equation 2.7 is difficult, as an infinite number of pair of points must be checked. Checking the Hessian of the
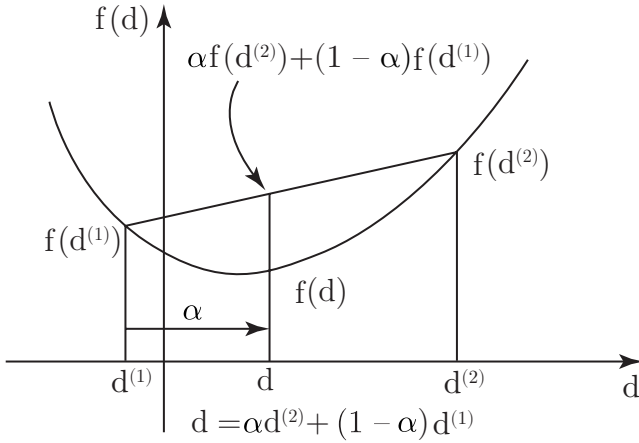
Figure 6 – Convexity check on unidimensional function

Source – Adapted from Arora (2004)

function is a simpler alternative. A function is said to be convex when its Hessian $\nabla^2 f(\mathbf{d})$, defined as:

$$
\nabla^2 f(\mathbf{d}) := \begin{bmatrix}
\dfrac{\partial^2 f(\mathbf{d})}{\partial d_1^2} & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_1 \partial d_2} & \cdots & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_1 \partial d_n} \\[2ex]
\dfrac{\partial^2 f(\mathbf{d})}{\partial d_2 \partial d_1} & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_2^2} & \cdots & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_2 \partial d_n} \\[2ex]
\vdots & \vdots & \ddots & \vdots \\[2ex]
\dfrac{\partial^2 f(\mathbf{d})}{\partial d_n \partial d_1} & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_n \partial d_2} & \cdots & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_n^2}
\end{bmatrix}, \quad (2.8)
$$

is at least positive semidefinite everywhere, that is, has non-negative

eigenvalues for all points in $\mathbb{S}$.

Thus, a Convex Optimization problem is one where $\mathbb{S}$, f and the inequality constraints g_i are convex and the equality constraints h_j are linear. This type of problem has an important characteristic that will be discussed in subsection 2.2.2.

2.2.2   Local or global optima

The distinction between global and local optima is relevant when considering non-convex functions that arise in practical engineering optimization. A local optimum is defined as a feasible point $\hat{\mathbf{d}}$ from the search space $\mathbb{S}$ such that sufficiently small neighborhoods surrounding $\hat{\mathbf{d}}$ contain no points that are both feasible and improving in objective function value (ZABINSKY, 2003).

When a point has the best objective value over all the search space, then it is referred to as global optimum. Figure 7 illustrates this concept using a one-dimensional function. This function is multimodal, that is, presents multiple local optima.
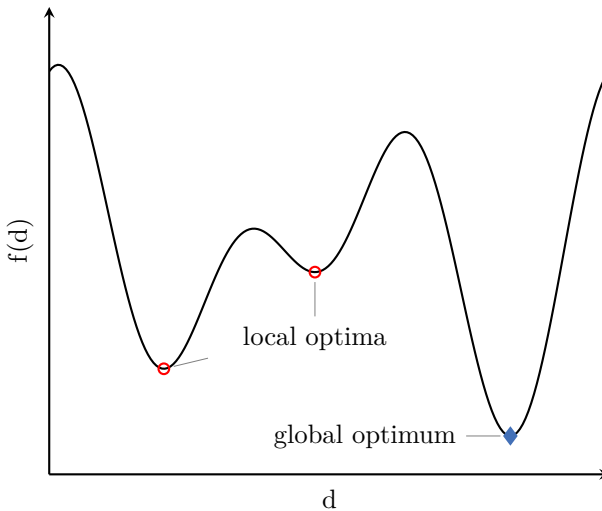
Figure 7 – Local and global minima

Mathematically, considering a minimization problem, it can be stated:

A point $\hat{\mathbf{d}} \in \mathbb{S}$ is a local optimum (minimum) of f over $\mathbb{S}$ if

$$\exists \delta > 0 : f(\hat{\mathbf{d}}) \leq f(\mathbf{d}) \text{ and } ||\mathbf{d} - \hat{\mathbf{d}}|| < \delta, \ \forall \mathbf{d} \in \mathbb{S}, \qquad (2.9)$$

and also:

A point $\hat{\mathbf{d}} \in \mathbb{S}$ is a global optimum (minimum) of f over $\mathbb{S}$ if

$$f(\hat{\mathbf{d}}) \leq f(\mathbf{d}) \ \forall \mathbf{d} \in \mathbb{S}. \qquad (2.10)$$

When trying to solve an optimization problem, one is usually interested in finding the global optimum, *i.e.* the best possible result. Often, a question arises if the solution found is indeed a global minimum. Considering Convex Optimization problems, a fundamental property is that any locally optimal point is also globally optimal. This property makes Convex Optimization problems somewhat easier to analyze, as there is a guarantee of global optimality (LUENBERGER, 1969). The other option to ensure global optimality, which is usually intractable, is an exhaustive search on the search space. This means that unless the problem can be shown as convex, there is no way to recognize if the solution found can be further improved.

Several functions in engineering optimization are non-convex and thus require special care when using optimization algorithms in order to avoid being trapped in local solutions. This distinctive behavior among algorithms is further discussed in section 2.3.

2.3 Aspects and classification of optimization algorithms

2.3.1 Algorithmic progression

Depending on the existence of random components in the optimization procedure, the algorithmic progression behaves differently. The algorithms can be classified in this regard as deterministic or stochastic:

- Deterministic: Classical mathematical optimization algorithms are usually deterministic. They iteratively improve the solution according to some deterministic rule;

- Stochastic: Algorithms of this category possesses a random component in its formulation. Contrary to deterministic algorithms, running the same optimization process multiple times may lead to different results.

2.3.2   Local Search

Local Search methods are the ones that converge to a local optimum. Most mathematical optimization algorithms from linear and non-linear programming fall in this category. They usually focus on identifying and iteratively following a descent direction. To accomplish this, they make use of the information from function evaluations, gradients and/or the Hessian. Some examples include (NOCEDAL; WRIGHT, 2006):

- Line Searches;

- Steepest Descent;

- Conjugate Gradients;

- Simplex Methods;

- Newton Methods;

- Quasi-Newton Methods;

- Sequential Quadratic Programming (SQP);

- Interior Point Methods.

2.3.3 Global optimization

Global optimization is distinguished from local optimization by its focus on finding the maximum or minimum over all the search space. When the objective function is convex, employing a local search such as those cited in subsection 2.3.2 yields the global optimum as discussed earlier. Moreover, this class of algorithms usually converges very fast.

On the other hand, when the function is non-convex, the solutions from this search procedure may be local optima. The final solution becomes dependent on the point where the search started. In order to address this issue and still use a local search for its fast convergence a restart procedure may be employed. The idea is to assume that the global optimum has been found when a minimum value is achieved from multiple runs and starting points (MUSELLI, 1997; TORII *et al.*, 2011).

Another approach to solve this type of problem is the use of metaheuristics. Metaheuristics are a class of algorithms well suited to search for both local and global optima. They are mostly stochastic and usually do not use gradients information. They also do not require the function to be continuous or differentiable. A solution is found by applying a set of rules and randomness. Nevertheless, convergence proof only exists in probabilistic terms, *i.e.*, as the number of iterations tends to infinity the minimum found so far tends to the global minimum. They also are usually computationally costlier, *i.e.*, need a larger number of function evaluations to converge, if compared to local searches or the methods that will be discussed in chapter 4 (YOUNIS; DONG, 2010). Still, reasonably good results can be achieved given enough time to run. Some algorithms include:

- Genetic Algorithms (GA) (GOLDBERG, 1989);

- Simulated Annealing (SA) (KIRKPATRICK *et al.*, 1983);

- Particle Swarm Optimization (PSO) (KENNEDY; EBERHART, 1995);

- Ant Colony Optimization (ACO) (COLORNI *et al.*, 1992);

- Search Group Algorithm (SGA) (GONÇALVES *et al.*, 2015);

- Imperialist Competitive Algorithm (ICA) (ATASHPAZ-GARGARI; LUCAS, 2007; CARLON *et al.*, 2015)

- Backtracking Search Algorithm (BSA) (CIVICIOGLU, 2013);

- Probabilistic Restart (PR) (LUERSEN; RICHE, 2004).

In this study, optimization takes place considering a non-convex, multimodal and expensive to evaluate black box function. By black box, it is meant that no information, other than the function response is available. This, therefore, rules out various local search techniques. Being multimodal also makes the use of local search improper. Metaheuristics could be a useful procedure, but given the fact that the function is expensive to evaluate, the process might become too time-consuming. A different approach is needed for solving this type of problem. The approach considered in this study is metamodeling.

The metamodeling approach consists on developing fast surrogate models for the objective and constraint functions. The models offer a cheap to compute transfer function between input and output and no gradient information is required. They are especially useful for complex, time-consuming simulations, and might be used in an optimization context. The structure of the model and its formulation may be exploited focusing specifically the search for the global optimum. It makes this approach the best candidate to solve the type of problem proposed. Further considerations on surrogate modeling and the use of this methodology for global optimization will be discussed in chapter 4.

3 Probability theory and uncertainties

For now, all components of the optimization problem were taken as deterministic. This does not map well with reality, due to the existence of uncertainties.

In the next few sections, concepts of probability, stochastic processes and reliability will be reviewed. Later in this chapter, the optimization problem will be reformulated by considering those uncertainties, under different approaches.

3.1 Initial probability concepts

3.1.1 Events

An event E is defined as a subset of the sample space Ω. The sample space contains all possible outcomes of a random quantity. The failure event E of a structural element, for example, may be modeled by the event:

$$E = \mathrm{R} - \mathrm{S} < 0, \tag{3.1}$$

where R represents resistance and S represents solicitation.

Considering the frame example from chapter 2, another event could be one where the elastic modulus value E_1, of the first steel element, is between 205 and 210 GPa. For every event considered, a probability of occurrence can be associated. This will be discussed in subsection 3.1.2.

3.1.2 Axiomatic probability measure

For the classical probability definition, it is necessary to understand the mathematical concept of σ-algebra. A σ-algebra is a subset

from a set which has some special properties such as having an empty element and being closed under union and complementation. With such structure, it is possible to represent events and a probability measure. Given a sample space Ω and an associated σ-algebra $\mathcal{F}$. The probability measure is a function $P : \mathcal{F} \to [0, 1]$, which follows the Kolmogorov axioms (KOLMOGOROV, 1950; SUDRET, 2007):

$$P(A) \geq 0, \ \forall A \in \mathcal{F} \tag{3.2}$$

$$P(\Omega) = 1 \tag{3.3}$$

$$P(A \cup B) = P(A) + P(B), \ \forall A, B \in \mathcal{F}, \ A \cap B = \varnothing. \tag{3.4}$$

The probability measures how probable an event occurrence is. An impossible event will have a probability of zero, while an event that always occurs will have unit probability. From these axioms, it is also possible to compute the probability of unions and/or intersections of multiple events.

3.1.3   Frequentist interpretation of probability

The axiomatic construction proposed in subsection 3.1.2 is purely mathematical. On the other hand, when considering real world experiments, a more practical view relates to the evaluation of scenarios. Under the frequentist interpretation of probability theory, the probability of an event is the limit of its empirical frequency. This frequency is calculated as the number of occurrences (n_A) of an event A divided by the number of trials n:

$$\text{Freq}(A) = \frac{n_A}{n}; \tag{3.5}$$

Thus, the probability of event A becomes:

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}. \tag{3.6}$$

3.1.4   Random variables

Random variables are a useful tool to characterize quantities subjected to random variations. Throughout the text, they will be denoted by capital letters. A real-valued random variable $X(\omega)$ for $\omega \in \Omega$ is defined as a mapping $X : \Omega \to \mathbb{R}$. In other words, it is a function that attributes to every sample point ω from the sample space Ω a real value. The sample space Ω can be finite or countable infinite, which results in a discrete random variable. When Ω possess an uncountable number of elements, the resulting random variable is called continuous.

A simple illustrative example of a random variable could be the model of fair coin toss. Calling it C, the random variable has two states:

$$C = \begin{cases} 1, & \text{if the outcome is heads;} \\ 0, & \text{if the outcome is tails.} \end{cases} \tag{3.7}$$

Clearly, the variable is discrete as $\Omega = \{0, 1\}$ and the event of a heads outcome can be written with the following notation:

$$P\big(\{\omega \in \Omega \mid C(\omega) = 1\}\big) \equiv P(C = 1) = \frac{1}{2}. \tag{3.8}$$

Note that the numbers 0 and 1 are artificially assigned numerical values and therefore, other values could have been associated with the events in question. In this way, it is possible to identify possible outcomes of a random phenomenon by numerical values. In most cases these values will simply be the outcomes of the phenomenon (THOFT-CHRISTENSEN; BAKER, 1982).

3.2   Density Functions

3.2.1   Cumulative Density Function

The probabilistic characteristics of a random variable can be fully described by its Cumulative Density Function (CDF). The CDF, or distribution function of X, evaluated at a real number x, is the

probability that the random variable X will take a value less than or equal to x. That is:

$$F_{\mathrm{X}}(\mathrm{x}) := \mathrm{P}(\mathrm{X} \le \mathrm{x}), \ \mathrm{x} \in \mathbb{R}. \tag{3.9}$$

3.2.2   Probability Density Function

It also is useful to write the probability density function (PDF). For real continuous variables it represents the probability of the random variable falling in an interval $[\mathrm{x}, \mathrm{x} + \mathrm{dx}]$. It may be computed by taking the derivative of the CDF:

$$f_{\mathrm{X}}(\mathrm{x}) := \frac{\mathrm{d}F_{\mathrm{X}}(\mathrm{x})}{\mathrm{dx}}, \ \mathrm{x} \in \mathbb{R}. \tag{3.10}$$

Figure 8 illustrates how an event probability can be computed from the PDF. Given a certain event $E = \{\mathrm{a} \le \mathrm{X} \le \mathrm{b}\} \subset \Omega$, its probability is given by:

$$\mathrm{P}(E) = \int_{\mathrm{a}}^{\mathrm{b}} f_{\mathrm{X}}(\mathrm{x})\mathrm{dx}, \tag{3.11}$$

and is represented by the hatched area in Figure 8.
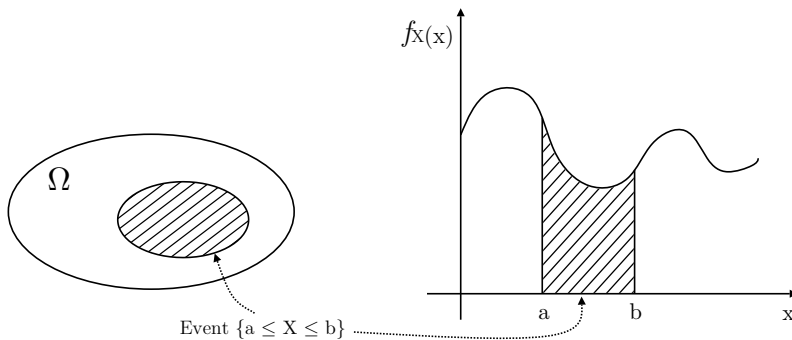
Figure 8 – Illustration of a PDF

3.3   Moments

In a mathematical context moments have the purpose of characterizing the shape from a sample of points. Two important moments are the mean and variance (MONTGOMERY; RUNGER, 2010).

3.3.1   Mean

The mean μ_X is the expected value of the random variable X:

$$\mu_X := \mathrm{E}\big[X\big] = \int_{-\infty}^{+\infty} x \, f_X(x) \, dx, \tag{3.12}$$

where E[.] is the mathematical expectancy operator.

Analogously, when considering functions of random variables, the expected value of a function q(X) is defined as:

$$\mu_{q(X)} := \mathrm{E}\big[q(X)\big] = \int_{-\infty}^{+\infty} q(x) \, f_X(x) \, dx. \tag{3.13}$$

If the mean of X has to be estimated based on a n-sized finite sample, the following unbiased estimator can be used:

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{3.14}$$

This estimation gives results that are more accurate as n becomes larger, *i.e.*, a bigger sample is evaluated. There is an error (or variance) associated with this estimative. This variance will be estimated by using Equation 3.18.

3.3.2   Variance

The variance, σ^2, is a measure of the random variable dispersion around the mean. For a continuous random variable X it can be calculated as:

$$\sigma_X^2 := \mathrm{E}\big[(X - \mu)^2\big] = \int_{-\infty}^{+\infty} (x - \mu)^2 \, f_X(x) \, dx. \tag{3.15}$$

Another important measure of dispersion, derived from this moment, is the standard deviation or standard error σ. Unlike the variance, it is expressed in the same units as the analyzed data. For random variables, it is computed by taking the square root of the variance:

$$\sigma_X := \sqrt{\sigma_X^2}. \tag{3.16}$$

If the variance of X has to be estimated based on a n-sized finite sample, the following unbiased estimator can be used (DEVORE, 2011):

$$\overline{V}_X := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2. \tag{3.17}$$

Another measure that shall prove useful is the estimated variance of the sample mean. The unbiased estimator can be written as:

$$\overline{V}_{\overline{X}} := \frac{\overline{V}_X}{n}. \tag{3.18}$$

This result shows that as the sample size n increases, the variance of the sample mean decreases.

3.3.3 Covariance and Correlation

Covariance is a measure of the correlation between two or more random variables. For two random variables X_1 and X_2 it can be written:

$$\text{Cov}(X_1, X_2) := E[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})] = E[X_1 X_2] - \mu_{X_1} \mu_{X_2}. \tag{3.19}$$

Additionally, a correlation coefficient between the variables can be calculated. It is defined as a normalized covariance with respect to the standard deviations of X_1 and X_2 and is given by:

$$\text{Cor}(X_1, X_2) := \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}}. \tag{3.20}$$

3.4 Gaussian Distribution

The normal or Gaussian probability distribution (named after Karl F. Gauss [1777–1855]) is widely used due to its simplicity and wide applicability (SCHAY, 2007). This distribution is the basis for many statistical methods, such as the ones that will be studied in chapter 4. A normal random variable is usually denoted by $X \sim \mathcal{N}(\mu, \sigma)$, where μ is the distribution mean and σ its standard deviation.

A standard normal random variable corresponds to $Z \sim \mathcal{N}(0, 1)$, that is, normal distribution with zero mean and unitary standard deviation. The standard normal PDF $\varphi(x)$ is defined by (BERTSEKAS; TSITSIKLIS, 2002):

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{3.21}$$

The standard normal CDF $\Phi(x)$ reads:

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt. \tag{3.22}$$

Moreover, considering the computer implementation, the following relation may be used in order to compute the CDF faster (CODY, 1969):

$$\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \tag{3.23}$$

where erf(.) is the error function, which is often available in programming environments.

One of the useful properties of the normal distribution is its preservation under linear transformation. Thus, the PDF and CDF can be calculated for arbitrary μ and σ using the results from the standard normal distribution:

$$f_X(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right), \tag{3.24}$$

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right). \tag{3.25}$$

In Figure 9 the plots of the PDF and CDF obtained from a Standard ($\mu = 0, \sigma = 1$) Normal Distribution are presented, respectively.

Std. Normal Distribution
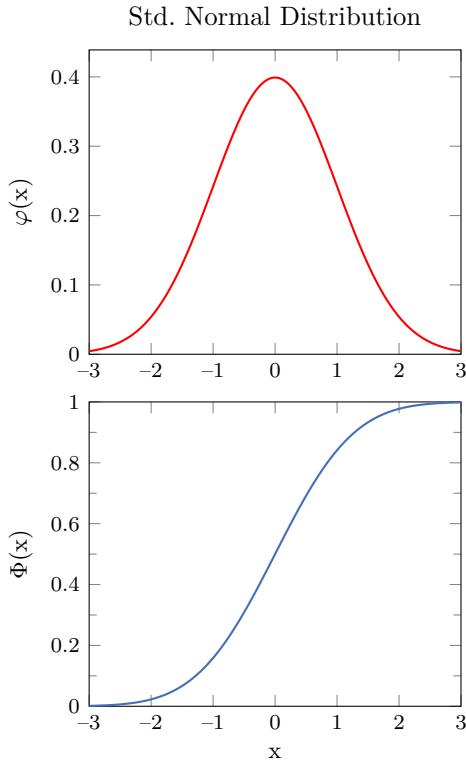
Figure 9 – PDF and CDF from standard normal distribution

3.5   Extension to multiple variables

3.5.1   PDF and CDF

The previous concepts of PDF and CDF might be extended to the case of multiple variables. Considering a random vector $\mathbf{X}$, in which each of its components is a random variable (PAPOULIS; PILLAI,

2002):

$$\mathbf{X} = [X_1, X_2, \ldots, X_n].$$ (3.26)

The probability that $\mathbf{X}$ is in a region $\mathbb{D} \subset \mathbb{R}^n$ of the n-dimensional space equals:

$$P(\mathbf{X} \in \mathbb{D}) = \int_{\mathbb{D}} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x},$$ (3.27)

where

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) := \frac{\partial^n F_{\mathbf{X}}(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n}$$ (3.28)

is the joint probability density function (PDF) of $\mathbf{X}$.

The joint cumulative density function is defined as:

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, x_2, \ldots, x_n) := P\big(X_1 \leq x_1, X_2 \leq x_2 \ldots X_n \leq x_n\big).$$ (3.29)

3.5.2 Expected value of arbitrary functions

Extending the expected value of an arbitrary function (Equation 3.13) to the multivariate case, it is possible to write(PAPOULIS; PILLAI, 2002; BERTSEKAS; TSITSIKLIS, 2002):

$$E[g(\mathbf{X})] :=$$
$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, x_2, \ldots, x_n) f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) \, dx_1 dx_2 \ldots dx_n.$$ (3.30)

Generalizations can be made, if the function is known. For example, if g is a linear function of the form:

$$g(\mathbf{X}) = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$ (3.31)

then, the expected value becomes:

$$E[a_1 X_1 + a_2 X_2 + \cdots + a_n X_n] = a_1 E[X_1] + a_2 E[X_2] + \cdots + a_n E[X_n]. \tag{3.32}$$

3.5.3   Correlation and Covariance

For a random vector $\mathbf{X}$, one can define the correlation and covariance matrices (PAPOULIS; PILLAI, 2002). The correlation matrix $\psi_{\mathbf{X}}$ is defined as:

$$\psi_{\mathbf{X}} := E[\mathbf{X}\mathbf{X}^{\mathrm{T}}] = E \begin{bmatrix} X_1^2 & X_1X_2 & \dots & X_1X_n \\ X_2X_1 & X_2^2 & \dots & X_2X_n \\ \vdots & \vdots & \vdots & \vdots \\ X_nX_1 & X_nX_2 & \dots & X_n^2 \end{bmatrix}$$

$$= \begin{bmatrix} E[X_1^2] & E[X_1X_2] & \dots & E[X_1X_n] \\ E[X_2X_1] & E[X_2^2] & \dots & E[X_2X_n] \\ \vdots & \vdots & \vdots & \vdots \\ E[X_nX_1] & E[X_nX_2] & \dots & E[X_n^2] \end{bmatrix}. \tag{3.33}$$

The covariance matrix represents the covariance between the elements of the vector, which in turn are random variables. It is also called dispersion matrix or variance–covariance matrix. Its definition reads:

$$\Psi_{\mathbf{X}} := \mathrm{E}[(\mathbf{X} - \mathrm{E}[\mathbf{X}])(\mathbf{X} - \mathrm{E}[\mathbf{X}])^{\mathrm{T}}] =$$

$$\begin{bmatrix} \sigma^2(\mathrm{X}_1) & \mathrm{Cov}(\mathrm{X}_1, \mathrm{X}_2) & ... & \mathrm{Cov}(\mathrm{X}_1, \mathrm{X}_n) \\ \mathrm{Cov}(\mathrm{X}_2, \mathrm{X}_1) & \sigma^2(\mathrm{X}_2) & ... & \mathrm{Cov}(\mathrm{X}_2, \mathrm{X}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\mathrm{X}_n, \mathrm{X}_1) & \mathrm{Cov}(\mathrm{X}_n \mathrm{X}_2) & ... & \sigma^2(\mathrm{X}_n) \end{bmatrix} . \qquad (3.34)$$

From Equation 3.20, one can write the following relation:

$$\Psi_{\mathbf{X}} = \sigma_{\mathbf{X}}^2 \psi_{\mathbf{X}}, \qquad (3.35)$$

which will be useful when describing the Kriging metamodel in chapter 4.

3.6   Stochastic Processes

A stochastic or random process, is a collection of random variables or random vectors indexed by a parameter t over the same sample space Ω. The indexed parameter may be time, *e.g.* when dealing with the motion of a structure subjected to seismic loading, or space, *e.g.* when considering the variability of the Young's modulus within a beam element.

Considering a random process $\mathrm{X}(\omega, \mathrm{t})$, for any parameter t_0 fixed, $\mathrm{X}(\omega, \mathrm{t}_0) \equiv \mathrm{X}_{\mathrm{t}_0}(\omega)$ is a random variable with prescribed properties. When the prescribed properties of the random variables do not change over the continuum, the process is said to be stationary. In this notation, ω is an outcome. It is possible to obtain a realization of the stochastic process, *i.e.*, its trajectory over the parameter t, by fixing an outcome ω_{f}. This is shown in Figure 10 for three different ω.

As it will be seen in chapter 4, the Kriging metamodel applies this concept in its formulation. It makes use of a stationary stochastic process to represent data points over its domain. Stochastic processes also arise when considering problems of time-dependent reliability, which
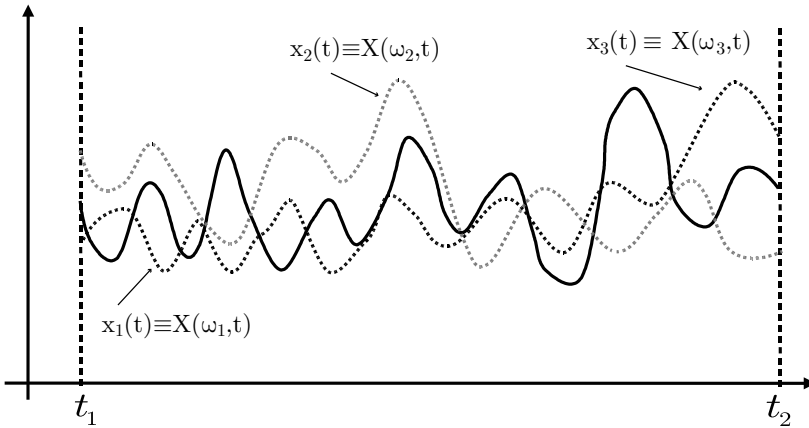
Figure 10 – Realizations from a stochastic process

Source – Adapted from Sudret (2007)

is a topic that will be studied for the proposed application discussed in subsection 3.10.1.

3.7   Structural Reliability

A common basis for the different levels of reliability methods is the introduction of a so-called limit state function G (or failure function, or g-function) which gives a mathematical definition of the failure event in mechanical terms (LEIRA, 2013). It is a function of random variables that determines a performance measure. This function divides the domain in a safe region and a failure region. It is constructed such that $G(\mathbf{x}) \leq 0$ represents failure and $G(\mathbf{x}) > 0$ represents safety or survival. The vector $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ represents realizations of the basic random variables $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ considered.

In a structural sense, the failure may represent any undesired performance such as cracking, corrosion, excessive deformations, exceeding load-carry capacity, local or global buckling, *etc.* (NOWAK; COLLINS, 2012). In the simple steel frame example discussed in chapter 2, failure
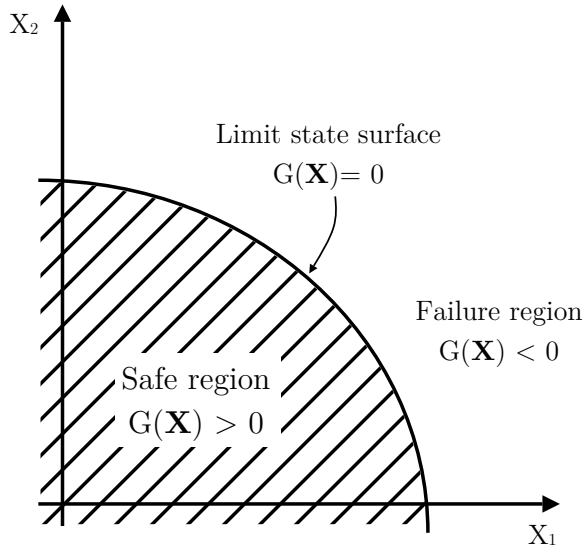
Figure 11 – Limit state representation

could be represented as a lateral displacement larger than a certain limit or a load combination larger than the resistant strength.

An example with two random variables is depicted in Figure 11. In that figure, considering structural reliability, the random variables X_1 and X_2 could be replaced by R (resistance) and S (solicitation), respectively. This two-variable formulation with R and S is known as the fundamental problem of structural reliability (BECK, 2014).

From the construction of $G(\mathbf{X})$ it is possible to calculate the failure probability P_f. The probability of failure represents the probability that the combination of realizations of the random variables lies in the failure region. That is:

$$P_f = P(G(\mathbf{X}) \leq 0), \tag{3.36}$$

or in integral form:

$$P_f = \int \cdots \int_{G(\mathbf{X}) \leq 0} f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x}, \qquad (3.37)$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the joint probability density function of the n-dimensional vector $\mathbf{X}$ of basic variables.

The solution of Equation 3.37 may be difficult. For instance, there may be a large number of variables involved, the limit state function may not be explicit (*i.e.*, it cannot be described by a single equation), or the solution may be impossible to calculate analytically. Several alternative methods have been proposed to solve it. These methods will be categorized and discussed in a uncertainty quantification context in section 3.9.

3.8   Types of uncertainties

Uncertainties can be classified according to their main source (LOPEZ; BECK, 2013):

- Uncertainty of model parameters: this type is related to the natural randomness of a model parameter, for example the uncertainty in the yield stress due to production variability or the wind loading on a structure which is varies in time and space.

- Measurement uncertainty: is the uncertainty caused by imperfect measurements of for example a geometrical quantity;

- Statistical uncertainty: caused by the lack of sufficient information about the observed quantity;

- Model uncertainty: a model is an idealization of a real phenomenon. Imperfect knowledge of the real behavior or uncertainty related on how to represent a given quantity results in this type of uncertainty.

In this study, uncertainties on the model parameters will be taken into account to improve the quality of the approximated response.

3.9   Uncertainty representation and quantification

In order to represent uncertainties, two approaches are often used: the probabilistic and possibilistic approach. The probabilistic approach uses concepts from probability theory and represents uncertainties by using random variables, processes and fields (FELLER, 1968; LOPEZ; BECK, 2013). Thus, the PDF carries all the information relevant to the uncertainty such as the mean, the variance and the median. The PDF, however, may not be completely available. Therefore, possibilistic approach makes use of intervals such as upper and lower bounds in random variables to describe incomplete or imperfect data.

This study will focus on the probabilistic representation. Under this representation, uncertainty quantification methods can be sorted in three categories (LOPEZ; BECK, 2013):

- Analytical methods: Comprehend methods that make use of space transformations, linearization and Taylor's series expansion to seek full PDF characterization, moments or the probability of failure (PARKINSON *et al.*, 1993; SANKARARAMAN, 2015). In Figure 12 a diagram of the distinct information sought by these methods is shown. Full characterization methods aim at obtaining the PDF of the response of the system (LOPEZ *et al.*, 2011a). Additionaly, response variability methods compute statistical moments of the response such as its mean and/or standard deviation (BEYER; SENDHOFF, 2007). Furthermore, reliability methods investigate the probability of failure of the system (LOPEZ *et al.*, 2011b);

- Numerical integration: Probabilistic characteristics of the random response of the system are evaluated using multidimensional numerical integration. This procedure is feasible for a limited number of variables. For high-dimensional problems, say more than 10 variables, the computational time increases considerably. Therefore, in most applications the numerical integration procedure is

applied only for validation of other methods with a small number of variables (VROUWENVELDER; KARADENIZ, 2010);

- Simulation methods: Uses multiple samples to simulate and estimate characteristics of the uncertain system. The Monte Carlo Simulation (MCS) is one the most intuitive and easy to implement simulation procedures (CAFLISCH, 1998; SANKARARAMAN, 2015). A usual drawback of MCS is the large number of simulations needed for accuracy convergence. On the other hand, this method can handle highly non-linear systems. It may also serve as a reference to analytical solution methods, which very easily become mathematically too complicated (DITLEVSEN; MADSEN, 1996).
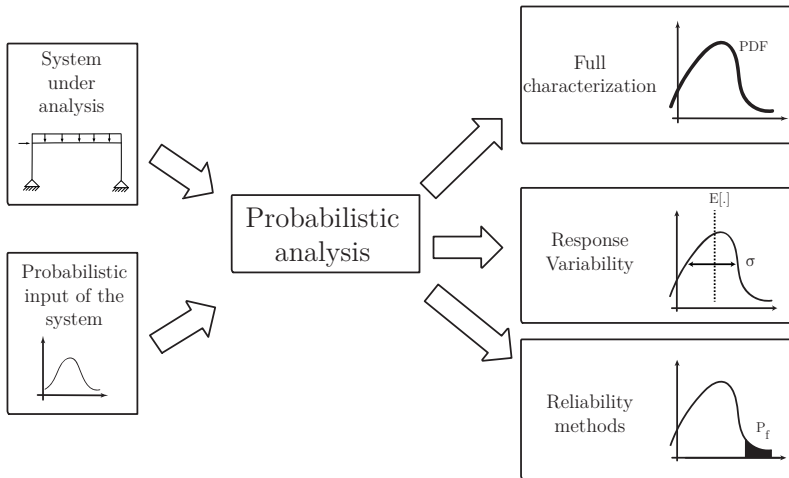
Figure 12 – Analytical methods

3.10   Formulation of designs in an uncertain context

When considering uncertainties, designers usually seek one or both of the following properties: reliability and robustness. A reliable

design means one that has a low or acceptable probability of failure. Robust designs are those which are less sensitive to inherent parameter variability, without removing the sources of uncertainty. Considering one or the other and possibly coupling it with constrained or unconstrained optimization leads to multiple problem formulations as seen in Table 1. For this study, the scope is limited to the Robust Design formulation. This formulation will be used when considering the Stochastic Kriging approach that will be seen in chapter 5.

Table 1 – Different uncertain formulations

		No objective function	Objective function d, x deterministic	Objective function d, x uncertain
No constraint function			Optimal design (optimization without constraint)	Robust design
Constraint function d, x deterministic		Admissible design (sizing)	Optimal and admissible design (optimization under constraint)	Robust and admissible design
Constraint function d, x uncertain		Reliable design (reliability)	Optimal and reliable design (RBDO)	Robust and reliable design (RBRDO)

increase of numerical efforts — Robustness — >

increase of numerical efforts / Reliability

Note – Gray tiles consider uncertainty

Source – Adapted from Lelièvre *et al.* (2016)

3.10.1   Robust Design Optimization - RDO

Robust Design Optimization (RDO) is usually employed as a multi-objective optimization problem, where the mean and variance of the system are to be minimized (BEN-TAL; NEMIROVSKI, 2002; BECK *et al.*, 2015). Mean and variance may be different minimizers and therefore could represent conflicting goals. Nevertheless, one may

aggregate the different objectives in a single function using a weighted sum of the objective functions (BEYER; SENDHOFF, 2007).

More generally, RDO may represent an optimization problem taking into account uncertainties and minimizing any statistic of the performance function. Thus, it is possible to formulate the problem as (LOPEZ; BECK, 2013):

Find $\mathbf{d}$ which minimizes:

$$\text{stat}[f(\mathbf{d}, \mathbf{X})], \tag{3.38}$$

where $\mathbf{d} \in \mathbb{R}^n$ is the design vector, $\mathbf{X} \in \mathbb{R}^m$ is the vector of random parameters of the system, and $\text{stat}[.]$ represents some statistic of the performance function $f(\mathbf{d}, \mathbf{X})$, such as:

$$\text{stat}[.] = \begin{cases} P_k : \text{the } k^{th} \text{ percentile of the performance function;} \\ E[.] : \text{the expected value;} \\ \sigma^2[.] : \text{the variance;} \\ \alpha E[.] + (1-\alpha)P_k \, ; \, \alpha \in [0,1] : \text{a multi}-\text{objective problem;} \\ \alpha E[.] + (1-\alpha)\sigma^2[.] \, ; \, \alpha \in [0,1] : \text{a conventional multi-objective problem.} \end{cases} \tag{3.39}$$

It can be noticed that now the parameters are random variables, in contrast with the deterministic formulation presented in chapter 2.

A practical application of this concept in engineering is on the optimum control of structures subject to seismic loading. It can be cited, for example, the project of damping systems (*e.g.* friction dumpers, tuned mass dumpers) in buildings subjected to wind or earthquake loads (ARFIADI; HADI, 2011; MIGUEL *et al.*, 2014; LOPEZ *et al.*, 2015).

In the present study, the objective is to minimize the expected value of a multivariate objective function with uncertain parameters. Thus, it can be formulated as a RDO problem with $\text{stat}[.] = E[.]$. This type of problem arises, for example, when considering robust optimal control problems of dynamical systems. An academic example problem

that will be presented in section 6.4 deals with the optimization of damping device parameters on shear frame buildings subjected to earthquake loads. This control problem involves minimizing the expected probability of structural failure, where "failure" is related to the probability of first passage from a safe region over a certain limiting barrier (TAFLANIDIS; SCRUGGS, 2009).

For now, looking at a simplified problem version, suppose the system performance is given by the function $f(\mathbf{d}, \mathbf{X})$, then the objective function for a robust-to-uncertainties design is given by (TAFLANIDIS; BECK, 2008):

$$E[f(\mathbf{d}, \mathbf{X})] = \int f(\mathbf{d}, \mathbf{x}) \, f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}, \qquad (3.40)$$

where $f_{\mathbf{X}}$ is the joint PDF of the probabilistic parameters. This multidimensional integral may be computed by simulation procedures. Using MCS, one may write:

$$\hat{E}_{Ns}[f(\mathbf{d}, \mathbf{X})] = \frac{1}{Ns} \sum_{i=1}^{Ns} f(\mathbf{d}, \mathbf{X}_i), \qquad (3.41)$$

where Ns is the number of sampled parameter vectors. Numerous evaluations may be needed to approximate the multi-dimensional integral accurately. Thus, the metamodeling approach described in chapter 4 may be used in order to solve such problem efficiently.

The approximation accuracy of Equation 3.41 depends on Ns, *i.e.*, as $Ns \to \infty$, $\hat{E}_{Ns}[f(\mathbf{d}, \mathbf{X})] \to E[f(\mathbf{d}, \mathbf{X})]$. Additionally, the error in this approximation may be estimated by the variance of the mean $\hat{E}_{Ns}$, calculated using Equation 3.18. This information can then be employed to help filter stochastic noise in the Kriging metamodel. This approach will be detailed using an example in section 5.2.

4 Deterministic metamodeling

Much of today's engineering analysis consists of running complex computer codes: supplying a vector of design variables (inputs) $\mathbf{d}$ and computing a vector of responses (outputs) $\mathbf{y}$. Although computer power has been steadily increasing, the expense of running many analysis codes remain non-trivial. For example, a single evaluation of a finite element analysis (FEA) model may take minutes to hours, if not longer. Thus, it often becomes impractical to perform a large number of such simulations, for example, as required during optimization. To address such a challenge, approximation or metamodeling techniques are often used.

The basic approach is to construct a surrogate model or approximate response surface of the analysis function, which is sufficiently accurate at a reasonable cost. Suppose one wants to create a surrogate model to a displacement-based FEA code. If the response (displacement) of the analysis, given a vector of inputs $\mathbf{d}$ (geometry, load, boundary conditions), is:

$$y = f(\mathbf{d}), \tag{4.1}$$

what is done is to construct a substitute model for the original model, *i.e.* a metamodel, such that:

$$\hat{y} = \hat{f}(\mathbf{d}) \approx f(\mathbf{d}), \tag{4.2}$$

over a given design domain. The terms metamodel, response surface and surrogate model is used interchangeably throughout the text to refer to the approximation in Equation 4.2.

4.1   Metamodel approaches

Although multiple metamodel approaches exist, here are listed some of the most actively researched methods:

- Polynomial Regression (PR) (BOX *et al.*, 1978; MONTGOMERY; MYERS, 1995);

- Radial Basis Functions (RBF) (DYN *et al.*, 1986);

- Kriging (SACKS *et al.*, 1989);

- Neural Networks (HAYKIN, 1998);

- Polynomial Chaos Expansion (SUDRET, 2008);

- Support Vector Regression (CLARKE *et al.*, 2005).

Early use of approximation methods in engineering revolved around the response surface methodology (RSM). The basis of this approach is to construct a metamodel using second-order polynomial regression with least squares fitting. It has been extensively used in many fields of engineering (JR., 1997; LIU *et al.*, 2000; TORII; LOPEZ, 2011; EOM *et al.*, 2011; TORII *et al.*, 2012). However, the use low-order polynomials makes this approach limited in terms of flexibility. It has been shown to produce better results only when applied locally or on convex functions (HUSSAIN *et al.*, 2002). Thus, the application of this approach is not suited for engineering highly non-linear and non-convex functions.

Among the other approaches, most of them are capable of fitting a non-linear and non-convex function to some extent. Nevertheless, for this study the surrogate of choice is Kriging. It has been gaining popularity in the last decade, and has been applied in numerous branches of science (THEODOSSIOU; LATINOPOULOS, 2006; HUANG *et al.*, 2011; XU *et al.*, 2012; ZHU *et al.*, 2014). It has been shown to produce very accurate surfaces and offers a estimate of the committed error. This estimative is useful because it enables the development of an integrated

global optimization procedure. More on this procedure will be discussed in section 4.4.

4.2 Kriging

Kriging is named after Danie Krige (KRIGE, 1951), a South African mining engineer who first developed the method in the field of geostatistics (CRESSIE, 1993). Initially restricted to the geostatistics, the method was later employed in deterministic simulation, mainly influenced by Sacks *et al.* (1989).

It acts as a regression model which exactly interpolates the observed input/output data. The advantage of an interpolative metamodel is the capacity of yielding globally accurate response surfaces while ensuring that previously known response values remains the same. This advantage is useful in optimization.

Contrary to classic regression where coefficients are estimated to completely describe what the function is, in kriging the focus is to estimate parameters that describe how the function typically behaves (JONES *et al.*, 1998). This difference may be emphasized by first looking at a regression procedure.

Suppose there is a deterministic black box function of n variables and n_s points to be evaluated. For each point $\mathbf{d}^{(i)} = (d_1^{(i)}, \ldots, d_n^{(i)})$, the function returns an associated value $y^{(i)} = y(\mathbf{d}^{(i)})$, for $i = 1, \ldots n_s$. From this scenario, one of the simplest ways to fit a response for this function is by the use of linear regression. This technique treats observations as if they are generated from the following model:

$$y(\mathbf{d}^{(i)}) = \sum_h \beta_h \, f_h(\mathbf{d}^{(i)}) + \epsilon^{(i)} \quad (i = 1 \ldots n_s), \tag{4.3}$$

where f_h are functional terms, β_h are coefficients to be estimated and $\epsilon^{(i)}$ are normally distributed independent error terms with zero mean and variance σ^2. One problem of this approach is that the form of functional terms to be used are not known (FORRESTER *et al.*, 2008). A large number of terms may lead to better approximation, similarly

to a Taylor series expansion. However, increasing the number of terms also increases the flexibility of the model and there is danger of over fitting the response (GELDER *et al.*, 2014).

Regarding the form selection, another method in literature tackles this limitation: Radial Basis Functions (RBF). Here the complex function is modeled considering linear combinations of radially symmetric basis centered on the known n_s sampled points. Each base ϕ_i has a weight parameter w_i which controls how far its radius of influence goes when computing a prediction. That is:

$$y(\mathbf{d}) = \sum_{i=1}^{n_s} w_i\, \phi(\|\mathbf{d} - \mathbf{d_i}\|), \tag{4.4}$$

from here, multiple basis can be chosen:

- linear $\phi(r) = r$

- cubic $\phi(r) = r^3$

- thin plate spline $\phi(r) = r^2 \ln(r)$

- Gaussian $\phi(r) = e^{-r^2/(2c^2)}$

- multiquadric $\phi(r) = \sqrt{r^2 + c^2}$

- inverse multiquadric $\phi(r) = 1/\sqrt{r^2 + c^2}$,

where the term c for the last three basis is a prescribed parameter. In RBF the most used basis is the multiquadric, the influence of the parameter c can be seen in Figure 13.

It is important to notice that in RBF the correlation parameter c is identical across all dimensions. For kriging this is not the case. Moreover, the use of Gaussian basis is a key aspect of the formulation.

Given this initial insight on polynomial regression, its limitations as well as the characteristics of RBF which shall be useful, it is possible to discuss the kriging metamodel on its own merits.
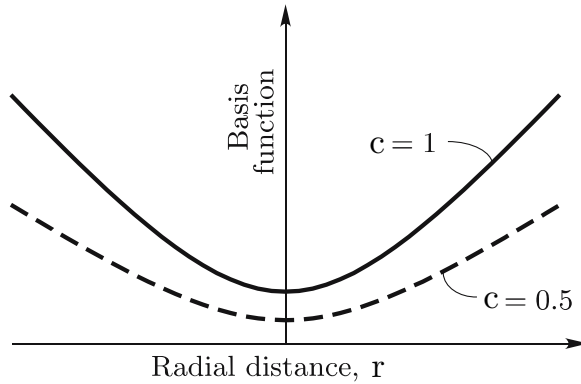
Figure 13 – Multiquadric base function

4.2.1 Formulation

The kriging metamodel formulation combines two functions: a global trend and a function that models local departures from the trend (SIMPSON *et al.*, 2001):

$$\hat{y}(\mathbf{d}) = M(\mathbf{d}) + Z(\mathbf{d}), \tag{4.5}$$

where $M(\mathbf{d})$ represents the global trend and $Z(\mathbf{d})$ represents the local departures. The global trend can be any sum of polynomial terms. However, it is usually taken as a constant term μ, which yields good prediction for most situations (SACKS *et al.*, 1989; CHEN *et al.*, 2003). The local departures are evaluated from a spatial correlation model.

When dealing with kriging, the interpolation is constructed using a stochastic process approach. In this approach, the function response at any position $\mathbf{d}$ is modeled as a realization of a stationary stochastic process. The response random variable $Y(\mathbf{d})$ has an assumed normal distribution with mean μ and variance σ^2. Considering an initial sampling plan $\mathbf{\Gamma} = \{\mathbf{d}^{(1)}, \mathbf{d}^{(2)}, \ldots, \mathbf{d}^{(n_s)}\}$:

The covariance between any two input points $\mathbf{d}^{(i)}$ and $\mathbf{d}^{(j)}$ is given by:

$$\mathrm{Cov}\left[Z(\mathbf{d}^{(i)}), Z(\mathbf{d}^{(j)})\right] = \sigma^2\ \mathbf{\Psi}\left(\mathbf{d}^{(i)}, \mathbf{d}^{(j)}\right), \qquad (4.6)$$

where $\mathbf{\Psi}$ is the correlation matrix and have the form:

$$\mathbf{\Psi}\left(\mathbf{d}^{(i)}, \mathbf{d}^{(j)}\right) = \sum_{k=1}^{n} \exp\left(-\theta_k\ \left|d_k^{(i)} - d_k^{(j)}\right|^{p_k}\right). \qquad (4.7)$$

The correlation function in Equation 4.7 is clearly similar to the Gaussian basis in the previous section. The difference is the number of parameters. While in RBF there was a single parameter, here the variance of the basis function can be controlled in each of the n dimensions of the design space by θ_k. The exponent p_k can also be varied across all dimensions but its value is usually held fixed with $p_k \in [1,2]$ (FORRESTER; KEANE, 2009). A plot with the influence of the exponent p_k on the correlation is shown in Figure 14. As it can be seen, this parameter relates with the "smoothness" of the correlation. Decreasing p_k makes the initial correlation drop faster, to the point where setting $p_k = 0.1$ leads to a near discontinuity.

This correlation basis is intuitive in the sense that when points move close together, $d_k^{(i)} - d_k^{(j)} \to 0$, then $\exp(-\theta_k\ |d_k^{(i)} - d_k^{(j)}|^{p_k}) \to 1$, displaying very close correlation. On the other hand, when distance between points is very large $d_k^{(i)} - d_k^{(j)} \to \infty$, then $\exp(-\theta_k\ |d_k^{(i)} - d_k^{(j)}|^{p_k}) \to 0$, exhibiting no correlation. The correlation variation with θ can be seen in Figure 15. There is a notion of "activity" associated with θ, that is, it affects how far a sample point influence extends.

Clearly, these additional Kriging parameters make the model construction more expensive, but this cost is compensated by the possibility of improved accuracy in the surrogate. This gain has been demonstrated in studies such as: Costa *et al.* (1999), Krishnamurthy (2005) and Kim *et al.* (2009). Its usefulness depends partially on the condition of expensiveness of the true function. Otherwise, the parameter estimation costs could overcome the costs of a large number of evaluations of the true function.
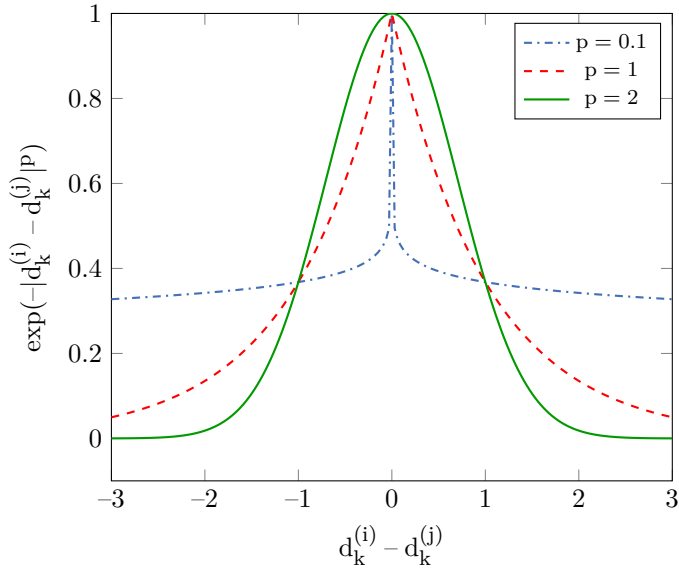
Figure 14 – Influence of p on correlation

Source – Adapted from Forrester e Keane (2009)

It can be noticed that this metamodel makes very few assumptions on the form of the landscape being approximated, which is one of the reasons it can emulate a varied range of functions so effectively (FORRESTER *et al.*, 2008).

The unknown parameters θ_k and p_k may be found by Maximum Likelihood Estimate (MLE) (MONTGOMERY; RUNGER, 2010). In this process, the parameters are chosen so that the model of the function's typical behavior is most consistent with the observed data. With the observations vector denoted as $\mathbf{y} = \{y^{(1)}, y^{(2)}, \ldots, y^{(n_s)}\}$, the likelihood function is written as (JONES *et al.*, 1998):

$$\mathcal{L} = \frac{1}{(2\pi)^{\frac{n_s}{2}} (\sigma^2)^{\frac{n_s}{2}} \sqrt{\det(\boldsymbol{\Psi})}} \exp\left[\frac{-(\mathbf{y} - \mathbf{1}\mu)^{\mathrm{T}} \boldsymbol{\Psi}^{-1} (\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2}\right], \quad (4.8)$$

where $\mathbf{1}$ is $1 \times n_s$ column vector of ones, $\boldsymbol{\Psi}$ is the correlation matrix
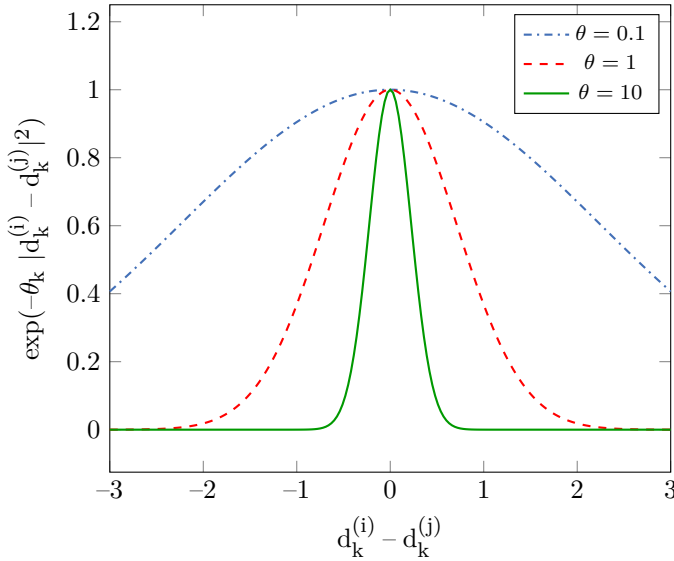
Figure 15 – Influence of θ on correlation

Source – Adapted from Forrester e Keane (2009)

from equation Equation 4.7, and $(.)^{\mathrm{T}}$ is the transpose operation.

It is usually more convenient to maximize the log-likelihood, which can be written by taking the natural logarithm and ignoring constant terms as:

$$\log(\mathcal{L}) = -\frac{n_s}{2}\ln(\sigma^2) - \frac{1}{2}\ln(\det(\boldsymbol{\Psi})) - \frac{(\mathbf{y}-\mathbf{1}\mu)^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\mathbf{y}-\mathbf{1}\mu)}{2\sigma^2}. \quad (4.9)$$

Taking the derivative of Equation 4.9 with respect to both σ^2 and μ and setting them to zero, it is possible to obtain the optimal parameters as functions of $\boldsymbol{\Psi}$:

$$\hat{\mu} = \frac{\mathbf{1}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\mathbf{y}}{\mathbf{1}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\mathbf{1}}, \quad (4.10)$$

$$\widehat{\sigma^2} = \frac{(\mathbf{y}-\mathbf{1}\hat{\mu})^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\mathbf{y}-\mathbf{1}\hat{\mu})}{n_s}. \quad (4.11)$$

Substituting Equations 4.10 and 4.11 back into Equation 4.9 and again ignoring constant terms, the so-called concentrated log-likelihood function is obtained:

$$\log(\mathcal{L})_{\text{concentrated}} = -\frac{n_{\text{s}}}{2}\ln(\widehat{\sigma^2}) - \frac{1}{2}\ln(\det(\mathbf{\Psi})). \qquad (4.12)$$

As it can be seen, this function depends only on $\mathbf{\Psi}$, hence, on the correlation parameters. This function is then maximized to obtain the estimates $\hat{\theta}_{\text{k}}$ and $\hat{\text{p}}_{\text{k}}$. Once the optimal parameters are found, it is possible to compute the estimates $\hat{\mu}$ and $\widehat{\sigma^2}$ using Equations 4.10 and 4.11.

The parameter optimization subproblem is not trivial as it usually contains a highly non-linear region and a long flat region at the same time (ZHAO *et al.*, 2010). For this reason, gradient-based algorithms are not well suited and metaheuristics, as those seen in subsection 2.3.3, are often applied (SONG *et al.*, 2013). Additionally, this is the most computationally intensive step in the metamodel creation. As can be seen in Equation 4.9, it requires the evaluation of the determinant of $\mathbf{\Psi}$ as well as the computation of its inverse. The use of Gaussian basis is advantageous in this regard, as it results in a symmetric positive definite correlation matrix. This enables the use of Cholesky decomposition of $\mathbf{\Psi}$, which increases the calculation efficiency (FORRESTER *et al.*, 2008; CASTRILLÓN-CANDÁS *et al.*, 2015). With this procedure, any computation that required matrix inversion becomes solvable by a forward and backward substitution on the decomposed matrix. Moreover, the determinant can be quickly obtained by taking the square of the product of the main diagonal terms of the decomposed matrix (FORD, 2014).

With the estimated parameters it is now possible to make function predictions at an unknown point $\mathbf{d}_{\text{u}}$ using the kriging predictor, as derived in Sacks *et al.* (1989):

$$\hat{y}(\mathbf{d}_{\text{u}}) = \hat{\mu} + \mathbf{r}^{\text{T}}\mathbf{\Psi}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \qquad (4.13)$$

where $\mathbf{r}$ is the vector of correlations of $\mathbf{d_u}$ with the other n_s sampled points.

One of the key benefits of kriging and other Gaussian process based models is the provision of an estimated error in its predictions. The Mean Squared Error (MSE), derived by Sacks *et al.* (1989) using the standard stochastic process approach reads:

$$\mathrm{s}^2(\mathbf{d}) = \widehat{\sigma^2} \left[1 - \mathbf{r}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\mathbf{r} + \frac{(1 - \mathbf{1}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\mathbf{r})^2}{\mathbf{1}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\mathbf{1}} \right]. \qquad (4.14)$$

Equation 4.14 has the intuitive property that it is zero at already sampled points. This is expected, as there is no uncertainty about known points in the deterministic case.

4.3   Sampling plan

The basis of metamodeling revolves around creating an approximation of an unknown objective function landscape based on a model fitted using a certain number of function evaluations. In order to achieve maximum efficiency in accessing the global response, the initial sampled points must fill the design space (PRONZATO; MÜLLER, 2012). Kriging accuracy increases with the existence of sampled points in the vicinity of where predictions are being made. Therefore, it is rather intuitive that a uniform level of model accuracy throughout the design space requires a uniform spread of points. Thus, the choice of an appropriate sampling technique is generally considered critical for the performance of any metamodeling approach (KALAGNANAM; DIWEKAR, 1997).

A commonly used sampling plan is called Latin Hypercube Sampling (LHS). This technique aims to sample the design space by ensuring uniformly spread projection of points on all axes. The logic behind this approach is that it is wasteful to sample a variable more than once at the same value. In Figure 16 a two-dimensional LHS or Latin Square is shown.

Despite displaying uniformity in each dimension separately this
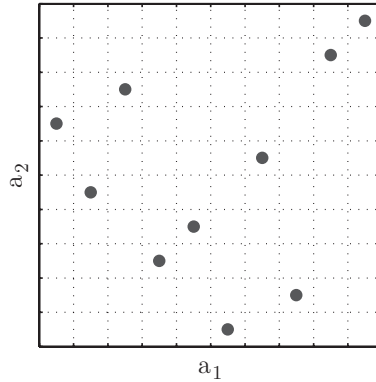
Figure 16 – Latin Square

sampling plan can still be improved in terms of "space-fillingness". Intuitively, placing all points on the main diagonal of the design space will fulfill the projection criteria, although it would not fill the space uniformly.

In this regard, Mitchell e Morris (1992) proposed a set of rules used the evaluate the "space-fillingness" property, based on the maximin metric introduced by Johnson *et al.* (1990). Let $\{\delta_1, \delta_2, \ldots \delta_m\}$ be the list of the unique values of distances between all possible pairs of points in a sampling plan $\mathbf{\Gamma}$, sorted in ascending order. Moreover, let $\kappa_1, \kappa_2, \ldots, \kappa_m$ be defined such that κ_j is the number of pairs of points in $\mathbf{\Gamma}$ separated by the distance δ_j. The Morris-Mitchel criteria considers a maximin plan as one that maximizes the distance δ and also minimizes the number of points J within that distance. This verification starts from the pair (δ_1, κ_1) and is extended to the following m pairs as needed in order to avoid ties. The distance measure between any two points $\mathbf{d}^\alpha$ and $\mathbf{d}^\beta$ is computed by the p-norm as follows (MORRIS; MITCHELL, 1995):

$$\delta(\mathbf{d}^\alpha, \mathbf{d}^\beta) := \left(\sum_{j=1}^{n} |d_j^\alpha - d_j^\beta|^p \right)^{\frac{1}{p}}, \qquad (4.15)$$

where usually p = 1 (Manhattan distance) or p = 2 (Euclidian distance). In Figure 17 a random sample plan (Figure 17a) and a LHS optimized by the Morris-Mitchel criteria (Figure 17b) are presented. Clearly, the second case shows points that are more distanced from each other, therefore improving the space filling on the unit cube.

(a) Random Sampling   (b) Morris-Mitchel LHS
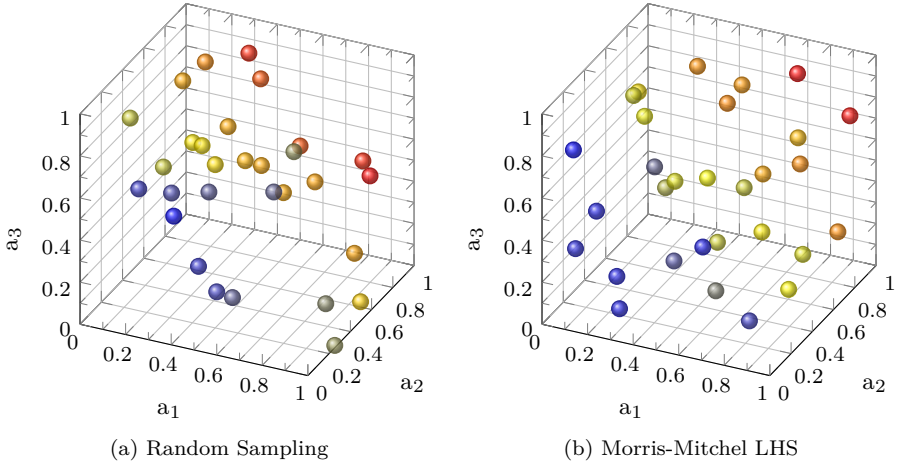
Figure 17 – Illustration of designs with different space-fillingness

4.4   Efficient Global Optimization

Efficient Global Optimization (EGO) is the name of the framework developed by Jones *et al.* (1998), which exploits the information provided by the kriging metamodel to iteratively add new points, improving the surrogate accuracy and at the same time seeking its minimum.

According to Jones (2001), methods for selecting search points can be divided in two categories: two-stage and one-stage approaches. In that study, seven methods are presented, as shown in Table 2. Most current and popular approaches are two-staged. They can be described by the following steps:

1. Construct initial sampling plan;

Table 2 – Selection methods

Kind of Response Surface			Method for selecting search points					
			Two-stage approach: first fits surface, then find the next iterate by optimizing an auxiliary function based on the surface				One-stage approach: evaluate hypothesis about optimum based on implications for the response surface	
			Minimize the Response Surface	Minimize a Lower Bounding Function	Maximize the Probability of Improvement	Maximize Expected Improvement	Goal seeking: find point that achieves a given target	Optimization: find point that minimizes an objective
Not interpolating (smoothing)	Quadratic polynomials and other regression models		1					
Interpolating	Fixed basis functions. No statistics.	Thin-plate splines, Multiquadrics	2	3	4	5	6	7
	Tuned basis functions. Statistical interpretation	Kriging						

Source – Adapted from Jones (2001)

2. Compute responses on sampled points and fit initial metamodel;

3. Iterate, adding infill points which optimize an auxiliary function.

These steps are repeated until a termination criterion is met, *e.g.*, maximum number of function evaluations. Methods in this category usually differ by the auxiliary function being optimized.

Alternatively, in one-stage approaches, step 2 is skipped and the initial metamodel is constructed based on the credibility of a hypothesis. This hypothesis may be, for example, that the surface interpolates the observed data and additionally a target response.

Methods 1 and 2 consist of adding points that minimize the surface using quadratic polynomials and kriging, respectively. Method 1 is promptly discarded considering that the regression model keeps its quadratic characteristics. Thus, adding new points does not necessarily improve the accuracy of the response. Method 2, on the other hand, is capable of finding a local optimum. Nevertheless, it lacks exploratory features in order to search for the global optimum. It may fail in case the surface minimum coincides with a sampled point of the function.

Methods 3 to 7 as in Table 2 are the ones that attempt to use kriging statistical information to guide the search for the global optimum.

Method 3 determines its next iterate by minimizing a "statistical lower bounding function". This function is usually expressed by the kriging predictor minus a number of standard errors. As noted by Jones (2001), in Method 3 the iterate is not dense, *i.e.* it does not guarantee global convergence (TORN; ZILINSKAS, 1989). The remaining methods 4 to 7 are all suitable to finding a global optimum. In this study, however, emphasis will be given to Method 4, which maximizes the probability of improvement and also on its promising modification called Enhanced Method 4 (JONES, 2001). This is due to its desirable characteristic of enabling the addition of multiple points to the surrogate. Although not implemented in this study, this feature may enable exploiting parallel computing.

4.5 Probability of Improvement - PI

The fourth method described by Jones (2001), consists on adding new points to the kriging metamodel which maximize the auxiliary Probability of Improvement measure. By the premise that kriging prediction $\hat{y}(\mathbf{d})$ is the realization of a random variable $Y(\mathbf{d})$, it is possible to define an improvement event as:

$$I(\mathbf{d}, y_{min}) := \max(0, y_{min} - Y(\mathbf{d})), \qquad (4.16)$$

where y_{min} is the current best solution. Thus, the probablity of improvement over this current solution can be written as:

$$P[I(\mathbf{d}, y_{min})] = \Phi\left(\frac{y_{min} - \hat{y}(\mathbf{d})}{s(\mathbf{d})}\right), \qquad (4.17)$$

where Φ denotes the normal cumulative density function as shown in Equation 3.22. A graphical interpretation of this probability can be seen in Figure 18. It shows the prediction along with a vertical Gaussian distribution with variance $s^2(\mathbf{d})$ centered around $\hat{y}(\mathbf{d})$. The probability of improvement is the area enclosed by the normal distribution below the current best observed value.
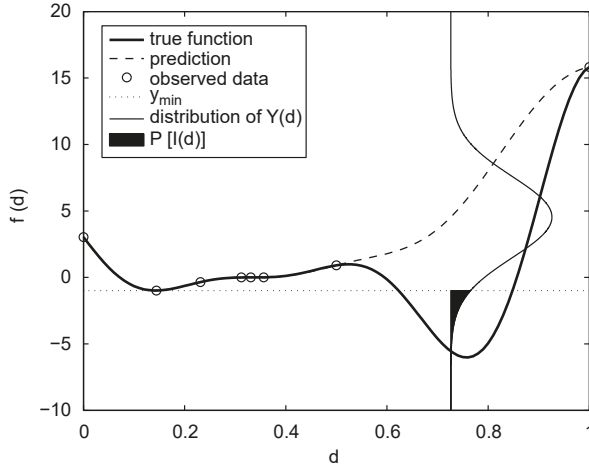
Figure 18 – A graphical interpretation of the probability of improvement

Source – Adapted from Forrester e Keane (2009)

This method is intuitively convergent. As the function is sampled more and more around the current best point, the standard error tends to zero. Consequently, the search is directed to unexplored regions, *i.e.*, with larger uncertainty (FORRESTER *et al.*, 2008).

4.6 Enhanced Method 4 - EM4

It turns out that the improvement event defined earlier is very modest. This causes the search to sample many points around the present best in a local manner before exploring other regions of the landscape.

To address this shortcoming it is possible define the improvement based on a target response as in:

$$I(\mathbf{d}, y_{\text{target}}) := \max(0, y_{\text{target}} - Y(\mathbf{d})), \qquad (4.18)$$

where y_{target} can be any arbitrary value, *i.e.*, a fraction of the present best solution.

This formulation can give a better balance between exploration and exploitation but it also makes the search very sensible to the choice of y_{target}. If the target improvement is too small, the search will be highly local, as discussed earlier, moving to a more global search only after searching nearly exhaustively around the current best point. Moreover, if the target improvement is too high, the search will be excessively global, and the algorithm will be slow to fine-tune any promising solutions (JONES, 2001).

In Enhanced Method 4 multiple targets are evaluated, representing small, moderate and big improvements. By selecting several search points in each iteration it is possible to search both globally and locally in each iteration. This infill criteria allows one to take advantage of any parallel computing capabilities.

The target creation procedure suggested by Jones (2001), can be calculated as follows:

$$y_{target} = y_{min} - \alpha(f_{max} - f_{min}), \qquad (4.19)$$

where y_{min} is minimum value of the surface, f_{max} and f_{min} are the maximum and minimum response from the sampled points, respectively, and α is the factor which controls the quantity of improvement being searched. An $\alpha = 0.001$ means a very modest improvement which usually results in points close to the current best. On the other hand, an $\alpha = 0.5$ seeks a very large improvement over the present best, corresponding to half the estimated amplitude of the function.

Searching multiple targets in each iteration results in multiple optimization procedures. This means that a larger computational effort is needed. However, this cost becomes less relevant if all optimizations are made in parallel. Although searching multiple targets is beneficial, adding a large number of points to the model may not be. When searching over multiple α, points tend to cluster around the present best and promising unexplored regions. Instead of adding the total amount of searched targets, a clustering procedure is employed and just the best target within a cluster is added to the model.

$$f(d) = \sin(d) + \sin(\tfrac{2}{3}d)$$
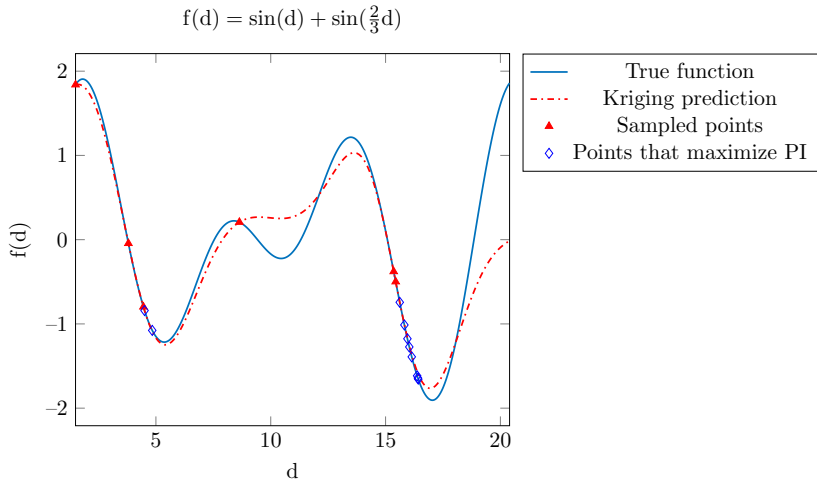
Figure 19 – Illustration of clustering over multiple targets

Figure 19 illustrates this situation. The same function presented in subsection 2.2.2 is shown. After the construction of a Kriging approximation with 6 sampled points, points that maximize the PI are searched. Those points, plotted as diamonds in the figure, form two clusters when considering 10 different targets of varying α. Under the Enhanced Method 4 approach, from these clusters, two points with the best improvement would then be added to the model. In this example, the clusters formed are easy to identify.

In practice, when considering multi-dimensional functions the clusters are identified by some arbitrary criteria, usually involving a "distance" threshold. The number of points added depends on this clustering threshold. The manner to cluster points and set targets is very important to this method and improvements for these procedures have been studied recently in Viana e Haftka (2010), Chaudhuri e Haftka (2014) and Chaudhuri *et al.* (2015).

5 Stochastic metamodeling

5.1 Kriging for noisy data

As discussed, kriging is an interpolative model, *i.e.*, sampled points are taken as exact, and deterministically represent the function output. However, when considering noisy data, such as the random variables used to describe uncertainty, the exact output ceases to exist. The realization of the random variables results in one of many possible objective function values.

One simple approach used by Beers e Kleijnen (2003) consists of using the average over a certain number of sample replications in the deterministic kriging formulation. They show that this approach performs better than the use of some regression models.

Figure 20 displays how the number of replications affects the approximated function. It shows a graph of a deterministic function and its noisy counterpart under two representations: the first shows the plot from a single function evaluation, the second shows the resulting plot from an average of 100 simulations. It is possible to see that the average gives a much better approximation than a single noisy simulation.

Another approach is to allow the kriging model to regress data. This may be achieved by introducing a regularization or regression parameter λ (POGGIO; GIROSI, 1990). Ideally, it should be set to the variance of the noise in the response (KEANE; NAIR, 2005). Although this variance is usually unknown, one approach is to add it to the list of parameters that need to be estimated in the MLE procedure (FORRESTER *et al.*, 2008). An alternative approach, which is proposed by this study, is to set λ as the estimated variance from probability theory over a number of output replications. More on this approach is discussed considering a practical example in section 5.2.

The λ parameter is added to the leading diagonal of the kriging
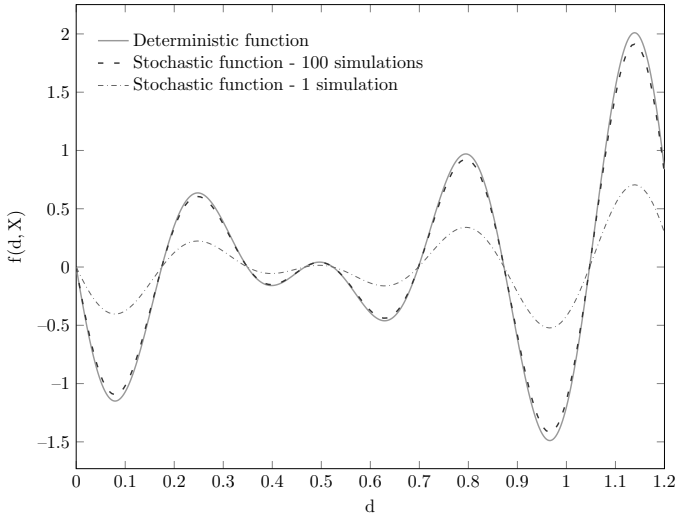
Figure 20 – Function plot - Problem 1

correlation matrix $\mathbf{\Psi}$, defined in Equation 4.7. Forrester *et al.* (2006) redefines the kriging prediction in order to keep using the infill criteria seen in section 4.4 while considering the regression parameter.

5.2   Kriging exploiting Monte Carlo variance

As seen in section 5.1, it is possible to use the kriging metamodel to filter or regress stochastic noise. To accomplish this, the regression parameter λ should be set to the variance of the noise.

Moreover, this study aims to solve the type of problem stated in subsection 3.10.1, *i.e.*, the minimization of an expected value $E[f(\mathbf{d}, \mathbf{X})]$. This expected value may be computed using Monte Carlo simulation. Ideally, the number of samples from the Monte Carlo simulation should be infinite. Since this is not possible, the estimated variance may be used as a measure of error when computing $\hat{E}[f(\mathbf{d}, \mathbf{X})]$ with a finite number of samples. The expression to compute this value originates from probability theory and was shown in Equation 3.18.

Thus, the proposed approach is to use the estimated variance

information from the Monte Carlo simulation in the regression parameter λ from kriging. The goal here is to use the noise filtering capabilities of kriging to reduce the number of function evaluations when adding new points to the stochastic metamodel. It becomes possible to construct the kriging model based on a "known" error that is enforced by the Monte Carlo procedure.

Responses may be evaluated for a certain input until the variance of the mean reaches an arbitrary target. If the target variance is set too high, kriging may not be able to correctly regress the data and the approximation error will be large. On the other hand, if the target variance is set too low, a large number of function evaluations are needed. This trade-off is precisely what is observed in Figure 21. It shows an one-dimensional function approximated over a number of different target variances. The function is written as:

$$f(d, X) = -(16(dX)^2 - 24dX + 5)\exp(-dX), \tag{5.1}$$

where $X \sim \mathcal{N}(1, 0.1)$.

For each point placed in the model, new replications are made and the variance estimate is updated until it becomes lower than the target. The number of Objective Function Evaluations (OFE) is presented for each curve. It gives an insight of the computational burden required for the optimization. As seen in the figure, the number of OFE grows exponentially when exponentially decreasing the target variance. This observation will be used when developing an adaptive targeting decay function in section 5.5.

In order to compare the regressing kriging prediction with the noiseless model, the approximate value of the Root Mean Squared Error (RMSE) was calculated. It serves as a measure of the model accuracy. It is computed as:

$$\text{RMSE} = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (\hat{y}_i - y_i)^2}, \tag{5.2}$$

Figure 21 – Comparison of multiple target variances

where n_c is the number of comparisons made between the approximated response $\hat{y}$ and the analytical function y. For the plots in Figure 21, 100 points are compared, which coincides with the points used for plotting. Observing the plots, it is possible to note that for this problem a variance of 10^{-3} yields the best approximation considering a reasonable number of OFEs. For reference, timings from a machine using a 3.5 GHz processor are also displayed.

More generally, it is clear from Figure 21 that the incorporation of the information from the Monte-Carlo estimation may be useful in lowering the computational cost of the optimization. How to achieve a good trade-off between accuracy and computational cost is matter that will be addressed in the following sections.

5.3 Stochastic Kriging - SK

More recently, Ankenman *et al.* (2010) proposed an extension to the kriging deterministic methodology to deal with stochastic simulation. Their main contribution was to fully account for the sampling variability that is inherent to a stochastic simulation. In order to accomplish this, they characterized both the intrinsic error inherent in a stochastic simulation and the extrinsic error that comes from the metamodel approximation. This approach is mainly used in the mathematical field of operations research for prediction (STAUM, 2009).

It will become evident how such methodology fits well with the class of problems we are aiming to solve. In these problems, the minimization of the objective function f_{obj} can be written as the minimization of the integral:

$$\min f_{obj}(\mathbf{d}) = \int f(\mathbf{d}, \mathbf{x}) \, f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}, \tag{5.3}$$

where $f_{\mathbf{X}}$ is a density function, $\mathbf{d}$ is the vector of design variables and $\mathbf{X}$ the paramaters that model the uncertainty about the system. A PDF describes the available system knowledge related to these parameters. This integral denotes the expected performance of the system. It

arises not only on robust design context as seen in subsection 3.10.1 but also in performance-based engineering (MOEHLE; DEIERLEIN, 2004; SPENCE; GIOFFRÈ, 2012; BECK *et al.*, 2014; SPENCE; KA-REEM, 2014; BOBBY *et al.*, 2016) and Optimal Design of Experiments (FEDOROV, 1972; STEINBERG; HUNTER, 1984; MÜLLER, 2005; PRONZATO, 2008). The proposed SK metamodeling approach may be useful for solving these and other problems that may be written as the minimization of any integral.

It is often impossible to solve such integrals analytically. Numerical integration such as quadrature procedures may be employed although its accuracy is reduced for high dimensional problems. Alternatively, simulation techniques may be utilized. The integral can be viewed as the expected value over the uncertainties domain. It can be approximated by a Monte Carlo integration as in:

$$\hat{E}_{n_s}[f(\mathbf{d}, \mathbf{X})] = \frac{1}{n_s} \sum_{i=1}^{n_s} f(\mathbf{d}, \mathbf{X}_i). \tag{5.4}$$

It is known that such a procedure requires a large amount of function evaluations. Considering a black-box expensive function, the computational cost may be prohibitive. In the proposed approach, stochastic kriging is used to construct an approximate model, in which is easier to perform the optimization procedure. This approximation is created with information from the multiple simulations, compensating the stochastic noise. Thus, Monte Carlo is applied to solve the stochastic layer of the problem while also using its information to help create the model approximation.

The SK formulation can be seen as an extension of the deterministic case:

$$\hat{y}(\mathbf{d}_i) = M(\mathbf{d}_i) + Z(\mathbf{d}_i) + \epsilon_j(\mathbf{d}_i), \quad j = 1, 2, ..., n_r, \quad , i = 1, 2, ..., n_s \tag{5.5}$$

where $M(\mathbf{d})$ is the usual average trend, $Z(\mathbf{d})$ accounts for the model uncertainty and is now referred as extrinsic noise and the additional

term ϵ_j, called intrinsic noise, accounts for the simulation uncertainty from the j-th replication where n_r is the total number of replications. The intrinsic noise is assumed independent and identically distributed (i.i.d) across replications and possesses a Gaussian distribution with zero mean and covariance matrix (ANKENMAN *et al.*, 2010):

$$(\Sigma_\epsilon)_{ii} = \text{Var} \left[\frac{1}{n_r} \sum_{j=1}^{n_r} \epsilon_j(\mathbf{d}_i) \right], \quad i = 1, 2, ..., n_s. \tag{5.6}$$

The assumptions of independence and normal error are briefly explored considering a function with a distribution different from the Gaussian in Appendix C. The variance is often unknown and needs to be estimated. Using point estimate from the replications sample leads to:

$$(\widehat{\Sigma}_\epsilon)_{ii} = \frac{1}{n_r(n_r - 1)} \sum_{j=1}^{n_r} (\hat{y}(\mathbf{d}_i) - \bar{\hat{y}}(\mathbf{d}_i)^2), \quad i = 1, 2, ..., n_s, \tag{5.7}$$

where $\bar{\hat{y}}(\mathbf{d}_i)$ represents the mean from the i-th point replicates. This is the same unbiased estimator seen in Equation 3.18. Assuming the intrinsic noise with a Gaussian distribution is advantageous for the formulation of the predictor and the estimated error. As shown by Ankenman *et al.* (2010), the Best Linear Unbiased Predictor (BLUP) of SK is:

$$\hat{y}(\mathbf{d}_u) = \hat{\mu} + \mathbf{r}^T (\mathbf{\Psi} + \Sigma_\epsilon)^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{5.8}$$

which is the usual Kriging prediction from Equation 4.13 with the added diagonal correlation matrix from the intrinsic noise. This leads to the same result discussed in section 5.1 with the use of a regularization parameter. That is, when the intrinsic noise is homogeneous across the domain:

$$\Sigma_\epsilon = \lambda I, \tag{5.9}$$

where λ is the regularization parameter and I an identity matrix.

It is important to note however, that there is the possibility of heterogeneous noise variance. In such case, sampled points provide different terms for the diagonal in the correlation matrix Σ_ϵ. Equation 5.9 is still valid in this case but λ becomes a vector o variances instead of a constant value. Kriging estimation with heterogeneous noise is a key aspect that enables the development of the adaptive variance target selection that will be explored in section 5.5.

Regarding the estimated error, it also becomes an extension of the deterministic case (FORRESTER *et al.*, 2008). It reads:

$$
s^2(\mathbf{d_u}) = \widehat{\sigma^2}\left[1 + \lambda(\mathbf{d_u}) - \mathbf{r}^{\mathrm{T}}(\boldsymbol{\Psi} + \Sigma_\epsilon)^{-1}\mathbf{r} + \frac{1 - \mathbf{1}^{\mathrm{T}}(\boldsymbol{\Psi} + \Sigma_\epsilon)^{-1}\mathbf{r}}{\mathbf{1}^{\mathrm{T}}(\boldsymbol{\Psi} + \Sigma_\epsilon)^{-1}\mathbf{1}})\right],
$$
(5.10)

where $\lambda(\mathbf{d_u})$ is the noise variance term that depends on $\mathbf{d_u}$ in the heterogeneous noise case. The optimal MLE parameters $\widehat{\sigma^2}$ and $\hat{\mu}$ become, respectively:

$$
\widehat{\sigma^2} = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})^{\mathrm{T}}(\boldsymbol{\Psi} + \Sigma_\epsilon)^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu})}{n_s},
$$
(5.11)

and

$$
\hat{\mu} = \frac{\mathbf{1}^{\mathrm{T}}(\boldsymbol{\Psi} + \Sigma_\epsilon)^{-1}\mathbf{y}}{\mathbf{1}^{\mathrm{T}}(\boldsymbol{\Psi} + \Sigma_\epsilon)^{-1}\mathbf{1}}.
$$
(5.12)

Newer developments in SK literature propose different manners to obtain the intrinsic variance. Kleijnen e Mehdad (2016) compares the point estimate from Equation 5.7 with two other methods, namely Distribution-free Bootstrapping (DB) and Parametric Bootstrapping (PB). They conclude that using DB leads to a more conservative variance estimate and that it may be faster than using Monte Carlo replications. Kamiński (2015) proposes a new update method for sequential observations in addition to the use of smoothed variance evaluations to compute the intrinsic noise. Chen e Kim (2014) consider the case of unbiased samples, where the classic estimator no longer holds. In a different direction, Plumlee e Tuo (2014) suggests an approach where the intrinsic

error is not assumed Gaussian distributed. Emulators are built via an approach called Quantile Kriging producing less instability in variance estimation.

Considering the use of SK for optimization a few contributions may be cited. Picheny *et al.* (2013) benchmarked different infill criteria for the noisy case. Differently from the deterministic case, when there is noise in the objective function, multiple evaluations from the same input will lead to different results. In order to attenuate these effects, different criteria may be employed taking into account the noise. From that study, a modification of the deterministic Expected Improvement criterion called Augmented Expected Improvement obtained good performance. Sun *et al.* (2014) developed a new algorithm based on Gaussian processes which aims to balance exploration and exploitation and is applied to discrete optimization. Qu e Fu (2014) developed a new approach to enhance the SK accuracy when gradients are available. It uses them to extrapolate additional responses. Additionally, considering optimization use, the update procedure in Kamiński (2015) may be used to avoid recomputation of model parameters at each iteration.

5.4   Infill criteria for noisy evaluations

As discussed in section 5.3 the use of a stochastic kriging model allows to filter noise without having to guess the functional form of the underlying trend. Although SK gives a good noise filtering model, the error estimates are no longer appropriate for use when choosing points at which to run new computer experiments (FORRESTER *et al.*, 2006).

Probability of Improvement and Expected Improvement are two popular infill criteria for the EGO algorithm. They can combine local exploitation and global exploration into a single figure of merit. Locatelli (1997) has proven that a search based on running new experiments at points of maximum expected improvement of a kriging interpolation converges toward the global optimum. This is an important property that is exploited in deterministic kriging optimization. However, in the

presence of noisy functions, the key condition required by Locatelli's proof no longer holds, namely, that the error is zero at a sampled point. Sampling twice the same point will give different results because of the underlying noise. Thus, it is important to take into account this added uncertainty when choosing the next infill point.

5.4.1 Augmented Expected Improvement - AEI

The AEI criterion was proposed by Huang *et al.* (2006). It is treated as an extension of the Expected Improvement (EI) criteria. In EI, the infill point selected is the one that maximizes the expected value of the improvement measurement. This improvement definition requires a target value to indicate the greediness of the search. Recalling the improvement definition from Equation 4.18:

$$I(\mathbf{d}, y_{\text{target}}) := \max(0, y_{\text{target}} - Y(\mathbf{d})), \tag{5.13}$$

the y_{target} for EI is usually chosen as the minimal solution found so far. For the AEI infill criteria this target is the so-called effective best solution $\mathbf{d}^{**}$ and is computed as:

$$\mathbf{d}^{**} = \operatorname{argmin}(\mathbf{d}^{(i)} + \alpha s_{\text{n}}^{(i)}) \quad \text{for } i = 1, ..., n_{\text{s}}, \tag{5.14}$$

where $\mathbf{d}^{(i)}$ represents each of the sampled points, $s_{\text{n}}^{(i)}$ the corresponding kriging error, obtained by taking the square root of the MSE in Equation 4.14 and α is an arbitrary constant which the authors recommend setting to 1.

The criterion can be calculated as the expected improvement over the effective best solution multiplied by a penalization term:

$$\text{AEI}(\mathbf{d}) = \text{E}[I(\mathbf{d}, \mathbf{d}^{**})] \left(1 - \frac{\sqrt{\lambda}}{\sqrt{s_{\text{n}}^2(\mathbf{d}) + \lambda}}\right), \tag{5.15}$$

where $E[I(\mathbf{d}, \mathbf{d}^{**})]$ represents the expected value of the already defined improvement, s_n^2 the kriging estimated error and λ the intrinsic output noise. For now, λ is considered a constant, however when dealing with multiple variances (heterogeneous noise) it becomes $\lambda(\mathbf{d})$. Such case will be seen in section 5.5.

The second term, which is a penalty factor, amplifies the importance of the kriging variance and thus enhances exploration, avoiding multiple simulations over the same input. This is a criterion defined heuristically, although it has been shown to be efficient in many cases (PICHENY; GINSBOURGER, 2014). Comparatively, EM4, as discussed in section 4.6, follows the same idea of computing an improvement but instead considers multiple targets corresponding to different levels of desired improvement. It does not have, however, a mechanism to compensate noise.

Multiple benchmark tests are conducted for two different functions in order to verify the advantage on using an infill criterion that accounts for the noisy output. The tests consist of multiple runs of the EGO algorithm coupled with two infill criteria. In this test, EM4 and AEI are compared. In the comparison, noisy functions are considered under the stochastic kriging framework. The parameter that relates both criteria is the number of infill points. Fixating the number of infill points makes it possible to perceive which criterion needs the least amount of iterations to achieve a solution close to the optimum. The variance target is also varied during the experiment. This target variable determines the Monte Carlo sample size and effectively the computational cost involved in each iteration. Lower targets mean a higher computational cost by iteration but also a closer approximation to the deterministic output.

In order to reduce the variability of results, the figures present the average of 10 independent runs of the EGO algorithm for each configuration. For both functions being benchmarked, error bars represent the 5 and 95 percentiles of the independent runs results.

The first function considered the one-dimensional case and is

here named Problem 1. It is the fifth global optimization benchmark problem presented in Gavana (2016). It is defined as follows:

$$\min f_{\mathrm{obj}}(d) = \int f(d, x) \, f_X(x) \, dx, \tag{5.16}$$

where

$$f(d, X) = -(1.4 - 3dX) \sin(18dX), \tag{5.17}$$

- Domain: $d \in [0, 1.2]$;

- Deterministic global optimum: -1.48907 at $d = 0.96609$;

- Noise: $X \sim \mathcal{N}(1, 0.3)$;

- Optimum with 10000 simulations = -1.48333.

Figure 22 presents the results for a number of different configurations. It is possible to notice the trend of improvement when considering more infill points and also lower variance target. For the target variance of 0.01, AEI performed significantly better than EM4 for 20 and 30 infill points. For a number of infill points larger than 50, both algorithms were similar for every variance target.

The same approach was followed for the benchmarks on the second function. This time the function being evaluated was a noisy tilted Brain. The tilt comes from the $5d_1$ term, which is added to the function in order to force the existence of a single global optimum. Writing it as the minimization of an integral:

$$\min f_{\mathrm{obj}}(\mathbf{d}) = \int f(\mathbf{d}, \mathbf{x}) \, f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}, \tag{5.18}$$

where,

$$f(\mathbf{d}, \mathbf{X}) = a(d_2 X_2 - b(d_1 X_1)^2 + c(d_1 X_1) - r)^2 \\ + s(1 - t) \cos(d_1 X_1) + s + 5(d_1 X_1), \tag{5.19}$$

$a = 1, \ b = 5.1/(4\pi^2), \ c = 5/\pi, \ r = 6, \ s = 10, \ t = 1/(8\pi)$
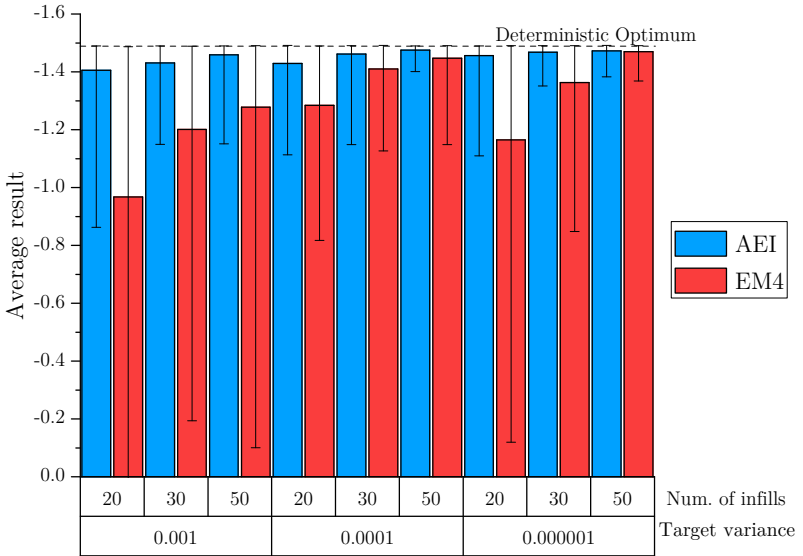
Figure 22 – Infill criterion on one-dimensional noisy function

- Domain: $d_1 \in [-5, 10]$, $d_2 \in [0, 15]$;

- Deterministic global optimum: -16.644 at $\mathbf{d} = \{-3.689, 13.630\}$;

- Noise: $X_i \sim \mathcal{N}(1, 0.1)$; $i = \{1, 2\}$;

- Optimum with 10000 simulations $= -16.6637$.

Figure 23 shows the results for a number of different configurations. Again, it is possible to notice the advantage of using AEI over EM4 for noisier outputs. Comparatively with the one-dimensional experiment, in this case EM4 performance is more pronounced for all configurations.

Overall, it is clear that the AEI modifications make it cope better with noise. This is mainly caused by the penalty parameter used, which includes both the kriging error and noise estimates. Moreover, lowering
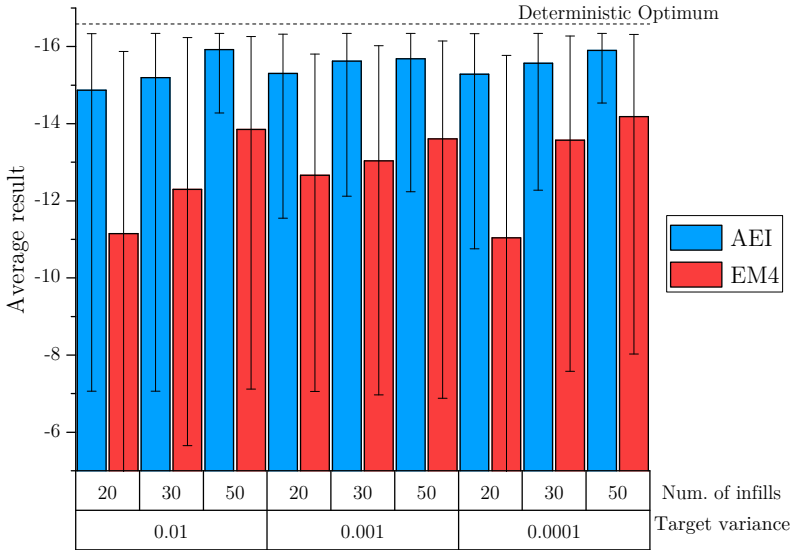
Figure 23 – Infill criterion on two-dimensional noisy function

the target variance makes the response approach the deterministic case where both criteria perform similarly.

5.4.2   Example

In order the better visualize the optimization procedure, an example is presented. The function being optimized is the one-dimensional stochastic Problem 1 function from Equation 5.17. This function is disturbed by a multiplicative noise following a Gaussian distribution. A plot of an observation of the noisy function over the domain can be seen in Figure 24.

At the first stage, a SK model is fitted from a certain number of observations. In this example the initial sample size is five. Each observation is replicated a number of times in order to achieve the desired variance target. Different targets imply in different error es-

Figure 24 – Noisy function

timates. In fact, the more a point is evaluated for a fixed input, the less uncertainty exists about its output. This difference is i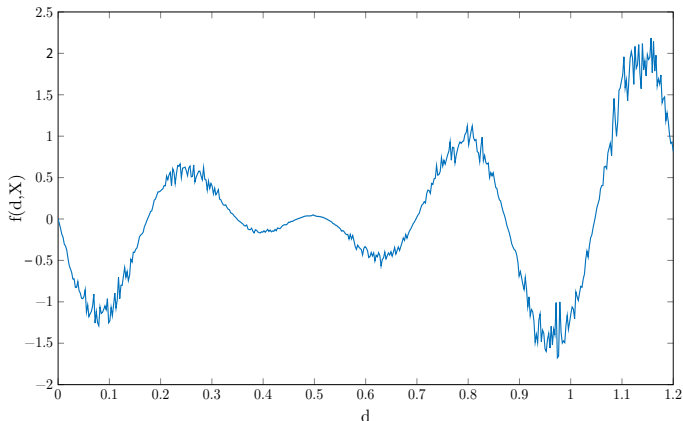llustrated in Figure 25. Figure 25a shows a kriging model based on a variance target of 1.00. Similarly, Figure 25b shows the kriging model resulting from the same initial sample, but in this case with a target variance of 0.01. In the first case, a lower number of simulations is needed. Nevertheless, it is possible to perceive the high uncertainty of the model caused by the noise. This uncertainty is illustrated by the dashed lines that correspond to the interval +s and -s from the prediction. Here s, the so-called Root Mean Squared Error, is a measure of the estimated error computed by taking the square root of the MSE presented in Equation 4.14.

In deterministic optimization, the estimated error is exactly zero at sampled points. However, in the presence of noise it only approaches zero as the number of Monte Carlo replications increases. As a result, the uncertainty is diminished depending on how lower the target variance is. That is precisely the reason why using deterministic infill criteria like EM4 or EI is not so valuable. The sampled points continue to display an estimated error, which is caused by the noise. If there is still uncertainty, the infill criteria will continue to look for a possible improvement and will possibly sample the same point multiple times. This behavior leads

to a stall of the optimization procedure.
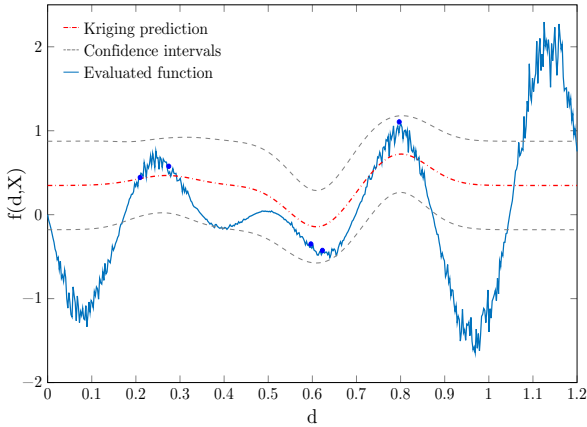
(a) Target var = 1.00
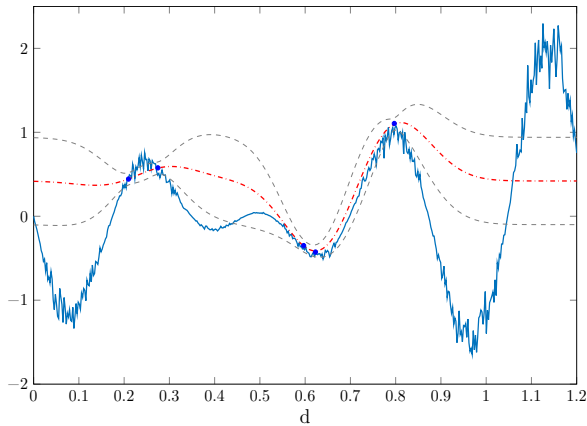
(b) Target var = 0.01

Figure 25 – Different model based on the estimated error

Still, comparing both cases in Figure 25 it can be seen that the RMSE interval is closer to the sampled points in Figure 25b while the it is barely disturbed in Figure 25a. Although a larger number of evaluations is needed in order to achieve a lower target variance, it can be seen that it may improve the prediction. Having a more accurate

description of the function with a lower estimated error on sampled points makes the infill criteria approach the deterministic case which is globally convergent.
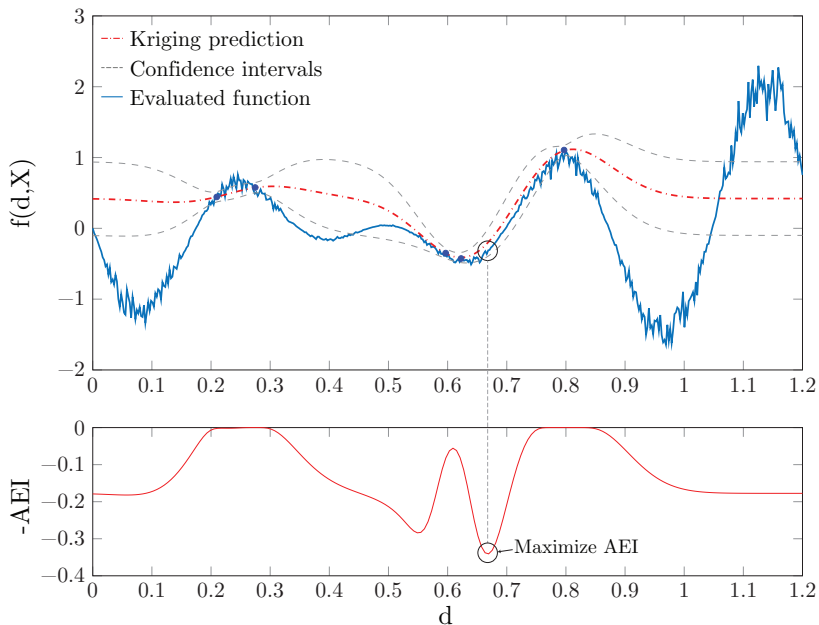
Figure 26 – AEI plot over the domain

After the SK model construction, the next step in the optimization procedure is the refinement of the model by adding infill points. Using the AEI criterion, the point that minimizes the utility function composed of an Expected Improvement term and a penalty term is added to the model. Continuing the example with the lower variance model, Figure 26 illustrates how the AEI infill works. The auxiliary plot presented below the function plot represents the negative of the AEI measure over the domain. The circled point indicates the selected infill point, which is shown to minimize the AEI measure.

Figure 27 shows the progress of the optimization in two moments. In Figure 27a the 5th infill is added to the model. The search still has not

found the optimal valley. This is mainly caused by the lack of knowledge between 0.8 and 1.2, where only predictions were computed. For this reason, it is important to have an initial sample design with the space-fillingness property. For this example, the initial sample was randomly selected. Nevertheless, the search proceeds and after some evaluations on the initial valley the algorithm starts exploring more uncertain regions. By the infill 20, it is possible to see in Figure 27b that the algorithm already sampled around the optimum multiple times. Therefore, the Augmented Expected Improvement is very low and directed in that single region.

When considering the previous model with target variance of 1.00, a different situation occurs. As seen in Figure 29, the optimization process stalls at the first valley found and never explores other regions. This is caused by the lack of information gained by each point added to the model. Each infill is inserted considering the unit target variance. Because the estimated error remains almost the same, the infill criterion becomes highly local, avoiding exploration of potentially better regions. The penalization from AEI only takes effect when s_n^2 is low, which is not the case when the target variance is high. Thus, there is a margin for optimization on the selection of the variance target. A higher target reduces computational cost but can become highly local. A lower target may need higher computational cost but the refined model is more easily exploited. A trade-off between these characteristics must be made in order to achieve a highly efficient optimization.

5.5   Adaptive target

As seen in the previous section, the variance target plays an important role when using the Monte Carlo approach coupled with Stochastic Kriging. Setting it to either too small or too high values may prejudice the optimization. By being too small, the target forces a large number of evaluations on a single point, effectively compromising the computational budget. Nevertheless, setting it too high may stall
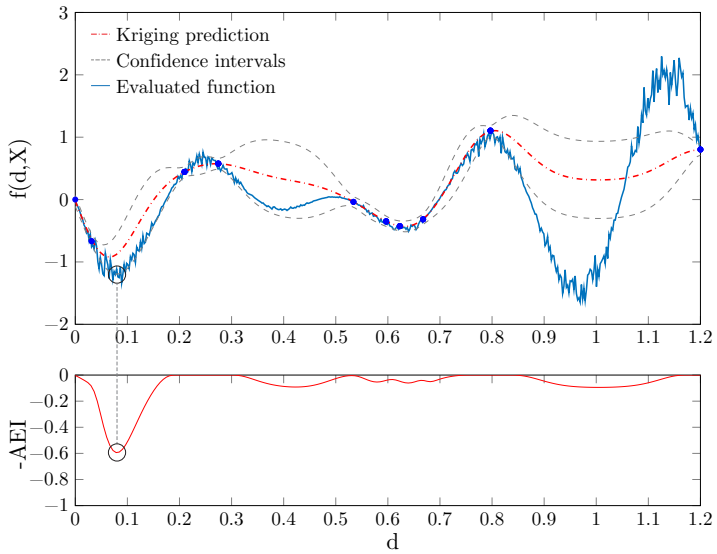
(a) Infill 5
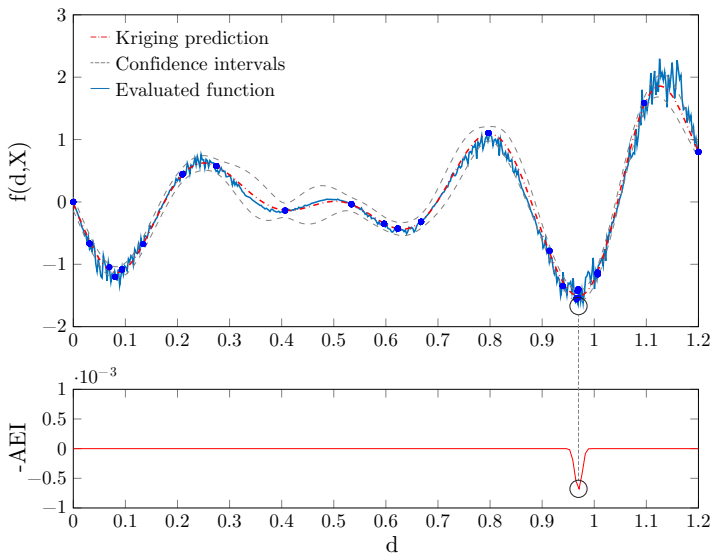
(b) Infill 20

Figure 27 – Refinement of the model seeking the optimal value

the search for the lack of information obtained in each infill. Moreover, another weakness of high variance target is the increase of total number of infill points. It is not desirable to have excessive points in the model. As seen in Equation 4.13, each new model update involves the inversion or Cholesky decomposition of a matrix. This matrix size equals the number of points in the model. This means that with a large number of points the operations on higher dimensional matrices may be too intensive to justify the gains from a lower number of evaluations of the real function. Thus, part of the success on the optimization approach resides in the target selection. However, which target to choose? A main contribution of this study is to offer a solution for this question. The proposed approach consists on an adaptive target selection. It aims to balance the trade-off between accuracy and computational cost. The idea is similar to an exploration versus exploitation aspect of any global optimization procedure. This approach starts with a high target with relatively low computational cost and gradually reduces it after the global trends of the functional surface has been explored.

The EGO algorithm starts, as usual, with the creation of the sampling plan. All initially sampled points are evaluated a single time, ignoring the default target variance. After the construction of the kriging model with the noisy sample plan, the infill stage begins. Each new infill point is added to the model after being simulated up to the corresponding target variance. A fluxogram of the optimization process, including the proposed adaptive target selection is shown in Figure 28.

The adaptation is parametrized by the number of points in the model contained in a hypercube centered in the infill point. When the infill is located within an unsampled region, its target variance is the default higher value. Moreover, when the infill is located in a region with existing sampled points, a lower target variance is used. This is done to allocate more computational budget on regions that need to be exploited. Thus, it indicates the purpose of the infill. Isolated infill points focus on exploring the landscape, where higher precision is not needed. When they start to group up, the focus changes to surface exploitation.
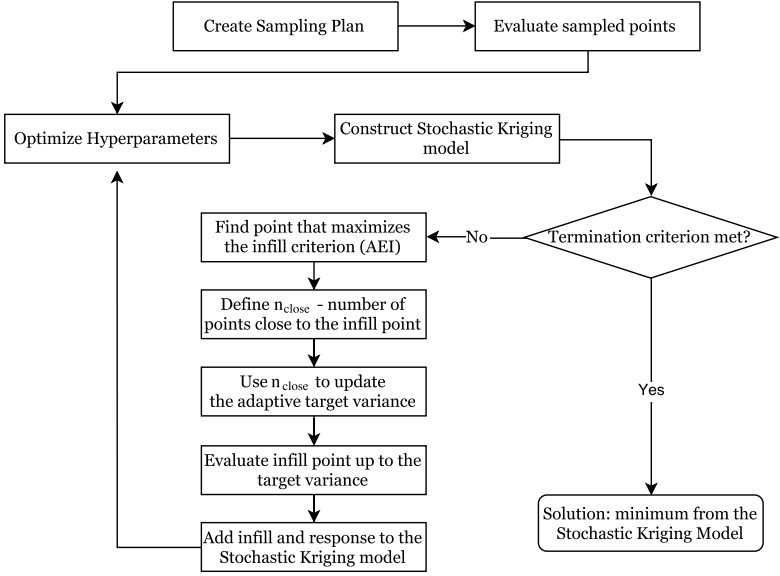
Figure 28 – Fluxogram of the algorithm

In this situation, the variance target value is diminished, increasing the model precision. By doing so, it also avoids the clustering of multiple inaccurate points that causes the stalling observed in Figure 29.

Supose **q** is the point that maximizes the infill criteria. The expression used to calculate the adaptive target value for the EGO iteration is:

$$\text{target}_{\text{adapt}} = \frac{\text{target}}{\exp(1/2 + 1/2 \cdot n + 9/19 \cdot n_{\text{close}} - 1/100 \cdot n_{\text{close}} \cdot n)}. \tag{5.20}$$

This expression involves the dimension of the problem n and n_{close}, the number of close points already sampled, which is calculated as:

$$n_{\text{close}} = \sum_{j=1}^{n_s} \left\{ 1 : \left| (q_i - d_i^{(j)}) \right| < r_{\text{hc}}, \quad \forall i \in \{1, 2, ..., n\} \right\}, \tag{5.21}$$
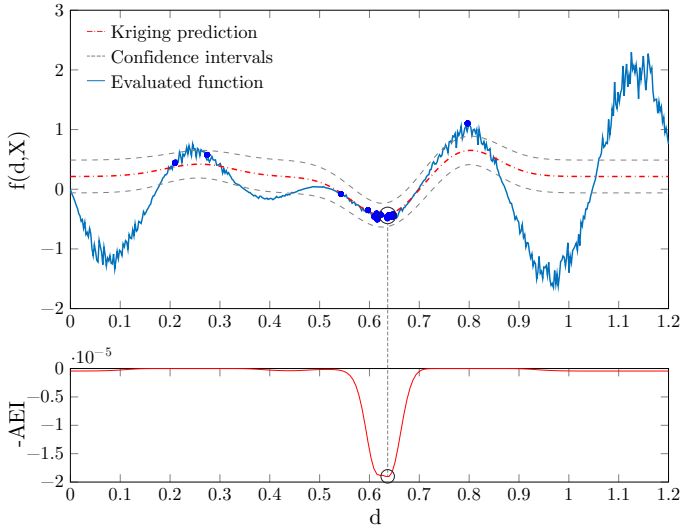
Figure 29 – Stalling at infill 20 with target variance 1.00

where r_{hc} corresponds to half the side-length of a hypercube around the infill point selected. This value is set to 0.1 for all benchmarks. This value is independent of the problem bounds because the domain is normalized to a unit hypercube before the metamodel creation.

Figure 30 presents the plot of the variance target decay function for different problem dimensions. The reasoning behind the construction of this function is twofold. First, it displays an exponential decay of the target value. Second, the curves show a higher decrease in the target value as the problem dimension increases.

The choice of an exponential decay seems to be the most intuitive considering how the number of function evaluations increases and the error decreases. This aspect was observed in Figure 21. If a linear model was used instead, the target would decrease very slightly with few closer points. Yet, it would drop abruptly as n_{close} approached its maximum value. This would cause, initially, an unnecessary number of infill points added to the model without a reasonable gain of information. Further in the optimization, the target would drop abruptly resulting in a large
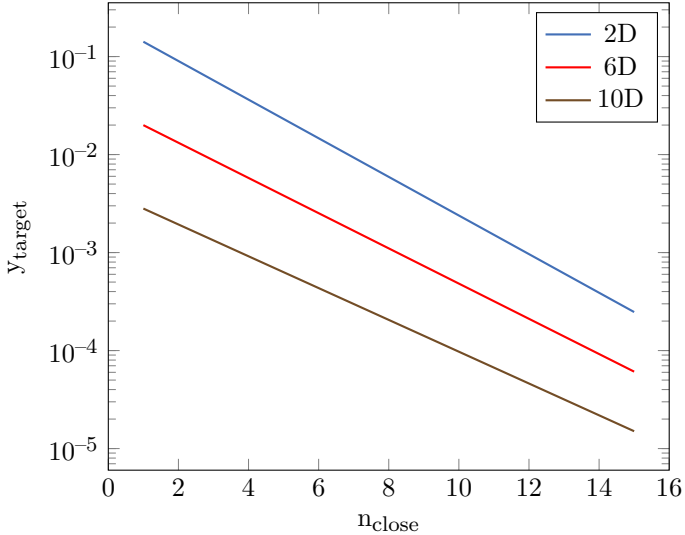
Figure 30 – Target decay varying with problem dimension

number of evaluations. With the proposed approach, the targets starts high initially and is progressively lowered when points begin to cluster around an optimum value.

The other aspect of the expression relates to the problem dimension. With low dimensional problems, the input space is relatively small. It becomes easier for the infill points to cluster. Thus, the targeting decay cannot be too aggressive at risk of expending too much computational resources. At higher dimensions n_{close} does not increase so fast. Thus, it allows for a more significant target decay.

As an example, Figure 31 presents the benchmark results comparing constant and adaptive target for a one-dimensional function. The benchmark function is Problem 1, which was already presented in Equation 5.17. The results represent the average of 100 independent runs with error bars displaying the 5 and 95 percentiles. Further numerical results for higher dimensional problems will be presented in chapter 6. For this one-dimensional case the adaptive approach at 100 evaluations is already very close to the deterministic optimum. By selectively ex-

pending the computational budget, the proposed approach is able to converge faster to the basin of attraction. This improvement is further noticeable when increasing the function noise, as seen in Figure 31b.
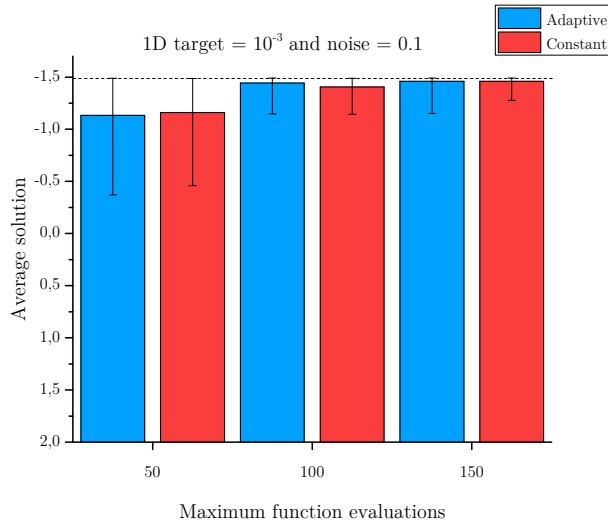
(a) Noise stardard deviation at 0.1
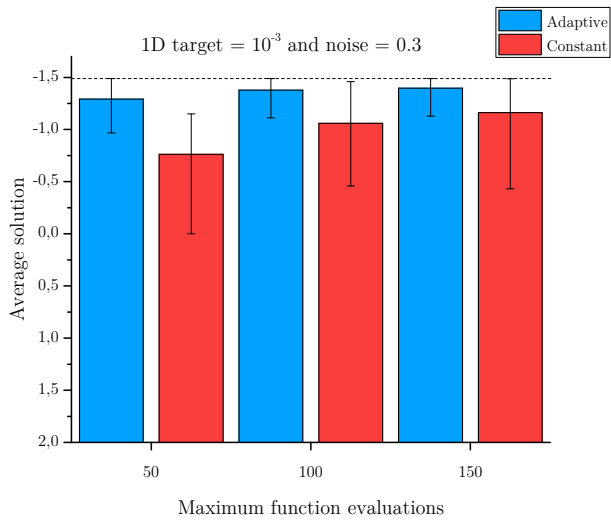
(b) Noise stardard deviation at 0.3

Figure 31 – 1D Function - Limited number of evaluations

6 Numerical results

For the testing of several different approaches in kriging optimization and evaluation of problems that are discussed in this chapter a specialized computing environment was developed. Although third-party packages such as UQLab (MARELLI; SUDRET, 2015) and DiceKriging (ROUSTANT *et al.*, 2012) exist, their use was discarded for the lack of desired features and/or control. Moreover, the development of this independent environment enhanced the author's understanding on the underlying topics. It also offered flexibility by the implementation of multiple approaches and set the basis for future developments and implementations. The implementation focus was on Kriging metamodeling using EGO's optimization framework. It was created using Matlab (THE MATHWORKS, INC., 2015) and was initially based on snippet codes from Forrester *et al.* (2008) for sampling plan creation and kriging prediction. EGO was implemented along with all methods from Table 2.

All methods, including the enhanced version of Method 4 were tested and validated by the optimization of benchmark functions. One-stage approaches, corresponding to Methods 6 and 7, were tested but later discarded due to the excessive computational resources needed to complete each optimization. Computing time was increased by a factor of 2 to 10 if compared to two-stage approaches. It relates to the fact that such procedure results in a parameter estimation problem with twice the dimension of the original. From one-stage approaches, Enhanced Method 4 performed the best and was the selected method to continue with the preliminary experiments. It showed good balance of exploration and exploitation, computational efficiency and has parallelization capabilities, as explained in section 4.6.

For the sampling plan creation step, a $n_s = 7n$ rule was used for all problems solved with EGO. This means that the initial sampling plan size started with seven times the number of dimensions of the problem.

Although some authors, such as Forrester e Keane (2009), suggest using $n_s = 10n$, both rules were tested and $n_s = 7n$ was considered enough regarding the final optimization result.

For the MLE optimization step, four metaheuristic algorithms were evaluated: Genetic Algorithm (GA) (GOLDBERG, 1989), Probabilistic Restart (PR) (LUERSEN; RICHE, 2004), Backtracking Search Algorithm (BSA) (CIVICIOGLU, 2013; SOUZA; MIGUEL, 2016; SOUZA *et al.*, 2016) and Particle Swarm Optimization (KENNEDY; EBERHART, 1995). They were tested for different combinations of parameters on different surfaces generated by Kriging approximations. From the conducted tests, BSA has shown too much emphasis in exploration and was slow to converge. PR and GA presented similar convergence but the execution time of GA was generally faster. Yet, PSO with the default parameters from the Matlab (THE MATHWORKS, INC., 2015) implementation obtained slightly more accurate results than GA and was generally the fastest method employed. Thus, it was selected as the default parameter optimization method.

In the following sections, the performance of EGO is evaluated by the use of benchmark test functions. These functions are stochastic. Here, the measure of the computational cost is the number of evaluations of the objective function. Since the benchmarked functions are not expensive to evaluate, it becomes pointless to discuss wall clock timings. Thus, this information is not presented. In section 6.1 the noisy output is addressed with Stochastic Kriging with adaptive variance target while section 6.2 aims to show how it compares against another algorithm.

6.1   EGO performance on stochastic benchmark functions

The main application of the metamodeling procedure deals with functions involving random variables, as discussed in subsection 3.10.1. Therefore, stochastic versions of the benchmark functions are used. This is accomplished by inserting a multiplicative random noise to the input variables. The noise follows a Gaussian distribution with a specified

standard deviation for each problem analyzed. These functions possess a single global optimum and most of them are multimodal, making them good candidates to assess the method's exploration and exploitation capabilities.

In this first preliminary study, deterministic kriging is used. However, parameter variability should be taken into consideration. This is addressed with the approach presented in section 5.1, where each point is replicated a number of times and the average output is accepted similarly as a deterministic response. For the presented problems, 100 replications are used when computing the estimated $\hat{\mathrm{E}}[f(\mathbf{d}, \mathbf{X})]$. Precision of this estimate increases with the number of replications, as seen in subsection 3.3.1. For the considered problems, this number of reevaluations is verified as enough to reduce the variability of the expected value estimate.

Moreover, it is important to stress that the optimization procedure presented depends on random quantities. Therefore, the resulting value is not deterministic and may change after running the algorithm again. For this reason, when dealing with stochastic algorithms, it is appropriate to present statistical results over a number of algorithm runs. Thus, for each problem average and standard deviation over multiple runs are presented along with the minimum found. Moreover, considering the response variability, for each problem the expected value of the deterministic global optimum considering 10000 simulations from the stochastic function is presented. It is obtained by taking the average output of those simulations where each evaluation uses the same known input, which minimizes the deterministic function. This value is different from the deterministic optimum and depends on the characteristics of the random variables present in the formulation.

Figures 33, 34, 36 show benchmark results. The benchmark compared constant with adaptive targeting for a limited number of function evaluations. Each figure represents the average results from multiple runs of the algorithm. The number of independent runs is 100 for the 1D function and 40 for the 2D and 6D functions. The benchmarks are

conducted considering two different noise values for each function. The iteration procedure stopped after the infill step reached a number of evaluations larger than the maximum permitted. Error bars from each figure are presented in order for the result variability to be observed. Each bar represents one standard deviation. Moreover, in order to reduce the influence of the initial sampling plan, the same plan is used for both constant and adaptive targeting.

6.1.1   Branin tilted

Writing the optimization problem as the minimization of an integral it becomes:

$$\min f_{\text{obj}}(\mathbf{d}) = \int f(\mathbf{d}, \mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}, \tag{6.1}$$

where

$$f(\mathbf{d}, \mathbf{X}) = a(d_2 X_2 - b(d_1 X_1)^2 + c(d_1 X_1) - r)^2 + s(1-t)\cos(d_1 X_1) + s + 5(d_1 X_1), \tag{6.2}$$

$a = 1, \; b = 5.1/(4\pi^2), \; c = 5/\pi, \; r = 6, \; s = 10, \; t = 1/(8\pi).$

This is a modified Branin function where a term $5d_1$ is added to the function. This modification forces the existence of a single global optimum value. The plot of the deterministic version of this function is shown in Figure 32.

- Domain: $d_1 \in [-5, 10], \; d_2 \in [0, 15]$;

- Deterministic global optimum: -16.644 at $\mathbf{d} = \{-3.689, 13.630\}$.

In the second function seen in Figure 33 both approaches improved the solution similarly with the increase of available evaluations. Nevertheless, the adaptive targeting presented a mean closer to the optimum for all maximal evaluations targets.
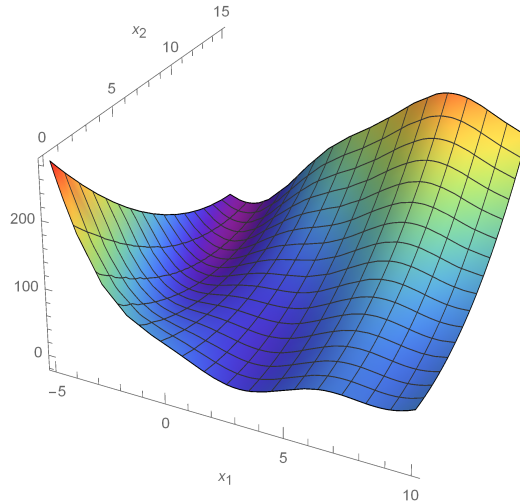
Figure 32 – Function plot - Branin tilted

6.1.2 Hartman 6D

This is multi-modal benchmark function with six local minima. Writing the stochastic version of the problem as minimization of an integral becomes:

$$\min \mathrm{f}_{\mathrm{obj}}(\mathbf{d}) = \int \mathrm{f}(\mathbf{d}, \mathbf{x}) \, f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{6.3}$$

where

$$\mathrm{f}(\mathbf{d}, \mathbf{X}) = -\sum_{i=1}^{4} \alpha_i \, \exp\left(-\sum_{j=1}^{6} \mathrm{A}_{ij}(\mathrm{d}_j \mathrm{X}_j - \mathrm{P}_{ij})^2\right), \tag{6.4}$$

with

$$\alpha = (1.0, 1.2, 3.0, 3.2)^{\mathrm{T}}, \quad \mathrm{A} = \begin{pmatrix} 10 & 3 & 17 & 3.50 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{pmatrix}$$
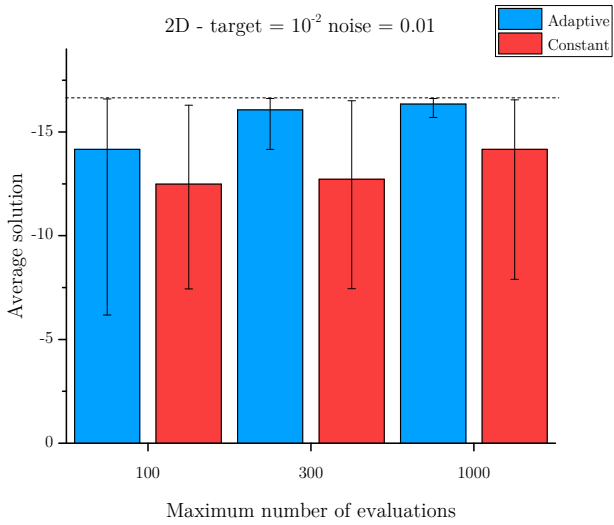
(a) Noise stardard deviation at 0.1
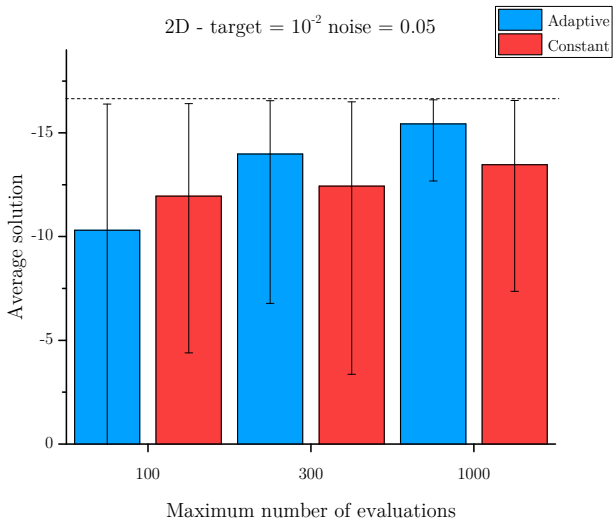
(b) Noise stardard deviation at 0.3

Figure 33 – 2D Function - Limited number of evaluations

and

$$
P = 10^{-4} \cdot \begin{pmatrix}
1312 & 1696 & 5569 & 124 & 8283 & 5886 \\
2329 & 4135 & 8307 & 3736 & 1004 & 9991 \\
2348 & 1451 & 3522 & 2883 & 3047 & 6650 \\
4047 & 8828 & 8732 & 5743 & 1091 & 381
\end{pmatrix} .
$$

- Domain: $d_i \in [0,1]$ for $i = \{1..6\}$;

- Deterministic global optimum: -3.32237 at $\mathbf{d} = \{0.20169, 0.150011, 0.476874, 0.275332, 0.311652, 0.6573\}$.

For the second function, the benchmark in Figure 34 presented the largest difference between both approaches. It can be noted that the adaptive target selection behaves better when noise is increased. As there is more noise, it takes more evaluations to reach a certain target and it becomes prohibitive to expend the computational budget on precise exploration points. That causes the difference seen in the second benchmark. Moreover, the adaptive target setting may enable the optimization to run longer than the constant counterpart may. Without the early termination proposed in the benchmark, a small constant target would try to exploit the whole domain instead of only the promising regions.

6.1.3 Levy 10D

The last benchmark problem is a n-dimensional multimodal stochastic function. As usual, the problem is written as an integral minimization:

$$
\min f_{\mathrm{obj}}(\mathbf{d}) = \int f(\mathbf{d}, \mathbf{x}) \, f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{6.5}
$$
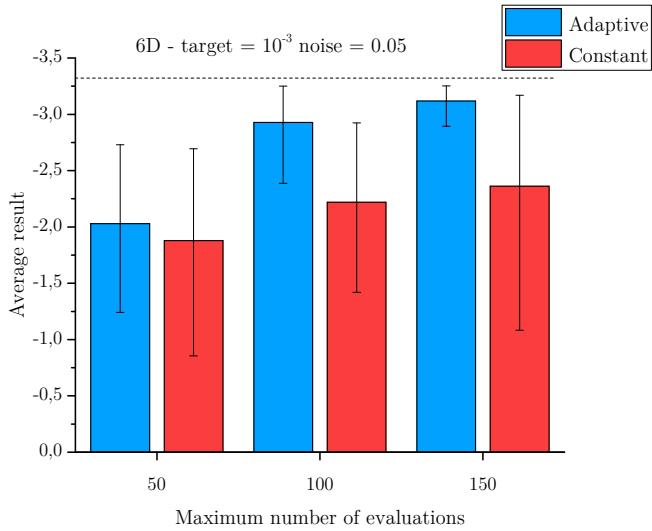
(a) Noise stardard deviation at 0.05
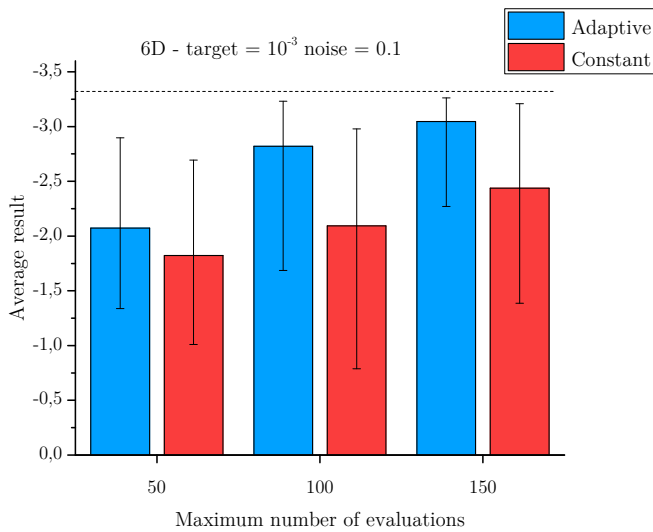
(b) Noise stardard deviation at 0.1

Figure 34 – 6D Function - Limited number of evaluations

where

$$f(\mathbf{d}, \mathbf{X}) = \sin^2(\pi w_1) + \sum_{i=1}^{n-1} (w_i - 1)^2 [1 + 10\sin^2(\pi w_i + 1)] + (w_k - 1)^2 [1 + \sin^2(2\pi w_n)],$$

(6.6)

with $w_i = 1 + \frac{d_i X_i}{4}$ for $i = \{1, ..., n\}$. The benchmark was evaluated considering a dimension of 10.

- Domain: $d_i \in [-10, 10]$ for $i = \{1..10\}$;

- Deterministic global optimum: 0 at $\mathbf{d} = \{1, 1, ..., 1\}$.
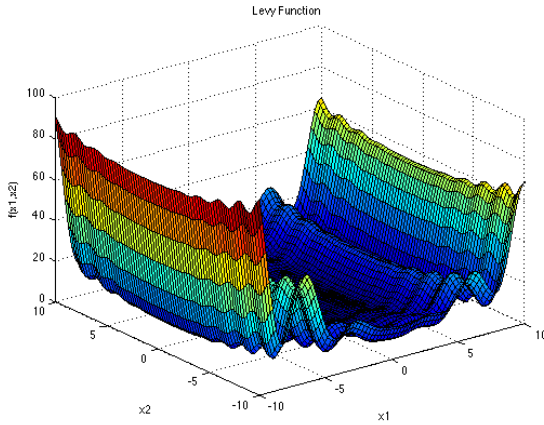
Figure 35 – Levy function for the 2-dimensional case

The benchmark results are presented in Figure 36. Similarly as the previous tests, the adapative targeting obtains better average results than the constat counterpart. Moreover, increasing the maximum number of function evaluations consistently decreases the variability of results, represented by the error bars. The method obtain reasonable results using a relatively small number of function evaluations even considering a very large 10-dimensional search space.
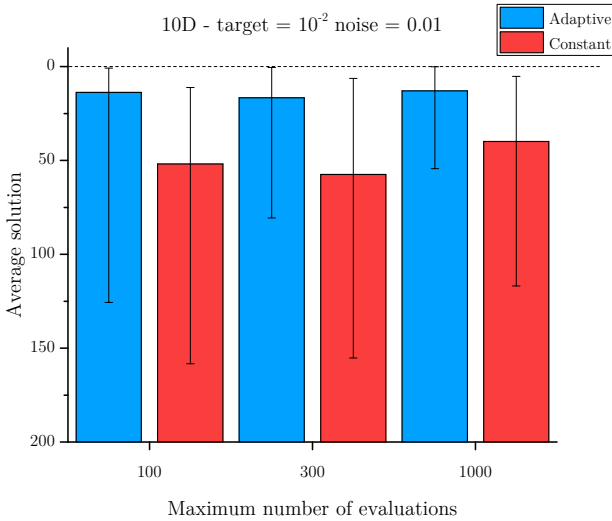
Figure 36 – 10D Function - Limited number of evaluations

6.2   Comparative against another approach

It became clear from previous sections that the EGO algorithm described might be successfully employed in the optimization of noisy functions. In this section, the presented method is compared against another optimization procedure, which also focuses on a lower computational cost. The chosen algorithm for the comparative is called Globalized Bounded Nelder–Mead (GBNM). It was developed by Luersen e Riche (2004) and consists of a probabilistic restart procedure on a Nelder-Mead local search (NELDER; MEAD, 1965).

A question remains as how to employ and fairly compare GBNM with EGO as it does not possess the same framework for dealing with noisy functions. It is common practice, when there is an output noise, to construct a mean convergence plot. From this plot, the number of evaluations needed for the mean to converge is identified. This number is then used as constant replication number for every sampled point. Figure 37 shows an example from a 1D noisy function. The mean

becomes stable after approximately 100 simulations. This is the same function compared in Figure 31. If 100 simulations were employed for each new point in the model, the computational cost would be much higher than the 150 upper bound used in the proposed approach.
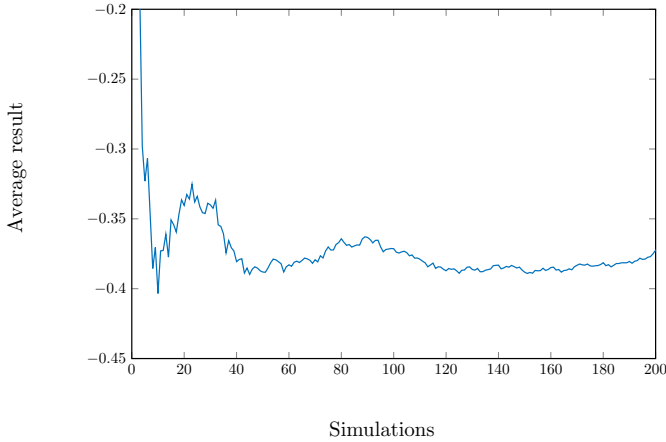
Figure 37 – Mean convergence for 1D noisy function

In Figure 38 a comparison is presented between GBNM and EGO with constant target variance. The problem being optimized is Problem 1. This is the same problem discussed in, section 5.5.

In this comparison the average solutions of 50 runs are presented for two different noise levels. The stopping criterion for both approaches is a maximum number of function evaluations. For the noise 0.1 both had a maximum of 150 evaluations. Error bars are presented with the 5 and 95 percentile values. It can be seen that EGO displays a reduced dispersion of the results. Moreover, by iteratively refining the metamodel the best result found matches the optimum value. The same does not happen when considering GBNM. The approach takes the minimum of multiple restarts considering the constant target. In some runs, due to the noise involved, the minimum found is less than the deterministic optimum.

For the second noise level, the maximum number of evaluations
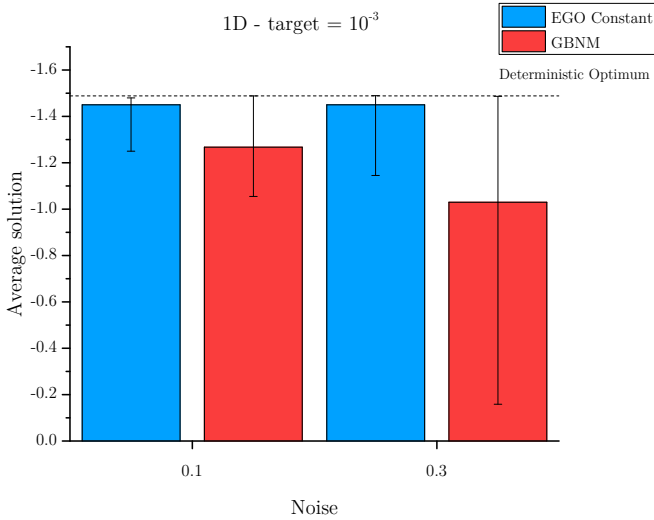
Figure 38 – Comparison between two approaches with 1D function

differs in each method. EGO maintains the 150 evaluations while GBNM increases the maximum to 1000 evaluations. This is done because even using the least computationally demanding parameters with GNBM, the number of evaluations cannot become lower. By comparing both approaches in the second noise level, it can be seen that GBNM remains with a large spread and that EGO produces results closer to the deterministic optimum on average.

6.3   Advantage against quadratures

Another benchmark is conducted in order to evaluate the performance of the proposed integral evaluation with adaptive targets against a quadrature approach. Reiterating, the problem being solved is written as:

$$\min f_{\text{obj}}(\mathbf{d}) = \int f(\mathbf{d}, \mathbf{x}) \, f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}, \qquad (6.7)$$

The benchmark considers two approaches. First, the minimization of the integral using EGO with the AEI infill criteria and adaptive targeting. Second, the same problem but instead using EGO with the deterministic EM4 infill criteria and evaluating the integral using a quadrature. Since the characteristics of the random variables in X are known for the problems analyzed, it is possible to compute the integral by the definition in Equation 6.7 instead of resorting to simulation. Therefore, the goal is to verify if quadratures can compete with the proposed solution in terms of performance and applicability.

For consistency reasons, the problems solved here are the same 1D and 2D functions already discussed in subsections 5.4.1 and 6.1.1. In the following benchmark, a Gaussian quadrature algorithm is employed. It follows the algorithm described in Davis (1984), using a product rule for multidimensional integration.

In both problems, the maximum number of infills is set to 30. From this setting, some experiments were conducted in order to verify the number of quadrature points that resulted in a similar response compared to a fixed Monte Carlo variance target. Considering the Gaussian quadrature procedure, the grid of points that is evaluated is a function of the chosen number of quadrature segments in each dimension.

For the 1D problem, called Problem 1, the noise is applied multiplicatively on the inputs by a random variable with unit mean and standard deviation 0.3, that is, $X \sim \mathcal{N}(1, 0.3)$. In this case, it is possible to compare both approaches under the same maximum number of evaluations. In the first situation 5 quadrature points are used. For the second, the number of quadrature points is increased to 20. The AEI stopping criterion is set to the same number of evaluations of the EM4 quadrature.

Figure 39 presents the results for the 1D case, where error bars represent the 5 and 95 percentiles from 50 independent runs. It can be seen that AEI Adaptive outperforms EM4 with quadrature. It can achieve results closer to the deterministic optimum with less dispersion.

When the maximum number of evaluations is increased to 600, both approaches lead to improved results. However, it can be seen that even by increasing the total budget, EM4 with quadrature performs worse than AEI with the initial budget.
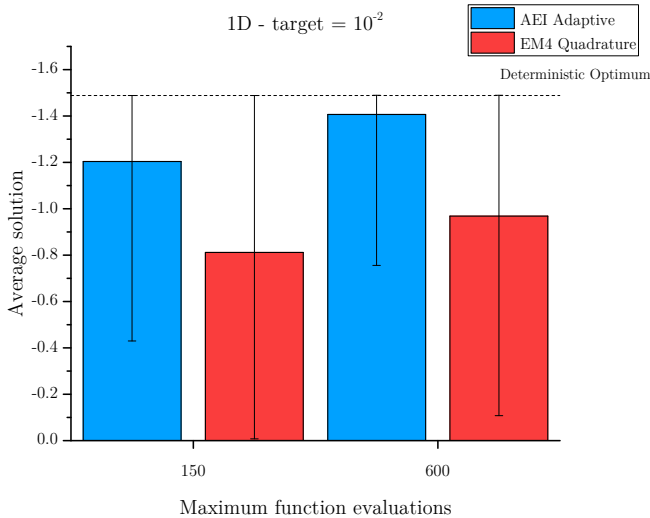
Figure 39 – 1D Quadrature

For the second problem, called Branin tilted, the noise multiplied to each input comes from the two- element Gaussian random vector $\mathbf{X} \sim \mathcal{N}(1, 0.01)$. Again, AEI with adaptive target heavily outperforms the quadrature approach, as seen in Figure 40. In the first situation, AEI is limited to 1000 evaluations while EM4 uses 20 quadrature segments in each dimension. Considering this number of segments in each dimension for this 2D problem results in a grid with $20^2 = 400$ different points to be evaluated on each infill step. The large spread and the distant average from the optimum indicate that the quadrature may not have enough precision. The second situation considers an increased number of quadrature segments. It is increased from 20 to 50 for each dimension, which improves the quadrature precision. However, it also

heavily increases the total number of evaluations considering a fixed number of infill points. In this situation, EM4 with quadratures can only obtain a result similar to AEI Adaptive by increasing 75 times the computation cost.
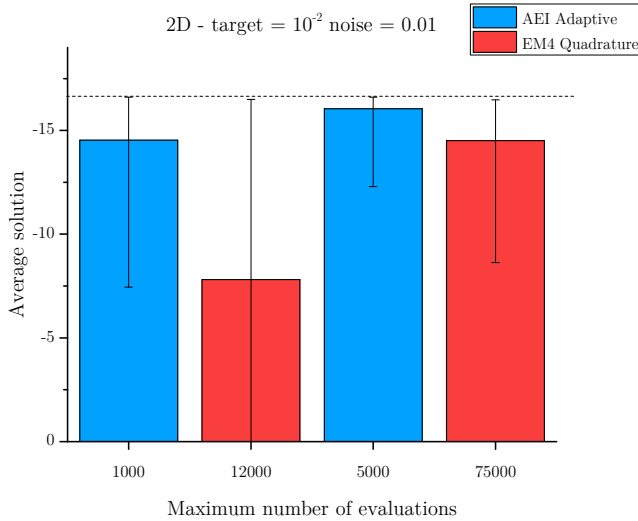
Figure 40 – 2D Quadrature

The trend of a higher number of evaluations required for the quadrature procedure is further increased when considering higher dimensional problems. Suppose one wants to solve a six dimensional problem. The number of evaluations would depend on the number of divisions on the hyper-rectangular region that is being integrated. If, for example, 20 segments for each dimension were used. This number already proved to be insufficient for the level of noise present in these benchmarks. Nevertheless, $20^6 = 64000000$ evaluations would be required. Moreover, that would be only one step in the EGO algorithm, thus being replicated for every infill point that is added to the model. It clearly becomes unfeasible for higher dimensions. Thus, AEI augmented with the adaptive target variance remains as a more applicable and

efficient approach to solving the proposed class of problems.

6.4   Application: Tuned-Mass Dumper system optimization

As discussed in section 5.3, there are several engineering problems where the proposed integral minimization could be employed. In this section, one application in the field of structural dynamics is shown. It involves the seismic vibration control of a structure. In order to achieve the best performance, one must determine the optimal parameters of a Tuned-Mass Dumper (TMD) that shall be installed on a structure subjected to seismic excitation. A structure from literature is analyzed and the TMD parameters are subject to the optimization procedure.

The TMD is a mass-spring-damper system, which aims to reduce the vibrational energy that is transfered to the primary members of the structure. There is a growing interest on the study of TMD to date. This is due to the fact that amongst the numerous passive control techniques, the TMD is one of the simplest and most reliable control devices (CHAKRABORTY; ROY, 2011).

In the TMD design optimization process, the design variables are usually the stiffness and damping of the energy dissipation system. Considering an n-story MDOF linear building structure with a mass damper installed at the top floor, the equation of motion of the combined system subjected to ground acceleration can be written as follows:

$$\mathcal{M}\ddot{\mathbf{z}}(t) + \mathcal{C}\dot{\mathbf{z}}(t) + \mathcal{K}\mathbf{z}(t) = -\mathbf{m}\,\ddot{z}_g(t), \tag{6.8}$$

where $\mathbf{z}$ is the (n+1) dimensional response vector representing the displacements relative to the ground:

$$\mathbf{z}(t) = \{z_1, z_2, ..., z_n, z_d\}, \tag{6.9}$$

$\ddot{z}_g$ is the ground acceleration, $\mathbf{m}$ is mass vector:

$$\mathbf{m} = \{m_1, m_2, ..., m_n, m_d\}, \tag{6.10}$$
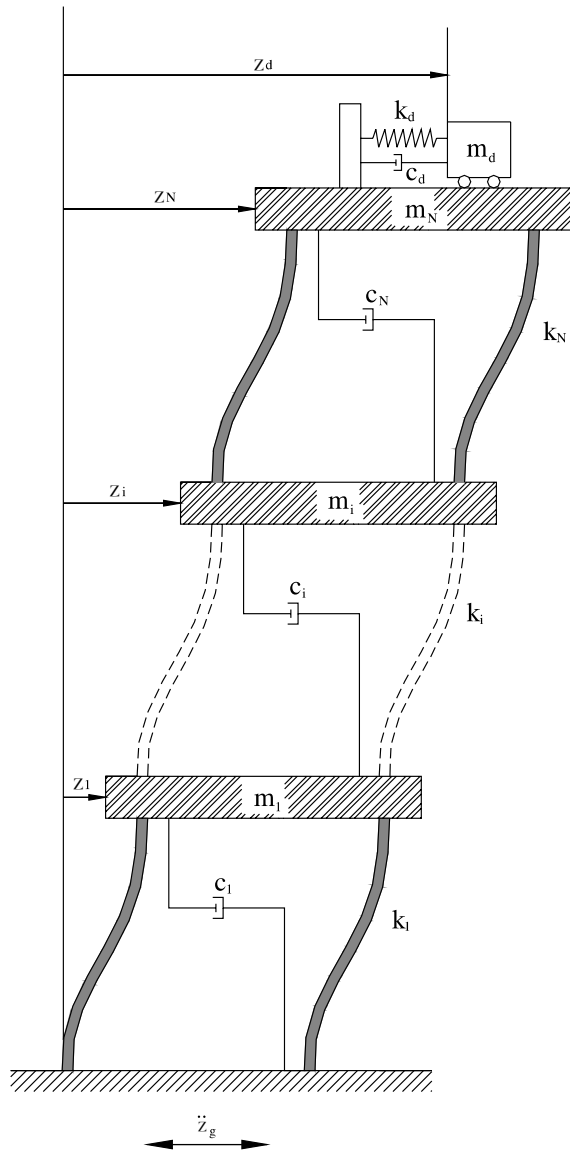
Figure 41 – TDM building

and $\mathcal{M}$, $\mathcal{C}$ and $\mathcal{K}$ are matrices corresponding to the mass, viscous damping and the stiffness of the structure, respectively. They can be written as:

$$\mathcal{M} = \begin{bmatrix} m_1 & & & & \\ & m_2 & & & \\ & & \ddots & & \\ & & & m_n & \\ & & & & m_d \end{bmatrix} \tag{6.11}$$

$$\mathcal{C} = \begin{bmatrix} (c_1 + c_2) & -c_2 & & & & \\ -c_2 & (c_2 + c_3) & -c_3 & & & \\ & -c_3 & & & & \\ & & \ddots & -c_n & & \\ & & -c_n & (c_n + c_d) & -c_d \\ & & & -c_d & c_d \end{bmatrix} \tag{6.12}$$

$$\mathcal{K} = \begin{bmatrix} (k_1 + k_2) & -k_2 & & & & \\ -k_2 & (k_2 + k_3) & -k_3 & & & \\ & -k_3 & & & & \\ & & \ddots & -k_n & & \\ & & -k_n & (k_n + k_d) & -k_d \\ & & & -k_d & k_d \end{bmatrix} \tag{6.13}$$

where m_i is the mass of the i-th floor, m_d is the mass of the damper, c_i is the damping of the i-th floor, c_d is the damping of the damper, k_i is the stiffness of the i-th floor, k_d is the stiffness of the damper, z_i is the displacement of the i-th floor relative to ground and z_d is the displacement of the damper relative to ground.

The structure being analyzed here is a classic example from literature. It has already been studied by several authors (HADI; ARFIADI, 1998; LEE *et al.*, 2006; LOPEZ *et al.*, 2015). It consists of a ten-story

shear frame structure as illustrated in Figure 41. Each story has an assumed height of 3 meters, resulting in a total height (h) of 30 meters. During the event of an earthquake, the displacement of the top floor should not exceed a certain barrier.

The goal in this problem is to minimize the expected value of the failure probability of the structure over the design parameters domain considering different barrier levels. The problem is solved for three different barrier levels, which are imposed as fractions of total height. The reliability index β is used to measure the reliability of the system. A higher β implies a lower failure probability. Thus, considering the design vector $\mathbf{d} = [k_d, c_d]$ and the stochastic parameters vector $\mathbf{X}$, the problem can be stated:

$$\text{Find} : \mathbf{d}$$
$$\text{which minimizes} : \int -\beta(\mathbf{d}, \mathbf{x}) \, f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} \qquad (6.14)$$
$$\text{subject to} : d_n^{\min} \leq d_n \leq d_n^{\max}, \quad n = \{1, 2\}.$$

The upper bound and the lower bound value of the stiffness (k_d) and damping coefficient (c_d) of the TMD are 0-4000 kN/m and 0–1000 kNs/m.

To determine the optimum TMD parameters, a stationary earthquake excitation is assumed, which can be modeled as a white noise signal with constant spectral density, S_0, filtered through the Kanai-Tajimi spectrum (KANAI, 1957; TAJIMI, 1960). The power spectral density function is given by:

$$s(\omega) = S_0 \left[\frac{\omega_f^4 + 4\omega_f^2 \xi_f^2 \omega^2}{(\omega^2 - \omega_f^2)^2 + 4\omega_f^2 \xi_f^2 \omega^2} \right] \qquad (6.15)$$

where ξ_f and ω_f are the ground damping and frequency, respectively. Their values are adopted as $\xi_f = 0.6$, $\omega_f = 37.3 \text{rad/s}$ (MOHEBBI *et al.*, 2013). The term S_0 acts as a scaling factor and in this context represents the amplitude of the bedrock excitation spectrum. Its value is adopted as $S_0 = 1 \times 10^{-3} \text{m}^2/\text{s}^3$. This combination of parameters corresponds to

an earthquake with 0.38g peak ground acceleration on a medium firm soil (CHEN; LUI, 2005).

Additionally to the seismic noise signal, another source of uncertainty is taken into account. The structural parameters from the floors and the TMD are considered random variables with a Gaussian distribution. Their assumed mean values and Coefficient of Variation (C.o.V) are shown in Table 3. In Figure 42 it can be seen how irregular the surface becomes when considering the output without simulation, *i.e.* the result of a single evaluation for each input.
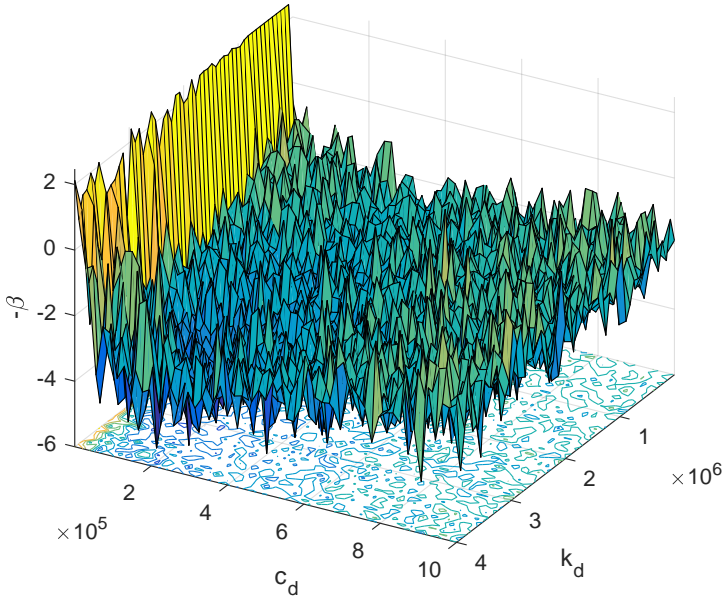
Figure 42 – Noisy reliability surface over design variables range

For the calculation of the reliability index the design life time (t_D) of the structure is considered to be 50 years. Moreover, the rate of arrival (v) of earthquake events is of 1 every 10 years and each event had the duration t_E of 50 seconds. Given this information, it is possible to calculate β considering the time-dependent reliability of oscillators. The underlying concepts as well as the formulation and assumptions that

Table 3 – Statistical properties of structural parameters

		Mean	C.o.V. [%]
Stifness [N/m]	Story	650.0×10^6	15
	TMD	k_d	15
Mass [kg]	Story	360.0×10^3	05
	TMD	108.0×10^3	05
Damping [Ns/m]	Story	6.20×10^6	25
	TMD	c_d	25

lead to the reliability index result are further explored in Appendix A.

Moreover, in order to obtain a faster objective function evaluation, the problem is rewritten using the state space formulation. Appendix B details the solution procedure based on the Lyapunov equation. By solving the Lyapunov equation of the problem for the covariance matrix, it is possible to extract the variance of the displacements and velocities from each degree of freedom. Therefore, it becomes straightforward to calculate the standard deviation of those quantities, which are in turn needed for the reliability index computation.

After having described the problem and the aspects leading to its solution, the problem nature is emphasized. The problem presented here is strictly academic. Aspects of the real structure are not taken into account and simplifications were made such as: assuming the structural response linear elastic, assuming that the excitation comes from a stationary process, assuming that the random variables of each floor are uncorrelated, *etc.* Nevertheless, it remains as a fairly complex problem comprehending the engineering fields of optimization, control, dynamics and reliability.

Table 4 shows the optimization results using the proposed Stochastic Kriging approach with adaptive variance targeting. The stopping criterion of the algorithm is a maximum number of function evaluations of 1000. Three cases labeled (a), (b) and (c) are considered employing

different barrier levels. The stiffness (k_d) and damping (c_d) are displayed along with the corresponding β_{max} found. Additionally, the β_{mean} result is shown, which represents the average of 25 independent runs of the algorithm. For comparison, the reliability index for the uncontrolled case is also displayed. It can be seen that decreasing the barrier level results in smaller values of β, thus increasing the failure probability. Notable increases in β were achieved. Taking for example the case (b), the system reliability increases from 1.37 without TMD to 4.24 by using the TDM with the reported parameters. It terms of failure probability it means decreasing P_f from 8.5×10^{-2} to 1.1×10^{-5}. Looking at case (c), the structure that would certainly fail without TMD achieves a reliability index of 2.48 (6.6×10^{-3} failure probability) when using the optimized TMD parameters. Moreover, β_{mean} results remained close to β_{max}. It shows that the proposed approach could obtain reasonable results in multiple runs despite all the random parameters and limited number of function evaluations.

Table 4 – TMD optimization results

Case	k_d(MN/m)	c_d(MNs/m)	β_{max}	β_{mean}	$\beta_{uncontrolled}$
(a) h/300	3.053	0.153	6.68	6.31	3.70
(b) h/400	2.963	0.152	4.24	3.99	1.37
(c) h/500	3.018	0.160	2.48	2.31	fail

In Figure 43, the Monte Carlo convergence for a single input is shown. The average value of $-\beta$ starts to converge around 150 simulations. Thus, it becomes clear that using the standard approach of using this fixed number to simulate every input would lead to a higher computational cost. Simulating only seven different points would cause the maximum number of evaluations of 1000 to be exceeded. However, using the proposed approach only points closer to the optimum are simulated a higher number of times. Moreover, by using the Monte Carlo variance estimates it regress the surface structure and avoid further

computational costs.
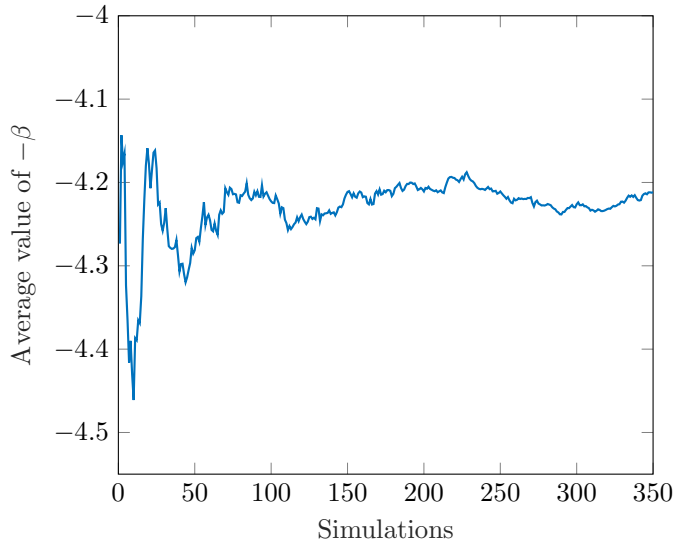
Figure 43 – Monte Carlo convergence curve

In Figure 44, the resulting surface generated by the proposed algorithm, considering the case (b) barrier level is illustrated. Comparing it against Figure 42 shows how the surface becomes smooth by the regression capabilities of Stochastic Kriging. The red dots in Figure 44 represent the points that were sampled. Some are scattered over the domain, which most likely belong to the initial sampling procedure using Latin Hypercubes. Yet, numerous points are concentrated close to minimum value of $-\beta$. Those are the points obtained by the infill step, with the maximization of the AEI criterion and considering adaptive variance target. This application possess the characteristics which the proposed algorithm is particularly efficient at solving, *i.e.*, problems with numerous random variables but a relatively small number of design variables ($n \leq 10$).
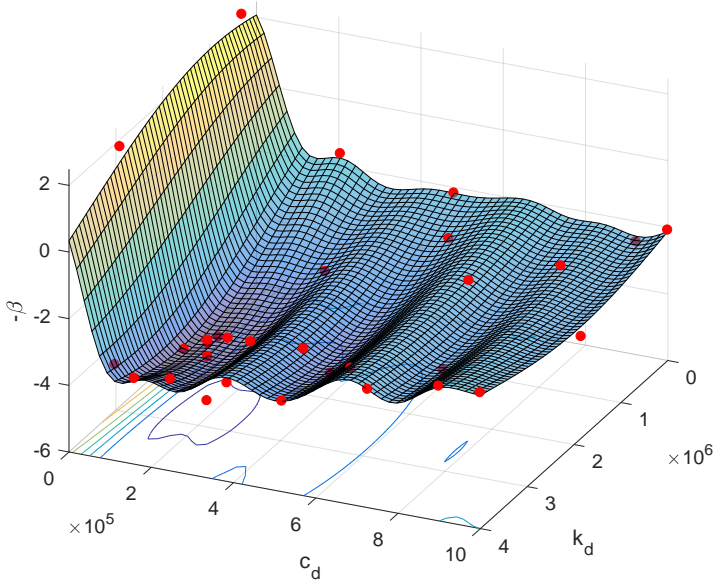
Figure 44 – Surface generated by Stochastic Kriging with sampled
            points

7 Conclusions and future studies

7.1 Conclusions

In this study an effective method based on Stochastic Kriging for the minimization of functions that depend on an integral was proposed. The method has been based on using variance estimates from Monte Carlo simulation to aid the regression metamodel construction. Moreover, modifications in the AEI infill criteria were proposed. As seen, considering a large fixed variance target may stall the optimization. Yet, if the variance target is too small, the cost becomes prohibitive. In this study, an adaptive targeting approach was employed. By being adapted based on proximity of known points it achieved a better balance between exploration and exploitation.

In order to assess the effectiveness of the proposed method, numerous benchmark tests were conducted. The proposed method was first tested against deceptive noisy benchmark functions from literature. Problems with up to 10 dimensions were analyzed and a comparison was made between the proposed adaptive variance targeting and the use of a fixed target. The proposed method obtained better results in all comparisons. Moreover, it was observed that higher dimensions and higher noise lead to a greater difference in results, favoring the proposed approach.

Another benchmark conducted compared the EGO algorithm using a constant target, against a successful algorithm from literature. The algorithm, called GBNM, was based on restarted local searches. It obtained worse results than EGO both in terms of efficiency, with a larger number of evaluations required as well as consistency, with a bigger 5 to 95 percentile range over multiple runs.

Consisting of an integral minimization approach, the proposed

method was also compared against a common tool used for this type of problem: quadratures. The quadrature was employed in a Deterministic Kriging environment as replacement for the Monte Carlo simulation used in Stochastic Kriging. The best infill criterion found for Deterministic Kriging was used, *i.e.* EM4. The quadrature used a simple Gauss rule for one dimension extended for multiple dimensions by the product of rules. It was observed that the approach using quadratures performed worse in the one-dimensional case and that the computational cost heavily increased with the increase of the problem dimension in reason of the high number of points by dimension needed for an accurate result and the use of product of quadrature rules.

Lastly, a structural engineering application was analyzed. It consisted of a ten-story shear frame subjected to seismic excitation. The problem had multiple random variables and the objective was to find the minimum expected value of the failure probability of the structure by selecting the optimal TMD parameters. The problem was successfully solved and the best results found were presented. The proposed approach was able to obtain good results efficiently and consistently for the three different cases considered.

Overall, the modification proposed to the EGO algorithm and its use case for the minimization of integrals by incorporating the Monte Carlo variance estimates yielded convincing results. Different types of problems were solved and the approach maintained its computational efficiency. Moreover, when comparing against other approaches it required a smaller number of function evaluations and presented less variability of the results over multiple independent runs.

The method limitation resides in the inherent size limitation of EGO coupled with the Kriging metamodel, *i.e.* the need of inversion or decomposition of the covariance matrix at each iteration step. This may render the task more computationally demanding than working directly on the objective function. Thus, it becomes more attractive for problems where each function evaluation is known to be computationally demanding. Nevertheless, the method usefulness and applicability was

shown in a practical engineering problem. It became evident how it can work efficiently for cases where the number of design variables is relatively small but there are numerous stochastic parameters involved in the objective function.

7.2   Future studies

As for further studies, a number of extensions may be pursued both in terms of formulation of the studied method as well the applicability to other types of problems. Here are listed a few suggestions:

- Use of gradients or Hessian information on the SK model;

- Application for problems with constraints;

- Application for multi-objective optimization problems;

- Study other forms of adaptivity of the target variance;

- Investigate efficiency gain by avoiding the computation of the Kriging hyperparameters at every iteration of EGO;

- Investigate other ways to estimate the variance of each sampled point.

References

ANDERSON, T. W.; DARLING, D. A. Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, JSTOR, p. 193–212, 1952.   Cited in page 155.

ANKENMAN, B.; NELSON, B. L.; STAUM, J. Stochastic kriging for simulation metamodeling. *Operations research*, INFORMS, v. 58, n. 2, p. 371–382, 2010.   Cited 2 times in pages 81 and 83.

Applied Science International, LLC. *Extreme Loading ®for Structures*: Nonlinear dynamic structual analysis software. 2016. Available from Internet: <http://www.extremeloading.com/els-applications/vulnerability-assessment/>. Date accessed: 2016-10-05.   Cited in page 21.

ARFIADI, Y.; HADI, M. N. S. Optimum placement and properties of tuned mass dampers using hybrid genetic algorithms. *Iran University of Science & Technology*, Iran University of Science & Technology, v. 1, n. 1, p. 167–187, 2011.   Cited in page 56.

ARORA, J. S. *Introduction to Optimum Design*. San Diego, USA: Elsevier Academic Press, 2004. ISBN 9780080470252.   Cited 2 times in pages 32 and 33.

ARORA, J. S. *Optimization of structural and mechanical systems*. Hackensack, USA: World Scientific, 2007. ISBN 9789812569622.   Cited in page 30.

ATASHPAZ-GARGARI, E.; LUCAS, C. Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In: IEEE. *Evolutionary computation, 2007. CEC 2007. IEEE Congress on*. Singapore, Singapore, 2007. p. 4661–4667.   Cited in page 38.

BECK, A. T. *Curso de Confiabiliade Estrutural*. São Carlos, SP, Brazil: Escola de Engenharia de São Carlos - Departamento de Engenharia de Estruturas, 2014.   Cited in page 51.

BECK, A. T.; GOMES, W. J. de S.; LOPEZ, R. H.; MIGUEL, L. F. F. A comparison between robust and risk-based optimization under

uncertainty. *Structural and Multidisciplinary Optimization*, Springer, v. 52, n. 3, p. 479–492, 2015.   Cited in page 55.

BECK, A. T.; KOUGIOUMTZOGLOU, I. A.; SANTOS, K. R. M. dos. Optimal performance-based design of non-linear stochastic dynamical rc structures subject to stationary wind excitation. *Engineering Structures*, Elsevier, v. 78, p. 145–153, 2014.   Cited in page 82.

BEERS, W. C. M. V.; KLEIJNEN, J. P. C. Kriging for interpolation in random simulation. *Journal of the Operational Research Society*, Nature Publishing Group, v. 54, n. 3, p. 255–262, 2003.   Cited in page 77.

BEN-TAL, A.; NEMIROVSKI, A. Robust optimization–methodology and applications. *Mathematical Programming*, Springer-Verlag, v. 92, n. 3, p. 453–480, 2002.   Cited in page 55.

BERTSEKAS, D. P.; TSITSIKLIS, J. N. *Introduction to probability.* Belmont, USA: Athena Scientific, 2002. v. 1.   Cited 3 times in pages 42, 45, and 47.

BEYER, H.-G.; SENDHOFF, B. Robust optimization–a comprehensive survey. *Computer methods in applied mechanics and engineering*, Elsevier, v. 196, n. 33, p. 3190–3218, 2007.   Cited 2 times in pages 53 and 56.

BOBBY, S.; SPENCE, S. M. J.; KAREEM, A. Data-driven performance-based topology optimization of uncertain wind-excited tall buildings. *Structural and Multidisciplinary Optimization*, v. 54, n. 6, p. 1379–1402, Dec 2016.   Cited in page 82.

BOX, G. E. P.; HUNTER, W. G.; HUNTER, J. S. *et al. Statistics for experimenters.* New York, USA: John Wiley and Sons, 1978.   Cited in page 60.

BROME, T. *Wyoming turbine collapse*: Collapse of turbine 11 at the Foote Creek Rim wind energy facility near Arlington, Wyoming, USA. 2010. Available from Internet: <http://www.windaction.org/posts/ 29881-wyoming-turbine-collapse#.WAgTUegrKhc>. Date accessed: 2016-10-18.   Cited in page 24.

CAFLISCH, R. E. Monte carlo and quasi-monte carlo methods. *Acta numerica*, Cambridge Univ Press, v. 7, p. 1–49, 1998.   Cited in page 54.

CARLON, A. G. *et al.* Desenvolvimento de um novo operador para algoritmos metaheurísticos baseado na maximização da rigidez aplicado à otimização de estruturas treliçadas. 2015. Cited in page 38.

CARRARO, F.; LOPEZ, R. H.; MIGUEL, L. F. F. Optimum design of planar steel frames using the search group algorithm. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, Springer, p. 1–14, 2016. Cited in page 29.

CASTRILLÓN-CANDÁS, J. E.; GENTON, M. G.; YOKOTA, R. Multi-level restricted maximum likelihood covariance estimation and kriging for large non-gridded spatial datasets. *Spatial Statistics*, Elsevier, 2015. Cited in page 67.

CHAKRABORTY, S.; ROY, B. K. Reliability based optimum design of tuned mass damper in seismic vibration control of structures with bounded uncertain parameters. *Probabilistic Engineering Mechanics*, Elsevier, v. 26, n. 2, p. 215–221, 2011. Cited in page 118.

CHAUDHURI, A.; HAFTKA, R. T. Efficient global optimization with adaptive target setting. *AIAA Journal*, American Institute of Aeronautics and Astronautics, v. 52, n. 7, p. 1573–1578, 2014. Cited in page 75.

CHAUDHURI, A.; HAFTKA, R. T.; IFJU, P.; CHANG, K.; TYLER, C.; SCHMITZ, T. Experimental flapping wing optimization and uncertainty quantification using limited samples. *Structural and Multidisciplinary Optimization*, v. 51, n. 4, p. 957–970, 2015. ISSN 1615-1488. Cited in page 75.

CHEN, V. C. P.; TSUI, K. L.; BARTON, R. R.; ALLEN, J. K. A review of design and modeling in computer experiments. *Statistics in industry*, v. 22, n. 2003, p. 231–261, 2003. ISSN 01697161. Cited in page 63.

CHEN, W. F.; LUI, E. M. *Handbook of Structural Engineering, Second Edition*. Boca Raton, FL, USA: CRC Press, 2005. ISBN 9781420039931. Cited in page 122.

CHEN, X.; KIM, K.-K. Stochastic kriging with biased sample estimates. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, ACM, v. 24, n. 2, p. 8, 2014. Cited in page 84.

CIVICIOGLU, P. Backtracking search optimization algorithm for numerical optimization problems. *Applied Mathematics and*

Computation, Elsevier, v. 219, n. 15, p. 8121–8144, 2013. Cited 2 times in pages 38 and 104.

CLARKE, S. M.; GRIEBSCH, J. H.; SIMPSON, T. W. Analysis of support vector regression for approximation of complex engineering analyses. *Journal of mechanical design*, American Society of Mechanical Engineers, v. 127, n. 6, p. 1077–1087, 2005. Cited in page 60.

CODY, W. J. Rational chebyshev approximations for the error function. *Mathematics of Computation*, v. 23, n. 107, p. 631–637, 1969. Cited in page 45.

COLORNI, A.; DORIGO, M.; MANIEZZO, V. An investigation of some properties of an ant algorithm. In: *Proc. Parallel Problem Solving from Nature Conference*. Brussels, Belgium: Elsevier, 1992. p. 509–520. Cited in page 37.

COSTA, J.-P.; PRONZATO, L.; THIERRY, E. A comparison between kriging and radial basis function networks for nonlinear prediction. In: *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*. Antalya, Turkey: NSIP, 1999. p. 726–730. Cited in page 64.

CRESSIE, N. *Statistics for spatial data*. New York: Wiley, 1993. v. 15. 105–209 p. Cited in page 61.

DAVIS, P. R. P. J. *Methods of numerical integration*. 2nd ed. ed. San Diego, USA: Academic Press, 1984. (Computer science and applied mathematics). ISBN 9780122063602,0122063600. Cited in page 115.

DEVORE, J. L. *Probability and Statistics for Engineering and the Sciences*. Boston, Massachusetts, USA: Cengage learning, 2011. Cited 2 times in pages 44 and 156.

DITLEVSEN, O.; MADSEN, H. O. *Structural reliability methods*. New York, USA: Wiley New York, 1996. v. 178. Cited in page 54.

DYN, N.; LEVIN, D.; RIPPA, S. Numerical procedures for surface fitting of scattered data by radial functions. *SIAM Journal on Scientific and Statistical Computing*, SIAM, v. 7, n. 2, p. 639–659, 1986. Cited in page 60.

EOM, Y.-S.; YOO, K.-S.; PARK, J.-Y.; HAN, S.-Y. Reliability-based topology optimization using a standard response surface method for three-dimensional structures. *Structural and multidisciplinary optimization*, Springer, v. 43, n. 2, p. 287–295, 2011. Cited in page 60.

FEDOROV, V. V. *Theory of optimal experiments.* New York, USA: Elsevier, 1972.   Cited in page 82.

FELLER, W. *An introduction to probability theory and its applications: volume I.* London, UK: John Wiley & Sons, 1968. v. 3.   Cited in page 53.

FLETCHER, R. *Practical Methods of Optimization.* 2nd. ed. New York, USA: Wiley-Interscience, 1987. ISBN 0-471-91547-5.   Cited in page 29.

FORD, W. *Numerical linear algebra with applications: Using MATLAB.* San Diego, USA: Academic Press, 2014.   Cited in page 67.

FORRESTER, A.; SOBESTER, A.; KEANE, A. *Engineering design via surrogate modelling: a practical guide.* Chichester, West Sussex, United Kingdom: John Wiley & Sons, 2008.   Cited 7 times in pages 61, 65, 67, 73, 77, 84, and 103.

FORRESTER, A. I. J.; KEANE, A. J. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, Elsevier, v. 45, n. 1, p. 50–79, 2009.   Cited 5 times in pages 64, 65, 66, 73, and 104.

FORRESTER, A. I. J.; KEANE, A. J.; BRESSLOFF, N. W. Design and analysis of "noisy" computer experiments. *AIAA journal*, v. 44, n. 10, p. 2331–2339, 2006.   Cited 2 times in pages 78 and 85.

GAJIC, Z.; QURESHI, M. T. J. *Lyapunov Matrix Equation in System Stability and Control.* San Diego, California, USA: Elsevier Science, 1995. (Mathematics in Science and Engineering). ISBN 9780080535678. Cited in page 153.

GAVANA, A. *1-D Test Functions.* 2016. Available from Internet: <http://infinity77.net/global_optimization/test_functions_1d. html#go_benchmark.Problem05>. Date accessed: 2016-08-28.   Cited in page 88.

GELDER, L. V.; DAS, P.; JANSSEN, H.; ROELS, S. Comparative study of metamodelling techniques in building energy simulation: Guidelines for practitioners. *Simulation Modelling Practice and Theory*, Elsevier, v. 49, p. 245–257, 2014.   Cited in page 62.

GOLDBERG, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Boston, USA: Addison-Wesley Publishing Company, 1989. (Artificial Intelligence). ISBN 9780201157673.   Cited 2 times in pages 37 and 104.

GONÇALVES, M. S.; LOPEZ, R. H.; MIGUEL, L. F. F. Search group algorithm: A new metaheuristic method for the optimization of truss structures. *Computers and Structures*, v. 153, p. 165–184, 2015. Cited in page 38.

HADI, M. N. S.; ARFIADI, Y. Optimum design of absorber for mdof structures. *Journal of Structural Engineering*, American Society of Civil Engineers, v. 124, n. 11, p. 1272–1280, 1998. Cited in page 120.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation.* 2nd. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998. ISBN 0132733501. Cited in page 60.

HUANG, D.; ALLEN, T. T.; NOTZ, W. I.; ZENG, N. Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of global optimization*, Springer, v. 34, n. 3, p. 441–466, 2006. Cited 2 times in pages 24 and 86.

HUANG, Z.; WANG, C.; CHEN, J.; TIAN, H. Optimal design of aeroengine turbine disc based on kriging surrogate models. *Computers & structures*, Elsevier, v. 89, n. 1, p. 27–37, 2011. Cited in page 60.

HUSSAIN, M. F.; BARTON, R. R.; JOSHI, S. B. Metamodeling: radial basis functions, versus polynomials. *European Journal of Operational Research*, Elsevier, v. 138, n. 1, p. 142–154, 2002. Cited in page 60.

JOHNSON, M. E.; MOORE, L. M.; YLVISAKER, D. Minimax and maximin distance designs. *Journal of statistical planning and inference*, Elsevier, v. 26, n. 2, p. 131–148, 1990. Cited in page 69.

JONES, D. A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization*, v. 21, n. 4, p. 345–383, 2001. Cited 4 times in pages 70, 71, 72, and 74.

JONES, D. R.; SCHONLAU, M.; WILLIAM, J. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, v. 13, p. 455–492, 1998. ISSN 09255001. Cited 3 times in pages 61, 65, and 70.

JR., C. W. C. Response surface methods for optimizing and improving reproducibility of crystal growth. In: *Macromolecular Crystallography Part A*. San Diego, USA: Academic Press, 1997, (Methods in Enzymology, v. 276). p. 74 – 99. Cited in page 60.

KALAGNANAM, J. R.; DIWEKAR, U. M. An efficient sampling technique for off-line quality control. *Technometrics*, Taylor & Francis, v. 39, n. 3, p. 308–319, 1997. Cited in page 68.

KAMIŃSKI, B. A method for the updating of stochastic kriging metamodels. *European Journal of Operational Research*, Elsevier, v. 247, n. 3, p. 859–866, 2015. Cited 2 times in pages 84 and 85.

KANAI, K. Semi-empirical formula for the seismic characteristics of the ground. v. 35, 07 1957. Cited in page 121.

KEANE, A.; NAIR, P. *Computational approaches for aerospace design: the pursuit of excellence.* Chichester, West Sussex, England: John Wiley & Sons, 2005. ISBN 9780470855485. Cited in page 77.

KENNEDY, J.; EBERHART, R. Particle swarm optimization. In: *Proceedings IEEE International Conference on Neural Networks.* Washington, USA: IEEE, 1995. v. 4, p. 1942–1948 vol.4. ISSN 19353812. Cited 2 times in pages 37 and 104.

KIM, B.-S.; LEE, Y.-B.; CHOI, D.-H. Comparison study on the accuracy of metamodeling technique for non-convex functions. *Journal of Mechanical Science and Technology*, Springer, v. 23, n. 4, p. 1175–1181, 2009. Cited in page 64.

KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by simulated annealing. *Science*, American Association for the Advancement of Science, v. 220, n. 4598, p. 671–680, 1983. ISSN 0036-8075. Cited in page 37.

KLEIJNEN, J. P. C.; MEHDAD, E. Estimating the variance of the predictor in stochastic kriging. *Simulation Modelling Practice and Theory*, Elsevier, v. 66, p. 166–173, 2016. Cited in page 84.

KOLMOGOROV, A. N. Foundations of the theory of probability. Chelsea Publishing Co., Oxford, England, 1950. Cited in page 40.

KRIGE, D. G. A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, Operational Research Society, v. 52, n. 6, p. 119–139, dez. 1951. Cited in page 61.

KRISHNAMURTHY, T. Comparison of response surface construction methods for derivative estimation using moving least squares, kriging and radial basis functions. In: *Proceedings of the 46th*

AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference, AIAA-2005-1821. Reno, NV, USA: American Institute of Aeronautics and Astronautics, 2005. p. 18–21. Cited in page 64.

LEE, C.-L.; CHEN, Y.-T.; CHUNG, L.-L.; WANG, Y.-P. Optimal design theories and applications of tuned mass dampers. *Engineering structures*, Elsevier, v. 28, n. 1, p. 43–53, 2006. Cited in page 120.

LEE, J. *A First Course in Combinatorial Optimization*. Cambridge, UK: Cambridge University Press, 2004. (Cambridge Texts in Applied Mathematics). ISBN 9780521010122. Cited in page 30.

LEIRA, B. J. *Optimal stochastic control schemes within a structural reliability framework*. Cham, Switzerland: Springer, 2013. (SpringerBriefs in Statistics). Cited in page 50.

LELIÈVRE, N.; BEAUREPAIRE, P.; MATTRAND, C.; GAYTON, N.; OTSMANE, A. On the consideration of uncertainty in design: optimization-reliability-robustness. *Structural and Multidisciplinary Optimization*, Springer, p. 1–15, 2016. Cited in page 55.

LIU, B.; HAFTKA, R. T.; AKGÜN, M. A. Two-level composite wing structural optimization using response surfaces. *Structural and Multidisciplinary Optimization*, Springer, v. 20, n. 2, p. 87–96, 2000. Cited in page 60.

LOCATELLI, M. Bayesian algorithms for one-dimensional global optimization. *Journal of Global Optimization*, v. 10, n. 1, p. 57–76, 1997. ISSN 1573-2916. Cited in page 85.

LOPEZ, R. H.; BECK, A. T. Optimization under uncertainties. In: *Optimization of Structures and Components*. Cham, Switzerland: Springer, 2013. p. 117–141. Cited 3 times in pages 52, 53, and 56.

LOPEZ, R. H.; CURSI, J. E. S. de; LEMOSSE, D. Approximating the probability density function of the optimal point of an optimization problem. *Engineering Optimization*, Taylor & Francis, v. 43, n. 3, p. 281–303, 2011. Cited in page 53.

LOPEZ, R. H.; LEMOSSE, D.; CURSI, J. E. S. de; ROJAS, J.; EL-HAMI, A. An approach for the reliability based design optimization of laminated composites. *Engineering Optimization*, Taylor & Francis, v. 43, n. 10, p. 1079–1094, 2011. Cited in page 53.

LOPEZ, R. H.; MIGUEL, L. F. F.; BECK, A. T. Tuned mass dampers for passive control of structures under earthquake excitations. In: *Encyclopedia of Earthquake Engineering*. Cham, Switzerland: Springer, 2015. p. 3814–3823. Cited 2 times in pages 56 and 120.

LUENBERGER, D. G. *Optimization by Vector Space Methods*. New York, USA: Wiley, 1969. (Professional Series). ISBN 9780471181170. Cited 2 times in pages 30 and 35.

LUERSEN, M. A.; RICHE, R. L. Globalized nelder–mead method for engineering optimization. *Computers & structures*, Elsevier, v. 82, n. 23, p. 2251–2260, 2004. Cited 3 times in pages 38, 104, and 112.

LUTES, L. D.; SARKANI, S. *Random Vibrations: Analysis of Structural and Mechanical Systems*. Burlington, Massachusetts, USA: Elsevier Science, 2004. ISBN 9780080470030. Cited in page 151.

MADSEN, H. O.; KRENK, S.; LIND, N. C. *Methods of structural safety*. New York, USA: Prentice-Hall, 1985. (Prentice-Hall international series in civil engineering and engineering mechanics). ISBN 9780135794753. Cited in page 149.

MARANO, G. C.; GRECO, R.; TRENTADUE, F.; CHIAIA, B. Constrained reliability-based optimization of linear tuned mass dampers for seismic control. *International Journal of Solids and Structures*, Elsevier, v. 44, n. 22, p. 7370–7388, 2007. Cited in page 152.

MARELLI, S.; SUDRET, B. Uqlab user manual. *Chair of Risk, Safety & Uncertainty Quantification, ETH Zürich*, 2015. Cited in page 103.

MELCHERS, R. E. *Structural reliability analysis and prediction*. New York, USA: John Wiley, 1999. (Civil Engineering Series). ISBN 9780471983248. Cited in page 148.

MIGUEL, L. F. F.; MIGUEL, L. F. F.; LOPEZ, R. H. Robust design optimization of friction dampers for structural response control. *Structural Control and Health Monitoring*, Wiley Online Library, v. 21, n. 9, p. 1240–1251, 2014. Cited in page 56.

MITCHELL, T. J.; MORRIS, M. D. The spatial correlation function approach to response surface estimation. In: *Proceedings of the 24th Conference on Winter Simulation*. New York, NY, USA: ACM, 1992. (WSC '92), p. 565–571. ISBN 0-7803-0798-4. Cited in page 69.

MÜLLER, P. Simulation based optimal design. *Handbook of Statistics*, v. 25, p. 509 – 518, 2005. ISSN 0169-7161. Bayesian Thinking. Cited in page 82.

MOEHLE, J.; DEIERLEIN, G. G. A framework methodology for performance-based earthquake engineering. In: *13th world conference on earthquake engineering.* [S.l.: s.n.], 2004. p. 3812–3814. Cited in page 82.

MOHEBBI, M.; SHAKERI, K.; GHANBARPOUR, Y.; MAJZOUB, H. Designing optimal multiple tuned mass dampers using genetic algorithms (GAs) for mitigating the seismic response of structures. *Journal of Vibration and Control*, Sage Publications Sage UK: London, England, v. 19, n. 4, p. 605–625, 2013. Cited in page 121.

MONTGOMERY, D. C.; MYERS, R. H. *Response surface methodology: process and product optimization using designed experiments.* Hoboken, USA: Wiley-Interscience, 1995. Cited in page 60.

MONTGOMERY, D. C.; RUNGER, G. C. *Applied statistics and probability for engineers.* New York, USA: John Wiley & Sons, 2010. Cited 2 times in pages 43 and 65.

MORRIS, M. D.; MITCHELL, T. J. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, v. 43, n. 3, p. 381 – 402, 1995. ISSN 0378-3758. Cited in page 69.

MULLUR, A. A.; MESSAC, A. Metamodeling using extended radial basis functions: a comparative approach. *Engineering with Computers*, Springer, v. 21, n. 3, p. 203–217, 2006. Cited in page 63.

MUSELLI, M. A theoretical approach to restart in global optimization. *Journal of Global Optimization*, v. 10, n. 1, p. 1–16, 1997. ISSN 1573-2916. Cited in page 37.

NASA. Extending the operational life of the International Space Station until 2024. *Office of Ispector General*, Audit Report IG-14-031, 2014. Cited in page 23.

NASA. *Orbital Debris.* 2014. Available from Internet: <https://www.nasa.gov/centers/johnson/partnerships/orbital-debris/#.WAQjlOgrKhc>. Date accessed: 2016-10-15. Cited in page 22.

NELDER, J. A.; MEAD, R. A simplex method for function minimization. *The computer journal*, Oxford University Press, v. 7, n. 4, p. 308–313, 1965. Cited 2 times in pages 23 and 112.

NHAMAGE, I. A. *et al.* Aperfeiçoamento do algoritmo de otimização híbrido pincus-nelder e mead para detecção de dano em estruturas a partir de dados vibracionais. 2014. Cited in page 23.

NOCEDAL, J.; WRIGHT, S. *Numerical Optimization*. 2nd. ed. New York, USA: Springer New York, 2006. (Springer Series in Operations Research and Financial Engineering). ISBN 9780387400655. Cited in page 36.

NOWAK, A. S.; COLLINS, K. R. *Reliability of structures*. Boca Raton, FL, USA: CRC Press, 2012. Cited in page 50.

PAPADIMITRIOU, C. H.; STEIGLITZ, K. *Combinatorial Optimization: Algorithms and Complexity*. Englewood Cliffs, USA: Dover Publications, 1982. (Dover Books on Computer Science). ISBN 9780486402581. Cited in page 30.

PAPOULIS, A.; PILLAI, S. U. *Probability, random variables, and stochastic processes*. Boston, Massachusetts, USA: Tata McGraw-Hill Education, 2002. Cited 2 times in pages 47 and 48.

PARKINSON, A.; SORENSEN, C.; POURHASSAN, N. A general approach for robust optimal design. *Journal of mechanical design*, American Society of Mechanical Engineers, v. 115, n. 1, p. 74–80, 1993. Cited in page 53.

PICHENY, V.; GINSBOURGER, D. Noisy kriging-based optimization methods: a unified implementation within the diceoptim package. *Computational Statistics & Data Analysis*, Elsevier, v. 71, p. 1035–1053, 2014. Cited in page 87.

PICHENY, V.; WAGNER, T.; GINSBOURGER, D. A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, Springer, v. 48, n. 3, p. 607–626, 2013. Cited in page 85.

PLUMLEE, M.; TUO, R. Building accurate emulators for stochastic simulations via quantile kriging. *Technometrics*, Taylor & Francis, v. 56, n. 4, p. 466–473, 2014. Cited in page 84.

POGGIO, T.; GIROSI, F. Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, American Association for the Advancement of Science, v. 247, n. 4945, p. 978–982, 1990. Cited in page 77.

PRONZATO, L. Optimal experimental design and some related control problems. *Automatica*, v. 44, n. 2, p. 303 – 325, 2008. ISSN 0005-1098. Cited in page 82.

PRONZATO, L.; MÜLLER, W. G. Design of computer experiments: space filling and beyond. *Statistics and Computing*, Springer, v. 22, n. 3, p. 681–701, 2012. Cited in page 68.

QU, H.; FU, M. C. Gradient extrapolated stochastic kriging. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, ACM, v. 24, n. 4, p. 23, 2014. Cited in page 85.

ROUSTANT, O.; GINSBOURGER, D.; DEVILLE, Y. Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, Foundation for Open Access Statistics, v. 51, n. i01, 2012. Cited in page 103.

SACKS, J.; WELCH, W. J.; MITCHELL, T. J.; WYNN, H. P. Design and analysis of computer experiments. *Statist. Sci.*, The Institute of Mathematical Statistics, v. 4, n. 4, p. 409–423, 11 1989. Cited 5 times in pages 60, 61, 63, 67, and 68.

SANKARARAMAN, S. Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction. *Mechanical Systems and Signal Processing*, Elsevier, v. 52, p. 228–247, 2015. Cited 2 times in pages 53 and 54.

SCHAY, G. *Introduction to Probability with Statistical Applications*. Boston, Massachusetts, USA: Birkhäuser Boston, 2007. ISBN 9780817644970. Cited in page 45.

SIMPSON, T. W.; MAUERY, T. M.; KORTE, J.; MISTREE, F. Kriging models for global approximation in simulation-based multidisciplinary design optimization. *AIAA Journal*, v. 39, n. 12, p. 2233–2241, 2001. ISSN 0001-1452. Cited in page 63.

SONG, H.; CHOI, K. K.; LAMB, D. A study on improving the accuracy of kriging models by using correlation model/mean structure selection and penalized log-likelihood function. In: *10th World Congress*

on Structural and Multidisciplinary Optimization. Florida, Orlando, USA: International Society for Structural and Multidisciplinary Optimization, 2013. Cited in page 67.

SOONG, T. T.; GRIGORIU, M. *Random vibration of mechanical and structural systems.* Eaglewood Cliffs, New Jersey, USA: PTR Prentice Hall, 1993. ISBN 9780137523610. Cited in page 151.

SORENSEN, D. C.; ZHOU, Y. Direct methods for matrix sylvester and lyapunov equations. *Journal of Applied Mathematics*, Hindawi Publishing Corporation, v. 2003, n. 6, p. 277–303, 2003. Cited in page 153.

SOUZA, J. V. de; MIGUEL, L. F. F. Determinação ótima do número e dos parâmetros de múltiplos atenuadores dinâmicos sintonizados em estruturas altas submetidas à ação dinâmica do vento. 2016. Cited in page 104.

SOUZA, R. R.; MIGUEL, L. F. F.; LOPEZ, R. H.; TORII, A. J.; MIGUEL, L. F. F. A Backtracking Search Algorithm for the Simultaneous Size, Shape and Topology Optimization of Trusses. *Latin American Journal of Solids and Structures*, scielo, v. 13, p. 2922 – 2951, 12 2016. ISSN 1679-7825. Cited in page 104.

SPENCE, S. M. J.; GIOFFRÈ, M. Large scale reliability-based design optimization of wind excited tall buildings. *Probabilistic Engineering Mechanics*, Elsevier, v. 28, p. 206–215, 2012. Cited in page 82.

SPENCE, S. M. J.; KAREEM, A. Performance-based design and optimization of uncertain wind-excited dynamic building systems. *Engineering Structures*, Elsevier, v. 78, p. 133–144, 2014. Cited in page 82.

STAUM, J. Better simulation metamodeling: The why, what, and how of stochastic kriging. In: IEEE. *Proceedings of the 2009 Winter Simulation Conference (WSC)*. Austin, TX, USA, 2009. p. 119–133. Cited in page 81.

STEINBERG, D. M.; HUNTER, W. G. Experimental design: review and comment. *Technometrics*, Taylor & Francis, v. 26, n. 2, p. 71–97, 1984. Cited in page 82.

SUDRET, B. Uncertainty propagation and sensitivity analysis in mechanical models–contributions to structural reliability and stochastic spectral methods. *Habilitationa diriger des recherches, Université*

Blaise Pascal, Clermont-Ferrand, France, 2007. Cited 2 times in pages 40 and 50.

SUDRET, B. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, Elsevier, v. 93, n. 7, p. 964–979, 2008. Cited in page 60.

SUDRET, B.; KIUREGHIAN, A. D. *Stochastic finite element methods and reliability: a state-of-the-art report*. California, Berkeley, USA: Department of Civil and Environmental Engineering, University of California Berkeley, CA, 2000. Cited in page 22.

SUN, L.; HONG, L. J.; HU, Z. Balancing exploitation and exploration in discrete optimization via simulation through a gaussian process-based search. *Operations Research*, INFORMS, v. 62, n. 6, p. 1416–1438, 2014. Cited in page 85.

TAFLANIDIS, A. A.; BECK, J. L. An efficient framework for optimal robust stochastic system design using stochastic simulation. *Computer Methods in Applied Mechanics and Engineering*, Elsevier, v. 198, n. 1, p. 88–101, 2008. Cited in page 57.

TAFLANIDIS, A. A.; SCRUGGS, J. T. Probabilistically-robust performance optimization for controlled linear stochastic systems. In: IEEE. *2009 American Control Conference*. St. Louis, MO, USA, 2009. p. 4557–4562. Cited in page 57.

TAJIMI, H. A statistical method of determining the maximum response of a building structure during an earthquake. In: *Proceedings of the 2nd World Conference on Earthquake Engineering*. Tokyo, Japan: [s.n.], 1960. p. 781–797. Cited in page 121.

TAKATS, G. *Destroyed building in Muzaffarabad after the earthquake (2005)*. 2005. Available from Internet: <https://www.flickr.com/photos/dfataustralianaid/10693803394/>. Date accessed: 2016-10-18. Cited in page 24.

THE MATHWORKS, INC. *MATLAB version 8.6.0.267246 (R2015b)*. Natick, Massachusetts, 2015. Cited 2 times in pages 103 and 104.

THEODOSSIOU, N.; LATINOPOULOS, P. Evaluation and optimisation of groundwater observation networks using the kriging methodology. *Environmental Modelling & Software*, Elsevier, v. 21, n. 7, p. 991–1000, 2006. Cited in page 60.

THOFT-CHRISTENSEN, P.; BAKER, M. J. *Structural reliability theory and its applications.* Berlin, Germany: Springer-Verlag, 1982. Cited in page 41.

TORII, A. J.; LOPEZ, R. H. Reliability analysis of water distribution networks using the adaptive response surface approach. *Journal of Hydraulic Engineering*, American Society of Civil Engineers, v. 138, n. 3, p. 227–236, 2011. Cited in page 60.

TORII, A. J.; LOPEZ, R. H.; BIONDINI, F. An approach to reliability-based shape and topology optimization of truss structures. *Engineering Optimization*, v. 44, n. 1, p. 37–53, 2012. Cited in page 60.

TORII, A. J.; LOPEZ, R. H.; LUERSEN, M. A. A local-restart coupled strategy for simultaneous sizing and geometry truss optimization. *Latin American Journal of Solids and Structures*, v. 8, p. 335 – 349, 09 2011. ISSN 1679-7825. Cited in page 37.

TORN, A.; ZILINSKAS, A. *Global optimization.* New York, NY, USA: Springer-Verlag New York, Inc., 1989. Cited in page 72.

VIANA, F. A. C.; HAFTKA, R. T. Surrogate-based optimization with parallel simulations using the probability of improvement. In: *13th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference.* Fort Worth, USA: American Institute of Aeronautics and Astronautics, 2010. p. 13–15. Cited in page 75.

VROUWENVELDER, T.; KARADENIZ, H. Overview of structural reliability methods. *Safety and Reliability of Industrial Products, Systems and Structures*, CRC Press, p. 181, 2010. Cited in page 54.

XU, B.; SUN, F.; LIU, H.; REN, J. Adaptive kriging controller design for hypersonic flight vehicle via back-stepping. *IET control theory & applications*, IET, v. 6, n. 4, p. 487–497, 2012. Cited in page 60.

YOUNIS, A.; DONG, Z. Trends, features, and tests of common and recently introduced global optimization methods. *Engineering Optimization*, Taylor & Francis, v. 42, n. 8, p. 691–718, 2010. Cited in page 37.

ZABINSKY, Z. B. *Stochastic Adaptive Search for Global Optimization.* New York, USA: Kluwer Academic Publishers, 2003. ISBN 978-1-4613-4826-9. Cited in page 34.

ZHAO, L.; CHOI, K. K.; LEE, I.; GORSICH, D. A metamodeling method using dynamic kriging and sequential sampling. In: *The 13th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference.* Fort Worth, TX, USA: American Institute of Aeronautics and Astronautics, 2010. p. 13–15. Cited in page 67.

ZHU, P.; ZHANG, L. W.; LIEW, K. M. Geometrically nonlinear thermomechanical analysis of moderately thick functionally graded plates using a local petrov–galerkin approach with moving kriging interpolation. *Composite Structures*, Elsevier, v. 107, p. 298–314, 2014. Cited in page 60.

APPENDIX A – Time-dependent reliability of oscillators

The classical reliability problem for an oscillator is depicted in Figure 45. For a specified excitation duration event t_E, a specific, scalar displacement response of the oscillator $Z(t)$ should not exceed a given critical response level. This critical response, defines a barrier $(\pm b)$, which should not be exceeded. This barrier can be the relative displacement between floors, displacement of top floor or any other critical displacement measure. It may also represent permanent damage like cracking of concrete, or ultimate failure due to loss-of-equilibrium.
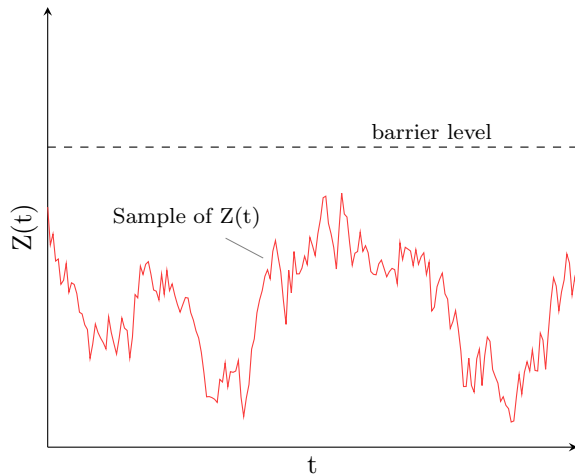

Figure 45 – Barrier

For a given barrier level b and excitation duration t_E, the failure probability is calculated following the classical Poisson model (MELCH-

ERS, 1999):

$$P_f(b, t_e) = 1 - \exp\left(-2\int_0^{t_e} v_z^+(b, t)dt\right),\tag{A.1}$$

where v_z^+ is the up-crossing rate. When considering a stationary excitation and a time-invariant barrier $b(t) = b$, the crossing rate becomes:

$$\int_0^{t_e} v_z^+(b, t)dt = v_z^+(b)\, t_E.\tag{A.2}$$

For a linear system excited by a Gaussian process, the response is Gaussian and the crossing rate is evaluated as:

$$v_z^+(b) = \frac{\sigma_{\dot{z}}}{\sigma_z}\frac{1}{2\pi}\exp\left(-\frac{b^2}{2\sigma_z^2}\right),\tag{A.3}$$

where σ_z and $\sigma_{\dot{z}}$ are the standard deviation of the displacement and of the velocity response, respectively.

The structural loading from an earthquake, which is the application topic, is described by the arrival of an unknown number of events. The same reasoning can be applied to many types of environmental loads such as winds, storms, sea weaves, *etc.* The arrival of the events is modeled as a Poisson process, and for a design life t_D, the failure probability becomes:

$$P_f(b, t_D) = \sum_{i=1}^{\infty} P_f(b, t_E|i)\, p_i(t_D)\tag{A.4}$$

where $P_f(b, t_E|i)$ is the conditional probability of failure, given the occurrence of exactly i events during the design life, and $p_i(t_D)$ is the probability of having exactly i events, given by the Poisson distribution:

$$p_i(t_D) = \frac{(v\, t_D)^i \exp(-v\, t_D)}{i!}\tag{A.5}$$

where v is the arrival rate of events. If independence among events is assumed, the conditional failure probability is given by:

$$P_f(b, t_E|i) = 1 - (1 - P_f(b, t_E))^i.\tag{A.6}$$

The failure probability can be converted into the more convenient reliability index (β) by using the classical First Order Reliability result (MADSEN *et al.*, 1985):

$$\beta = -\Phi^{-1}(P_f(b, t_D)), \tag{A.7}$$

where Φ is the standard Gaussian cumulative distribution function. Clearly, this index is a function of the vector of design parameters $\mathbf{d}$ as well as the vector of structural parameters $\mathbf{X}$. Thus, one can write $\beta = \beta(\mathbf{d}, \mathbf{X})$.

APPENDIX B – Solving TMD Random Response Based on Lyapunov Equation

Recalling from Equation 6.8, the equation of motion for the n-floor shear building with a single TMD at the top is:

$$\mathcal{M}\ddot{\mathbf{z}}(t) + \mathcal{C}\dot{\mathbf{z}}(t) + \mathcal{K}\mathbf{z}(t) = -\mathbf{m}\,\ddot{z}_g(t).$$
(B.1)

Moreover, the process of the excitation of the base is assumed to be a white noise process $w(t)$ filtered by the Kanai-Taijimi filter. It can be written as (LUTES; SARKANI, 2004):

$$
\begin{cases}
\ddot{z}_f(t) + 2\xi\omega_f\dot{z}_f + \omega_f^2 z_f = -w(t) \\
\ddot{z}_g(t) = \ddot{z}_f(t) + w(t) = 2\xi_f\omega_f\dot{z}_f + \omega_f^2 z_f,
\end{cases}
$$
(B.2)

where ξ_f and ω_f are, respectively, the damping ratio and the frequency of the filter.

For numerical analysis, *i.e.* with MATLAB, it is often convenient to write second order differential equations as a set of first order equations. Introducing the space state vector:

$$\mathbf{s} = \{z_1, z_2, ..., z_n, z_d, z_f, \dot{z_1}, \dot{z_2}, ..., \dot{z_n}, \dot{z_d}, \dot{z_f}\},$$
(B.3)

the state space system description can replace Equations B.1 and B.2 with the equivalent formulation (SOONG; GRIGORIU, 1993):

$$\dot{\mathbf{s}} = \boldsymbol{A}\mathbf{s} + \mathbf{f}$$
(B.4)

where $\mathbf{f}$ is the vector $\mathbf{f} = \{0, 0, ..., -\mathrm{w(t)}\}$, and $\boldsymbol{A}$ it the system matrix of size $(2\mathrm{n} + 4) \times (2\mathrm{n} + 4)$, written as:

$$
\begin{bmatrix}
\mathbf{0} & \mathbf{I} \\
\boldsymbol{H}_\mathrm{k} & \boldsymbol{H}_\mathrm{c}
\end{bmatrix}
\tag{B.5}
$$

where $\mathbf{0}$ and $\mathbf{I}$ are null and identity submatrices of size $(\mathrm{n} + 2) \times (\mathrm{n} + 2)$, respectively. Additionaly, $\boldsymbol{H}_\mathrm{k}$ and $\boldsymbol{H}_\mathrm{c}$ are respectively written as (MARANO *et al.*, 2007):

$$
\boldsymbol{H}_\mathrm{k} =
\begin{bmatrix}
 & & & & \omega_\mathrm{f}^2 \\
 & & & & \omega_\mathrm{f}^2 \\
 & -(\mathcal{M}^{-1}\mathcal{K}) & & & \vdots \\
 & & & & \omega_\mathrm{f}^2 \\
0 & \cdots & \cdots & 0 & -\omega_\mathrm{f}^2
\end{bmatrix},
\tag{B.6}
$$

and

$$
\boldsymbol{H}_\mathrm{c} =
\begin{bmatrix}
 & & & & 2\xi\omega_\mathrm{f} \\
 & & & & 2\xi\omega_\mathrm{f} \\
 & -(\mathcal{M}^{-1}\mathcal{C}) & & & \vdots \\
 & & & & 2\xi\omega_\mathrm{f} \\
0 & \cdots & \cdots & 0 & -2\xi\omega_\mathrm{f}
\end{bmatrix}.
\tag{B.7}
$$

The covariance matrix of the vector $\mathbf{s}$ is defined as:

$$
\boldsymbol{\Gamma}_\mathbf{s} = \mathrm{E}[(\mathbf{s} - \mu_\mathbf{s})(\mathbf{s} - \mu_\mathbf{s})^\mathrm{T}].
\tag{B.8}
$$

Due to the excitation considered being a zero-mean process, the response is also a zero-mean process. Therefore, the covariance of the response is given by:

$$
\boldsymbol{\Gamma}_\mathbf{s} = \mathrm{E}[\mathbf{s}\mathbf{s}^\mathrm{T}].
\tag{B.9}
$$

For the stationary case, the Lyapunov equation can be used to solve for the covariance matrix (GAJIC; QURESHI, 1995). It states:

$$\boldsymbol{A}\boldsymbol{\Gamma_s} + \boldsymbol{\Gamma_s}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B} = 0 \tag{B.10}$$

where the input matrix $\boldsymbol{B}$ for the excitation considered is a null matrix, except for the last element $\boldsymbol{B}_{(2n+4,2n+4)} = 2\pi S_0$.

After solving the Lyapunov equation, it is possible to extract the variances of the displacements and velocities from all the degrees of freedom of the structure. This is done by taking the elements from the main diagonal of $\boldsymbol{\Gamma_s}$. This procedure of finding the variances is considerably more efficient than using a time-stepping algorithm, such as Newmark (SORENSEN; ZHOU, 2003). This advantage is beneficial in an optimization environment, where the Lyapunov equation must be solved numerous times.

APPENDIX C – Intrinsic noise assumption

Here an experiment is conducted in order to verify the plausibility of a Stochastic Kriging assumption. As seen in section 5.3, SK considers the intrinsic noise to be i.i.d from a Gaussian distribution. The goal here is to compute multiple error measures from a Monte Carlo integration procedure, and verify the normality hypothesis. The integral analyzed can be written as:

$$f(x) = \int_0^1 \exp(x) \cos^2(x) dx, \tag{C.1}$$

The exact value of this integral can be derived analytically as:

$$\text{exact} = \frac{1}{10} \left(e(5 + 2\sin(2) + \cos(2)) - 6 \right) \approx 1.14037, \tag{C.2}$$

where this value is used to compute the error from the Monte Carlo integration. Considering the random variable in the function, it is possible to obtain an estimate of the exact solution as:

$$\frac{1}{n_s} \sum_{i=1}^{n_s} \exp(x) \cos^2(x), \quad \text{with } x \sim U(0,1) \tag{C.3}$$

where U is the uniform distribution and n_s is the number of samples.

The analysis is conducted by taking 5 batches of 2000 independent Monte Carlo runs and subtracting the results from the exact value of the integral. Each batch corresponds to an increasing value of n_s. Each error vector is then tested for normality using the Anderson-Darling test (ANDERSON; DARLING, 1952).

Figure 46 presents the p-value results of the normality test along with an histogram of the data for each fixed number of evaluated samples.

It can be seen from the p-values, that for this case the hypothesis of normality cannot be rejected with 95% confidence in any case. Moreover, the more samples are used more probably normal the errors become. This relates to the Central Limit Theorem, which states that, in most situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution (DEVORE, 2011).

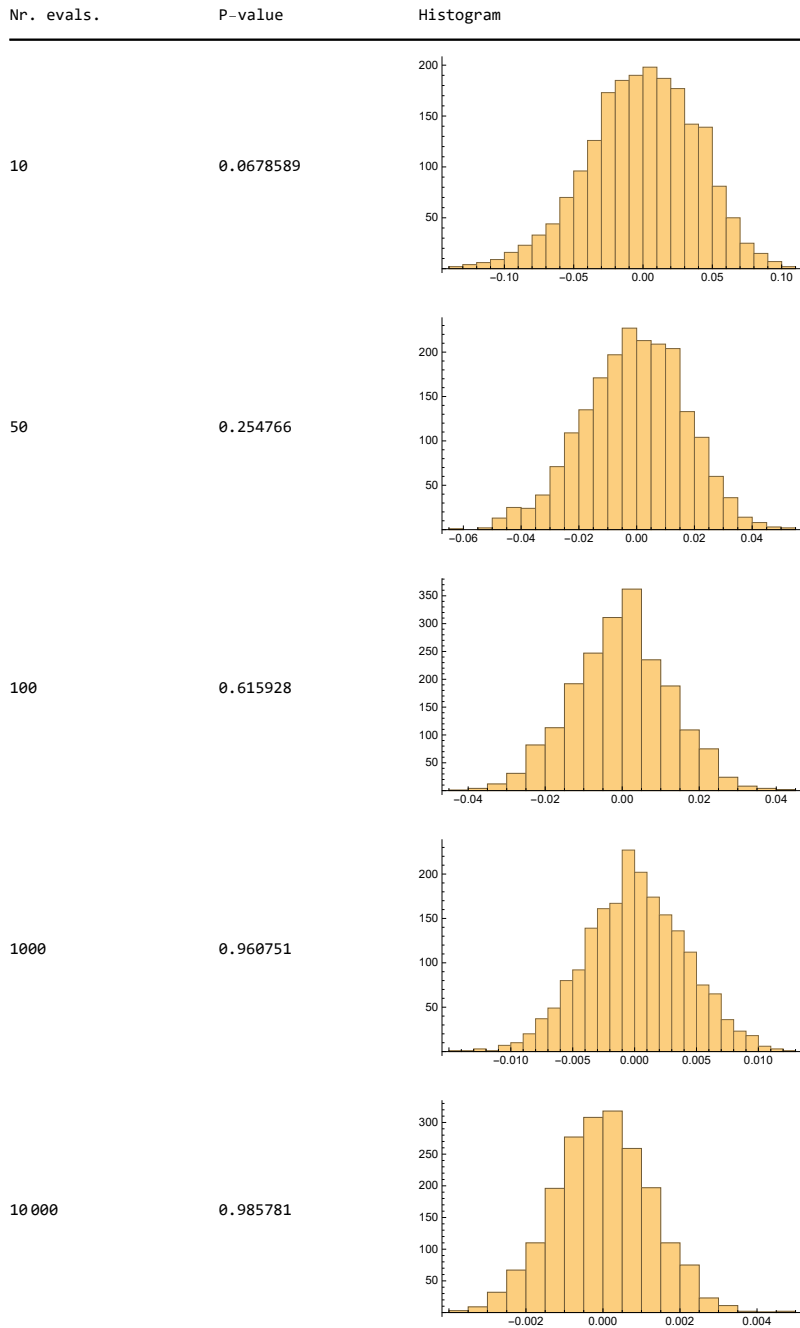| Nr. evals.	P−value	Histogram
10	0.0678589	
50	0.254766	
100	0.615928	
1000	0.960751	
10 000	0.985781	

Figure 46 – Histograms of the intrinsic errors