OPTIMIZATION UNDER UNCERTAINTY: ADAPTIVE VARIANCE REDUCTION, ADAPTIVE METAMODELING, AND INVESTIGATION OF ROBUSTNESS MEASURES

A Dissertation

Submitted to the Graduate School

of the University of Notre Dame

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

by

Juan Camilo Medina

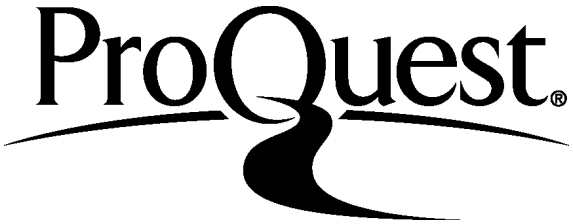
_____

Alexandros A. Taflanidis, Director

Graduate Program in Aerospace and Mechanical Engineering

Notre Dame, Indiana

July 2014



OPTIMIZATION UNDER UNCERTAINTY: ADAPTIVE VARIANCE REDUCTION, ADAPTIVE METAMODELING, AND INVESTIGATION OF ROBUSTNESS MEASURES

Abstract

by

Juan Camilo Medina

This dissertation offers computational and theoretical advances for optimization under uncertainty problems that utilize a probabilistic framework for addressing such uncertainties, and adopt a probabilistic performance as objective function. Emphasis is placed on applications that involve potentially complex numerical and probability models. A generalized approach is adopted, treating the system model as a "black-box" and relying on stochastic simulation for evaluating the probabilistic performance. This approach can impose, though, an elevated computational cost, and two of the advances offered in this dissertation aim at decreasing the computational burden associated with stochastic simulation when integrated with optimization applications.

The first one develops an adaptive implementation of importance sampling (a popular variance reduction technique) by sharing information across the iterations of the numerical optimization algorithm. The system model evaluations from the current

iteration are utilized to formulate importance sampling densities for subsequent iterations with only a small additional computational effort. The characteristics of these densities as well as the specific model parameters these densities span are explicitly optimized. The second advancement focuses on adaptive tuning of a kriging metamodel to replace the computationally intensive system model. A novel implementation is considered, establishing a metamodel with respect to both the uncertain model parameters as well as the design variables, offering significant computational savings. Additionally, the adaptive selection of certain characteristics of the metamodel, such as support points or order of basis functions, is considered by utilizing readily available information from the previous iteration of the optimization algorithm.

The third advancement extends to a different application and considers the assessment of the appropriateness of different candidate robust designs. A novel robustness measure is introduced, the probability of dominance, defined as the likelihood that a given design will outperform its competing designs. This new measure ultimately provides a rational approach to quantify the preference towards each candidate design. The existence of a model prediction error is also addressed within the definition of this measure.

To Kelly and my family

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

I would like to express my appreciation to everybody who offered some form of support for me to reach this milestone. In particular, I owe a great debt of gratitude to my current advisor Professor Alexandros Taflanidis for his guidance, support, and patience throughout the years. Dr. Taflanidis has been an excellent mentor and a role model not only academically, but personally. In addition, my former advisor Dr. John Renaud (R.I.P.) will always be remembered for his guidance and support as well. Special thanks to my dissertation committee members: Dr. Panos Antsaklis, Dr. Timothy Ovaert, and Dr. Jim Schmiedeler for their time and assistance through the dissertation process.

I would like to thank the University of Notre Dame and all its benefactors for their generosity and for offering me the Dean's fellowship, which has allowed me to eat more than ramen noodles while I do research. Special thanks to all my labmates from the NDDAL and HIPAD for all the great times.

Lastly, and more importantly, I would like to give special thanks to my fiancée Kelly, my parents, my siblings, and all my family for their love and care.

CHAPTER 1:

INTRODUCTION

1.1 Background and Motivation

In any engineering design application, the performance predictions for the system under consideration involve some level of uncertainty (Helton 1997; Beck and Gomes 2012; Der Kiureghian and Ditlevsen 2009) stemming from the incomplete knowledge about the system itself and its environment (representing future excitations) – what is generally referenced as epistemic uncertainty – or from the inherent variability in the system/excitation description (for example existence of disturbances of stochastic nature) – what is generally referenced as aleatory uncertainty. Explicitly accounting for these uncertainties is exceptionally important for providing optimal configurations that exhibit robust performance, and toward this goal, increased attention has been given in recent years to non-deterministic techniques for the design of engineering systems (Schuëller and Jensen 2008; Doltsinis 2004). For this purpose, various frameworks have been established for addressing uncertainties in the analysis/design of such systems; examples include interval analysis, fuzzy sets, and probability logic (Ayyub and Gupta 1997; Jaynes 2003). The latter, i.e. a probability logic approach, is perhaps the most popular and is widely acknowledged to provide a rational and consistent framework for

incorporating our incomplete knowledge about the *true system* and its environment into the characterization of the *system model* properties (Beck and Taflanidis 2013). This is established by employing probability distributions to describe the relative plausibility (likelihood) of different uncertain model parameters values (Jaynes 2003). In this setting, the probabilistic performance is described though some statistical measure (i.e., an expectation operator) over these distributions. This dissertation offers theoretical and computational advances in the design of engineering systems when such a probabilistic framework is adopted to characterize the uncertainties in their description, focusing particularly on applications with potentially complex numerical and probability models for which evaluation of the probabilistic performance can typically be established only through stochastic simulation techniques.

1.1.1 Traditional Approaches in Optimization under Uncertainty

Within a probabilistic framework, various formulations have been proposed for the design of engineering systems (Schuëller and Jensen 2008; Doltsinis 2004; Beyer and Sendhoff 2007; Fang et al.; Park et al. 2006; Valdebenito and Schuëller 2010). Undoubtedly, two are traditionally recognized as popular approaches: Reliability-Based Design Optimization (RBDO) and Robust Design Optimization (RDO).

1.1.1.1 Reliability-Based Design Optimization (RBDO)

RBDO utilizes the reliability of the system as either the objective function or more commonly as a constraint (Valdebenito and Schuëller 2010; Aoues 2008; Royset 2006; Jia and Taflanidis 2013; Tu et al. 1999; Enevoldsen and Sørensen 1994; Frangopol

1985). The design of systems that must satisfy some target reliability constraint generally leads to simpler optimization problems, which has been a reason for its increased popularity. Advances in computational/computer science have motivated research over the past decade to directly incorporate reliability as a system objective in the design problem formulation since that approach offers a more rational design framework for many types of problems (Helton 1997; Beck and Gomes 2012; Der Kiureghian and Ditlevsen 2009). The system reliability is quantified by the probability that its performance will not exceed some acceptable threshold, or more generally that a specific limit state is not reached. The selection of this threshold (or more generally the characteristics of the limit state function) has a direct impact on the obtained optimal design (i.e. changing the threshold will dramatically change that design) (Taflanidis et al. 2010). Analytical approximations such as the First Order Reliability Method (FORM) (Chiralaksanakul 2005) or Second Order Reliability Method (SORM) (Der Kiureghian et al. 1987) are commonly employed to estimate the system reliability within the context of the optimization algorithm employed to solve the RBDO problem. Both of these methods utilize only information from the most probable failure point (also known as design point), defined as the model parameter configuration that has the highest likelihood in producing system failure (i.e., satisfy the chosen limit state function), to approximate the system reliability. Thus, approximation of the latter is transformed into an optimization problem for identifying these design point(s).

A relative drawback of RBDO is its lack of theoretical/computational generality since it requires the definition of a single limit state function to describe system

performance, leading to a description of the performance that has a binary distinction (dividing the uncertain domain into two different domains, the safe domain and the failure domain), and the predominant computational tools for its solution employ the concept of design points.

Furthermore, depending on the nature of the limit state function, these computational tools (FORM or SORM) may not always (i.e. for all design configurations) provide accurate approximations, something that can be exploited by the numerical optimization algorithm itself, converging to regions in the design space that correspond to a large reliability approximation simply because of the inaccuracy of the algorithms (FORM or SORM) used to calculate this reliability. This has incentivized researchers to use methods relying on stochastic simulation for solving RBDO problems (Royset 2004; Taflanidis and Beck 2008), although such approaches are not the standard.

1.1.1.2 Robust Design Optimization (RDO)

RDO methods on the other hand, seek to address modeling uncertainties by explicitly attempting to reduce the standard deviation of the system response (Doltsinis 2004; Beyer and Sendhoff 2007; Tsompanakis et al. 2008; Lee et al. 1996). Various approaches have been proposed in the literature for formulating and solving the RDO optimization problems, mainly differentiated by (i) the methodologies adopted for approximating the statistical measures involved and (ii) the assumptions utilized to obtain the optimization objective. Typically, RDO is formulated by considering as objectives the mean value as well as the standard deviation of the performance (though

different statistical measures can be also utilized) (Beyer and Sendhoff 2007; Lee and Park 2001), either within a multi-objective setting (Doltsinis 2004; Marler and Arora 2004) or, and perhaps more commonly, by formulating a single objective through the introduction of an appropriate weight between these two competing objectives (Lee and Park 2001; Dunning and Kim 2013). Different designs can then be attained through this problem formulation either by varying the weight, establishing a different compromise between the mean and standard deviation of the performance (response), or approaching the problem as a truly multi-objective optimization. Frequently, this optimization approach will provide designs that are fundamentally different in the design space, but exhibit similar robust traits (i.e. similar values for the mean or variance). Selecting the most suitable design among a set of such candidate robust designs is then a challenging task, as it requires either some measure of appropriateness, or that a design stands out from the rest because of its superior performance (mean value) and robustness (standard deviation), which is not typically the case as the two competing objectives usually conflict with one another.

The relationship between RBDO and RDO as well as their comparative advantages and disadvantages has been examined in some studies (Beyer and Sendhoff 2007; Yao et al. 2011). In this dissertation, some attention is placed on RDO, especially in the context of evaluating the appropriateness of different solutions obtained within the RDO optimization formulation, which will be a secondary objective of this dissertation. The focus is on a more generalized optimization under uncertainty problem.

1.1.2 Generalized Framework for Optimization under Uncertainty

Beyond RDO and RBDO, a more general framework for optimization under uncertainty problems can be established by considering *any* statistical measure (i.e., expectation operator) of *any* desired system response function. This measure can range from system fragility, to system risk, or to expected life-cycle cost (Aoues 2008; Royset 2006; Taflanidis and Beck 2009). In this framework, the objective function corresponds to a multivariate probabilistic integral (over the domain of the uncertain model parameters), and no other assumptions should be employed to support generality/versatility of the developed tools. Given that for problems involving complex system models, such probabilistic integrals can rarely be calculated or approximated analytically, stochastic simulation [i.e., Monte-Carlo Simulation (MCS)]-based techniques provide the only practical alternative approach for the estimation of the objective function (Taflanidis and Beck 2008; Robert and Casella 2004; Kroese et al. 2011). In MCS, N samples are initially drawn from a predefined distribution (proposal density). Then, the system's performance is evaluated for each one of those samples. Subsequently, the desired expected value is approximated through the statistics of the samples. The popularity of stochastic simulation has been increasing in recent years (Royset 2004; Taflanidis and Beck 2008; Jensen et al. 2009) due to their general applicability as well as the potential of exploiting advances in parallel/distributed computing for efficiently performing the required independent evaluations of the system performance (Papadrakakis and Lagaros 2002). Nonetheless, there exist a number of significant challenges associated with this framework, and the primary

objective of this dissertation is to formulate novel computational approaches to address some of these challenges.

1.1.3 Challenges in Optimization under Uncertainty

The main challenges associated with optimization under uncertainty problems when stochastic simulation is utilized to estimate the objective function are: the existence of an unavoidable estimation error in the approximation (demonstrated in Figure 1.1) and the high computational cost associated with each objective function estimation (N model evaluations are needed to estimate the objective function once).

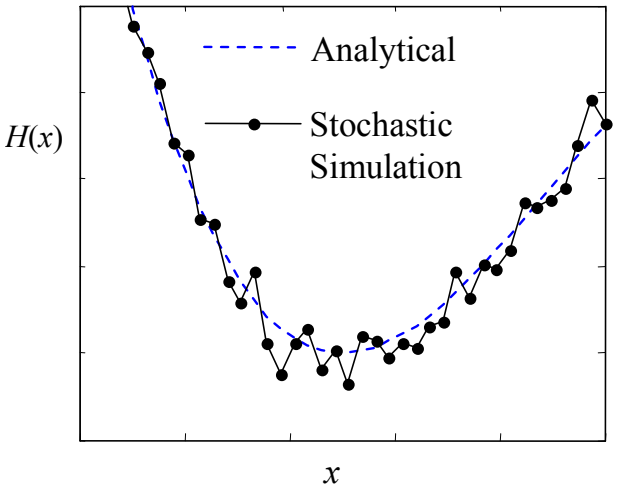


Figure 1.1 Analytical and simulation-based evaluation of an objective function $H(x)$. Note the estimation error in the evaluation based on stochastic simulation

The estimation error may be considered as noise in the objective function and conflicts with classical deterministic optimization algorithms that assume the objective function is exactly known (and commonly assumed as smooth). There exist two distinct approaches through which the impact of the estimation error can be reduced: by

reducing either its *relative* or *absolute* importance. The latter (reduction of absolute importance) can be achieved by using a large enough sample set N or some variance reduction technique to reduce the overall estimation error (Schuëller and Jensen 2008; James 1985). Either of these two requires an additional computational burden either directly (increasing N) or indirectly (implementing and tuning an appropriate variance reduction technique); thus, the focus is frequently on the reduction of the relative importance of the estimation error. This can be achieved by using Common Random Numbers (CRN), to create a consistent estimation error in the comparisons between different design configurations (Glasserman and Yao 1992). This contributes to the increase of the correlation between the estimates, which ultimately decreases the variance of their difference. Important requirements for improving the efficiency of stochastic comparisons when using CRN (Glasserman and Yao 1992) are continuity and monotonicity of the system output with respect to the system model parameters. If that output is sufficiently smooth, these two aforementioned requirements can be typically guaranteed as long as the design choices compared are not too far apart in the design domain, i.e., when the system configurations compared are not too different so that the model parameter values have a similar effect on the system response (Taflanidis and Beck 2010).

The use of CRNs can be integrated in different ways within the context of numerical optimization algorithms (Kleinman et al. 1999). The most straightforward is perhaps through adoption of an Exterior Sampling Approximation (ESA). ESA employs the same stream of random numbers throughout all iterations in the optimization

process, typically with N selected sufficiently large to reduce the absolute influence of the estimation error, thus transforming the problem into a large scale deterministic design problem that can be solved by any conventional algorithm.

The reduction of the absolute importance of the estimation error is an important feature of ESA since the designs compared within the algorithm extend over the entire design domain (and thus can be potentially very different), which does not necessarily guarantee high efficiency of using CRNs (thus cannot rely only on reduction of relative importance of absolute error) as discussed in the previous paragraph. Frequently, gradient-free algorithms are selected in this context since they circumvent difficulties in obtaining derivative information (Beck 1999; Lagaros 2002). Another drawback of this approach consists of its dependence on the exact stream of random numbers used. The quality of the solution obtained through ESA is formally assessed by solving the optimization problem multiple times for different independent random samples streams (Taflanidis and Beck 2008; Spall 2003). Even though the computational cost for the ESA deterministic optimization is typically smaller than the original stochastic problem, the overall efficiency is expected to be worse because of the requirement to perform the optimization multiple times.

The alternative sampling approach is interior sampling (Taflanidis and Beck 2008; Spall 2003). This approach uses a different stream of random numbers at each iteration of the numerical optimization algorithm, although CRN can be still be utilized within each separate iteration if more than one evaluation of the objective function is required, for example if numerical differentiation is employed (more on this will be

discussed in following chapters). It should be stressed that reduction of the absolute importance of the estimation error is not necessary when interior sampling is adopted (the approach can be implemented with a lower sample size N compared to exterior sampling) since the impact of the random number stream is addressed by changing it at each iteration of the numerical optimization. A challenge in this case is that convergence cannot be evaluated based on the objective function value estimates due to the high variability in these estimates (error is not consistent among them since CRN has not been utilized).

These computational challenges associated with stochastic simulation techniques are particularly important when complex numerical models are employed for describing the performance of engineering systems, something that is undeniably becoming an increasingly popular trend due to recent advances in computers and computational science that have encouraged researchers to develop models of higher fidelity/accuracy (Gano 2005; Shan and Wang 2010). When such models are adopted within the stochastic-simulation-based optimization framework, there is a significant increase in the computational burden for each evaluation of the objective function, creating an incentive to develop techniques that can improve the efficiency through either employing adaptive variance reduction techniques to improve the estimation accuracy without requiring a large value for N, or by adopting surrogate modeling (metamodeling) approaches to approximate the initial computational intensive high fidelity model (Beyer and Sendhoff 2007; Shan and Wang 2010; Shapiro 2003; Polak 2008). An additional challenge for such applications is that the system model frequently

needs to be treated as a "black-box" (Shan and Wang 2010), thus requiring either numerical differentiation/approximation (Spall 2003; Spall 1998b) to obtain derivative information or the use of gradient-free algorithms (Jin and Branke 2005).

1.2 Research Objectives

Motivated by the previous challenges, the two general goals of this dissertation are

(a) To improve the computational efficiency of stochastic simulation techniques when implemented within a numerical optimization approach to solve design problems that employ a probabilistic objective function.

(b) To provide a novel robustness measure to assess the appropriateness of a set of candidate designs established within the RDO formulation.

Improvement of efficiency is sought through two distinct approaches: (a.1) by formulating an adaptive importance sampling implementation to serve as a variance reduction technique integrated within the numerical optimization approach or (a.2) by establishing an adaptive surrogate modeling framework that considers simultaneous approximation of the system performance with respect to both the design variables and the model parameters. The first approach is more general and can be implemented for problems with arbitrary numbers of model parameters and design variables, whereas the second one has the fundamental assumption that the number of design variables and model parameters is not large [less than 50 (Simpson et al. 2001c)].

Thus, three distinct specialized ***objectives*** are identified for this dissertation:

1) Formulate an adaptive importance sampling framework for design under uncertainty optimization problems. This framework will improve efficiency of stochastic-simulation-based evaluations of the objective function by sharing information across the iterations of the numerical optimization algorithm.
2) Establish an adaptive surrogate modelling methodology with a similar goal, utilizing specifically a kriging metamodel. Consider a novel implementation of the surrogate model in the augmented model parameters and design variables space and develop efficient techniques for sharing of information between iterations to improve the accuracy of the developed metamodel.
3) Offer a new measure for assessing the appropriateness of different RDO designs and examine the impact of model prediction errors on this measure as well as the RDO problem itself.

The first two objectives share a common foundation as they are based on exploiting readily available information across the iterations of the optimization algorithm in order to improve the numerical efficiency/accuracy during subsequent iterations. To facilitate the sharing of information across these iterations, algorithms that rely on local searches (such as gradient-based algorithms) need to be considered. With this choice, the system configurations in adjacent iterations are similar (particularly as the algorithm approaches a local optimum); therefore, information obtained at the current iteration can be the basis for making decisions for subsequent ones. The third objective pertains to developing a novel robustness measure that can quantify the preference of a robust design among a set of candidate designs. Below some further

background related to each of these objectives is offered, identifying the knowledge gaps this dissertation ultimately aims to fill.

1.2.1 Adaptive Importance Sampling (IS)

For improving the accuracy of stochastic simulation methods, many variance reduction techniques exist, and the most popular is importance sampling (IS) (Robert and Casella 2004; Kroese et al. 2011; Papadrakakis and Lagaros 2002; Schuëller 2004). IS relies on the introduction of proposal densities to concentrate the computational effort in regions of the uncertain model parameter space that have a higher contribution to the overall probabilistic performance, i.e. near the maxima of the associated integrand. Selection of efficient IS densities, i.e. densities that lead to significant improvement in accuracy, however, is a nontrivial task since numerical problems (namely divergence of IS-based estimators) may arise when these densities are not *properly* chosen (Robert and Casella 2004). The divergence typically occurs when the IS densities do not cover the entire region of importance for the integrand. This is particularly true for problems that include a large number of uncertain variables since the IS densities need to be properly chosen for all of them (Au and Beck 2003). In such applications, it is generally advisable to formulate IS densities for a smaller subset of parameters, consisting of only the most influential variables (Taflanidis and Beck 2008; Pradlwarter 2007). Thus, two aspects can be considered as the main challenges related to IS: (a) selection of efficient IS densities, including their characteristics (balancing between improving accuracy and avoiding numerical problems) and (b) selection of the specific variables for IS to target in

high-dimensional problems. Many adaptive approaches have been proposed to address (a), relying primarily on information obtained through samples that are distributed proportional to the probabilistic integrand of interest (Ang 1992; Tang and Chen 2009; Au and Beck 1999; Cappé et al. 2008). For (b), one of the proposed solutions relies primarily on the qualitative characterization of the influence of the different model parameters (Taflanidis and Beck 2008; Pradlwarter 2007). More importantly, little attention has been given to efficient integration of these approaches within the context of optimization applications to minimize the computational burden in selecting IS densities.

The proposed research focuses on bridging the foregoing knowledge gap. It focuses on the adaptive implementation of importance sampling (IS) across the iterations of the numerical optimization. An optimal formulation of IS densities is developed, extending to the selection of the type/shape of the IS distribution and of the exact model parameters to concentrate in the IS formulation, and to the rules for the efficient sharing of information across the iterations of the numerical optimization to minimize the computational burden in developing the IS densities.

1.2.2 Adaptive Surrogate Modeling in the Augmented Design Variable and Model Parameter Space

Surrogate models (also known as metamodels) replace a computationally expensive model (normally a high fidelity model) with an inexpensive approximation (Gano 2005; Kleijnen 1987) based on information (evaluation of the expensive model)

gathered at some specially selected points, generally known as support points. Several techniques exist within this greater family, such as polynomial response surfaces, radial basis functions, kriging, and neural networks (Simpson et al. 2001c; Jin et al. 2001). All of them have their own relative advantages and disadvantages, and the focus of this research effort is on the kriging metamodeling approach since it is generally acknowledged to offer accurate and computationally efficient approximations (Jin et al. 2003; Simpson et al. 2001b). Its advantages include: it is an exact interpolator, meaning that the error at the supports points is zero; it corresponds to the Best Linear Unbiased Predictor, BLUP, (Sacks et al. 1989); it facilitates an easy calculation of gradient information; and it not only gives the prediction but also the associated uncertainty, namely the local variance of the prediction error which can be explicitly incorporated in the probabilistic performance assessment (Jia and Taflanidis 2013).

A critical issue for the accuracy/efficiency of surrogate models is the selection of support points, an approach generally known as design of experiments (DoE). The most popular methodology to accomplish this goal is to adopt a space filling technique such as Latin Hypercube (LHC) sampling to evenly populate the domain of interest (Wang and Shan 2007; Simpson et al. 2001a). This is expected to result in a smaller global error between the surrogate model and the function that is approximated, i.e. it minimizes the Integrated Mean Squared Error (IMSE) between these two (Park 1994; Van Beers and Kleijnen 2004). Adaptive DoE has also been considered (Picheny et al. 2010; Bect et al. 2012); in this case, a small number of initial designs is established (through a space-filling technique), and then, based on this information, each additional design is

obtained by formulating an optimization problem to select the support point that is expected to minimize the IMSE.

This optimization includes a substantial computational burden but for problems with highly complex numerical models (i.e. when the computational cost for performing such an optimization is negligible when compared to the computational cost for evaluating the system model) it can ultimately contribute to significant computational savings without compromising the accuracy of the surrogate model.

Undoubtedly the most popular implementation of surrogate modeling concepts within optimization applications is for the direct approximation of the objective function (Simpson et al. 2001c; Jin et al. 2003; Wang and Shan 2007; Persson and Ölvander 2013; Lee et al. 2006; Huang et al. 2006). Nevertheless, surrogate modeling approaches have been also considered within optimization under uncertainty design problems for directly approximating the system response in the uncertain model parameter space (Jin et al. 2003; Persson and Ölvander 2013; Eldred et al. 2002; Choi et al. 2001; Taflanidis 2012). The design of experiments in this case has an even increased complexity since it becomes crucial to obtain an accurate approximation in target regions of the model parameters, the regions that contribute more to the integrand representing the probabilistic performance (Picheny et al. 2010; Dubourg et al. 2011a). In other words, minimization of IMSE is not necessarily an appropriate objective for DoE, and modifications are needed to obtain accurate approximations in a target region. Some of the aforementioned approaches have considered interaction-DoE formulations with this goal in mind (Youn and Choi 2004) but have been constrained within RBDO formulations

where the target region is conveniently just the failure domain. Additionally, these efforts have not considered a setting relying on stochastic simulation (instead FORM and SORM solutions were adopted in the RBDO framework). Finally, the potential development of surrogate models in the augmented design variables and model parameter space has not received appropriate attention. This concept has been recently introduced (Taflanidis 2009), primarily out of necessity since surrogate modeling was considered in the context of an augmented design formulation in which the design variables themselves were uncertain. A more thorough investigation is needed to properly explore the potential of such an approach.

The proposed research focuses on addressing these challenges and knowledge gaps. It considers implementation of kriging for generalized optimization under uncertainty problems when stochastic simulation is utilized to evaluate the objective function. Formulation of kriging within the augmented space of the design variables and model parameters is established, and an adaptive selection of support points and tuning of the kriging characteristics is considered by sharing information across the iterations of the numerical optimization algorithm (similar to the ideas established for adaptive IS). Specifically for DoE, a hybrid approach is established combining a space filling stage and stage targeting a specific region (where the integrand is considered important). In contrast to other approaches (Picheny et al. 2010; Zhao et al. 2011; Dubourg et al. 2011b), the proposed DoE targeteted support point selection is established with no additional computational burden, relying explicitly on information readily available within the optimization algorithm itself. The same approach is

introduced for automatic selection of the order of basis functions needed for the kriging approximation. Finally, the explicit consideration of the kriging prediction in the probabilistic performance assessment is also addressed.

1.2.3 Assessing Appropriateness of RDO Designs

As outlined earlier, RDO problems typically lead to different candidate designs through modification of the relative importance (to the overall performance objective) of the mean and standard deviation of the system performance (response). Selecting the most suitable design among such sets of candidate robust designs is generally a challenging task as the two competing objectives usually conflict with one another. The common approach (Beyer and Sendhoff 2007) for choosing the preferred design is to use the distance (defined through some appropriate norm) from the ideal state, also known as utopia point, corresponding to the minimum mean and standard deviation treated as single-objectives for each instance. This requires definition of a proper norm (to quantify distance), and establishing a proper weighting of the two objectives for this purpose is not always straightforward.

An alternative formulation is considered in this dissertation based on a novel robustness measure, termed probability of dominance, for assessing the appropriateness of different candidate designs. This measure quantifies the likelihood that a particular design will outperform the rival designs within a candidate set. The impact of prediction errors between the real system and the assumed numerical model are also addressed within this setting of assessing appropriateness/robustness of

different designs. Though this error has been previously considered in RBDO problems (Taflanidis and Beck 2010), that consideration was rather simple, focusing on direct incorporation of a specific probabilistic model for the prediction error within the optimal design problem. The research here extends to a deeper evaluation of the impact of the prediction error on RDO design and on the probability of dominance, including the consideration of different probability models for this error.

1.2.4 Outline of the Dissertation

In Chapter 2, the optimization under uncertainty design problem is discussed. The generalized formulation and the RDO description are presented. Furthermore, some common computational tools that are employed to facilitate different aspects of this investigation are introduced. These are: Importance Sampling (IS), rejection sampling, Kernel Density Estimation (KDE), and probabilistic global sensitivity analysis.

Chapters 3 and 4 are devoted to the development of methods that increase the computational efficiency of existing optimization algorithms when stochastic simulation is employed to estimate the objective function. In particular, Chapter 3 focuses on an adaptive integration of IS within the numerical optimization algorithms. The challenges of adaptively selecting the characteristics of the IS proposal density as well as the most appropriate model parameters for the application of IS are addressed. The framework is illustrated in an example considering the optimization of passive and semi-active dampers for the suspension of a half-car model driving on a rough road. Chapter 4 then discusses an adaptive kriging implementation. Formulation of the metamodel in the

augmented design-uncertain model parameter space is considered, and important issues such as the automatic selection of the order of basis functions for the kriging approximation and the selection of support points for a targeted region are addressed. The framework is illustrated for the same example as in Chapter 3.

Chapter 5 then deals with assessing the appropriateness of different candidate designs within the RDO formulation. For this purpose, a novel measure, the probability of dominance, is introduced and the impact of prediction errors for the system model is investigated. Two different examples are considered to illustrate the framework; the first corresponds to the optimization of a tuned mass damper and the second to topology optimization for minimum compliance.

Lastly, Chapter 6 discusses the significance of the contributions presented in this dissertation as well as future potential extensions.

CHAPTER 2:

OPTIMIZATION UNDER UNCERTAINTY PROBLEM

In this chapter, the general optimization under uncertainty problem formulation is initially outlined with some emphasis on RDO problems. Then a number of numerical tools that will be used throughout the dissertation are discussed, namely (a) Importance sampling for improving the numerical accuracy in estimation of probabilistic integrals, (b) rejection sampling for obtaining samples from arbitrary distributions, (c) kernel density estimation for approximating distributions based on samples and (d) global sensitivity analysis for identifying the importance of individual model parameters to the overall probabilistic performance.

A common nomenclature used in this chapter and throughout the proposal is presented in Table 2.1. In the remaining chapters, individual nomenclatures for the terminology established within that chapter are additionally offered.

TABLE 2.1

RELEVANT NOMENCLATURE FOR CHAPTER 2

γ	Maximum of the quotient used in rejection sampling	$\hat{\cdot}$	(Hat) Estimate using stochastic simulation
δ	Coefficient of variation	$\cdot_i$	(Subscript) i^{th} component of vector
$\boldsymbol{\theta}$	uncertain model parameters	$\cdot^j$	(Superscript) j^{th} sample
Θ	Uncertain Space	$K(.)$	One dimensional kernel
$\{\boldsymbol{\theta}^j\}$	Sample set	$k_d(.)$	Kernel density estimate using fix-window bandwidth
$\pi(\boldsymbol{\theta}\|\mathbf{x})$	PDF proportional to the integrand of the objective function for given $\mathbf{x}$	N	Number of samples for stochastic simulation
$\pi(\theta_i\|\mathbf{x})$	Marginal distribution for θ_i based on $\pi(\boldsymbol{\theta}\|\mathbf{x})$	N_s	Total number of kernel samples
$\tilde{\pi}(\theta_i\|\mathbf{x})$	KDE estimate for marginal distribution for θ_i	n_θ	Dimension of $\boldsymbol{\theta}$
σ_i	Standard deviation of samples from $\pi(\boldsymbol{\theta}\|\mathbf{x})$ for θ_i	n_x	Dimension of $\mathbf{x}$
$D(.\|\|.)$	Relative entropy between two arguments	$p(\boldsymbol{\theta})$	Probability density function for $\boldsymbol{\theta}$
$E[.]$	Expectation operator	$r(.)$	Importance sampling quotient
$E_p[.]$	Expectation under probability distribution p	t_i	fixed-window bandwidth parameter for KDE for θ_i
$f(\boldsymbol{\theta})$	Importance sampling probability density function	u	Uniformly distributed random variable
$f^p(\boldsymbol{\theta})$	Proposal density in rejection sampling for generating samples from $\pi(\boldsymbol{\theta}\|\mathbf{x})$	$Var_p[.]$	Variance under probability distribution p
$h(\mathbf{x},\boldsymbol{\theta})$	Performance function	X	Admissible design space
$H(\mathbf{x})$	Objective function	$\mathbf{x}$	Design variables
$\hat{H}(\mathbf{x}\|\{\boldsymbol{\theta}^j\})$	Estimate of objective function through stochastic simulation using set $\{\boldsymbol{\theta}^j\}$	$\mathbf{x}^*$	Optimal design solution

2.1 Optimization under Uncertainty Problem Formulation

Consider a system that involves some controllable parameters that define its design, referred to as design variables, and let $\mathbf{x} = [x_1, x_2, ..., x_{n_x}] \in X \subset \mathbb{R}^{n_x}$ be the design vector, where X denotes the bounded admissible design space. Let $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_{n_\theta}] \in \Theta \subset \mathbb{R}^{n_\theta}$ be the vector of uncertain model parameters (random variables) of the system, where Θ denotes the set of their possible values (uncertain space). In addition, let $h(\mathbf{x}, \boldsymbol{\theta}) : \mathbb{R}^{n_x} \times \mathbb{R}^{n_\theta} \to \mathbb{R}$ be the performance function characterizing the favorability of the system-model response. Note that this ultimately corresponds to an augmented model description, and generally a distinction of separate excitation (disturbance), system, and performance evaluation models is established in most engineering applications (these can be then augmented into a single system). The first two models provide the system response, denoted by $\mathbf{z}(\mathbf{x}, \boldsymbol{\theta}) \subset \mathbb{R}^{n_z}$ herein, and the performance evaluation model $h(\mathbf{x}, \boldsymbol{\theta}) \subset \mathbb{R}$ assesses the favorability of that response. For many practical engineering applications, evaluation of the system model response corresponds frequently to a computationally demanding and black-box type of task. The developments offered in this dissertation are particularly appropriate for such applications. Even for these types of problems, the performance evaluation model is typically of lower complexity, a characteristic that is exploited in Chapter 4. Note also that in most engineering applications, $h(\mathbf{x}, \boldsymbol{\theta})$ is a strictly positive function, although that assumption is not explicitly enforced herein.

To address our incomplete knowledge about the system and excitation characteristics, a Probability Density Function (PDF) $p(\boldsymbol{\theta})$ is assigned to the model parameters. Selection of $p(\boldsymbol{\theta})$ depends on the specific design problem considered; for example, it can be established as a state-of-knowledge model based on information theory principles (Jaynes 2003). The focus on this dissertation is not on how to select that PDF but rather on how to solve the resultant design problem.

The performance function $h(\mathbf{x},\boldsymbol{\theta})$ depends upon both the design variables $\mathbf{x}$ as well as uncertain parameters $\boldsymbol{\theta}$, and it is a random variable itself (stemming from the uncertainty in $\boldsymbol{\theta}$). The probabilistic performance is given by the expected value of $h(\mathbf{x},\boldsymbol{\theta})$ over the uncertain model parameter space

$$H(\mathbf{x}) \triangleq E_p[h(\mathbf{x},\boldsymbol{\theta})] = \int_{\Theta} h(\mathbf{x},\boldsymbol{\theta})\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{2.1}$$

where $E[.]$ corresponds to the expectation operator and $E_p[.]$ denotes explicitly the expectation under probability model $p(\boldsymbol{\theta})$. Different selections for $h(\mathbf{x},\boldsymbol{\theta})$ facilitate different descriptions for the probabilistic performance. The formulation in Eq. (2.1) is thus very powerful, and it can be used to describe problems ranging from probability of failure, to system risk, to expected life-cycle cost, to RDO (Aoues 2008; Royset 2006; Taflanidis and Beck 2009).

Within this setting, the density that is proportional to the absolute value of the integrand has a particular importance, as it provides critical information about the local and global behavior of the probabilistic performance. This density is defined as

$$\pi(\boldsymbol{\theta} \mid \mathbf{x}) = \frac{|\, h(\mathbf{x}, \boldsymbol{\theta}) \,|\, p(\boldsymbol{\theta})}{\int_{\Theta} |\, h(\mathbf{x}, \boldsymbol{\theta}) \,|\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto |\, h(\mathbf{x}, \boldsymbol{\theta}) \,|\, p(\boldsymbol{\theta}), \tag{2.2}$$

where $\propto$ denotes proportionality and the denominator in the first equality corresponds simply to a normalization constant [so that integration of $\pi(\boldsymbol{\theta} \mid \mathbf{x})$ over Θ yields one] and is referenced as ζ_n henceforth (but not explicitly needed).

The typical convention that smaller values of $h(\mathbf{x}, \boldsymbol{\theta})$ correspond to a more favorable performance, i.e. $h(\mathbf{x}, \boldsymbol{\theta})$ corresponds to a certain cost function, is adopted herein. We are then interested in the optimal design problem

$$\mathrm{x}^{*} = \arg\min_{\mathbf{x} \in X} H(\mathbf{x}), \tag{2.3}$$

where any deterministic constraints have been incorporated in the definition of the admissible design space X.

2.2 Solution through Stochastic Simulation

For problems with complex system models, analytical evaluation or the probabilistic integral in Eq. (2.1) is frequently impossible, and as discussed in the introduction, stochastic simulation (i.e., Monte Carlo (MC) estimation) is considered throughout this research effort for this purpose. In its simplest form, Direct Monte Carlo simulation (advances will be discussed later) is a set of random samples (numbers) $\{\boldsymbol{\theta}^j\}$ that are independent identically distributed (i.i.d) according to $p(\boldsymbol{\theta})$, denoted $\boldsymbol{\theta}^j \sim p(\boldsymbol{\theta})$, and is employed to obtain the approximation of the performance function given by the sample average of the system performance function

$$\hat{H}(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N} h(\mathbf{x},\boldsymbol{\theta}^j). \tag{2.4}$$

The notation $\hat{H}(\mathbf{x}\,|\,\{\boldsymbol{\theta}^j\})$ is also utilized to explicitly denote the dependence of $\hat{H}(\mathbf{x})$ on the sample set $\{\boldsymbol{\theta}^j\}$. Herein, superscripts are employed to denote sample and {.} to denote sample set. The stochastic simulation process is illustrated in Figure 2.1.

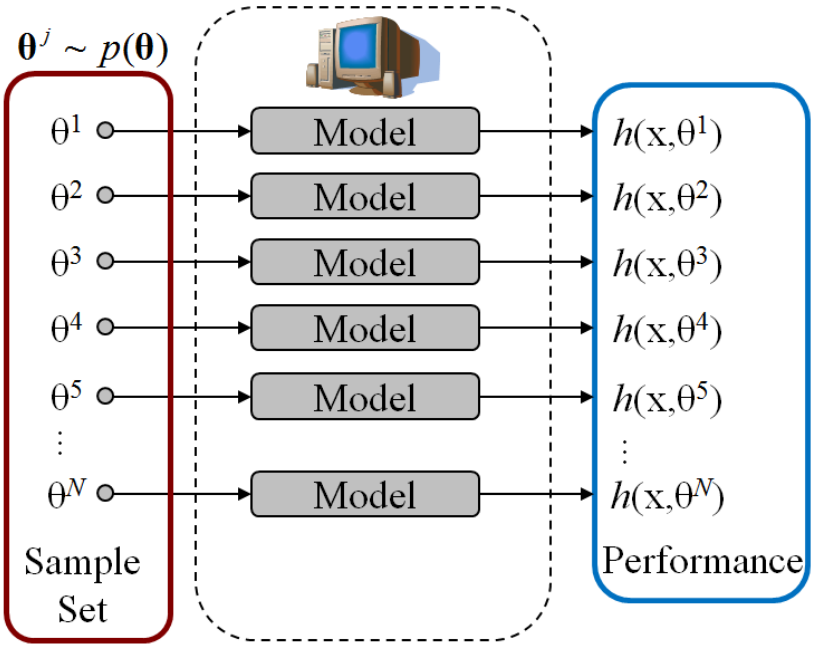


Figure 2.1 Illustration of the stochastic simulation process

The accuracy of $\hat{H}$ can be determined by the Coefficient of Variation (CoV) δ, which can be obtained utilizing the sample set $\{\boldsymbol{\theta}^j\}$.

$$\begin{aligned}\delta &= \frac{1}{\sqrt{N}}\frac{\sqrt{Var_p[h(\mathbf{x},\boldsymbol{\theta})]}}{H(\mathbf{x})} = \frac{1}{\sqrt{N}}\frac{\sqrt{E_p\left[\left(h(\mathbf{x},\boldsymbol{\theta})\right)^2\right]-H(\mathbf{x})^2}}{H(\mathbf{x})},\\ &= \frac{1}{\sqrt{N}}\sqrt{\frac{E_p\left[\left(h(\mathbf{x},\boldsymbol{\theta})\right)^2\right]}{H^2(\mathbf{x})}-1} \approx \frac{1}{\sqrt{N}}\sqrt{\frac{\frac{1}{N}\sum_{j=1}^{N}\left(h(\mathbf{x},\boldsymbol{\theta}^j)\right)^2}{\hat{H}^2(\mathbf{x})}-1},\end{aligned} \tag{2.5}$$

where $Var_p[.]$ stands for variance under probability model $p(\boldsymbol{\theta})$ and is defined as

$$Var_p[h(\mathbf{x},\boldsymbol{\theta})] = E_p\left[h\left(\mathbf{x},\boldsymbol{\theta}\right)^2\right]-E_p\left[h\left(\mathbf{x},\boldsymbol{\theta}\right)\right]^2 = E_p\left[h\left(\mathbf{x},\boldsymbol{\theta}\right)^2\right]-H(\mathbf{x})^2, \tag{2.6}$$

where $H(\mathbf{x})$ can be approximated by Eq.(2.4) and the second moment similarly by

$$E_p[h(\mathbf{x},\boldsymbol{\theta})^2] \simeq \frac{1}{N}\sum_{j=1}^{N}h^2(\mathbf{x},\boldsymbol{\theta}^j). \tag{2.7}$$

Thus, the stochastic simulation approach provides not only an estimate for the objective function but simultaneously a measure of the accuracy of that estimate. The latter can be utilized to guide the selections for the sample size used (choose N to establish a required accuracy threshold). Additionally, due to the independence of the samples $\{\boldsymbol{\theta}^j\}$, the approach can easily exploit parallel implementation for the evaluation of the system performance $\{h(\mathbf{x},\boldsymbol{\theta}^j); j=1,...,N\}$, a characteristic that has contributed to the increasing popularity of such stochastic simulation-based techniques for uncertainty propagation (stemming from the increased popularity of multi-core computer architecture and high-performance clusters).

Finally, optimization in Eq. (2.3) may be performed by substituting $H(\mathbf{x})$ for $\hat{H}(\mathbf{x})$ in Eq. (2.4) as the objective function, leading to the stochastic optimization problem

$$\mathrm{x}^{*} = \arg\min_{\mathbf{x}\in X} \hat{H}(\mathbf{x}). \tag{2.8}$$

A review of appropriate techniques and algorithms for this problem is presented in Section 1.1.3.

2.3 Robust Design Optimization (RDO)

RDO problems are of particular interest in this research (see Chapter 5). In this case, two different statistical measures of $h(\mathbf{x},\boldsymbol{\theta})$ are typically employed to characterize and address the impact of uncertainty in the description of the system model (Doltsinis 2004; Park et al. 2006). The first measure is the mean value for $h(\mathbf{x},\boldsymbol{\theta})$ given by

$$\mu(\mathbf{x}) = E_p\left[h(\mathbf{x},\boldsymbol{\theta})\right] = \int_{\Theta} h(\mathbf{x},\boldsymbol{\theta})\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{2.9}$$

whereas the second measure is the standard deviation of $h(\mathbf{x},\boldsymbol{\theta})$, given by

$$\sigma(\mathbf{x}) = \sqrt{E_p\left[h\left(\mathbf{x},\boldsymbol{\theta}\right)^2\right] - E_p\left[h\left(\mathbf{x},\boldsymbol{\theta}\right)\right]^2} = \sqrt{\int_{\Theta} h(\mathbf{x},\boldsymbol{\theta})^2\, p(\boldsymbol{\theta}) d\boldsymbol{\theta} - \mu(\mathbf{x})^2}. \tag{2.10}$$

Note that $\sigma^2(\mathbf{x})$ corresponds to the variance of $h(\mathbf{x},\boldsymbol{\theta})$, $Var_p[h(\mathbf{x},\boldsymbol{\theta})]$, whereas $\mu(\mathbf{x})$ is ultimately equivalent to the objective function used in the generalized optimization under uncertainty problem formulation, $H(\mathbf{x})$.

RDO establishes a compromise between the minimization of the mean value of the performance and the minimization of the spread of this performance around this

mean value. The former directly influences the robustness characteristics of the approach by implicitly providing a measure of variability in the model. These two statistical measures define the feasible objective space given by $\mathbf{F}(\mathbf{x}) := [\mu(\mathbf{x})\sigma(\mathbf{x})] \in \mathsf{F} \subset \mathbb{R}^2$. Often, these conflict with each other; therefore, the RDO problem can be viewed as a multi-objective minimization of $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$. A key feature of the objective space F is the utopia point, corresponding to the minimum of $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$, when each of them is viewed as a separate objective. This point corresponds to the best solution in F, but in general is unattainable. The compromise between $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ is typically expressed through the Pareto front, composed of the designs for which no other feasible design exists that will simultaneously improve both competing objectives (in other words, it cannot make one better without making the other worse) (Marler and Arora 2004).

Several different methods have been proposed for formulating/solving the multi-objective RDO problem (Marler and Arora 2004; Deb and Gupta 2006). A popular approach is to employ the weighed sum approximation as the objective function in which a single objective is formed by employing a weighted linear combination of $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$. It should be stressed that through this approach the entire Pareto front cannot necessarily be obtained; only solutions lying on its convex approximation are ultimately identified. Nevertheless, this is the most popular formulation for RDO problems, and the one adopted in this study, leading to the optimization problem of finding

$$\mathrm{x}^{*} = \arg\min_{\mathbf{x}\in X} \; H_{RDO}(\mathbf{x}), \tag{2.11}$$

where, again, any deterministic constraints are incorporated in the definition of the admissible design space X and the objective function is defined as

$$H_{RDO}(\mathbf{x}) = (1-w)\frac{\mu(\mathbf{x})}{\mu_n} + w\frac{\sigma(\mathbf{x})}{\sigma_n}, \tag{2.12}$$

where $w \in [0,1]$ is the weight parameter and μ_n, σ_n are normalization constants for the mean and standard deviation, respectively. Commonly, μ_n and σ_n are selected as the minimum mean and standard deviation (Lee and Park 2001), denoted by μ_o and σ_o, that can be obtained by solving Eq. (2.11) when considering $w = 0$ and $w = 1$, respectively. The point [μ_o σ_o] is in fact the utopia point. Moreover, the solutions to the probabilistic integrals in Eqs. (2.9) and (2.10) are challenging to obtain since close form solutions often do not exist. Thus, different approaches have been proposed in the literature for solving the optimization problem in Eq. (2.11), mainly differentiated by the methodologies adopted for approximating the aforementioned integrals. References (Doltsinis 2004; Beyer and Sendhoff 2007) offer a detailed review of these methods in the context of RDO. Within this study, a stochastic simulation approach, as outlined in the previous section, is adopted for evaluation of the necessary statistical measures.

2.4 Computational tools

This Section reviews a variety of computational tools that are frequently utilized in this investigation.

2.4.1 Importance Sampling (IS)

The premise of importance sampling is to introduce a *proposal density* $f(\boldsymbol{\theta})$ that (when selected properly) concentrates the computational effort within a stochastic simulation setting in regions of the uncertain model parameter space that have higher contributions to the overall probabilistic performance, i.e. near the maxima of the associated integrand.

For this purpose the integral in Eq. (2.1) is first transformed to

$$H(\mathbf{x}) = \int_{\Theta} h(\mathbf{x},\boldsymbol{\theta}) \frac{p(\boldsymbol{\theta})}{f(\boldsymbol{\theta})} f(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int_{\Theta} \left(h(\mathbf{x},\boldsymbol{\theta}) r(\boldsymbol{\theta}) \right) f(\boldsymbol{\theta}) d\boldsymbol{\theta} = E_f[h(\mathbf{x},\boldsymbol{\theta}) r(\boldsymbol{\theta})]. \quad (2.13)$$

In this case, $r(\boldsymbol{\theta}) = p(\boldsymbol{\theta}) / f(\boldsymbol{\theta})$ is defined as the IS quotient. The estimate for Eq. (2.1) is then obtained though

$$\hat{H}(\mathbf{x}) = \frac{1}{N} \sum_{j=1}^{N} h(\mathbf{x},\boldsymbol{\theta}^j) r(\boldsymbol{\theta}^j), \quad (2.14)$$

where $\{\boldsymbol{\theta}^j\}$ are independent identically distributed (i.i.d) samples of the model parameters that follow the distributions $f(\boldsymbol{\theta})$ [instead of $p(\boldsymbol{\theta})$ utilized in direct Monte Carlo estimation]. The coefficient of variation of the estimate in Eq. (2.14) is now computed as

$$\begin{aligned} \delta &= \frac{1}{\sqrt{N}} \frac{\sqrt{Var_f[h(\mathbf{x},\boldsymbol{\theta}) r(\boldsymbol{\theta})]}}{H(\mathbf{x})} = \frac{1}{\sqrt{N}} \frac{\sqrt{E_f\left[\left(h(\mathbf{x},\boldsymbol{\theta}) r(\boldsymbol{\theta}) \right)^2 \right] - H(\mathbf{x})^2}}{H(\mathbf{x})}, \\ &= \frac{1}{\sqrt{N}} \sqrt{\frac{E_f\left[\left(h(\mathbf{x},\boldsymbol{\theta}) r(\boldsymbol{\theta}) \right)^2 \right]}{H(\mathbf{x})^2} - 1} \approx \frac{1}{\sqrt{N}} \sqrt{\frac{\frac{1}{N} \sum_{j=1}^{N} \left(h(\mathbf{x},\boldsymbol{\theta}^j) r(\boldsymbol{\theta}^j) \right)^2}{\hat{H}(\mathbf{x})^2} - 1}, \end{aligned} \quad (2.15)$$

where, like in Eq. (2.5), the approximation in the last equality is established by using stochastic simulation to calculate the necessary probabilistic integrals.

The optimal (but impractical) choice for IS density is (Robert and Casella 2004; Kroese et al. 2011)

$$f^*(\boldsymbol{\theta}) = \frac{|h(\mathbf{x},\boldsymbol{\theta})|\, p(\boldsymbol{\theta})}{\int_{\Theta} |h(\mathbf{x},\boldsymbol{\theta})|\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto \pi(\boldsymbol{\theta}\,|\,\mathbf{x}), \tag{2.16}$$

which is equal to the density introduced in Eq. (2.2). If properly chosen, IS can lead to significant improvement in accuracy, especially when the integrand in Eq. (2.1) has a significantly different shape than $p(\boldsymbol{\theta})$ because creating samples according to the latter distribution (corresponding to a direct Monte Carlo approach) concentrates the computational effort in regions with small importance to $H(\mathbf{x})$. In addition, it means that $h(\mathbf{x},\boldsymbol{\theta})$ exhibits strong sensitivity with respect to $\boldsymbol{\theta}$, high enough to create a difference between $p(\boldsymbol{\theta})$ and $h(\mathbf{x},\boldsymbol{\theta})p(\boldsymbol{\theta})$. In the design space X, this is anticipated to occur more prominently in regions closer to local minima, where smaller values of $h(\mathbf{x},\boldsymbol{\theta})$ are anticipated for regions in which the density $p(\boldsymbol{\theta})$ has larger values, leading to bigger differences between $p(\boldsymbol{\theta})$ and the integrand.

Based on Eqs. (2.16) and (2.15), the common practical guidelines (Taflanidis and Beck 2008; Robert and Casella 2004) for developing IS densities are: (i) concentrating most of the probability distribution (and thus the effort in the stochastic simulation) close to the local maxima of the integrand in Eq. (2.1), while also (ii) avoiding large values for the IS quotient $r(\boldsymbol{\theta})$ stemming from small values of $f(\boldsymbol{\theta})$ compared to $p(\boldsymbol{\theta})$ in regions where the integrand is still relatively important. The latter means that the

tails of distributions are of significance for the IS formulation and may even lead, at the extreme case where $r(\boldsymbol{\theta})$ is not bounded, to convergence problems for the estimator in Eq. (2.14) (Robert and Casella 2004). This challenge can be avoided by having IS densities with big enough bandwidth to cover most of the regions for which $p(\boldsymbol{\theta})$ is important. An excessively large spread, however, reduces the effectiveness of the IS densities, i.e. increases the coefficient of variation. Ultimately, selection of IS densities (especially their spread) is always a compromise between the aforementioned (i)-(ii) goals.

2.4.2 Rejection Sampling

Rejection sampling is a versatile algorithm for obtaining independent samples from a particular distribution (frequently referenced as the *target distribution*) based on samples that have been generated from a different distribution (Robert and Casella 2004). In this investigation, it is utilized to obtain samples from the auxiliary density function $\pi(\boldsymbol{\theta}|\mathbf{x})$ of Eq. (2.2).

In particular, implementation of rejection sampling is considered here in the context of the estimation of the objective function using stochastic simulation, when evaluations of $h(\mathbf{x},\boldsymbol{\theta})$ are readily available for a set $\{\boldsymbol{\theta}^j\}$ corresponding to a proposal density $f^p(\boldsymbol{\theta})$. This proposal density may correspond to $f(\boldsymbol{\theta})$ in the context of evaluating Eq. (2.14) or $p(\boldsymbol{\theta})$ in the context of evaluating Eq. (2.4). In this case, and since the evaluations have been already established, the rejection sampling simplifies to the following procedure. First estimate

$$\gamma = s_f \cdot \max_j \left[|h(\mathbf{x}, \boldsymbol{\theta}^j)| \frac{p(\boldsymbol{\theta}^j)}{f^p(\boldsymbol{\theta}^j)} \right], \tag{2.17}$$

where $s_f > 1$ is a factor to guarantee that γ is larger than the maximum of the quantity in brackets estimated over the entire domain Θ and not simply over the samples available. Then for each sample compare

$$\frac{|h(\mathbf{x}, \boldsymbol{\theta}^j)| \frac{p(\boldsymbol{\theta}^j)}{f^p(\boldsymbol{\theta}^j)}}{u^j} > \gamma, \tag{2.18}$$

where u^j are samples uniformly distributed in $[0,1]$. If the inequality at Eq. (2.18) holds, the sample $\boldsymbol{\theta}^j$ corresponds to a sample from $\pi(\boldsymbol{\theta}|\mathbf{x})$. The efficiency of this approach, evaluated by the number of samples obtained, is on average $N/(\zeta_n \gamma)$ (Robert and Casella 2004), where ζ_n is the normalization constant for $\pi(\boldsymbol{\theta}|\mathbf{x})$. This shows that s_f should be selected as small as possible. In general, the efficiency of rejection sampling depends on how similar $f^p(\boldsymbol{\theta})$ is to the target density.

2.4.3 Approximation of Density Based on Samples and Kernel Density Estimation (KDE)

Approximation of an unknown probability density function utilizing only information from samples from it is a problem encountered frequently in engineering applications (McLachlan and Peel 2004; Martinez and Martinez 2007). Parametric approaches can be implemented for this purpose; this involves choosing a family of densities parameterized through some vector and then selecting values for that vector based on the information in the available samples (match statistical characteristics or maximize likelihood of samples) (Martinez and Martinez 2007). The challenge with

these parametric approaches is that for the simpler ones, the efficiency depends on how closely the family of densities considered matches the target density, whereas for the more advanced ones that in general facilitate more robust approximations, the definition of their characteristics involves some extra computational burden. For example, mixture models typically require an explicit optimization for the number of densities considered and potentially a clustering of the samples each mixture model approximates.

Beyond parametric approaches, nonparametric techniques also exist and generally enjoy greater robustness and efficiency. Kernel density estimation is widely acknowledged as one of the most powerful density estimation techniques in this family and is the one ultimately preferred in this dissertation. KDE is obtained by centering some smooth, symmetric kernel over each of the samples (see Figure 2.2). This kernel has a specific shape and bandwidth, i.e. the region over each sample that the kernel ultimately impacts. As long as sufficient samples are available, KDE may efficiently approximate challenging distributions of interest (Epanechnikov 1969), even distributions with multiple modes, i.e. with distinct and well separated regions of high probability density values. For addressing multivariate KDE, the product rule may be utilized since it provides enhanced flexibility in separately selecting the form of the kernel in each dimension (Epanechnikov 1969; Silverman 1986). If N_s samples $\{\boldsymbol{\theta}^s\}$ are available from the target distribution, with θ_i^s denoting the i^{th} component of the s^{th}

sample [so the samples available for θ_i are $\{\theta_i^s\}$], the fixed-bandwidth product KDE takes the form

$$k_d(\boldsymbol{\theta}) = \frac{1}{N_s}\sum_{s=1}^{N_s}\left(\prod_{i=1}^{n_\theta}\left\{\frac{1}{t_i}\left[K\left(\frac{\theta_i - \theta_i^s}{t_i}\right)\right]\right\}\right), \tag{2.19}$$

where $K(.)$ is the chosen univariate kernel and t_i is the bandwidth parameter defining the spread of the Kernel for the i^{th} dimension. The selection of bandwidth is a critical issue in providing satisfactory approximations; large bandwidth leads to over-smoothing of the derived densities, whereas small bandwidth leads to spurious noise in the approximation.

Several univariate kernels have been proposed in the literature (Scott 1992); one of the simplest and most popular is the Gaussian kernel given by

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{-t/2}. \tag{2.20}$$

For this kernel, an optimal selection of the bandwidth parameter that minimizes the asymptotic mean integrated error (AMISE) between the approximation in Eq. (2.19) and the target distribution (assumed in this case to be a Gaussian distribution in order to obtain a close form solution) is (Epanechnikov 1969)

$$t_i^{AMISE} = \left(\frac{4}{N_s(n_\theta + 2)}\right)^{\frac{1}{n_\theta + 4}} \sigma_i, \tag{2.21}$$

where σ_i is the standard deviation of the samples $\{\theta_i^s\}$. This ultimately leads to a simple closed-form solution for the bandwidth [as opposed to more computationally

demanding approaches (Mugdadi and Ahmad 2004)], which is the main reason for its popularity. Figure 2.2 illustrates the concept of sample-based KDE.

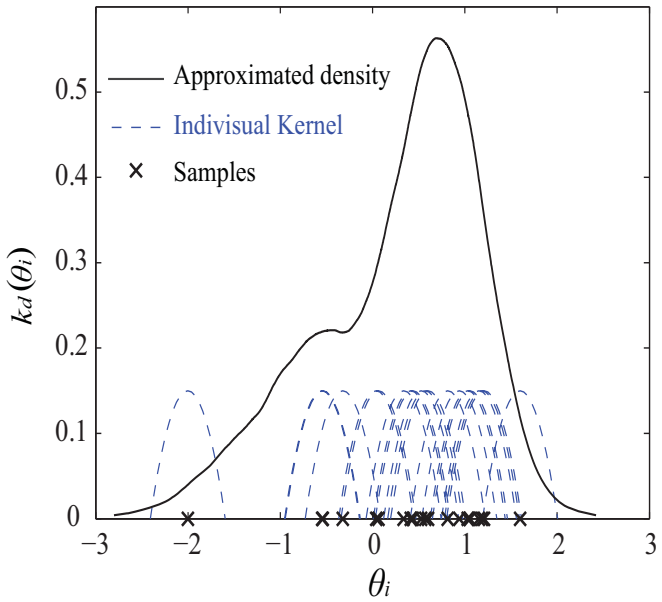


Figure 2.2 Example of Kernel Density Estimation based on samples

In some instances, generation of samples from $k_d(\boldsymbol{\theta})$ are also needed in this dissertation. This task is straightforward since the distribution given by Eq. (2.19) corresponds ultimately to an N_s-fold mixture of the parent kernel distribution $K(.)$; it simply requires sampling an index s from a discrete uniform distribution in $[1\ N_s]$ and then generating samples from distribution $K\left((\theta_i - \theta_i^s)/t_i\right)$ independently for each component of $\boldsymbol{\theta}$.

2.4.4 Probabilistic Global Sensitivity Analysis

Global sensitivity analysis techniques aim to quantify the importance of each of the different model parameters θ_i to the overall probabilistic performance. Various such approaches exist, such as Sobol indices or Analysis of Variance (ANOVA) (Sobol 2001; Archer et al. 1997). In this investigation, a recently proposed global sensitivity analysis (Taflanidis and Jia 2011; Jia and Taflanidis 2011) that utilizes the definition in Eq. (2.2) is adopted for this purpose since it can be implemented using information readily available from the evaluations of $h(\mathbf{x},\boldsymbol{\theta})$ within a stochastic simulation setting.

The sensitivity analysis is established by comparing the marginal distributions $p(\theta_i)$ and $\pi(\theta_i \mid \mathbf{x})$. The importance for each model parameter is then quantified by using the relative information entropy, which is a measure of the difference between $\pi(\theta_i,\mathbf{x})$ and $p(\theta_i)$, and is expressed through

$$D\left(\pi(\theta_i \mid \mathbf{x}) \,\|\, p(\theta_i)\right) = \int_{\mathbb{R}} \pi(\theta_i \mid \mathbf{x}) \log\left[\frac{\pi(\theta_i \mid \mathbf{x})}{p(\theta_i)}\right] d\theta_i. \tag{2.22}$$

For efficient calculation of the relative information entropy, approximation of $\pi(\theta_i \mid \mathbf{x})$ through its samples $\{\theta_i^s\}$ and KDE has been proposed (Jia and Taflanidis 2011). For this purpose, a sample set $\{\boldsymbol{\theta}^s\}$ from $\pi(\boldsymbol{\theta} \mid \mathbf{x})$ is first obtained. Projection of this set to each model parameter leads then to samples $\{\theta_i^s\}$ from $\pi(\theta_i \mid \mathbf{x})$ that can be used to obtain a KDE-based approximation as in Eq. (2.19). Since we are interested in scalar variables, the approximation in this case simplifies to

$$\tilde{\pi}(\theta_i \mid \mathbf{x}) = \frac{1}{N_s} \sum_{s=1}^{N_s} \left(\frac{1}{t_i} \left[K\left(\frac{\theta_i - \theta_i^s}{t_i} \right) \right] \right), \tag{2.23}$$

where $\tilde{\pi}(\theta_i \mid \mathbf{x})$ denotes the KDE-based approximation to $\pi(\theta_i \mid \mathbf{x})$ and t_i is the optimal AMISE bandwidth given for the Gaussian kernel by Eq. (2.21) with $n_\theta = 1$ (since the density approximation is implemented for scalar quantities here), which simplifies to $t_i = 1.06 \cdot N_s^{-1/5} \sigma_i$, with σ_i corresponding to the standard deviation of the samples $\{\theta_i^s\}$.

The approximation for the relative information entropy in Eq. (2.22) is finally (Beirlant et al. 1997)

$$D\left(\pi(\theta_i \mid \mathbf{x}) \,\|\, p(\theta_i)\right) \approx \int_{b_{li}}^{b_{ui}} \tilde{\pi}(\theta_i \mid \mathbf{x}) \log\left(\frac{\tilde{\pi}(\theta_i \mid \mathbf{x})}{p(\theta_i)} \right) d\theta_i, \tag{2.24}$$

where the last scalar integral can be numerically evaluated and $[b_{li}, b_{ui}]$ is the region for which samples for $\pi(\theta_i \mid \mathbf{x})$ are available (Jia and Taflanidis 2011). Parameters with higher values for $D(\pi(\theta_i \mid \mathbf{x}_k) \,\|\, p(\theta_i))$ have higher importance to the probabilistic performance (Jia and Taflanidis 2011).

CHAPTER 3:

OPTIMIZATION UNDER UNCERTAINTY WITH ADAPTIVE IMPORTANT SAMPLING

This chapter offers an adaptive implementation of importance sampling (IS) across the iterations of the numerical optimization described through Eq. (2.8). Recall the two main challenges in IS: (a) selection of efficient IS densities, including their characteristics (balancing between improving accuracy and avoiding numerical problems) and (b) selection of the specific variables for IS to target in high-dimensional problems. The advances offered here address both these challenges.

Although the proposed IS approach can be combined with any optimization algorithm relying on local search (such as gradient-based techniques), the Simultaneous Perturbation Stochastic Approximation (SPSA) (Kleinman et al. 1999; Spall 1998a) with Common Random Numbers (CRN) is chosen for this purpose. Similar to the work of Ang et al. (Ang 1992), the formulation of IS densities is based on samples that are distributed according to the integrand of the objective function, with Kernel density estimation (KDE) adopted for density approximation and its characteristics optimally selected by maximizing the anticipated accuracy for the performance objective estimate if these densities were to be implemented as IS densities. The main novel contribution of this work is the development of new approaches for facilitating this adaptive formulation of

IS densities without adding to the computational burden using only system evaluations that have already been performed (when evaluating the objective function in previous iterations of the optimization algorithm). A new, robust selection of the KDE characteristics is additionally proposed considering higher order statistics for the expected accuracy, and to avoid numerical problems when trying to develop an IS density for all model parameters; a relative prioritization is introduced. This is established by integrating a recently developed global sensitivity analysis (Taflanidis and Jia 2011; Vetter 2012) to quantify the relative importance of individual model parameters to the overall probabilistic performance. Development of IS densities is then considered only for the most important ones. Ultimately, the global sensitivity comparison provides a relative prioritization for each uncertain model parameter, whereas the optimization for the expected accuracy provides an optimal selection of the characteristics of the IS densities and of the exact number of model parameters considered in the IS formulation. Within the optimization framework, this methodology is implemented across the different iterations (searching for the system optima), and guidelines for the efficient sharing of information are established, imposing ultimately a minimum additional computational burden (no new system response evaluations) for performing the global sensitivity analysis and selecting the IS density characteristics.

The nomenclature specific to this chapter is reviewed in Table 3.1.

TABLE 3.1

RELEVANT NOMENCLATURE FOR CHAPTER 3

$\mathbf{\Delta}^{SP}$	Perturbation direction for gradient approximation in SPSA	$\tilde{f},\ \tilde{r}$	Functions f, r established for $\tilde{\mathbf{w}}, \tilde{\boldsymbol{\xi}}$
δ_{thresh}	Target value of coefficient of variation δ	$k(.)$	Pilot kernel estimate
$\tilde{\delta}$	Coefficient of variation for new proposal density	$k_a(.)$	Kernel density estimate using adaptive bandwidth
$\boldsymbol{\theta}^s$	Total number of available samples for Kernel formulation	$\cdot_k$	(Subscript) k^{th} iteration of the optimization algorithm
$\boldsymbol{\theta}^a$	Samples from $\pi(\boldsymbol{\theta}\mid\mathbf{x})$	N_s	Total number of kernel samples
λ^s	local bandwidth factor for s^{th} sample	N_{req}	Estimated value of N to establish δ_{thresh}
$\boldsymbol{\xi}$	Remaining components of $\boldsymbol{\theta}$ excluding $\mathbf{w}$	N_a	Number of kernel samples obtained in current iteration
$\tilde{\boldsymbol{\xi}}, \tilde{\mathbf{w}}$	Choices of $\boldsymbol{\xi}$ and $\mathbf{w}$ for next iteration of algorithm	N_{min}	Minimum number of kernel samples obtained in current iteration
$\pi(\boldsymbol{\theta}\mid\mathbf{x})$	Density proportional to integrand	n_w	Dimension of $\mathbf{w}$
$\tilde{\pi}(\boldsymbol{\theta}\mid\mathbf{x})$	KDE approximation to density proportional to integrand	n_ξ	Dimension of $\boldsymbol{\xi}$
τ	KDE sensitivity factor	$q(\mathbf{w})$	Importance sampling density for subset $\mathbf{w}$
ω^s	Adaptive-window bandwidth parameter for s^{th} Kernel	s_e	Allowable entropy percentage reduction for definition of D_{min}
c^{SP}	Step size for gradient approximation for SPSA	s_o	Outlier detection scaling
D_{min}	Cut-off value for entropy	t_i	bandwidth parameter for θ_i
$\mathbf{d}$	Vector of kernel characteristics	t_i^{AMISE}	AMISE optimal bandwidth
$f(\boldsymbol{\theta})$	Importance sampling probability density function	$\mathbf{w}$	Current subset of $\boldsymbol{\theta}$ for which IS is considered

3.1 Optimization Formulation

The iterative numerical optimization algorithm for solving Eq. (2.8) is described within this chapter as

$$\mathbf{x}_{k+1} = G_{opt}(\mathbf{x}_k \mid \{\boldsymbol{\theta}^j\}_k), \tag{3.1}$$

indicating that solution at iteration $k+1$ is obtained based on the solution at iteration k through the function G_{opt} representing the recursive relations of the optimization algorithm. Sample set $\{\boldsymbol{\theta}^j\}_k$ is used within the iteration, and this sample set ultimately changes from iteration to iteration (interior sampling). Herein the subscript k denotes the iteration of the numerical optimization algorithm. The goal is to share information across these iterations about the behavior of the probabilistic integrand to improve the stochastic-simulation-based evaluation of the objective function. This ultimately means that the designs at iterations $k+1$ and k cannot be significantly different as then this information would be erroneous. As such, algorithms relying on local searches need to be adopted for the iterative optimization described by Eq. (3.1). The specific algorithm considered here, is the Simultaneous Perturbation Stochastic Approximation (SPSA) (Kleinman et al. 1999; Spall 1998a) with Common Random Numbers (CRN), an algorithm that is appropriate for problems in which the system simulation for estimating $h(\mathbf{x},\boldsymbol{\theta})$ is a "black box" for the designer and as such, numerical differentiation is the only feasible approach for obtaining derivative information. There is a wealth of literature regarding SPSA, and the details of the method, i.e. the definition of G_{opt}, are presented in Appendix A (Spall 2003; Spall 1998a).

3.2 Adaptive Importance Sampling Foundation

In the proposed framework, IS densities are formulated by sharing information across the iterations of the optimization algorithm described through Eq. (3.1) (an illustration of the sharing process can be observed in Figure 3.1). Both the characteristics of the IS densities, and the subset of $\boldsymbol{\theta}$ for which they are formulated are adaptively chosen in this process. At iteration k, let $\mathbf{w}_k$ denote the subset of $\boldsymbol{\theta}$ for which IS is formulated and $q_k(\mathbf{w}_k)$ correspond to the selection of their joint density. If $\boldsymbol{\xi}_k$ corresponds to the remaining components of $\boldsymbol{\theta}$ (such that $\boldsymbol{\theta} = [\mathbf{w}_k \;\; \boldsymbol{\xi}_k]$), the proposal density in Eq. (2.13) for the k^{th} iteration is transformed to $f_k(\boldsymbol{\theta}) = q_k(\mathbf{w}_k p(\boldsymbol{\xi}_k \mid \mathbf{w}_k))$ leading to IS quotient

$$r_k(\boldsymbol{\theta}) = \frac{p(\boldsymbol{\theta})}{f_k(\boldsymbol{\theta})} = \frac{p(\mathbf{w}_k)p(\boldsymbol{\xi}_k \mid \mathbf{w}_k)}{q_k(\mathbf{w}_k)p(\boldsymbol{\xi}_k \mid \mathbf{w}_k)} = \frac{p(\mathbf{w}_k)}{q_k(\mathbf{w}_k)} = r_k(\mathbf{w}_k). \tag{3.2}$$

Note that this quotient is only a function of $\mathbf{w}_k$ (and not of entire vector $\boldsymbol{\theta}$); the simplified notation $r_k(\mathbf{w}_k)$ is also used henceforth. The estimate for objective function becomes

$$\hat{H}(\mathbf{x}_k \mid \{\boldsymbol{\theta}^j\}_k) = \frac{1}{N_k}\sum_{j=1}^{N_k} h(\mathbf{x}_k, \boldsymbol{\theta}_k^j) r_k(\mathbf{w}_k^j), \tag{3.3}$$

where $\mathbf{w}_k^j$ are samples from $q_k(\mathbf{w}_k)$, $\boldsymbol{\xi}_k^j$ are samples from $p(\boldsymbol{\xi}_k \mid \mathbf{w}_k)$, and N_k corresponds to the number of samples used. Recall that $\{\boldsymbol{\theta}^j\}_k = \{\mathbf{w}_k^j \;\; \boldsymbol{\xi}_k^j\}_k$ is the sample set employed in the k^{th} iteration. Note that the underlying assumption for Eq. (3.2) is that the marginal distributions can be readily obtained based on the joint

distribution. It is important to point out that this is not always straightforward and a transformation of probability spaces, for example transformation to standard Gaussian space (Au and Beck 2003; Melchers 1999), might be required to accomplish it.

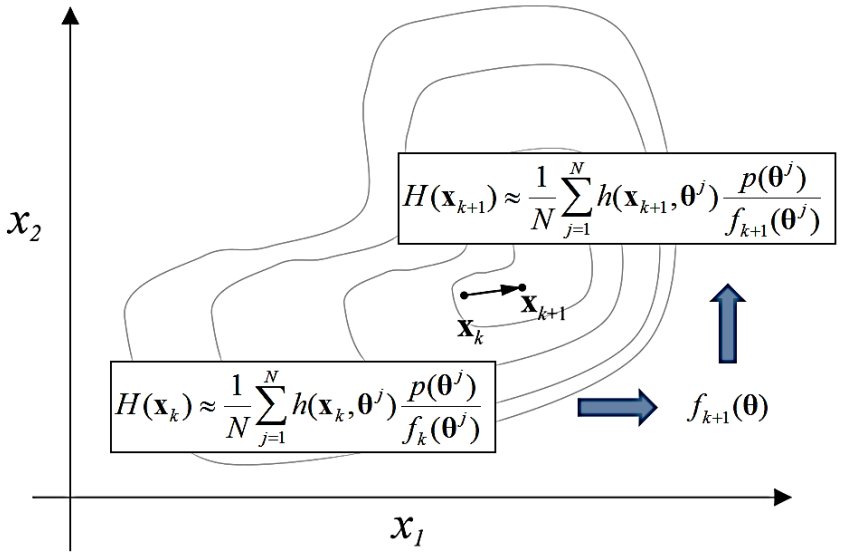


Figure 3.1 Illustration of sharing of information between iterations k and $k+1$ for an arbitrary problem to obtain efficient proposal densities

During this evaluation of the objective function, SPSA (the optimizer) requires evaluation of the perturbed designs $h(\mathbf{x}_k + c_k^{SP}\boldsymbol{\Delta}_k^{SP}, \boldsymbol{\theta})$ and $h(\mathbf{x}_k - c_k^{SP}\boldsymbol{\Delta}_k^{SP}, \boldsymbol{\theta})$, rather than evaluations of $h(\mathbf{x}_k, \boldsymbol{\theta})$, to construct the gradient approximation. Therefore, the IS formulation could be established based on either of the perturbed designs $\mathbf{x}_k \pm c_k^{SP}\boldsymbol{\Delta}_k^{SP}$; however, the choice here is to use both and linearly interpolate for $h(\mathbf{x}_k, \boldsymbol{\theta})$ to obtain the performance given a set $\{\boldsymbol{\theta}^j\}_k$. This choice does not require additional function evaluations of the system's model.

$$h(\mathbf{x}_k,\boldsymbol{\theta}) \approx \frac{1}{2}\left(h(\mathbf{x}_k + c_k^{SP}\boldsymbol{\Delta}_k^{SP},\boldsymbol{\theta}) + h(\mathbf{x}_k - c_k^{SP}\boldsymbol{\Delta}_k^{SP},\boldsymbol{\theta})\right). \tag{3.4}$$

The liner interpolation of $h(\mathbf{x}_k,\boldsymbol{\theta})$ approaches the true value of the performance function as c_k^{SP} approaches zero ($c_k^{SP} \to 0$). The approximation of the objective function at $\mathbf{x}_k$ is then established using the interpolated values of Eq. (3.4) in Eq. (3.3). The estimate for the coefficient of variation, which will be used to evaluate the accuracy, is then provided by Eq. (2.15) considering the specific IS formulation in Eq. (3.2) and ultimately corresponds to

$$\delta_k \approx \frac{1}{\sqrt{N_k}}\sqrt{\frac{\frac{1}{N_k}\sum_{j=1}^{N_k}\left(h(\mathbf{x}_k,\boldsymbol{\theta}_k^j)r_k(\mathbf{w}_k^j)\right)^2}{\hat{H}(\mathbf{x}_k)^2} - 1}. \tag{3.5}$$

Note that the approximation for $h(\mathbf{x}_k,\boldsymbol{\theta})$ in Eq. (3.4) is only used to make decisions for the IS density formulation; it is not used in the optimization algorithm which requires $h(\mathbf{x}_k + c_k^{SP}\boldsymbol{\Delta}_k^{SP},\boldsymbol{\theta})$ and $h(\mathbf{x}_k - c_k^{SP}\boldsymbol{\Delta}_k^{SP},\boldsymbol{\theta})$ to be separately used for the gradient approximation in Eq. (A.2).

3.2.1 Proposal Density and Approximation through Samples

Based on Eq. (2.16), the target IS PDF at the k^{th} iteration is

$$\pi(\boldsymbol{\theta} \mid \mathbf{x}_k) \propto |h(\mathbf{x}_k,\boldsymbol{\theta})|\, p(\boldsymbol{\theta}), \tag{3.6}$$

where again $\propto$ denotes proportionality. Approximation to this PDF through samples from it using KDE is considered. To obtain the necessary samples from $\pi(\boldsymbol{\theta} \mid \mathbf{x}_k)$, the readily available evaluations of $h(\mathbf{x}_k,\boldsymbol{\theta})$ for set $\{\boldsymbol{\theta}^j\}_k$ are used within the rejection

sampling approach discussed in Section 2.4.2. This is established by using $q_k(\mathbf{w}_k)p(\boldsymbol{\xi}_k \mid \mathbf{w}_k) = f_k(\boldsymbol{\theta})$ as the proposal density $f^p(\boldsymbol{\theta})$ within the algorithm, leading to

$$\gamma_k = s_f \cdot \max_j \left[\mid h(\mathbf{x}_k, \boldsymbol{\theta}^j) \mid r_k(\mathbf{w}_k^j) \right], \tag{3.7}$$

and acceptance of samples based on comparison (samples are accepted when below inequality holds)

$$\frac{\mid h(\mathbf{x}_k, \boldsymbol{\theta}^j) \mid r_k(\mathbf{w}_k^j)}{u^j} > \gamma_k \,. \tag{3.8}$$

A relaxation is further proposed here to guarantee that enough samples are obtained to inform the formulation of IS densities. This relaxation is established by allowing $s_f < 1$ and ultimately corresponds to modification of $\pi(\boldsymbol{\theta} \mid \mathbf{x}_k)$ to

$$\pi(\boldsymbol{\theta} \mid \mathbf{x}_k) \propto \min\left[\gamma_k, \mid h(\mathbf{x}_k, \boldsymbol{\theta}) \mid r_k(\mathbf{w}_k)\right] q_k\left(\mathbf{w}_k\right) p(\boldsymbol{\xi}_k \mid \mathbf{w}_k), \tag{3.9}$$

where $\min[i, j]$ corresponds to the minimum of the two arguments i and j. If $\gamma_k > \max(\mid h(\mathbf{x}_k, \boldsymbol{\theta}) \mid r_k(\mathbf{w}_k))$, Eq. (3.9) equals Eq. (3.6). Thus, the selection of γ_k (and consequently of s_f) is a compromise between obtaining sufficient samples from $\pi(\boldsymbol{\theta} \mid \mathbf{x}_k)$ and also having a proposal density Eq. (3.9) that is similar to the optimal one given by Eq. (3.6).

This rejection sampling approach leads to a set of N_a samples, denoted $\{\boldsymbol{\theta}^a\}_k$, from the density Eq. (3.9). Rather than relying on samples only from the current iteration, samples from multiple previous iterations can be utilized, with the ultimate goal to obtain a sufficient total number of samples N_s to better inform the IS

formulation. This augmented set will be denoted as $\{\boldsymbol{\theta}^s\}_k$, composed of the $\{\boldsymbol{\theta}^a\}_k$ as well as samples obtained in previous iterations, with the selection of the exact number of total samples discussed later. An approximation for $\pi(\boldsymbol{\theta}|\mathbf{x}_k)$ denoted $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{x}_k)$ may then be established using these samples. Thus, the questions that need to be address are:

(a) How can we *adaptively* construct the IS densities based on the obtained samples? This pertains to both the type of density as well as its characteristics.

(b) How can we *adaptively* select the new subset of $\boldsymbol{\theta}$, denoted $\tilde{\mathbf{w}}_k$, for which these densities should be established (note that this will not be necessarily the same as the subset chosen for the current iteration $\mathbf{w}_k$)?

Section 3.2.2 examines question (a), while section 3.2.3 examines question (b).

3.2.2 Selection of IS Densities and Its Characteristics

The sample-based density approximation approach considered in this work is kernel density estimation (KDE) (introduced in section 2.4.3), which belongs to the greater family of non-parametric density estimation methods. KDE may efficiently approximate challenging distributions of interest (Epanechnikov 1969), even distributions with multiple modes, something that is of importance since $\pi(\boldsymbol{\theta}|\mathbf{x}_k)$ can have multiple modes, even if it $p(\boldsymbol{\theta})$ does not. This is the reason KDE is preferred here over alternative parametric density approximation approaches (Silverman 1986) that have been considered for IS formulation (Au and Beck 1999; Melchers 1989).

Because the approximation at the tails of the IS distribution is of importance (recall the discussion in section 2.4.1), following the recommendations in (Ang 1992; Cappé et al. 2008), an adaptive kernel bandwidth formulation is considered (Abramson 1982), and for addressing multivariate KDE, the product rule is adopted herein since it provides enhanced flexibility in separately selecting the form of the kernel in each dimension (Silverman 1986; Scott and Sain 2005).

For the subset $\tilde{\mathbf{w}}_k \subset \boldsymbol{\theta}$ with dimension $n_{\tilde{w}}$ and based on the available samples $\{\tilde{\mathbf{w}}_k^s\}_k$ (corresponding to the respective subset of $\{\boldsymbol{\theta}^s\}_k$), the product kernel in Eq. (2.19) with adaptive bandwidth transforms to

$$q(\tilde{\mathbf{w}}_k) = k_a(\tilde{\mathbf{w}}_k \mid \mathbf{d}) = \frac{1}{N_s}\sum_{s=1}^{N_s}\left(\prod_{i=1}^{n_{\tilde{w}}}\left\{\frac{1}{\lambda^s t_i}\left[K\left(\frac{\tilde{w}_{ki}-\tilde{w}_{ki}^s}{\lambda^s t_i}\right)\right]\right\}\right), \tag{3.10}$$

where $K(.)$ is the chosen kernel, $\mathbf{d}$ is the vector of kernel-characteristics (defined later), λ^s is a local bandwidth factor, t_i is the fixed-window bandwidth parameter defining the spread of the kernel for the i^{th} variable $\tilde{w}_{ki}$ (respective component of $\tilde{\mathbf{w}}_k$), and $\{\tilde{w}_{ki}^s\}_k$ are the corresponding samples for it. For bounded parameters, boundary correction KDE approaches should be used to circumvent approximation problems close to the boundaries (Jia and Taflanidis 2011; Karunamuni and Zhang 2008). The chosen product kernel facilitates an easy implementation of such corrections since it allows them to be established independently for each parameter (Jia and Taflanidis 2011).

For λ^s equal to 1 (or to any constant), the approximation in Eq. (3.10) corresponds to the fixed-bandwidth kernel [similar to the one discussed in Eq. (2.19) but

considering only the $\tilde{\mathbf{w}}_k$ subset of $\boldsymbol{\theta}$]. The introduction of λ^s establishes the so-called adaptive kernel bandwidth (Abramson 1982), aimed at improving the approximation at the tails of the distribution by adaptively changing the bandwidth, which is selected larger for the regions with lower values of the probability density. This factor is given by

$$\lambda^s = \left\{ \left[\prod_{k=1}^{N_s} k(\tilde{\mathbf{w}}_k^s) \right]^{1/N_s} / k(\tilde{\mathbf{w}}_k^s) \right\}^{\tau} = \left(\omega^s\right)^{\tau}, \tag{3.11}$$

where $0 < \tau < 1$ is a sensitivity factor, and ω^s is defined as

$$\omega^s = \left[\prod_{k=1}^{N_s} k(\tilde{\mathbf{w}}_k^s) \right]^{1/N_s} / k(\tilde{\mathbf{w}}_k^s), \tag{3.12}$$

and is calculated using the fix-bandwidth kernel approximation (referenced as the pilot estimate), established similarly to Eq. (3.10) but with $\lambda^s = 1$,

$$k(\tilde{\mathbf{w}}_k) = \frac{1}{N_s} \sum_{s=1}^{N_s} \left(\prod_{i=1}^{n_{\tilde{w}}} \left\{ \frac{1}{t_i} \left[K\left(\frac{\tilde{w}_{ki} - \tilde{w}_{ki}^s}{t_i} \right) \right] \right\} \right). \tag{3.13}$$

It is generally acknowledged that the method is insensitive to the fine details of the pilot estimate (Abramson 1982). This means that the preliminary selection for t_i in $k(\tilde{\mathbf{w}}_k)$ has small impact on the final adaptive kernel (optimization of this bandwidth will be discussed later). Among the different kernel choices, the popular Gaussian kernel is employed in this investigation Eq. (2.20) .

The optimal selection of the bandwidth parameter (for the fix-bandwidth kernel) based on the AMISE, following the discussion in Section 2.4.1, is

$$t_i^{AMISE} = \left(\frac{4}{N_s(N_{\tilde{w}}+2)} \right)^{\frac{1}{N_{\tilde{w}}+4}} \sigma_i, \tag{3.14}$$

where σ_i is the standard deviation of the samples $\{\tilde{w}_i^s\}_k$. This selection may be used to formulate the pilot estimate to calculate the local bandwidth factor Eq.(3.11), but may not be necessarily appropriate for $k_a(\tilde{\mathbf{w}}_k \mid \mathbf{d})$ when utilized as IS density (see discussion later). It also shows that σ_i corresponds to a proper normalization for t_i.

3.2.2.1 Optimal Proposal Densities

The vector $\mathbf{d}$ for the kernel characteristics is composed of the sensitivity factor τ and the bandwidth parameter t_i for each component (or a version normalized by σ_i), leading to the definition: $\mathbf{d} = [\tau \;\; t_i / \sigma_i \;;\; i = 1, \cdots, n_{\tilde{w}}]$. Since the goal here is to use the established distribution as the IS density, this vector $\mathbf{d}$ should be selected to optimize the expected coefficient of variation (Ang 1992), or equivalently $E_{\tilde{f}_k}[(h(\mathbf{x}_k, \boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k \mid \mathbf{d}))^2]$, referenced henceforth as objective target for the IS density optimization, where $\tilde{f}_k$ and $\tilde{r}_k$ correspond to the new proposal density and IS quotient respectively. Both of them are defined based on $k_a(\tilde{\mathbf{w}}_k \mid \mathbf{d})$. Note that in the study by Ang et al. (Ang 1992), its application within optimization algorithms was not considered. Adoption of the approach advocated in Ang's study requires knowledge of $h(\mathbf{x}_k, \boldsymbol{\theta})$ for $\{\boldsymbol{\theta}^s\}_k$, something that does not hold for the implementation proposed in this investigation since the kernels are based on samples obtained over multiple iterations

(it is only true for the subset $\{\boldsymbol{\theta}^a\}_k$). Thus, a different approximation is proposed hereafter relying on the following transformation:

$$
\begin{aligned}
E_{\tilde{f}_k}\left[\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k\,|\,\mathbf{d})\right)^2\right] &= \int_{\Theta}\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k\,|\,\mathbf{d})\right)^2\tilde{f}_k(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&= \int_{\Theta} h(\mathbf{x}_k,\boldsymbol{\theta})^2\frac{p(\tilde{\mathbf{w}}_k)}{k_a(\tilde{\mathbf{w}}_k\,|\,\mathbf{d})}\left(\frac{p(\tilde{\mathbf{w}}_k)}{k_a(\tilde{\mathbf{w}}_k\,|\,\mathbf{d})}\tilde{f}_k(\boldsymbol{\theta})\right)d\boldsymbol{\theta} \\
&= \int_{\Theta} h(\mathbf{x}_k,\boldsymbol{\theta})^2\frac{p(\tilde{\mathbf{w}}_k)}{k_a(\tilde{\mathbf{w}}_k\,|\,\mathbf{d})}p(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&= \int_{\Theta} h(\mathbf{x}_k,\boldsymbol{\theta})^2\frac{p(\tilde{\mathbf{w}}_k)}{k_a(\tilde{\mathbf{w}}_k\,|\,\mathbf{d})}p(\boldsymbol{\theta})\frac{f_k(\boldsymbol{\theta})}{f_k(\boldsymbol{\theta})}d\boldsymbol{\theta} \\
&= \int_{\Theta} h(\mathbf{x}_k,\boldsymbol{\theta})^2\frac{p(\tilde{\mathbf{w}}_k)}{k_a(\tilde{\mathbf{w}}_k\,|\,\mathbf{d})}r_k(\mathbf{w}_k)f_k(\boldsymbol{\theta})d\boldsymbol{\theta}.
\end{aligned}
\tag{3.15}
$$

The last expression may then be approximated through stochastic simulation using the readily available performance evaluations for sample set $\{\boldsymbol{\theta}^j\}_k$ [which is distributed according to $f_k(\boldsymbol{\theta})$] as

$$
\hat{E}_{\tilde{f}_k}\left[\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k)\right)^2\right] = \frac{1}{N_k}\sum_{j=1}^{N_k} h(\mathbf{x}_k,\boldsymbol{\theta}_k^j)^2 r_k(\mathbf{w}_k^j)\frac{p(\tilde{\mathbf{w}}_k^j)}{k_a(\tilde{\mathbf{w}}_k^j\,|\,\mathbf{d})}. \tag{3.16}
$$

This then leads to an optimization problem for selection of vector $\mathbf{d}$,

$$
\mathbf{d}^* = \underset{\mathbf{d}\in\mathbb{R}^{(N_{\tilde{\mathbf{w}}}+1)+}}{\arg\min}\,\hat{E}_{\tilde{f}_k}\left[\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k\,|\,\mathbf{d})\right)^2\right], \tag{3.17}
$$

which may be solved by any standard algorithm. Note that the evaluation of the target function in this problem simply requires estimation of $\{k_a(\tilde{\mathbf{w}}_k^j\,|\,\mathbf{d})\}$ for the new values of vector $\mathbf{d}$, something that involves minimal computational burden. As such, obtaining a local optimum for Eq. (3.17) is a straightforward numerical optimization task. Since Eq. (3.16) involves $k_a(\tilde{\mathbf{w}}_k^j\,|\,\mathbf{d})$ in the denominator and some elements of sets $\{\tilde{\mathbf{w}}_k^j\}_k$ and

$\{\tilde{\mathbf{w}}_k^s\}_k$ are the same (i.e., points where kernel density is estimated are identical to points used to establish these densities), ill-posed problems may arise in the optimization because at these points $k_a(\tilde{\mathbf{w}}_k^j) \to \infty$ when $t_i \to 0$, leading to zero contribution in the approximation for $E_{\tilde{f}_k}[(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k \mid \mathbf{d}))^2]$. This should be attributed to the fact that the samples used to estimate $E_{\tilde{f}_k}[(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k \mid \mathbf{d}))^2]$ through stochastic simulation have correlation to the samples utilized to form the kernel density. The remedy here, as in (Ang 1992), is to substitute $k_a(\tilde{\mathbf{w}}_k^j \mid \mathbf{d})$ with $k_{-a}(\tilde{\mathbf{w}}_k^j \mid \mathbf{d})$, which corresponds to the kernel estimate using all kernels $\{\tilde{\mathbf{w}}_k^s\}_k$ except for the samples that correspond to $\tilde{\mathbf{w}}_k^j$, if such samples exists (this automatically circumvents the correlation problem). It is then straightforward to prove that $k_{-a}(\tilde{\mathbf{w}}_k^j \mid \mathbf{d}) \to k_a(\tilde{\mathbf{w}}_k^j \mid \mathbf{d})$ for N_s large enough (Ang 1992); as such this remedy does not influence the accuracy of the approximation in Eq. (3.17).

Once selection of $\mathbf{d}^*$ has been established, the corresponding expected coefficient of variation may be obtained by

$$\tilde{\delta}_k \approx \frac{1}{\sqrt{N_k}} \sqrt{\frac{\hat{E}_{\tilde{f}_k}\left[\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k \mid \mathbf{d})\right)^2\right]}{\hat{H}(\mathbf{x}_k)^2} - 1}. \tag{3.18}$$

It is generally expected that this approximation will underestimate the value for the coefficient of variation. This should be attributed to the fact than the approximation in Eq. (3.16) (obtained through stochastic simulation) is used for $E_{\tilde{f}_k}[(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k \mid \mathbf{d}))^2]$ and that $\mathbf{d}$ has been explicitly optimized to minimize this approximation. This will contribute to exploiting any vulnerabilities of the approximation

to produce an estimated value lower than the true one. In other words, the optimization will create a bias, which will be bigger for smaller values of N_k (and will in general approach zero if N_k is sufficiently large).

3.2.2.2 Robust Proposal Densities Selection

The previously described bias in the estimation when optimizing Eq. (3.17) can lead to erroneous solutions, especially when the number of available samples is low (so decisions are made without sufficient information). To circumvent this challenge, a robust optimization for $\mathbf{d}^*$ is established, motivated by the generalized RDO formulation; a weighed combination of the mean and standard deviation of $\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k)\right)^2$ is considered. This leads then to the target function

$$\begin{aligned} &= (1-w)E_{\tilde{f}_k}\left[\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k)\right)^2\right] + w\sqrt{Var_{\tilde{f}_k}\left[\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k)\right)^2\right]} \\ &= E_{\tilde{f}_k}\left[\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k)\right)^2\right]\left((1-w)+w\sqrt{\frac{E_{\tilde{f}_k}\left[\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k)\right)^4\right]}{E_{\tilde{f}_k}\left[\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k)\right)^2\right]^2}-1}\right), \end{aligned} \tag{3.19}$$

where w is the relative weight parameter that controls the degree of robustness. In this expression, $E_{\tilde{f}_k}[(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k\,|\,\mathbf{d}))^2]$ may be approximated through Eq. (3.16) and an approximation for $E_{\tilde{f}_k}[(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k\,|\,\mathbf{d}))^4]$ may be similarly established as

$$\hat{E}_{\tilde{f}_k}\left[\left(h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k\,|\,\mathbf{d})\right)^4\right] = \frac{1}{N_k}\sum_{j=1}^{N_k} h(\mathbf{x}_k,\boldsymbol{\theta}_k^j)^4 r_k(\mathbf{w}_k^j)\frac{p^3(\tilde{\mathbf{w}}_k^j)}{k_a^3(\tilde{\mathbf{w}}_k^j\,|\,\mathbf{d})}. \tag{3.20}$$

3.2.3 Global Sensitivity Analysis and Selection of Model Parameters for Adaptive IS Densities

The remaining question is the selection of the exact model parameters corresponding to the vector $\tilde{\mathbf{w}}_k$. The probabilistic sensitivity analysis described in Section 2.4.4 is employed for this purpose. Recall that this analysis can identify the importance of each of the uncertain model parameters θ_i to the overall probabilistic performance and can be efficiently performed in the context of the proposed adaptive IS implementation using readily available information. It is established by comparing the marginal distributions of the prior distribution $p(\theta_i)$ and the auxiliary density $\pi(\theta_i \mid \mathbf{x}_k)$, where the latter is approximated through a fix-bandwidth KDE utilizing samples $\{\boldsymbol{\theta}^s\}_k$, leading to

$$\tilde{\pi}(\theta_i \mid \mathbf{x}_k) = \frac{1}{N_s}\sum_{s=1}^{N_s}\left(\frac{1}{t_i}\left[K\left(\frac{\theta_i - \theta_i^s}{t_i}\right)\right]\right), \tag{3.21}$$

with $t_i = 1.06 \cdot N_s^{-1/5}\sigma_i$ and σ_i corresponding to the standard deviation of the samples $\{\theta_i^s\}_k$. Note that this time, the goal is to approximate the values for $\pi(\theta_i \mid \mathbf{x}_k)$; therefore, the choice of the bandwidth that minimizes the integrated square error is appropriate. In addition, implementation of an adaptive kernel for this case would necessarily increase the computational complexity (since a new pilot-estimate would be required for each individual model parameter), whereas use of fixed-bandwidth kernels has been proven accurate for this type of application (Jia and Taflanidis 2011). As such, the fixed-bandwidth kernel is preferred for the entropy evaluation. The relative

information entropy $D(\pi(\theta_i \mid \mathbf{x}_k) \| p(\theta_i))$ can be approximated through the numerical integration of Eq. (2.24). Parameters with higher values for this entropy will have higher importance to the probabilistic performance (Jia and Taflanidis 2011), and formulation of IS densities should focus on them. Although this approach provides a relative prioritization of the different model parameters, a remaining question is how many model parameters should be included in the set $\tilde{\mathbf{w}}_k$? To answer this question, the anticipated coefficient of variation Eq. (3.18) is used through the following approach:

1. The model parameters are first ordered based on their respective values for $D(.)$ [ranked in decreasing importance from high to low values for $D(.)$]. Based on some threshold D_{min}, the maximum number of model parameters n_{max} for which IS will be considered is then defined. Parameters having entropy $D(.)$ lower than D_{min} are disqualified as potential candidates for the IS formulation.
2. For different values of m up to n_{max}, i.e. $m = 1, \cdots, n_{max}$, define $\tilde{\mathbf{w}}_k$ to consist of the m most influential components of $\boldsymbol{\theta}$; then perform optimization in Eq. (3.17) and calculate the optimal value of the objective target, meaning the second moment Eq. (3.16) or if the robust approach is adopted, the expression in Eq. (3.19).
3. Select as m the number of model parameters that yields the smallest value for the objective target.

This procedure allows for an adaptive selection of the number of parameters considered for the IS density formulation to yield the highest expected accuracy. The motivation for introducing D_{min} is to reduce the computational burden [requirement to

perform optimization Eq. (3.17)] since for parameters of low importance, we do not anticipate benefits from establishing IS densities. D_{min} may be adaptively selected based on the highest entropy value among all parameters and some pre-defined allowable percentage reduction $s_e < 1$,

$$D_{min} = s_e \max_i \left[D\left(\pi(\theta_i \mid \mathbf{x}_k) \,\|\, p(\theta_i)\right)\right]. \tag{3.22}$$

Through this definition, the selection is always normalized with respect to the importance of the most influential parameter.

3.3 Optimization under Uncertainty with Adaptive IS

3.3.1 Consideration for Implementation across Iterations

In the proposed scheme, the adaptive IS density formulation is integrated across the iterations of the optimization algorithm. At iteration k, corresponding to system configuration $\mathbf{x}_k$, the evaluations of the system performance function $h(\mathbf{x}_k, \boldsymbol{\theta})$ are utilized to formulate the IS densities that will be adopted for iteration $k+1$, corresponding to system configuration $\mathbf{x}_{k+1}$, with the sample set $\{\boldsymbol{\theta}^s\}_k$ used to develop the densities including samples that have been obtained in the last few iterations (not just in iteration k). If the systems configuration corresponding to $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ are not drastically different, where different should be interpreted as leading to different performance functions for the same model parameter values, these IS densities for $\mathbf{x}_k$ are expected to be efficient for $\mathbf{x}_{k+1}$ as well. This is the motivation for

adopting an algorithm that relies on local searches for solving the optimization problems, since the design configurations in subsequent iterations of these algorithms are similar, especially as the algorithm approaches local minima (when the steps in the design space are in general small). At initial iterations of the algorithm, the efficiency of the IS implementation by sharing information across iterations might be smaller since the step sizes in the design space can be large, leading to a large difference between $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$. But this is of small concerns since (i) the values of the objective function are also larger, so the coefficient of variation is expected to be low, reducing the need for adopting IS (smaller problems in accuracy), and (ii) obtaining high accuracy estimates is more important closer to a local optimum, where convergence/stopping criteria are utilized. Note that an alternative implementation can consider updating the IS density only every few iterations (rather than at every single iteration) when the algorithm has nearly converged (small steps in design space) in order to reduce the overall burden for identifying the optimal IS density, namely the optimization described by Eq. (3.17). Since the proposed adaptive IS formulation has small cost, this should not be considered a burden for most applications.

This discussion provides also guidelines for selection of the number of samples in the set used to inform the KDE-based proposal density formulation $\{\boldsymbol{\theta}^s\}_k$. During the initial iterations (when steps within the design space are large) samples from only the immediate iteration (or from not more than a few iterations back) should be utilized. When a local minimum is approached, samples from multiple iterations can be retained

since use of larger N_s better informs the IS densities. All these samples are expected to provide valuable information for the current iteration, as they correspond to design configurations similar to the one in the current iteration.

A final question that needs to be addressed is how the improvement in accuracy established through the IS density implementation can be utilized within the optimization algorithm. There are two approaches. The first one is to rely simply on the fact that the improved accuracy will contribute to greater robustness characteristics and faster convergence for the optimization algorithm itself, thus utilizing the benefits in an indirect way. The second one, advocated here, is to directly choose the number of samples needed N_{req} in order to establish a specific coefficient of variation δ_{thresh} based on the estimate for the coefficient of variation from the current iteration, given by Eq. (3.5). If N_k samples were used in the current iteration and a coefficient of variation δ_k was established, then assuming similar efficiency between subsequent IS densities, the required number of samples to attain the required δ_{thresh} corresponds to

$$N_{req} \geq N_k \left(\frac{\delta_k}{\delta_{thresh}} \right)^2 . \tag{3.23}$$

Note that an alternative approach would have been to use the anticipated coefficient of variation $\tilde{\delta}_k$ in Eq. (3.18) based on the new proposed IS density and not the one used at the beginning of the iteration, which may be significantly different than the new one obtained through use of this new IS density. As discussed earlier, $\tilde{\delta}_k$ will always underestimate the true coefficient of variation; as such, a selection based on δ_k is more

appropriate. As the algorithm converges to the optimal solution and design changes established between different iterations are small, it is expected that δ_k will provide a good estimate for the coefficient of variation even under the new proposed IS density (i.e. $\tilde{\delta}_k = \delta_k$). Thus, this simplification is expected to have a small impact on the resultant optimal selection for the number of samples. An erroneous estimate may be obtained, though, by this selection when the IS densities in the current iteration do not properly cover the entire Θ space that is of importance to the integrand. This could lead to large values for the expression $h(\mathbf{x}_k, \boldsymbol{\theta}_k^j) r_k(\mathbf{w}_k^j)$ for some samples (belonging in the aforementioned region) and ultimately to large values for δ_k. These samples will be included, though, as kernels in subsequent iterations [since they correspond to the highest values for $h(\mathbf{x}_k, \boldsymbol{\theta}_k^j) r_k(\mathbf{w}_k^j)$], and these problematic regions will be ultimately well represented by future IS densities. This means that due to the influence stemming from these samples (if and when they exist), the current estimate for δ_k does not represent well the estimated coefficient of variation for the following iterations, creating an artificially high requirement for the number of samples to attain the desired accuracy δ_{thresh}. Thus, a detection of such outlier samples is introduced before estimating δ_k though Eq. (3.23). This is established by neglecting all samples for which

$$\left| h(\mathbf{x}_k, \boldsymbol{\theta}_k^j) r_k(\mathbf{w}_k^j) - \hat{H}(\mathbf{x}_k \mid \{\boldsymbol{\theta}^j\}_k) \right| > s_o \sigma_k, \tag{3.24}$$

where σ_k is the standard deviation of $h(\mathbf{x}_k, \boldsymbol{\theta}_k^j) r_k(\mathbf{w}_k^j)$, estimated through all available samples, and s_o is an outlier detection scaling (generally larger than 6). Note that this

outlier detection impacts only the estimation for δ_k used to update N_k through Eq. (3.23) and does not impact any other aspects of the algorithm.

It should be finally noted that instead of using a specific threshold δ_{thresh}, an adaptive selection can be adopted using smaller values as the optimum design is approached (Polak 2008). Since the focus here is to investigate the advantages of the proposed adaptive formulation of IS densities, this latter choice is avoided because it adds another level of complexity in the comparisons.

3.3.2 Algorithm for Adaptive IS Implementation

When combining the previous ideas, one can formulate the following optimization algorithm utilizing adaptive IS. Define first the bounded design space X, as well as the characteristics of the SPSA algorithm (or for any other algorithm used), and select the initial point $\mathbf{x}_1$ and the number of initial samples N_1 for the stochastic simulation. Choose the minimum number of samples obtained per iteration N_{min} and factor s_f for the stochastic sampling, the number of total samples for the density approximation N_s (or if this changes across iterations, the adaptive rules for choosing it), the percentage reduction s_e for the definition of minimum importance based on the entropy evaluation, the desired accuracy δ_{thresh} for selection of N_k, the outlier scaling factor s_o and the objective target for the selection of optimal density characteristics, i.e. $E_{\tilde{f}_k}[h(\mathbf{x}_k,\boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k))^2]$ or its robust expression in Eq. (3.19). In the latter case, choose also

the robustness factor w_r. A general flow diagram of the adaptive IS algorithm can be observed in Figure 3.2.

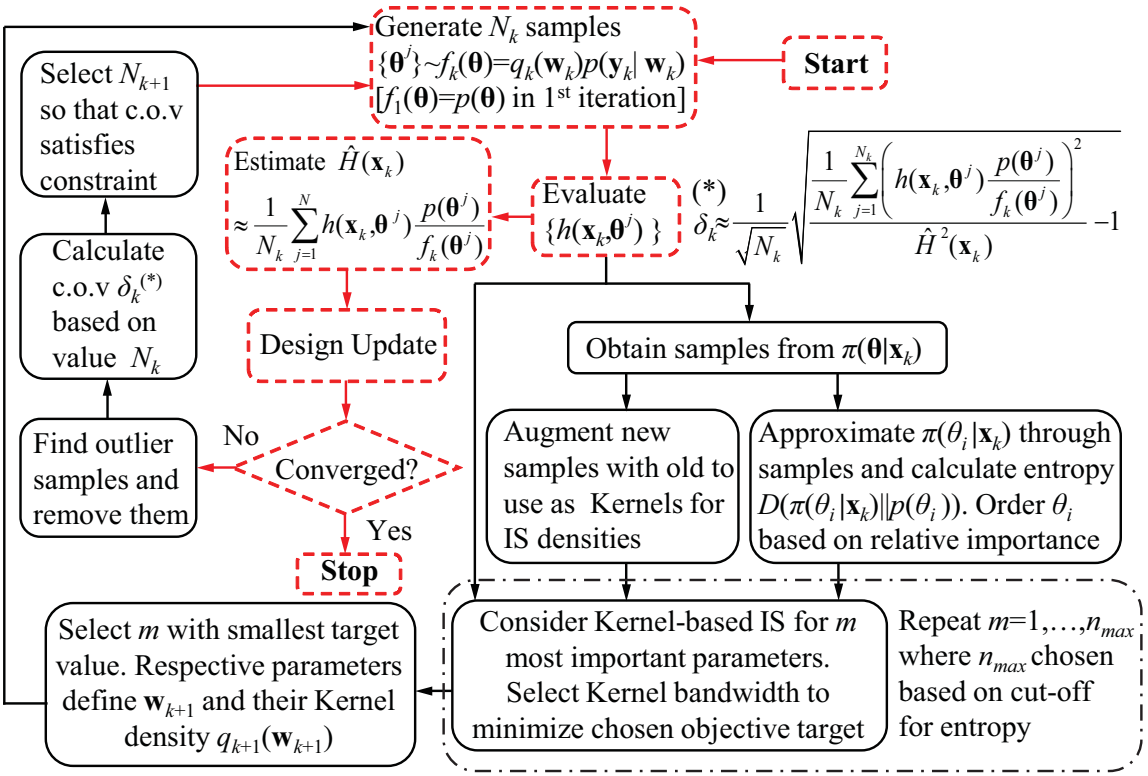


Figure 3.2 Flow diagram for adaptive Importance Sampling (IS), Components of diagram corresponding to approach without the IS formulation are indicated with dotted lines.

At iteration k of the numerical optimization algorithm [Eq. (3.1)], and assuming updating of IS is established, perform the following steps:

- Step 1 (performance function evaluation): Obtain samples $\{\boldsymbol{\theta}^j\}_k = \{\boldsymbol{\theta}^j : j = 1,...,N_k\}$ for $\boldsymbol{\theta}$ from $f_k(\boldsymbol{\theta}) = q_k(\mathbf{w}_k)p(\boldsymbol{\xi}_k | \mathbf{w}_k)$ [with $f_1(\boldsymbol{\theta})$ for the first iteration corresponding to $p(\boldsymbol{\theta})$] and evaluate $h(\mathbf{x}_k \pm c_k\Delta_k^{SP}, \boldsymbol{\theta})$, which

are required for SPSA, and the IS quotient in Eq. (3.2) for all of the samples. The value of $h(\mathbf{x}_k, \boldsymbol{\theta})$ can be obtained using Eq. (3.4).

- Step 2 (optimizer): Apply the recursive formulas to determine the $k+1$ design, for SPSA based on Eq. (A.1) and Eq. (A.2), calculate the objective function in Eq. (3.3), and check the stopping criteria. If the stopping criteria have not been satisfied, perform the following steps to select the future IS densities.
- Step 3 (stochastic sampling): Calculate γ_k through Eq. (3.7), simulate N_k uniform random numbers $\{u^j\}$ and obtain samples $\{\boldsymbol{\theta}^a\}_k$ from $\pi(\boldsymbol{\theta} \mid \mathbf{x}_k)$ through rejection sampling utilizing Eq. (3.8). If the number of samples obtained is less than N_{min}, modify γ_k in Eq. (3.8) so that N_{min} samples are obtained. This can be established by selecting γ_k to correspond to the N_{min} largest value of ratio $| h(\mathbf{x}_k, \boldsymbol{\theta}^j) | r_k(\mathbf{w}_k^j) / u^j$.
- Step 4 (augmentation of samples): Combine these samples $\{\boldsymbol{\theta}^a\}_k$ with the samples obtained from previous iterations to obtain kernel set $\{\boldsymbol{\theta}^s\}_k$. If the total number of samples available is greater than N_k, remove the oldest (meaning from previous iterations) samples to obtain set $\{\boldsymbol{\theta}^s\}_k$, consisting of exactly N_s samples.
- Step 5 (identification of importance): Using samples $\{\boldsymbol{\theta}^s\}_k$ calculate for each θ_i the approximation $\hat{\pi}(\theta_i \mid \mathbf{x}_k)$ given by Eq. (3.21) with t_i equal to $1.06 \cdot N_s^{-1/5} \sigma_i$

and σ_i corresponding to the standard deviation of the samples $\{\theta_i^s\}_k$. Then use this approximation to calculate the entropy through Eq.(2.24) .

- Step 6 (definition of minimum importance): Select D_{min} based on Eq. (3.22) and define n_{max} as the number of model parameters that have values for entropy higher than D_{min}.
- Step 7 (selection of characteristics of IS): For different values of m up to n_{max}, $m = 1, \cdots, n_{max}$, define $\tilde{\mathbf{w}}_k$ to consist of the m-most influential components of $\boldsymbol{\theta}$. Establish a pilot estimate for $k(\tilde{\mathbf{w}}_k)$ through Eq. (3.13) using for t_i the estimate in Eq. (3.14), and calculate ω^s through Eq. (3.12). Then perform optimization in Eq. (3.17) utilizing the chosen target as the objective function, either the second moment $E_{\tilde{f}_k}[h(\mathbf{x}_k, \boldsymbol{\theta})\tilde{r}_k(\tilde{\mathbf{w}}_k))^2]$ or the robust expression in Eq. (3.19), using the estimate in Eq. (3.16) and, if required, the one in Eq. (3.20) (depending on the target selection). Find the optimal characteristics $\mathbf{d}^*$ but also keep in memory the optimal target value established.
- Step 8 (final selection of IS densities): Select as optimal index m for the IS densities the one corresponding to the minimum value for the chosen target function. Select the IS density for the next iteration as $f_{k+1}(\boldsymbol{\theta}) = q_{k+1}(\mathbf{w}_{k+1})p(\xi_{k+1} \mid \mathbf{w}_{k+1})$, where $\mathbf{w}_{k+1}$ corresponds to the uncertain paramters for that index m, ξ_{k+1} are the remaining components of $\boldsymbol{\theta}$ (excluding

$\mathbf{w}_{k+1}$) and $q_{k+1}(.)$ is the sample-based density (for index m) with projected efficiency identified as optimal in the previous step.

- Step 9 (selection of number of samples): If adaptive selection for N_k has been chosen, perform first the outlier detection in Eq. (3.24), then calculate the current coefficient of variation through Eq. (3.5) and finally set N_{k+1} to the minimum value N_{req} required so that the coefficient of variation in this iteration is less than δ_{thresh} based on Eq. (3.23). Subsequently proceed to iteration $k+1$.

It is worth noting that this approach establishes an adaptive optimal selection for the number of parameters to formulate IS densities as well as the characteristics of such IS densities with the ultimate goal to minimize the anticipated coefficient of variation. More importantly, efficient IS densities are established with small computational effort since no new system evaluations are required. Thus, for problems with computationally intensive numerical models (when the computational complexity should be considered under the scope that the system model will need to be evaluated multiple times for each objective function estimation), it is expected to contribute to a great reduction of computational effort.

3.4 Case Study: Half-Car Suspension Model Driving on a Rough Road

The framework is illustrated next with an example considering the optimization of the damper characteristics for the suspension of a half-car nonlinear model riding on a rough road. Two design cases are considered, the first corresponding to a passive

damper and the second to a semi-active damper with skyhook feedback law, which is a common approach for implementing semi-active damping in vehicle suspension systems (Rill 2011; Verros et al. 2005). For the semi-active case, the actuator dynamics are not explicitly modeled in order to maintain a similar model between the two design cases since the ultimate goal is to focus on the impact of the modeling uncertainty on the design problem. Note that for all results reported herein, the optimal design configuration is dependent upon the specific probability models and performance quantification chosen. Various studies exist that examine the implications of such choices on the optimal design configuration and the degree of system robustness attained in comparison to deterministic design approaches, either in a generalized design setting (Doltsinis 2004; Beyer and Sendhoff 2007; Valdebenito and Schuëller 2010) or in the context of the specific application considered here (Taflanidis et al. 2010), i.e., controller optimization for systems under a stochastic excitation while considering parametric modeling uncertainty. As discussed in Section 2.1, such comparisons are out of the scope of the present study; the focus of this investigation is on the computational efficiency established while attaining the optimal design (through the adaptive IS formulation) given the specific probability and numerical models chosen.

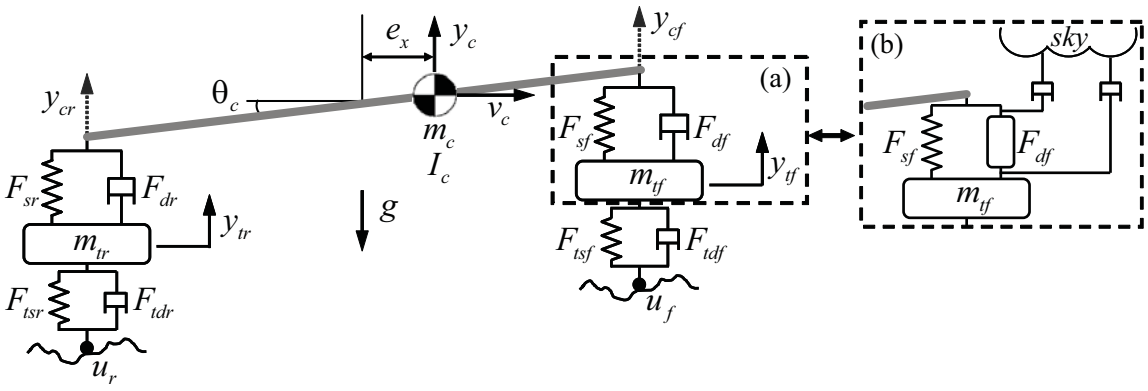


Figure 3.3: Schematics of the half-car suspension model with (a) Passive, (b) Skyhook suspension

3.4.1 Description of Simulation and Probability Models

3.4.1.1 Half-Car Simulation Model

The half-car model assumed here is shown in Figure 3.3. The chassis is represented as a rigid body connected to the tires at the ends by a combination of a spring and a dashpot. Furthermore, the tires are connected to the ground by another spring/dashpot combination. The numerical model is developed by using the small angle assumption. In this context, let y_c, y_{tf}, y_{tr}, and θ_c denote the vertical displacements of the chassis' center of mass, the front and rear tire, and the angular displacement (pitch) of the chassis, respectively. These correspond to the primary state variables for the system. The vertical displacement of the front and rear suspensions are denoted by y_{cf} and y_{cr}, respectively, and are related to y_c and θ_c by

$$
\begin{aligned}
y_{cf} &= y_c + \theta_c (1 - e_x) L / 2, \\
y_{cr} &= y_c - \theta_c (1 + e_x) L / 2,
\end{aligned}
\tag{3.25}
$$

where $e_x \in [0,1]$ is the eccentricity between the geometric center of the chassis and its center of mass, while L is the distance between the two suspensions. Also let m_c, m_{tf}, m_{tr} denote the masses for the chassis, the front and rear tires, respectively, and let I_c denote the moment of inertia of the chassis. To simplify notation represent the location of a given component by subscript $o \coloneqq \{f, r\}$ (either front or rear), with the tire spring and dashpot forces denoted by F_{to}, F_{ho}, and the suspension spring and dashpot forces denoted by F_{so}, F_{do}.

The equations of motion for the coupled system are then

$$
\begin{aligned}
m_c \ddot{y}_c + F_{sf} + F_{df} + F_{sr} + F_{dr} &= m_c g, \\
m_{tr} \ddot{y}_{tr} - F_{sr} - F_{dr} + F_{tr} + F_{hr} &= m_{tr} g, \\
m_{tf} \ddot{y}_{tf} - F_{sf} - F_{df} + F_{tf} + F_{hf} &= m_{tf} g, \\
I_c \ddot{\theta}_c + \left(F_{sf} + F_{df}\right)(1 - e_x) L / 2 - \left(F_{sr} + F_{dr}\right)(1 + e_x) L / 2 &= 0,
\end{aligned}
\tag{3.26}
$$

where g denotes gravity acceleration. Linear characteristics are assumed here for the tire spring forces and nonlinear characteristics for the suspension ones (Verros et al. 2005; Szaszi et al. 2002).

$$
\begin{aligned}
F_{so} &= K_o{}^l (y_{so} - y_{to}) + K_o{}^n (y_{so} - y_{to})^3, \\
F_{to} &= K_{to} (y_{to} - u_o),
\end{aligned}
\tag{3.27}
$$

where u_o is the vertical elevation of the road surface above a reference level, K_o^l and K_o^n are the linear and nonlinear stiffness coefficients of the suspension spring,

respectively, and K_{to} is the linear stiffness coefficient of the tire. The nonlinear part of the suspension spring is aimed to capture spring hardening effects (Szaszi et al. 2002). The tire dashpot force can be expressed as a function of the damping coefficient C_{to}, and the velocity of the road surface input $\dot{u}_o$,

$$F_{ho} = C_{to}(\dot{y}_{to} - \dot{u}_o). \tag{3.28}$$

The suspension's damper force when acting passively is modeled as

$$F_{do} = C_o^{\,l}(\dot{y}_{so} - \dot{y}_{to}) - C_o^{\,n}\left|\dot{y}_{so} - \dot{y}_{to}\right|, \tag{3.29}$$

where C_o^l and C_o^n are the damping coefficients for the linear and nonlinear component of the force, respectively. The nonlinear part of the force captures the increased stroke reaction when the suspension is moving downward and is taken as 40% of its linear counterpart based on the suggestions in (Verros et al. 2005; Szaszi et al. 2002), i.e. $C_o^n = 0.4C_o^l$. The design variables for the passive case are thus

$$\mathbf{x} = [C_f^l \quad C_r^l]^T. \tag{3.30}$$

When considered as a semi-active suspension with an idealized skyhook implementation the damper, the damper force F_{do} is modeled as

$$F_{do} = C_o^s \dot{y}_{so} - C_o^t \dot{y}_{to}. \tag{3.31}$$

The design variables for the skyhook case correspond to

$$\mathbf{x} = [C_f^t \quad C_r^s \quad C_r^t \quad C_r^s]^T. \tag{3.32}$$

Details on the technical realization of the skyhook semi-active force may be found in (Rill 2011), and as stressed earlier, an idealized implementation is assumed in this

example (actuator dynamics not considered). Note that alternative skyhook implementations could have been also formalized (Verros et al. 2005).

3.4.1.2 Excitation (Rough Road) Modeling and Response Evaluation

The road surface input u_o is modeled as a zero-mean Gaussian stationary stochastic process with Power Spectral Density $S(\Omega)$ given by (Verros et al. 2005; Robson 1979)

$$S(\Omega) = \kappa_i \left(\frac{\Omega}{\Omega_c} \right)^{-w}, \; w = \begin{cases} 2, & if \;\; \Omega \le \Omega_c \\ 1.5, & if \;\; \Omega > \Omega_c \end{cases}, \tag{3.33}$$

where Ω corresponds to the wavenumber (in cycles/m) and $\Omega_c = 1/2\pi$ is the critical wavenumber. Parameter κ_i corresponds to the roughness coefficient, whose value is defined by the International Standard Organization (ISO). A time-domain realization for u_f and u_r may be obtained by using the spectral representation method assuming that the car drives with a constant horizontal velocity v_c. This leads to the expressions

$$\begin{aligned} u_o(t) &= \sum_{n=1}^{N} \sqrt{2S(n\Delta\Omega)\Delta\Omega}\, \sin(n\Delta\Omega \cdot v_c t + \varphi_n + s), \\ \dot{u}_o(t) &= \sum_{n=1}^{N} n\Delta\Omega \cdot v_c \sqrt{2S(n\Delta\Omega)\Delta\Omega}\, \cos(n\Delta\Omega \cdot v_c t + \varphi_n + s), \end{aligned} \tag{3.34}$$

where $\Delta\Omega = 2\pi / L_r$, with L_r being the total length of road considered, taken here as $5km$, φ_n are independent random variables following a uniform distribution in the interval $[0, 2\pi)$ and $s = 0$ for front tire, while $s = L / v_c$ for the rear tire.

A numerical model for the system of Eqs. (3.26) is developed in SIMULINK (Klee 2007), and finally the response statistics for any quantity of interest under the random road excitation (and assuming the drive is long enough for the vehicle response to reach stationary conditions) are obtained through the simulation results. For instance, the standard deviation σ_{z_i} for some zero mean response quantity z_i is equivalent to the root mean square value RMS_{z_i}, which can be obtained by

$$\sigma_{z_i} = RMS_{z_i} = \sqrt{\frac{1}{T}\int_0^T z_i^2(t)dt}, \tag{3.35}$$

where $T = L_r / v_c$.

3.4.1.3 Probability Models Selection

With respect to the model parameters, all model characteristics apart from the design variables are treated as uncertain, leading to the definition of $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta} = \left[m_s \quad m_{tf} \quad m_{tr} \quad I_s \quad v_c \quad C_{tf} \quad K_{tf} \quad C_{tr} \quad K_{tr} \quad K_f^l \quad K_f^n \quad K_r^l \quad K_r^n \quad e_x \quad \kappa_i \right]. \tag{3.36}$$

The probability models are selected as follows. Independent lognormal distributions are assumed for m_s, m_{tf}, m_{tr} and I_s with median values $580kg$, $40kg$, $40kg$ and $1100kg\,m^2$, respectively, and coefficients of variation of 20%. Velocity v_c is also assumed to follow a lognormal distribution with median $60km/h$ and coefficient of variation 20%, whereas the eccentricity e_x is assumed to follow a uniform distribution in interval $[0.2, 0.8]$. The road roughness κ_i is assumed to follow a lognormal distribution with median $64{\cdot}10^{-6}\,m^2/cycle/m$ (corresponding to average road based on ISO

standards) and coefficient of variation 10%. The linear and nonlinear components of the suspension springs $[K_o^l \; K_o^n]$ for a particular location are assumed to follow a correlated lognormal distribution with correlation coefficient 40%. These distributions are independent for the front and rear of the car, but with common characteristics, median $23.5\,kN$ for K_o^l and $335\,kN$ for K_o^n and coefficient of variation 20%. Similarly, the tire stiffness and damping distributions are assumed as independent lognormal distributions that have common characteristics; the median for C_{to} is $20\,Ns/m$ and for K_{to} is $190\,kN$, with coefficient of variation 20% for each. All the aforementioned median values for the car model are based on the ones used in (Szaszi et al. 2002). A summary of the probability models employed in this problem can be seen in Table 3.2.

For the nominal vehicle model (i.e. a deterministic model with model parameters equal to the median values of the uncertain PDFs), the natural frequencies of the linearized system (ignoring the nonlinear suspension force) are calculated as [1.011 1.636 11.633 11.642 Hz]. Also, the ratio of RMS values of the nonlinear to linear components of the suspension spring forces for the vehicle with passive dampers with damping coefficients equal to $600\,Ns/m$ is estimated as 22%, indicating that the vehicle exhibits significant nonlinear behavior. Thus, estimation of its response statistics would be impractical with any other approach, apart from the simulation-based setting discussed here. The computational burden for one simulation, that is one evaluation of the response, is on average $1.5s$ on a $3\,GHz$ Xeon CPU (care was taken to establish a model that balances between numerical accuracy and efficiency).

TABLE 3.2

SUMMARY OF PROBABILITY MODELS FOR THE HALF-CAR SUSPENSION PROBLEM.

m_c	lognormal μ =580 kg cv=20%	m_{tf} m_{tr}	Lognormal μ =40 kg cv=20%	κ_i	lognormal μ =64e-6 m^2/cycle/m, cv=10%
K_f^l K_f^n	Corr. Lognormal μ for K_f^l 23.5 kN/m μ for K_f^n 435 kN/m^3 cv=20% , ρ =40%	K_r^l K_r^n	Corr. Lognormal μ for K_r^l 23.5 kN/m μ for K_r^n 435 kN/m^3 cv=20% , ρ =40%	C_{tf} C_{tr}	lognormal μ =20Ns/m, cv=20%
				K_{tf} K_{tr}	Lognormal μ =190 kN/m, cv=20%
v_c	lognormal μ =60 km/h, cv=20%	e_x	Uniform in [0.1, 0.4]	I_c	Lognormal μ =1100 kg m^2, cv=20%

In this table μ corresponds to median, cv to coefficient of variation and ρ to correlation coefficient

3.4.1.4 Probabilistic Performance Quantification

The performance measure $h(\mathbf{x},\boldsymbol{\theta})$ is selected as the normalized linear combination of the fragilities related to the root mean square of the vertical acceleration at the center of mass $(RMS_{\ddot{y}_c})$, which in turn is a measure of passenger comfort (Dahlberg 1979; Baumal et al. 1998), and of the root mean square of the suspension's damping forces at the rear and front of the car ($RMS_{F_{df}}$ and $RMS_{F_{dr}}$ respectively) – a measure of suspension fatigue. Thus, the response vector is

$$\mathbf{z}(\mathbf{x},\boldsymbol{\theta}) = \begin{bmatrix} RMS_{\ddot{y}_s} \\ RMS_{F_{df}} \\ RMS_{F_{dr}} \end{bmatrix}, \tag{3.37}$$

and the performance function is

$$h(\mathbf{x},\boldsymbol{\theta}) = \frac{1}{3}\sum_{i=1}^{3}\Phi\left[\frac{\ln z_i - \ln b_i}{\sigma_{\varepsilon_i}}\right], \tag{3.38}$$

where $\Phi[.]$ corresponds to the standard Gaussian Cumulative Distribution Function (CDF) , b_i is the threshold related to each response quantity of interest, taken here as $1.0 m/s^2$ for the acceleration, $240\,N$ for the dampers in the passive suspension and $160N$ in the skyhook suspension (more sensitive implementation for semi-active actuators, leading to lower threshold selection), and σ_{ε_i} is the coefficient of variation for the fragilities, assumed as 5% for all of them. Note that when σ_{ε_i} tends to a value of zero, the above CDF approaches the indicator function which is one if z_i exceeds b_i and zero otherwise. Ultimately the introduction of the fragilities through the CDF can be equivalently viewed as addressing unmodeled uncertainties (Taflanidis and Beck 2010): rather than having a binary distinction of the performance, i.e. perform acceptably when the response is smaller than threshold b_i and unacceptably when not, as in the case of the indicator function, the lognormal fragility introduces a preference for the response between the two extreme cases [0 or 1], with σ_{ε_i} defining the smoothness of that transition. This can be equivalently considered as the threshold corresponding to an uncertain quantity $b_i\varepsilon_i$ (because we cannot specify exactly when the response is unacceptable) or the model response having a prediction error over the real system response $z_i\varepsilon_i$ (this case will be further discussed in Chapter 5), with ε_i having (for both aforementioned cases) a lognormal distribution with median equal to one and

logarithmic standard deviation σ_{ε_i}. This then leads to the failure probability (fragility for each response quantity)

$$P[z_i > b_i\varepsilon_i] = P[\varepsilon_i \le z_i / b_i] = P[\ln(\varepsilon_i) \le \ln(z_i) - \ln(b_i)] = \Phi\left[\frac{\ln(z_i) - \ln(b_i)}{\sigma_{\varepsilon_i}}\right], \quad (3.39)$$

where the last equality is based on the fact that under the stated assumptions, $\ln(\varepsilon_{bi})$ follows a Gaussian distribution (which is a symmetric distribution) with zero mean and standard deviation equal to $\sigma_{\varepsilon b}$. The last fragility is the one ultimately appearing in Eq. (3.38), whereas the initial definition for this formulation is further exploited in Chapter 4 (in the case study considered there).

Finally, the objective function $H(\mathbf{x})$ is the average failure probability over the three different response quantities $RMS_{\ddot{y}_c}$, $RMS_{F_{df}}$ and $RMS_{F_{dr}}$, and is constrained within the $[0,1]$ range.

3.4.1.5 Complexity of the Adopted Models and of the Resultant Optimization Problem

To properly evaluate the degree of complexity of the adopted numerical model, a comparison to similar design problems discussed in the literature needs to be considered. Traditionally, vehicle suspension design has predominantly dealt with randomness only due the stochastic disturbance, i.e. the rough road (Sharp and Crolla 1987; Rakheja et al. 1994; Tamboli 1999), with many of the studies adopting a linear (or linearized) model and only a small subset of authors (Szaszi et al. 2002; Su et al. 1991) incorporating nonlinear elements. In the case of linear systems, existing approaches

tend to favor frequency-domain analysis for evaluating the performance of the system, something that is no longer feasible when nonlinearities are introduced. Few studies have incorporated probabilistic uncertainties in the model parameter description, but all of them focused on analysis and not on design (Gao et al. 2007; Gao et al. 2008). This discussion makes evident that the current model, which incorporates a large number of nonlinear components and relies on time-domain numerical simulation to evaluate the system performance while addressing parametric probabilistic uncertainties in the design problem formulation (not just due to the stochastic characteristics of the rough road), should be considered of significant complexity when compared against the relevant studies in field.

3.4.2 Optimization Details

The passive design case, also referenced herein as D_1, has a design domain with upper bounds of $[1000 \quad 1000]\,Ns/m$ and lower bounds $[0 \quad 0]\,Ns/m$. For the semi-active case, also referenced herein as D_2, the upper bounds are defined as $[450 \quad 3500 \quad 450 \quad 3500]\,Ns/m$, and the lower bounds as $[0 \quad 0 \quad 0 \quad 0]\,Ns/m$. The difference in upper bounds for the damping coefficients stems from the fact that the tire mass is smaller than the mass of the chassis; thus, the skyhook damper is expected to have a more significant effect on it. For both problems, the search domain is transformed into the respective normalized space for the design variables (the normalized design variables vary between $[0\ 1]$ within the chosen design domain).

For SPSA, common values suggested in the literature are considered since the focus here is not to evaluate the efficiency of SPSA but rather focus on the adaptive IS implementation. β^{SP} is selected as 0.602, ζ^{SP} as 0.101, and A^{SP} as 5% of the maximum allowed iterations, taken here as 200. The parameter c^{SP} is chosen as 0.0002 in the normalized space, and a^{SP} is selected so that the maximum step of the first iteration in any direction in the design space is 10% of the domain. The starting point for design problem D_1 (passive) is set as $\mathbf{x}_1 = [200 \quad 200]^T$ and for design problem D_2 (skyhook) is $\mathbf{x}_1 = [50 \quad 450 \quad 50 \quad 450]^T$. These design choices ultimately correspond to small dampers and are expected to contribute to high values for the objective function (limited vibration suppression). Due to its importance for selecting step sizes, the gradient in the first iteration $\mathbf{g}_1$ [Eq. (A.1)] is averaged over six different trials (corresponding to different random samples for the model parameters and $\mathbf{\Delta}_k^{SP}$ defining the perturbation direction). A blocking rule is implemented such that a step size larger than 10% of the search domain is allowed.

Convergence of the algorithm is determined through the infinity norm $\left\|\mathbf{x}_{k+1} - \mathbf{x}_k\right\|_\infty = \max_i |\, x_{i,k+1} - x_{i,k}\,|$, with convergence criteria satisfied if this norm does not exceed 0.04% for passive and 0.08% for skyhook of the search space in at least 4 subsequent iterations. Two additional stopping criteria are set; the total number of iterations is set to 200 [meaning 400 evaluations of the probabilistic integral given by Eq. (2.1)] and the total number of allowed simulations (calls to the system model) is defined as $500,000$. The total number of allowable simulations was chosen to

correspond to roughly $48\,h$ of simulation-time in a double-quad $3\,GHz$ Xeon workstation (utilizing parallel simulations in all eight cores simultaneously), corresponding to a time-intensive operation.

For the IS implementation, N_1 is taken as 150, while the minimum number of samples required per iteration N_{min} is 50. The factor s_f for the stochastic sampling is set to 1.1, the number of total samples for the density approximation N_s to 200, the percentage reduction s_e for definition of minimum importance to 5%, and the outlier scaling factor to 8 (note that rarely were outliers detected). For the threshold δ_{thresh} used to select N_k, the primary value considered is 10%, but results for two other values are presented, 5% and 20%. The value of 10% is a common choice for stochastic optimization applications implementing interior sampling; 5% would have been a choice more appropriate for exterior sampling implementation, whereas 20% is generally considered to correspond to high variability of the estimates that can lead to problems within the numerical optimization. The second moment is primarily used as the objective target for the IS selection, but results are also presented for the robust target function of Eq. (3.19), with parameter w_r selected as $3/4$. This is equivalent to the standard deviation weighted three times higher in the objective target selection.

Apart from the fully adaptive IS implementation (i.e. adaptively selecting the optimum kernel characteristics as well as the optimum number of model parameters for which IS should be performed), two additional cases are considered: optimization considering IS for all dimensions in $\boldsymbol{\theta}$ (the characteristics of the bandwidth are still

adaptively chosen based on the proposed scheme) and optimization without implementing IS. These three cases are denoted, respectively, as AIS (adaptive IS implementation), FIS (IS implementation for full vector $\boldsymbol{\theta}$), and MCS (Monte Carlo Simulation). The efficiency of the proposed IS scheme (AIS) is evaluated against these latter two. Comparison against MCS illustrates the efficiency of the adaptive IS formulation itself, established by sharing information across the iterations of the optimization algorithm, whereas comparison against FIS illustrates the advantages of the adaptive selection of the number of model parameters for IS through integration of global sensitivity analysis tools. The efficiency is evaluated through two different criteria: a) the number of system simulations required to obtain satisfactory accuracy for the objective function estimate and b) the number of system simulations required to converge to the optimal solution. A fourth case with robust selection of IS densities is also considered, which is denoted as AISR.

To facilitate more consistent comparisons between these different approaches for performing stochastic simulation, the aforementioned two types of convergence are distinguished and separately discussed herein: reaching the maximum number of 200 iterations or $500,000$ system model evaluations (or equivalently $30\,hr$ of computer time) are referenced as reaching *stopping criteria* (hard convergence), whereas satisfaction of the convergence for $\left\|\mathbf{x}_{k+1}-\mathbf{x}_{k}\right\|_{\infty}$ is referenced as reaching *convergence criteria*. Since the latter is ultimately a user-defined selection, in other words choosing different convergence thresholds would lead to different results, many of the

discussions focus on the hard convergence behavior of the cases compared. This allows for a comparison of the proposed IS scheme independent of the characteristics chosen for the SPSA algorithm. It should be stressed, additionally, that no special attention was given to tuning these characteristics; rather common values were used based on literature suggestions. This means that these results should not be directly used to assess the efficiency of the SPSA algorithm itself.

To judge the quality of the obtained solutions, the optimization problem was solved using exterior sampling with a very large number of samples ($N = 20,000$ for both design problems) and appropriate proposal densities to establish a small coefficient of variation around the optimal solution (less than 2%). The proposal densities were selected different for each design problem, meaning the passive and skyhook implementations. This selection is justified later. The results of this optimization provide a benchmark solution for comparing to the results obtained from the SPSA application. An efficient algorithm (based on metamodeling concepts) for costly global optimization problems was selected since the choice for a very large N imposes a significant computational cost for calculating the objective function. This algorithm was implemented through the powerful TOMLAB optimization toolbox (Holmström et al. 2006). The optimal solution for the passive case was found to be $\mathbf{x}^* = [713 \quad 606]\, Ns/m$ with respective performance $\hat{H}(\mathbf{x}^*) = 0.0104\ (1.4\%)$, whereas for skyhook, it is $\mathbf{x}^* = [186 \quad 1651 \quad 197 \quad 1512]\, Ns/m$ with respective performance $\hat{H}(\mathbf{x}^*) = 0.00066\ (6.6\%)$. It is immediately evident that the skyhook implementation

provides, as expected, a higher level of vibration-suppression as the performance objective retains a significantly lower optimal value. This means that the computational challenges for calculating the probabilistic performance (at least in regions closer to the optimum) are expected in this case to be significantly higher since the objective corresponds ultimately to a fairly rare event. This also means that the optimal IS density characteristics for the two design problems around the optimal solution are very different (different sensitivity with respect to $\boldsymbol{\theta}$), which is the reason why the proposal densities were chosen differently for each problem when implementing the exterior sampling solution.

3.4.3 Results and Discussion

To facilitate the discussions in this section, the following notations are used

- N^{tot} : Total number of model evaluations until convergence criteria are satisfied.
- N^{totc} : Total number of model evaluations until stopping criteria (maximum allowable iterations) are satisfied.
- N_k^{tot} : Total number of model evaluations up to the k^{th} iteration of the algorithm.
- $\hat{H}(\mathbf{x})$: Estimate of objective function for $\mathbf{x}$ using random samples.
- $\hat{H}(\mathbf{x} \mid \{\boldsymbol{\theta}^c\})$: Estimate of objective function for $\mathbf{x}$ using a specific set of samples $\{\boldsymbol{\theta}^c\}$ (see also the discussion in next paragraph for the formal definition of $\{\boldsymbol{\theta}^c\}$).
- n_k^{is} : Number of dimensions for which IS is adaptively formulated at iteration k.

The first three are related to computational effort for evaluating the efficiency of the different algorithms. Table 3.3 shows results for a sample run of the optimization algorithm for all cases (AIS, AISR, FIS, MCS) and target thresholds (5%, 10%, 20%) for the two design problems discussed (passive, skyhook). The optimum converged solutions $\mathbf{x}^*$, their respective performance $\hat{H}(\mathbf{x}^*)$, and the computational effort are reported. The performance is evaluated employing a large number of samples to obtain a small coefficient of variation (close to 2% for all cases reported). CRN are used across all comparisons for each design problem (but different for the skyhook and passive applications for the reasons discussed earlier) to enable a consistent comparison. This sample set is denoted $\{\boldsymbol{\theta}^c\}$, and the respective evaluation for $H(\mathbf{x})$ when this set is employed is denoted as $\hat{H}(\mathbf{x}\,|\,\{\boldsymbol{\theta}^c\})$. The computational effort in Table 3.3 is reported with respect to both the number of system simulations required for converging to the optimal solution N^{totc} as well as the number of simulations needed to reach the stopping criteria of the algorithm N^{tot}.

TABLE 3.3

RESULTS FOR A SAMPLE RUN OF THE OPTIMIZATION ALGORITHM.

δ_{thresh}	Results	AIS		AISR		FIS		MCS	
		Passive (D_1)	Skyhook (D_2)	Passive (D_1)	Skyhook (D_2)	Passive (D_1)	Skyhook (D_2)	Passive (D_1)	Skyhook (D_2)
5%	$\mathbf{x}^*$ (Ns/m)	709 566	180 1627 196 1459	708 578	182 1606 194 1483	718 570	185 1555 200 1195	710 542	164 1466 179 999
	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$	0.0105	0.00066	0.0104	0.00066	0.0105	0.00074	0.0108	0.00130
	N^{totc}	224,662	366,662	278,464	429,248	n/c+	n/c+	n/c+	n/c+
	N^{tot}	407,866	500,000	450,828	500,000	500,000	500,000	500,000	500,000
10%	$\mathbf{x}^*$ (Ns/m)	709 574	182 1588 205 1390	703 567	175 1645 190 1495	702 568	179 1630 210 1270	686 516	181 1459 209 1146
	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$	0.0105	0.00068	0.0105	0.00065	0.0105	0.00069	0.0115	0.00078
	N^{totc}	60,256	99,150	46,036	150,466	162,710	250,594	n/c+	n/c+
	N^{tot}	114,038	249,352	114,958	247,238	370,362	500,000	500,000	500,000
20%	$\mathbf{x}^*$ (Ns/m)	714 572	186 1517 202 1514	705 599	174 1622 194 1509	717 562	178 1622 195 1381	719 576	180 1652 194 1479
	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$	0.0104	0.00070	0.0103	0.00067	0.0106	0.00069	0.0104	0.00066
	N^{totc}	38,000	69,144	33,784	45,948	95,306	92,664	73,324	n/c+
	N^{tot}	57,142	94,654	46,880	86,630	138,434	228,922	155,040	500,000

+n/c stands for no-convergence, meaning that stopping criteria was reached. In this case the optimal solution reported is the one at the last iteration. Optimum design, corresponding probability of failure and computational effort are reported

Various trends are evident from Table 3.3, and many of them are discussed in more detail later (after results from multiple runs of the optimization algorithm are additionally presented); as expected, decreasing the threshold δ_{thresh} significantly

increases the computational effort (N^{tot}) since there is an inverse proportional relationship for the required N_k to achieve δ_{thresh}, and it also impacts the convergence (N^{totc}). The computational challenges are significantly higher for the D_2 design case (skyhook) since it corresponds to a design problem with significantly lower value for the objective function (average failure probability). It is interesting to note that for the FIS and MCS cases and for small δ_{thresh}, the total number of evaluations is exceeded before the 200 iterations are reached or the convergence criteria are satisfied. This is never true for AIS and AISR which show great computational savings over both MCS and FIS. This is a first strong indication of the efficiency of the adaptive IS scheme, especially of the selection of the optimal number of dimensions; AIS and AISR outperform FIS by a significant margin in terms of computational efficiency, whereas when FIS faces robustness problems over MCS (for example D_1 design case for $\delta_{thresh} = 20\%$ - this trend will be explored further later), both AIS and AISR still perform with improved levels of efficiency. With respect to optimal solution $\mathbf{x}^*$ and the respective performance, all cases examined (at least when convergence has been reached before stopping criteria are enforced) tend to achieve similar performance to the one established through the exterior sampling solution discussed earlier. This validates the efficiency of the SPSA optimization scheme adopted.

The optimization is then performed 5 different times for each case, starting from always the same initial point but using different random samples, and the results are averaged. This aims to reduce the influence of the random samples used for $\boldsymbol{\theta}$ and of

the random sequence Δ_k^{SP} for the design vector. For a given optimization run, though, the random sequence Δ_k^{SP} was kept the same for all the different algorithms in order to simplify direct comparisons. The results are presented in Figures 3.3-3.7 and in Tables 3.4 and 3.5. For the results in the figures, no convergence criterion was utilized; the optimization was carried out until the stopping criteria were enforced (200 iterations or $500,000$ total system evaluations).

Figure 3.4 presents for the D_1 (left column) and D_2 (right column) design problems the number of optimum IS dimensions n_k^{is} selected for the AIS and AISR cases. Results only for threshold 10% are reported since all other cases exhibit almost identical trends. Figure 3.5 and Figure 3.6 present for D_1 and D_2 design problems, respectively, the number of evaluations needed per iteration N_k for each δ_{thresh} selection (rows of figure) for each of the four cases considered (curves in each subplot).

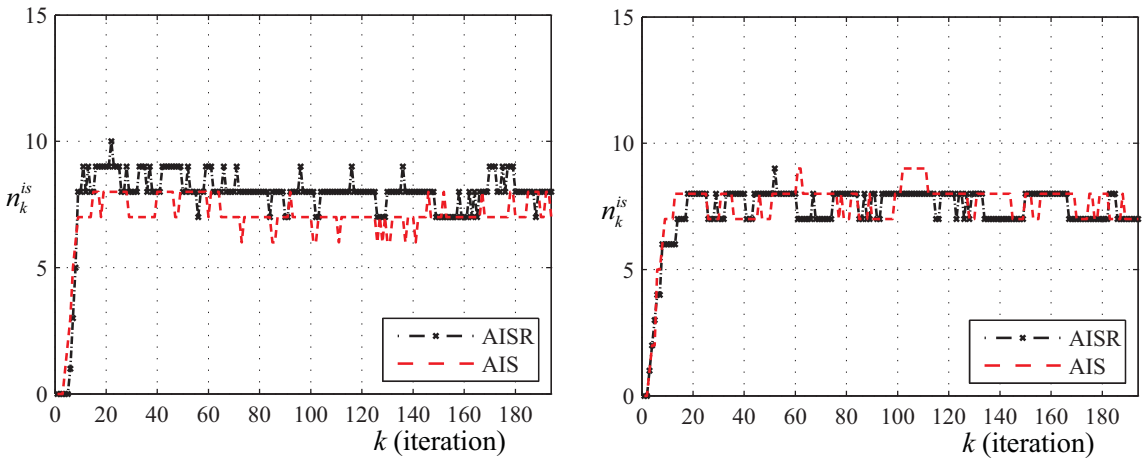


Figure 3.4 Optimum number of IS dimensions n_k^{is} selected for each k for the AIS and AISR cases. Results are reported for the

D_1 (left column) and D_2 (right column) design problems but only for $\delta_{thresh} = 10\%$ case

Figure 3.7 presents for design problem D_1 (passive) the probabilistic performance $\hat{H}(\mathbf{x}_k)$ attained as a function of the total number of simulations performed N_k^{tot} up to each stage of the optimization algorithm for each δ_{thresh} selection (rows of figure) for each of the four cases considered (curves in each subplot). In the left column, the estimates $\hat{H}$ reported are directly from the optimization algorithm itself; in other words, different random samples are used within each estimate, which makes comparison problematic (as it will be discussed next), whereas in the right column, the same random samples were used, so it corresponds to $\hat{H}(\mathbf{x} \mid \{\boldsymbol{\theta}^c\})$. Figure 3.8 presents the same results for the D_2 (skyhook) case.

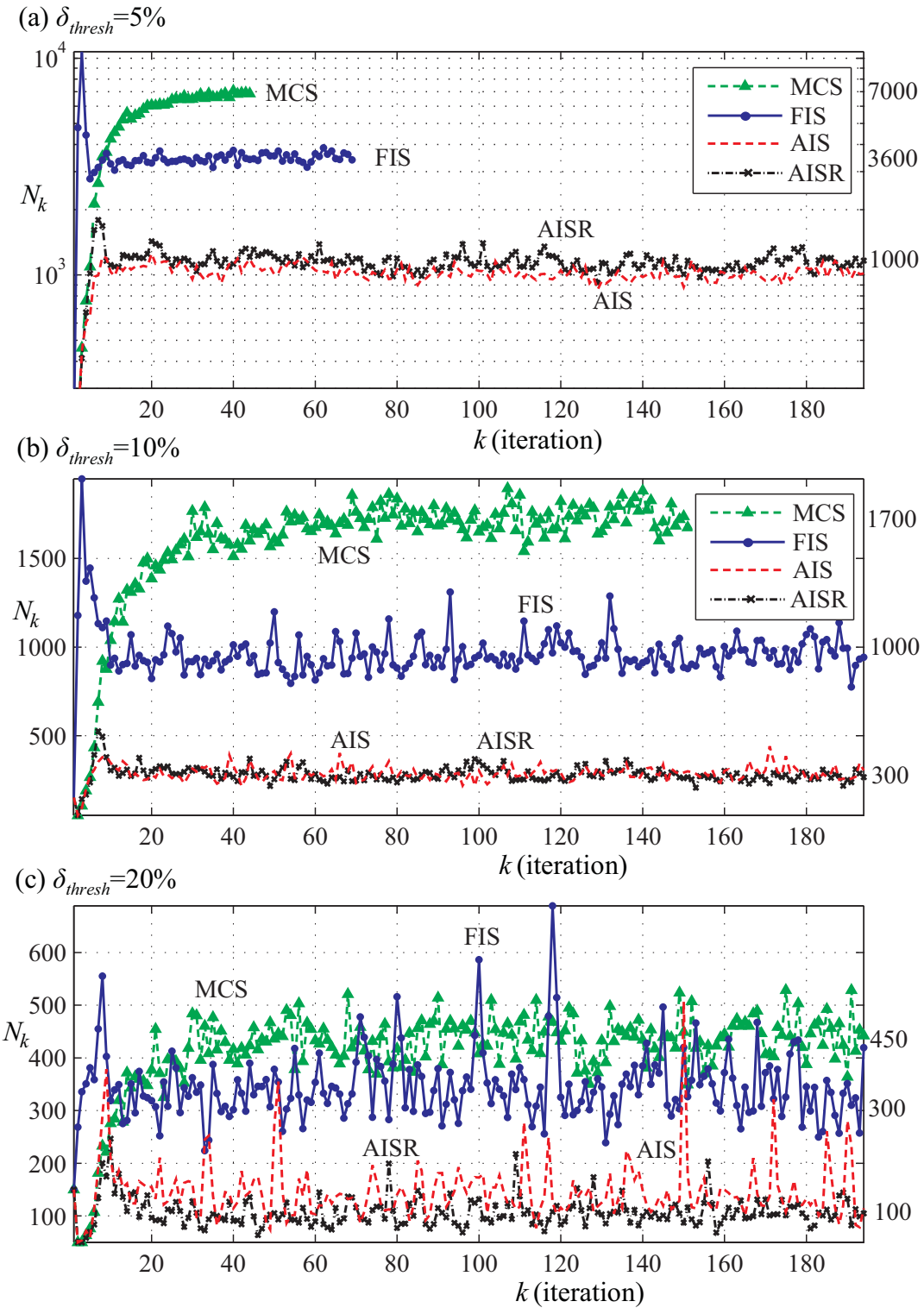


Figure 3.5 Number of evaluations N_k needed per k to establish each chosen δ_{thresh} for the D_1 (passive) design problem.

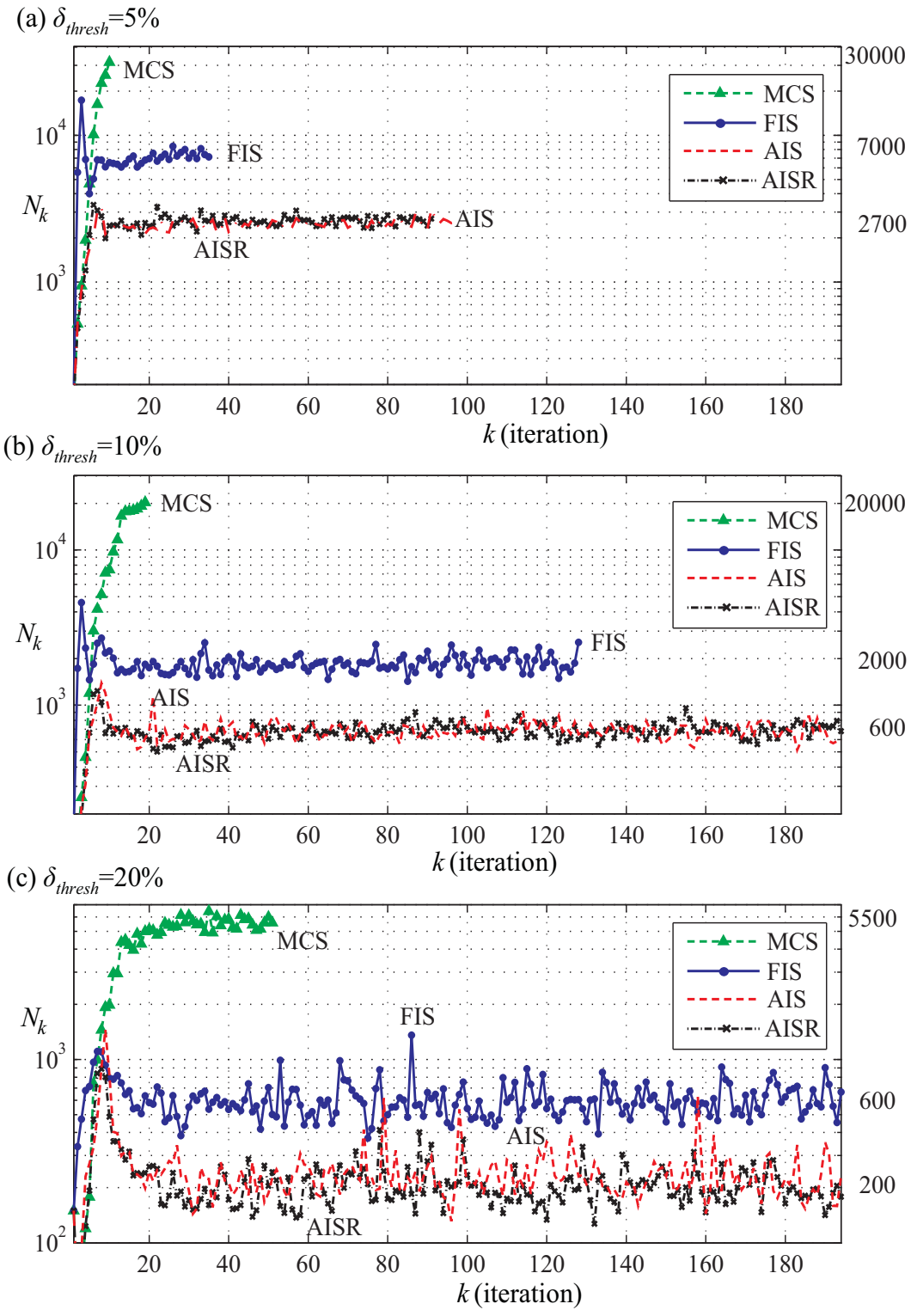


Figure 3.6 Number of evaluations N_k needed per iteration k to establish each chosen δ_{thresh} for the D_2 (skyhook) design problem.

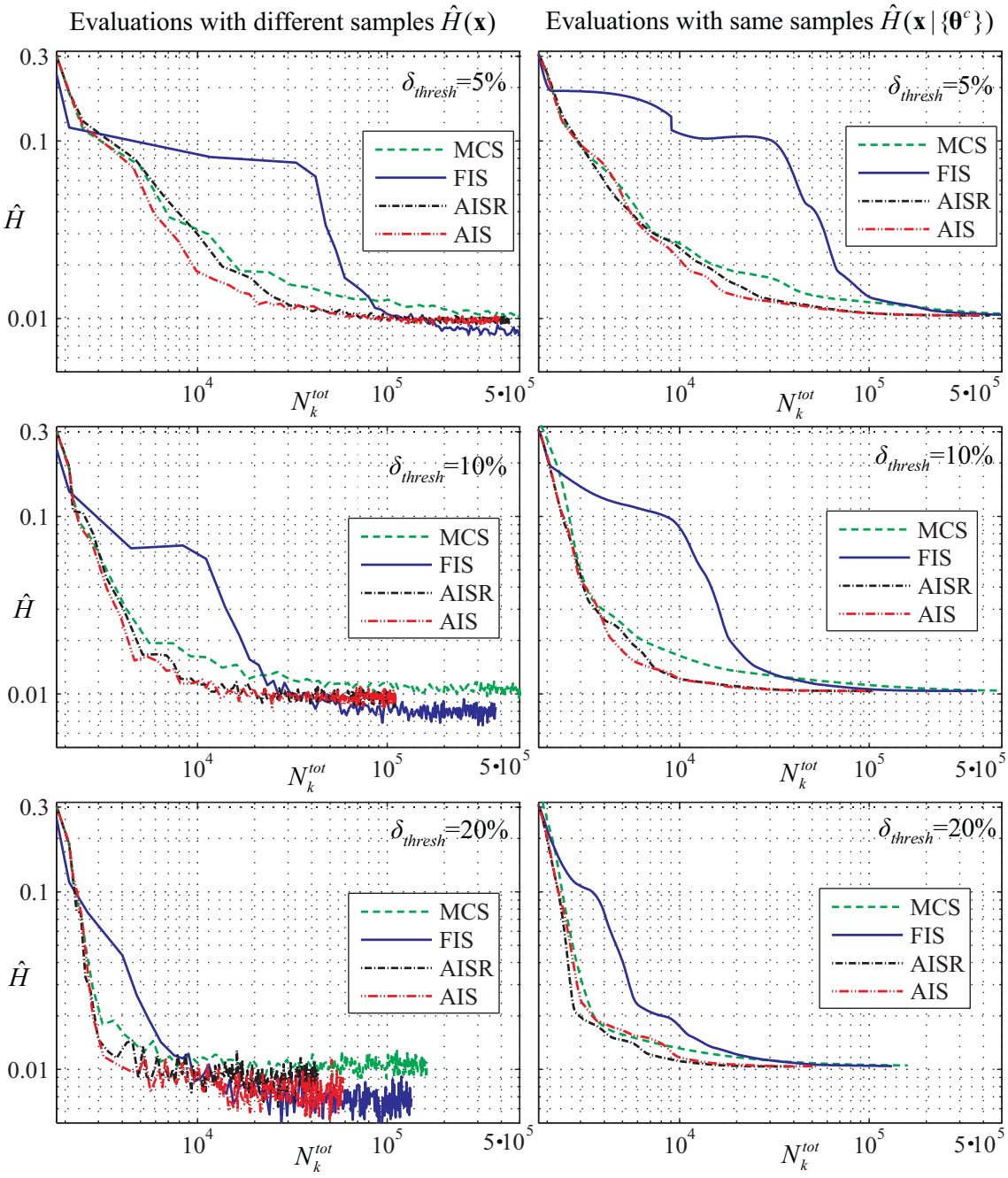


Figure 3.7 Objective function value $\hat{H}$ as a function of the total number of system simulations N_k^{tot} up to each stage of the optimization algorithm for the D_1 (passive) design problem.

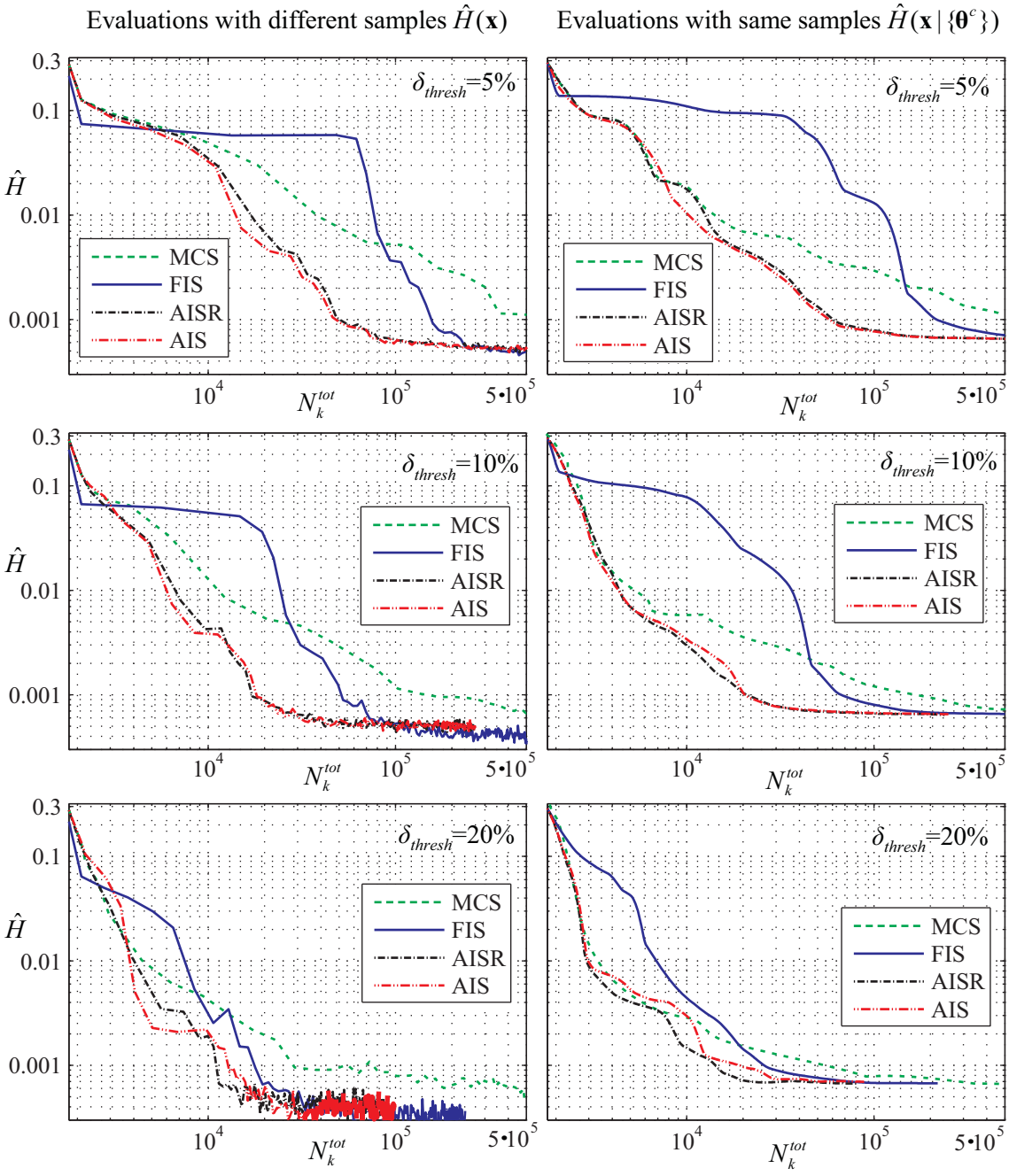


Figure 3.8 Objective function value $\hat{H}$ as a function of the total number of system simulations N_k^{tot} up to each stage of the optimization algorithm for the D_1 (skyhook) design problem.

Table 3.4 presents the N^{tot} for each case considered, that is the average number of simulations needed untill the stopping criteria are satisfied. Finally, Table 3.5 presents results for convergence based on the criteria defined earlier; in particular, it contains the optimum objective function and the computational effort needed for convergence N^{totc}. In parenthesis, the coefficient of variation for each result (over the 5 runs) is also reported in both of these tables. It is evident that all coefficients of variation are small, which shows that there is no strong dependence on the random samples utilized and, as such, the comparisons discussed next are meaningful.

TABLE 3.4

AVERAGE RESULTS OF COMPUTATIONAL EFFORT NEEDED UNTILL STOPPING CRITERIA ARE SATISFIED.

δ_{thresh}	N^{tot}							
	AIS		AISR		FIS		MCS	
	Passive (D_1)	Skyhook (D_2)	Passive (D_1)	Skyhook (D_2)	Passive (D_1)	Skyhook (D_2)	Passive (D_1)	Skyhook (D_2)
5%	396,108 (0.031)	500,000 (.)	444,902 (0.028)	500,000 (.)	500,000 (.)	500,000 (.)	500,000 (.)	500,000 (.)
10%	112,421 (0.034)	267,732 (0.057)	110,760 (0.040)	264,438 (0.059)	375,365 (0.010)	500,000 (.)	500,000 (.)	500,000 (.)
20%	58,616 (0.054)	99,078 (0.083)	43,533 (0.054)	86,696 (0.037)	135,651 (0.018)	236,842 (0.058)	162,862 (0.045)	500,000 (.)

Note: The coefficient of variation for each result is reported in parenthesis (over the 5 runs)

Figure 3.7 and Figure 3.8 demonstrate the efficiency of SPSA since both design problems converge very quickly (within 40 iterations) to objective function values close to the optimum. In initial iterations, the number of dimensions chosen for IS is zero (in other words, the AIS and AISR cases are algorithmically identical to MCS). As the algorithm moves across the design space to regions with lower objective function values (designs corresponding in this case to failure events being less likely) where there is stronger sensitivity of the response h to the $\boldsymbol{\theta}$ values (and thus potential benefits from implementing IS), n_k^{is} quickly increases. Still, the optimum number of IS dimensions remains far from the maximum number of dimensions (15 in this problem) which demonstrates a definite preference for the number of model parameters for which to perform IS and ultimately algorithmic differences between AIS (or AISR) and FIS.

Looking next at the computational effort needed per iteration (Figure 3.5 and Figure 3.6 or Table 3.4), it is evident that the skyhook design creates significantly higher challenges whereas as the algorithm moves to designs with smaller values of the objective function (correlate the results in Figures 4 and 5 to those reported in Figure 3), the required number of N_k quickly increases. In many instances, this N_k reaches some form of plateau (equilibrium). The values of that plateau are reported on the right of each figure to make the comparisons between the different cases and algorithms easier. For MCS, changes in the threshold value lead to a reduction of N_k based on the coefficient of variation proportionality relationship given in Eq. (2.15), but the same does not hold for any case implementing IS (stronger influence is evident). The reason

for this is that in the proposed scheme, the IS densities are adaptively chosen utilizing information from the system evaluations performed at each iteration; increasing δ_{thresh} leads to a reduction of N_k and thus, of the information available to adaptively select the IS densities. This poses challenges in selecting the density characteristics for minimizing the second moment in Eq. (3.15) and thus, decreases efficiency. As expected, FIS has bigger challenges; for $\delta_{thresh} = 20\%$ (when little information is available to make the IS decisions) and for the D_1 design problem, these lead to N_k values that frequently correspond to higher computational effort than performing direct MCS. For the D_2 design, the higher sensitivity makes IS in general more beneficial, leading to an improvement for FIS compared to MCS even for $\delta_{thresh} = 20\%$. It is interesting to note that the trend for higher computational burden for FIS as compared to MCS holds for all cases examined during the initial iterations when the objective function has higher values and the sensitivity of the response to $\boldsymbol{\theta}$ is evidently small. This agrees with the well-known challenges of finding efficient IS densities in high-dimensional problems. Due to the low sensitivity when the objective function is relatively high, IS is not greatly beneficial during early iterations, and trying to find an appropriate density over the entire $\boldsymbol{\theta}$ faces significantly challenges, even though these IS densities are adaptively selected to optimize the anticipated computational efficiency. This does not mean that the optimal IS distribution (given by Eq. (2.16)) does not exist, rather that the proposed KDE-based approach is not efficient for approximating it.

The AIS and AISR algorithms do not share, though, the same challenges as FIS since the additional adaptive selection of the dimensions for which to perform IS greatly improves efficiency. The computational effort for AIS never exceeds that of MCS (in the first iterations, n_k^{is} is set to zero as this is found to be the optimal case), and in most cases, it provides a significant computational advantage; as soon as the algorithm has reached regions in the design space corresponding to high sensitivity for the integrand, the values for N_k in the AIS and MCS cases have significant differences, close to 5 times for D_1 and 30 times for D_2 (the more challenging case). In addition, AIS always out-performs FIS, as expected. It is important to note that even for low values of the objective function (skyhook case), the efficiency of the proposed IS scheme is high, even though one would expect that formulation of IS densities could be challenging (rare events are analyzed). Efficient IS densities should be attributed to their sequential development, sharing information across the iterations of the optimization algorithm. Another important comment relates to the fact that using samples for the kernels (in the IS formulation) obtained over multiple iterations, not just relying on the last one, greatly improves the robustness of the proposed IS scheme. When the latter is not adopted and $\{\boldsymbol{\theta}^s\}_k$ is set to $\{\boldsymbol{\theta}^a\}_k$ (samples only from the current iteration are used as kernels), significant challenges (not great improvement of computational effort) were found for the $\delta_{thresh} = 20\%$ and 10% cases when the number of such samples is typically small. Maintaining samples from multiple iterations evidently solves this challenge and increases robustness. Note that this has nothing to do with the selection of the density

characteristics; in this case information from the samples in the current iteration are always used, but with the samples used as kernels.

Moving now to the comparison between AIS and AISR, there are small differences for the $\delta_{thresh} = 5\%$ and $\delta_{thresh} = 10\%$ cases, with AIS providing marginally better performance and bigger differences exhibited for the $\delta_{thresh} = 20\%$ case in favor of AISR. For this threshold, the reduced information (smaller N_k) on which the selection of IS densities is based leads to considerable challenges, with frequently larger N_k values for the AIS case. The robust version, which adds higher order statistical information in the selection of IS characteristics, does not suffer from similar deficiencies, verifying the advantages that this formulation has to offer. When there is not sufficient information to make the IS selections, adding this information for the anticipated efficiency constrains the algorithm from making erroneous density selection choices. It is also important to note that the differences between the AIS and AISR cases seem to stem primarily from the selection of the optimal IS density characteristics, not the number of IS dimensions, since the latter seem common (bottom row of Figure 3.4) for both cases. All these trends are further validated by the results in Table 3.4; AIS and AISR offer large improvements over FIS and MCS, whereas AISR behaves evidently better than AIS for the $\delta_{thresh} = 20\%$ case.

Moving now to a comparison of the computational burden as related to the quality of the solution obtained (Figure 3.7 and Figure 3.8), it is evident that FIS faces significant challenges at early stages of the optimization, as the number N_k^{tot} needed to

reach regions corresponding to lower objective function values is very large. This should be attributed, as reported earlier, to the attempt to formulate IS densities for all model parameters when the sensitivity of the response to $\boldsymbol{\theta}$ is small. Another interesting trend is the apparent bias in the estimates of the objective function when IS is implemented for the $\delta_{thresh} = 20\%$ case (also evident for the $\delta_{thresh} = 10\%$); the values obtained during the optimization algorithm (left column of Figures 6 and 7) are significantly smaller than the high accuracy values obtained for the $\{\boldsymbol{\theta}^c\}$ sample set. The problem is more evident for FIS (compared to AIS and AISR) and should be attributed to the fact that the accuracy threshold is set to high values (20% coefficient of variation), leading to erroneous results as IS is enforced (this again pertains to the fact that limited information is available for evaluating the statistical information). As expected, the variability of the estimates obtained (judged by the smooth characteristics of the curve) is higher for the high δ_{thresh} or the lower objective function values. This also shows that the quality of the results needs to be carefully evaluated, not relying on the objective function estimates stemming directly from the optimization algorithm. This is also the reason why such high coefficients of variation need to be avoided. Overall, the results in Figure 3.7 and Figure 3.8 verify the previous trends; AIS and AISR greatly outperform FIS and MCS, whereas improvement is evident for AISR over AIS in the $\delta_{thresh} = 20\%$ case. In the latter case, the robust selection avoids the artificial exploitation of the vulnerabilities related to the selection of kernels characteristic based on erroneous statistical estimates. The overall differences are smaller for the passive case compared to the skyhook case,

something that should be attributed to the simpler characteristics (smaller number of design variables, larger values for optimal objective function value) of the design problem. It should also be pointed out that the faster convergence rate in the first iterations for the passive design problem should not be considered as strange; the SPSA convergence characteristics are not dependent on the dimensionality of the design problem, this holds for the behavior of the algorithm in the long run, not in early iterations where that dimensionality actually has an impact.

Comparing, finally, the different cases for δ_{thresh}, it is evident that the $\delta_{thresh} = 20\%$ case leads to faster convergence. This should be considered as a validation of the SPSA optimization scheme. Despite the low accuracy estimates obtained for the objective function, through the adoption of common random numbers, good convergence behavior is achieved. The use of $\delta_{thresh} = 20\%$ may impact the variability of the estimates, though, as discussed in the previous paragraph, which is a feature one needs to be careful about. Finally, the convergence results from Table 3.5 follow similar trends, showing great computational savings for AIS and AISR, whereas the converged solution for all cases seems to establish similar levels for the performance. The main difference between the algorithms is the computational effort employed to achieve this performance.

TABLE 3.5

AVERAGE RESULTS FOR THE OPTIMUM PROBABILISTIC PERFORMANCE AND THE COMPUTATIONAL EFFORT FOR CONVERGENCE.

δ_{thresh}	Results	AIS		AISR		FIS		MCS	
		Passive (D_1)	Skyhook (D_2)	Passive (D_1)	Skyhook (D_2)	Passive (D_1)	Skyhook (D_2)	Passive (D_1)	Skyhook (D_2)
5%	N^{totc}	171,866 (0.238)	280,420 (0.291)	218,076 (0.209)	328,431 (0.192)	467,786 (0.331)	458,831 (0.142)	427,748 (0.377)	n/c^{+}
	$\hat{H}(\mathbf{x}^{*} \mid \{\boldsymbol{\theta}^{c}\})$	0.01042 (0.007)	0.000665 (0.046)	0.01043 (0.005)	0.000661 (0.033)	0.01052 (0.022)	0.000712 (0.092)	0.01081 (0.051)	0.00119 (0.230)
10%	N^{totc}	461,555 (0.332)	105,742 (0.292)	64,603 (0.205)	103,136 (0.372)	208,177 (0.291)	239,221 (0.250)	172,261 (0.287)	n/c^{+}
	$\hat{H}(\mathbf{x}^{*} \mid \{\boldsymbol{\theta}^{c}\})$	0.01042 (0.006)	0.000671 (0.024)	0.01041 (0.008)	0.000658 (0.023)	0.01043 (0.011)	0.000667 (0.052)	0.01072 (0.046)	0.00074 (0.11)
20%	N^{totc}	39,368 (0.234)	58,640 (0.357)	33,733 (0.218)	52,871 (0.298)	91,162 (0.368)	110,735 (0.236)	81,990 (0.386)	462,317 (0.157)
	$\hat{H}(\mathbf{x}^{*} \mid \{\boldsymbol{\theta}^{c}\})$	0.01043 (0.005)	0.000708 (0.048)	0.01038 (0.007)	0.000671 (0.042)	0.0105 (0.012)	0.000681 (0.077)	0.01044 (0.005)	0.00067 (0.048)

Notes: The coefficient of variation for each result is also reported in parenthesis (over the 5 runs). $^{+}$n/c stands for no-convergence, meaning that stopping criteria were satisfied prior to convergence criteria. In this case, the performance reported is the one at the last iteration

Overall, the study shows potential for improved efficiency (reduction of computational burden) through the proposed adaptive IS scheme as it provides great computational savings for the stochastic simulation utilizing readily available system model evaluations. Some of the key results in this illustrative example are the following:

- The adaptive selection of the IS dimensions offers great robustness and circumvents the challenges associated with trying to formulate efficient IS

densities for all dimensions (FIS). AIS (AISR) was never outperformed by either MCS or FIS, as expected since the latter two are actually subsets of AIS.

- Making decisions based on reduced available information (meaning larger values for δ_{thresh}) has some impact on the adaptive IS formulation, but this impact can be limited if the robust version (AISR) is utilized for making IS selections.
- Even for low values of the objective function (skyhook case), corresponding to high sensitivity to $\boldsymbol{\theta}$, the efficiency of the proposed IS scheme is high. This should be attributed to the fact that IS densities are formulated by sequentially sharing the information across iterations of the optimization algorithm. In other words, this high sensitivity does not need to be captured within a single iteration, rather as the optimization algorithm converges to a design.
- Adaptive IS performed at similar level of efficiency across all considered cases, independently of the problem characteristics (sensitivity level, number of design variables, thresholds chosen).
- Developing IS densities based on samples (for the kernels) obtained over multiple iterations, not just relying on the last iteration, greatly improves the robustness of the proposed IS scheme when the number of such samples (from the current iteration) is small.

3.5 Summary

In this chapter, adaptive implementation of Importance Sampling (IS) was considered to reduce the computational burden associated with optimization under

uncertainty problems adopting stochastic simulation. Even though IS methods have attracted significant attention over the past decades, there had been no investigation aiming at their efficient integration within numerical optimization algorithms by sharing information across the iterations of the latter. Two different challenges were addressed for IS implementation: selection of the shape and characteristics of the IS densities and selection of the model parameters for which IS is to be applied.

The proposed framework has a general form (no restrictions on the number of design variables or model parameters or the complexity of the numerical and probability models considered) and can be combined with any algorithm based on local searches. The Simultaneous Perturbation Stochastic Approximation (SPSA) was chosen here for this purpose, an algorithm that is appropriate for complex engineering problems for which analytical expressions for the gradient might be impractical to derive. The formulation of IS densities was based on samples that are distributed according to the optimal IS density (the integrand of the integral corresponding to the probabilistic performance). Rejection sampling was considered for obtaining these samples utilizing readily available system model evaluations (no additional computational effort was needed). Kernel density estimation was considered for description of the target density through available samples with density characteristics that were optimally selected by maximizing the anticipated accuracy for the estimate of the performance objective if these densities were to be implemented as IS densities. A framework was successfully established for facilitating this adaptive formulation of IS densities without adding to the computational burden using only system evaluations

that have been already performed (when evaluating the objective function). To avoid the numerical problems when trying to develop IS for all model parameters, a relative prioritization was also introduced by integrating a recently developed global sensitivity analysis to quantify the relative importance of individual model parameters to the overall probabilistic performance. Development of IS densities was considered only for the most important ones. Ultimately, the global sensitivity comparison provided a relative prioritization for each uncertain model parameter, whereas the optimization for the expected accuracy provided an optimal selection of the characteristics of the IS densities and of the exact number of model parameters considered in the IS formulation.

The illustrative example demonstrated the efficiency and robustness of the proposed adaptive IS formulation and stressed the importance of prioritization of the appropriate subset of model parameters for forming the IS densities. The improvement in computational cost was remarkable when compared to direct stochastic simulation, and the adaptive selection of the IS dimensions circumvented the challenges associated with trying to formulate efficient IS densities for all dimensions. When making decisions based on reduced available information (limited number of available samples for calculating the anticipated efficiency), the robust selection of the kernel characteristics offered significant advantages. Overall, the adaptive IS implementation, coupled with SPSA for the stochastic optimization, was shown to yield a significant increase in algorithmic efficiency without compromising the quality of the results. Even within challenging optimization settings (great sensitivity of the response to the uncertain

model parameters), the proposed scheme provided results that demonstrate great robustness and efficiency (actually greater for such cases, since the benefits from the IS formulation are expected to be higher).

CHAPTER 4:

OPTIMIZATION UNDER UNCERTAINTY WITH ADAPTIVE KRIGING IN AUGMENTED SPACE

This chapter considers the implementation of surrogate modeling to approximate the system performance (or more appropriately the system response) for optimization under uncertainty problems that rely on stochastic simulation. As discussed in Chapter 1, kriging is adopted as a surrogate model since it has been proven highly efficient for approximating complex response functions while simultaneously providing gradient information and a local approximation of the standard deviation of the metamodeling prediction error. The latter is incorporated within the probabilistic performance quantification in the illustrative example considered, and the benefits from this incorporation are demonstrated.

Although the framework can also be implemented for approximating the system response with respect to just the model parameters, an augmented formulation is considered here that offers a greater generality as well as larger potential computational savings. Rather than building a separate kriging model for each design choice examined (something that would be required if the surrogate model was established only with respect to the model parameters), an augmented input space is considered. Specifically, a sub-region of the design space is defined, and a kriging

metamodel is built to approximate the system response (output) with respect to both the design variables and the uncertain model parameters. High-fidelity model evaluations are obtained at properly selected support points (experiments), and the kriging model is then developed employing this information. This metamodel is then used within a stochastic simulation setting to approximate the system performance when estimating the objective function and its gradient for specific values of the design variables, where the stochastic simulation is ultimately established with respect to the uncertain model parameters. This information (i.e. estimate of objective function and gradient) is then used to search for a local optimum within the previously established design sub-domain. Only when the optimization algorithm drives the search outside this sub-domain is a new metamodel generated, and the process is iterated until convergence is obtained.

This framework provides great computational savings since the high-fidelity model is only utilized for calculating the response for the chosen support points (as long as the design choices remain within the considered sub-domain in the design space). The main challenge is how to adaptively tune the kriging model, and various enhancements are proposed toward this goal. For selecting the basis functions of the metamodel, the global probabilistic sensitivity analysis introduced in Section 2.4.4 is utilized. Higher order basis functions are assigned to the more important variables, contributing to increased approximation accuracy. Furthermore, an adaptive DoE approach is developed for selecting the support points, populating more densely those regions in the model parameter space that have higher contribution to the integrand

quantifying the probabilistic performance. This leads to a kriging model with enhanced accuracy in those regions, something that ultimately improves the accuracy of the objective function estimates. This adaptive DoE is accomplished with a very small computational burden utilizing readily available information.

Before advancing further, it should be pointed out that the underlying assumption in this chapter is that the input space for formulation of the kriging metamodel is not of high dimension (say below 50 (Simpson et al. 2001c)) since otherwise it would require a significant amount of information (number of support points) to get an accurate approximation, making the implementation impractical. This is an important constraint when considering the applicability of this framework, and even though the approach discussed here can offer remarkable computational savings compared to the adaptive IS formulation discussed in the previous chapter, it is important to remember that the IS formulation suffers from no such constraints.

The nomenclature specific to this chapter is reviewed in Table 4.1.

TABLE 4.1

RELEVANT NOMENCLATURE FOR CHAPTER 4

$\hat{\varepsilon}(\mathbf{y})$	Kriging prediction error	n	Total number of support points
$\boldsymbol{\theta}^a$	Samples from $\hat{\pi}(\boldsymbol{\theta}\mid\mathbf{x})$	n_2	Number of support points in second stage of Hybrid DoE
μ_i^z	Mean of output over observation set	n_1	Number of support points in first stage of hybrid DoE
$\hat{\pi}(\boldsymbol{\theta}\mid\mathbf{x})$	Kriging approximation of $\pi(\boldsymbol{\theta}\mid\mathbf{x})$	n_y	Dimension of $\mathbf{y}$
σ_i^z	Standard deviation of output	$p(\mathbf{x})$	Artificial probability density function for $\mathbf{x}$
$\sigma_{\hat{\varepsilon}}^2(\mathbf{y})$	Variance of prediction error	$\mathbf{R}$	Correlation Matrix (kriging)
$\boldsymbol{\varphi}$	Weight parameter vector for correlation function	$\mathbf{r}$	Correlation vector
$\mathbf{a}^*, \mathbf{b}^*$	Optimal kriging solutions	s_e	Scaling parameter to determine D_{min}
D_{min}	Threshold for selecting the IS parameters	s^r	Scaling parameter for trust region reduction
$D_{\min}^{re}$	Threshold for selecting the order of kriging basis functions	s_e^{re}	Scaling parameter to determine $D_{\min}^{re}$
$\mathbf{F}$	Basis functions matrix (kriging)	$\underline{\cdot}$	(underscore) Normalized versions for output quantities
$\overline{\mathbf{f}}$	Basis functions vector	$\mathbf{y}$	Augmented input vector
$H_{kring}(\mathbf{x})$	Objective function estimated with kriging	X_k	Design subdomain at iteration for k^{th} iteration
$\hat{h}(\mathbf{x},\boldsymbol{\theta})$	Approximation of performance function through kriging	$\mathbf{x}_k^u$	Upper subdomain bounds for k^{th} iteration
$\cdot_i$	(Subscript) i^{th} component of vector	$\mathbf{x}_k^l$	Lower subdomain bounds for k^{th} iteration
$\cdot^j$	(Superscript) j^{th} sample	$\mathbf{Z}$	Output (observation) matrix
$\cdot_k$	(Subscript) k^{th} iteration of the optimization algorithm	$\mathbf{z}$	Output vector
$\mathbf{L}_k$	Subdomain length vector	$\hat{\mathbf{z}}(\mathbf{x},\boldsymbol{\theta})$	Median kriging approximation of output vector

4.1 Optimization Problem Formulation and Augmented Input Space Definition

Metamodeling techniques for design problems under uncertainty are frequently used for approximating the objective function (output) with respect to the design variables (input) over the design domain (Jin et al. 2003; Wang and Shan 2007; Persson and Ölvander 2013; Lee et al. 2006; Huang et al. 2006). This approach requires the probabilistic objective in Eq. (2.1) to be evaluated at several design locations in the design space (corresponding to the support points). Since for each of these values of the objective function evaluation, N simulations of the computationally expensive numerical model are required, the overall savings established are relatively small. In contrast, the formulation adopted here considers directly approximation of the system response $z(\mathbf{x},\boldsymbol{\theta})$, and in particular, the input considered extends to both the design variables as well as the model parameters.

This augmented input space is defined as a tensor product between the design and uncertain spaces $X \otimes \Theta$ (Kharmanda et al. 2002), whereas to improve accuracy of the metamodel, a smaller set of the design domain is considered instead of the entire domain, establishing an iterative approach similar to the one described by Eq. (3.1). This approach is denoted herein by

$$\mathbf{x}_{k+1} = G_{krig}(\mathbf{x}_k \mid \{\boldsymbol{\theta}^j\}_k), \tag{4.1}$$

where the function G_{krig} represents the kriging-based optimization recursive relations and the notation $\{\boldsymbol{\theta}^j\}_k$ is used to denote the sample set used within the k^{th} iteration. Note that this sample set ultimately changes from iteration to iteration, i.e. the proposed

formulation corresponds to an interior sampling approach. The design subdomain in the k^{th} iteration, also known as the trust region, of the optimization algorithm will be denoted X_k in the remaining of this chapter. To simplify some of the discussions, an equivalent distribution for $\mathbf{x}$ $p(\mathbf{x})$ is utilized, corresponding to a uniform distribution in X_k. This should not be interpreted as $\mathbf{x}$ being uncertain, simply as a tool to have uniformity in some of the methodologies and terminology analogous to $\boldsymbol{\theta}$, which has its own distribution $p(\boldsymbol{\theta})$ (since it is actually an uncertain variable).

Thus, the input vector $\mathbf{y}$ for the kriging metamodel is composed of the design and uncertain model parameter vectors

$$\mathbf{y} = [\mathbf{x} \quad \boldsymbol{\theta}], \tag{4.2}$$

while the output vector corresponds to the system response vector $\mathbf{z}(\mathbf{x},\boldsymbol{\theta})$. The motivation for establishing a kriging approximation for the response and not for the overall performance function $h(\mathbf{x},\boldsymbol{\theta})$ (which would have been the alternative choice) is the fact that, as discussed in Section 2.1, the computational complexity of the performance evaluation model for estimating $h(\mathbf{x},\boldsymbol{\theta})$ based on $\mathbf{z}(\mathbf{x},\boldsymbol{\theta})$ is typically small. Establishing an approximation for $\mathbf{z}(\mathbf{x},\boldsymbol{\theta})$ and then using the actual performance evaluation model to estimate $h(\mathbf{x},\boldsymbol{\theta})$ circumvents one level of approximations and can ultimately offer significant improvements in accuracy (Jin et al. 2003; Persson and Ölvander 2013). This approach further allows, as will be demonstrated in the illustrative example, the explicit consideration of the local kriging prediction error within the definition of the performance function in the metamodeling formulation.

The kriging model provides an approximation for the response vector $\hat{\mathbf{z}}(\mathbf{y}) = \hat{\mathbf{z}}(\mathbf{x}, \boldsymbol{\theta})$, and through this, an approximation to the performance function is established, denoted by $\hat{h}(\mathbf{y}) = \hat{h}(\mathbf{x}, \boldsymbol{\theta})$. Simultaneously, gradient information can be also obtained for both of these quantities as will be demonstrated in the next section. Using this information, the numerical optimization scheme G_{krig} can be formulated establishing ultimately a local search within the trust region X_k (defined by its upper and lower bounds $\mathbf{x}_k^l$ and $\mathbf{x}_k^u$, respectively). In this step, the objective function and its gradient are then approximated as

$$H_{krig}(\mathbf{x}) = \int_{\Theta} \hat{h}(\mathbf{x}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{4.3}$$

$$\nabla H_{krig}(\mathbf{x}) = \nabla \left(\int_{\Theta} \hat{h}(\mathbf{x}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \right) = \int_{\Theta} \nabla \left(\hat{h}(\mathbf{x}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \right) = \int_{\Theta} p(\boldsymbol{\theta}) \nabla \hat{h}(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{4.4}$$

where for Eq. (4.4), we assume that the functions $\hat{h}(\mathbf{x}, \boldsymbol{\theta}) p(\boldsymbol{\theta})$ and $(\partial \hat{h}(\mathbf{x}, \boldsymbol{\theta}) / \partial x_i) p(\boldsymbol{\theta})$ (appearing within the gradient vector) are continuous on the domain $X \times \Theta$ and bounded, so the differentiation and expectation operators can commute (Spall 2003). The independence between $p(\boldsymbol{\theta})$ and $\mathbf{x}$ was utilized in the last equality. These probabilistic integrals can be then evaluated through stochastic simulation. Due to the computational efficiency of the kriging metamodel, a large number of samples can be utilized. Furthermore, IS can be considered for cases in which the computational burden (despite this simplicity) is considered large. Following the IS formulation of Section 2.4.1, the approximations to Eqs. (4.3) and (4.4) are

$$\hat{H}_{krig}(\mathbf{x}) = \frac{1}{N} \sum_{j=1}^{N} \hat{h}(\mathbf{x}, \boldsymbol{\theta}^j) \frac{p(\boldsymbol{\theta}^j)}{f_k(\boldsymbol{\theta}^j)}, \tag{4.5}$$

$$\nabla \hat{H}_{krig}(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N}\frac{p(\boldsymbol{\theta}^j)}{f_k(\boldsymbol{\theta}^j)}\nabla \hat{h}(\mathbf{x},\boldsymbol{\theta}^j), \tag{4.6}$$

where sample set $\{\boldsymbol{\theta}^j\}_k$ used in the stochastic simulation corresponds to independent identically distributed (i.i.d) samples of the model parameters that follow the distribution $f_k(\boldsymbol{\theta})$, corresponding to the proposal density used in the k^{th} iteration of the algorithm.

Utilizing this information, especially the gradient approximation in Eq. (4.6), an appropriate gradient-based algorithm is adopted to establish a local search within X_k. Two possible outcomes can occur for the optimization described by Eq. (4.1): (i) converge to a local optimum within X_k or (ii) reach the boundary of the search domain, which means that the local search should stop to avoid extrapolations. This prompts the optimization algorithm to advance to the next iteration $\mathbf{x}_{k+1}$ and generate a new kriging model if the overall optimization has not converged. Note the local search optimization searching for the optimal solution within X_k [utilizing Eqs. (4.5) and (4.6)] has evidently its own inner iterations, but we are interested here in the iterations of the exterior optimization algorithm characterized by Eq. (4.1). In addition, if there is an overlapping region between X_k and X_{k+1}, the supports point laying on the intersection of these regions can be reused at iteration $k+1$. The concept of stochastic simulation in augmented space is illustrated in Figure 4.1.

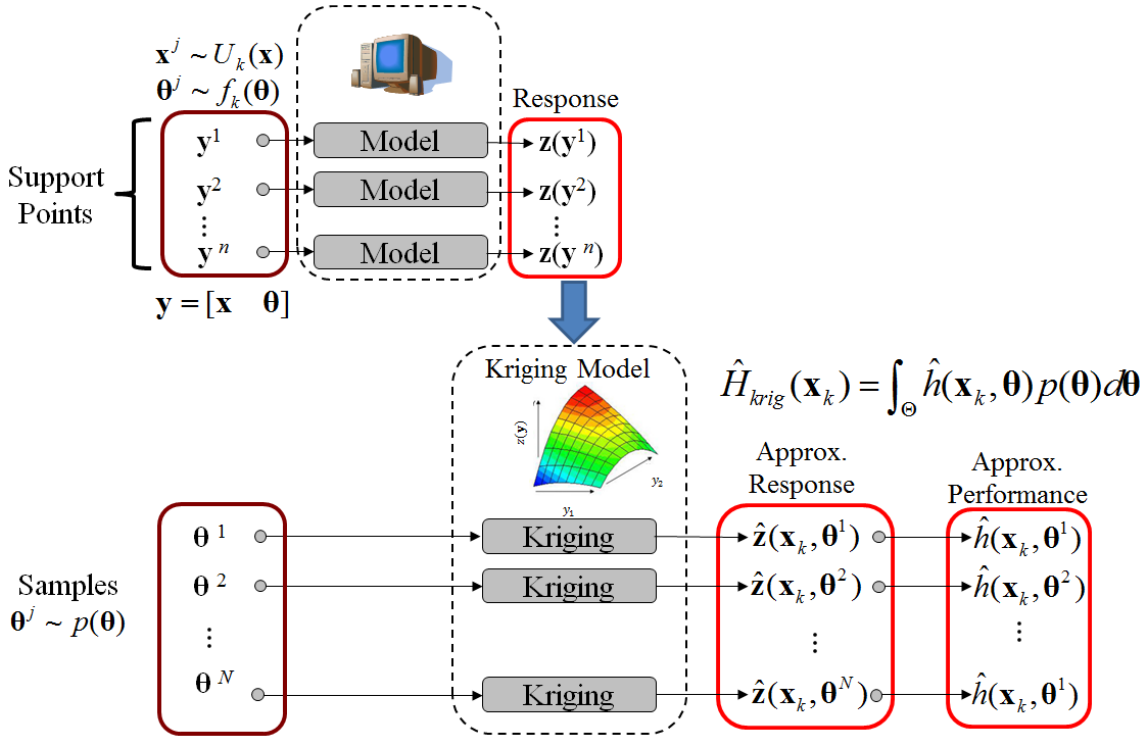


Figure 4.1 Illustration of stochastic simulation in augmented space

In the next section, the kriging metamodel implementation is briefly reviewed, and then the proposed adaptive kriging formulation is presented in detail.

4.2 Review of Kriging Metamodeling Formulation

The kriging approximation is established with respect the input/output pair $\mathbf{y}$ / $\mathbf{z}$ given a set of n support points (selection discussed later), also known as the training or observation set, consisting of evaluations of the response vector $\{\mathbf{z}^h; h = 1,...,n\}$ for different input conditions $\{\mathbf{y}^h; h = 1,...,n\}$, which will be denoted by $\mathbf{Z} \in \mathbb{R}^{n \times n_z}$ and $\mathbf{Y} \in \mathbb{R}^{n \times n_y}$, respectively. The corresponding input and output matrices are $\mathbf{Z} = [\mathbf{z}^1 \ \cdots \ \mathbf{z}^n]^T$ and $\mathbf{Y} = [\mathbf{y}(\mathbf{x}^1) \ \cdots \ \mathbf{y}(\mathbf{x}^n)]^T$, respectively.

Utilizing kriging, the i^{th} component of the response output vector $z_i(\mathbf{y})$ can be approximated though the median prediction from the surrogate model $\hat{z}_i(\mathbf{y})$ and the corresponding error $\hat{\varepsilon}_i(\mathbf{y})$ such that

$$z_i(\mathbf{y}) = \hat{z}_i(\mathbf{y}) + \hat{\varepsilon}_i(\mathbf{y}), \tag{4.7}$$

where $\hat{\varepsilon}_i(\mathbf{y})$ is a Gaussian prediction error with zero mean and variance $\sigma^2_{\hat{\varepsilon}_i}(\mathbf{y})$. The probabilistic performance can be estimated through the median kriging model predictions $\hat{z}_i(\mathbf{y})$, while also adjusting for the existence of $\hat{\varepsilon}_i(\mathbf{y})$.

It is common to establish some normalization for $\mathbf{z}$. The typical choice is to transform each of its components into zero mean and unit variance under the statistics of the observation set (Lophaven et al. 2002b)), leading to

$$\underline{z}_i = \frac{z_i - \mu_i^z}{\sigma_i^z}, \tag{4.8}$$

where

$$\begin{aligned} \mu_i^z &= \frac{1}{n}\sum_{l=1}^{n} z_i^l, \\ \sigma_i^z &= \sqrt{\frac{1}{n}\sum_{l=1}^{n}\left(z_i^l - \mu_i^z\right)^2}. \end{aligned} \tag{4.9}$$

The surrogate model is then formulated for $\underline{z}_i$, and the predictions are transformed to z_i by the inverse of Eq. (4.8). The corresponding (normalized) vector for the output is denoted by $\underline{\mathbf{z}}$ and the observation matrix $\underline{\mathbf{Z}}$.

4.2.1 . Scalar Case

The main assumption behind the kriging model is that the true (scalar) function $\underline{z}_i(\mathbf{y})$ is one realization of a stochastic process (Sacks et al. 1989)

$$\underline{z}_i(\mathbf{y}) = \sum_{j=1}^{p} \alpha_{ij} \overline{f}_{ij}(\mathbf{y}) + S_i(\mathbf{y}) = \overline{\mathbf{f}}_i \boldsymbol{\alpha}_i^T + S_i(\mathbf{y}), \tag{4.10}$$

where $\overline{\mathbf{f}}_i(\mathbf{y}) = [\overline{f}_{i1}(\mathbf{y}) \ \cdots \ \overline{f}_{ip}(\mathbf{y})]$ is the basis function vector, $\boldsymbol{\alpha}_i = [\alpha_{i1} \ \cdots \ \alpha_{ip}]$ is an unknown coefficients vector, p the total number of bases considered, and $S_i(\mathbf{y})$ is a Gaussian stochastic process with zero mean and stationary covariance

$$\mathrm{cov}_i(\mathbf{y}^I, \mathbf{y}^J) = \tilde{\sigma}_i^2 R_i(\mathbf{y}^I, \mathbf{y}^J), \tag{4.11}$$

where $R_i(\mathbf{y}^I, \mathbf{y}^J)$ is the chosen spatial correlation function and $\tilde{\sigma}_i^2$ is the process variance. While $\overline{\mathbf{f}}_k \boldsymbol{\alpha}_k^T$ provides a "global" model in the $\mathbf{Y}$ space function, $S_k(\mathbf{y})$ creates a "localised" deviation weighting the points in the training set $\mathbf{Y}$ that are closer to the target point $\mathbf{y}$.

The common choice for basis functions is polynomials of some low order, typically up to second order. For example, a complete second order polynomial is

$$\overline{\mathbf{f}}_i(\mathbf{y}) = [1 \ \ y_1 \ \cdots \ y_{n_y} \ \ y_1^2 \ y_1 y_2 \ \cdots \ y_{n_y}^2 \ ]. \tag{4.12}$$

Improved accuracy however, can be established if the order of the polynomial for the different components of the vector $\mathbf{y}$ is explicitly optimized (Jia and Taflanidis 2013).

For the correlation function, different choices have been proposed in the literature (Rasmussen and Williams 2006), ultimately impacting the smoothness and

differentiability of the derived metamodel. The specific correlation function adopted later in the examples is a generalized exponential function, leading to

$$R_i(\mathbf{y}^I, \mathbf{y}^J) = \prod_{i=1}^{n_y} \exp[-\varphi_i \,|\, \mathbf{y}_i^I - \mathbf{y}_i^J \,|^{\varphi_{n_y+1}}], \quad \boldsymbol{\varphi} = [\varphi_1 \ \cdots \ \varphi_{n_y+1}], \tag{4.13}$$

where $\boldsymbol{\varphi}$ corresponds to the vector of weights. The proper tuning of these weights contributes to the accuracy of kriging (Jones 2001) and to its ability to approximate highly complex functions.

For the set of n observations (support points) with input matrix $\mathbf{Y}$ and corresponding output vector for the i^{th} dimension $\underline{\mathbf{Z}}_i \in \mathbb{R}^{n\times 1}$, with $\underline{\mathbf{Z}}_i = [\underline{z}_i(\mathbf{y}^1)\cdots\underline{z}_i(\mathbf{y}^n)]^T$, we define the basis matrix $\mathbf{F}_i \in \mathbb{R}^{n\times p}$ such that $\mathbf{F}_i = [\overline{\mathbf{f}}_i(\mathbf{y}^1)\cdots\overline{\mathbf{f}}_i(\mathbf{y}^n)]^T$ and the correlation matrix $\mathbf{R}_i \in \mathbb{R}^{n\times n}$ with the IJ-element defined as $\mathbf{R}_i^{IJ} = R_i(\mathbf{y}^I, \mathbf{y}^J);\ I, J = 1,\cdots n$. For a new input vector $\mathbf{y}$, let $\mathbf{r}_i(\mathbf{y}) = [R_i(\mathbf{y}, \mathbf{y}^1)\cdots R_i(\mathbf{y}, \mathbf{y}^n)]^T$ represent the vector of correlations between the input and each of the elements of $\mathbf{Y}$. Then the median kriging predictions, corresponding to the best linear unbiased predictor (BLUP) (Lophaven et al. 2002b), is

$$\hat{\underline{z}}_i(\mathbf{y}) = \overline{\mathbf{f}}_i(\mathbf{y})^T \boldsymbol{\alpha}_i^* + \mathbf{r}_i(\mathbf{y})^T \boldsymbol{\beta}_i^*, \tag{4.14}$$

where

$$\boldsymbol{\alpha}_i^* = (\mathbf{F}_i^T \mathbf{R}_i^{-1} \mathbf{F}_i)^{-1} \mathbf{F}_i^T \mathbf{R}_i^{-1} \underline{\mathbf{Z}}_i, \quad \boldsymbol{\beta}_i^* = \mathbf{R}_i^{-1}(\underline{\mathbf{Z}}_i - \mathbf{F}_i \boldsymbol{\alpha}_i^*). \tag{4.15}$$

Note that for a given training set, $\boldsymbol{\alpha}_i^*$ and $\boldsymbol{\beta}_i^*$ are fixed; thus, to predict $\hat{z}_i(\mathbf{y})$ at a new input, one only needs to compute $\mathbf{f}_i(\mathbf{y})$ and $\mathbf{r}_i(\mathbf{y})$ and perform the matrix multiplications in the aforementioned equations. Additionally, all matrix inversions (that

could potentially increase computational complexity) are independent of $\mathbf{y}$ and need to be performed only once. Also, the process is solely based on matrix multiplications; thus, when the output needs to be calculated for multiple inputs $\{\mathbf{y}^j; j=1,\ldots,N\}$, as established for examples within a stochastic-simulation-based setting, the entire process can easily be extended to provide simultaneously (meaning here using a single matrix expression) the output $\hat{z}_i(\mathbf{y})$ for all different $\mathbf{y}^j$. In that case, the estimation is established by substituting vectors $\mathbf{f}_i(\mathbf{y})$ and $\mathbf{r}_i(\mathbf{y})$ with the matrices that have rows equal to $\mathbf{f}_k(\mathbf{y}^j)$ and $\mathbf{r}_k(\mathbf{y}^j)$, leading to

$$\begin{aligned} &\underline{\hat{\mathbf{z}}}\,(\mathbf{y}) = \tilde{\mathbf{f}}_k^{\,T}\boldsymbol{\alpha}_k^* + \tilde{\mathbf{r}}_k^{\,T}\boldsymbol{\beta}_k^*, \\ &\tilde{\mathbf{f}}_i = [\overline{\mathbf{f}}_i(\mathbf{y}^1)\ \ldots\ \overline{\mathbf{f}}_i(\mathbf{y}^N)], \quad \tilde{\mathbf{r}}_k = [\mathbf{r}_k(\mathbf{y}^1)\ \ldots\ \mathbf{r}_k(\mathbf{y}^N)]. \end{aligned} \tag{4.16}$$

The weight vector $\boldsymbol{\varphi}$, along with the process variance σ_i^2 in Eq. (4.11), can be obtained using the Maximum Likelihood Estimation (MLE) principle. The likelihood is defined here as the probability of the n observations given the weights $\boldsymbol{\varphi}$, and maximizing this likelihood with respect to $\boldsymbol{\varphi}$ ultimately corresponds to the optimization problem (Simpson et al. 2001c; Sacks et al. 1989)

$$\boldsymbol{\varphi}^* = \min_{\boldsymbol{\varphi}} \left[|\mathbf{R}_i|^{\frac{1}{n}} \tilde{\sigma}_i^2 \right] = \min_{\boldsymbol{\varphi}} \left[\ln(|\mathbf{R}_i|) + 2n \ln(\tilde{\sigma}_i) \right], \tag{4.17}$$

where |.| stands for determinant of a matrix and $\tilde{\sigma}_i^2$ is the least square estimate of the process variance given by (Sacks et al. 1989)

$$\tilde{\sigma}_i^2 = (\underline{\mathbf{Z}}_i - \mathbf{F}_i\boldsymbol{\alpha}_i^*)^T \mathbf{R}_i^{-1} (\underline{\mathbf{Z}}_i - \mathbf{F}_i\boldsymbol{\alpha}_i^*)/n. \tag{4.18}$$

This optimization corresponds to a multidimensional nonlinear minimization problem. However, it can be solved very efficiently using the modified Hooke and Jeeves method as in the popular DACE toolbox (Lophaven et al. 2002a).

For the type of application considered here, where higher accuracy is sought after at targeted regions of the input space, one could be tempted to formulate a different optimization problem to replace the one given by Eq.(4.17), trying to explicitly incorporate this goal in the selection of $\boldsymbol{\varphi}$, that is formulate an optimization problem for $\boldsymbol{\varphi}$ with higher weight for the error in such regions.

The computational burden that such an approach would involve [see for example the cost associated with optimization of moving least square response surface characteristics (Taflanidis et al. 2013; Loweth et al. 2011)] is generally prohibitive. The MLE approach, even though it targets global accuracy over the entire set of support points, provides a highly efficient tuning for the kriging parameters (optimization within a few minutes even for problems with many support points and numbers of basis functions), and as such, it is the one always preferred. Note that if a larger number of support points is available in the aforementioned targeted regions (which will be the case here through the DoE discussed in the next section), the MLE approach indirectly provides some greater preference toward reducing the error in these regions.

The prediction error variance for the normalized output, denoted $\hat{\phi}_i^2(\mathbf{y})$ for selection $\mathbf{y}$, is given by

$$\hat{\phi}_i^2(\mathbf{y}) = \tilde{\sigma}_i^2[1 + \mathbf{u}^T(\mathbf{F}_i^T\mathbf{R}_i^{-1}\mathbf{F}_i)^{-1}\mathbf{u} - \mathbf{r}_i(\mathbf{y})^T\mathbf{R}_i^{-1}\mathbf{r}_i(\mathbf{y})],$$
$$\text{where } \mathbf{u} = \mathbf{F}_i^T\mathbf{R}_i^{-1}\mathbf{r}_i(\mathbf{y}) - \mathbf{f}_i(\mathbf{y}). \tag{4.19}$$

Note that this variance is a function of the input vector $\mathbf{y}$ and not constant over the entire input space as is typically the case with most metamodeling techniques. This is why the term *local* is frequently used to describe the error variance estimate provided by kriging. Unfortunately, though, calculation of the estimate in Eq. (4.19) has an increased computational burden, typically much higher than the kriging prediction given by Eq. (4.14) (Lophaven et al. 2002a). This is the reason why explicit consideration of this error in the analysis is frequently ignored.

The performance of the metamodel can be validated directly by the process variance Eq. (4.18) which represents the mean squared error (Lophaven et al. 2002a), or by calculating different error statistics (such as coefficient of determination or mean error) using a leave-one-out cross-validation approach. This approach is established by removing sequentially each of the observations from the database, using the remaining support points to predict the output for that one and then evaluating the error between the predicted and real responses. The validation statistics are then obtained by averaging the errors established over all observations.

Eventually, the kriging approximation for the actual output can be established through the inverse of Eq. (4.8) using the statistics from Eq. (4.9) .

$$\hat{z}_i(\mathbf{y}) = \sigma_i^z\hat{\underline{z}}_i(\mathbf{y}) + \mu_i^z, \tag{4.20}$$

where the approximation for the normalized output $\hat{\underline{z}}_i(\mathbf{y})$ is given by Equation (4.14). Similarly, the prediction error variance may be obtained by scaling the normalized prediction error variance in Eq. (4.19) by $(\sigma_i^z)^2$,

$$\sigma^2_{\hat{\varepsilon}_i}(\mathbf{y}) = \hat{\phi}^2(\mathbf{y})(\sigma_i^z)^2. \tag{4.21}$$

Gradient information can be also easily derived by differentiating directly Eq. (4.14) and noting that Eqs. (4.15) are independent of $\mathbf{y}$. This leads to

$$\nabla\hat{\underline{\mathbf{z}}} = \mathbf{J}_f(\mathbf{y})^T\boldsymbol{\alpha}_i^* + \mathbf{J}_r(\mathbf{y})^T\boldsymbol{\beta}_i^*, \tag{4.22}$$

where $\mathbf{J}_f$ and $\mathbf{J}_r$ are the Jacobians of $\mathbf{f}_k$ and $\mathbf{r}_k$, respectively, with elements

$$\begin{aligned} J_f(\mathbf{y})_{lj} &= \frac{\partial f_{ij}(\mathbf{y})}{\partial y_l}, \qquad j = 1,...,p; l = 1,...,n_y, \\ J_r(\mathbf{y})_{lj} &= \frac{\partial R_i(\mathbf{y},\mathbf{y}^j)}{dy_l}, \quad j = 1,\cdots,n; l = 1,...,n_y, \end{aligned} \tag{4.23}$$

which do not depend on the output. Similarly, the following expression for the partial derivatives of the prediction error variance can be obtained by differentiating Eq. (4.19) (Lophaven et al. 2002a),

$$\nabla\hat{\phi}_i^2 = 2\tilde{\sigma}_i^2((\mathbf{F}_i^T\mathbf{R}_i^{-1}\mathbf{F}_i)(\mathbf{F}_i^T\mathbf{R}_i^{-1}\frac{\partial\mathbf{r}_i}{\partial y_j} - \mathbf{J}_f\boldsymbol{\beta}^*) - \mathbf{R}_i^{-1}\frac{\partial\mathbf{r}_i}{\partial y_j}), \quad j = 1,\cdots,p. \tag{4.24}$$

Note that both the gradients in Eqs. (4.22) and (4.24) need to be scaled by (σ_i^z) and $(\sigma_i^z)^2$, respectively, to obtain the information for the initial (not normalized) output.

Overall, some noteworthy aspects of the kriging metamodel are:

(a) The kriging predictions for any input $\mathbf{y}$ that belongs in the initial data-set $\mathbf{Y}$ can match the exact corresponding output; this behavior corresponds to an exact

interpolator (Simpson et al. 2001c). This is not necessarily true for all metamodeling approaches, as some have regression characteristics (response surfaces for example).

(b) Kriging provides the variance for the prediction error, which is also a function of the location $\mathbf{y}$, as opposed to other surrogate modeling techniques that either do not provide explicitly the variance for the error or keep it constant over the entire input space (Taflanidis et al. 2013).

(c) The kriging implementation involves only matrix manipulations with matrix inversions that need to be performed only once (for the definition in Eq. (4.15)).

(d) The optimization in Eq. (4.17) for the parameters that correspond to the correlation function can be performed highly efficiently.

(e) Selecting the polynomial order for the basis functions for the kriging model is not straightforward as there are no simplified approaches to accomplish this. In general, different basis function selections need to be considered and their accuracies compared (Jia and Taflanidis 2013). The latter can be obtained through a cross validation approach which is, in general, cumbersome.

(b) Kriging provides the variance for the prediction error which is also a function of the location $\mathbf{y}$, as opposed to other surrogate modeling techniques that either do not provide explicitly the variance for the error, or keep it constant over the entire input space (Taflanidis et al. 2013).

(c) The kriging implementation involves only matrix manipulations with matrix inversions that need to be performed only once (for the definition in Eq. (4.15)).

(d) The optimization in Eq. (4.17) for the parameters that correspond to the correlation function can be performed highly efficiently.

(e) Selecting the polynomial order for the basis functions for the kriging model is not straightforward as there are no simplified approaches to accomplish this. In general different basis function selections need to be considered and their accuracy compared (Jia and Taflanidis 2013). The latter can be obtained through a cross validation approach which is, in general, cumbersome.

4.2.2 Vector Case

Moving now to predictions for the entire vector $\mathbf{z}$, multiple metamodels can be considered, i.e. separate metamodels for each component (this ultimately corresponds to repeating the process for the scalar case multiple times), or a single metamodel can be implemented for all components. This corresponds to using a common choice of basis functions and correlation function for all outputs. In this case, which is the one adopted in the examples considered in this dissertation, the extension from the scalar case (repeating for each component and augmenting the results) is straightforward and yields

$$\begin{aligned} \hat{\underline{\mathbf{z}}}(\mathbf{y}) &= \overline{\mathbf{f}}(\mathbf{y})^T \boldsymbol{\alpha}^* + \mathbf{r}(\mathbf{y})^T \boldsymbol{\beta}^*, \\ \hat{\phi}_i^2(\mathbf{y}) &= \tilde{\sigma}_i^2 [1 + \mathbf{u}^T (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{u} - \mathbf{r}(\mathbf{y})^T \mathbf{R}^{-1} \mathbf{r}(\mathbf{y})]. \end{aligned} \tag{4.25}$$

with the matrices defined as

$$\begin{aligned} \boldsymbol{\alpha}^* &= (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{R}^{-1} \underline{\mathbf{Z}}, \\ \boldsymbol{\beta}^* &= \mathbf{R}^{-1} (\underline{\mathbf{Z}} - \mathbf{F} \boldsymbol{\alpha}^*), \\ \mathbf{u} &= \mathbf{F}^T \mathbf{R}^{-1} \mathbf{r}(\mathbf{y}) - \mathbf{f}(\mathbf{y}). \end{aligned} \tag{4.26}$$

The only difference from the scalar case is replacing $\underline{\mathbf{Z}}_i$ with $\underline{\mathbf{Z}} \in \mathbb{R}^{n \times n_z}$, where $\underline{\mathbf{Z}} = [\underline{\mathbf{Z}}_1 \cdots \underline{\mathbf{Z}}_{n_z}]$, and the optimal weights can be provided by the minimization problem (Lophaven et al. 2002b)

$$\boldsymbol{\varphi}^* = \min_{\boldsymbol{\varphi}} \left[|\mathbf{R}|^{\frac{1}{n}} \tilde{\sigma}^2 \right] = \min_{\boldsymbol{\varphi}} \left[\ln(|\mathbf{R}|) + 2n \ln(\tilde{\sigma}) \right], \tag{4.27}$$

with $\tilde{\sigma}^2 = \sum_{i=1}^{n_z} \tilde{\sigma}_i^2$ and $\tilde{\sigma}_i^2 = (\underline{\mathbf{Z}}_i - \mathbf{F}\boldsymbol{\alpha}_i^*)^T \mathbf{R}^{-1} (\underline{\mathbf{Z}}_i - \mathbf{F}\boldsymbol{\alpha}_i^*) / n$. Compared to the optimization for the scalar case, the optimization for the vector case requires evaluation of $\tilde{\sigma}_i^2$ for each of the n_z dimensions. Apart from this characteristic, the rest of the approach is similar to the scalar case and can also be performed using the open source DACE toolbox (Lophaven et al. 2002b).

The required gradients can be expressed as

$$\nabla \hat{\underline{\mathbf{z}}} = \mathbf{J}_f(\mathbf{y})^T \boldsymbol{\alpha}^* + \mathbf{J}_r(\mathbf{y})^T \boldsymbol{\beta}^*, \tag{4.28}$$

$$\nabla \hat{\phi}_i^2 = 2\tilde{\sigma}_i^2 ((\mathbf{F}_i^T \mathbf{R}_i^{-1} \mathbf{F}_i)(\mathbf{F}_i^T \mathbf{R}_i^{-1} \nabla \mathbf{r}_i - \mathbf{J}_f \boldsymbol{\beta}^*) - \mathbf{R}_i^{-1} \nabla \mathbf{r}_i), \tag{4.29}$$

where the Jacobians are the same as provided in Eq. (4.23) since they do not depend on the system output information.

4.3 Adaptive Kriging Formulation

In the proposed framework, a kriging approximation is developed by sharing information across the iterations of the optimization algorithm described through Eq. (4.1). This approximation is established in the augmented input space and is implemented within a stochastic simulation setting to approximate the system response

for different values of the model parameters for specific values of the design variables, as needed for estimates Eqs. (4.5) and (4.6). Within such a setting the focus is here on the adaptive DoE as well as the adaptive selection of the polynomial order of basis functions.

At iteration k let $\mathbf{x}_k$ denote the design variable vector that has been identified at the end of the previous iteration of the numerical optimization Eq. (4.1). Evaluation of the approximation to the system performance will be also available for $\mathbf{x}_k$, $\{\hat{h}(\mathbf{x}_k, \boldsymbol{\theta}^j); j = 1, ..., N\}$ for the sample set $\{\boldsymbol{\theta}^j\}_k$ used to estimate the objective function through Eq. (4.5). A localized box-bounded design subdomain X_k (also known as trust region) is defined; this domain is centered on $\mathbf{x}_k$ and has an appropriate length for each design variable (defining the length vector $\mathbf{L}_k$) that ultimately prescribes the upper and lower bounds for the design vector, $\mathbf{x}_k^l$ and $\mathbf{x}_k^u$, respectively. Any appropriate technique may be adopted for selecting the length vector $\mathbf{L}_k$ (Rodrigues et al. 1998; Rodríguez et al. 2000). A relevant recommendation for this is that the length is gradually reduced as iterations progress to regions closer to the minima, where one needs higher accuracy approximations (Thomas et al. 1992). A kriging metamodel is then established within subdomain X_k for the augmented input vector $\mathbf{y}$.

4.3.1 Design of Experiments (DoE)

A critical issue for the kriging implementation is the DoE to obtain the n observations (support points). As discussed in Chapter 1, space filling techniques or

adaptive design of experiments are commonly preferred for this task. However, the former may not provide the necessary accuracy in regions of importance, while the latter may significantly increase the computational cost. Therefore, hybrid DoE is proposed in this investigation instead.

Due to their distinct nature, the two different components of the input vector $\mathbf{y}$ have different characteristics/demands related to their accuracies. For instance, in the case of $\mathbf{x}$, accurate approximations are needed within the entire domain X_k since the metamodel is ultimately used to compare different design choices within this entire domain to converge to the optimal design configuration. This indicates that a space filling technique should be considered, and Latin Hypercube (LHS) sampling is adopted here for this purpose (Wang and Shan 2007; Simpson et al. 2001a). This formulation partitions the domain of interest into equal intervals that are uniformly distributed, and one sample is obtained per interval. In higher dimensions, these samples are randomly permuted and paired to preserve the uniformity of the sampling density across the dimensions.

An accurate approximation is needed on the other hand, for $\boldsymbol{\theta}$ over the domain in the uncertain model parameters space Θ that provides higher contribution to the integrand in the evaluation of the objective function (4.3). Thus, a target region DoE is needed for $\boldsymbol{\theta}$. Adaptive DoE has been proposed in the literature for this purpose (Picheny et al. 2010), but this approach has an added computational overhead (to identify these regions) and as such, is not advocated for the implementation considered here. Instead, the target region is approximated to correspond to the important region

for the integrand for $\mathbf{x}_k$ that is proportional to the density $\pi(\boldsymbol{\theta}|\mathbf{x}_k)$, which is given by Eq. (2.2). Since this requires knowledge of $h(\mathbf{x}_k,\boldsymbol{\theta})$, an approximation is established considering the density $\hat{\pi}(\boldsymbol{\theta}|\mathbf{x}_k)$, for which $h(\mathbf{x}_k,\boldsymbol{\theta})$ is replaced by the kriging prediction $\hat{h}(\mathbf{x}_k,\boldsymbol{\theta})$,

$$\hat{\pi}(\boldsymbol{\theta}|\mathbf{x}) \propto |\hat{h}(\mathbf{x},\boldsymbol{\theta})|\, p(\boldsymbol{\theta}). \tag{4.30}$$

This density can then be readily approximated based on a sample set, denoted $\{\boldsymbol{\theta}^a\}_k$, that can be obtained utilizing the readily available evaluations of $\hat{h}(\mathbf{x}_k,\boldsymbol{\theta}^j)$ for the sample set $\{\boldsymbol{\theta}^j\}_k$ (established in the previous iteration) from density $f_k(\boldsymbol{\theta})$. Such samples can be obtained through the rejection sampling approach discussed in Section 2.4.2. This is established by accepting the samples for which the following holds.

$$\frac{|\hat{h}(\mathbf{x}_k,\boldsymbol{\theta}^j)|\, p(\boldsymbol{\theta}^j)}{u^j f_k(\boldsymbol{\theta}^j)} > \max_j \left[|\hat{h}(\mathbf{x}_k,\boldsymbol{\theta}^j)| \frac{p(\boldsymbol{\theta}^j)}{f_k(\boldsymbol{\theta}^j)} \right]. \tag{4.31}$$

The sample set $\{\boldsymbol{\theta}^a\}_k$ ultimately represents the region in the Θ space that contributes more to the probabilistic performance for $\mathbf{x}_k$ and as such, corresponds to a good approximation (especially if X_k is not overly large so that behavior with respect to for $\mathbf{x}_k$ is an adequate representation for the entire domain) for the target region in which higher accuracy is sought after in the kriging metamodel. Any sample-based density approximation approach, for example the KDE discussed in Section 2.4.3 or an alternative parametric approach (also briefly discussed in that section) such as a

Gaussian mixture (Scott 1992), can be utilized for this purpose. This density will be denoted $f_k^s(\boldsymbol{\theta})$ herein.

Because of the importance of this approximation and ultimately of the number of samples in the set $\{\boldsymbol{\theta}^a\}_k$ for providing sufficient information for this approximation, a modification is further introduced to provide a sufficient number of available samples. Upon convergence to $\mathbf{x}_k$, utilizing the stochastic simulation estimates in Eqs. (4.5) and (4.6), an additional sample set beyond the N samples in $\{\boldsymbol{\theta}^j\}_k$ is generated to obtain a large sample set consisting of N_p samples for which $\hat{h}(\mathbf{x}_k,\boldsymbol{\theta})$ is evaluated. The rejection sampling in Eq. (4.31) is then performed over this sample set. Given that evaluation of $\hat{h}(\mathbf{x}_k,\boldsymbol{\theta})$ involves a small computational effort, this requirement may be easily satisfied. Additionally, a relaxation approach such as the one presented in Eq. (3.9) could have been considered.

Finally, it is important to consider that the kriging metamodel needs to have sufficient accuracy even in regions beyond this specific target region since erroneous approximations in such regions can impact the estimation result (these regions may become erroneously important because of such errors). This consideration leads to the following two-stage *hybrid DoE,* with the first stage aiming to obtain satisfactory global accuracy in the broader domain Θ of importance and the second stage aiming to obtain higher accuracy in the target region; initially (first stage) n_1^s samples are obtained, adopting a space-filling approach within the domain of importance based on $p(\boldsymbol{\theta})$ (for example, 4-5 standard deviations away from the median values for each model

parameters). Then additional n_2^s are obtained from a density approximation based on samples set $\{\boldsymbol{\theta}^a\}_k$. The total number of support points is then $n = n_1^s + n_2^s$.

4.3.2 Selection of Basis Functions' Order

Another important question for the kriging approximation is the order of the polynomial basis functions for each component of the input vector $\mathbf{y}$. Selecting the same higher order (for example second order) for all components might reduce the accuracy of the kriging metamodel; ultimately, components that exhibit higher sensitivity should have higher order polynomials associated with them (Jia and Taflanidis 2013). The optimization to identify the best basis function selection is a challenging task as discussed earlier. This challenge is circumvented herein by integrating the global sensitivity analysis discussed in Section 2.4.4 and selecting second order polynomial functions only for the most important components (linear polynomial functions are considered for the rest).

For the augmented vector $\mathbf{y}$, this sensitivity analysis is established in the following way. As soon as the evaluations of the system performance for the support points are computed, the auxiliary density function

$$\pi(\mathbf{y}) = \pi(\mathbf{x},\boldsymbol{\theta}) \propto |\, h(\mathbf{x},\boldsymbol{\theta}) \,|\, p(\boldsymbol{\theta}) p(\mathbf{x}) \propto |\, h(\mathbf{x},\boldsymbol{\theta}) \,|\, p(\boldsymbol{\theta}), \tag{4.32}$$

is considered (recall $p(\mathbf{x})$ is simply a uniform distribution within X_k), and samples from it are obtained through rejection sampling. The n_2^s support points available from the second stage of the hybrid DoE are considered for this purpose within the rejection sampling approach discussed in Section 2.4.2. Recall that these points

follow a uniform distribution for $\mathbf{x}$ and distribution $f_k^s(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$. Thus, the samples from the target distribution in Eq. (4.32) correspond simply to samples for which the following inequality holds,

$$\frac{|h(\mathbf{x}^j,\boldsymbol{\theta}^j)|\,p(\boldsymbol{\theta}^j)}{u^j f_k^s(\boldsymbol{\theta}^j)} > \max_{j=1,\ldots,n_2^s}\left[|h(\mathbf{x}^j,\boldsymbol{\theta}^j)|\frac{p(\boldsymbol{\theta}^j)}{f_k^s(\boldsymbol{\theta}^j)}\right]. \tag{4.33}$$

This approach leads to total of N_s samples $\{\mathbf{y}^s\}_k$ and to KDE approximation for the density in Eq. (4.32),

$$\tilde{\pi}(\mathrm{y}_i) = \frac{1}{N_s}\sum_{s=1}^{N_s}\left(\frac{1}{t_i}\left[K\left(\frac{y_i - y_i^s}{t_i}\right)\right]\right), \tag{4.34}$$

with bandwidth $t_i = 1.06\cdot N_s^{-1/5}\sigma_i$ with σ_i corresponding to the standard deviation of the samples $\{y_i^s\}$. The relative information entropy $D(\pi(y_i)\,||\,p(y_i))$ can be approximated through the numerical integration of Eq. (2.24) and used to rank the different components of the input vector according to their importance.

A threshold $D_{\min}^{re}$ can then be set to determine the importance of the input vector components. If the value of relative entropy is larger than this threshold, that particular parameter is assigned a higher order basis function (e.g. quadratic). In contrast, if the relative entropy value is less that the threshold, the parameter is assigned a lower order basis function (e.g. linear). Note that similarly to the approach discussed in Section 3.2.3, the aforementioned threshold can be adaptively selected to correspond to a fraction of the highest relative entropy value. If the allowable percentage reduction of the maximum entropy among the entire input vector is $s_e^{re} < 1$,

$$D_{\min}^{re} = s_e^{re} \max_i \left[D\left(\pi(y_i) \,\|\, p(y_i)\right)\right], \tag{4.35}$$

and this formulation ultimately leads to consideration of higher order basis functions for parameters that correspond to relative entropy values at least equal to s_e^{re} of the maximum entropy of the input vector.

4.4 Optimization under Uncertainty with Adaptive Kriging

4.4.1 Considerations for Implementation across Iterations

As discussed earlier in the proposed framework, the kriging approximation is developed by sharing information across the iterations of the optimization algorithm described through Eq. (4.1). The remaining questions for a successful implementation are the following:

a) How is the proposal density $f_k(\boldsymbol{\theta})$ for the estimations in Eqs. (4.5) and (4.6) established?

b) How is convergence evaluated?

c) What are the recommendations for the selection of the length vector $\mathbf{L}_k$ defining the trust region?

Starting with the proposal density $f_k(\boldsymbol{\theta})$, this density may be selected based on the information from the sample set $\{\boldsymbol{\theta}^a\}_k$ from distribution $\pi(\boldsymbol{\theta} \mid \mathbf{x}_k)$, which actually corresponds to the optimal IS for design configuration $\mathbf{x}_k$. This is similar to the concept advocated in (Cannamela et al. 2008) with the difference being that information readily

available is employed [set $\{\boldsymbol{\theta}^a\}_k$] rather than constructing a metamodel solely for the purpose of formulating IS densities. The density $\pi(\boldsymbol{\theta}|\mathbf{x}_k)$ is expected to provide a satisfactory accuracy for the entire domain X_k if, as discussed previously, $\mathbf{x}_k$ provides an adequate representation of the behavior of the integrand for different design configurations within X_k. Recall that exploiting the efficiently of the kriging metamodel, a large number of samples N can be used in this case for the stochastic-simulation-based evaluation of the objective function and its gradient, described in Eqs. (4.3) and (4.4), respectively. As such, no special attention needs to be placed on a highly efficient IS formulation; improvement in accuracy is primarily sought by adopting a larger number of N, though considerable advantages are also expected from the IS implementation. For example, the simple KDE approach presented in Section 2.4.3 can be implemented with no special attention given to the optimization of the bandwidth characteristics discussed in Chapter 3. Still, IS should be formulated only for the more important model parameters, and the global sensitivity analysis discussed in Section 2.4.4 can be incorporated for this purpose. Utilizing the readily available sample set $\{\boldsymbol{\theta}^a\}_k$, the entropy for each model parameter θ_i is estimated through Eq. (2.24) utilizing the approximation to $\pi(\boldsymbol{\theta}|\mathbf{x}_k)$ established through Eq. (2.23). The IS formulation is considered for only for the most important parameters, corresponding to entropy larger than some pre-defined allowable percentage reduction $s_e < 1$ of the maximum entropy among the entire group of model parameters,

$$D_{min} = s_e \max_i \left[D\left(\pi(\theta_i \mid \mathbf{x}_k) \,\|\, p(\theta_i)\right)\right]. \tag{4.36}$$

Ultimately, this implementation is identical to the one considered for selection of the order of the basis functions in Section 4.3.2 (same sample set and entropy calculation approach utilized) with the only difference being that only the $\boldsymbol{\theta}$ component of the kriging input vector $\mathbf{y}$ is considered ($\mathbf{x}$ is not included).

Moving now to the convergence of the algorithm, this is established when the new identified optimum x_k^* is a local optimum of the trust region X_k. To further improve the quality of the obtained solution, a second optimization stage is proposed; upon convergence, the number of support points is increased to establish a higher accuracy kriging metamodel, and the optimization described by Eq. (4.1) is repeated. This allows the use of a smaller number of support points $n = n_1^s + n_2^s$ in the initial iterations, until convergence is established. Ultimately, we are not concerned with obtaining high accuracy estimates for the kriging metamodel at the initial iterations; establishing an approximate descent direction in the design domain toward the optimal design is sufficient (greedy optimization approach).

Finally, with respect to the length vector selection $\mathbf{L}_k$, it can be initially considered as a specific fraction s_1^l of the design domain X, i.e. $\mathbf{L}_1 = s_1^l X$. At each iteration, a specific reduction s^r of this proportionality can be implemented, leading to $s_k^l = (s^r)^{k-1} s_1^l$ and $\mathbf{L}_k = s_k^l X$. Upon initial convergence, a further reduction by s_f^r can be established to localize the search around the candidate optimum.

Figure 4.2 provides an example of how the algorithm progresses through the design space. The squares are the trust regions X_k for each iteration. The gray dots show the intermediate steps needed to find a local optimum within the trust region (only using evaluations of the kriging model). The dash-dot line shows the second stage of the optimization that starts when the first stage has encountered an interior point local optimum. This stage has a significantly reduced length, and the number of support points within the domain is increased in order to increase the accuracy of the kriging model near the optimum point.

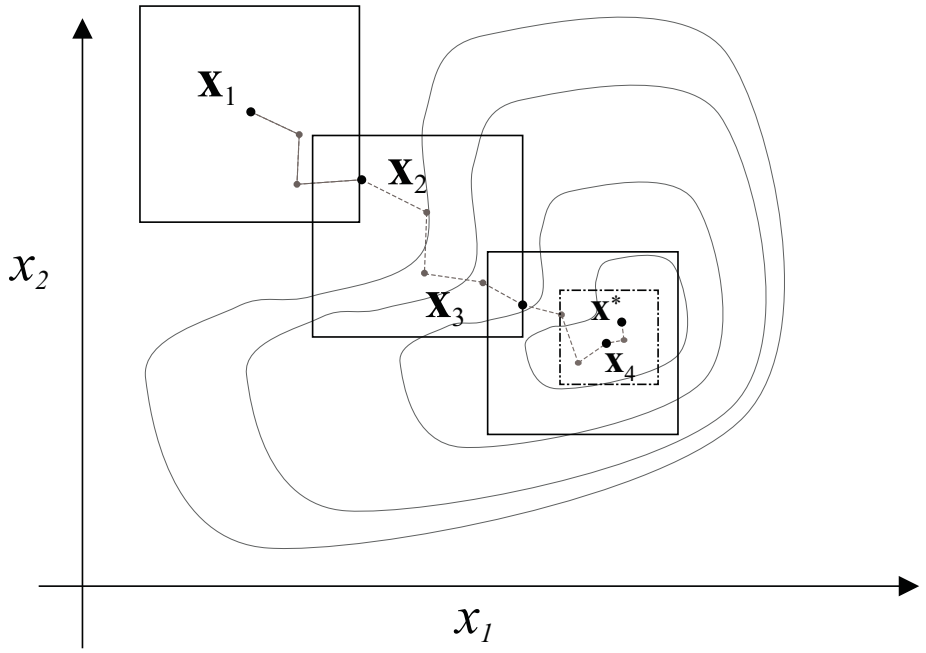


Figure 4.2: Illustration of the evolution of trust region

4.4.2 Algorithm for Adaptive Kriging Implementation

When combining the previous ideas, one can formulate the following optimization algorithm utilizing adaptive kriging. First, define the bounded design space

X, the starting point of the algorithm $\mathbf{x}_1$, the number of support points for the hybrid DoE approach n_1^s and n_2^s, respectively, including the number of support points for the second optimization stage n_1^f and n_2^f. Select the number of samples N for the estimation of the objective function and its gradient utilizing stochastic simulation, the number of samples N_p for which $\hat{h}(\mathbf{x}_k, \boldsymbol{\theta})$ will be obtained, the allowable percentage reduction $s_e < 1$ of the maximum entropy for the IS formulation, and the similar value for the basis function formulation $s_e^{re} < 1$. Finally, choose the fraction parameter s_1^l defining the initial trust region, its reduction s^r per iteration and the final reduction upon convergence s_f^r. A general flow diagram of the adaptive kriging algorithm can be observed in **Error! Reference source not found.**.

At iteration k of the numerical optimization algorithm [Eq. (4.1)] perform the following steps:

- Step 1 (trust region definition): Define box-bounded search domain X_k centered around $\mathbf{x}_k$ with length vector given by $\mathbf{L}_k = (s^r)^{k-1} s_1^l X$. If convergence has been established (last iteration), further reduce length vector by s_f^r. Adjust (truncate) trust region if it exceeds the design domain bounds X.

- Step 2 (support points): Employing the hybrid DoE for $\boldsymbol{\theta}$, obtain n_1^s (n_1^f if convergence has been established) samples using a space-filling approach (LHS) in the region of importance for $p(\boldsymbol{\theta})$, then obtain n_2^s (n_2^f if convergence has

been established) samples from density $f_k^s(\boldsymbol{\theta})$ [$p(\boldsymbol{\theta})$ in first iteration]. For $\mathbf{x}$, obtain $n = n_1^s + n_2^s$ ($n = n_1^f + n_2^f$ if convergence has been established) samples using a space filling approach (LHS) in X_k.

- Step 3 (evaluation of model response). For all the support points, evaluate the model response $\{\mathbf{z}(\mathbf{x}^j, \boldsymbol{\theta}^j); j = 1, ..., n\}$ and ultimately the system performance function $\{h(\mathbf{x}^j, \boldsymbol{\theta}^j); j = 1, ..., n\}$.
- Step 4 (selection of basis functions): Based on the evaluations of the performance function on the support points from the second stage $\{h(\mathbf{x}^j, \boldsymbol{\theta}^j); j = 1, ..., n_2^s\}$, obtain samples from $\pi(\mathbf{x}, \boldsymbol{\theta})$ through rejection sampling as in Eq. (4.33). Then calculate the entropy for each component of the output vector $D(\pi(y_i) \| p(y_i))$ using the approximation in Eq. (4.34) obtained through these samples. Consider higher order (quadratic) basis functions only for components of the input vector with relative entropy higher than the value given by Eq. (4.35) and lower order (linear) basis functions for the rest.
- Step 5 (kriging model): Employing the information in steps 1-4, build the kriging model in augmented input space through the approach discussed in Sections 4.2.1 (scalar case) and 4.2.2 (vector case).
- Step 6 (trust region local optimum): Simulate a set of N samples from distribution $f_k(\boldsymbol{\theta})$ [$p(\boldsymbol{\theta})$ in first iteration] and perform the optimization described by Eq. (4.1) utilizing estimations in Eq. (4.3) and (4.4) and employing a

gradient-based algorithm. Identify local optimum x_k^*. Note that this local optimization corresponds to exterior sampling but the overall approach [Eq. (4.1)]] to interior sampling.

- Step 7 (information for $\mathbf{x}_{k+1}$ and proposal density formulation for DoE): Consider $\mathbf{x}_{k+1} = \mathbf{x}_k^*$ and evaluate the response and the performance function through the kriging approximation for N_p samples. Obtain sample set $\{\boldsymbol{\theta}^a\}_{k+1}$ through Eq. (4.31) and establish $f_{k+1}^s(\boldsymbol{\theta})$ either through KDE or through parametric density estimation.
- Step 8 (IS proposal density for iteration $k+1$): Utilizing the same sample set $\{\boldsymbol{\theta}^a\}_{k+1}$, formulate, for example through KDE, the IS proposal density $f_{k+1}(\boldsymbol{\theta})$. Perform global sensitivity analysis by calculating the entropy through Eq. (2.24) using the available samples to establish the approximation given by Eq. (2.23). Consider formulation of IS only for model parameters with entropy values higher than in Eq. (4.36).
- Step 9 (convergence check): If $\mathbf{x}_{k+1}$ is on the boundary of X_k, convergence has not been established and proceed back to Step 1 and advance to $k+1$. If not, convergence has potentially been attained, and the second optimization stage needs to be implemented by repeating steps 1-6 with $n = n_1^f + n_2^f$ and s_f^r.

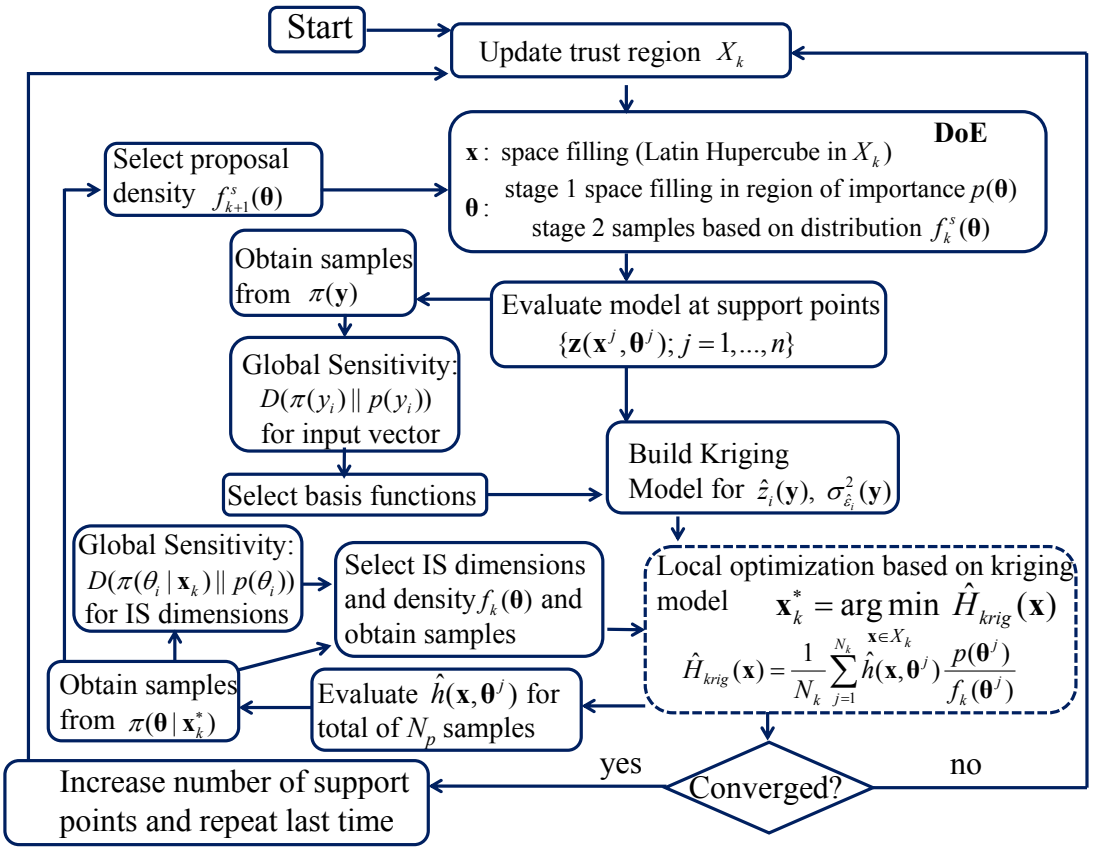


Figure 4.3 Flow diagram for adaptive kriging implementation in the augmented input space

4.5 Case Study: Half-Car Suspension Model Driving on a Rough Road

The framework is illustrated next with the same example considered in Chapter 3 (Section 3.4), where the optimization of the damper characteristics for the suspension of a half-car nonlinear model riding on a rough road is considered. The same numerical and probability models are considered here again (same as in Sections 3.4.1.1, 3.4.1.2, and 3.4.1.3), where two different design cases are examined: D_1 corresponding to a passive suspension and D_2 investigating a semi-active skyhook suspension implementation.

4.5.1 Transformation of Performance Function and Gradient Evaluation

Recall that the response quantities of interest z_{ri} for the half-car suspension model correspond to the Root Mean Square (RMS) of the vertical acceleration at the center of mass ($RMS_{\ddot{y}_c}$) and the suspension's damping forces at the front and rear of the car ($RMS_{F_{df}}$ and $RMS_{F_{dr}}$ respectively). Since the performance function formulation described by Eq. (3.37) involves the log values of these quantities, the kriging approximation is formulated here directly for these log values. As such, the response vector for the kriging implementation corresponds to

$$\mathbf{z}(\mathbf{x},\boldsymbol{\theta}) = \begin{bmatrix} \ln(RMS_{\ddot{y}_s}) \\ \ln(RMS_{F_{df}}) \\ \ln(RMS_{F_{dr}}) \end{bmatrix}. \tag{4.37}$$

Furthermore, the prediction error from the kriging metamodeling can be directly incorporated into the performance function definition. This is established by considering the transformation of the probability of exceedance for response quantity z_{ri}, described by Eq. (3.39), when the log of that quantity is expressed by the kriging relationships given by Eq. (4.7). This leads to

$$\begin{aligned} P[z_{ri} > b_i\varepsilon_i] &= P[\varepsilon_i \le z_{ri}/b_i] = P[\ln(\varepsilon_i) \le \ln(z_{ri}) - \ln(b_i)], \\ &= P[\ln(\varepsilon_i) \le \hat{z}_i + \hat{\varepsilon}_i - \ln(b_i)] = P[\ln(\varepsilon_i) - \hat{\varepsilon}_i \le \hat{z}_i - \ln(b_i)], \\ &= \Phi\left(\frac{\hat{z}_i - \ln(b_i)}{\sqrt{\sigma_{\hat{\varepsilon}_i}^2 + \sigma_{\varepsilon_i}^2}}\right), \end{aligned} \tag{4.38}$$

where $\hat{z}_i$ corresponds to the kriging approximation for $\ln(z_{ri})$ and the last equality is based on the fact that since $\ln(\varepsilon_i)$ and $\hat{\varepsilon}_i$ are zero mean independent Gaussian

variables with variances $\sigma_{\varepsilon_i}^2$ and $\sigma_{\hat{\varepsilon}_i}^2$, respectively, their sum (or difference in this case) is also a Gaussian variable with zero mean and variance $\sigma_{\varepsilon_i}^2 + \sigma_{\hat{\varepsilon}_i}^2$. This leads to the following approximation to the performance function,

$$\hat{h}(\mathbf{x},\boldsymbol{\theta}) = \frac{1}{3}\sum_{i=1}^{3}\Phi\left(\frac{\hat{z}_i - \ln(b_i)}{\sqrt{\sigma_{\hat{\varepsilon}_i}^2 + \sigma_{\varepsilon_i}^2}}\right), \tag{4.39}$$

which is obtained by using the kriging metamodel predictions while also incorporating its prediction error. The gradient of this expression is also needed and can be derived as

$$\begin{aligned}\nabla\hat{h}(\mathbf{x},\boldsymbol{\theta}) &= \frac{1}{3}\sum_{i=1}^{3}\nabla\Phi\left(\frac{\hat{z}_i - \ln(b_i)}{\sqrt{\sigma_{\hat{\varepsilon}_i}^2 + \sigma_{\varepsilon_i}^2}}\right), \\ &= \frac{1}{3}\sum_{i=1}^{3}\phi\left[\frac{\hat{z}_i - \log(b_i)}{\sqrt{\sigma_{\hat{\varepsilon}_i}^2 + \sigma_{\varepsilon_i}^2}}\right]\nabla\left(\frac{\hat{z}_i - \ln(b_i)}{\sqrt{\sigma_{\hat{\varepsilon}_i}^2 + \sigma_{\varepsilon_i}^2}}\right), \\ &= \frac{1}{3}\sum_{i=1}^{3}\phi\left[\frac{\hat{z}_i - \log(b_i)}{\sqrt{\sigma_{\hat{\varepsilon}_i}^2 + \sigma_{\varepsilon_i}^2}}\right]\left(\frac{\nabla\hat{z}_i}{\sqrt{\sigma_{\hat{\varepsilon}_i}^2 + \sigma_{\varepsilon_i}^2}} - \frac{(\hat{z}_i - \ln(b_i))}{(\sigma_{\hat{\varepsilon}_i}^2 + \sigma_{\varepsilon_i}^2)^{3/2}}\nabla\sigma_{\hat{\varepsilon}_i}^2\right),\end{aligned} \tag{4.40}$$

where $\phi[.]$ corresponds to the Gaussian Probability Density Function (PDF), which by definition is the derivative of the Gaussian CDF $\Phi[.]$, and all required gradients in the last equation were given in Section 4.2.1.

Ignoring the prediction error, the above two quantities are simplified to

$$\hat{h}(\mathbf{x},\boldsymbol{\theta}) = \frac{1}{3}\sum_{i=1}^{3}\Phi\left(\frac{\hat{z}_i - \ln(b_i)}{\sigma_{\varepsilon_i}}\right), \tag{4.41}$$

$$\nabla\hat{h}(\mathbf{x},\boldsymbol{\theta}) = \frac{1}{3}\sum_{i=1}^{3}\phi\left[\frac{\hat{z}_i - \log(b_i)}{\sigma_{\varepsilon_i}}\right]\frac{\nabla\hat{z}_i}{\sigma_{\varepsilon_i}}. \tag{4.42}$$

4.5.2 Optimization Details

The design domain X is selected similar to the ones in Section 3.4.2. For the passive design case, the design domain has upper bounds $[1000 \quad 1000]Ns/m$ and lower bounds $[0 \quad 0]Ns/m$; for the skyhook implementation, the upper bounds are $[400 \quad 4000 \quad 400 \quad 4000]Ns/m$ and the lower bounds are $[0 \quad 0 \quad 0 \quad 0]Ns/m$. For the trust region definition, the length of the initial region $\mathbf{L}_1$ is selected as 20% of the design domain X, i.e. $s_1^l = 0.2$ (recall $\mathbf{L}_1 = s_1^l X$), with a reduction in size of 5% with every iteration, i.e. $s^r = 0.95$. When the optimization has reached the last stage, the reduction in the trust region is set to 50% or $s_f^r = 0.5$. For the local search within X_k, the trust-region-reflective algorithm is adopted, employed through the MATLAB optimization toolbox (Coleman et al. 1999). Recall that the local search corresponds to an exterior sampling (same samples are utilized within the trust-region-reflective algorithm), and the overall approach corresponds to an interior sampling (different sample set is generated whenever the algorithm is initiated at the beginning of each local optimization). Note that the main contribution of this investigation resides in the increased efficiency for the estimation utilizing the adaptive kriging, or in other words, on the development of the kriging implementation used within the local search. Thus, a standard approach was adopted for the local optimization. The maximum number of overall iterations is set to 20, while the maximum number of iterations for the search is set to 30. These are hard convergence bounds that are only activated when divergence problems exist (and are introduced exactly for that purpose).

The number of support points for the first hybrid DoE stage is selected as $n_1^s = 200$ for both design cases, and the second hybrid DoE stage is selected as $n_2^s = 400$ for D_1 and $n_2^s = 700$ for D_2. The increased number of samples in D_2 was chosen to accommodate the larger number of design variables as well as the greater sensitivity of the response expected in the $\boldsymbol{\theta}$ domain since, as seen in previous Chapter, the optimum for that case corresponds to rare event simulation. The number of support points for the second optimization stage, for increasing the accuracy of the kriging approximation, is taken as double the ones for the initial iterations, i.e. $n_1^f = 400$ for both design cases, while $n_2^f = 800$ for D_1 and $n_2^f = 1400$ for D_2.

The number of samples for the estimates in Eqs. (4.5) and (4.6) for the local search is taken as $N = 2000$ for both the D_1 and D_2 problems and increased to $N = 10000$ in the second optimization stage. The total number of simulations for $\hat{h}(\mathbf{x}_k, \boldsymbol{\theta})$ to inform the selection of IS densities and the sampling density for the second stage of the DoE approach is set to $N_p = 10000$. For the selection of order of basis functions and of the dimensions to formulate IS densities, the identification of importance discussed earlier in Sections 4.3.2 and 4.4.1 is adopted. For the basis functions, the cutoff entropy with respect to the maximum entropy over the entire input vector is set to $s_e^{re} = 0.4$. It means that input parameters with relative entropy values less than 40% of the maximum entropy value are going to be assigned linear basis functions, while those with values greater than 40% of the maximum entropy are

assigned quadratic basis functions. The IS cutoff is set to $s_e = 0.3$, meaning that uncertain model parameters with relative entropy values less than 30% of the maximum entropy value are neglected for importance sampling. A parametric approach is employed to construct the IS density (due to its simplicity); a Gaussian density with mean and variance of the available samples is considered.

Apart from the fully adaptive kriging implementation [(a)adaptive selection of order of basis functions and (b) hybrid DoE] that additionally incorporates (c) the kriging error in the objective function formulation, three additional cases are examined, neglecting each time one of the aforementioned three components (a-c). This leads to a total of four different optimization approaches, and through comparison of each of the latter three to the first one can directly demonstrate some of the advantages of the proposed adaptive kriging formulation. Approaches (i) denoting Adaptive Kriging (AK) , and (ii) denoting Adaptive Kriging with Error (AKE) correspond to the proposed algorithm that adaptively employs the probabilistic sensitivity analysis to select the order of the basis function for the kriging model and also employs the proposed hybrid DoE. The only difference is that approach (i) does not include the kriging error in the objective function while approach (ii) does. Approach (iii) denoting Quadratic Kriging (QK) uses quadratic basis functions for all the parameters of the kriging mode, i.e. does not employ the probabilistic sensitivity analysis to select the order of basis function, but employs the hybrid DoE; the kriging error is also not included in the objective function for this case. Lastly, approach (iv) denoting Latin hypercube Kriging (LK), employs the traditional sampling technique of LHS for both $\mathbf{x}$ and $\boldsymbol{\theta}$ (i.e. does not differentiate in the

second optimization stage) and uses quadratic basis functions for all the parameters of the kriging mode without including the kriging error in the objective function. This last configuration is the baseline case, where none of the proposed advances are employed.

The optimization is performed 5 different times for each design case starting from a different initial design $\mathbf{x}_1$; the initial points for the different trials are reported in Table 4.2.

TABLE 4.2

INITIAL DESIGNS FOR THE DIFFERENT OPTIMIZATION TRIALS

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
D_1, $\mathbf{x}_1$	[100,100]	[100,900]	[900,900]	[900,100]	[100,500]
D_2, $\mathbf{x}_1$	[50,200, 50,200]	[350,3000, 350,3000]	[350,200, 350,200]	[50,3000 50,3000]	[50,200, 350,3000]

4.5.3 Results and Discussion

To facilitate a better understanding of the different optimization approaches in terms of the kriging implementation, initially a cross-validation is discussed extending to the kriging metamodel itself (approximation of the system responses) as well as to the estimate of the objective function. After that preliminary comparison, the results of the optimization are discussed in detail. The cross-validation is used to provide insight for this comparison.

4.5.3.1 Cross-Validation of Kriging Metamodel

The cross-validation of the metamodel is examined within the context in which it is used for the numerical optimization; the support points are obtained in the augmented input space, then the kriging metamodel is established and finally it is used to estimate the objective function for specific values of the design variables through the stochastic simulation approach given by Eq. (4.5). The design case D_2 is examined as it corresponds to a more challenging implementation (higher sensitivity in the model parameter space), and as such, it can demonstrate more clearly specific patterns and trends. The domain for the design variables has lower and upper bounds, $[120,1400,120,1400]\,Ns/m$ and $[200,2200,200,2200]\,Ns/m$, respectively.

A total of 900 support points are obtained using the two-stage DoE, and the kriging metamodel is then optimized for the system response following the procedure outlined in Section 4.2. The cross-validation is established with respect to (i) the predicted response (for each component z_i) as well as (ii) the resultant performance function. A leave-one-out cross validation is adopted. This is established by choosing a set of validation points corresponding to input $\{\mathbf{y}^{jp}; jp=1,\cdots,N_E\}$ and output $\{\mathbf{z}(\mathbf{y}^{jp}); jp=1,\cdots,N_E\}$, then removing sequentially each of these points from the observation set, and using the remaining support points to *predict* the output for that one left out in order to evaluate the error between the *predicted* and real values. The validation statistics are then obtained by averaging the errors established over all N_E observations. For scalar outputs, a common measure of the fit is the coefficient of

determination R_{zi}^2, corresponding to a normalized squared error metric, which for response quantity $z_i(\mathbf{y})$ is given by

$$R_{z_i}^2 = 1 - \frac{SSE_{z_i}}{SST_{z_i}},$$
$$SSE_{z_i} = \sum_{jp=1}^{N_E} \left(z_i(\mathbf{y}^{jp}) - \hat{z}_i(\mathbf{y}^{jp}) \right)^2, \tag{4.43}$$
$$SST_{z_i} = \sum_{jp=1}^{N_E} \left(z_i(\mathbf{y}^{jp}) - \frac{1}{N_E} \sum_{p=1}^{N_E} z_i(\mathbf{y}^{p}) \right)^2.$$

A similar approach holds for the performance function with $z_i(\mathbf{y}^{jp})$ replaced by $h(\mathrm{x}^{jp}, \boldsymbol{\theta}^{jp})$ and $\hat{z}_i(\mathbf{y}^{jp})$ by $\hat{h}(\mathrm{x}^{jp}, \boldsymbol{\theta}^{jp})$. In this case, the coefficient of determination is denoted by R_h^2. This coefficient of determination expresses the portion of the variability in the data (validation points) that can be explained by the kriging metamodel with values close to 1 corresponding to high accuracy.

Results are presented in Table 4.3 for both $R_{z_i}^2$ and R_h^2 using a set of 500 validation points. All three different kriging implementation approaches discussed in the previous subsection are considered (note that in this case considering or not a prediction error has no impact, which excludes that case).

TABLE 4.3

CROSS-VALIDATION ERROR STATISTICS FOR THE DIFFERENT KRIGING CASES

Kriging approach	$R^2_{z_i}$	R^2_h
Kriging with hybrid DoE and entropy selected basis function (AK)	[0.8745 0.9942 0.9802]	0.88293
Kriging with hybrid DoE and all quadratic basis function (QK)	[0.8808 0.9944 0.9819]	0.86099
Kriging with LHC and all quadratic basis functions (LK)	[0.9696 0.9465 0.9493]	0.93568

The results (compare first two rows in the table) indicate that the adaptive selection of basis functions provides a marginally worse accuracy with respect to the system response, but is better with respect to the performance function. This should be of no surprise; the adaptive selection is based on information from the performance function (sensitivity of the integrand that involves that function) and not the system response, and as such, it is expected to provide primarily an improved accuracy for that. What is surprising at least at a first glance, is the significantly improved accuracy when the hybrid DoE is not adopted, rather the space filling approach (LHC) is utilized, indicating a preference for the latter (LHC) and an inability of the proposed DoE to enhance the accuracy of the metamodel. This trend is again easy to explain; the adopted accuracy measure, the coefficient of determination, assesses the quality of the metamodel approximation over the support points and not over the target region of

importance, which is ultimately what we are interested in and the motivation for introducing the hybrid DoE in the first place. The comparison between the AK and LK approaches is established over different sample sets; the uniformity of the support points for LK contributes ultimately to better performance for the coefficient of determination over these support points, but this does not necessarily mean better accuracy in estimating the objective function. A weighted squared error could have been utilized for this purpose (Picheny et al. 2010) to evaluate the goodness of fit over the targeted region rather than globally over the domain of the support points, but the calculation of this error typically involves an additional computational burden, so this integration within the proposed framework is difficult to justify. Finally, it is also clear that in the targeted region, the log of the RMS acceleration of the center of mass is the response quantity that shows the greater variability and is ultimately the more challenging one to approximate for the kriging model; the coefficient of determination is much lower for that quantity when looking at the AK and QK approaches. This should be attributed to greater sensitivity of that acceleration to the dynamical system characteristics.

A second validation step is then performed by assessing the accuracy of the objective function estimate provided by Eq. (4.3) for the different approaches considered. This is established for a specific value of the design vector corresponding to $\mathbf{x}_t = [160, 1800, 160, 1800]$ (test design). Since AK and QK provide similar performance, the comparison is established only with respect to the AK and LK approaches. Additionally, for both of them, the inclusion or not of the prediction error is considered

when defining the performance function for the system model. The validation is performed by comparing against the estimate provided by the actual numerical model. For the stochastic simulation (to obtain the estimates for the objective function), a large sample set of $N = 10000$ samples is utilized following the guidelines presented in Section 3.4.3; the same samples (CRN) are used across all evaluations to enable a consistent comparison, and appropriate importance sampling densities are established to get a small coefficient of variation (< 4%) in the obtained estimates.

The results are reported in Table 4.4, and the goodness of fit is evaluated by comparing the estimate obtained from the kriging implementation to the estimates provided when the actual system model is utilized. Immediately, the superiority of the proposed DoE is evident. The LHS approach, despite offering a globally better approximation as indicated by the better coefficient of determination values, provides a significant overestimation of the objective function. This provides an important validation of the proposed adaptive DoE focusing on the targeted region (integrand of the probabilistic performance). Additionally, it is evident that inclusion of the prediction error from the kriging constitutes a significant improvement in accuracy. For the proposed implementation (compare AK and AKE) when the error is not included, a significant underestimation of the probabilistic performance is provided. This shows the importance, at least from an accuracy perspective, of including the kriging error in the estimation of the performance function, despite the fact that , as discussed earlier, estimation of the local variance for this error is a computationally more demanding task (at least when compared to the estimation of just the kriging mean prediction). Note

also that the error increases the estimated probabilistic performance; this trend is anticipated based on similar studies (Jia and Taflanidis 2013) since the performance corresponds to a rare event.

A final interesting comparison is the computational cost required for the different implementations; the $N = 10000$ evaluation of the actual model needs 1 Hr on an Intel Xeon CPU at 3.2 Ghz, whereas the kriging implementation requires 180 secs when no error is included, and 240 when the error is included. This directly shows the great computational savings accomplished in this example through the kriging implementation.

TABLE 4.4

PROBABILISTIC PERFORMANCE $\hat{H}(\mathbf{x}\,|\,\{\boldsymbol{\theta}^c\})$

Case	AK	AKE	QK	QKE	Actual system model
$\hat{H}(\mathbf{x}\,|\,\{\boldsymbol{\theta}^c\})$	0.000306	0.000742	0.000964	0.002265	0.000614

Statistics estimated through stochastic simulation, and obtained through the different kriging implementation approaches and the actual system model.

4.5.3.2 Optimization Results

The discussion moves now to the more interesting aspect, the optimization through the proposed kriging formulation. The results are reported in Table 4.5 for D_1 and Table 4.6 for D_2. Recall that the different trials in these tables correspond to different initial conditions for the algorithm. In particular, the optimal solution $\mathbf{x}^*$, the total number of simulations of the system model until convergence is established N_s^{Tot}, the objective function value obtained through the use of the kriging metamodel $\hat{H}_{krig}(\mathbf{x}^*)$, which is obtained directly from the optimization algorithm, and the objective function value obtained through the use of the actual system model $\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$ are reported. For the latter, and following the same approach as in Chapter 3, the same samples (common random numbers) are used across all comparisons for each design problem (but different for the skyhook and passive applications for the reasons discussed earlier) to enable a consistent comparison. The notation $\hat{H}(\mathbf{x} \mid \{\boldsymbol{\theta}^c\})$ is used to explicitly denote the dependence on this sample set $\{\boldsymbol{\theta}^c\}$. $N = 10000$ samples are used in this comparison, which facilitates a small coefficient of variation for $\hat{H}(\mathbf{x} \mid \{\boldsymbol{\theta}^c\})$, close to 2% for case D_1 and 4% for case D_2. A three-fold comparison can be established based on these results:

(a) Comparison between $\hat{H}_{krig}(\mathbf{x}^*)$ and $\hat{H}(\mathbf{x} \mid \{\boldsymbol{\theta}^c\})$ (the simplest one) shows the accuracy of the kriging implementation.

(b) Comparison of the N_s^{Tot} for different approaches shows the computational efficiency for convergence of the algorithm

(c) Comparison between $\hat{H}(\mathbf{x} \mid \{\boldsymbol{\theta}^c\})$ and the actual optimal solution of the problem shows the robustness of the approach in converging to the true optimum. Recall that the optimal solution reported in Section 3.4.2 for this problem, obtained through exterior sampling and using a costly global optimization approach, was found to be $\mathbf{x}^* = [713 \quad 606]\, Ns/m$ with respective performance $\hat{H}(\mathbf{x}^*) = 0.0104\ (1.4\%)$ for the passive implementation and $\mathbf{x}^* = [186 \quad 1651 \quad 197 \quad 1512]\, Ns/m$ with respective performance $\hat{H}(\mathbf{x}^*) = 0.00066\ (6.6\%)$ for the skyhook implementation. These are the benchmarks that the obtained solutions should be compared against.

TABLE 4.5

OPTIMIZATION RESULTS FOR THE DIFFERENT KRIGING CASES AND DESIGN PROBLEM D_1

Case (i) AK				
Trial	$\mathbf{x}^*$	N_s^{Tot}	$\hat{H}_{krig}(\mathbf{x}^*)$	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$
1	[690.38 633.95]	7302	0.0099	0.0106
2	[714.14 575.75]	8296	0.0108	0.0106
3	[706.4 626.61]	2721	0.0102	0.0105
4	[725.68 434.08]	2951	0.0133	0.0139
5	[719.1 616.94]	4227	0.0097	0.0105
Case (ii) AKE				
Trial	$\mathbf{x}^*$	N_s^{Tot}	$\hat{H}_{krig}(\mathbf{x}^*)$	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$
1	[700.97 614.99]	7402	0.0108	0.0105
2	[711.71 620.79]	7258	0.0109	0.0105
3	[689.94 649.38]	2728	0.0107	0.0107
4	[735.12 608.86]	6847	0.0113	0.0106
5	[704.11 629.68]	4134	0.0116	0.0105
Case (iii) QK				
Trial	$\mathbf{x}^*$	N_s^{Tot}	$\hat{H}_{krig}(\mathbf{x}^*)$	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$
1	[694.85 605.67]	5952	0.0098	0.0106
2	[694.86 631.7]	5489	0.0105	0.0106
3	[705.13 596.22]	2717	0.0106	0.0105
4	[712.88 500.24]	2959	0.0096	0.0119
5	[691.94 610.94]	4312	0.0094	0.0106
Case (iv) LK				
Trial	$\mathbf{x}^*$	N_s^{Tot}	$\hat{H}_{krig}(\mathbf{x}^*)$	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$
1	[928.93 700.96]	11096	0.0465	0.0257
2	[705.83 590.1]	11382	0.0140	0.0106
3	[908.74 994.6]	7990	0.0710	0.0528
4	[784.64 748.48]	6155	0.0210	0.0150
5	[399.66 900.13]	11202	0.0599	0.0569

TABLE 4.6

OPTIMIZATION RESULTS FOR THE DIFFERENT KRIGING CASES AND DESIGN PROBLEM D_2

Case (i) AK				
Trial	$\mathbf{x}^*$	N_s^{Tot}	$\hat{H}_{krig}(\mathbf{x}^*)$	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$
1	[122.58 2121.1 142.87 1884.5]	8817	0.000253	0.000795
2	[161.32 1484.9 145.74 1725.3]	6607	0.000232	0.000908
3	[149.47 1780.3 168.64 1568.5]	12046	0.000209	0.000733
4	[125.65 2183.2 113.17 1656.2]	3736	0.000200	0.001126
5	[159.14 1541 160.21 1532.8]	8698	0.000221	0.000894
Case (ii) AKE				
Trial	$\mathbf{x}^*$	N_s^{Tot}	$\hat{H}_{krig}(\mathbf{x}^*)$	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$
1	[136.98 2091.9 154.5 1814.3]	5721	0.000541	0.000657
2	[159.46 1819 189.88 1521.3]	4751	0.000529	0.000634
3	[158.65 1886.3 153.62 1762.9]	5512	0.000590	0.000633
4	[173.19 1731.5 147.58 1808.1]	7076	0.000449	0.000640
5	[171.33 1847.5 170.02 1708]	4703	0.000586	0.000624

TABLE 4.6 (CONTINUED)

Case (iii) QK				
Trial	$\mathbf{x}^*$	N_s^{Tot}	$\hat{H}_{krig}(\mathbf{x}^*)$	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$
1	[134.09 2075 140.58 1860]	10440	0.000255	0.000721
2	[138.28 2050.5 182.73 1390.7]	12912	0.000357	0.000710
3	[143.18 2027.5 140.02 1751.2]	9023	0.000294	0.000718
4	[125.13 2239.2 147.55 1746.4]	3705	0.000381	0.0007423
5	[137.57 2028.6 147.17 1691.6]	6558	0.000274	0.000743
Case (iv) LK				
Trial	$\mathbf{x}^*$	N_s^{Tot}	$\hat{H}_{krig}(\mathbf{x}^*)$	$\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$
1	[148.73 2649.4 195.09 1765.9]	19778	0.000241	0.001758
2	[220.35 2011.5 84.054 1860.4]	19720	0.001824	0.002798
3	[157.52 1838.1 145.37 2400.7]	19153	0.000024	0.000903
4	[189.96 1887.6 71.355 2174.7]	18763	0.000319	0.00119
5	[136.33 1937 306.38 859.23]	19591	0.001193	0.001908

The discussion will focus initially on the proposed adaptive formulation with integration of the prediction error (AKE). The results for AKE demonstrate a remarkable *computational efficiency* and *robustness* in converging to the optimal solution. The identified solution $\mathbf{x}^*$ is always in the vicinity of the benchmark optimum solution, and, more importantly the attained performance is always comparable to or even better

than the benchmark performance, and even to the performance reported in the adaptive IS implementation in Section 3.4.3. This is accomplished with a small number of model evaluations, not exceeding $7,500$ for any design case and trial. Note that the differences between these trials are well expected since there is a strong dependence of the optimization approach on the initial conditions (which is the point of difference in the different trials). Note that there is no dependence of the number of model evaluations needed for convergence on the dimension of the design vector n_x (actually convergence for case D_2 is established with a smaller average computational effort than case D_1). This is expected since the kriging is formulated in the augmented input space, which has similar dimensions for both design cases, and thus, there is no direct impact of n_x on the results. Overall, the reported efficiency corresponds to tremendous computational savings when compared to the approaches discussed in Section 3, something that is accomplished primarily through the proposed formulation of the kriging metamodel in the augmented input space. It should be stressed once again that there is a constraint on this approach since the dimension of the augmented input space cannot be excessively large.

Moving now to the accuracy of the kriging implementation (still for AKE) assessed through the comparison between $\hat{H}_{krig}(\mathbf{x}^*)$ and $\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$, there is an overall good agreement. The differences are significantly bigger for case D_2, as expected, since the skyhook implementation is characterized by greater sensitivity of the performance/response in the regions of the model parameters, something that

contributes to ultimately greater challenges for the kriging metamodel. When comparing the kringig accuracy when the prediction error is not included in the performance function estimate (comparing $\hat{H}_{krig}(\mathbf{x}^*)$ and $\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$ for AK) it is evident that the explicit consideration of that error provides significantly improved estimates, i.e. closer values of $\hat{H}_{krig}(\mathbf{x}^*)$ and $\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$ especially for design case D_2, which has the largest associated error and thus impact of that error on the predicted probabilistic performance. Finally, an under-prediction trend [$\hat{H}_{krig}(\mathbf{x}^*)$ smaller than $\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$] is evident in the results, especially when the error is not included in the probabilistic performance assessment. This is somewhat anticipated since kriging corresponds to an interpolation/smoothing scheme. Since the performance for both design cases corresponds to an example in which failure is not common, the smoothing of the response will contribute to estimates moving toward the non-failure region and ultimately to a reduced estimate of the fragility of those responses. The prediction error ultimately circumvents, at least partially, this challenge.

The more interesting comparison is, however, between AKE and the three alternative approaches (AK/QK/LK) in terms of computational efficiency (comparison of N_s^{Tot} for same trial) and more importantly, robustness (comparison of $\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$ for same trial). In all instances, it is shown that the other three approaches do not share the robustness of the proposed AKE implementation, as they converge for some trials to a suboptimal performance $\hat{H}(\mathbf{x}^* \mid \{\boldsymbol{\theta}^c\})$. The differences are perhaps more evident for LK

and secondary for AK. This is an important result; it shows that a space-filling DoE, even though it might provide a good global accuracy as reported earlier, leads to significant errors in regions of the model parameters that are of importance for the probabilistic performance and ultimately to erroneous identified optimal designs. Similarly, ignoring the prediction error not only decreases the accuracy of the estimated performance as shown already, but, and perhaps more importantly, can provide erroneous optimal solutions.

Overall, the study shows potential improved efficiency (reduction of computational burden) established through the proposed adaptive kriging scheme as it provides great computational savings and robustness for convergence to the optimal solution. Some of the key results in this illustrative example are the following:

- By considering a kriging implementation in the augmented input space, remarkable computational savings are established.
- The proposed hybrid DoE for a targeted region greatly enhances the accuracy of the kriging implementation and its ability to avoid converging to suboptimal solutions. This can be established simply by sharing information across the iterations of the optimization algorithm, with a small additional computational burden.
- Explicitly incorporating the prediction error not only improves the accuracy of the estimated objective function through the kriging metamodel but also supports a more robust optimization (avoid convergence to suboptimal solutions).

- The adaptive selection of basis functions provides a smaller advantage when compared to the other advances considered here.
- For the cross-validation of the kriging metamodel, there is a need for methods that, with small computational burden, provide weighted error statistics. Such weights need to stress the importance of getting accurate estimates in regions in the model parameter space that contribute more to the probabilistic performance.

4.6 Summary

In this chapter, adaptive implementation of kriging metamodeling was considered to reduce the computational burden associated with optimization under uncertainty problems adopting stochastic simulation. Two important aspects for tuning of the kriging metamodel were adaptively addressed within this implementation: design of experiments (DoE) for selecting support points aimed at improving the accuracy of a targeted region, the one contributing most to the probabilistic performance, and selection of the order of basis functions for the different inputs of the metamodel. Additionally, a novel implementation was introduced, formulating the kriging metamodel in the augmented model parameter and design variable space, and the prediction error from the kriging metamodel was explicitly considered within the evaluation of the probabilistic performance. Althrough kriging has been considered before for approximating the system response with respect to the uncertain model parameters within RBDO problems, no attention has been given in generalized

optimization under uncertainty problems or to the formulation of metamodels in an augmented input space. As such, the proposed enhancements constitute significant advances in this field.

By considering the augmented space, higher efficiency can be achieved. The system response is simultaneously approximated with respect to both the model parameters and the design variables, where the latter are considered to belong to some trust region, and the same metamodel is used for all evaluations of the objective function (for specific design variable values) while the search is constrained within this trust region. Only new model evaluations are necessary when the search reaches the boundary of this region, whereas the overlapping support points (between the trust regions in subsequent iterations) are maintained to improve accuracy.

For the design of experiments, a hybrid approach was proposed for the model parameters, combining a space-filling first stage (Latin hypercube was considered for this purpose) and a second stage targeting the region that has larger contributions to the probabilistic performance. The latter was approximated through samples from the integrand (corresponding to this performance) obtained using readily available information from the previous iteration of the optimization algorithm (no additional computational burden involved). The selection of the order of the basis functions was also adaptively addressed by integrating the global sensitivity analysis introduced in Section 2.4.4. Higher order basis functions were considered only for the parameters that were identified to have higher importance, and the sensitivity analysis was integrated in the proposed framework, again utilizing readily available information.

The illustrative example showed the computational efficiency (convergence with small number of evaluations of the system model) as well as robustness (convergence to solutions that are close to the true optimum) established through the proposed adaptive kriging implementation, especially from the adaptive DoE. It was also shown that incorporating the kriging prediction error in the performance function estimate can greatly improve the accuracy of the probabilistic performance estimates as well as the robustness of the optimization algorithm itself.

CHAPTER 5:

PROBABILITY OF DOMINANCE AS ROBUSTNESS MEASURE FOR ASSESSING APPROPRIATENESS OF ROBUST DESIGN OPTIMIZATION SOLUTIONS

This chapter considers the appropriateness of different candidate designs within an RDO formulation. Toward this goal, a new robustness measure, termed probability of dominance, is introduced. Rather than utilizing resultant statistical measures (i.e. mean and standard deviation) or comparing each design to an established benchmark (utopia point), the probability of dominance compares simultaneously the entire candidate set and examines the favorability of the design for all different model parameter values. It is defined as the likelihood that the performance of a design will outperform all other designs in the set. A multi-stage implementation is also proposed that facilitates increased robustness in the selection by considering the comparison among smaller subsets within the initial larger set of candidate designs. The impact of prediction errors between the real system and the assumed numerical model is also addressed. This extends to proper modeling of the prediction error influence, including selection of a probability model for it, as well as its impact on the probability of dominance and on the RDO formulation. Two different error models are considered, corresponding to either additive or multiplicative influence of the prediction error, and comparisons are drawn

between them. The correlation between the errors for different design configurations is also explored.

The characteristics of the RDO formulation have been already introduced in Section 2.3. In the following section, the probability of dominance and the multi-stage formulation are discussed. Subsequently, the prediction error and its impact in both the probability of dominance and the RDO designs are examined. Lastly, two illustrative examples are presented. The first one considers the design of a tuned mass damper (TMD) for vibration mitigation under harmonic excitations and the second one a topology optimization problem of minimum compliance. Extensive comparisons are discussed in these two examples, with some emphasis on the influence of the prediction error on the robustness traits of the obtained solutions.

The nomenclature specific to this chapter is reviewed in Table 5.1.

TABLE 5.1

RELEVANT NOMENCLATURE FOR CHAPTER 5

Γ	Set of candidate designs	d	Dimension of subsets of Γ
λ	Decay rate for exponential correlation function	$h_s(\mathbf{x})$	Performance function for real model
$\mu(\mathbf{x})$	Mean for system model	$I(\Gamma_k, \boldsymbol{\theta})$	Indicator function for dominance of $\mathbf{x}_k$ against set Γ and given $\boldsymbol{\theta}$
$[\mu_o \ \ \sigma_o]$	Utopia point for system model	$\cdot_k$	(Subscript) characteristic associated with the k^{th} design
ν_i	Weight for defining norm for exponential correlation	M_{kl}^d	Margin of dominance for design $\mathbf{x}_k$ within set S_{kl}^d
ρ_{kj}	Correlation function for the prediction errors associated with designs $\mathbf{x}_k$ and $\mathbf{x}_j$	m	Number of designs in Γ
$\boldsymbol{\Sigma}_{kl}^d$	Covariance matrix for vector of the difference between error for k^{th} design and errors from each of the designs in S_{kl}^d	$P_D(\mathbf{x}_k \mid \Gamma)$	Probability of dominance of design $\mathbf{x}_k$ over set Γ
$\sigma(\mathbf{x})$	Standard deviation for system model	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	Probability of dominance calculated through MCS
σ_e	Standard deviation for prediction error	$p(e)$	Probability model for prediction error
ψ	Parameter associated with prediction error variance	S^d	Set of all possible d -dimensional subsets of Γ
e	Prediction error	S_k^d	Set of d -dimensional subsets of Γ including design $\mathbf{x}_k$
$\bar{e}$	Natural logarithm of prediction error	S_{kl}^d	l^{th} subset of d -dimensional subsets of Γ including design $\mathbf{x}_k$
D_t	Threshold for dominance	$\cdot_s$	(Subscript) value for real system
$D_d(\mathbf{x}_k)$	Degree of dominance of design $\mathbf{x}_i$ for d -dimensional subsets	w	Weight used in the RDO formulation
$Dh_{kl}^d(\boldsymbol{\theta})$	Vector of the difference between performance value for k^{th} design and performance for each of the designs in S_{kl}^d	$\mathbf{x}_j^{dlk}$	j^{th} design within the S_{kl}^d set

5.1 Probability of Dominance

5.1.1 Definition

Recall that RDO can lead to different candidate designs, establishing a different compromise between the different objectives (minimization of mean and standard deviation of response). Assume $\Gamma = \{\mathbf{x}_m;\ d = 1, \cdots, n_m\}$ to correspond to a set of such candidate designs and consider the task of quantifying the appropriateness of these designs. The explicit notation * to denote optimal design for $\mathbf{x}$ is not used hereafter since the method can in principal be used for any set of candidate designs.

A new measure, termed probability of dominance, is introduced for assessment of this appropriateness. This robust measure is defined as the likelihood of candidate design $\mathbf{x}_k$ to outperform its rival designs within Γ under probability model $p(\boldsymbol{\theta})$ for the uncertain model parameters. It should be noted that this definition is independent of the design formulation for obtaining the candidate design configurations, i.e. it can be extended to other types of optimization under uncertainty applications beyond RDO problems. Based on the total probability theorem, this probability of dominance P_D is given by

$$P_D(\mathbf{x}_k \mid \Gamma) = \int_{\Theta} P_D(\mathbf{x}_k \mid \Gamma, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{5.1}$$

where $P_D(\mathbf{x}_k \mid \Gamma, \boldsymbol{\theta})$ stands for the probability of dominance given the model configuration $\boldsymbol{\theta}$ and is equal to the probability that the k^{th} design will outperform all other designs. Dominance is, thus, probabilistically attained when a particular design $\mathbf{x}_k$

has a better performance $h(\mathbf{x},\boldsymbol{\theta})$ relative to other designs under a specific realization of the uncertain parameters $\boldsymbol{\theta}$.

Assuming that a smaller performance is preferable, we have

$$P_D(\mathbf{x}_k \mid \Gamma,\boldsymbol{\theta}) = P\left[\bigcap_{m=1,\, m\neq k}^{n_m} \left(h(\mathbf{x}_k,\boldsymbol{\theta}) < h(\mathbf{x}_m,\boldsymbol{\theta})\right)\right]. \tag{5.2}$$

In the case that no prediction error is assumed for the model (the case with prediction error will be addressed later), this is also given by

$$P\left[\bigcap_{m=1,\, m\neq k}^{n_m} h(\mathbf{x}_k,\boldsymbol{\theta}) < h(\mathbf{x}_m,\boldsymbol{\theta})\right] = P\left[h(\mathbf{x}_k,\boldsymbol{\theta}) = \min_{m=1,\ldots,n_m}\left(h(\mathbf{x}_m,\boldsymbol{\theta})\right)\right]. \tag{5.3}$$

and it finally corresponds to the indicator function

$$P_D(\mathbf{x}_i \mid \Gamma,\boldsymbol{\theta}) = I(\Gamma_k,\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } h(\mathbf{x}_k,\boldsymbol{\theta}) = \min_{m=1,\ldots,n_m}\left(h(\mathbf{x}_m,\boldsymbol{\theta})\right) \\ 0 & \text{otherwise} \end{cases}, \tag{5.4}$$

with the notation $I(\Gamma_k,\boldsymbol{\theta})$ used to symbolize the indicator function for the dominance of design k among the designs in set Γ.

The high-dimensional probability integral in (5.1) can be readily calculated through stochastic (Monte Carlo) simulation (similar to the ideas discussed in Section 2.2 for the objective function), leading to

$$\hat{P}_D(\mathbf{x}_k \mid \Gamma) = \frac{1}{N}\sum_{j=1}^{N} P_D(\mathbf{x}_k \mid \Gamma,\boldsymbol{\theta}^j), \tag{5.5}$$

where the N samples for $\boldsymbol{\theta}$, with $\boldsymbol{\theta}^j$ denoting the j^{th} sample, are simulated from distribution $p(\boldsymbol{\theta})$. Since we are ultimately interested in the dominant design, corresponding to a high probability, it is anticipated that this approach will yield

satisfactorily results (adequate accuracy) without significant computational burden (necessity to select high values for N). The accuracy of the estimation can be assessed through the coefficient of variation of the estimator (Robert and Casella 2004).

5.1.2 Multistage Implementation

The probability of dominance can be viewed as a weighted measure of the hyper-volume of the region over Θ where a particular design has a performance that is the best among the candidate designs, the weight provided by the relative likelihood $p(\boldsymbol{\theta})$ of the model parameters $\boldsymbol{\theta}$ within Θ. The dominant design is selected as the one corresponding to the maximum $P_D(\mathbf{x}_k \mid \Gamma)$,

$$\mathbf{x}_D = \underset{k}{\operatorname{argmax}}\, P_D(\mathbf{x}_k \mid \Gamma). \tag{5.6}$$

An advantage of this measure is that it is independent of the design process, which allows application to any arbitrary group of designs. However, a caveat of the approach is the fact that the result is a function of the exact designs contained in the set Γ, with the final outcome potentially changing even when non-dominant designs are introduced (or removed) into the set of candidate designs. This can happen if the new design only dominates (that is, provides better performance) in regions of the uncertain space that the previous best design used to dominate, thus detracting from the favorability of the previously dominant design only (and not of any other designs). This may ultimately switch the preference without this new design emerging as the new dominant one.

To circumvent this challenge and provide enhanced robustness in the choice of a most appropriate design, a multi-stage formulation is developed. Rather than looking at the dominance over the entire set Γ, the dominance within d-dimensional subsets of Γ containing d total designs is considered (simplest choice for d is 2). A design is then termed as d-dominant if it is the dominant one within the n_d different d-dimensional subsets that include it. To formalize these ideas, we denote as S^d the set of d-dimensional subsets of d and as S_k^d the set of d-dimensional subsets including design $\mathbf{x}_k$ and use subscript l to distinguish between the different elements of these sets (so S_l^d denotes the l^{th} subset of Γ containing d designs). For instance, if $\Gamma = \{A, B, C, D\}$, then $S_C^2 = \{[C\ A], [C\ B], [C\ D]\}$ (three candidate sets including two designs that have C as a member), $S_C^3 = \{[C\ A\ B], [C\ A\ D]\}$ (two candidate sets including three designs that have C as a member) and $S^3 = \{[C\ A\ B], [C\ A\ D], [D\ A\ B],\ \}$ (three candidate sets including three total designs). The probability of dominance over subset S_{kl}^d is then

$$P_D(\mathbf{x}_l \mid S_{kl}^d) = \int_\Theta P_D(\mathbf{x}_k \mid S_{kl}^d, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int_\Theta P\left[\bigcap_{m=1, m\neq i}^{d} \left(h(\mathbf{x}_k, \boldsymbol{\theta}) < h(\mathbf{x}_m^{dlk}, \boldsymbol{\theta}) \right) \right] p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (5.7)$$

where $\mathbf{x}_m^{dlk}$ denotes the m^{th} design within the S_{kl}^d set. This probability can be evaluated through stochastic simulation as in Eq. (5.5), employing Eq.(5.4) to describe the probability function within the integrand. The transformed equations are

$$\hat{P}_D(\mathbf{x}_i \mid S_{kl}^d) = \frac{1}{N} \sum_{j=1}^{N} P_D(\mathbf{x}_k \mid S_{kl}^d, \boldsymbol{\theta}^j), \quad (5.8)$$

with

$$P_D(\mathbf{x}_k \mid S_{kl}^d, \boldsymbol{\theta}) = I(S_{kl}^d, \boldsymbol{\theta}) = \begin{cases} 1 & \text{if } h(\mathbf{x}_k, \boldsymbol{\theta}) = \min\limits_{m=1,\ldots,n_m} \left(h(\mathbf{x}_m^{dlk}, \boldsymbol{\theta}) \right) \\ 0 & \text{otherwise} \end{cases}. \tag{5.9}$$

The design $\mathbf{x}_k$ is dominant within the set S_{kl}^d if it has the largest probability of dominance within the set, and it is considered to be d -dominant if it dominates each subset within S_k^d. To quantify these notions, the margin of dominance between a particular design $\mathbf{x}_k$ and the rest of the designs within set S_{kl}^d is defined as

$$M_{kl}^d = P_D(\mathbf{x}_k \mid S_{kl}^d) - \max_{m=1,\ldots,d, m \neq i} P_D(\mathbf{x}_m \mid S_{kl}^d), \tag{5.10}$$

while the degree of dominance is defined as

$$D_d(\mathbf{x}_k) = \min_{l=1,\ldots,n_d} M_{kl}^d. \tag{5.11}$$

The margin of dominance is defined as the difference between the probability of dominance for $\mathbf{x}_k$ compared to the design among the remaining designs in S_{kl}^d that has the highest probability of dominance. Positive values for this margin correspond to dominance of $\mathbf{x}_k$ in subset S_{kl}^d. The degree of dominance corresponds to the minimum of the margin of dominance for $\mathbf{x}_k$ among all subsets that include it. A positive value for it corresponds to d -dimensional dominance for $\mathbf{x}_k$. Nonetheless, it can be the case that no design emerges as dominant. In the previously discussed example, this will happen for the two-dimensional ($d = 2$) comparisons if A dominates B and D, but is dominated by C, and C is dominated at least by B. If, on the other hand, A dominates B, C, and D (always within the context of individual comparisons), then A is the dominant design for the $d = 2$ dimensional subsets.

The overall preferred design should be taken as the one that is dominant within the lowest possible d -dimensional subsets considered. This does not necessarily mean that the chosen design will be dominant for higher values of d -dimensional subsets, but it exhibits the most robust dominance characteristics for the set Γ . For example, assume that a candidate design exhibits $d = 2$ dimensional dominance; then if some other design is removed from Γ , the candidate design will still enjoy the dominance property. If an additional design is added, then if the older dominant design outperforms the newly added, it will also maintain its $d = 2$ dimensional dominance. Further robustness characteristics can be established by also considering the degree of dominance and establishing a minimum threshold D_t , i.e. design can be considered as dominant only if $D_d(\mathbf{x}_k) > D_t$.

Note that for all the computations required to identify the dominant design, the same sample set $\boldsymbol{\theta}^j$ can be used. As such, the performance function evaluations (which could be computationally intensive for complex models) used in the estimation of Eq. (5.8) can be the same ones used for Eq. (5.5); thus, only performance comparisons are needed (not re-computation of the performance). Furthermore, to reduce the overall burden, the results from the comparison within the set Γ can be used to guide the priority of comparisons; designs for which $P_D(\mathbf{x}_k \mid \Gamma)$ is small are not expected to dominate within any considered d -dimensional set and can be given lower priority in the comparisons.

Finally, the computational framework for identifying the dominant design within set $\Gamma = \{\mathbf{x}_k;\ i = 1,\cdots,n_m\}$ is the following:

Step 1: Generate a set of N samples $\{\boldsymbol{\theta}^j\}$ from $p(\boldsymbol{\theta}^j)$ and evaluate the performance for each one to obtain the set $\{h(\mathbf{x}_k,\boldsymbol{\theta}^j);\ k = 1,\cdots,n_m;\quad j = 1,\cdots,N\}$, if it is not readily available from the optimization stage.

Step 2: Calculate the probability of dominance for entire set Γ for each design $k = 1,\cdots,n_m$ by Eq. (5.8) utilizing the performance function evaluations from Step 1. Re-order the designs based on their dominance $P_D(\mathbf{x}_k \mid \Gamma)$ to obtain set $\{\underline{\mathbf{x}}_c;\ c = 1,\cdots,n_m\}$, and ignore any designs as deemed appropriate.

Step 3: Set $d = 2$ and start the comparison from the smallest possible subset class.

Step 4: Set $c = 1$ and start by considering the most dominant design based on the values of $P_D(\mathbf{x}_k \mid \Gamma)$.

Step 5: For all subsets in S_c^d, including $\underline{\mathbf{x}}_c$, estimate the probability of dominance for all designs included in it $\{P_D(\underline{\mathbf{x}}_m \mid S_{cl}^d); l = 1,...,n_m, m = 1,..,d\}$ using approximation Eq. (5.8), calculated through the readily available performance function evaluations from Step 1. For the first subset comparisons (when $d = 2$), this corresponds to comparison of a pair of designs, and the probability of dominances for the compared designs are complimentary.

Step 6: Calculate the margin of dominance with Eq. (5.10) and the degree of dominance with Eq. (5.11). If the latter is greater than zero (or the chosen threshold D_t), $\underline{\mathbf{x}}_c$

is the overall dominant design, and the process can stop. Else, set $c = c + 1$ and go back to Step 5, understanding that a lot of the necessary probabilities $P_D(\underline{\mathbf{x}}_m \mid S^d_{(c+1)l})$ have already been computed (when considering the smaller valued c's). Perform Step 5 for the new $\underline{\mathbf{x}}_c$ only if it could emerge as a dominant one, meaning if it has not yet been dominated by any of the previously considered designs.

Step 7: If all designs have been examined, i.e. $c = n_m$, and no dominant design has been identified for the *d*-dimensional subsets, set $d = d + 1$ and go back to Step 4.

Through this approach, the design exhibiting the lowest possible d -dimensional dominance is readily identified. As will be illustrated later, this frequently (but not necessarily always) corresponds to 2 -dimensional dominance and to a design having a large associated value of $P_D(\mathbf{x}_k \mid \Gamma)$. This design is guaranteed to outperform all other designs within the considered class of candidate designs and is the approach advocated here for selecting the preferred design among Γ . The degree of dominance $D_d(.)$ can be further used to evaluate the preference among different designs, even those that are non-dominant. The process will be illustrated in the examples considered later.

5.2 Impact of the prediction error on the Probability of Dominance

The previous section assumed no error between the response of the actual system, denoted $h_s(\mathbf{x})$ herein, and that of its assumed model, $h(\mathbf{x}, \boldsymbol{\theta})$ [in other words, it

was assumed that $h_s(\mathbf{x}) = h(\mathbf{x}, \boldsymbol{\theta})$]. The notation $h_s(\mathbf{x})$ is used to stress the fact that the real system is dependent on the exact design selection $\mathbf{x}$, and its performance is independent of the numerical model assumed and thus of $\boldsymbol{\theta}$ (although $\boldsymbol{\theta}$ is obviously related to the system properties). This assumption, i.e. $h_s(\mathbf{x}) = h(\mathbf{x}, \boldsymbol{\theta})$, is the reason why the probability $P_D(\mathbf{x}_k \mid S_{kl}^d, \boldsymbol{\theta})$ ultimately corresponds to the indicator function $I(S_{kl}^d, \boldsymbol{\theta})$ in the analysis discussed in Section 5.1.1.

In any engineering application, the adopted model cannot describe exactly the behavior of the actual system (Beck 2010); a prediction error always exists, and its incorporation in probabilistic analysis can have an impact on the design choices (Taflanidis and Beck 2010). This was illustrated in the definition of the probabilistic performance in the example considered in Chapters 3 and 4.

This error e will be taken here to directly quantify the difference between $h_s(\mathbf{x})$ and $h(\mathbf{x}, \boldsymbol{\theta})$ (note that in Chapters 3 and 4, the error in the system response vector was considered, and the error in $h(\mathbf{x}, \boldsymbol{\theta})$ needs to be generally addressed in this chapter). Since this error is unknown, it can be probabilistically described by assigning a probability density function $p(e)$ for it (Taflanidis and Beck 2010), treating it like any other uncertain model parameter. This prediction error can be viewed as describing (a) un-modeled characteristics of the system model or (b) un-modeled uncertainties for $\boldsymbol{\theta}$ (Beck and Taflanidis 2013; Beck 2010). In this section, the influence of this error on the probability of dominance, as well as on the actual formulation of the RDO problem, will be investigated. The two most common error models will be considered for the impact

of the error; these are either additive or multiplicative influence on $h(\mathbf{x},\boldsymbol{\theta})$. Furthermore, the assumption that e is independent of $\boldsymbol{\theta}$ will be utilized (this ultimately means independence between $p(e)$, or equivalently $h(\mathbf{x},\boldsymbol{\theta})$, and e). The latter assumption introduces the largest amount of uncertainty in the problem formulation (Jaynes 2003; Taflanidis and Beck 2010) and is the reason adopted here (establishing a greater robustness), as it assumes the least amount of information imposed upon the problem description.

5.2.1 Correlation of Errors between Designs

Before moving forward, the correlation between errors for different design selections needs to be addressed, as this correlation is important in assessing the probability $P_D(\mathbf{x}_k \mid S_{kl}^d,\boldsymbol{\theta})$. This correlation is expected to depend on the distance between $\mathbf{x}_k$ and $\mathbf{x}_m$; if the two designs are very similar, any prediction errors for the system numerical model are expected to be the same for both of them since the system models examined (under design configurations $\mathbf{x}_k$ and $\mathbf{x}_m$) are similar. Note that assuming no correlation introduces the largest amount of uncertainty in the problem formulation, but in this case, this is not a reasonable choice. In the example considered later, an exponential function is assumed for the correlation coefficient ρ_{km} between designs k and m given by

$$\rho_{km} = exp(-\lambda \parallel \mathbf{x}_k - \mathbf{x}_m \parallel_v^2) = exp\left[-\lambda \sum_{i=1}^{n_x} v_i^2 (x_{ki} - x_{mi})^2 \right], \tag{5.12}$$

where $\| \mathbf{x}_k - \mathbf{x}_m \|_v$ stands for the weighted Euclidean norm with v_i weight for each component of vector $\mathbf{x}$, while λ is the assumed decay rate of the exponential function. Exact selection of these terms is discussed in the examples later.

It should be stressed that, even though the use of exponential functions to describe correlation is very popular in engineering applications (Lophaven et al. 2002a), this choice is ultimately ad-hoc (especially due to the dependence on the chosen norm and decay rate λ), and if more information is available for the prediction error and its relationship to the design configuration, some different correlation structure can be adopted (Papadimitriou and Lombaert 2012) (note that this type of information is normally not available in most practical engineering problems).

For the analysis in this section, a correlation coefficient ρ_{km} between designs is assumed to be defined by the designer through any appropriate method chosen (Papadimitriou and Lombaert 2012), whereas in the examples, the impact of this correlation on the final design choice is examined using the correlation function in Eq.(5.12). Next, the different modeling choices for the prediction error and their impact on the probability of dominance are discussed.

5.2.2 Additive Prediction Error

The most commonly used assumption for the prediction error is an additive influence (this agrees with the assumption commonly used for Bayesian system identification applications (Beck and Taflanidis 2013)), leading to

$$h_s(\mathbf{x}) = h(\mathbf{x}, \boldsymbol{\theta}) + e. \tag{5.13}$$

The mean for the system model is updated as

$$\mu_s(\mathbf{x}) = E[h_s(\mathbf{x})] = E[h(\mathbf{x},\boldsymbol{\theta}) + e] = E[h(\mathbf{x},\boldsymbol{\theta})] + E[e] = E[h(\mathbf{x},\boldsymbol{\theta})] = \mu(\mathbf{x}), \quad (5.14)$$

where $E[e]$ is taken equal to zero in order to establish unbiased predictions. Similarly, the variance is transformed to

$$\begin{aligned} \sigma_s^2(\mathbf{x}) &= E[h_s^2(\mathbf{x})] - \mu_s^2(\mathbf{x}) = E[(h(\mathbf{x},\boldsymbol{\theta}) + e)^2] - \mu^2(\mathbf{x}), \\ &= E[h(\mathbf{x},\boldsymbol{\theta})^2] + 2E[h(\mathbf{x},\boldsymbol{\theta})\, e] + E[e^2] - \mu^2(\mathbf{x}), \\ &= (E[h(\mathbf{x},\boldsymbol{\theta})^2] - \mu^2(\mathbf{x})) + E[e^2] + 2E[h(\mathbf{x},\boldsymbol{\theta})\,]E[e], \\ &= \sigma^2(\mathbf{x}) + E[e^2] = \sigma^2(\mathbf{x}) + \sigma_e^2\,, \end{aligned} \quad (5.15)$$

where σ_e^2 is the variance of the prediction error and for the second-to-last equality, the independence between e and $h(\mathbf{x},\boldsymbol{\theta})$ was utilized. These equations show that to characterize the statistics of interest, knowledge of only the mean and variance of the error is necessary. Therefore, to establish unbiased predictions, the mean needs to be zero. This also agrees with modeling intuition; the error in Eq. (5.13) should be zero mean (the opposite would indicate a consistent bias in the predictions of the numerical model). Thus, the minimum additional information needed to quantify the error statistically is its variance. Based on this information, the probability model for $p(e)$ is chosen as Gaussian with zero mean and variance σ_e^2. This Gaussian model is the probability model that maximizes the entropy, or equivalently, it incorporates the largest amount of uncertainty in the definition of $p(e)$, assuming that only its mean and variance are known (Jaynes 2003), and is the reason why it is chosen.

Therefore, the only remaining question for the complete description of $p(e)$ is: how should its variance be defined? Note that based on Eq. (5.13), the error in this case

(and so its variance) has the same units as the performance function. Rather than choosing the variance to be constant in the design space, a reasonable assumption is to take the error statistics to be dependent upon the statistic of the system model. If the error is assumed to represent un-modeled uncertainties, σ_e^2 should be set as a fraction of the variance of the system model.

$$\sigma_e^2 = \psi^2 \sigma^2(\mathbf{x}), \tag{5.16}$$

where ψ is a proportionality constant. Note that this constant will typically have a small value since the error is expected to be small (at least if the engineering model predictions are expected to be close to the real system behavior). Through Eq. (5.16), the error statistics are scaled based on the variability observed for the system model and stemming from the uncertainty in $\boldsymbol{\theta}$; if that variability is small, so will be the error, showing that in fact the error does primarily address un-modeled uncertainties. If, on the other hand, the error is taken to also represent un-modeled characteristics of the system, a more reasonable assumption is to select it proportional to the second moment of the response, scaling the overall magnitude of the response.

$$\sigma_e^2 = \psi^2 E[h(\mathbf{x},\boldsymbol{\theta})^2] = \psi^2 \sigma^2(\mathbf{x}) + \psi^2 \mu^2(\mathbf{x}). \tag{5.17}$$

The latter component in Eq.(5.17), i.e. $\psi^2 \mu^2(\mathbf{x})$, can be considered to address the un-modeled characteristics of the real system, with the first component representing the un-modeled uncertainties. Both these assumptions ultimately lead to an error variance that is dependent upon the design configuration $\sigma_e^2 = \sigma_e^2(\mathbf{x})$. Modeling assumption Eq.

(5.16) will be referred to hereafter as variability-proportional and modeling assumption Eq. (5.17) as moment-proportional.

Once the error has been defined, the important question is how to incorporate it in the analysis for the probability of dominance; In other words, how is $P_D(\mathbf{x}_i \mid S_{il}^d, \boldsymbol{\theta})$ modified by considering the influence of the prediction error? In this case, $P_D(\mathbf{x}_i \mid S_{il}^d, \boldsymbol{\theta})$ is expressed with respect to the real system behavior,

$$P_D(\mathbf{x}_k \mid S_{kl}^d, \boldsymbol{\theta}) = P\left[\bigcap_{m=1, m\neq i}^{d} \left(h_s(\mathbf{x}_k) < h_s(\mathbf{x}_m^{dlk})\right) \mid \boldsymbol{\theta}\right], \tag{5.18}$$

and utilizing Eq. (5.13) to introduce the impact of the prediction error, Eq. (5.18) is transformed to

$$P\left[\bigcap_{m=1, m\neq i}^{d} \left(h_s(\mathbf{x}_k) < h_s(\mathbf{x}_m)\right) \mid \boldsymbol{\theta}\right] = P\left[\bigcap_{m=1, m\neq i}^{d} \left(h(\mathbf{x}_k, \boldsymbol{\theta}) + e_k < h(\mathbf{x}_m, \boldsymbol{\theta}) + e_m\right)\right], \tag{5.19}$$

where the superscript dlk is omitted herein for notational simplicity and a subscript is used for e to denote the design configuration to which each error term corresponds. This then leads to

$$\begin{aligned} P\left[\bigcap_{m=1, m\neq i}^{d} \left(h(\mathbf{x}_k, \boldsymbol{\theta}) + e_i < h(\mathbf{x}_m, \boldsymbol{\theta}) + e_m\right)\right] &= P\left[\bigcap_{m=1, m\neq i}^{d} \left(e_k - e_m < h(\mathbf{x}_m, \boldsymbol{\theta}) - h(\mathbf{x}_k, \boldsymbol{\theta})\right)\right], \\ &= P\left[\begin{bmatrix} e_k - e_1 \\ e_k - e_2 \\ \cdots \\ e_k - e_d \end{bmatrix} \leq \begin{bmatrix} h(\mathbf{x}_1, \boldsymbol{\theta}) - h(\mathbf{x}_k, \boldsymbol{\theta}) \\ h(\mathbf{x}_2, \boldsymbol{\theta}) - h(\mathbf{x}_k, \boldsymbol{\theta}) \\ \cdots \\ h(\mathbf{x}_d, \boldsymbol{\theta}) - h(\mathbf{x}_k, \boldsymbol{\theta}) \end{bmatrix}\right], \end{aligned} \tag{5.20}$$

where the different inequalities in the latter expression are examined component-wise.

Considering now the vector

$$\begin{bmatrix} e_k - e_1 \\ e_k - e_2 \\ \cdots \\ e_k - e_d \end{bmatrix}, \tag{5.21}$$

and based on the assumed probability model for each of the error terms, it follows a Gaussian distribution with zero mean and covariance matrix with elements (Johnson and Wichern 2002) (the m^{th} and q^{th} entry is presented below)

$$\begin{aligned} E[(e_k - e_m)(e_k - e_q)] &= E[e_k^2] - E[e_k e_q] - E[e_m e_k] + E[e_m e_q], \\ &= \sigma_{ek}^2 - \sigma_{ek}\sigma_{eq}\rho_{kq} - \sigma_{ek}\sigma_{em}\rho_{mk} + \sigma_{em}\sigma_{eq}\rho_{mq}, \end{aligned} \tag{5.22}$$

where a subscript is used to denote the relationship of the error standard deviation to the design configuration since such a relationship has been introduced based on Eq. (5.16) or Eq. (5.17). For the diagonal elements of the covariance matrix, this relationship simplifies to

$$E[(e_k - e_m)^2] = \sigma_{ek}^2 + \sigma_{em}^2 - 2\sigma_{ek}\sigma_{em}\rho_{km}. \tag{5.23}$$

Since the probability model for the error vector in Eq. (5.21) is known, the probability in Eq. (5.20) corresponds to the associated cumulative distribution function (CDF). This eventually leads to

$$P_D(\mathbf{x}_k \mid S_{kl}^d, \boldsymbol{\theta}) = F_G\left[Dh_{kl}^d(\boldsymbol{\theta}), 0, \boldsymbol{\Sigma}_{kl}^d \right], \tag{5.24}$$

and probability of dominance given by

$$P_D(\mathbf{x}_k \mid S_{kl}^d) = \int_{\Theta} F_G\left[Dh_{kl}^d(\boldsymbol{\theta}), 0, \boldsymbol{\Sigma}_{kl}^d \right] p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{5.25}$$

where $F_G\left[Dh_{kl}^d(\boldsymbol{\theta}),0,\boldsymbol{\Sigma}_{kl}^d\right]$ stands for the CDF of the multivariate Gaussian distribution with zero mean, and covariance matrix $\boldsymbol{\Sigma}_{kl}^d$ with elements Eq. (5.22) obtained by comparing the error for the k^{th} design to the error of each other element of S_{kl}^d. This CDF is centered and evaluated at the vector of differences between the performances of each of these designs to the performance of the k^{th} design (recall this vector does not include the k^{th} design), given by

$$Dh_{kl}^d(\boldsymbol{\theta}) = \begin{bmatrix} h(\mathbf{x}_1^{dlk},\boldsymbol{\theta}) - h(\mathbf{x}_k,\boldsymbol{\theta}) \\ h(\mathbf{x}_2^{dlk},\boldsymbol{\theta}) - h(\mathbf{x}_k,\boldsymbol{\theta}) \\ \cdots \\ h(\mathbf{x}_d^{dlk},\boldsymbol{\theta}) - h(\mathbf{x}_k,\boldsymbol{\theta}) \end{bmatrix}. \tag{5.26}$$

The introduction of the prediction error leads to modification of the initial indicator function, which was imposing a binary distinction for the system performance (either dominates or not), to a preference function. This function is expressed as the likelihood of dominance based on the difference between the performance of the design of interest and the performance of the other designs under consideration given by Eq. (5.26), as well as the statistics of the prediction error, introduced through the covariance matrix $\boldsymbol{\Sigma}_{kl}^d$. Note that this ultimately corresponds to analytical integration over the prediction error on the initial probabilistic integral Eq. (5.18). Figure 5.1 illustrates this concept for a two-dimensional example (comparison of two designs). It is evident that increasing the prediction error variance (equivalent in this case to increasing β) provides a smoother transition between the two extreme cases, as the amount of uncertainty in the system description increases (reducing the confidence in

assessing the real-system dominance when knowing the dominance characteristics for the model).

Computationally, the probability of dominance with the prediction error can still be evaluated by Monte Carlo simulation as in Eq. (5.8) substituting $P_D(\mathbf{x}_k \mid S_{kl}^d, \boldsymbol{\theta}^j)$ with $F_G\left[Dh_{kl}^d(\boldsymbol{\theta}^j), 0, \boldsymbol{\Sigma}_{kl}^d\right]$. The latter multivariate CDF can be estimated by any standard numerical approach. In the example later, the highly efficient algorithm proposed by Genz is adopted (Genz 1992). For assessing the dominance over the entire set, the approach is the same as above and simply needs to utilize Γ instead of S_{il}^d.

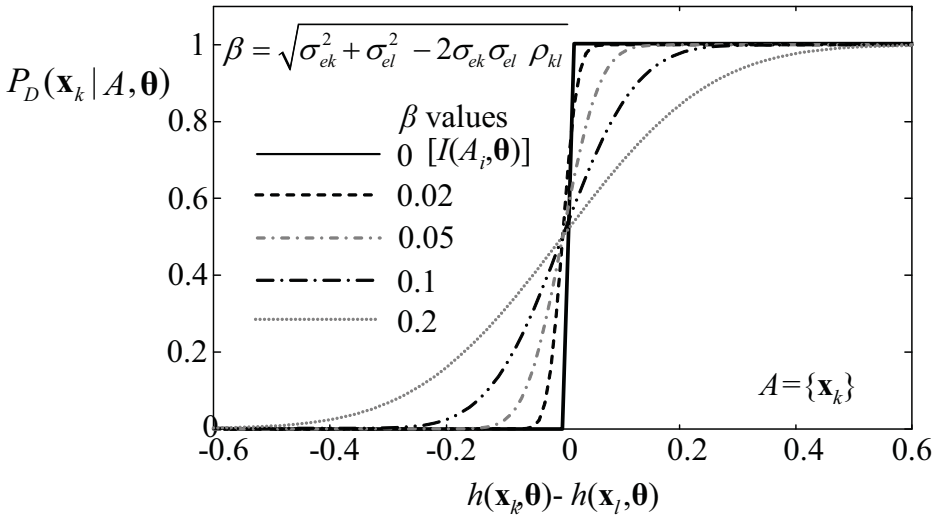


Figure 5.1 Likelihood function for comparison of $\mathbf{x}_k$ against $A = \{\mathbf{x}_l\}$ for different values of $\beta^2 = \sigma_{ek}^2 + \sigma_{el}^2 - 2\sigma_{ek}\sigma_{el}\rho_{kl}$

5.2.3 Multiplicative Prediction Error

Modeling prediction errors through a multiplicative assumption is also common in the engineering literature (Taflanidis et al. 2013; Porter et al. 2007), especially for performance functions that must take strictly positive values (note that this was the

case considered in the example in Sections 3 and 4 for quantifying the probabilistic performance). The relationship between system and model performance in this case is

$$h_s(\mathbf{x}) = h(\mathbf{x},\boldsymbol{\theta}) \cdot e, \tag{5.27}$$

with $e > 0$ corresponding to a dimensionless quantity. This modeling assumption for the prediction error addresses both un-modeled uncertainties and un-modeled characteristics in the numerical model since it directly scales the entire performance function (the option of addressing only un-model uncertainties does not exist here). The mean and variance for the real system, respectively, are

$$\mu_s(\mathbf{x}) = E[h_s(\mathbf{x})] = E[h(\mathbf{x},\boldsymbol{\theta}) \cdot e] = E[h(\mathbf{x},\boldsymbol{\theta})] \cdot E[e] = E[h(\mathbf{x},\boldsymbol{\theta})] = \mu(\mathbf{x}), \tag{5.28}$$

$$\begin{aligned} \sigma_s^2(\mathbf{x}) &= E[h_s^2(\mathbf{x})] - \mu_s^2(\mathbf{x}) = E[(h(\mathbf{x},\boldsymbol{\theta})\, e)^2] - \mu^2(\mathbf{x}) = E[h(\mathbf{x},\boldsymbol{\theta})^2]\, E[e^2] - \mu^2(\mathbf{x}), \\ &= \sigma^2(\mathbf{x}) + E[h(\mathbf{x},\boldsymbol{\theta})^2]\left(E[e^2] - 1\right) = \sigma^2(\mathbf{x}) + E[h(\mathbf{x},\boldsymbol{\theta})^2]\,\sigma_e^2, \end{aligned} \tag{5.29}$$

where $E[e]$ is taken equal to one in order to establish unbiased predictions and the independence between e and $\boldsymbol{\theta}$ was used in the third equality in both relationships. Similarl to the case of additive prediction error, only the mean and variance of the error are necessary to determine the statistics of interest for the real system behavior. Therefore, to establish unbiased predictions, the mean of the error needs to be one. Since in this case the error e is dimensionless and directly scales the model performance, its variance should be independent of the characteristics of the model, so it is assumed to be $\sigma_e^2 = \psi^2$, with ψ typically having a small value (similar justification as for additive error). Note that for the same value of ψ, there is an equivalence for the statistics of the real-system response between the multiplicative error and the additive

error with moment-proportional variance modeled through Eq. (5.17). This is consistent also with the fact that both of these cases address both (i) un-modeled uncertainties and (ii) un-modeled system characteristics.

Furthermore, taking the log of Eq. (5.27), we have

$$\ln(h_s(\mathbf{x})) = \ln(h(\mathbf{x},\boldsymbol{\theta})) + \ln(e), \tag{5.30}$$

which means then that the $log(e)$ has an additive influence on the log of the performance. Based on this observation, the common assumption for the multiplicative prediction error probability model is a lognormal distribution. This also agrees with the most appropriate model based on the maximum entropy principle if the available knowledge is assumed to be on the log of the error (rather than the error itself) to circumvent the $e > 0$ constraint. This means that $\bar{e} = log(e)$ follows a Gaussian distribution with mean and variance

$$\begin{aligned} \mu_{\bar{e}} &= -\sigma_{\varepsilon}^2 / 2, \\ \sigma_{\bar{e}}^2 &= \ln(\sigma_e^2 + 1) \approx \sigma_e^2 \end{aligned} \tag{5.31}$$

where the approximation in the second equation is established assuming that σ_e^2 is small and the first equality is necessary so that $E[e] = 1$.

The modification for $P_D(\mathbf{x}_i \mid S_{il}^d, \boldsymbol{\theta})$ is in this case

$$\begin{aligned} P_d(\mathbf{x}_k \mid S_{kl}^d, \boldsymbol{\theta}) &= P\left[\bigcap_{m=1, m\neq i}^{d} \left(h_s(\mathbf{x}_k) < h_s(\mathbf{x}_m^{dlk}) \right) \mid \boldsymbol{\theta} \right], \\ &= P\left[\bigcap_{m=1, m\neq i}^{d} \left(\log(h_s(\mathbf{x}_k)) < \log(h_s(\mathbf{x}_m^{dlk})) \right) \mid \boldsymbol{\theta} \right], \\ &= P\left[\bigcap_{m=1, m\neq i}^{d} \left(\log(h(\mathbf{x}_k,\boldsymbol{\theta}) + \varepsilon_k) < \log(h(\mathbf{x}_m^{dlk},\boldsymbol{\theta}) + \varepsilon_m^{dlk}) \right) \mid \boldsymbol{\theta} \right], \end{aligned} \tag{5.32}$$

which is then analyzed in a similar way to the additive error since $\bar{e} = log(e)$ is Gaussian, leading to

$$P_D(\mathbf{x}_k \mid S_{kl}^d, \boldsymbol{\theta}) = F_G\left[Dl_{kl}^d(\boldsymbol{\theta}), 0, \boldsymbol{\Sigma}_{kl}^d\right], \tag{5.33}$$

where $Dl_{dl}^k(\boldsymbol{\theta})$ is the vector of the difference between the logarithms of the performance

$$Dl_{kl}^d(\boldsymbol{\theta}) = \begin{bmatrix} \log(h(\mathbf{x}_1^{dlk}, \boldsymbol{\theta})) - \log(h(\mathbf{x}_k, \boldsymbol{\theta})) \\ \log(h(\mathbf{x}_2^{dlk}, \boldsymbol{\theta})) - \log(h(\mathbf{x}_k, \boldsymbol{\theta})) \\ \cdots \\ \log(h(\mathbf{x}_d^{dlk}, \boldsymbol{\theta})) - \log(h(\mathbf{x}_k, \boldsymbol{\theta})) \end{bmatrix} = \begin{bmatrix} \log(h(\mathbf{x}_1^{dlk}, \boldsymbol{\theta}) / h(\mathbf{x}_k, \boldsymbol{\theta})) \\ \log(h(\mathbf{x}_2^{dlk}, \boldsymbol{\theta}) / h(\mathbf{x}_k, \boldsymbol{\theta})) \\ \cdots \\ \log(h(\mathbf{x}_d^{dlk}, \boldsymbol{\theta}) / h(\mathbf{x}_k, \boldsymbol{\theta})) \end{bmatrix}, \tag{5.34}$$

and the covariance matrix $\boldsymbol{\Sigma}_{kl}^d$ has elements

$$E[(\bar{e}_k - \bar{e}_m)(\bar{e}_k - \bar{e}_q)] = \sigma_{\bar{e}}^2 (1 + \bar{\rho}_{mq} - \bar{\rho}_{kq} - \bar{\rho}_{mk}), \tag{5.35}$$

where $\bar{\rho}_{km}$ is the correlation between $\bar{e}_k = log(e_k)$ and $\bar{e}_m = \ln(e_m)$, which can be readily obtained based on the correlation of e_k and e_m (Law et al. 1991), through

$$\bar{\rho}_{km} = \frac{\ln\left(1 + \rho_{km}\sigma_e^2\right)}{\ln(1 + \sigma_e^2)} \approx \rho_{km} \frac{\sigma_e^2}{\sigma_e^2} = \rho_{km} \tag{5.36}$$

where the assumption that σ_e is small was used for the approximation in the second equality.

Thus, the evaluation under the impact of multiplicative error follows exactly the same approach as for the additive error, simply by replacing the CDF in Eq. (5.24) with a similar CDF given by Eq. (5.33).

5.2.4 Impact of Prediction Error on the RDO Formulation

Beyond the impact on the probability of dominance, the influence of the prediction error on the RDO designs can also be examined. In the objective function Eq. (2.12), the statistics of the real system are substituted, leading to

$$H_{RDO}(\mathbf{x}) = (1-w)\frac{\mu_s(\mathbf{x})}{\mu_{sn}} + w\frac{\sigma_s(\mathbf{x})}{\sigma_{sn}}, \tag{5.37}$$

with $[\mu_{sn}\ \sigma_{sn}]$ corresponding to the new normalization, selected based on the utopia point for the real system $[\mu_{so}\ \sigma_{so}]$. For multiplicative prediction error or for additive prediction error with the scaled moment-proportional variance of Eq. (5.17) this leads to

$$\begin{aligned} H_{RDO}(\mathbf{x}) &= (1-w)\frac{\mu(\mathbf{x})}{\mu_{so}} + w\frac{\sqrt{\sigma^2(\mathbf{x}) + E[h(\mathbf{x},\boldsymbol{\theta})^2]\psi^2}}{\sigma_{so}} \\ &= (1-w)\frac{\mu(\mathbf{x})}{\mu_s} + w\frac{\sqrt{E[h(\mathbf{x},\boldsymbol{\theta})^2](1+\psi^2) - \mu^2(\mathbf{x})}}{\sigma_{so}}, \end{aligned} \tag{5.38}$$

with $\mu_{so}=\mu_o$, $\sigma_{so} \neq \sigma_o$. For the additive error with the scaling of Eq. (5.16), the new objective function is

$$H_{RDO}(\mathbf{x}) = (1-w)\frac{\mu(\mathbf{x})}{\mu_{so}} + w\frac{\sigma(\mathbf{x})}{\sigma_{so}}\sqrt{1+\psi^2} = (1-w)\frac{\mu(\mathbf{x})}{\mu_o} + w\frac{\sigma(\mathbf{x})}{\sigma_o}, \tag{5.39}$$

with $\mu_{so}=\mu_o$, $\sigma_{so} = \sqrt{1+\psi^2}\sigma_o$, and thus, it remains identical to the objective function with respect to the system-model performance Eq. (2.12) as long as the standard deviation is scaled with respect to the new utopia point. Note that as mentioned earlier,

the expressions of the additive error with moment scaling and the multiplicative error yield the same model variance.

5.2.5 Summary for modeling/impact of the prediction error

This section addressed the influence of the model prediction error on the probability of dominance as well as on the RDO formulation. The modeling of this error was also extensively discussed, and various reasonable modeling assumptions were presented, the most important being that the adopted numerical model for the system provides unbiased predictions for the expected performance. Additive and multiplicative impacts of the error were considered, and the assumed probability models were suggested to be (i) Gaussian with zero mean and variance scaled according to the statistics of the numerical model and (ii) lognormal with mean unity and a constant variance (independent of the model statistics). The additive prediction error can possibly be chosen to address only un-modeled uncertainties, with its variance scaled according to the response variance of the assumed numerical model (because of the modeled uncertainties $\boldsymbol{\theta}$. In this case, the formulation of the RDO problem is not impacted at all. The additive error can also be selected to address both un-modeled uncertainties and un-modeled system characteristics, with its variance scaled according to the second moment of the numerical model. This is equivalent to the multiplicative prediction error in terms of how it impacts the statistics of the real system performance and ultimately the RDO formulation.

Lastly, the impact of the modeling error on the probability of dominance can be readily addressed by substituting the indicator function describing the favorability of the design given the system model with a likelihood function dependent upon the assumed probability model of the error and the model performance.

5.3 Illustrative Examples

The proposed framework is now illustrated with two examples for which RDO formulations are very popular in the literature (this is the motivation for using a different example in this section, as opposed to the one discussed in the previous two sections). The first example considers the design of a tuned mass damper (TMD) for vibration mitigation of harmonic excitations, and extensive comparisons are established for a variety of different cases. These cases mainly vary in the probability models utilized for $p(\boldsymbol{\theta})$ and the error assumption. The second example discusses the topology optimization problem for minimum compliance, and the discussion focuses only on specific cases of interest.

5.3.1 Case Study: Tuned Mass Damper (TMD) Design

5.3.1.1 TMD Simulation Model

Tuned mass dampers (TMDs) are widely used to mitigate vibrations for a variety of different structural systems (Chang 1999; Kareem and Kline 1995; Kwon and Park 2004). They consist (as seen in Figure 5.2) of a secondary mass m_d connected to the

primary mass through a spring with stiffness $k_s = \omega_d^2 m_s$ and a dashpot with damping coefficient $c_d = 2\zeta_d \omega_d m_d$, where ω_d represents the damper's natural frequency and ζ_d its damping ratio. Through the proper tuning of these two design variables (stiffness and damping), significant reduction of the response can be achieved for a variety of dynamic environmental excitations.

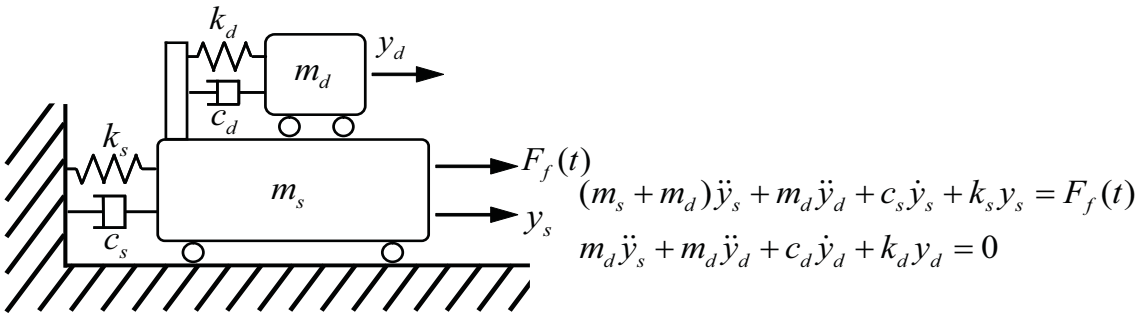


Figure 5.2 Schematic of single-degree-of-freedom system with tuned mass damper. The equation of motion is also shown.

The RDO of TMDs has been discussed in the literature (Marano et al. 2008; Mohtat and Dehghan-Niri 2011; Mohtat and Dehghan-Niri 2011; Zang et al. 2005). Here the design of a TMD to mitigate the vibrations of a bridge structure, modeled as a single-degree-of-freedom (SDOF) system under harmonic (monochromatic) excitation is examined, similar to the problem discussed in (Zang et al. 2005). Figure 5.2 shows the details of a SDOF system with mass m_s, stiffness $k_s = \omega_s^2 m_s$, and damping coefficient $c_s = 2\zeta_s \omega_s m_s$ (where ω_s represents the natural frequency and ζ_s the damping ratio), equipped with a TMD, under the impact of dynamic excitation $F_f(t)$. The mass of the damper m_d is taken as 0.5% of the median (i.e., nominal) mass of the structure. The

design variables correspond to the damper stiffness and damping coefficient $\mathbf{x} = [k_d \quad c_d]^T$. The uncertain model parameters correspond to the mass, stiffness, and damping ratio of the structure, $\boldsymbol{\theta} = [m_s \quad k_s \quad \zeta_s]^T$. For all of these, independent lognormal distributions are assumed with coefficients of variation γ_c and median values 17500 Kg for m_s, 3 MN/m for k_s and 3% for ζ_s. Two cases are considered for γ_c, representing different levels of assumed uncertainty in $\boldsymbol{\theta}$, the first corresponding to a 30% coefficient of variation and the second to a 10% coefficient of variation. These cases will be termed *large variability* and *small variability*, respectively.

The performance function is taken as the maximum of the amplification factor, corresponding to the maximum of the dynamic response for a monochromatic excitation over every potential frequency ω. This ultimately leads to (Chang 1999)

$$h(\mathbf{x},\boldsymbol{\theta}) = \max_{\beta_f}\left[\frac{\sqrt{[\varphi_f^2 - \beta_f^2]^2 + \left(2\zeta_d\varphi_f\beta_f\right)^2}}{\sqrt{E_s^2 + D_s^2}}\right];\ \varphi_f = \frac{\omega_d}{\omega_s},\ \beta_f = \frac{\omega}{\omega_s}$$
$$E_s = [1-\beta_f^2][\varphi_f^2 - \beta_f^2] - 4\zeta_s\zeta_d\varphi_f\beta_f^2 - \mu\varphi_f^2\beta_f^2 \qquad (5.40)$$
$$D_s = 2\beta_f\left\{\zeta_s[\varphi_f^2 - \beta_f^2] + [1-(1+\mu)\beta_f^2]\zeta_d\varphi_f\right\} \quad ,$$

where the auxiliary variables φ_f and β_f correspond to the tuning ratio for the damper and the non-dimensional excitation frequency, respectively, while $\mu = m_s / m_d$ is defined as the mass ratio.

5.3.1.2 Results and Discussion

The RDO weight parameter w is varied from 0 to 1 with higher emphasis on lower weight values, $w = [0, 0.1, 0.2, 0.3, 0.4, 0.6, 1]$ (this creates a balanced Pareto front as will be shown later), leading ultimately to seven different designs configurations. The statistics needed for the objective function evaluation are calculated using stochastic simulation as in Eqs. (2.4) and (2.7). The total number of samples to generate the RDO designs was taken as $N_r = 2000$, and an exterior sampling approach is adopted as discussed in Chapter 2, using the same samples over the entire optimization procedure, thus transforming the RDO Eq. (2.11) into a standard deterministic optimization problem that can be solved by any numerical optimization approach; for this case, the design domain is discretized with small intervals and an exhaustive search was employed. The upper bounds of the design space X are selected as $k_d = [1\times10^4 \;\; 1.5\times10^4] N/m$ and $c_d = [100 \;\; 500] Ns/m$, while the lower bounds are zero. The following cases are considered

i. No prediction error, termed as NE.

ii. Additive prediction error with variance scaling according to Eq. (5.16) and ψ chosen as 10%. This case is termed as AE.

iii. Additive prediction error with variance scaling according to Eq. (5.17) and ψ chosen as 5%. This case is termed as AE_t.

iv. Multiplicative prediction error (ME) with ψ chosen the same as in AE_t.

As discussed in Section 5.2, the NE and AE cases and ME and AE_t cases lead to identical RDO designs when normalization is established with respect to the utopia point (as was done here). The higher value for ψ chosen for AE compared to the value for AE_t is selected so that similar values are established for the error variance σ_e^2 between AE and AE_t.

The results of the optimization are shown in Table 5.2, Figure 5.3 and Figure 5.4. In particular, Table 5.2 shows the optimal solutions for k_d and c_d and the different weights, as well as the performance $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ under that optimal design configuration. For the AE_t and ME designs, the variance for the real system $\sigma_s(\mathbf{x})$ is reported (that is, under the influence of the prediction error). For the AE design this information is omitted since it merely corresponds to multiplication of $\sigma(\mathbf{x})$ by $\sqrt{1+\psi^2}$. Figure 5.3 then shows the different optimal solutions in the design space X whereas Figure 5.4 the Pareto-front in the normalized (with respect to utopia point) objective space F. The arrow in this figure represents the closest distance of any design from the utopia point. The closest point corresponds to the $w = 0.4$ design for all cases considered, and it would have been the preferred design based on this traditional approach for evaluation of appropriateness among RDO design.

TABLE 5.2

RDO DESIGNS RESULTS FOR ALL CASES CONSIDERED FOR THE TMD IMPLEMENTATION

	Large variability				Small variability			
	NE , AE		AE_t , ME		NE , AE		AE_t , ME	
w	$\mathbf{x}_k$ $k_d\ (N/m)$ $c_d\ (Ns/m)$	$\mu(\mathbf{x})$ $\sigma(\mathbf{x})$	$\mathbf{x}_k$ $k_d\ (N/m)$ $c_d\ (Ns/m)$	$\mu(\mathbf{x})$ $\sigma_s(\mathbf{x})$	$\mathbf{x}_k$ $k_d\ (N/m)$ $c_d\ (Ns/m)$	$\mu(\mathbf{x})$ $\sigma(\mathbf{x})$	$\mathbf{x}_k$ $k_d\ (N/m)$ $c_d\ (Ns/m)$	$\mu(\mathbf{x})$ $\sigma_s(\mathbf{x})$
0	14873 211.4	14.8 4.41	14873 211.4	14.81 4.47	14705 142.05	12.4 1.84	14705 142.05	12.49 1.95
0.1	14536 244.43	14.8 4.32	14536 241.12	14.83 4.39	14367 198.19	12.7 1.47	14409 184.98	12.65 1.67
0.2	14240 274.15	14.8 4.26	14282 274.15	14.87 4.33	14156 247.73	13.0 1.28	14240 227.91	12.91 1.49
0.3	14029 303.87	14.9 4.22	14029 300.57	14.91 4.29	13987 277.45	13.2 1.22	14071 257.64	13.12 1.42
0.4	13776 323.69	14.9 4.20	13818 320.39	14.95 4.27	13903 297.27	13.3 1.20	13987 280.76	13.27 1.39
0.6	13438 356.71	15.0 4.18	13480 356.71	15.02 4.25	13734 326.99	13.5 1.18	13818 307.18	13.45 1.37
1	12974 396.35	15.1 4.17	12974 393.04	15.11 4.25	13565 356.71	13.7 1.17	13649 340.2	13.65 1.36

Focusing first on these RDO results, it is evident that the uncertainty in the system description has an important impact on the robust design configuration. Different values for w lead to different optimal designs, with the higher variability case (larger uncertainty in system description) contributing, as expected, to higher values of the system statistics (mean and standard deviation). The variation of the normalized values for these statistics is, though, smaller for this case (Figure 5.4), and the same principle applies for the optimal design configurations. This is easy to explain; because of the higher uncertainty in the system description, the TMD has smaller comparative

efficiency (more challenging to regulate the performance over the entire Θ region of importance) and as such, the differences between the performance for different w values become smaller. The prediction error has an impact on the optimal design configuration (as seen by the differences in designs between AE_t / ME and NE / AE), especially for the small variability case, something that is evident in Figure 5.3 and Figure 5.4. This should be attributed to the fact that since for that case the impact of the system uncertainty is smaller (smaller variability in system characteristics), which allows the uncertainty induced by the error to influence more prominently the resultant design/performance.

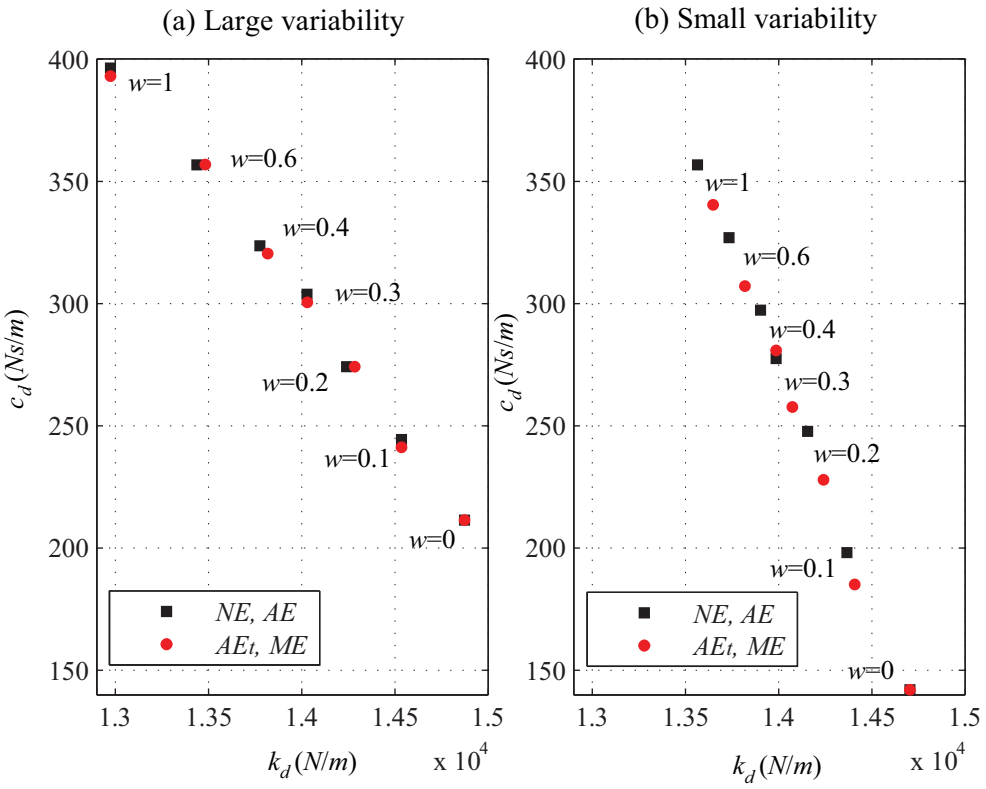


Figure 5.3 Design space along with Pareto optimal designs that compose the set Γ for (a) large variability case, and (b) small variability

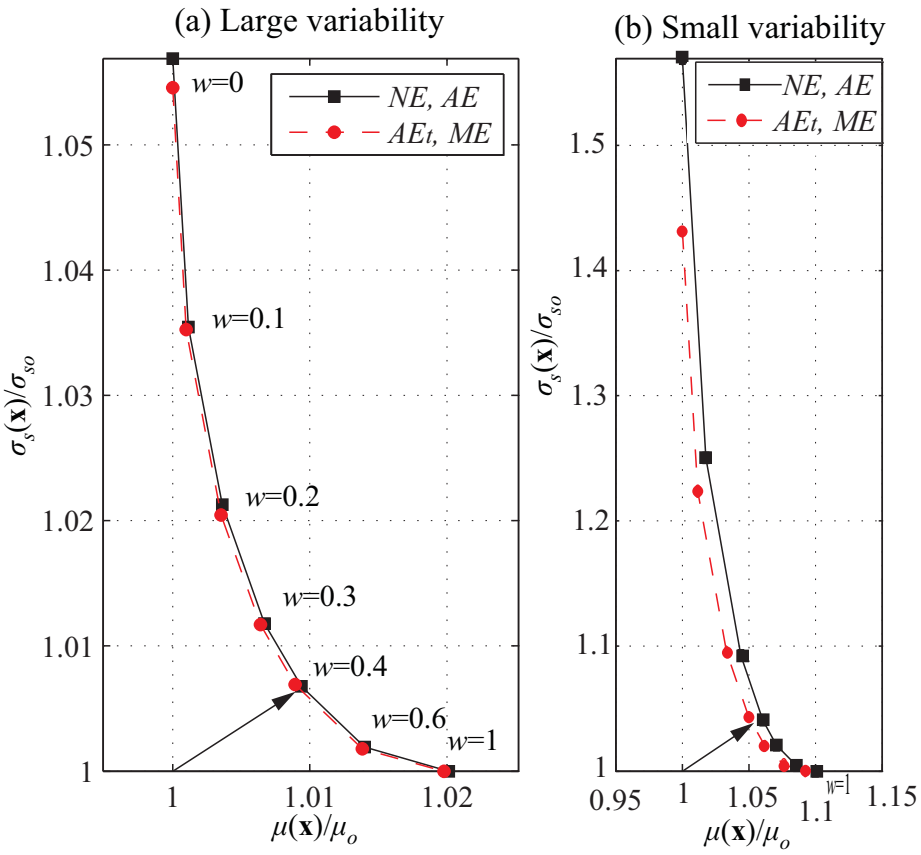


Figure 5.4 Pareto-front for the TMD RDO implementation of the (a) large and (b) small variability cases. Closest design to utopia point indicated with an arrow.

The probability of dominance is examined next. This further requires definition of the correlation characteristics for the prediction error. The exponential function in Eq. (5.12) is adopted, with normalization weight v_i selected as one over the length of the design domain for each design variable. The decay rate λ is selected so that correlation decrease is equal to $\rho_o = 5\%$ when the total change in the weighted distance reaches $\alpha = 30\%$, leading to

$$\lambda = \frac{-\log(\rho_o)}{\alpha}. \tag{5.41}$$

This assumption leads to the following correlation matrix ρ_{km} between the different designs for the AE design case (similar are the matrices for the AE_t / ME cases) for the large and small variability cases, respectively,

$$\rho_{km} = \begin{bmatrix} 1 & 0.89 & 0.67 & 0.44 & 0.28 & 0.12 & 0.03 \\ 0.89 & 1 & 0.91 & 0.72 & 0.54 & 0.28 & 0.09 \\ 0.67 & 0.91 & 1 & 0.93 & 0.79 & 0.51 & 0.21 \\ 0.44 & 0.72 & 0.93 & 1 & 0.95 & 0.73 & 0.38 \\ 0.28 & 0.54 & 0.79 & 0.95 & 1 & 0.89 & 0.56 \\ 0.12 & 0.28 & 0.51 & 0.73 & 0.89 & 1 & 0.83 \\ 0.03 & 0.09 & 0.21 & 0.38 & 0.56 & 0.83 & 1 \end{bmatrix}, \quad (5.42)$$

$$\rho_{km} = \begin{bmatrix} 1 & 0.78 & 0.44 & 0.26 & 0.17 & 0.08 & 0.03 \\ 0.78 & 1 & 0.84 & 0.64 & 0.5 & 0.3 & 0.16 \\ 0.44 & 0.84 & 1 & 0.94 & 0.84 & 0.63 & 0.41 \\ 0.26 & 0.64 & 0.94 & 1 & 0.97 & 0.84 & 0.63 \\ 0.17 & 0.5 & 0.84 & 0.97 & 1 & 0.94 & 0.77 \\ 0.08 & 0.3 & 0.63 & 0.84 & 0.94 & 1 & 0.94 \\ 0.03 & 0.16 & 0.41 & 0.63 & 0.77 & 0.94 & 1 \end{bmatrix}. \quad (5.43)$$

The probability of dominance over Γ as well as the degree of dominance for 3 different values of $d = \{2, 3, 4\}$ are then calculated and shown in Table 5.3 though Table 5.6. Rather than just identifying the dominant design based on the procedure outlined in Section 5.1.2, here the degree of dominance for different d values is reported to facilitate a more thorough comparison. In particular, Table 5.3 and Table 5.5 show the results for the NE and AE designs for the large and small variability cases, respectively. Note that even though the designs corresponding to these cases (AE and NE) are the same, the probability of dominance will not be, since the latter design

includes a prediction error (that influences this probability). Subsequently, Table 5.4 and Table 5.6 show the results for the AE_t and ME in a similar fashion. The results will also be different in this case because of the different impact of the prediction error. 10000 samples were used to calculate the statistics reported in these tables, facilitating a high-accuracy comparison.

Based on the probability of dominance $P_D(\mathbf{x}_k \mid \Gamma)$ over the entire set Γ, the design corresponding to $w = 0$ is deemed as the dominant one for all cases considered. For the small variability case, this dominance is preserved even when looking at lower dimensional subsets. For all subsets considered, i.e. $d = \{2, 3, 4\}$, design $w = 0$ emerges as the more dominant one (degree of dominance for it has a positive value), although the margin of dominance is somewhat reduced, dropping to just 15% for $d = 2$ dimensional subsets. The design corresponding to $w = 0.1$ emerges as the second most robust design for all cases considered. It is interesting to note that this design would have also been the second preferable one when looking at the entire set Γ. Introduction of the model prediction error has a small influence on the estimated statistics, with the overall trend corresponding to relative reduction of the preference, while degrees of dominance trend closer to 0 (note that this does not always hold). No significant changes are reported for the degree of preference toward the $w = 0$ design. This should be attributed to the fact that the main rival design is $w = 0.1$, which has high correlation to $w = 0$ based on the assumed correlation function for the prediction error, As such, this error has only a small impact on the calculated statistics.

TABLE 5.3

PROBABILITY AND DEGREE OF DOMINANCE FOR NE AND AE (LARGE VARIABILITY)

	NE				AE			
w	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$D_4(\mathbf{x}_k)$	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$D_4(\mathbf{x}_k)$
0	0.370	-0.260	-0.203	-0.159	0.352	-0.047	-0.038	-0.001
0.1	0.048	-0.145	-0.515	-0.480	0.084	0.032	-0.342	-0.314
0.2	0.047	-0.083	-0.476	-0.476	0.079	-0.032	-0.379	-0.366
0.3	0.069	0.074	-0.366	-0.377	0.063	-0.094	-0.428	-0.398
0.4	0.038	-0.074	-0.483	-0.483	0.074	-0.129	-0.427	-0.392
0.6	0.059	-0.129	-0.498	-0.469	0.092	-0.172	-0.416	-0.369
1	0.369	-0.262	-0.196	-0.159	0.257	-0.210	-0.228	-0.190

TABLE 5.4

PROBABILITY AND DEGREE OF DOMINANCE FOR ME AND AE_t (LARGE VARIABILITY)

	ME				AE_t			
w	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$D_4(\mathbf{x}_k)$	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$D_4(\mathbf{x}_k)$
0	0.345	0.015	-0.017	0.015	0.329	-0.008	-0.045	0.01
0.1	0.093	-0.015	-0.342	-0.315	0.092	0.008	-0.331	-0.303
0.2	0.079	-0.076	-0.399	-0.373	0.081	-0.047	-0.382	-0.356
0.3	0.067	-0.110	-0.442	-0.394	0.068	-0.083	-0.426	-0.378
0.4	0.079	-0.128	-0.409	-0.382	0.081	-0.103	-0.394	-0.364
0.6	0.099	-0.150	-0.37	-0.353	0.101	-0.127	-0.361	-0.335
1	0.237	-0.168	-0.243	-0.206	0.247	-0.144	-0.216	-0.179

TABLE 5.5

PROBABILITY AND DEGREE OF DOMINANCE FOR NE AND AE (SMALL VARIABILITY)

	NE				AE			
w	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$D_4(\mathbf{x}_k)$	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$D_4(\mathbf{x}_k)$
0	0.579	0.158	0.278	0.299	0.577	0.161	0.281	0.301
0.1	0.197	-0.158	-0.381	-0.381	0.200	-0.161	-0.376	-0.377
0.2	0.091	-0.553	-0.684	-0.685	0.082	-0.554	-0.690	-0.691
0.3	0.048	-0.737	-0.819	-0.819	0.048	-0.724	-0.809	-0.810
0.4	0.022	-0.829	-0.891	-0.891	0.021	-0.810	-0.877	-0.876
0.6	0.021	-0.877	-0.916	-0.893	0.024	-0.853	-0.899	-0.877
1	0.039	-0.920	-0.898	-0.875	0.046	-0.896	-0.875	-0.853

TABLE 5.6

PROBABILITY AND DEGREE OF DOMINANCE FOR ME AND AE_t (SMALL VARIABILITY)

	ME				AE_t			
w	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$D_4(\mathbf{x}_k)$	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$D_4(\mathbf{x}_k)$
0	0.521	0.153	0.248	0.271	0.513	0.134	0.236	0.260
0.1	0.173	-0.153	-0.363	-0.358	0.172	-0.134	-0.355	-0.350
0.2	0.079	-0.434	-0.606	-0.596	0.081	-0.416	-0.594	-0.583
0.3	0.053	-0.517	-0.674	-0.658	0.054	-0.503	-0.664	-0.648
0.4	0.036	-0.538	-0.683	-0.670	0.037	-0.526	-0.675	-0.660
0.6	0.041	-0.561	-0.666	-0.680	0.042	-0.549	-0.659	-0.670
1	0.093	-0.594	-0.648	-0.617	0.097	-0.584	-0.638	-0.605

TABLE 5.7

PROBABILITY AND DEGREE OF DOMINANCE FOR SOME DESIGN CASES WHEN NO CORRELATION IS ASSUMED FOR PREDICTION ERROR

	Small variability						Large variability		
	AE			*ME*			*AE*		
w	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$
0	0.551	0.160	0.258	0.388	0.127	0.153	0.243	0.013	0.026
0.1	0.202	-0.160	-0.313	0.205	-0.127	-0.158	0.144	-0.013	-0.076
0.2	0.082	-0.538	-0.594	0.122	-0.267	-0.298	0.118	-0.042	-0.115
0.3	0.056	-0.614	-0.667	0.092	-0.361	-0.373	0.111	-0.072	-0.137
0.4	0.042	-0.704	-0.726	0.076	-0.435	-0.432	0.114	-0.092	-0.159
0.6	0.034	-0.775	-0.795	0.063	-0.506	-0.496	0.123	-0.13	-0.185
1	0.030	-0.833	-0.817	0.050	-0.573	-0.535	0.143	-0.180	-0.180

For the high variability case, the trends change. Similarly, the $w = 0$ design maintains dominance when looking at the entire set Γ. However, it would not emerge as dominant for some of the cases considered when looking at the lower dimensional subsets. For the NE case, the $w = 0.3$ design emerges as the dominant one since it exhibits $d = 2$ dominance, although with relatively small degree of dominance (7.4%). It is interesting to note than when looking at higher values for d, $w = 0.3$ quickly loses its dominance, and even when looking at the entire set Δ, its associated probability of dominance is rather small (7%). Another interesting characteristic for the NE case is the fact that no design exhibits dominance for $d = 3, 4$ dimensional subsets, with all of them having negative values for the degree of dominance. Overall $w = 0$, $w = 1$ and $w = 0.3$ exhibit the highest probability of dominance with robust selection

corresponding to $w = 0$ if one is merely looking at the entire set Γ and to $w = 0.3$ based on the proposed multistage approach, focusing on lower d -dimensional dominance. This discussions demonstrate that looking solely at $\hat{P}_D(\mathbf{x}_k \mid \Gamma)$ could not be enough, and looking at dominance within smaller subsets becomes necessary. Comparisons among S^2 provide the more trustworthy information since it is equivalent to one-on-one comparisons of the designs. Looking at larger dimensional subsets, conclusive information is not necessarily provided for selecting a particular design, though some valuable insight is provided.

The introduction of the error in this case has a bigger impact (compared to the low variability case); for AE, the statistics change significantly when compared with the NE case, even though the compared designs are identical. This demonstrates the importance of explicitly accounting for the prediction error in the comparisons for the probability of dominance. For AE, the design $w = 0.1$ exhibits the lower dimensional dominance ($d = 2$), though with a very small margin (degree of dominance 3%), whereas two other designs, $w = 0.2$ and $w = 0.1$, show good performance (degree of dominance as large as -5%). For this instance, (dominance by only a small margin) one would be interested in looking at higher values for *d* before assessing overall robustness, and in this case, it is evident that $w = 0$ exhibits better robustness characteristics. The previous discussion indicates that beyond the degree of dominance being positive, its value should also be taken into account by the designer in assessing the preferences toward different designs. Looking next at the other cases, for the ME

and AE_t cases, the characteristics for dominance also change (compared to the NE case), although with smaller margins. For ME, $w=0$ emerges as the dominant design, exhibiting both $d=2$ and $d=4$ dominance (but not $d=3$ for which no design emerges as dominant, with $w=0$ having the better performance overall). For AE_t, designs $w=0$ and $w=0.1$ are practically equal. In this case, the decision needs to be based in comparisons of higher order values for d, which tilts the preference toward $w=0$. The big impact of the prediction error on the preference toward different designs for the high variability case can be explained based on the inability of the TMD to efficiently suppress vibrations over the entire important Θ region; thus, close performance is established by a few different designs, and this is why the different preference for this performance given by the different error-model selections can ultimately shift the dominance toward different designs.

To further examine the influence of the correlation of the error on the results, Table 5.7 shows the probability of dominance for some of the design cases considered (AE and ME for the small variability case and AE for the large variability case) if no such correlation is assumed. Comparison of these results to the result in Table 5.3 (for AE and large variability), Table 5.5 (for AE and small variability), and Table 5.6 (for ME and small variability) shows small differences for the AE case when small variability is assumed for the system description, but large differences for the other two cases. For the large variability case (AE), the differences are large enough to impact the dominance of the designs over the $d=2$ and 3 dimensional subsets; when no

correlation is considered for the prediction error, the design $w = 0$ emerges as the dominant design, which was not the case before. This discussion shows that the assumed correlation of the prediction error can have an impact on the calculated statistics, so it should be carefully chosen.

It is also interesting to note that in no case examined does the design selected through the degree of dominance coincide with the one corresponding to the minimum distance from the utopia point ($w = 0.4$ in all cases as previously discussed). The latter design corresponds to rather small values for the probability of dominance and the degree of dominance for different values of *d* in all cases and would have been avoided through the proposed approach. This demonstrates the utility of the new methodology for offering an alternative approach in assessing the robustness characteristics of different proposed designs.

Finally, the connection between RDO and RBDO is investigated with respect to the preference provided by the probability of dominance. For this purpose, the probability of failure P_f for different thresholds for the amplification factor b_{AF} is plotted in Figure 5.5. This probability is defined as the probability that the amplification factor will exceed the threshold b_{AF}. The measure of comparison (failure probability) is motivated by RBDO (since that is the statistical measure used in the problem formulation in that case), whereas the designs compared are ultimately the ones coming from RDO. Results are shown for cases AE and NE only and for small variability since the trends for ME and AE_t are similar, whereas the curves for the high variability case

for the different designs examined were found to be identical (curves practically coincide with one another), stemming from the aforementioned inability of the TMD to regulate effectively the performance over the entire Θ domain.

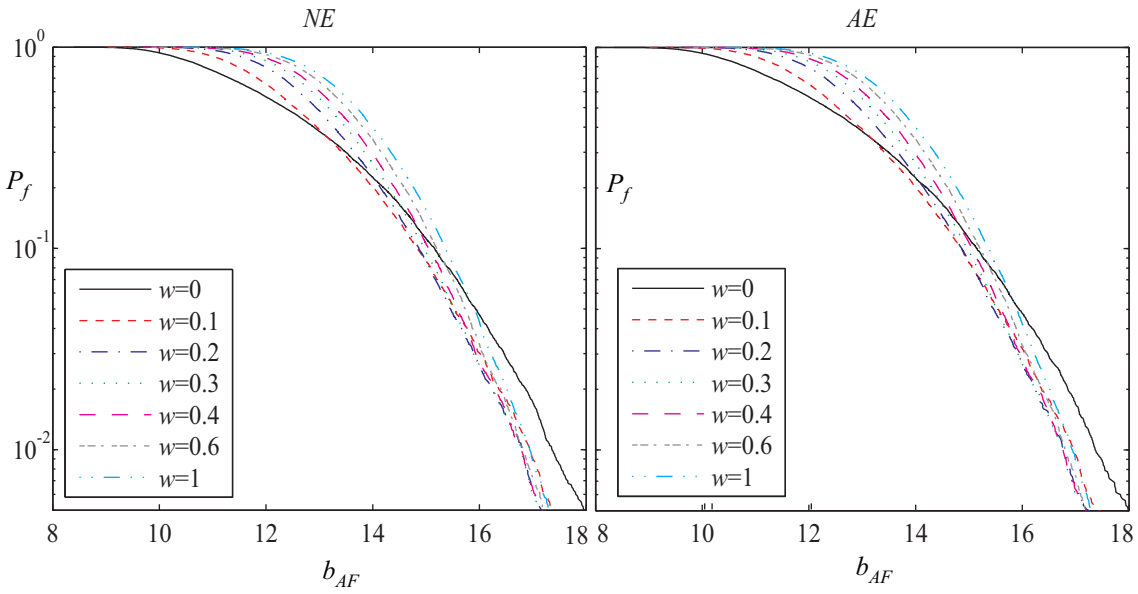


Figure 5.5 Probability of failure P_f as function of failure threshold b_{AF} for NE and AE designs for small variability case

The results indicate a correlation for smaller threshold values between the RBDO criterion and the RDO designs reported to be the dominant ones (corresponding to high probability of dominance), as the latter correspond to the smallest failure probability (best RBDO performance) among the candidate designs. This is not the case for rare events. This demonstrates that a modification of the probability of dominance is needed if the intention is to establish robust performance against rare events. This is understandable since the definition adopted in Eq. (5.1) provides equal importance for all responses and does not necessarily focus on responses that exceed certain

thresholds (thus not giving priority to rare events). The latter can be achieved by incorporating an additional appropriate weight in the definition of the probability of dominance, which is out of the scope of the current investigation. It should be stressed that this discussion follows the general trends reported when comparing between RDO and RBDO approaches; highlighting the differences of the two approaches for rare events. Also, it should be noted that the prediction error has only a small influence on the calculated probabilities of failure. This is anticipated as the values assumed for this error are in general small (the underlying assumption is that the numerical model approximates the system fairly well).

5.3.2 Case Study: Topology Optimization for Minimum Compliance

5.3.2.1 Topology Optimization for Minimum Compliance Simulation Model

Topology optimization searches for the structural configuration (shape and connectivity of the structure) contained within a given design domain X that for some given boundary and loading conditions provides the most favorable response in terms of some chosen performance measure. Details for deterministic topology optimization problems are provided in Appendix B

The RDO formulation has been employed for addressing uncertainties within topology optimization applications (Dunning and Kim 2013; Chen et al. 2010; Dunning et al. 2011; Tootkaboni et al. 2012). A problem similar to the one presented in (Chen et al. 2010) is adopted in this investigation: a rectangular design domain with an aspect ratio

of $2:1$ is subjected to a central point load at the bottom, with simply supported conditions (Figure 5.6). The design domain is discretized into 100×50 elements (resulting in a number of elements of $N_e = 5000$), and the mass fraction is taken as $m_f = 0.3$ (meaning that 30% of the domain may be occupied). The uncertain model parameters correspond to:

(i) The loading conditions, where the load in the horizontal direction F_x is assumed to follow a uniform distribution in $[-1,1]$ and the load in the vertical direction F_y a uniform distribution in $[0,2]$

(ii) The Young's moduli for each element, assumed to follow a lognormal distribution with median 1 and coefficient of variation $(c.o.v)$ of 0.2.

The total number of uncertain parameters is 5002, while the performance function is directly taken as the compliance of the structure, i.e. $h(\mathbf{x},\boldsymbol{\theta}) = c(\mathbf{x},\boldsymbol{\theta})$. The gradient of the performance function is also easy to obtain (i.e. this does not fall under the general category of "black-box" problems), as shown in Appendix B.

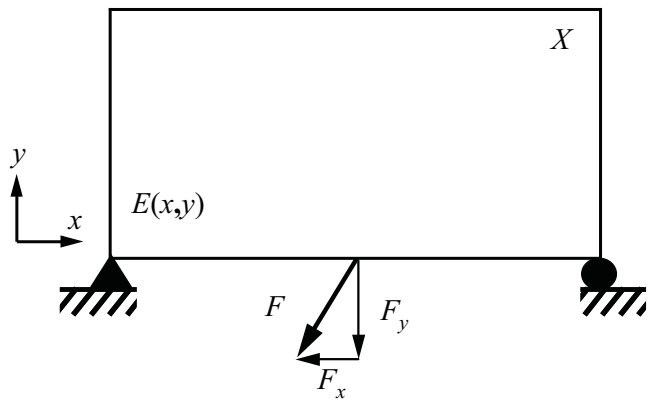


Figure 5.6 Topology optimization RDO problem schematics

The solution to the optimization problem Eq. (2.11) is obtained with the Method of Moving Asymptotes (MMA) (Svanberg 1987), which is a popular algorithm to solve topology optimization problems. To avoid a solution with intermediate densities, a penalization factor of $p_p = 3$ is employed. In addition, to avoid the problem in which a checker-board pattern emerges causing artificial stiffness, a filter with a size of 1.2 is employed. More details about the selection of these parameters can be found in (Bendsøe and Sigmund 2003). Since the optimal design is sought employing a gradient-based algorithm, the gradient for the mean Eq. (2.9) and standard deviation of the response Eq. (2.10) must be provided. Assuming that the functions $c(\mathbf{x},\boldsymbol{\theta})p(\boldsymbol{\theta})$ and $\frac{\partial c(\mathbf{x},\boldsymbol{\theta})}{\partial x_e}p(\boldsymbol{\theta})$ are continuous on the domain $X \times \Theta$ and bounded, while also noting that $p(\boldsymbol{\theta})$ does not depend on $\mathbf{x}$, the differentiation and the expectation operators commute (Spall 2003), leading to the following expressions for the elements of these gradients

$$\frac{\partial \mu(\mathbf{x})}{\partial x_e} = \frac{\partial E[c(\mathbf{x},\boldsymbol{\theta})]}{\partial x_e} = E\left[\frac{\partial c(\mathbf{x},\boldsymbol{\theta})}{\partial x_e}\right], \quad e = 1,\ldots,N_e, \tag{5.44}$$

$$\begin{aligned}\frac{\partial \sigma(\mathbf{x})}{\partial x_e} &= \frac{\partial}{\partial x_e}\left(\sqrt{E\left[c(\mathbf{x},\boldsymbol{\theta})^2\right] - E\left[c(\mathbf{x},\boldsymbol{\theta})\right]^2}\right) \\ &= \frac{E\left[c(\mathbf{x},\boldsymbol{\theta})\frac{\partial c(\mathbf{x},\boldsymbol{\theta})}{\partial x_e}\right] - E\left[c(\mathbf{x},\boldsymbol{\theta})\right]E\left[\frac{\partial c(\mathbf{x},\boldsymbol{\theta})}{\partial x_e}\right]}{\sqrt{E\left[c(\mathbf{x},\boldsymbol{\theta})^2\right] - E\left[c(\mathbf{x},\boldsymbol{\theta})\right]^2}}, \quad e = 1,\ldots,N_e\end{aligned}. \tag{5.45}$$

All expected values required within the above expressions are obtained through stochastic simulation employing again exterior sampling (as in the TMD optimization case) with ta otal of $N = 2000$ samples. Note that for the case corresponding to the multiplicative error or the moment-proportional additive error, the partial derivatives for the system variance, needed in the objective function Eq. (5.38), are transformed to

$$\frac{\partial \sigma_s(\mathbf{x})}{\partial x_e} = \frac{E\left[c(\mathbf{x},\boldsymbol{\theta})\frac{\partial c(\mathbf{x},\boldsymbol{\theta})}{\partial x_e}\right](1+\psi^2) - E\left[c(\mathbf{x},\boldsymbol{\theta})\right]E\left[\frac{\partial c(\mathbf{x},\boldsymbol{\theta})}{\partial x_e}\right]}{\sqrt{E\left[c(\mathbf{x},\boldsymbol{\theta})^2\right](1+\psi^2) - E\left[c(\mathbf{x},\boldsymbol{\theta})\right]^2}}, \quad e = 1,\ldots,N_e \quad (5.46)$$

5.3.2.2 Results and Discussion

In this example, the value of w was varied from 0 to 1 with equal steps of 0.2, and only the following three cases are considered:

i. No prediction error, termed as NE.

ii. Additive prediction error with variance scaled according to Eq. (5.16) and ψ chosen as 10%. This case is termed as AE.

iii. Multiplicative prediction error with ψ chosen as 5%. This case is termed as ME.

The resultant optimal solutions were very similar for the NE / AE and ME cases (almost identical topologies); thus, the various results are primarily reported for the former only. The optimal topologies for different values of w are shown in Figure 5.7 and the statistics (mean and standard deviation) in Table 5.8. A number of 5000 samples were used for the statistics reported in this table in order to provide a higher

accuracy comparison. Figure 5.8 shows the Pareto front in the normalized (with respect to utopia point) objective space F for all three design cases considered. The arrow in this figure represents the closest distance of any design from the utopia point. The closest point corresponds to the $w = 1$ design for all cases, and it would have been the preferred design based on this traditional approach for evaluation of appropriateness among RDO design.

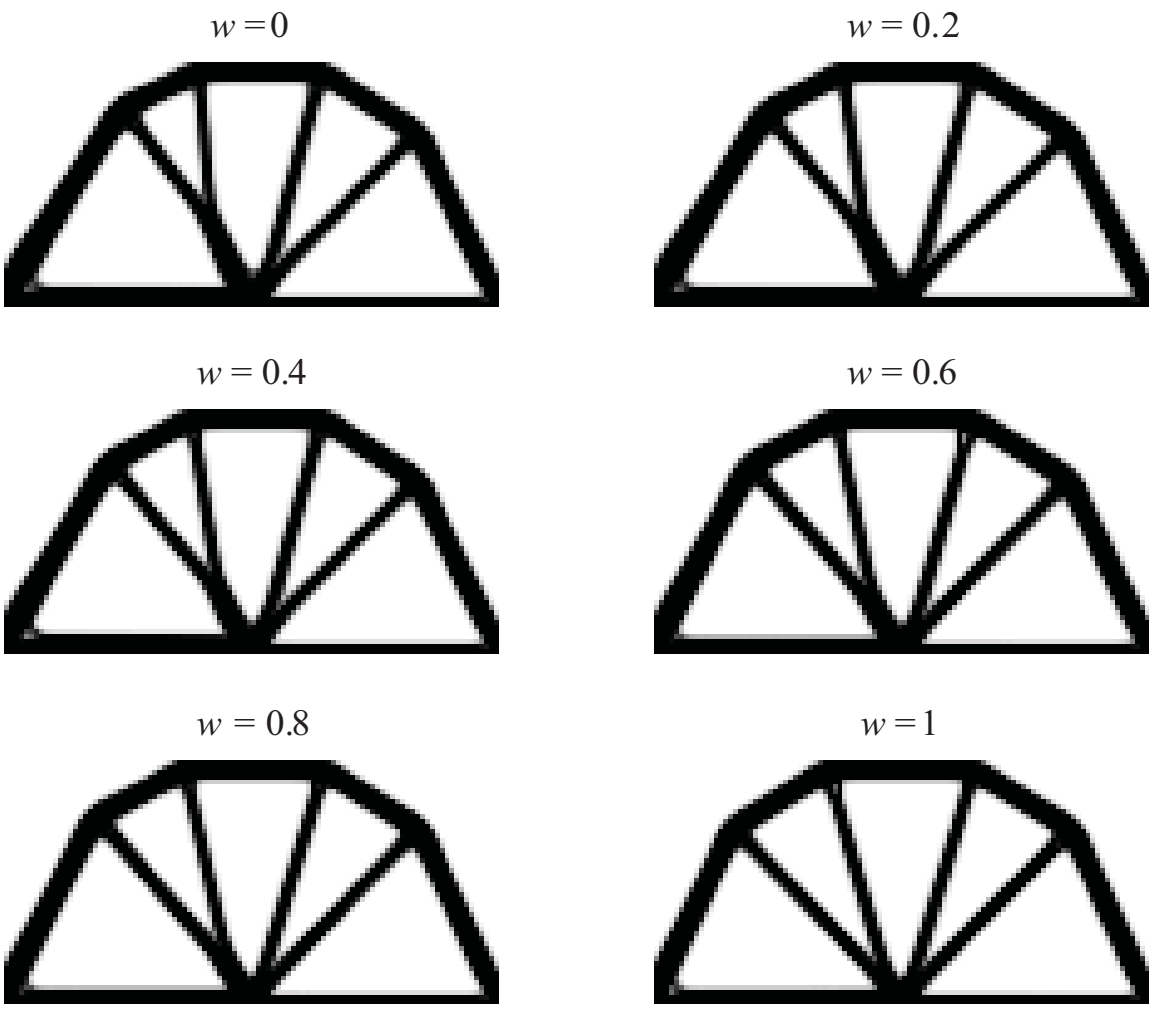


Figure 5.7 Optimal topologies for different values of w for the NE/AE designs

TABLE 5.8

MEAN AND STANDARD DEVIATION OF OPTIMAL SOLUTION FOR THE NE CASE FOR THE TOPOLOGY OPTIMIZATION PROBLEM

w	$\mu(\mathbf{x}_k)$	$\sigma(\mathbf{x}_k)$
0	39.01	29.85
0.2	39.03	29.75
0.4	39.08	29.66
0.6	39.32	29.59
0.8	39.40	29.50
1	39.44	29.30

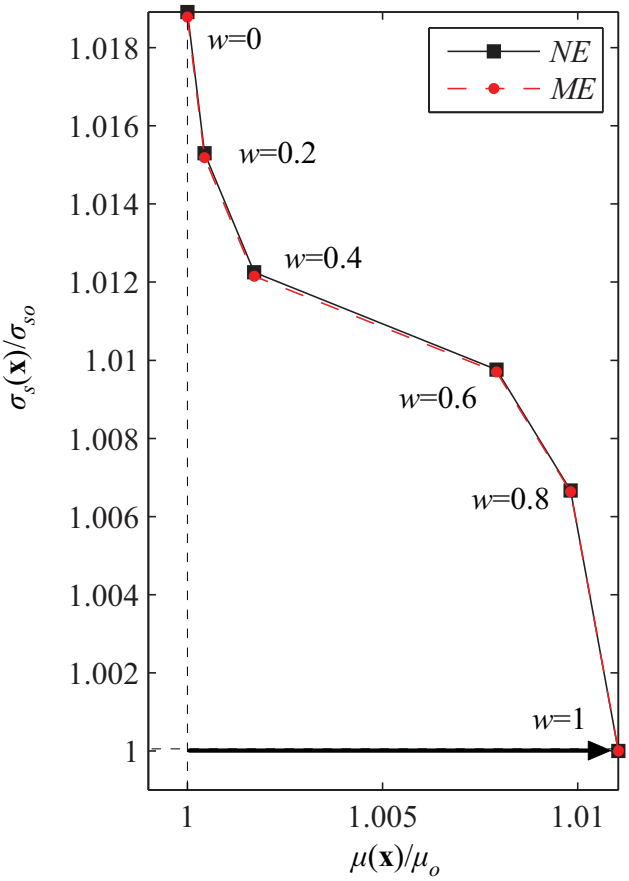


Figure 5.8 Pareto-front for the Topology RDO implementation. Closest design to utopia point indicated with an arrow.

These results agree with the ones presented in (Chen et al. 2010), where asymmetry is observed in the optimum designs due to asymmetric boundary conditions. It is possible to observe that the different designs have a similar statistical response even though they correspond to distinct topologies. In fact, the values for μ and σ between the different designs are so similar that it would be very challenging to select a preferred design from the set solely based on these statistics. Thus, the probability of dominance, which in this case corresponds to the probability of having a lower compliance over Θ, provides a rational measure for assessing the appropriateness of the candidate designs. For the correlation of the prediction error needed to calculate this probability (and similar to the TMD example), the exponential function in Eq. (5.12) is adopted, with normalization weight ν_l selected as one and decay rate λ selected so that correlation-decay is equal to $\rho_o = 5\%$ when the total change in the weighted distance between designs reaches $\alpha = 30\%$ [as described by Eq. (5.41)]. This assumption leads then to the following correlation matrix ρ_{im} between the different designs for the AE design case (similar result holds for the *ME* design),

$$\rho_{km} = \begin{bmatrix} 1 & 0.63 & 0.34 & 0.11 & 0.07 & 0.03 \\ 0.63 & 1 & 0.67 & 0.19 & 0.11 & 0.05 \\ 0.34 & 0.67 & 1 & 0.37 & 0.20 & 0.06 \\ 0.11 & 0.19 & 0.37 & 1 & 0.74 & 0.17 \\ 0.07 & 0.11 & 0.20 & 0.74 & 1 & 0.27 \\ 0.03 & 0.05 & 0.06 & 0.17 & 0.27 & 1 \end{bmatrix}. \tag{5.47}$$

The probability of dominance over Γ as well as the degree of dominance for 2 different values of $d = \{2,3\}$ are shown in Table 5.9. As in the TMD example, rather than

just identifying the dominant design based on the procedure outlined in Section 3.2, the degree of dominance for different d values is reported to facilitate a more thorough comparison.

TABLE 5.9

PROBABILITY AND DEGREE OF DOMINANCE FOR NE, AE, AND ME FOR THE TOPOLOGY OPTIMIZATION PROBLEM

	NE			NE			ME		
w	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$	$\hat{P}_D(\mathbf{x}_k \mid \Gamma)$	$D_2(\mathbf{x}_k)$	$D_3(\mathbf{x}_k)$
0	0.371	0.030	-0.005	0.207	0.005	-0.043	0.267	0.055	0.065
0.2	0.009	-0.030	-0.47	0.143	-0.005	-0.126	0.158	-0.055	-0.162
0.4	0.164	-0.084	-0.280	0.161	-0.016	-0.081	0.163	-0.092	-0.126
0.6	0.162	-0.174	-0.362	0.140	-0.060	-0.174	0.122	-0.202	-0.297
0.8	0.002	-0.214	-0.523	0.140	-0.072	-0.173	0.115	-0.227	-0.307
1	0.290	-0.180	-0.218	0.207	-0.076	-0.095	0.172	-0.237	-0.249

For the NE case, design $w = 0$ emerges as the dominant one when looking at the entire set Γ, with that preference also upheld over $d = 2$ (S^2) subsets and $d = 3$ (S^3) dimensional subsets. For the $d = 3$ case, no design actually emerges as dominant (values of $D_3(\mathbf{x}_k)$ are negative for all of them, for the $w = 0$ design the value of $D_3(\mathbf{x}_k)$ is very close to zero, and much larger than all other designs, indicating a strong

preference for it, especially when combined with the information for dominance over Γ and over the two-dimensional subsets). It is interesting to note that design $w = 0.2$ is almost dominant over the two-dimensional subsets (value of $D_2(\mathbf{x}_k)$ only slightly negative) but has poor performance when looking at the three-dimensional subsets or at the entire set Γ. This stresses the importance of looking at different subsets to evaluate the preference over a specific design, especially when the dominance properties it exhibits have a narrow margin (values close to zero). When incorporating the additive prediction error in the analysis (AE case in Table 5.9), we see significant differences in the calculated probabilities. This should be attributed to the fact that the compared designs have very similar statistical characteristics (as shown in Table 5.8). Design $w = 0$ still emerges as the preferred one when looking at lower dimensional subsets ($d = \{2,3\}$, nothing though that for $d = 3$ there is no dominance) but with much smaller margins compared to the NE case, whereas when looking at the entire set Γ, it practically ties with the $w = 1$ design. Changing the probability model for the prediction error to multiplicative influence (ME case) changes its overall impact. Though the prediction error still has a strong influence towards the calculated probability of dominance (changes over the NE case), the relative impact as compared to the AE case changes without a change in the overall preference ($w = 0$ is still the dominant design with actually increased robustness margins when looking at the lower dimensional subsets). This validates the previous conclusions that the prediction error model may have an impact on the results.

Finally, it is also interesting to note that, like the TMD example, in no case examined does the design selected through the degree of dominance coincide with the one corresponding to the minimum distance from the utopia point ($w = 1$ in all cases as previously discussed). Though the $w = 1$ design has competitive values when considering the probability of dominance over the entire set Γ, for some cases considered it corresponds to rather small values for the probability of dominance for lower values of d and would have been avoided through the proposed approach. This further demonstrates the utility of the new methodology for assessing the robustness characteristics of different proposed designs.

5.4 Summary

In this chapter, a novel framework to assess the appropriateness of robust designs was proposed within RDO problems. Traditionally, RDO employs the mean and standard deviation of a response function to obtain robust designs with different weightings of these competitive objectives, ultimately resulting in different optimal designs. The approach is based on a new robustness measure, termed probability of dominance. This measure is defined as the likelihood that each design will outperform the rival designs within the candidate set. Given that the resultant probability is a function of the exact designs composing the set, a multi-stage approach was also formulated for enhancing the robustness of the chosen solution. This approach compares designs within smaller dimensional subsets and searches for the design that dominates within all subsets that include it. The latter is ultimately characterized by a

positive value of the degree of dominance, with that degree defined by comparing the probability of dominance for the candidate design against the probability of dominance of all other designs within all different subsets to which the candidate designs belongs to.

The impact of prediction errors between the real system and the assumed numerical model was also investigated. Two different error models were explored; these correspond to either an additive or multiplicative influence, and reasonable assumptions for selecting the error characteristics were discussed in detail. The additive prediction error can be ultimately chosen to address two different sources of modeling uncertainty. It can be selected to address only un-modeled uncertainties when its variance is scaled according to the response variance of the assumed numerical model. If this assumption is selected, the formulation of the RDO problem is not impacted at all. This error can also be also selected to address both un-modeled uncertainties and un-modeled system characteristics when its variance is scaled according to the second moment of the numerical model. This in turn is equivalent to the multiplicative prediction error in terms of how it impacts the statistics of the real system performance and ultimately the RDO formulation. In addition, it was shown that when modeling errors are considered, the impact on the probability of dominance corresponds to substitution of the indicator function, describing the favorability of each design, by a likelihood function dependent upon the assumed probability model of the error and the model performance.

Two RDO illustrative examples were presented, the first considering the design of a tuned mass damper (TMD) for vibration mitigation of harmonic excitations and the second a topology optimization for minimum compliance. It was shown in these examples that the new approach provides an alternative to the established popular methodologies for assessing the appropriateness of candidate designs. Thus it ultimately equips designers with an additional tool. The results also showed that looking solely at the probability of dominance (over the entire set) does not always lead to the most appropriate choice. Comparison among smaller dimensional subsets can guide decisions better and further facilitate a deeper understanding of the preference toward each design, even when none emerges as the dominant one. It was also demonstrated that beyond the degree of dominance being positive, its value should also be taken into account by the designer in assessing the preferences for different designs. Small values for the degree of dominance should be carefully examined by the designer when evaluating the different designs. Moreover, it was shown that explicitly including the prediction error in the comparison can have a significant impact on the assessment. In this case, the assumed correlation of the prediction error between different designs needs to be carefully considered. The influence of the prediction error is especially critical when the compared designs have similar statistical performance despite corresponding to different design configurations. In general, the proposed method showed promising results and provided a rational framework for assessing the appropriateness of a set of candidate robust designs.

CHAPTER 6:

CONCLUSIONS AND FUTURE WORK

6.1 Contributions of This Dissertation

This dissertation offered computational and theoretical advances for design optimization problems under uncertainty that utilize the probabilistic system performance as objective function. Special focus was placed on applications that involve potentially complex numerical and probability models, and a generalized approach was adopted treating the system model as a "black-box" and relying on stochastic simulation for evaluating the probabilistic performance. Two generalized goals were identified: (a) to improve the efficiency of stochastic simulation techniques when implemented within the context of numerical optimization algorithms relying on local searches, and (b) to provide a novel robustness measure to assess the appropriateness of a set of candidate designs established within the RDO formulation while addressing the existence of prediction error between model and system performance.

To achieve these generalized goals, three distinct research objectives were set to be accomplished:

1) Formulate an adaptive importance sampling framework for design under uncertainty optimization problems. This framework improves the efficiency of

stochastic-simulation-based evaluations of the objective function by sharing information across the iterations of the numerical optimization algorithm.

2) Establish an adaptive surrogate modelling methodology with a similar goal, utilizing specifically a kriging metamodel. Consider a novel implementation of the surrogate model in the augmented model parameters and design variables space and develop efficient techniques for sharing of information between iterations to improve the accuracy of the developed metamodel.

3) Offer a new measure for assessing the appropriateness of different RDO designs and examine the impact of model prediction errors on this measure as well as the RDO problem itself.

The first two objectives shared a common foundation and directly addressed the goal (a), while the third objective addressed goal (b). The common foundation was established by sharing information across the iterations of the optimization algorithm to improve computational efficiency of the stochastic-simulation-based evaluation of the objective function. Two main tasks were considered within this foundation: (i) *generation of important samples* from a density proportional to the integrand, representing the domain in the uncertain model parameter space that has larger contribution to the integrand representing the probabilistic performance, (ii) *integration of a global probabilistic sensitivity analysis* that can identify the importance toward the overall probabilistic performance of different uncertain model parameters. Both tasks were accomplished with small additional computational burden, using the system model

evaluations readily available within the stochastic-simulation-based evaluation of the objective function

Chapter 2 of this dissertation introduced the simulation-based optimization under uncertainty problem in a general form as well as the popular RDO formulation. It also discussed some common computational tools that were utilized to achieve the required goals. Chapters 3-5 were devoted to achieve the three different research objectives, and their contributions are summarized in the following sections.

6.1.1 Optimization under Uncertainty with Adaptive Importance Sampling

In Chapter 3, which discussed the first research objective, an adaptive Importance Sampling (IS) formulation was introduced. It uses the *important samples* to formulate IS densities utilizing Kernel Density Estimation (KDE) and considers an explicit optimization of the KDE characteristics using model response evaluations readily available from the previous iteration of the optimization algorithm. The *global sensitivity analysis* was used to select the model parameters that span the IS density, avoiding the well-known problems in establishing IS densities for high dimensions. The novel contributions include:

- A complete adaptive framework was established for formulating IS densities by sharing information across the iterations of optimization algorithms relying on local search (so that design configurations examined in subsequent iterations are not drastically different and one can provide information for the subsequent). Rules and guidelines for this sharing of information were established and a

robust approach was established for the adaptive selection of the number of samples to achieve a predetermined accuracy level at each iteration.

- A probabilistic global sensitive analysis was integrated within this framework to provide a prioritization of the importance for the different model parameters. This prioritization was employed to determine on what model parameters IS should focus. Incorporating this information can help circumvent numerical problems that arise when trying to formulate IS for a large number of model parameters.
- A new approximation for the expected coefficient of variation in subsequent iterations was derived for any considered IS density. This approximation was established using readily available model response evaluations and was used to optimize for the characteristics of KDE proposal densities as well as for selecting the optimal number of parameters to consider in the IS formulation. The aforementioned probabilistic sensitivity analysis and relative prioritization of the model parameters were used to reduce the computational cost of this optimization. This concurrent optimization of density characteristics and number of model parameters provides a powerful framework for rational selection of optimal IS densities.
- A robust formulation of the aforementioned optimization was established for applications in which not enough information is available to obtain an accurate approximation for the expected value of the coefficient of variation (i.e. when a greedy optimization approach is adopted tolerating a large coefficient of

variation and as such, a reduced number of samples of the system model response). Higher order statistics (variance) was considered for this purpose. This formulation ultimately offers enhanced robustness in selection of optimal IS densities.

In the illustrative example considered, it was shown that the proposed adaptive IS framework can provide significant computational savings since it leads to an adaptive selection of IS densities utilizing readily available information. The robust formulation can avoid vulnerabilities in identifying erroneous optimal IS characteristics because of numerical errors associated with the approximation for the expected coefficient of variation when limited samples are available, whereas the adaptive optimal selection of parameters for the IS offers enhanced robustness and efficiency when compared to establishing IS for all model parameters.

The framework has ultimately no constraints as it can be implemented in problems with an arbitrary number of model parameters and design variables (as long as an appropriate numerical optimization algorithm is adopted to address the latter).

6.1.2 Optimization under Uncertainty with Adaptive Kriging in Augmented Space

Chapter 4 offered an adaptive implementation of a surrogate modeling within the same context considered in Chapter 3. In particular, a kriging metamodel was adopted due to its ability to efficiently approximate complex functions, offer estimates for the local error variance and provide gradient information. The formulation was established in the augmented model parameter and design variable space. The

proposed framework used the *important samples* to approximate the target region for the selection of support points (Design of Experiments – DoE) with no additional computational effort or requirement to establish an adaptive DoE. The *global sensitivity analysis* was used to select the basis functions for the kriging implementation; higher order basis functions were considered only for the more important components of the augmented input vector. The prediction error associated with the kriging metamodel was also explicitly considered in the formulation of the objective function and the design optimization.

The novel contributions are:

- A formal framework was established for creating a kriging metamodel in the augmented input space for generalized design under uncertainty problems. This pertains to rules for both the kriging formulation as well as the optimization for the design variables using the derived metamodel. Within this framework the metamodel is utilized to approximate the model response simultaneously with respect to both the uncertain model parameters and the design variable, with the latter belonging in some trust region (considered for obtaining the support points for the metamodel). This approximation is then used to estimate the objective function and its gradient for specific values of the design variables by performing a stochastic simulation with respect to the model parameters. This information is then utilized to establish a local search (within the selected trust region) for the optimal design variables, and only when this local search moves to a design configuration on the boundary of the trust region are new model

evaluations needed. The model response approximations for the latter design configuration are used to identify the important samples to guide decisions in the next iteration, which starts with the task of obtaining new support points and a new metamodel.

- A DoE approach for improving the accuracy in a target region was established without requiring any additional computational cost, as opposed to previous research efforts. Rather, the readily available model response evaluations are utilized to obtain important samples and approximate a distribution for this task. A hybrid sampling is also established by comparing this stage with an additional space-filling stage to provide an adequate approximation over the entire domain in the model parameter space.
- An adaptive methodology was established for selecting the order of the basis functions for the metamodel. This is accomplished, again, with no additional computational effort or cross-validation formulations, something that facilitates a novel approach for such a selection. The global sensitivity analysis is utilized for this purpose.
- The prediction error was explicitly considered in the design formulation, and the benefits from such a consideration were clearly demonstrated.

The same illustrative example as for the adaptive IS was considered, and the adaptive kriging was shown to offer remarkable computational savings, facilitating convergence to the optimal solution with a very small number of model evaluations (even when compared against the adaptive IS formulation). The DoE for the target

region and the inclusion of the prediction error were shown to provide increased robustness characteristics in identifying an optimal solution. It should be stressed that, contrary to the adaptive IS, this approach has a significant constraint since the number of the augmented input space (design variables and model parameters) can be only moderately large.

6.1.3 Probability of Dominance as Robustness Measure

In Chapter 5, an approach for assessing the appropriateness of a set of candidate robust designs within RDO was introduced. The approach is based on the definition of a new robustness measure, termed probability of dominance, given by the likelihood that a chosen design will outperform its rivals within a set of candidate designs under some given uncertain conditions. A multi-stage approach was also formulated for enhancing robustness of the chosen solution; this approach compares designs within smaller dimensional subsets and searches for the design that dominates within all subsets that include it. This ultimately leads to the concept of the degree of dominance, defined by comparing the probability of dominance for the candidate design against the probability of dominance of all other designs within all different subsets to which the candidate designs belongs. The impact of the prediction error between the real system and the assumed numerical model was also investigated in detail. Two different error models were investigated; these correspond to both an additive or multiplicative influence, and assumptions for selecting the error characteristics were discussed. The additive prediction error can ultimately be chosen to address two different sources of modeling

uncertainty. It can be selected to address only un-modeled uncertainties, when its variance is scaled according to the response variance of the numerical model. The error can also be also selected to address both un-modeled uncertainties and un-modeled system characteristics, when its variance is scaled according to the second moment of the numerical model. This in turn is equivalent to the multiplicative prediction error in terms of how it impacts the statistics of the real system performance. Moreover, the prediction error was incorporated into the RDO formulation and its repercussions investigated. The original contributions are:

- The robustness measure proposed corresponds to an entirely new approach for assessing appropriateness of designs and a measure that is directly related to the uncertainty description (probability space) that provided the parent RDO problem formulation.
- The multi-stage selection process facilitates a rational decision framework, offering the designer with an enhanced metric (the degree of dominance) for assessing the performance of the different candidate designs.
- An efficient computational approach was developed for calculating the degree of dominance and ultimately the preferred design.
- The prediction error was formally considered in the definition of the probability of dominance as well as for the formulation of the RDO problem itself. This ultimately facilitates enhanced robustness in assessing the appropriateness of each design but also explicitly addresses in the analysis/design an important source of uncertainty, the fact that the system model can never provide exact

predictions for the real model. Though this uncertainty is an integral component of Bayesian updating approaches, it had been typically ignored in analysis/design due to the challenges associated with its quantification.

- For addressing these challenges, an extensive discussion was presented for the rational selection of probability model and statistical characteristics for the model prediction error. The connection of these statistics to two different sources for this error was examined – un-modeled uncertainties and un-model system characteristics.

The framework was demonstrated with two different illustrative examples, from popular RDO applications. It was shown in these examples that the new approach provides an attractive alternative to the established popular methodologies for assessing the appropriateness of the candidate designs, equipping designers with an additional decision tool. The results demonstrated the importance of the multi-stage formulation (consideration of lower-dimensional sets for the comparison) for supporting decisions with robust characteristics. Furthermore, it was shown that explicitly including the prediction error in the comparison can have a significant impact on the assessment. The influence of the prediction error is especially critical when the compared designs have similar statistical performance despite corresponding to different design configurations.

6.2 Future Research

There are a number of potential extensions of the research efforts summarized below.

6.2.1 Optimization under Uncertainty with Adaptive Importance Sampling

The proposed adaptive IS focused on the integrand corresponding to the objective function for establishing the desired densities. Within the context of gradient-based optimization algorithms, an interesting extension corresponds to formulation of IS directly for the gradient (or its numerical approximation), since the latter is what is ultimately utilized by that algorithm. Of course, this task is fairly challenging to implement since each partial derivative needs to be separately treated, indicating that a separate IS density in needed for each component of the design vector, something that becomes impractical in problems with a large number of design variables. A potential remedy is to use the sensitivity analysis to determine a group of designs variables that have a higher contribution in the optimization and then focus on them for the IS. Either separate densities or a density that establishes a compromise can be considered.

6.2.2 Optimization under Uncertainty with Adaptive Kriging in Augmented Space

The proposed adaptive kriging implementation in the augmented space can provide significant computational savings; an important unaddressed question is how the number of the support points can also be adaptively selected. The answer to this question is challenging since it was shown that cross-validation techniques do not

provide accurate information for the performance of kriging metamodels in targeted domains of interest.

6.2.3 Probability of Dominance as Robustness Measure

The probability of dominance showed promising results and provided a rational framework for assessing the appropriateness of a set of candidate robust designs. It was shown that in some cases, the probability of dominance can identify a preferred design, while in other cases the distinction is not completely apparent. A closer examination in needed in order to determine the trustworthiness of the predictions from this framework. Moreover, it was shown that incorporation of some weight might be necessary for some applications, such as when the focus is on reliability of rare events. The exact characteristics of this weight implementation require further investigation.

APPENDIX A:

SIMULTANEOUS PERTURBATION STOCHASTIC APPROXIMATION (SPSA)

SPSA is based on the premise that one properly chosen simultaneous random perturbation in all components of $\mathbf{x}$ provides as much information for optimization purposes in the long run as a full set of finite differences. Thus, it uses only two evaluations of the objective function, in a direction randomly chosen at each iteration, to form an approximation to the gradient vector. It also uses the notion of stochastic approximation (Kushner and Yin 2003) which can significantly improve the computational efficiency of stochastic search applications. The latter approximation is performed by establishing (through proper recursive formulas) an equivalent averaging across the iterations of the algorithm.

At iteration k, corresponding to design $\mathbf{x}_k$, the implementation of SPSA takes the iterative form

$$\mathbf{x}_{k+1} = \mathbf{x}_k - a_k^{SP}\mathbf{g}_k\left(\mathbf{x}_k, \{\boldsymbol{\theta}^j\}_k\right), \tag{A.1}$$

where $\mathbf{x}_1 \in X$ is the chosen point to initiate the algorithm and the i^{th} component for the CRN simultaneous perturbation approximation to the gradient vector in the k^{th} iteration, and $\mathbf{g}_k(\mathbf{x}_k, \{\boldsymbol{\theta}^j\}_k)$ is given by

$$g_{k,i}\left(\mathrm{x}_k,\{\boldsymbol{\theta}_k^j\}\right)=\frac{\hat{H}\left(\mathbf{x}_k+c_k^{SP}\boldsymbol{\Delta}_k^{SP}\mid\{\boldsymbol{\theta}^j\}_k\right)-\hat{H}\left(\mathbf{x}_k-c_k^{SP}\boldsymbol{\Delta}_k^{SP}\mid\{\boldsymbol{\theta}^j\}_k\right)}{2c_k^{SP}\Delta_{k,i}^{SP}},\qquad(\mathrm{A.2})$$

where $\{\boldsymbol{\theta}^j\}_k$ is the sample set used at iteration k and $\boldsymbol{\Delta}_k^{SP}\subset\mathbb{R}^{n_x}$ is a vector of mutually independent random variables that defines the random direction of simultaneous perturbation for $\mathbf{x}$. A symmetric Bernoulli ±1 distribution is typically chosen for the components of $\boldsymbol{\Delta}_k^{SP}$, assuming that the components of $\mathbf{x}$ are properly normalized. Note that the set of random numbers $\{\boldsymbol{\theta}^j\}$ is selected the same for the two estimates of the objective function at each iteration used to calculate the gradient in Eq. (A.2). As discussed in Section 1.1.3, this reduces the variance of the difference of these estimates, thus creating a consistent estimation error (Taflanidis and Beck 2008; Glasserman and Yao 1992) and improving the accuracy of Eq. (A.2). However, this sample set is not the same across the iterations of the optimization algorithm (i.e. approach corresponding to interior sampling).

The selection of the sequences $\{c_k^{SP}\}$ and $\{a_k^{SP}\}$ for the SPSA algorithm using CRN is discussed in detail in (Kleinman et al. 1999). A choice that guarantees asymptotic convergence to $\mathbf{x}^*$ is $a_k^{SP}=a^{SP}/(k+A^{SP})^{\beta^{SP}}$ and $c_k^{SP}=c^{SP}/k^{\zeta^{SP}}$, where $4\zeta^{SP}-\beta^{SP}$, $\beta^{SP}-\zeta^{SP}>1$, with $\{A^{SP},\zeta^{SP}\}>0$ and $0<\beta^{SP}<1$. This selection leads to a rate of convergence that asymptotically approaches $k^{-\beta^{SP}/2}$ when CRN are used (Kleinman et al. 1999). The asymptotically optimal choice for β^{SP} is, thus, 1. For applications in which efficiency using a small number of iterations is sought, the use of smaller values for β^{SP}

are suggested (Spall 2003). Regarding the rest of the parameters for the sequences $\{c_k^{SP}\}$ and $\{a_k^{SP}\}$, *A* is typically set to $5\text{-}20\%$ of the maximum number of iterations selected for the algorithm, and the initial step c is chosen "close" to the standard deviation of the prediction error for the stochastic estimate in Eq. (2.4). The value of a can be determined based on the estimate of the gradient in the first step $\mathbf{g}_1$ and the desired step size for the first iteration. Some initial trials are generally needed in order to make a good selection for a^{SP}, especially when little prior insight is available for the sensitivity of the objective function to each of the design variables. Convergence of the iterative process is based on the value $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|$ of the last few steps for an appropriate selected vector norm. Note that contrary to deterministic optimization problems, convergence cannot be based on the objective function values in subsequent iterations due to the estimation error that makes this comparison impractical; since interior sampling is used, the random numbers used to obtain these two estimates are not the same, creating a significant potential impact of the estimation error on the intended comparison. As such, convergence needs to rely on comparisons of the design vectors. Additional stopping criteria related to the computational effort can be set, such as total allowed iterations of the algorithm or total allowed function calls to calculate the model response. Blocking rules can also be applied in order to avoid potential divergence of the algorithm, especially in the first iterations, and a proper normalization of the design vector $\mathbf{x}$ is strongly suggested (Spall 2003).

APPENDIX B:

DETERMINISTIC TOPOLOGY OPTIMIZATION

Topology optimization searches for the structural configuration (shape and connectivity of the structure) contained within a given design domain X that for some given boundary and loading conditions provides the most favorable response in terms of some chosen performance measure (Bendsøe and Sigmund 2003). A popular approach to obtain the solution is to discretize the domain into finite elements, e, and select the densities of each element as design variables $\mathbf{x}_e \in [0,1], e = 1, \cdots, N_e$ (typically interpolated between two states: either 0 (void) or 1 (full material)), where N_e corresponds to the total number of elements. The performance objective is typically taken as the minimization of the compliance $c(\mathbf{x})$ for a given structural domain, subject to a prescribed mass fraction constraint m_f such that $m_X \leq m_f$, where m_X denotes the domain's mass fraction, which can be computed as

$$m_X = \frac{1}{N_e} \sum_{e=1}^{N_e} x_e \,. \tag{B.1}$$

For a linear elastic material, the compliance is a function of the global displacement U (found through the finite element method) and the global stiffness matrix K (Bendsøe and Sigmund 2003). Here the problem is relaxed by employing the

Solid Isotropic Material with Penalization (SIMP) formulation (Rietz 2001), leading to the compliance function given by (Sigmund 2001)

$$c(\mathbf{x}) = U^T K U = \sum_{e}^{N_e} x_e^{p_p} u_e^T k_o u_e, \quad e = 1, ..., N_e, \tag{B.2}$$

where p_p is a penalization factor employed to avoid intermediate densities (i.e. to drive each element toward a particular state – either void or full material), u_e is the local displacement vector for a given element, and k_o is the element's stiffness matrix which in turn is a function of the poison ratio v and the modulus of elasticity E_e of the element.

The gradient of the compliance can be analytically obtained (Bendsøe and Sigmund 2003) as

$$\frac{\partial c(\mathbf{x})}{\partial x_e} = -p_p (x_e)^{p_p - 1} u_e^T k_o u_e, \quad e = 1, ..., N_e. \tag{B.3}$$

The solution to a deterministic topology optimization problem analogous to the problem examined in Chapter 5 is presented next, although it is not relevant to the discussion in that chapter. That problem deals with the topology optimization of a Michell-type structure where a rectangular design domain is subject to a central point load at the bottom with one of the lower corners fixed and the other one simply supported (Figure B.1 (a)). The solution to the problem is obtained with the Method of Moving Asymptotes (MMA) (Svanberg 1987), which is a popular algorithm to solve topology optimization problems, and the result is shown in B.1 (b). All the optimization

parameters such as the discretization, penalization factor, and filter radius are the same as in Chapter 5.

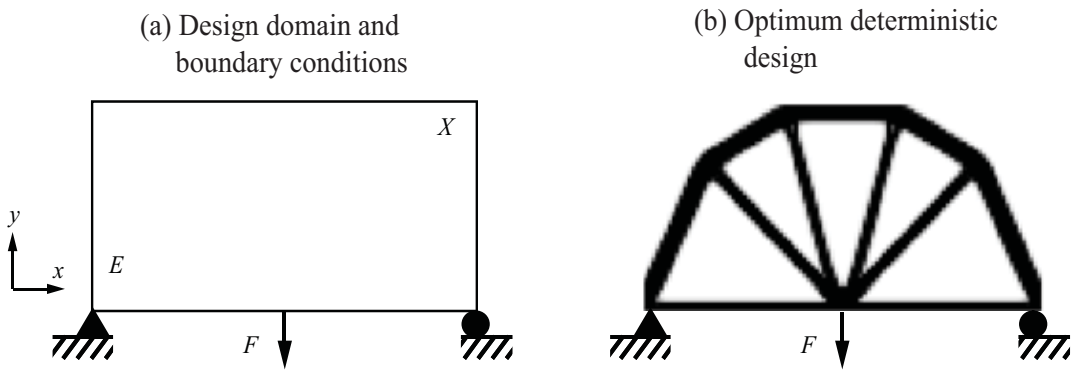


Figure B.1: Solution to the deterministic problem, a) design domain, loads, and supports. b) Optimum design. Discretization, penalization factor and filter size remain as in Chapter 5.

BIBLIOGRAPHY

Abramson, I. S. (1982). "On bandwidth variation in kernel estimates-a square root law." *The Annals of Statistics,* 1217-1223.

Ang, G. L. (1992). "Optimal importance-sampling density estimator." *J.Eng.Mech.,* 118(6), 1146.

Aoues, Y. (2008). "Reliability-based optimization of structural systems by adaptive target safety–Application to RC frames." *Struct.Saf.,* 30(2), 144.

Archer, G., Saltelli, A., and Sobol, I. (1997). "Sensitivity measures, ANOVA-like techniques and the use of bootstrap." *Journal of Statistical Computation and Simulation,* 58(2), 99-120.

Au, S. K., and Beck, J. L. (2003). "Important sampling in high dimensions." *Struct.Saf.,* 25(2), 139.

Au, S. K., and Beck, J. L. (1999). "A new adaptive importance sampling scheme for reliability calculations." *Struct.Saf.,* 21(2), 135.

Ayyub, B. M., and Gupta, M. M. (1997). *Uncertainty analysis in engineering and sciences: fuzzy logic, statistics, and neural network approach.* Springer, .

Baumal, A. E., McPhee, J. J., and Calamai, P. H. (1998). "Application of genetic algorithms to the design optimization of an active vehicle suspension system." *Comput.Methods Appl.Mech.Eng.,* 163(1–4), 87-94.

Beck, J. L. (1999). "Multi-criteria optimal structural design under uncertainty." *Earthquake Engineering Structural Dynamics,* 28(7), 741.

Beck, J. L. (2010). "Bayesian system identification based on probability logic." *Structural Control and Health Monitoring,* 17(7), 825-847.

Beck, J. L., and Taflanidis, A. A. (2013). "Prior and posterior robust stochastic predictions for dynamical systems using probability logic." *International Journal for Uncertainty Quantification,* 3(4),.

Beck, A., and Gomes, W. J. d. S. (2012). "A comparison of deterministic, reliability-based and risk-based structural optimization under uncertainty." *Prob.Eng.Mech.,* 28 18-29.

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). "Sequential design of computer experiments for the estimation of a probability of failure." *Statistics and Computing,* 22(3), 773-793.

Beirlant, J., Dudewicz, E. J., Györfi, L., and Van der Meulen, E. C. (1997). "Nonparametric entropy estimation: An overview." *International Journal of Mathematical and Statistical Sciences,* 6 17-40.

Bendsøe, M. P., and Sigmund, O. (2003). *Topology optimization: theory, methods, and applications.* Springer Verlag, .

Beyer, H., and Sendhoff, B. (2007). "Robust optimization–a comprehensive survey." *Comput.Methods Appl.Mech.Eng.,* 196(33), 3190-3218.

Cannamela, C., Garnier, J., and Bertrand Iooss. (2008). "Controlled Stratification for Quantile Estimation." *The Annals of Applied Statistics,* 2(4), 1554-1580.

Cappé, O., Douc, R., Guillin, A., Marin, J. M., and Robert, C. (2008). "Adaptive importance sampling in general mixture classes." *Statistics and Computing,* 18(4), 447-459.

Chang, C. (1999). "Mass dampers and their optimal designs for building vibration control." *Eng.Struct.,* 21(5), 454-463.

Chen, S., Chen, W., and Lee, S. (2010). "Level set based robust shape and topology optimization under random field uncertainties." *Structural and Multidisciplinary Optimization,* 41(4), 507-524.

Chiralaksanakul, A. (2005). "First-order approximation methods in reliability-based design optimization." *Journal of Mechanical Design,* 127(5), 851.

Choi, K., Youn, B. D., and Yang, R. (2001). "Moving least square method for reliability-based design optimization." *4th World congress of structural and multidisciplinary optimization, Dalian, China,* .

Coleman, T., Branch, M. A., and Grace, A. (1999). *Optimization Toolbox for Use with MATLAB: User's Guide, Version 2.* Math Works, Incorporated, .

Dahlberg, T. (1979). "Optimization Criteria for Vehicles Travelling on a Randomly Profiled Road-a Survey." *Veh.Syst.Dyn.,* 8(4), 239-352.

Deb, K., and Gupta, H. (2006). "Introducing robustness in multi-objective optimization." *Evol.Comput.,* 14(4), 463-494.

Der Kiureghian, A., Lin, H., and Hwang, S. (1987). "Second-order reliability approximations." *J.Eng.Mech.,* 113(8), 1208-1225.

Der Kiureghian, A., and Ditlevsen, O. (2009). "Aleatory or epistemic? Does it matter?" *Struct.Saf.,* 31(2), 105-112.

Doltsinis, I. (2004). "Robust design of structures using optimization methods." *Comput.Methods Appl.Mech.Eng.,* 193(23-26), 2221.

Dubourg, V., Sudret, B., and Bourinet, J. (2011a). "Reliability-based design optimization using kriging surrogates and subset simulation." *Structural and Multidisciplinary Optimization,* 44(5), 673-690.

Dubourg, V., Deheeger, F., and Sudret, B. (2011b). "Metamodel-based importance sampling for the simulation of rare events." *ArXiv e-Prints: 1104.3476,* .

Dunning, P. D., Kim, H. A., and Mullineux, G. (2011). "Introducing Loading Uncertainty in Topology Optimization." *AIAA J.,* 49(4), 760.

Dunning, P. D., and Kim, H. A. (2013). "Robust Topology Optimization: Minimization of Expected and Variance of Compliance." *AIAA J.,* 51(11), 2656-2664.

Eldred, M. S., Giunta, A. A., Wojtkiewicz Jr, S. F., and Trucano, T. G. (2002). "Formulations for surrogate-based optimization under uncertainty." *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Atlanta, GA,* .

Enevoldsen, I., and Sørensen, J. D. (1994). "Reliability-based optimization in structural engineering." *Struct.Saf.,* 15(3), 169-196.

Epanechnikov, V. A. (1969). "Non-parametric estimation of a multivariate probability density." *Theory of Probability & its Applications,* 14(1), 153-158.

Fang, J., Smith, S. M., and Elishakoff, I. "Structural Optimization under Uncertainty." *Mathematical Problems in Engineering,* 4 I87-200.

Frangopol, D. M. (1985). "Structural optimization using reliability concepts." *J.Struct.Eng.,* 111(11), 2288.

Gano, S. (2005). "Simulation-based design using variable fidelity optimization".

Gao, W., Zhang, N., and Du, H. P. (2007). "A half-car model for dynamic analysis of vehicles with random parameters." *5th Australasian Congress on Applied Mechanics, ACAM,* .

Gao, W., Zhang, N., Ji, J. C., and Du, H. P. (2008). "Dynamic analysis of vehicles with uncertainty." *Proc.Inst.Mech.Eng.Pt.D: J.Automobile Eng.,* 222(5), 657-664.

Genz, A. (1992). "Numerical Computation of Multivariate Normal Probabilities." *Journal of Computational and Graphical Statistics,* 1(2), 141-149.

Glasserman, P., and Yao, D. (1992). "Some guidelines and guarantees for common random numbers." *Management Science,* 38(6), 884-908.

Helton, J. C. (1997). "Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty." *Journal of Statistical Computation and Simulation,* 57(1-4), 3-76.

Holmström, K., Göran, A., and Edvall, M. (2006). "User's Guide for TOMLAB 7." *Tomlab Optimization Inc, Www.Tomopt.com/tomlab/, San Diego, CA,* .

Huang, D., Allen, T. T., Notz, W. I., and Zeng, N. (2006). "Global optimization of stochastic black-box systems via sequential kriging meta-models." *J.Global Optimiz.,* 34(3), 441-466.

James, B. A. P. (1985). "Variance Reduction Techniques." *J.Oper.Res.Soc.,* 36(6), 525-530.

Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge university press, .

Jensen, H. A., Valdebenito, M. A., Schuëller, G. I., and Kusanovic, D. S. (2009). "Reliability-based optimization of stochastic systems using line search." *Comput.Methods Appl.Mech.Eng.,* 198(49–52), 3915-3924.

Jia, G., and Taflanidis, A. A. (2013). "Non-parametric stochastic subset optimization for optimal-reliability design problems." *Comput.Struct.,* .

Jia, G., and Taflanidis, A. A. (2011). "Relative entropy estimation through stochastic sampling and stochastic simulation techniques." *Proceedings of the Second International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering, Chania, Greece,* 6-9.

Jia, G., and Taflanidis, A. A. (2013). "Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment." *Comput.Methods Appl.Mech.Eng.,* 261–262(0), 24-38.

Jin, R., Chen, W., and Simpson, T. W. (2001). "Comparative studies of metamodelling techniques under multiple modelling criteria." *Structural and Multidisciplinary Optimization,* 23(1), 1-13.

Jin, R., Du, X., and Chen, W. (2003). "The use of metamodeling techniques for optimization under uncertainty." *Structural and Multidisciplinary Optimization,* 25(2), 99-116.

Jin, Y., and Branke, J. (2005). "Evolutionary optimization in uncertain environments-a survey." *Evolutionary Computation, IEEE Transactions on,* 9(3), 303-317.

Johnson, R. A., and Wichern, D. W. (2002). *Applied multivariate statistical analysis.* Prentice hall Upper Saddle River, NJ, .

Jones, D. R. (2001). "A taxonomy of global optimization methods based on response surfaces." *J.Global Optimiz.,* 21(4), 345-383.

Kareem, A., and Kline, S. (1995). "Performance of multiple mass dampers under random loading." *J.Struct.Eng.,* 121(2), 348-361.

Karunamuni, R. J., and Zhang, S. (2008). "Some improvements on a boundary corrected kernel density estimator." *Statistics & Probability Letters,* 78(5), 499-507.

Kharmanda, G., Mohamed, A., and Lemaire, M. (2002). "Efficient reliability-based design optimization using a hybrid space with application to finite element analysis." *Structural and Multidisciplinary Optimization,* 24(3), 233-245.

Klee, H. (2007). *Simulation of dynamic systems with MATLAB and Simulink.* CRC Press, Inc., Boca Raton, FL.

Kleijnen, J. (1987). "Statistical tools for simulation practitioners." *Marcel Decker, New York,* .

Kleinman, N. L., Spall, J. C., and Naiman, D. Q. (1999). "Simulation-based optimization with stochastic approximation using common random numbers." *Journal of Management Science,* 45(11), 1570-1578.

Kroese, D. P., Taimre, T., and Botev, Z. I. (2011). *Handbook of Monte Carlo Methods.* Wiley, .

Kushner, H. J., and Yin, G. (2003). *Stochastic approximation and recursive algorithms and applications.* Springer Verlag, New York.

Kwon, S., and Park, K. (2004). "Suppression of bridge flutter using tuned mass dampers based on robust performance design." *J.Wind Eng.Ind.Aerodyn.,* 92(11), 919-934.

Lagaros, N. D. (2002). "Structural optimization using evolutionary algorithms." *Computers Structures,* 80(7-8), 571.

Law, A. M., Kelton, W. D., and Kelton, W. D. (1991). *Simulation modeling and analysis.* McGraw-Hill New York, .

Lee, K., Eom, I., Park, G., and Lee, W. (1996). "Robust design for unconstrained optimization problems using the Taguchi method." *AIAA J.,* 34(5), 1059-1063.

Lee, K., and Park, G. (2001). "Robust optimization considering tolerances of design variables." *Comput.Struct.,* 79(1), 77-86.

Lee, K., Park, G., and Joo, W. (2006). "A global robust optimization using kriging based approximation model." *JSME International Journal, Series C: Mechanical Systems Machine Elements & Manufacturing,* 49(3), 779-803.

Lophaven, S. N., Nielsen, H. B., and Søndergaard, J. (2002a). "Aspects of the matlab toolbox DACE." *Rep. No. IMM-REP-2002-13,* Informatics and Mathematical Modelling, Technical University of Denmark, DTU, .

Lophaven, S. N., Nielsen, H. B., and Søndergaard, J. (2002b). "DACE - A MATLAB Kriging toolbox." *Rep. No. IMM-TR-2002-12,* Informatics and Mathematical Modeling, Technical University of Denmark.

Loweth, E. L., De Boer, G. N., and Toropov, V. V. (2011). "Practical recommendations on the use of moving least squares metamodel building." *Proceedings of the Thirteenth International Conference on Civil, Structural and Environmental Engineering Computing,* Civil-Comp Press, Stirlingshire, UK, .

Marano, G. C., Sgobba, S., Greco, R., and Mezzina, M. (2008). "Robust optimum design of tuned mass dampers devices in random vibrations mitigation." *J.Sound Vibrat.,* 313(3), 472-492.

Marler, R. T., and Arora, J. S. (2004). "Survey of multi-objective optimization methods for engineering." *Structural and Multidisciplinary Optimization,* 26(6), 369-395.

Martinez, W. L., and Martinez, A. R. (2007). *Computational Statistics Handbook with MATLAB.* Chapman & Hall/CRC, Boca Raton, FL.

McLachlan, G., and Peel, D. (2004). *Finite mixture models.* John Wiley & Sons, New York, NY.

Melchers, R. E. (1989). "Importance sampling in structural systems." *Struct.Saf.,* 6(1), 3.

Melchers, R. E.,. (1999). *Structural reliability : analysis and prediction, 2nd ed.* Wiley, New York, NY.

Mohtat, A., and Dehghan-Niri, E. (2011). "Generalized framework for robust design of tuned mass damper systems." *J.Sound Vibrat.,* 330(5), 902-922.

Mugdadi, A. R., and Ahmad, I. A. (2004). "A bandwidth selection for kernel density estimation of functions of random variables." *Comput.Stat.Data Anal.,* 47(1), 49-62.

Papadimitriou, C., and Lombaert, G. (2012). "The effect of prediction error correlation on optimal sensor placement in structural dynamics." *Mechanical Systems and Signal Processing,* 28(0), 105-127.

Papadrakakis, M., and Lagaros, N. D. (2002). "Reliability-based structural optimization using neural networks and Monte Carlo simulation." *Comput.Methods Appl.Mech.Eng.,* 191(32), 3491-3507.

Park, G., Lee, T., Lee, K. H., and Hwang, K. (2006). "Robust design: an overview." *AIAA J.,* 44(1), 181-191.

Park, J. (1994). "Optimal Latin-hypercube designs for computer experiments." *Journal of Statistical Planning and Inference,* 39(1), 95-111.

Persson, J. A., and Ölvander, J. (2013). "Comparison of Different Uses of Metamodels for Robust Design Optimization." *51st AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition,* .

Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., and Kim, N. (2010). "Adaptive designs of experiments for accurate approximation of a target region." *Journal of Mechanical Design,* 132(7), 071008.

Polak, E. (2008). "Efficient sample sizes in stochastic nonlinear programming." *J.Comput.Appl.Math.,* 217(2), 301.

Porter, K., Kennedy, R., and Bachman, R. (2007). "Creating fragility functions for performance-based earthquake engineering." *Earthquake Spectra,* 23(2), 471-489.

Pradlwarter, H. J. (2007). "Application of line sampling simulation method to reliability benchmark problems." *Struct.Saf.,* 29(3), 208.

Rakheja, S., Afework, Y., and Sankar, S. (1994). "An analytical and experimental investigation of the driver-seat-suspension system." *Veh.Syst.Dyn.,* 23(1), 501-524.

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

Rietz, A. (2001). "Sufficiency of a finite exponent in SIMP (power law) methods." *Structural and Multidisciplinary Optimization,* 21(2), 159.

Rill, G. (2011). *Road Vehicle Dynamics: Fundamentals and Modeling.* CRC Press, .

Robert, C. P., and Casella, G. (2004). *Monte Carlo Statistical Methods.* Springer, New York.

Robson, J. D. (1979). "Road surface description and vehicle response." *Int.J.Veh.Des.,* 1(1), 25-35.

Rodrigues, J., Renaud, J. E., and Watsen, L. (1998). "Convergence of trust region augmented Lagrangian methods using variable fidelity approximation data." *Struct.Opt.,* 15(3-4), 141-156.

Rodríguez, J. F., Renaud, J. E., Wujek, B. A., and Tappeta, R. V. (2000). "Trust region model management in multidisciplinary design optimization." *J.Comput.Appl.Math.,* 124(1), 139-154.

Royset, J. O. (2006). "Optimal design with probabilistic objective and constraints." *J.Eng.Mech.,* 132(1), 107.

Royset, J. O. (2004). "Reliability-based optimal design using sample average approximations." *Prob.Eng.Mech.,* 19(4), 331.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). "Design and analysis of computer experiments." *Statistical Science,* 4(4), 409-423.

Schuëller, G. I. (2004). "A critical appraisal of reliability estimation procedures for high dimensions." *Prob.Eng.Mech.,* 19(4), 463.

Schuëller, G. I., and Jensen, H. A. (2008). "Computational methods in optimization considering uncertainties–an overview." *Comput.Methods Appl.Mech.Eng.,* 198(1), 2-13.

Scott, D. W., and Sain, S. R. (2005). "Multi-dimensional density estimation." *Handbook of Statistics,* 24 229-261.

Scott, D. W. (1992). *Multivariate density estimation: theory, practice, and visualization.* John Wiley & Sons, Inc, New York, NY.

Shan, S., and Wang, G. G. (2010). "Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions." *Structural and Multidisciplinary Optimization,* 41(2), 219-241.

Shapiro, A. (2003). "Monte Carlo sampling approach to stochastic programming." *ESAIM Proceedings,* 13 65.

Sharp, R., and Crolla, D. (1987). "Road vehicle suspension system design-a review." *Veh.Syst.Dyn.,* 16(3), 167-192.

Sigmund, O. (2001). "A 99 line topology optimization code written in Matlab." *Structural and Multidisciplinary Optimization,* 21(2), 120-127.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* Chapman & Hall/CRC, .

Simpson, T. W., Lin, D. K., and Chen, W. (2001a). "Sampling strategies for computer experiments: design and analysis." *International Journal of Reliability and Applications,* 2(3), 209-240.

Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F. (2001b). "Kriging models for global approximation in simulation-based multidisciplinary design optimization." *AIAA J.,* 39(12), 2233-2241.

Simpson, T., Peplinski, J., Koch, P., and Allen, J. (2001c). "Metamodels for Computer-based Engineering Design: Survey and recommendations." *Engineering with Computers,* 17 129-150.

Sobol, I. M. (2001). "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates." *Math.Comput.Simul.,* 55(1–3), 271-280.

Spall, J. C. (2003). *Introduction to stochastic search and optimization: estimation, simulation, and control.* Wiley-Interscience, New York.

Spall, J. C. (1998a). "Implementation of the simultaneous perturbation algorithm for stochastic optimization." *IEEE Trans.Aerospace Electron.Syst.,* 34(3), 817.

Spall, J. C. (1998b). "An overview of the simultaneous perturbation method for efficient optimization." *Johns Hopkins APL Tech.Dig.,* 19(4), 482.

Su, H., Rakheja, S., and Sankar, T. S. (1991). "Stochastic analysis of nonlinear vehicle systems using a generalized harmonic linearization technique." *Prob.Eng.Mech.,* 6(3–4), 175-183.

Svanberg, K. (1987). "The method of moving asymptotes - a new method for structural optimization." *Int J Numer Methods Eng,* 24(2), 359-373.

Szaszi, I., Gaspar, P., and Bokor, J. (2002). "Nonlinear active suspension modeling using linear parameter varying approach." *10th IEEE Mediterranean Conference on Control and Automation, Lisbon, Portugal,* IEEE, .

Taflanidis, A. A. (2012). "Stochastic subset optimization incorporating moving least squares response surface methodologies for stochastic sampling." *Adv.Eng.Software,* 44(1), 3-14.

Taflanidis, A. A., and Beck, J. L. (2008). "An efficient framework for optimal robust stochastic system design using stochastic simulation." *Comput.Methods Appl.Mech.Eng.,* 198(1), 88.

Taflanidis, A. A., and Beck, J. L. (2010). "Reliability-based design using two-stage stochastic optimization with a treatment of model prediction errors." *J.Eng.Mech.,* 136(12), 1460-1473.

Taflanidis, A. A., and Beck, J. L. (2009). "Life-cycle cost optimal design of passive dissipative devices." *Struct.Saf.,* 31(6), 508-522.

Taflanidis, A. A., Scruggs, J. T., and Beck, J. L. (2010). "Robust stochastic design of linear controlled systems for performance optimization." *Journal of Dynamic Systems, Measurement, and Control,* 132(5), 051008.

Taflanidis, A. A., Jia, G., Kennedy, A. B., and Smith, J. M. (2013). "Implementation/optimization of moving least squares response surfaces for approximation of hurricane/storm surge and wave responses." *Nat.Hazards,* 66(2), 955-983.

Taflanidis, A. A. (2009). "Stochastic subset optimization for reliability optimization and sensitivity analysis in system design." *Computers Structures,* 87(5-6), 318.

Taflanidis, A. A., and Jia, G. (2011). "A simulation-based framework for risk assessment and probabilistic sensitivity analysis of base-isolated structures." *Earthquake Engineering Structural Dynamics,* 40(14), 1629.

Taflanidis, A. A., Vetter, C., and Loukogeorgaki, E. (2013). "Impact of modeling and excitation uncertainties on operational and structural reliability of tension leg platforms." *Appl.Ocean Res.,* 43(0), 131-147.

Tamboli, J. A. (1999). "Optimum design of a passive suspension system of a vehicle subjected to actual random road excitations." *J.Sound Vibrat.,* 219(2), 193.

Tang, Y., and Chen, J. (2009). "Robust design of sheet metal forming process based on adaptive importance sampling." *Structural and Multidisciplinary Optimization,* 39(5), 531-544.

Thomas, H., Vanderplaats, G., and Shyy, Y. (1992). "A study of move limit adjustment strategies in the approximation concepts approach to structural synthesis." *Proc. 4th AIAA/USAF/NASA/OAI Symposium on Multidisciplinary Analysis and Optimization, Cleveland, OH, AIAA-92-4839,* .

Tootkaboni, M., Asadpoure, A., and Guest, J. K. (2012). "Topology optimization of continuum structures under uncertainty-A Polynomial Chaos approach." *Comput.Methods Appl.Mech.Eng.,* 201-204 263.

Tsompanakis, Y., Lagaros, N. D., and Papadrakakis, M. (2008). *Structural Design Optimization Considering Uncertainties: Structures & Infrastructures Book, Vol. 1, Series, Series Editor: Dan M. Frangopol.* Taylor & Francis, .

Tu, J., Choi, K. K., and Park, Y. H. (1999). "A new study on reliability-based design optimization." *Journal of Mechanical Design,* 121(4), 557-564.

Valdebenito, M. A., and Schuëller, G. I. (2010). "A survey on approaches for reliability-based optimization." *Structural and Multidisciplinary Optimization,* 42(5), 645-663.

Van Beers, W. C., and Kleijnen, J. P. (2004). "Kriging interpolation in simulation: a survey." *Simulation Conference, 2004. Proceedings of the 2004 Winter,* IEEE, .

Verros, G., Natsiavas, S., and Papadimitriou, C. (2005). "Design optimization of quarter-car models with passive and semi-active suspensions under random road excitation." *J.Vibrat.Control,* 11(5), 581-606.

Vetter, C. (2012). "Global sensitivity analysis for stochastic ground motion modeling in seismic-risk assessment." *Soil Dyn.Earthquake Eng.,* 38 128.

Wang, G. G., and Shan, S. (2007). "Review of metamodeling techniques in support of engineering design optimization." *Journal of Mechanical Design,* 129 370.

Yao, W., Chen, X., Luo, W., van Tooren, M., and Guo, J. (2011). "Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles." *Prog.Aerospace Sci.,* 47(6), 450-479.

Youn, B. D., and Choi, K. K. (2004). "A new response surface methodology for reliability-based design optimization." *Comput.Struct.,* 82(2), 241-256.

Zang, C., Friswell, M. I., and Mottershead, J. E. (2005). "A review of robust optimal design and its application in dynamics." *Comput.Struct.,* 83(4–5), 315-326.

Zhao, L., Choi, K., and Lee, I. (2011). "Metamodeling method using dynamic Kriging for design optimization." *AIAA J.,* 49(9), 2034-2046.