Summer 2014

Modified Bayesian Kriging for noisy response problems and Bayesian confidence-based reliability-based design optimization

Nicholas John Gaul
University of Iowa

MODIFIED BAYESIAN KRIGING FOR NOISY RESPONSE PROBLEMS AND BAYESIAN CONFIDENCE-BASED RELIABILITY-BASED DESIGN OPTIMIZATION

by

Nicholas John Gaul

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Mechanical Engineering
in the Graduate College of
The University of Iowa

August 2014

Thesis Supervisors: Professor Kyung K. Choi
Professor Mary Kathryn Cowles

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Nicholas John Gaul

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Mechanical Engineering at the August 2014 graduation.

Thesis Committee: _____
Kyung K. Choi, Thesis Supervisor

_____
Mary Kathryn Cowles, Thesis Supervisor

_____
David Lamb

_____
Jia Lu

_____
Sharif Rahman

To my family and friends

If you can’t explain it simply, you don’t understand it well enough.
Albert Einstein

ABSTRACT

The objective of this study is to develop a new modified Bayesian Kriging (MBKG) surrogate modeling method that can be used to carry out confidence-based reliability-based design optimization (RBDO) for problems in which simulation analyses are inherently noisy and standard Kriging approaches fail. The formulation of the MBKG surrogate modeling method is presented, and the full conditional distributions of the unknown MBKG parameters are derived and coded into a Gibbs sampling algorithm. Using the coded Gibbs sampling algorithm, Markov chain Monte Carlo is used to fit the MBKG surrogate model.

A sequential sampling method that uses the posterior credible sets for inserting new design of experiment (DoE) sample points is proposed. The sequential sampling method is developed in such a way that the new DoE sample points added will provide the maximum amount of information possible to the MBKG surrogate model, making it an efficient and effective way to reduce the number of DoE sample points needed. Therefore, it improves the posterior distribution of the probability of failure efficiently.

Finally, a confidence-based RBDO method using the posterior distribution of the probability of failure is developed. The confidence-based RBDO method is developed so that the uncertainty of the MBKG surrogate model is included in the optimization process.

A 2-D mathematical example was used to demonstrate fitting the MBKG surrogate model and the developed sequential sampling method that uses the posterior credible sets for inserting new DoE. A detailed study on how the posterior distribution of the probability of failure changes as new DoE are added using the developed sequential sampling method is presented. Confidence-based RBDO is carried out using the same 2-D mathematical example. Three different noise levels are used for the example to compare how the MBKG surrogate modeling method, the sequential sampling method,

and the confidence-based RBDO method behave for different amounts of noise in the response. A comparison of the optimization results for the three different noise levels for the same 2-D mathematical example is presented.

A 3-D multibody dynamics (MBD) engineering block-car example is presented. The example is used to demonstrate using the developed methods to carry out confidence-based RBDO for an engineering problem that contains noise in the response. The MBD simulations for this example were done using the commercially available MBD software package RecurDyn. Deterministic design optimization (DDO) was first done using the MBKG surrogate model to obtain the mean response values, which then were used with standard Kriging methods to obtain the sensitivity of the responses. Confidence-based RBDO was then carried out using the DDO solution as the initial design point.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AMV	Advanced Mean Value Method
BGR	Brooks, Gelman, and Rubin
CAE	Computer Aided Engineering
CDF	Cumulative Distribution Function
CFD	Computational Fluid Dynamics
CMV	Conjugate Mean Value Method
DDO	Deterministic Design Optimization
DKG	Dynamic Kriging
DoE	Design of Experiment
DRM	Dimension Reduction Method
FE	Finite Element
FORM	First-Order Reliability Method
HMV	Hybrid Mean Value Method
HMV+	Enhanced Hybrid Mean Value Method
LCVT	Latinized Centroidal Voronoi Tessellation
MARS	Multivariate Adaptive Regression Splines
MBKG	Modified Bayesian Kriging
MCMC	Markov Chain Monte Carlo
MCS	Monte Carlo Simulation
MBD	Multibody Dynamics
MLS	Moving Least Squares
MPP	Most Probable Point
MV	Mean Value Method
NN	Neural Networks
OKG	Ordinary Kriging

PDF	Probability Density Function
PMA	Performance Measure Approach
PRS	Polynomial Response Surface
RBF	Radial Basis Functions
RBDO	Reliability-Based Design Optimization
RIA	Reliability Index Approach
SORM	Second-Order Reliability Method
UKG	Universal Kriging
VSVM	Virtual Support Vector Machine

CHAPTER 1

INTRODUCTION

This study presents a new surrogate modeling method for reliability-based design optimization (RBDO) of problems whose simulation analyses are inherently noisy. A new modified Bayesian Kriging (MBKG) surrogate modeling method is proposed for handling performance measures of noisy simulations, which will accurately represent the true underlying unknown performance function without noise. Using Bayesian methods allows for a way to naturally characterize the uncertainty of the predicted values by providing the distribution of the predicted values. Credible sets produced by the surrogate model are used to develop a sequential sampling method and also to create a conservative RBDO design. The likelihood for the MBKG is proposed, and the prior distributions to be used with the likelihood to fit the Bayesian model are presented. A Markov chain Monte Carlo algorithm for fitting the MBKG model is developed as follows. Using the prior distributions, the full conditional distributions of the MBKG parameters are derived; where possible, conjugate prior distributions are used to help simplify the full conditionals and to ease the computational burden of fitting the surrogate model. Computer simulations of engineering models are computationally expensive; therefore, it is necessary to reduce the number of design of experiment (DoE) samples needed. An efficient DoE sampling method using the credible sets of the MBKG model will be developed to systematically reduce the uncertainty in the surrogate model. Finally, a confidence-based RBDO method using the posterior distribution of the probability of failure will be developed.

Section 1.1 presents the background and motivation of the proposed research; Section 1.2 gives a brief introduction to Bayesian statistics; Section 1.3 discusses the objectives of the proposed research; and Section 1.4 presents the thesis organization.

1.1 Background and Motivation

1.1.1 Reliability-Based Design Optimization

Customers want products that are low cost and also reliable, whether they are purchasing a small kitchen appliance, smartphone, tablet, computer, vehicle, or heavy machinery. At the same time, manufacturers want to make products that cost less to make and are reliable in order to reduce both warranty and production costs. The objective to reduce cost naturally led to the development of optimization methods. Just carrying out optimization without consideration of uncertainty is referred to as deterministic design optimization (DDO). Deterministic design optimization solutions, i.e., designs of products, are only approximately 50% reliable. Safety factors are often used when designing products to try to ensure that they will last and be reliable; however, estimation of the safety factor could be heuristic, which in turn increases manufacturing cost or reduces reliability. Both of these realities have brought about the need for RBDO.

Most reliability analysis methods can be classified into groups, the first being sensitivity-based methods and the second being sampling-based methods. The literature is rich with numerous sensitivity-based methods that have been developed using the most probable point (MPP) [Lee et al. 2010; Lee et al. 2008; Rahman and Wei 2006; Youn and Choi 2003; Haldar and Mahadevan 2000; Tu et al. 1999]. Some common sensitivity-based methods are the first-order reliability method (FORM) [Haldar and Mahadevan 2000; Hohenbichler et al. 1987; Madsen et al. 1985], the second-order reliability method (SORM) [Haldar and Mahadevan 2000; Hohenbichler and Rackwitz 1988; Hohenbichler et al. 1987; Madsen et al. 1985], and the dimension reduction method (DRM) [Lee et al. 2010; Lee et al. 2008; Rahman and Wei 2008; Rahman and Wei 2006]. While these methods can be computationally cheaper than sampling-based methods, one pitfall is that they may not be as accurate as sampling-based methods for highly nonlinear problems [Lee et al. 2011]. Another shortcoming of these methods is that they require the

sensitivity of the performance measures to be available. Obtaining the sensitivity, while possible for some problems, can be a daunting if not impossible task for some problems that are highly nonlinear and/or coupled with fluid structure interaction, e.g., crash and blast problems.

To overcome the shortfalls of the sensitivity-based methods, sampling-based methods have been developed. A brute force approach would be to do direct Monte Carlo simulation (MCS) using the computer-aided engineering (CAE) simulation models, e.g., finite element (FE) models and computational fluid dynamic (CFD) models, to calculate the reliability. While this method can be highly accurate using a large number of MCS points [Haldar and Mahadevan 2000], its limitation is the large number of MCS points required due to the computational cost of the CAE model simulations, thus rendering it impractical. In order to overcome the impractical computational cost of direct MCS using CAE simulations and to overcome the pitfalls of the sensitivity-based methods, the use of surrogate models is becoming a more common practice [Shi et al. 2012; Song et al. 2011; Zhao 2011; Zhao et al. 2011; An and Choi 2012; Rajashekhar and Ellingwood 1993]. There are three advantages of using surrogate models. The first is that the sensitivity of the performance measure is not needed to construct the surrogate. The second is that surrogate models are computationally inexpensive to use for evaluating large numbers of MCS points compared to the computational cost of the CAE models. The third is that surrogate models can be built using a limited number of CAE model simulations, therefore reducing the overall computational cost of performing a reliability analysis. The literature is full of numerous surrogate modeling methods that have been developed; these will be discussed in the next section.

1.1.2 Surrogate Modeling Methods

Surrogate modeling methods have been under development for decades and are still being actively developed. The main reason for all the development is that there is not one surrogate modeling method that works for every problem. However, the use of surrogate models for product design is becoming a more common practice during the design stage to help engineers gain a quick understanding of a problem due to their relative ease of use, inexpensive computational time, and ready availability in a number of commercial software programs, e.g., HyperStudy, Isight, Matlab, and Mathematica.

As previously mentioned, there are numerous surrogate modeling methods in the literature, such as polynomial response surface (PRS) [Forrester and Keane 2009; Forrester et al. 2008; Fang et al. 2005; Jin et al. 2001; Box and Draper 1987; Rajashekhar and Ellingwood 1993; Mullur and Messac 2006; Simpson et al. 2001; Wang and Shan 2007], polynomial chaos [Wiener 1938; Hu and Youn 2011; Isukapalli 1999; Kewlani and Iagnemma 2008; Wei et al. 2008], moving least squares (MLS) [Forrester and Keane 2009; Breitkopf et al. 2005; Lancaster and Salkauskas 1981; Levin 1998], multivariate adaptive regression splines (MARS) [Jin et al. 2001; Friedman 1991; Friedman and Roosen 1995; Lewis and Stevens 1991; Simpson et al. 2001; Wang and Shan 2007], support vector machine and support vector regression [Forrester and Keane 2009; Forrester et al. 2008; Burges 1998; Hearst et al. 1998; Wang and Shan 2007], virtual support vector machine (VSVM) [Song et al. 2011], radial basis functions (RBF) [Forrester and Keane 2009; Buhmann 2003; Forrester et al. 2008; Fang et al. 2005; Jin et al. 2001; Park and Sandberg 1991; Dyn et al. 1986; Mullur and Messac 2006; Wang and Shan 2007], neural networks (NN) [Agatonovic-Kustrin and Beresford 2000; Almeida 2002; Fonseca et al. 2003; Liu and Fang 2009; Sakata et al. 2010; Ukrainec et al. 1989; van der Merwe et al. 2007; Zobel et al. 2008; Galushkin 2007; Gallant 1993; Forrester et al. 2008; Simpson et al. 2001; Wang and Shan 2007], ordinary Kriging (OKG) and universal Kriging (UKG) [Krige 1951; Cressie 1991; Sacks et al. 1989; Beers and

Kleijnen 2003; Forrester and Keane 2009; Forrester et al. 2008; Mullur and Messac 2006; Jin et al. 2001; Simpson et al. 2001; Wang and Shan 2007], and dynamic Kriging (DKG) [Zhao 2011; Zhao et al. 2011]. There have been surrogate modeling methods developed that use Bayesian methods [Shi et al. 2012; An and Choi 2012; Romero et al. 2012; Romero et al. 2003; Romero 2008; Omre and Halvorsen 1989; Currin et al. 1991]. However, none of the methods studied claim to do a full Bayesian analysis when creating the surrogate model; they are only borrowing concepts from the Bayesian methods.

As previously described, a surrogate model is used to do the MCS prediction for the reliability analysis for the sampling-based RBDO method to ease the computational burden. Thus, the accuracy of the reliability analysis depends on the accuracy of the surrogate model. Kriging has become a popular surrogate modeling method because it offers flexibility in the choice of both the mean structure and the correlation function used [Forrester and Keane 2009; Forrester et al. 2008]. Dynamic Kriging was developed because it was found to be more accurate than OKG or UKG, giving better reliability analysis results [Zhao 2011; Zhao et al. 2011]. However, OKG, UKG, and DKG are typically formulated and used as interpolation methods and therefore break down when the response data contains noise [Forrester and Keane 2009; Forrester et al. 2008; Sakata et al. 2007; Sakata et al. 2008]. There are approximation and regression methods, e.g. MLS and MARS, which can be used when the response data contains noise. The formulation of Kriging can even be modified to change it to a regression method [Forrester and Keane 2009; Forrester et al. 2008; Sakata et al. 2007; Sakata et al. 2008].

The disadvantage of these methods is that they do not allow for a direct way to separate the noise from the data. Thus, when using them to predict response values, it is not known how much noise there may be in the predicted response value. If these regression surrogate models are used for MCS prediction for reliability analysis, this will affect the amount of uncertainty and variability that there is in the reliability analysis. It is not clear or easy to distinguish if the variability in the reliability analysis is due to the

noise in the predicted response or due to the uncertainty in the surrogate model itself. Another disadvantage of these methods is that they do not have a systematic way of characterizing the uncertainty of the surrogate model, the noise in the response value, or the uncertainty in the predicted response values using the surrogate model. Therefore, a surrogate modeling method that can systematically characterize all of these uncertainties as well as accurately predict the true underlying response value without noise needs to be investigated. Bayesian statistical methods allow for a natural and systematic way of characterizing uncertainty when predicting unknown parameters of distributions as well as predicting parameters that depend on these unknown distribution parameters. A brief introduction of Bayesian statistical methods is given in the next section.

1.2 Bayesian Statistics

1.2.1 Likelihood and Prior Distributions

In real-world problems there is often data available or obtainable that is known or assumed to come from a given distribution type, e.g., the data is known to follow a normal distribution. The distribution from which the data comes is referred to as the sampling distribution or data distribution [Gelman et al. 2004; Hamada et al. 2008]. However, even when the distribution type of the data is known, the parameter values of the distribution are often unknown. When the data distribution is considered as a function of the unknown distribution parameters for a given data set, it is referred to as the likelihood function [Gelman et al. 2004; Cowles 2013; Hamada et al. 2008; Bayes and Price 1763].

The goal in Bayesian statistics is to come up with an estimate of the unknown parameter values using both the available data and prior knowledge about the unknown parameters. In Bayesian statistics the unknown parameters are treated as if they are random variables. Any prior knowledge or belief about the unknown parameter is

expressed using a probability distribution. The definition of subjective probability of an event given by Cowles is: “A probability of an event or of the truth of a statement is a number between 0 and 1 that quantifies a particular person’s subjective opinion as to how likely that event is to occur (or to have already occurred) or how likely the statement is to be true” [2013]. Similarly, a subjective probability distribution quantifies one’s knowledge about an unknown parameter that may take on any value in a continuum. These probability distributions that express one’s knowledge about the unknown parameters are referred to as the prior distributions or simply as the priors [Cowles 2013; Gelman et al. 2004; Hamada et al. 2008; Bayes and Price 1763]. A Bayesian analysis is carried out to update one’s subjective probability distribution of the unknown parameters by combining prior information with the new information contained in the data. The next section will describe how Bayes’ rule is used with the likelihood and prior distributions to calculate the posterior distribution.

1.2.2 Bayes’ Rule and Posterior Distributions

Bayes’ rule provides a way to mathematically and systematically update one’s subjective probability distribution on model parameters. All knowledge available before the current data is observed is encapsulated in the prior. The updating occurs by incorporating the new data contained in the likelihood. Bayes’ rule states that the posterior probability distribution of model parameters given the data is proportional to the product of the likelihood and prior probability distribution [Bayes and Price 1763; Cowles 2013; Gelman et al. 2004; Hamada et al. 2008]. Mathematically this can be written as shown in Eq. (1.1).

$$f\left(Parameter \mid Data\right) \propto f\left(Parameter\right) \times f\left(Data \mid Parameter\right) \tag{1.1}$$

The final probability distribution given on the left side of Eq. (1.1) is called the posterior distribution, often referred to as the posterior. The posterior distribution expresses the current state of knowledge about model parameters. The posterior distribution can be used to obtain desired values about the event, e.g., the mean or median of the posterior could be used as a point estimate for the unknown parameter value. The posterior variance reveals the amount of uncertainty that remains about the parameter value—the larger the variance, the larger the uncertainty about what the true parameter value is. The posterior can also be used to give probability intervals, called credible sets, which are believed to contain the true parameter value with the specified probability, e.g., the 95% credible set for the parameter can easily be obtained from the posterior. Credible sets have a slightly different meaning than confidence intervals. For the 95% credible set, the probability that the true value of the parameter is in that interval is 95%. A 95% confidence interval means that if the same experiment or test is repeated many times to generate many different data sets, and each data set is used to generate a 95% confidence interval, only 95% of the confidence intervals generated would capture the true parameter value, while 5% of the confidence intervals generated would not capture the true parameter value [Cowles 2013; Gelman 2004]. The next section will discuss the use of conjugate priors and Markov chain Monte Carlo for updating the posterior distribution.

1.2.3 Conjugate Priors and Markov Chain Monte Carlo

As described in the previous section, the product of the prior distribution and likelihood is used in Bayes' rule to construct the posterior distribution of the unknown parameter(s). When possible, it is desirable to choose the prior distribution from a parametric family that takes on the same functional form as the likelihood function for the unknown parameter(s); priors of such a form are called conjugate priors. The parameters of the prior distribution are then chosen such that the prior reflects the known

information and beliefs about the unknown parameter(s). When applying Bayes' rule to the likelihood with a conjugate prior, the posterior distribution will belong to the same parametric family as the prior, i.e., the posterior distribution type will be the same distribution type as the prior. The parameter values of the posterior distribution will be a combination of the prior parameter values and the data used for the Bayesian analysis [Cowles 2013; Gelman 2004; Hamada et al. 2008].

There are scenarios in which a conjugate prior does not exist for a given problem, e.g., a conjugate prior does not adequately reflect the prior knowledge and belief about the unknown parameter(s), or there are multiple unknown parameters for which a conjugate joint distribution does not exist. For such scenarios, any distribution type that reflects the prior knowledge and belief about the unknown parameter(s) can be used as the prior. The use of such priors, however, leads to posterior distributions that most likely are not from a known distribution family type. Bayes' rule can still be applied in such cases but has to be done using a numerical method.

Markov chain Monte Carlo (MCMC) is a numerical method that can be used to draw samples from high-dimensional and nonstandard distribution types. One disadvantage of MCMC is that the samples drawn from the distribution are not independent, and this needs to be taken into consideration when using the samples for inference [Cowles 2013; Gelman 2004; Gilks et al. 1998; Hamada et al. 2008; Tierney 1998; Feller 1968]. The Markov property states that a sample drawn at a given time point conditional on the sample drawn at the time point immediately before it is independent of all the earlier samples drawn. Under certain regularity conditions, it can be shown that a Markov chain will converge in distribution to samples drawn from the target, i.e., posterior distribution [Cowles 2013; Gelman 2004; Gilks et al. 1998; Tierney 1998; Feller 1968]. There is a debate over whether it is better to run one long Markov chain or to run multiple shorter parallel Markov chains starting at different initial values to attempt to determine if the Markov chain has converged to the target distribution as well

as how to best assess whether convergence has occurred [Kass et al. 1998; Cowles and Carlin 1996]. One common and well-accepted way of diagnosing convergence in distribution when using MCMC is to use the Brooks, Gelman, and Rubin (BGR) diagnostic [Gelman and Rubin 1992; Brooks and Gelman 1998]. The BGR diagnostic requires running at least two parallel Markov chains starting at over-dispersed initial values. The BGR diagnostic uses the samples drawn from the parallel Markov chains to calculate credible sets of the individual chains using an increasing number of samples from the chain. The parallel chains are also pooled together to form one sample set that is used to calculate credible sets using an increasing number of samples. If widths of the credible sets calculated using the two different methods stabilize and become approximately equal, the MCMC chains are likely to have converged in distribution [Gelman and Rubin 1992; Brooks and Gelman 1998; Lunn et al. 2000; Lunn et al. 2009; Gelman 2013; Cowles 2013]. It is possible for the BGR diagnostic to misdiagnose convergence, i.e., the BGR diagnostic shows that convergence has been achieved when in actuality convergence has not yet been reached. The next section will discuss the objectives of the proposed study and how Bayesian statistical methods will be used in creating and fitting a surrogate model for problems with noisy simulation analyses, as well as capturing the uncertainty in the surrogate model and predicted values using credible sets.

1.3 Objectives of the Proposed Study

The first objective of this study is to develop a modified Bayesian Kriging (MBKG) surrogate modeling method that can accurately model the true underlying response value for noisy simulations. The posterior credible sets of the MBKG surrogate model will capture and show the uncertainty in the predicted values. The posterior credible sets will be used for carrying out confidence-based RBDO. The different Kriging

methods, OKG, UKG, and DKG, have been shown to be more accurate while using less data than other existing surrogate modeling methods [Zhao 2011; Zhao et al. 2011; Forrester and Keane 2009; Forrester et al. 2008]. This is why this study proposes using a modified Bayesian Kriging method.

The second objective is to develop an efficient method for selecting new DoE samples using the credible sets. This will allow for a systematic and mathematical way to continuously improve the MBKG surrogate model and measure the amount of improvement during sequential DoE sampling. Using the credible sets, new DoE samples will be added in areas where the uncertainty of the surrogate model is large, thereby decreasing the uncertainty of the surrogate model.

The third and final objective is to develop a confidence-based RBDO method using the posterior distribution of the probability of failure. The confidence-based RBDO method will be used to generate conservative reliable optimal designs. In order to perform optimization, the sensitivity of the probability of failure is needed and thus will be derived. The next section describes the organization of the thesis.

1.4 Organization of Thesis

Chapter 2 presents fundamental concepts of reliability analysis and reliability-based design optimization for both sensitivity-based and sampling-based methods.

Chapter 3 presents the existing conventional Kriging methods and the proposed modified Bayesian Kriging method. The prior distributions and corresponding full conditional distributions for the modified Bayesian Kriging method are also presented.

Chapter 4 presents three different examples demonstrating the use of the modified Bayesian Kriging method for fitting problems with noisy responses.

Chapter 5 presents the proposed sequential sampling method that uses the posterior credible sets for inserting new design of experiment sample points for updating the modified Bayesian Kriging surrogate model.

Chapter 6 presents the confidence-based reliability-based design optimization method that uses the posterior distribution of the probability of failure. A mathematical example using different amounts of noise are used to demonstrate the method. A 3-D multibody dynamics engineering example is used to demonstrate the method.

Chapter 7 presents the conclusions of the study and the future research to be carried out to enhance the modified Bayesian Kriging method to make carrying out confidence-based RBDO more efficient.

CHAPTER 2

DESIGN UNDER UNCERTAINTY

2.1 Introduction

This chapter gives a summary of the fundamental concepts in design under uncertainty, including sensitivity-based and sampling-based RBDO methods. Sections 2.2 and 2.3 present the basic ideas of reliability analysis and inverse reliability analysis methods. Section 2.4 discusses the sensitivity-based method using both the FORM and DRM methods. Section 2.5 introduces the sampling-based RBDO that is used for problems when the sensitivity cannot be calculated. As discussed in the previous chapter, the sampling-based method uses MCS for calculating the probability of failure and also the probabilistic sensitivity for the performance measures that are used for optimization.

2.2 Reliability Analysis

In order to perform a reliability analysis, the calculation of the probability of failure is required. The probability of failure, denoted by P_F, is calculated using the multi-dimensional integral [Madsen et al. 1986]

$$P_F \equiv P[G(\mathbf{X}) > 0] = \int_{G(\mathbf{X})>0} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{2.1}$$

where $\mathbf{X}=\{X_1, X_2, \cdots, X_{nr}\}^{\mathrm{T}}$ is an nr dimensional random vector, nr is the number of random variables, $G(\mathbf{X})$ is the performance measure function that is defined such that $G(\mathbf{X}) > 0$ is failure, and $f_{\mathbf{X}}(\mathbf{x})$ is the joint probability density function (PDF) of the random input variables X. For most real-world engineering problems, the exact evaluation of Eq. (2.1) is very difficult if not impossible to carry out since $f_{\mathbf{X}}(\mathbf{x})$ is usually non-Gaussian due to correlation in the random variables and $G(\mathbf{X})$ can be highly

nonlinear. The integration domain of Eq. (2.1) generally cannot be expressed analytically due to the nonlinearity of $G(\mathbf{X})$. To handle the non-Gaussian distribution and highly nonlinear $G(\mathbf{X})$, a transformation of the random variables, $\mathbf{X}$, from the X-space to the independent standard normal U-space is carried out [Rosenblatt 1952; Hogg et al. 2005]. To deal with highly nonlinear performance measures, $G(\mathbf{X})$ is approximated using first-order Taylor series expansion in the FORM and SORM methods for the sensitivity-based RBDO methods, and probabilistic MCS methods are used for sampling-based RBDO methods. The next section will introduce the transformation of variables.

2.2.1 Random Variable Transformation

For an nr dimensional random vector $\mathbf{X}$ that has a joint cumulative distribution function (CDF) $F_{\mathbf{X}}(\mathbf{x})$, let $T:\mathbf{X}\to\mathbf{U}$ denote the transformation from X-space to U-space that is defined by the Rosenblatt transformation [Rosenblatt 1952] as

$$T:\begin{cases} u_1=\Phi^{-1}\left[F_{X_1}\left(x_1\right)\right] \\ u_2=\Phi^{-1}\left[F_{X_2}\left(x_2\left|x_1\right.\right)\right] \\ \quad\vdots \\ u_{nr}=\Phi^{-1}\left[F_{X_{nr}}\left(x_{nr}\left|x_1,x_2,\cdots,x_{nr-1}\right.\right)\right] \end{cases} \tag{2.2}$$

where $F_{X_i}\left(x_i\left|x_1,x_2,\cdots,x_{i-1}\right.\right)$ is the conditional CDF given by

$$F_{X_i}\left(x_i\left|x_1,x_2,\cdots,x_{i-1}\right.\right)=\frac{\int_{-\infty}^{x_i} f_{X_1X_2\cdots X_i}(x_1,x_2,\cdots,x_{i-1},\xi)d\xi}{f_{X_1X_2\cdots X_{i-1}}(x_1,x_2,\cdots,x_{i-1})} \tag{2.3}$$

where ξ is the random variable being integrated over and $\Phi(\bullet)$ is the standard normal CDF given by

$$\Phi(u) = \int_{-\infty}^{u} \phi(\xi) d\xi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \exp\left(-\frac{1}{2}\xi^2\right) d\xi \tag{2.4}$$

where ϕ is the standard normal PDF.

The inverse transformation of Eq. (2.2) is expressed as

$$T^{-1}: \begin{cases} x_1 = F_{X_1}^{-1}\left[\Phi(u_1)\right] \\ x_2 = F_{X_2}^{-1}\left[\Phi(u_2 | x_1)\right] \\ \vdots \\ x_{nr} = F_{X_{nr}}^{-1}\left[\Phi(u_{nr} | x_1, x_2, \cdots, x_{nr-1})\right] \end{cases} \tag{2.5}$$

If the random variables of the $\mathbf{X}$ vector are independent, then the joint PDF is given as the product of the marginal PDF's $f_{X_i}(x_i)$ as

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1) \times f_{X_2}(x_2) \times \cdots \times f_{X_{nr}}(x_{nr}) \tag{2.6}$$

In this case the Rosenblatt transformation and the inverse transformation simplify to

$$u_i = \Phi^{-1}\left[F_{X_i}(x_i)\right] \text{ and } x_i = F_{X_i}^{-1}\left[\Phi(u_i)\right] \tag{2.7}$$

where $F_{X_i}(x_i)$ are the marginal CDFs. Table 2.1 shows five representative distributions and their corresponding transformations, assuming the random variables are independent.

Table 2.1 Probability Distribution and Its Transformation between X- and U-space

	Parameters	PDF	Transformation
Normal	$\mu = \text{mean}$ $\sigma = \text{standard deviation}$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-0.5[\frac{x-\mu}{\sigma}]^2}$	$X = \mu + \sigma U$
Log-normal	$\bar{\sigma}^2 = \ln[1+(\frac{\sigma}{\mu})^2]$, $\bar{\mu} = \ln(\mu) - 0.5\bar{\sigma}^2$	$f(x) = \frac{1}{\sqrt{2\pi}x\bar{\sigma}} e^{-0.5[\frac{\ln x - \bar{\mu}}{\bar{\sigma}}]^2}$	$X = \exp(\bar{\mu} + \bar{\sigma}U)$
Weibull	$\mu = \nu\Gamma(1+\frac{1}{k})$, $\sigma^2 = \nu^2[\Gamma(1+\frac{2}{k}) - \Gamma^2(1+\frac{1}{k})]$	$f(x) = \frac{k}{\nu}(\frac{x}{\nu})^{k-1} e^{-(\frac{x}{\nu})^k}$	$X = \nu[-\ln(\Phi(-U))]^{\frac{1}{k}}$
Gumbel	$\mu = \nu + \frac{0.577}{\alpha}, \sigma = \frac{\pi}{\sqrt{6}\alpha}$	$f(x) = \alpha e^{-\alpha(x-\nu) - e^{-\alpha(x-\nu)}}$	$X = \nu - \frac{1}{\alpha}\ln[-\ln(\Phi(U))]$
Uniform	$\mu = \frac{a+b}{2}$, $\sigma = \frac{b-a}{\sqrt{12}}$	$f(x) = \frac{1}{b-a}$, $a \le x \le b$	$X = a + (b-a)\Phi(U)$

2.2.2 First-Order Reliability Method (FORM) and Second-Order Reliability Method (SORM)

To calculate the probability of failure of the performance measure function $G(\mathbf{X})$ using FORM and SORM, it is first necessary to find the MPP, which is defined as the point $\mathbf{u}^*$ on the limit state function, $g(\mathbf{u}) = 0$, closest to the origin in the standard normal U-space as shown in Figure 2.1. Using the Rosenblatt transformation, the performance measure function in the U-space is defined as $g(\mathbf{u}) \equiv G(\mathbf{x}(\mathbf{u})) = G(\mathbf{x})$. Thus, the MPP can be found by solving the following optimization problem:

$$\begin{aligned} &\text{minimize} && \|\mathbf{u}\| \\ &\text{subject to} && g(\mathbf{u}) = 0 \end{aligned} \tag{2.8}$$

The distance from the MPP to the origin is commonly called the Hasofer-Lind reliability index [Hasofer and Lind 1974] and is denoted by β_{HL}. Using this reliability

index, FORM can approximate the probability of failure using a linear approximation of the performance function given as

$$P_F^{\text{FORM}} \cong \Phi(-\beta_{\text{HL}}) \tag{2.9}$$

A quadratic approximation of the performance measure function in the U-space and the rotational transformation from the standard normal U-space to the rotated standard normal V-space is used in SORM to calculate the probability of failure [Breitung 1984; Hohenbichler and Rackwitz 1988; Rahman and Wei 2006].

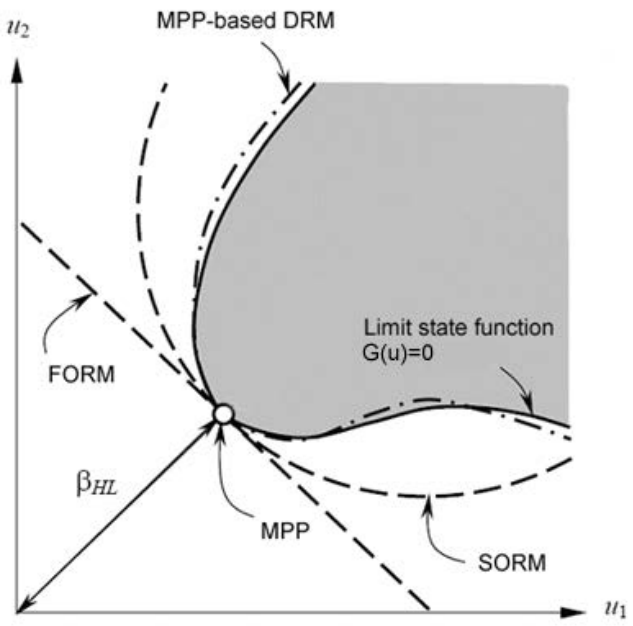

Figure 2.1 MPP and Reliability Index β_{HL} in the U-Space [Source: Wei 2006]

2.3 Inverse Reliability Analysis

The reliability analysis using FORM and SORM presented in the previous section is called the reliability index approach (RIA) [Tu et al. 1999] since it finds the reliability index β_{HL} using Eq. (2.8). The RIA method has the advantage that the probability of failure of the performance measure function can be calculated at the given design. However, the inverse reliability analysis in the performance measure approach (PMA) [Tu et al. 1999; Tu et al. 2001; Choi et al. 2001; Youn et al. 2003] is known to be numerically more efficient and stable than RIA. The probability of failure is not directly calculated in PMA. Instead, PMA determines if a given design satisfies the probabilistic constraint for a given target probability of failure P_F^{Tar}. The optimization problem for PMA is defined as

$$\begin{aligned} &\text{maximize} \quad g(\mathbf{u}) \\ &\text{subject to} \quad \|\mathbf{u}\| = \beta_t \end{aligned} \tag{2.10}$$

where β_t is the target reliability index. This is called inverse reliability analysis because Eq. (2.10) is the inverse problem of Eq. (2.8). The probabilistic constraint is satisfied for the given target reliability index β_t when the performance measure function value at the MPP, $g(\mathbf{u}^*)$, is less than zero, i.e., $G(\mathbf{X}) < 0$ is defined as safe.

There are several methods that have been developed to find the MPP using inverse reliability analysis with the given target reliability index β_t; the different methods are the mean value (MV) method, the advanced mean value (AMV) method [Wu et al. 1990; Wu 1994], the hybrid mean value (HMV) method [Youn et al 2003], and the enhanced hybrid mean value (HMV+) method [Youn et al. 2005].

The MV method creates a linear approximation of the performance measure function using the function value and gradient information at the mean value in the standard normal U-space. Thus, the MV method is a crude method for finding the MPP

of the inverse reliability analysis. Since no additional function evaluations and sensitivity information is needed, the MV method can be a good approximation to determine which constraints are active.

The AMV method uses the MPP obtained from the MV method for the first iteration. AMV then uses the gradient information at the MPP provided by the MV method to find the next candidate MPP; the iteration continues until the approximate MPP converges to the correct MPP. For convex performance measure functions, the AMV method is known to be an efficient method.

To overcome the issues AMV has with concave functions, the HMV method was developed. The HMV method uses the conjugate mean value (CMV) method for concave performance measure functions [Youn et al. 2003], and AMV is still used for convex performance measure functions. The HMV+ method uses interpolation between the two previous MPP candidate points for concave performance measure functions instead of using the CMV method [Youn et al. 2005].

2.4 MPP-based RBDO

2.4.1 MPP-Based RBDO Using FORM

The mathematical formulation of a general RBDO problem is expressed as

$$\begin{aligned} &\text{minimize} && \text{Cost}(\mathbf{d}) \\ &\text{subject to} && P[G_i(\mathbf{X}) > 0] \le P_{F_i}^{\text{Tar}},\ i = 1,\cdots,nc \\ & && \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U,\quad \mathbf{d} \in \mathbb{R}^{nd} \text{ and } \mathbf{X} \in \mathbb{R}^{nr} \end{aligned} \tag{2.11}$$

where $\mathbf{d} = \{d_i\}^{\mathrm{T}} = \boldsymbol{\mu}(\mathbf{X}), i = 1 \sim nd$ is the design vector, $\mathbf{X} = \{X_i\}^{\mathrm{T}}$ is the vector of random variables, and nc, nd, and nr are the number of probabilistic constraints,

design variables, and random variables, respectively. Using inverse reliability analysis, the i^{th} probabilistic constraint can be rewritten as

$$P[G_i(\mathbf{X}) > 0] - P_{F_i}^{\text{Tar}} \le 0 \Rightarrow G_i(\mathbf{x}^*) \le 0 \tag{2.12}$$

where $G_i(\mathbf{x}^*)$ is the i^{th} probabilistic constraint evaluated at the MPP, $\mathbf{x}^*$, in the X-space, and $P_{F_i}^{\text{Tar}}$ is the target probability of failure for the i^{th} performance measure.

Using FORM, Eq. (2.11) can be rewritten to give

$$\begin{aligned} &\text{minimize} && \text{Cost}(\mathbf{d}) \\ &\text{subject to} && P[G_i(\mathbf{X}) > 0] \le P_{F_i}^{\text{Tar}} = \Phi(-\beta_{t_i}),\ i = 1, \cdots, nc \\ & && \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^{nd} \text{ and } \mathbf{X} \in \mathbb{R}^{nr} \end{aligned} \tag{2.13}$$

where β_{t_i} is the target reliability index for the i^{th} constraint, and the probabilistic constraint can be written as

$$P[G_i(\mathbf{X}) > 0] - \Phi(-\beta_{t_i}) \le 0 \Rightarrow G_i(\mathbf{x}^*_{\text{FORM}}) \le 0 \tag{2.14}$$

where $\mathbf{x}^*_{\text{FORM}}$ is the FORM-based MPP.

To solve the optimization problem given in Eq. (2.13), the sensitivity of the probabilistic constraint in Eq. (2.14) with respect to the design variables $d_i = \mu(X_i)$ is required. Using the chain rule, the sensitivity of the probabilistic constraint with respect to the design variable can be written as

$$\frac{\partial G(\mathbf{x}^*)}{\partial \mathbf{d}} = \left.\frac{\partial G}{\partial \mathbf{d}}\right|_{\mathbf{x}=\mathbf{x}^*} = \sum_{i=1}^{nr} \left.\frac{\partial G}{\partial x_i}\right|_{\mathbf{x}=\mathbf{x}^*} \left.\frac{\partial x_i}{\partial \mathbf{d}}\right|_{\mathbf{x}=\mathbf{x}^*} = \left[\frac{\partial \mathbf{x}}{\partial \mathbf{d}}\right]^{\text{T}}_{\mathbf{x}=\mathbf{x}^*} \left.\frac{\partial G}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^*} \tag{2.15}$$

which can be simplified to give Eq. (2.16) [Gumbert et al. 2003; Hou et al. 2004].

$$\frac{\partial G(\mathbf{x}^*)}{\partial \mathbf{d}} = \left[\frac{\partial \mathbf{x}}{\partial \mathbf{d}}\right]^{\mathrm{T}}_{\mathbf{x}=\mathbf{x}^*} \left.\frac{\partial G}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^*} = \left.\frac{\partial G}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^*} \tag{2.16}$$

2.4.2 MPP-Based RBDO Using DRM

The dimension reduction method was developed to approximate multi-dimensional integration of a function using a function with reduced dimension [Xu and Rahman 2004; Rahman and Xu 2004]. The univariate dimension reduction method is a decomposition of an nr dimensional performance measure function into the summation of one-dimensional functions. Thus, an nr dimensional performance measure function $G(\mathbf{X})$ can be decomposed into a summation of one-dimensional functions at the MPP of the random vector $\mathbf{X}$ given as

$$G(\mathbf{X}) \cong \hat{G}(\mathbf{X}) \equiv \sum_{i=1}^{nr} G(x_1^*, \cdots, x_{i-1}^*, X_i, x_{i+1}^*, \cdots, x_{nr}^*) - (nr-1)G(\mathbf{x}^*) \tag{2.17}$$

where $\mathbf{x}^* = \{x_1^*, x_2^*, \cdots, x_{nr}^*\}^{\mathrm{T}}$ is the FORM-based MPP of the performance measure function $G(\mathbf{X})$ obtained from Eq. (2.10), and nr is the number of random variables. The univariate DRM gives more accurate reliability analysis results compared to FORM [Lee et al. 2010; Lee et al. 2008].

2.5 Sampling-Based RBDO

The MPP-based RBDO methods, FORM, SORM, and DRM, calculate the probability of failure of the performance measure function using approximations. For highly nonlinear problems, these approximations can be inaccurate and thus give unreliable designs. The MPP-based RBDO methods also require the sensitivity information of the performance measure functions, which can be difficult if not

impossible to calculate for real-world engineering problems. Therefore, a sampling-based RBDO method has been developed that uses surrogate models, MCS, and score functions to calculate the probability of failure and the sensitivity of the probabilistic constraints in RBDO [Shi et al. 2012; Song et al. 2011; Zhao 2011; Zhao et al. 2011; Lee et al. 2011].

2.5.1 Sampling-Based Probability of Failure

In sampling-based RBDO, the reliability analysis at both the component and system levels involves the calculation of the probability of failure, denoted by P_F. The probability of failure is calculated using a multi-dimensional integration and can be written as

$$P_F(\boldsymbol{\psi}) \equiv P[\mathbf{X} \in \Omega_F] = \int_{\mathbb{R}^{nr}} I_{\Omega_F}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\psi}) d\mathbf{x} = E\left[I_{\Omega_F}(\mathbf{X}) \right] \tag{2.18}$$

where $\boldsymbol{\psi}$ is a vector of the distribution parameters, which typically includes the mean, $\boldsymbol{\mu}$, and the standard deviation, $\boldsymbol{\sigma}$, of the random input variables $\mathbf{X} = \{X_1, \cdots, X_{nr}\}^{\mathrm{T}}$, $P[\bullet]$ represents a probability measure, Ω_F is the failure set, $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\psi})$ is the joint probability density function (PDF) of $\mathbf{X}$, and $E[\bullet]$ is the expectation operator. The failure set is defined as $\Omega_F \equiv \{\mathbf{x} : G_i(\mathbf{x}) > 0\}$ for component-level reliability analysis of the i^{th} constraint function $G_i(\mathbf{x})$. For series system-level and parallel system-level reliability analysis of nc performance measure functions, the failure set is $\Omega_F \equiv \left\{\mathbf{x} : \bigcup_{i=1}^{nc} G_i(\mathbf{x}) > 0\right\}$ and $\Omega_F \equiv \left\{\mathbf{x} : \bigcap_{i=1}^{nc} G_i(\mathbf{x}) > 0\right\}$, respectively. $I_{\Omega_F}(\mathbf{x})$ in Eq. (2.18) is called the indicator function and is defined as

$$I_{\Omega_F}(\mathbf{x}) \equiv \begin{cases} 1, & \mathbf{x} \in \Omega_F \\ 0, & \text{otherwise} \end{cases} \tag{2.19}$$

In this study, since the mean of the random variables $\mathbf{X}$, $\boldsymbol{\mu} = \{\mu_1, \cdots, \mu_{nr}\}^{\mathrm{T}}$ is the design variable vector, the vector of distribution parameters $\boldsymbol{\psi}$ can be replaced with $\boldsymbol{\mu}$ for the computation of the probability of failure given in Eq. (2.18).

2.5.2 Probabilistic Sensitivity Analysis

For the derivation of the sensitivity of the probability of failure, the following four regularity conditions need to be satisfied [Rubinstein and Shapiro 1993; Rahman 2009; Lee et al. 2011; Zhao 2011].

1. The joint PDF $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})$ is continuous.
2. The mean $\mu_i \in \mathrm{M}_i \subset \mathbb{R}, i = 1, \cdots, nr$, where M_i is an open interval on $\mathbb{R}$.
3. The partial derivative $\partial f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu}) / \partial \mu_i$ exists and is finite for all $\mathbf{x}$ and μ_i. In addition, $P_F(\boldsymbol{\mu})$ is a differentiable function of $\boldsymbol{\mu}$.
4. There exists a Lebesgue integrable dominating function, $r(\mathbf{x})$ for all $\boldsymbol{\mu}$ such that

$$\left| g(\mathbf{x}) \frac{\partial f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} \right| \le r(\mathbf{x}) \tag{2.20}$$

With the four conditions satisfied, taking the partial derivative of Eq. (2.18) with respect to μ_i and using the interchangeability between the differential and integral operators, the sensitivity of the probability of failure is given as [Rahman 2009; Lee et al. 2011]

$$\begin{aligned}\frac{\partial P_F(\boldsymbol{\mu})}{\partial \mu_i} &= \frac{\partial}{\partial \mu_i}\int_{\mathbb{R}^{nr}} I_{\Omega_F}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu}) d\mathbf{x} \\ &= \int_{\mathbb{R}^{nr}} I_{\Omega_F}(\mathbf{x}) \frac{\partial f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} d\mathbf{x} \\ &= \int_{\mathbb{R}^{nr}} I_{\Omega_F}(\mathbf{x}) \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu}) d\mathbf{x} \\ &= E\left[I_{\Omega_F}(\mathbf{x}) \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} \right]\end{aligned} \tag{2.21}$$

The partial derivative of the log function of the joint PDF in Eq. (2.21) with respect to μ_i is called the first-order score function for μ_i and is denoted as

$$s^{(1)}_{\mu_i}(\mathbf{x};\boldsymbol{\mu}) \equiv \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i}. \tag{2.22}$$

As shown in Eq. (2.22), using the first-order score function in the proposed probabilistic sensitivity analysis does not depend on the sensitivity of the performance measure function $G(\mathbf{x})$. The sensitivity of the joint input distribution is used instead and can be calculated analytically. This is shown in Figure 2.2 [Song 2013]; assume the horizontal axis represents the multi-dimensional random variable $\mathbf{x} = [x_1, x_2, ..., x_{nr}]^{\mathrm{T}}$ with the failure region for the j^{th} constraint $G_j(\mathbf{x})$ defined as $G_j(\mathbf{x}) > 0$. The joint input PDF $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})$ is shown in Figure 2.2. When doing deterministic design optimization, at the current design point $\boldsymbol{\mu}$, the sensitivity of the constraint function $G_j(\mathbf{x})$ at point A has to be used. However, when doing RBDO, probabilistic constraints are used, and the sensitivity of the probability of failure with respect to the current design point $\boldsymbol{\mu}$ is used. The probability of failure is shown in the figure as the volume of the grey shaded region under the joint input PDF $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})$. The joint input PDF will move as the design point $\boldsymbol{\mu}$ moves, and the rate of change of the probabilistic constraint will depend on the slope of the natural logarithm of the joint input PDF $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})$ at point B as shown in Figure 2.2 and Eq. (2.22).

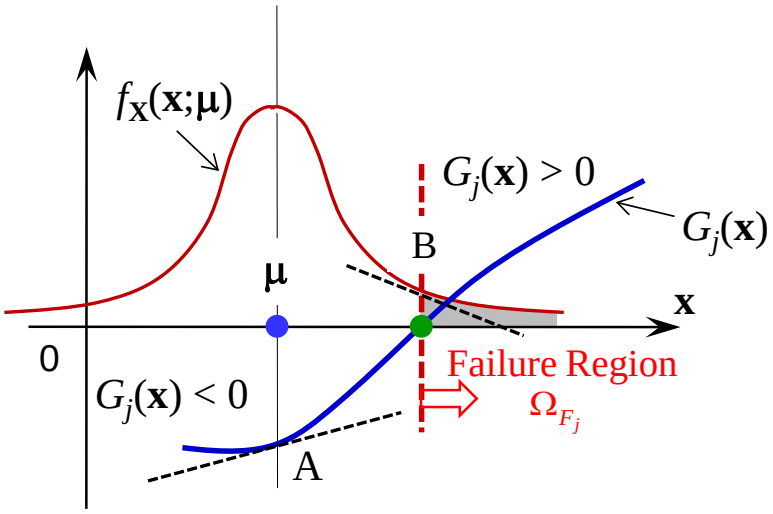

Figure 2.2 Probabilistic Sensitivity Analysis

For independent input random variables, the first-order score function for μ_i in Eq. (2.22) can be written as

$$s_{\mu_i}^{(1)}(\mathbf{x};\boldsymbol{\mu}) \equiv \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} = \frac{\partial \ln f_{X_i}(x_i;\mu_i)}{\partial \mu_i} \tag{2.23}$$

The marginal PDF and CDF are available analytically and are listed in Table 2.2 for some common distributions. Since they are available analytically, the derivation of the first-order score function for independent random variables is straightforward and is listed in Table 2.3 [Rahman 2009; Lee et al. 2011].

Table 2.2 Marginal PDF, CDF, and Parameters

	PDF, $f_X(x)$	CDF, $F_X(x)$	Parameters
Normal	$\frac{1}{\sqrt{2\pi}\sigma}e^{-0.5[\frac{x-\mu}{\sigma}]^2}$	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	μ, σ
Log-normal	$\frac{1}{\sqrt{2\pi}x\bar{\sigma}}e^{-0.5[\frac{\ln x-\bar{\mu}}{\bar{\sigma}}]^2}$	$\Phi\left(\frac{\ln x-\bar{\mu}}{\bar{\sigma}}\right)$	$\bar{\sigma}^2=\ln[1+(\frac{\sigma}{\mu})^2]$, $\bar{\mu}=\ln(\mu)-0.5\bar{\sigma}^2$
Gumbel	$\alpha e^{-\alpha(x-\nu)-e^{-\alpha(x-\nu)}}$	$\alpha e^{-e^{-\alpha(x-\nu)}}$	$\mu=\nu+\frac{0.577}{\alpha}$, $\sigma=\frac{\pi}{\sqrt{6}\alpha}$
Weibull	$\frac{k}{\nu}(\frac{x}{\nu})^{k-1}e^{-(\frac{x}{\nu})^k}$	$1-e^{-(\frac{x}{\nu})^k}$	$\mu=\nu\Gamma(1+\frac{1}{k})$, $\sigma^2=\nu^2[\Gamma(1+\frac{2}{k})-\Gamma^2(1+\frac{1}{k})]$

Table 2.3 First-Order Score Function for μ_i for Independent Random Variables

Marginal Distribution	First-Order Score Function, $s_{\mu_i}^{(1)}(\mathbf{x};\boldsymbol{\mu})$
Normal	$\frac{x_i-\mu_i}{\sigma_i^2}$
Log-normal	$-\frac{1}{\bar{\sigma}_i}\frac{\partial\bar{\sigma}_i}{\partial\mu_i}+\frac{1}{\bar{\sigma}_i^2}\left(\frac{\ln x_i-\bar{\mu}_i}{\bar{\sigma}_i}\right)\times\left[\bar{\sigma}_i\frac{\partial\bar{\mu}_i}{\partial\mu_i}+(\ln x_i-\bar{\mu}_i)\frac{\partial\bar{\sigma}_i}{\partial\mu_i}\right]$
Gumbel	$\alpha_i-\alpha_i e^{-\alpha_i(x_i-\nu_i)}$
Weibull	$\frac{1}{k_i}\frac{\partial k_i}{\partial\mu_i}-\frac{1}{\nu_i}\frac{\partial\nu_i}{\partial\mu_i}+\frac{\partial k_i}{\partial\mu_i}\ln\frac{x_i}{\nu_i}-\frac{(k_i-1)}{\nu_i}\frac{\partial\nu_i}{\partial\mu_i}-(\frac{x_i}{\nu_i})^{k_i}\left(\frac{\partial k_i}{\partial\mu_i}\ln\frac{x_i}{\nu_i}-\frac{k_i}{\nu_i}\frac{\partial\nu_i}{\partial\mu_i}\right)$

For bivariate correlated input random variables $\mathbf{X}=\{X_i,X_j\}^{\mathrm{T}}$, the joint PDF of $\mathbf{X}$ using copula functions can be expressed as [Noh et al. 2009; Noh et al. 2010; Lee et al. 2011]

$$f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu}) = \frac{\partial^2 C(u,v;\theta)}{\partial u \partial v} f_{X_i}(x_i;\mu_i) f_{X_j}(x_j;\mu_j) = C_{,uv}(u,v;\theta) f_{X_i}(x_i;\mu_i) f_{X_j}(x_j;\mu_j) \quad (2.24)$$

where C is the copula function; $u = F_{X_i}(x_i;\mu_i)$ and $v = F_{X_j}(x_j;\mu_j)$ are marginal CDFs for X_i and X_j, respectively; and θ is the correlation coefficient between X_i and X_j. The partial derivative of the copula function with respect to u and v is called the copula density function and is written as

$$c(u,v;\theta) \equiv \frac{\partial^2 C(u,v;\theta)}{\partial u \partial v} = C_{,uv}(u,v;\theta). \quad (2.25)$$

Using Eq. (2.24), the first-order score function in Eq. (2.22) for bivariate correlated variables $\mathbf{X}$ is given as

$$s^{(1)}_{\mu_i}(\mathbf{x};\boldsymbol{\mu}) \equiv \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} = \frac{\partial \ln c(u,v;\theta)}{\partial \mu_i} + \frac{\partial \ln f_{X_i}(x_i;\mu_i)}{\partial \mu_i} \quad (2.26)$$

Table 2.4 lists the derivation of the first term on the right-hand side of Eq. (2.26).

Table 2.4 Log-Derivative of Copula Density Function

Copula Type	$\dfrac{\partial \ln c(u,v;\theta)}{\partial \mu_i}$
Clayton	$\left(-\dfrac{1+\theta}{u}+\dfrac{(2\theta+1)u^{-(1+\theta)}}{u^{-\theta}+v^{-\theta}-1}\right)\dfrac{\partial u}{\partial \mu_i}$
AMH	$\left[\dfrac{-\theta^2(1-v)+\theta(v+1)}{1+\theta^2(1-u)(1-v)-\theta(2-u-v-uv)}-\dfrac{3\theta(1-v)}{1-\theta(1-u)(1-v)}\right]\dfrac{\partial u}{\partial \mu_i}$
Frank	$\theta\left[\dfrac{2\left(e^{\theta(1+u)}-e^{\theta(u+v)}\right)}{e^{\theta}-e^{\theta(1+u)}-e^{\theta(1+v)}+e^{\theta(u+v)}}+1\right]\dfrac{\partial u}{\partial \mu_i}$
FGM	$\left[\dfrac{2\theta(2v-1)}{1+\theta(1-2u)(1-2v)}\right]\dfrac{\partial u}{\partial \mu_i}$
Gaussian	$\left[\dfrac{\Phi^{-1}(u)}{\phi(\Phi^{-1}(u))}+\dfrac{\theta\Phi^{-1}(v)-\Phi^{-1}(u)}{\phi(\Phi^{-1}(u))(1-\theta^2)}\right]\dfrac{\partial u}{\partial \mu_i}$
Independent	0

Thus, the sensitivity of the probability of failure in Eq. (2.21) can be easily calculated for bivariate correlated random variables.

As described previously, surrogate models are used to calculate the probability of failure because it is computationally too expensive to carry out direct MCS of CAE models. If the surrogate model for the j^{th} constraint function $G_j(\mathbf{x})$ is denoted as $\hat{G}_j(\mathbf{x})$ then the probability of failure can be approximated as

$$P_{F_j} \equiv P\left[G_j(\mathbf{x})>0\right] \cong \frac{1}{M}\sum_{m=1}^{M} I_{\hat{\Omega}_{F_j}}(\mathbf{x}^{(m)}) \le P_{F_j}^{\text{Tar}} \tag{2.27}$$

where M is the number of MCS samples, $\mathbf{x}^{(m)}$ is the m^{th} realization of $\mathbf{X}$, and the failure region $\hat{\Omega}_{F_j}$ for the surrogate model is defined as $\hat{\Omega}_{F_j} \equiv \{\mathbf{x}:\hat{G}_j(\mathbf{x})>0\}$. The

sensitivity of the probabilistic constraint in Eq. (2.21) can be approximated using [Rahman 2009; Lee et al. 2011]

$$\frac{\partial P_{F_j}}{\partial \mu_i} \cong \frac{1}{M}\sum_{m=1}^{M} I_{\hat{\Omega}_{F_j}}(\mathbf{x}^{(m)}) s_{\mu_i}^{(1)}(\mathbf{x}^{(m)};\boldsymbol{\mu}). \tag{2.28}$$

The number of MCS samples and the target probability of failure controls the accuracy of the MCS. Based on the 95% confidence interval of the estimated probability of failure, the percentage error can be defined as [Haldar and Mahadevan 2000]

$$\varepsilon_{MCS} = \sqrt{\frac{(1-P_F^{\text{Tar}})}{M \times P_F^{\text{Tar}}}} \times 200\% \tag{2.29}$$

Thus, for a small non-zero target probability of failure, the number of MCS samples used needs to be increased to maintain a given target accuracy level.

This section only considered component-level probability of failure. However, as long as the failure set is defined appropriately for the system level, then Eqs. (2.27) and (2.28) can be used for system-level RBDO. The accuracy and sensitivity of the probability of failure depend not only on the number of MCS samples used but also on the accuracy of the surrogate model used. Thus, to perform confidence-based RBDO, a surrogate modeling method that captures the uncertainty of the probability of failure due to the uncertainty in the surrogate model is required. It is developed in the next chapter.

CHAPTER 3

MODIFIED BAYESIAN KRIGING

3.1 Introduction

As previously discussed, Kriging has become a popular surrogate modeling choice. However, because Kriging is typically formulated to be used as an interpolation method and is known to break down when the response values contain noise, it is not ideally suited for creating surrogate models for problems that contain noise. There is a way to modify the correlation matrix used in Kriging so that it becomes a regression method, but it does not offer a convenient way to separate the noise in the response values from the true underlying response values. In this chapter a modified Bayesian Kriging (MBKG) method is developed that can handle response values that contain noise, so that predicted response values without noise can be obtained from the MBKG model. The Kriging and dynamic Kriging methods are briefly reviewed in Section 3.2. The proposed modified Bayesian Kriging (MBKG) method is developed in Section 3.3, and the full conditional distributions used to fit the MBKG surrogate model are derived in Section 3.4.

3.2 Kriging and Dynamic Kriging Methods

3.2.1 Kriging Method

The Kriging method is based on the assumption that the response values of interest are a realization from a stochastic process. Consider N design of experiment (DoE) samples, denoted as $\mathbf{x}_{DoE} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)^{\mathrm{T}}$, and the corresponding N responses, denoted as $\mathbf{y} = (y_1, y_2, ..., y_N)^{\mathrm{T}}$, where $\mathbf{x}_{DoE} \in \mathbb{R}^m$ and m is the number of input variables, i.e., the spatial dimension of the input variables. The Kriging model for the responses is composed of two parts and is expressed mathematically as

$$\mathbf{y} = \mathbf{F}\boldsymbol{\beta} + \mathbf{Z} \tag{3.1}$$

where $\mathbf{F}\boldsymbol{\beta}$ is the mean structure of the responses, $\mathbf{F} = [\mathbf{f}_j(\mathbf{x}_i)], i = 1,2,...,N,\ j = 1,2,...,K$ is a $N \times K$ design matrix where $\mathbf{f}_j(\mathbf{x}_i)$ is the j^{th} basis function evaluated at the i^{th} DoE sample point, and $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_K]^{\text{T}}$ are the regression coefficients from the generalized least square regression method. The second part, $\mathbf{Z}$, is a realization from a stationary Gaussian random process with zero mean and covariance function given by

$$\Sigma(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2 R(\boldsymbol{\theta}, \mathbf{x}_i, \mathbf{x}_j) \tag{3.2}$$

where σ^2 is the process variance, R is the spatial correlation function, $\boldsymbol{\theta}$ is a vector containing the correlation function parameters, and $\mathbf{x}_i$, $\mathbf{x}_j$ are two DoE samples [Zhao 2011]. There are a number of different correlation functions that can be used, e.g., Gaussian, exponential, general exponential, linear, spherical, cubic, and spline. However, for engineering problems, the Gaussian correlation function is commonly used since it is infinitely differentiable and provides a smooth response surface. The parameters of the Kriging model that best fit the DoE samples and corresponding response values are determined by using the maximum likelihood estimation (MLE) method.

3.2.2 Dynamic Kriging Method

As previously mentioned, there are a number of different correlation functions to choose from, and there are also numerous basis functions to choose from. It has been shown that for different problems and data some correlation functions fit better than others; it has also been shown that a larger basis function set is not necessarily better than a smaller basis function set [Song et al. 2013; Zhao 2011; Zhao et al. 2011]. The latest proposed method chooses the mean structure from a choice of three mean structures and the correlation function from seven choices that best fit the data [Song et al. 2013]. This

dynamic Kriging method provides more flexibility and an automated way to build a Kriging model that can fit data from a wide range of problems. However, it is an interpolation method and breaks down when the response values contain noise.

3.3 Modified Bayesian Kriging (MBKG)

3.3.1 Modified Bayesian Kriging Formulation

The basic assumption of Kriging is that the response values are realizations from a Gaussian random process. Thus, fitting a Kriging model is ideally suited for use with Bayesian statistical methods. Kriging is commonly referred to as a Bayesian method, though it is typically fitted using MLE methods rather than Bayesian methods. Fitting a Kriging model using Bayesian methods is ideal because the assumption is that the response values are realizations from a Gaussian random process. As described in Section 1.2, the first part of a Bayesian analysis is determining the likelihood for the data. Once the likelihood is known, the next step is deciding the prior distributions to be used for the unknown parameters in the likelihood. Then, using the likelihood and prior distributions, the Bayesian analysis can be carried out to determine the posterior distribution of the unknown parameters.

The modified Bayesian Kriging method assumes the response values come from a stationary Gaussian random process with a constant mean structure in the following form:

$$\mathbf{y} \sim MVN\left(\mu_c \mathbf{1} + \boldsymbol{\varphi}, \sigma^2 \lambda \mathbf{I}\right) \tag{3.3}$$

where $\mathbf{y}$ is the vector of response values at the DoE sample points. The above formulation states that the response values are conditionally independent given parameters defining a variance of $\sigma^2\lambda$ and a mean value of $\mu_c\mathbf{1}+\boldsymbol{\varphi}$, which is composed of two parts; the first part is a constant value of μ_c, and the second part is $\boldsymbol{\varphi}$, which

depends on the spatial correlation of the DoE samples $\mathbf{x}$. The second assumption is that $\boldsymbol{\varphi}$ also follows a Gaussian random process with a zero mean and a covariance matrix that depends on the spatial correlation of the DoE samples $\mathbf{x}$; this is expressed mathematically as

$$\boldsymbol{\varphi} \sim MVN\left(\mathbf{0}, \sigma^2 \boldsymbol{\Psi}\right) \tag{3.4}$$

where $\sigma^2\boldsymbol{\Psi}$ is the covariance of the Gaussian process, σ^2 is the variance of the $\boldsymbol{\varphi}$ values, and $\boldsymbol{\Psi}$ is the spatial correlation matrix. Similar to the Kriging method, the spatial correlation matrix is a function of the DoE samples $\mathbf{x}$ via the correlation function that is used. As described in Section 3.2.1, there are a number of different correlation functions that can be used, and the Gaussian correlation function is the most commonly used for engineering problems and is defined as

$$\Psi^{(i)} = \exp\left(-\sum_{j=1}^{k} \theta_j \left| x_j^{(i)} - x_j \right|^2 \right) \tag{3.5}$$

where θ_j is the j^{th} correlation function parameter corresponding to the j^{th} dimension, k is the number of spatial dimensions, x_j is the value of the DoE sample for the j^{th} dimension, and superscript i is for the i^{th} DoE sample.

In this study, seven different correlation functions from the literature are considered and are listed in Table 3.1 [Song et al. 2013; Lophaven et al. 2002].

Table 3.1 Correlation Functions

Name	$\Psi^{(i)}$
Gaussian	$\exp\left(-\sum_{j=1}^{k}\theta_j\left\|x_j^{(i)}-x_j\right\|^{p_j}\right)$
Exponential	$\exp\left(-\theta_j\left\|x_j^{(i)}-x_j\right\|\right)$
General Exponential	$\exp\left(-\theta_j\left\|x_j^{(i)}-x_j\right\|^{\theta_{n+1}}\right), 0<\theta_{n+1}\le 2$
Linear	$\max\left\{0,1-\theta_j\left\|x_j^{(i)}-x_j\right\|\right\}$
Spherical	$1-1.5\xi_j+0.5\xi_j^3,\ \xi_j=\min\left\{1,\theta_j\left\|x_j^{(i)}-x_j\right\|\right\}$
Cubic	$1-3\xi_j^2+2\xi_j^3,\ \xi_j=\min\left\{1,\theta_j\left\|x_j^{(i)}-x_j\right\|\right\}$
Spline	$\begin{cases}1-15\xi_j^2+30\xi_j^3, & \text{for } 0\le\xi_j\le 0.2\\ 1.25(1-\xi_j)^3, & \text{for } 0<\xi_j<1\\ 0, & \text{for } \xi_j\ge 1,\end{cases}$ where $\xi_j=\theta_j\left\|x_j^{(i)}-x_j\right\|$

The unknown MBKG parameters that need to be determined in order to fit the model to the data, i.e., fit the model to the DoE samples $\mathbf{x}$ and their corresponding response values $\mathbf{y}$, are μ_c, σ^2, λ, and $\boldsymbol{\theta}$. The dimension of the $\boldsymbol{\theta}$ vector is the same as the spatial dimension of $\mathbf{x}$. From the Bayesian perspective, the likelihood of the response values is the Gaussian process given in Eq. (3.3), and the prior for the $\boldsymbol{\varphi}$ vector is the Gaussian process given in Eq. (3.4). The prior distributions for the remaining unknown parameters are presented in the next section.

The modified Bayesian Kriging method can also be formulated to have a mean structure that is a function of the DoE samples $\mathbf{x}$. For this formulation Eq. (3.3) is modified to give:

$$\mathbf{y} \sim MVN\left(\boldsymbol{\mu} + \boldsymbol{\varphi}, \sigma^2 \lambda \mathbf{I}\right) \tag{3.6}$$

where $\mu_c \mathbf{1}$ is replaced by the vector $\boldsymbol{\mu}$, which is a function of the DoE sample points. The $\boldsymbol{\mu}$ vector is expressed as:

$$\boldsymbol{\mu} = \mathbf{F}\boldsymbol{\beta} \tag{3.7}$$

where $\mathbf{F}$ is the design matrix composed of the polynomial basis functions evaluated at the DoE sample points. For first order the design matrix $\mathbf{F}$ is expressed as

$$\mathbf{F} = \begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_n^{(1)} \\ 1 & x_1^{(2)} & & x_n^{(2)} \\ \vdots & & \ddots & \vdots \\ 1 & x_1^{(m)} & \cdots & x_n^{(m)} \end{bmatrix}_{m \times (n+1)} \tag{3.8}$$

where m is the number of DoE and n is the number of variables, i.e., the dimension of the problem, and $\boldsymbol{\beta}$ is a vector of unknown coefficients to be determined when fitting the MBKG surrogate model.

3.3.2 Prior Distributions for MBKG Parameters

For the three unknown parameters μ_c, σ^2, and λ, semi-conjugate prior distributions are used to help improve the efficiency of fitting the MBKG surrogate model. For the μ_c parameter, the conjugate prior distribution is a normal distribution and is expressed as

$$\mu_c \sim N\left(\mu_p, \sigma_p^2\right) \tag{3.9}$$

where μ_p and σ_p^2 are the prior mean and variance of the μ_c parameter, respectively. The conjugate prior distribution for the σ^2 parameter is an Inverse-Gamma distribution and is expressed as

$$\sigma^2 \sim InverseGamma\left(\alpha_\sigma, \beta_\sigma\right) \tag{3.10}$$

where α_σ and β_σ are the prior parameters for the distribution. The conjugate prior for the λ parameter is also an Inverse-Gamma distribution and is expressed as

$$\lambda \sim InverseGamma\left(\alpha_\lambda, \beta_\lambda\right) \tag{3.11}$$

where α_λ and β_λ are the prior parameters for the distribution. The $\boldsymbol{\theta}$ parameters are embedded in the correlation matrix, and because of this there is no known conjugate distribution type that can be used for the prior distribution. Thus, the prior distribution for each of the $\boldsymbol{\theta}$ parameters is chosen to be a uniform distribution and is expressed as

$$\theta_j \sim \mathrm{U}\left(a_{\theta_j}, \mathrm{b}_{\theta_j}\right) \tag{3.12}$$

where θ_j is the j^{th} correlation function parameter, and a_{θ_j} and b_{θ_j} are the prior parameters for θ_j.

If the mean structure is not constant, then the $\boldsymbol{\mu}$ vector is used as in Eq. (3.6) and the conjugate prior is a multivariate normal distribution and is expressed as

$$\boldsymbol{\beta} \sim MVN\left(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right) \tag{3.13}$$

where $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\Sigma}_\beta$ are the prior mean vector and covariance matrix for the distribution.

Prior to fitting the MBKG model, the response values $\mathbf{y}$ are normalized such that they have a zero mean. The normalization is expressed as

$$y_i = \frac{\tilde{y}_i - mean(\tilde{\mathbf{y}})}{std(\tilde{\mathbf{y}})} \tag{3.14}$$

where y_i is the normalized response value of the i^{th} response, $\tilde{y}_i$ is the un-normalized response value for the i^{th} response, $mean(\tilde{\mathbf{y}})$ is the mean of the un-normalized response values, and $std(\tilde{\mathbf{y}})$ is the standard deviation of the un-normalized response value. The DoE samples are also normalized in the same way as the response values. In the literature it is has been found that, when fitting surrogate models a better fit can usually be achieved when the DoE samples and response values are normalized in some way to help restrict the range of the possible values than when they are not normalized [Forrester et al. 2008].

In addition to the normal distribution being a conjugate prior for the μ_c parameter, it is also a good choice because of the normalization of the response values. It is expected that the overall mean of the normalized response values should be close to zero. Thus, a normal distribution with a zero mean and a standard deviation of 0.5 is a prior distribution that reflects this knowledge. This prior distribution can still be considered a relatively noninformative prior distribution; thus, it can be used in general when creating a surrogate model for any problem.

The Inverse-Gamma distribution has positive support, meaning that the random variable of an Inverse-Gamma distribution can only take on positive values. From Eqs. (3.3) and (3.4) it can be seen that both parameters σ^2 and λ are variance parameters to a Gaussian process; therefore, by definition they can only take on positive values. With the Inverse-Gamma distribution being a positive value distribution and a conjugate prior for both σ^2 and λ, it is ideally suited to be used as the prior distribution.

As mentioned previously, there is no known conjugate prior distribution for the $\boldsymbol{\theta}$ parameters. Thus, a uniform distribution was chosen mostly because it gives a direct way to limit the possible values that the $\boldsymbol{\theta}$ parameters can take on. Due to the normalization of

the DoE samples, the maximum distance in each of the spatial directions from any given DoE sample to all the others is roughly three. After studying the Gaussian correlation function closely, it was found that a θ value of five and greater produces a correlation value that is asymptotically approaching zero regardless of the distance between the DoE samples. This is shown in Figure 3.1; the vertical axis is the squared distance between DoE sample points, and the horizontal axis is the correlation function parameter θ. As seen in the figure, when $\theta \approx 2.5$ and the squared distance is greater than one, the correlation is smaller than 0.1. Also seen in the figure is that when $\theta = 5$, the correlation is smaller than 0.1 except for squared distances that are approximately less than 0.5. Given this study, it was determined that an appropriate upper bound for θ would be $\theta = 5$ and an appropriate lower bound would be $\theta = 0.15$; these are the bounds used in the uniform prior distribution.

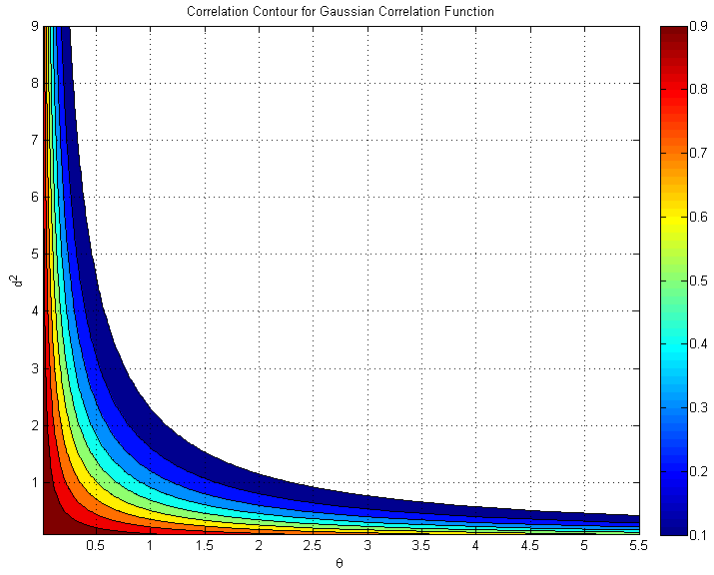

Figure 3.1 Correlation Contour of the Gaussian Correlation Function

Once the likelihood and prior distributions have been determined, the next step in a Bayesian analysis that uses the Gibbs sampling algorithm [Gelfand and Smith 1990] for Markov Chain Monte Carlo (MCMC) is to determine the full conditional distributions for the unknown parameters. The next section will show the derivation of the full conditional distributions.

3.4 Full Conditional Distributions for MBKG Parameters

3.4.1 Joint Distribution of the MBKG Parameters

The joint posterior distribution of the unknown parameters in a Bayesian model is needed to derive the full conditional distributions of the unknown parameters being estimated. The full conditional distribution for the unknown parameter is derived from the joint distribution such that the full conditional distribution is only a function of the unknown parameters that the parameter of interest depends on. For the likelihood and prior distributions that were presented in Section 3.3, the joint posterior distribution of the MBKG surrogate model for the constant mean structure is the product of the likelihood and all the prior distributions, which is given as

$$\begin{aligned} f\left(\mu_c, \sigma^2, \boldsymbol{\theta}, \lambda, \boldsymbol{\varphi} \mid \mathbf{y}\right) \propto & \prod_{j=1}^{k}\left[\mathrm{U}\left(a_{\theta_j}, \mathrm{b}_{\theta_j}\right)\right] * \mathit{InverseGamma}\left(\alpha_\sigma, \beta_\sigma\right) \\ & * N\left(\mu_p, \sigma_p^2\right) * \mathit{InverseGamma}\left(\alpha_\lambda, \beta_\lambda\right) \\ & * \mathit{MVN}\left(\mathbf{0}, \sigma^2 \boldsymbol{\Psi}\right) * \mathit{MVN}\left(\boldsymbol{\mu} + \boldsymbol{\varphi}, \sigma^2 \lambda \mathbf{I}\right) \end{aligned} \tag{3.15}$$

The joint posterior distribution of the MBKG surrogate model considering the mean structure is expressed as

$$f\left(\boldsymbol{\mu},\sigma^2,\boldsymbol{\theta},\lambda,\boldsymbol{\varphi}\,|\,\mathbf{y}\right)\propto\prod_{j=1}^{k}\left[\mathrm{U}\left(a_{\theta_j},\mathrm{b}_{\theta_j}\right)\right]*InverseGamma\left(\alpha_\sigma,\beta_\sigma\right)$$
$$*MVN\left(\boldsymbol{\mu}_\beta,\boldsymbol{\Sigma}_\beta\right)*InverseGamma\left(\alpha_\lambda,\beta_\lambda\right) \tag{3.16}$$
$$*MVN\left(\mathbf{0},\sigma^2\boldsymbol{\Psi}\right)*MVN\left(\boldsymbol{\mu}+\boldsymbol{\varphi},\sigma^2\lambda\mathbf{I}\right)$$

Substituting in the mathematical expression for each of the distributions in Eq. (3.15) gives the joint distribution as a function of the unknown parameters and the prior parameters and is given in Eq. (3.17).

The Bayesian model proposed for the MBKG surrogate modeling method is not a new Bayesian model. The full conditional distributions of this Bayesian model have been previously derived by others. However, the use of this Bayesian model as a surrogate model is a new concept that has not previously been done. The methods developed in this study have been coded in MATLAB so that they can be easily integrated with existing RBDO methods that were coded in MATLAB. In order to code the Gibbs sampling algorithm needed to fit the MBKG surrogate model using MCMC, the full conditional distributions are needed. For this reason they have been derived again and are presented in this study for completeness. To derive the full conditional distribution for a parameter of interest, only the terms from the joint distribution that contains the parameter of interest are used. The full conditional distribution of the parameter of interest is proportional to the terms taken from the joint distribution. The next sections derive the full conditional distributions for each of the MBKG parameters. Note that the derivations in the next sections only show partial steps of the derivation.

$$
\begin{aligned}
f\left(\mu_c,\sigma^2,\boldsymbol{\theta},\lambda,\boldsymbol{\varphi}\,|\,\mathbf{y}\right) \propto & \prod_{j=1}^{k}\left[\frac{1}{b_{\theta_j}-a_{\theta_j}}\right] \\
& * \frac{\beta_\sigma^{\alpha_\sigma}}{\Gamma(\alpha_\sigma)}(\sigma^2)^{-(\alpha_\sigma+1)}\exp\left(-\beta_\sigma\cdot\frac{1}{\sigma^2}\right) \\
& * \frac{1}{\sqrt{2\pi}\sigma_p}\exp\left(-\frac{1}{2\sigma_p^2}\left(\mu_c-\mu_p\right)^2\right) \\
& * \frac{\beta_\lambda^{\alpha_\lambda}}{\Gamma(\alpha_\lambda)}\lambda^{-(\alpha_\lambda+1)}\exp\left(-\beta_\lambda\cdot\frac{1}{\lambda}\right) \\
& * \left(2\pi\right)^{-d/2}\left|\boldsymbol{\Sigma}\right|^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{\varphi}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi}\right) \\
& * \left(2\pi\right)^{-d/2}\left|\sigma^2\lambda\mathbf{I}\right|^{-1/2} \\
& * \exp\left(-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)^T\left(\sigma^2\lambda\mathbf{I}\right)^{-1}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)\right)
\end{aligned}
\tag{3.17}
$$

3.4.2 Full Conditional for μ_c

Taking only the terms that depend on μ_c from the joint distribution in Eq. (3.17) gives the terms that are a function of the full conditional for μ_c, which gives

$$
\begin{aligned}
f\left(\mu_c\,|\,\mathbf{y},\sigma^2,\boldsymbol{\theta},\lambda,\boldsymbol{\varphi}\right) \propto & \exp\left(-\frac{1}{2\sigma_p^2}\left(\mu_c-\mu_p\right)^2\right) \\
& * \exp\left(-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)^T\left(\sigma^2\lambda\mathbf{I}\right)^{-1}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)\right)
\end{aligned}
\tag{3.18}
$$

Using index notation and grouping terms that are a function μ_c in Eq. (3.18), the right-hand side can be written to give Eq. (3.19). Note that the full conditional in Eq. (3.18) is written as proportional to, so that constant terms, i.e., terms that are not a function of μ_c, can be dropped from the equation.

$$\propto \exp\left(-\frac{1}{2\sigma_p^2}\left(\mu_c-\mu_p\right)^2\right)\exp\left(-\frac{1}{2}\frac{1}{\sigma^2\lambda}\left(I_i I_{ij} I_j \mu_c^2 + 2 I_i I_{ij}\varphi_j \mu_c - 2 y_i I_{ij} I_j \mu_c\right)\right) \underbrace{\exp\left(-\frac{1}{2}\frac{1}{\sigma^2\lambda}\left(y_i I_{ij} y_j - 2 y_i I_{ij}\varphi_j + \varphi_i I_{ij}\varphi_j\right)\right)}_{\text{constant}} \tag{3.19}$$

After removing the constant term and defining the intermediate variables A and B, Eq. (3.19) can be further simplified to give

$$\propto \exp\left(-\frac{1}{2\sigma_p^2}\left(\mu_c-\mu_p\right)^2\right)\exp\left(-\frac{1}{2}\frac{1}{\sigma^2\lambda}\left(\underbrace{I_i I_{ij} I_j}_{A}\mu_c^2 + -2\underbrace{\left(y_i I_{ij} I_j - I_i I_{ij}\varphi_j\right)}_{B}\mu_c\right)\right) \tag{3.20}$$

Precision is defined as the inverse of the variance so that prior precision and unknown precision can be expressed as

$$\tau_p^2 = \frac{1}{\sigma_p^2} \qquad \tau^2 = \frac{1}{\sigma^2\lambda} \tag{3.21}$$

Rewriting Eq. (3.20) in terms of precision and the intermediate variables A and B gives

$$\propto \exp\left(-\frac{1}{2}\tau_p^2\left(\mu_c^2 - 2\mu_p\mu_c + \mu_p^2\right)\right)\exp\left(-\frac{1}{2}\tau^2\left(A\mu_c^2 + -2B\mu_c\right)\right) \tag{3.22}$$

Once again terms that depend on μ_c are grouped together, and constant terms are dropped from the expression so that Eq. (3.22) is further simplified to give

$$\propto \exp\left(-\frac{1}{2}\left(\tau_p^2 + \tau^2 A\right)\left(\mu_c^2 - 2\frac{\left(\tau_p^2\mu_p + \tau^2 B\right)}{\left(\tau_p^2 + \tau^2 A\right)}\mu_c\right)\right) \tag{3.23}$$

The final step is then completing the square and dropping constant terms from the expression so that Eq. (3.23) is simplified to give

$$\propto \exp\left(-\frac{1}{2}\left(\tau_p^2+\tau^2 A\right)\left(\mu_c-\frac{\left(\tau_p^2\mu_p+\tau^2 B\right)}{\left(\tau_p^2+\tau^2 A\right)}\right)^2\right) \tag{3.24}$$

Equation (3.24) can be recognized as a kernel of a normal distribution that has a mean value of

$$\frac{\left(\tau_p^2\mu_p+\tau^2 B\right)}{\left(\tau_p^2+\tau^2 A\right)} \tag{3.25}$$

The precision and variance of this normal distribution can be seen to be

$$\text{precision}=\left(\tau_p^2+\tau^2 A\right)\Rightarrow \text{var}=\frac{1}{\left(\tau_p^2+\tau^2 A\right)} \tag{3.26}$$

The intermediate variables are defined as

$$\begin{aligned} A &= I_i I_{ij} I_j & B &= y_i I_{ij} I_j - I_i I_{ij}\varphi_j \\ \tau_p^2 &= \frac{1}{\sigma_p^2} & \tau^2 &= \frac{1}{\sigma^2\lambda} \end{aligned} \tag{3.27}$$

The full conditional distribution of μ_c written in terms of the precision, intermediate variables, prior parameters, and MBKG parameters can be expressed as

$$f\left(\mu_c \mid \mathbf{y},\sigma^2,\boldsymbol{\theta},\lambda,\boldsymbol{\varphi}\right)\sim N\left(\frac{\left(\tau_p^2\mu_p+\tau^2 B\right)}{\left(\tau_p^2+\tau^2 A\right)},\left(\tau_p^2+\tau^2 A\right)\right) \tag{3.28}$$

3.4.3 Full Conditional for σ^2

Taking only the terms that depend on σ^2 from the joint distribution in Eq. (3.17) gives the terms that are a function of the full conditional for σ^2, which gives

$$\begin{aligned} f\left(\sigma^2 \mid \mathbf{y}, \mu_c, \boldsymbol{\theta}, \lambda, \boldsymbol{\varphi}\right) \propto (\sigma^2)^{-(\alpha_\sigma+1)} \exp\left(-\beta_\sigma \cdot \frac{1}{\sigma^2}\right) |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\varphi}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varphi}\right) \\ * \left|\sigma^2 \lambda \mathbf{I}\right|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi})^T \left(\sigma^2 \lambda \mathbf{I}\right)^{-1} (\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi})\right) \end{aligned} \tag{3.29}$$

After grouping σ^2 terms and constant terms, the right-hand side of Eq. (3.29) can be rewritten to give

$$\begin{aligned} &\propto (\sigma^2)^{-(\alpha_\sigma+1)} \left(\sigma^2\right)^{-n/2} \left(\sigma^2\right)^{-n/2} \exp\left(-\beta_\sigma \cdot \frac{1}{\sigma^2}\right) \\ &* \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi})^T \left(\sigma^2 \lambda \mathbf{I}\right)^{-1} (\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi})\right) \\ &* \exp\left(-\frac{1}{2}\boldsymbol{\varphi}^T \left(\sigma^2 \boldsymbol{\Psi}\right)^{-1} \boldsymbol{\varphi}\right) * \underbrace{\lambda^{-n/2} |\boldsymbol{\Psi}|^{-1/2} |\mathbf{I}|^{-1/2}}_{\text{constant}} \end{aligned} \tag{3.30}$$

Dropping the constant terms and further simplifying Eq. (3.30) can finally be written as

$$\propto (\sigma^2)^{-(\alpha_\sigma+n+1)} \exp\left(-\left(\beta_\sigma + \frac{1}{2}\boldsymbol{\varphi}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\varphi} + \frac{1}{2}\frac{1}{\lambda}(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi})^T (\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi})\right) \cdot \frac{1}{\sigma^2}\right) \tag{3.31}$$

Equation (3.31) can be identified as the kernel of an Inverse-Gamma distribution with the first parameter of the distribution given as $\alpha_\sigma + n$ and the second parameter of the distribution given as

$$\beta_\sigma + \frac{1}{2}\boldsymbol{\varphi}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\varphi} + \frac{1}{2}\frac{1}{\lambda}(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi})^T (\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}) \tag{3.32}$$

The full conditional distribution of σ^2 written in terms of the prior parameters and MBKG parameters can be expressed as

$$f\left(\sigma^2 \mid \mathbf{y}, \mu_c, \boldsymbol{\theta}, \lambda, \boldsymbol{\varphi}\right) \sim InverseGamma\left(\alpha_\sigma + n, \beta_\sigma + \frac{1}{2}\boldsymbol{\varphi}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\varphi} + \frac{1}{2}\frac{1}{\lambda}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)^T\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)\right) \tag{3.33}$$

3.4.4 Full Conditional for λ

Taking only the terms that depend on λ from the joint distribution in Eq. (3.17) gives the terms that are a function of the full conditional for λ, which gives

$$\begin{aligned} f\left(\lambda \mid \mathbf{y}, \mu_c, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\varphi}\right) \propto \lambda^{-(\alpha_\lambda+1)} \exp\left(-\beta_\lambda \cdot \frac{1}{\lambda}\right)\left|\sigma^2\lambda\mathbf{I}\right|^{-1/2} \\ * \exp\left(-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)^T\left(\sigma^2\lambda\mathbf{I}\right)^{-1}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)\right) \end{aligned} \tag{3.34}$$

Grouping terms that are a function of λ and dropping constant terms, the right-hand side of Eq. (3.34) can be rewritten to give

$$\propto \lambda^{-(\alpha_\lambda+\frac{n}{2}+1)} \exp\left(-\left(\beta_\lambda + \frac{1}{2}\frac{1}{\sigma^2}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)^T\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)\right)\cdot\frac{1}{\lambda}\right) \tag{3.35}$$

Equation (3.35) can be identified to be a kernel to an Inverse-Gamma distribution with the first parameter given as $\alpha_\lambda + n/2$ and the second parameter given as

$$\beta_\lambda + \frac{1}{2}\frac{1}{\sigma^2}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)^T\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right) \tag{3.36}$$

The full conditional distribution of λ as a function of the prior parameters and MBKG parameters can be expressed as

$$f\left(\lambda \mid \mathbf{y}, \mu_c, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\varphi}\right) \sim InverseGamma\left(\alpha_\lambda + \frac{n}{2}, \beta_\lambda + \frac{1}{2}\frac{1}{\sigma^2}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)^T\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)\right) \tag{3.37}$$

3.4.5 Full Conditional for θ_j

Taking only the terms that depend on θ_j from the joint distribution in Eq. (3.17) gives the terms that are a function of the full conditional for θ_j, which gives

$$f\left(\theta_j \mid \mathbf{y}, \mu_c, \sigma^2, \boldsymbol{\theta}_{j-}, \lambda, \boldsymbol{\varphi}\right) \propto \left|\boldsymbol{\Sigma}\right|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\varphi}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi}\right) \tag{3.38}$$

The covariance matrix $\boldsymbol{\Sigma}$ in Eq. (3.38) is calculated using the correlation matrix $\boldsymbol{\Psi}$, $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Psi}$, and the correlation matrix depends on the correlation function parameter θ_j. There is no known distribution type in which the random variable, i.e., θ_j in this case, appears in this form. To draw samples from this nonstandard distribution, the Metropolis-Hastings algorithm [Metropolis et al. 1953; Hasting 1970] will be used. For numerical stability, the natural logarithm of the full conditional is used with the Metropolis-Hastings algorithm. The natural logarithm of Eq. (3.38) can be written as

$$Ln\left[f\left(\theta_j \mid \mathbf{y}, \mu_c, \sigma^2, \boldsymbol{\theta}_{j-}, \lambda, \boldsymbol{\varphi}\right)\right] \propto -\frac{n}{2}Ln\left[\sigma^2\right] - \frac{1}{2}Ln\left[\left|\boldsymbol{\Psi}\right|\right] + \left(-\frac{1}{2}\boldsymbol{\varphi}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi}\right) \tag{3.39}$$

3.4.6 Full Conditional for $\boldsymbol{\varphi}$

Taking only the terms that depend on $\boldsymbol{\varphi}$ from the joint distribution in Eq. (3.17) gives the terms that are a function of the full conditional for $\boldsymbol{\varphi}$, which gives

$$f\left(\boldsymbol{\varphi}\,|\,\mathbf{y},\mu_c,\sigma^2,\boldsymbol{\theta},\lambda\right) \propto \exp\left(-\frac{1}{2}\boldsymbol{\varphi}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi}\right) * \exp\left(-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)^T\left(\sigma^2\lambda\mathbf{I}\right)^{-1}\left(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\varphi}\right)\right) \tag{3.40}$$

After grouping and rearranging terms that are a function of $\boldsymbol{\varphi}$ and dropping constant terms, the right-hand side of Eq. (3.40) can be written as

$$\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\varphi}^T\underbrace{\left(\boldsymbol{\Sigma}^{-1}+\left(\sigma^2\lambda\mathbf{I}\right)^{-1}\right)}_{\mathbf{A}}\boldsymbol{\varphi}-\underbrace{\frac{-2}{\sigma^2\lambda}\left(\boldsymbol{\mu}^T-\mathbf{y}^T\right)}_{\mathbf{b}^T}\boldsymbol{\varphi}\right]\right) \tag{3.41}$$

The intermediate variables $\mathbf{A}$ and $\mathbf{b}$ are defined in Eq. (3.41), which then simplifies to

$$\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\varphi}^T\mathbf{A}\boldsymbol{\varphi}-\mathbf{b}\boldsymbol{\varphi}\right]\right) \tag{3.42}$$

After completing the square and dropping constant terms, Eq. (3.42) can be expressed as

$$\propto \exp\left(-\frac{1}{2}\left(\boldsymbol{\varphi}-\boldsymbol{\gamma}\right)^T\mathbf{A}\left(\boldsymbol{\varphi}-\boldsymbol{\gamma}\right)\right) \tag{3.43}$$

where the intermediate variable, $\boldsymbol{\gamma}$ is defined as

$$\boldsymbol{\gamma}=\frac{1}{2}\mathbf{A}^{-1}\mathbf{b} \tag{3.44}$$

From Eq. (3.41) the intermediate variables $\mathbf{A}$ and $\mathbf{b}$ are defined as

$$\begin{aligned}\mathbf{A} &= \boldsymbol{\Sigma}^{-1} + \left(\sigma^2\lambda\mathbf{I}\right)^{-1} \\ \mathbf{b} &= \frac{-2}{\sigma^2\lambda}\left(\boldsymbol{\mu} - \mathbf{y}\right)\end{aligned} \tag{3.45}$$

Equation (3.43) is identified as a kernel to a multivariate normal distribution with the mean vector $\boldsymbol{\gamma}$ and the covariance matrix $\mathbf{A}^{-1}$; note that Eq. (3.43) is written in terms of the precision and is hence the reason the covariance matrix is $\mathbf{A}^{-1}$ and not $\mathbf{A}$.

Finally the full conditional of $\boldsymbol{\varphi}$ as a function of the intermediate variables, prior parameters, and MBKG parameters can be expressed as

$$f\left(\boldsymbol{\varphi} \mid \mathbf{y}, \mu_c, \sigma^2, \boldsymbol{\theta}, \lambda\right) \sim MVN\left(\boldsymbol{\gamma}, \mathbf{A}^{-1}\right) \tag{3.46}$$

3.4.7 Full Conditional for $\boldsymbol{\beta}$

Taking only the terms that depend on $\boldsymbol{\beta}$ in Eq. (3.16) gives the terms that are a function of the full conditional for $\boldsymbol{\beta}$, which gives

$$\begin{aligned}f\left(\boldsymbol{\beta} \mid \mathbf{y}, \sigma^2, \lambda, \boldsymbol{\theta}, \boldsymbol{\varphi}\right) \propto \exp&\left(-\frac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta\right)^T \left(\boldsymbol{\Sigma}_\beta\right)^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta\right)\right) \\ *\exp&\left(-\frac{1}{2}\left(\mathbf{y} - \boldsymbol{\mu} - \boldsymbol{\varphi}\right)^T \left(\sigma^2\lambda\mathbf{I}\right)^{-1}\left(\mathbf{y} - \boldsymbol{\mu} - \boldsymbol{\varphi}\right)\right)\end{aligned} \tag{3.47}$$

After grouping and rearranging terms that are a function of $\boldsymbol{\beta}$ and dropping constant terms, the right-hand side of Eq. (3.47) can be written as

$$\begin{aligned}\propto \exp&\left(-\frac{1}{2}\left(\boldsymbol{\beta}^T \left(\boldsymbol{\Sigma}_\beta\right)^{-1}\boldsymbol{\beta} - 2\left(\boldsymbol{\mu}_\beta\right)^T \left(\boldsymbol{\Sigma}_\beta\right)^{-1}\boldsymbol{\beta}\right)\right) \\ *\exp&\left(-\frac{1}{2}\frac{1}{\sigma^2\lambda}\left(2\left(\boldsymbol{\varphi}^T - \mathbf{y}^T\right)\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{\mu}\right)\right)\end{aligned} \tag{3.48}$$

Substituting in the value $\mathbf{F\beta}$ for $\boldsymbol{\mu}$ and rearranging gives

$$\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\beta}^T \underbrace{\left(\frac{1}{\sigma^2\lambda}\mathbf{F}^T\mathbf{F} + \left(\boldsymbol{\Sigma}_\beta\right)^{-1} \right)}_{\mathbf{A}} \boldsymbol{\beta} - \underbrace{2\left(\left(\boldsymbol{\mu}_\beta\right)^T \left(\boldsymbol{\Sigma}_\beta\right)^{-1} - \frac{1}{\sigma^2\lambda}\left(\boldsymbol{\varphi}^T - \mathbf{y}^T\right)\mathbf{F} \right)}_{\mathbf{b}^T} \boldsymbol{\beta} \right] \right) \tag{3.49}$$

The intermediate variables $\mathbf{A}$ and $\mathbf{b}$ are defined in Eq. (3.49), which then simplifies to

$$\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\beta}^T\mathbf{A}\boldsymbol{\beta} - \mathbf{b}^T\boldsymbol{\beta} \right] \right) \tag{3.50}$$

After completing the square and dropping constant terms, Eq. (3.50) can be expressed as

$$\propto \exp\left(-\frac{1}{2}\left[\left(\boldsymbol{\beta} - \boldsymbol{\gamma}\right)^T \mathbf{A}\left(\boldsymbol{\beta} - \boldsymbol{\gamma}\right) \right] \right) \tag{3.51}$$

where the intermediate variable $\boldsymbol{\gamma}$ is defined as

$$\boldsymbol{\gamma} = \frac{1}{2}\mathbf{A}^{-1}\mathbf{b} \tag{3.52}$$

From Eq. (3.49) the intermediate variables $\mathbf{A}$ and $\mathbf{b}$ are defined as

$$\begin{aligned} \mathbf{A} &= \left(\frac{1}{\sigma^2\lambda}\mathbf{F}^T\mathbf{F} + \left(\boldsymbol{\Sigma}_\beta\right)^{-1} \right) \\ \mathbf{b} &= 2\left(\left(\boldsymbol{\mu}_\beta\right)^T \left(\boldsymbol{\Sigma}_\beta\right)^{-1} - \frac{1}{\sigma^2\lambda}\left(\boldsymbol{\varphi}^T - \mathbf{y}^T\right)\mathbf{F} \right) \end{aligned} \tag{3.53}$$

Equation (3.51) is identified as a kernel to a multivariate normal distribution with the mean vector $\boldsymbol{\gamma}$ and the covariance matrix $\mathbf{A}^{-1}$; note that Eq. (3.51) is written in terms of the precision and is hence the reason the covariance matrix is $\mathbf{A}^{-1}$ and not $\mathbf{A}$.

Finally, the full conditional of $\boldsymbol{\beta}$ as a function of the intermediate variables, prior parameters, and MBKG parameters can be expressed as

$$f\left(\boldsymbol{\beta} \mid \mathbf{y}, \sigma^2, \lambda, \boldsymbol{\theta}, \boldsymbol{\varphi}\right) \sim MVN\left(\boldsymbol{\gamma}, \mathbf{A}^{-1}\right) \tag{3.54}$$

The next chapter will present examples fitting a MBKG surrogate model to some example problems.

CHAPTER 4

APPLICATIONS OF MODIFIED BAYESIAN KRIGING (MBKG)

4.1 Introduction

This chapter presents fitting a modified Bayesian Kriging (MBKG) surrogate model to three different mathematical examples. The first example, in Section 4.2, presents fitting an MBKG surrogate model to a simple one-dimensional problem whose response contains noise. The section discusses the details of assessing the convergence of the Markov chain Monte Carlo (MCMC) analysis carried out to fit the MBKG surrogate model. Section 4.3 presents two examples, each fitting a MBKG surrogate model to a two-dimensional problem whose response contains noise.

4.2 One-Dimensional Quadratic Mathematical Example

This section demonstrates fitting a one-dimensional quadratic function that contains noise using the MBKG surrogate modeling method. The true quadratic function without noise is defined as

$$f(x) = 0.6x^2 - 3x + 4 \tag{4.1}$$

where x is the variable. For this mathematical example, the noise is generated using a normal distribution with zero mean and a standard deviation of 0.1. The noise, ε, is generated as realizations from the distribution, which is expressed as

$$\varepsilon \sim N\left(0, 0.1^2\right) \tag{4.2}$$

The response values with noise are then calculated as a summation of the true response value without noise given by Eq. (4.1) and a realization of ε. This is expressed as

$$y_i = f(x_i) + \varepsilon_i \tag{4.3}$$

where y_i is the i^{th} response value with noise, x_i is the i^{th} DoE sample point, and ε_i is the i^{th} realization of noise.

The red curve in Figure 4.1 shows a plot of Eq. (4.1). For this example, 25 DoE samples are used for fitting the surrogate model. To generate the uniformly spaced and randomly generated 25 DoE samples, the Latinized Centroidal Voronoi Tessellation (LCVT) [Burkardt et al. 2002; Saka et al. 2006] method was used. The responses for the 25 DoE samples were generated using Eq. (4.1) with realizations taken from Eq. (4.2). The 25 response values that contain noise are shown as the black asterisks in Figure 4.1.

The 25 DoE samples and responses with noise are first fitted using ordinary Kriging to demonstrate how Kriging is not well suited for use with response values that contain noise. Figure 4.2 shows a plot of Eq. (4.1) as the red curve, the 25 response values calculated using Eq. (4.3) are shown as the black asterisks, and the ordinary Kriging surrogate model is shown as the light purple curve. As seen in the figure, the ordinary Kriging surrogate model passes through the 25 DoE samples because Kriging is an interpolation method.

The ordinary Kriging model shown in Figure 4.2 will clearly not provide accurate predictions of the response values of the true underlying function, which does not contain noise. In order to carry out RBDO, a regression surrogate model is needed to capture the underlying true function without noise. In addition, a surrogate model that can generate either confidence intervals or credible sets that contain the uncertainty of the surrogate model is needed in order to perform confidence-based RBDO. This is why, in this study, an MBKG surrogate modeling method was developed; it will be used to predict the true function without noise and also provides credible sets that can be used for confidence-based RBDO.

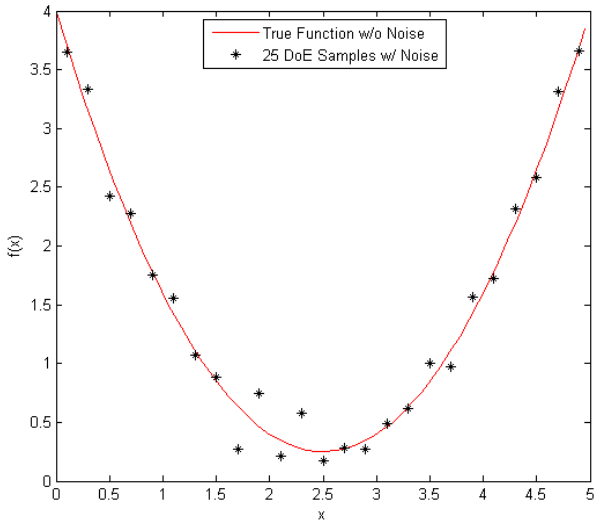

Figure 4.1 Eq. (4.1) without Noise and 25 DoE Samples

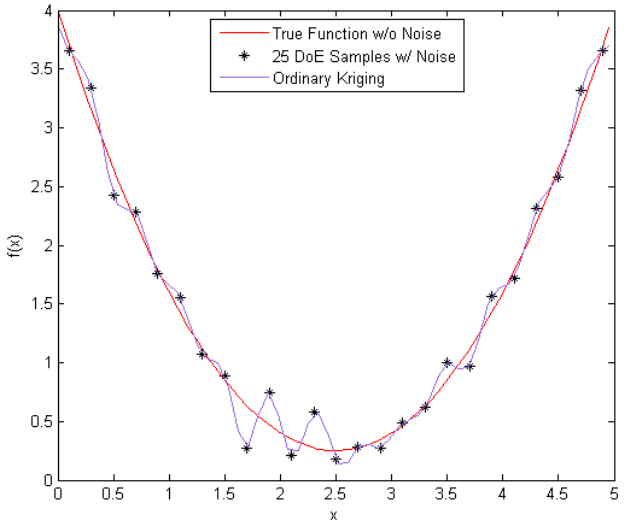

Figure 4.2 Ordinary Kriging Fit of Eq. (4.1) Using 25 DoE

Recall from Chapter 3 that, when carrying out a Bayesian analysis, the prior parameters for the prior distributions of the unknown parameters are needed. For the

example in this section, Table 4.1 lists the prior parameters used for fitting the MBKG surrogate model. As described in Section 3.3.2, the response values are normalized such that they have a zero mean. This implies that the MBKG μ_c parameter should have a mean value close to zero. For this reason the prior parameters for the μ_c prior distribution were chosen so that the prior has a mean value of zero. The standard deviation was chosen so that the 95% interval given by the prior distribution would be about ±1. This prior reflects that the subjective probability of the μ_c parameter being in this interval is 95%. From previous experience of fitting ordinary Kriging models to mathematical examples, the σ^2 parameter has typically been seen to take on values between about five and 20. Thus, a prior distribution that has a 95% interval with a lower bound of about five and an upper bound of about 20 seemed to be a reasonable prior. For the problems being considered, it is assumed that the variance of the noise in the response is much smaller than the spatial variance of the response data. With this assumption the λ parameter would be between zero and one. The prior for the λ parameter was chosen to reflect this assumption. However, the prior being used does not restrict the value of λ from being greater than one. This makes it possible for the value of λ to take on values larger than one should the data provide information where this is the case. Recall from Section 3.3.2 that a uniform prior with lower and upper bounds of 0.15 and 5, respectively, was chosen for the θ parameter. These values were chosen based on a study done about the relationship between the correlation value, the distance between DoE samples, and the θ parameter for the Gaussian correlation function. The reader is referred back to Section 3.3.2 for the details.

Table 4.1 Prior Parameter Values

Unknown Parameter	Prior Distribution	Parameter 1	Parameter 2	95% Interval	
μ_c	Normal	$\mu_p = 0$	$\sigma_p^2 = 0.5$	–1.39	1.39
σ^2	Inverse-Gamma	$\alpha_\sigma = 8.25$	$\beta_\sigma = 72.5$	4.91	20.04
λ	Inverse-Gamma	$\alpha_\lambda = 1.9548$	$\beta_\lambda = 0.2747$	0.05	1.20
θ	Uniform	$a_\theta = 0.15$	$b_\theta = 5$	0.27	4.88

To fit the MBKG surrogate model, Markov chain Monte Carlo (MCMC) is used. As described in Section 1.2.3, when using MCMC it is desirable to run at least three parallel Markov chains so that they can be used with the Brooks, Gelman, and Rubin (BGR) diagnostic [Gelman and Rubin 1992; Brooks and Gelman 1998] to help assess if convergence in drawing samples from the posterior distribution has been achieved. Also recall that, when doing this, the three parallel Markov chains should be started at over-dispersed initial values. For this example, three parallel chains are used, and the initial values for the Markov chains are taken from the prior distributions such that the three starting points come from the left tail, right tail, and the middle of the prior distribution. This is done in an effort to help make sure that the samples drawn cover the entire posterior distribution so that no modes of the posterior distribution are missed. The initial values for the Markov chains used for this example are given in Table 4.2.

Table 4.2 Initial Values for Markov Chains

Unknown Parameter	Chain 1 Initial Values	Chain 2 Initial Values	Chain 3 Initial Values
μ_c	0.00	–1.00	1.00
σ^2	4.00	9.00	17.0
λ	0.06	0.50	0.80
θ	0.20	2.50	5.00

Recall from Chapter 3 that the MBKG formulation has a $\boldsymbol{\varphi}$ vector that is unknown and needs to be estimated when fitting the MBKG surrogate model. The length of the $\boldsymbol{\varphi}$ vector, i.e., the number of unknown parameters in the vector, is equal to the number of DoE samples used to fit the model. For this example, 25 DoE samples are used to fit the model; thus the $\boldsymbol{\varphi}$ vector has 25 unknown parameters that need to be fitted in addition to the other four unknown parameters μ_c, σ^2, λ, and θ. The initial values used for the unknown $\boldsymbol{\varphi}$ vector are calculated using the initial values of μ_c and the 25 response values in the $\mathbf{y}$ vector. The initial values $\boldsymbol{\varphi}^{(i)}$ for the i^{th} chain are calculated as

$$\boldsymbol{\varphi}^{(i)} = \mathbf{y}^{DoE} - \mu_c^{(i)}\mathbf{1} \tag{4.4}$$

where $\boldsymbol{\varphi}^{(i)}$ are the initial values for the i^{th} chain, $\mathbf{y}^{DoE}$ are the 25 response values for the corresponding 25 DoE samples, $\mu_c^{(i)}$ is the initial value of μ_c for the i^{th} chain, and $\mathbf{1}$ is vector filled with the value of one and has the same length as the $\mathbf{y}^{DoE}$ vector.

Using the prior parameters in Table 4.1, the initial values in Table 4.2, and the initial values calculated using Eq. (4.4), MCMC is carried out to fit the MBKG surrogate model. For this example, 100,000 iterations were run using MCMC for each of the three parallel chains, giving a total of 300,000 iterations. In Bayesian analysis, when using MCMC, one iteration means that a sample was drawn for each of the unknown parameters using the full conditional distributions as described in Sections 1.2 and 3.4.

Another method to help assess if convergence in the MCMC chains has been achieved is to plot the history of the samples drawn for each of the MCMC iterations for the unknown parameters. A history plot for each of the unknown parameters can be created, and the three parallel chains for the parameter are plotted on the same plot. If the MCMC iterations are not drawing samples from the same posterior distribution for the three chains, then the history plots may show this clearly. Figure 4.3, Figure 4.4, Figure 4.5, and Figure 4.6 show the history plots of the first 5000, 200, 2000, and 2000 iterations for the four unknown parameters μ_c, σ^2, λ, and θ, respectively. Note that the history plots start at the first iteration of the MCMC chains, i.e., the first sample drawn from the full conditionals. The initial values given in Table 4.2 are the 0^{th} iteration of the MCMC chains, and it is common practice in Bayesian analysis that the 0^{th} iteration is not shown in the history plots. In all the history plots, the three chains are plotted in three different colors: chain one is plotted in red, chain two is plotted in blue, and chain three is plotted in green. Figure 4.3 shows that, from the first iteration to about the 1500^{th} iteration, chain three (green) looks to be drawing samples from a distribution that is wider than the distributions that the samples for chains one (red) and two (blue) are being drawn from. This is an indication that the three chains are not yet drawing samples from the same posterior distribution. However, after about iteration 1500 in Figure 4.3, the three chains look like white noise and look as though they may be drawing samples from the same posterior distribution.

Figure 4.4 shows that chain one (red) starts off at relatively larger values than chains two (blue) and three (green). However, by about 40 iterations, the three chains are starting to come together, though they are not well mixed yet. In Figure 4.5, chain one (red) and two (blue) come together quickly, and by about 200 iterations they look as if they may be drawing samples from the same posterior distribution. Chain three (green), however, does not merge with chains one and two until after about 1400 iterations. After about 1400 iterations, it appears that the three chains are drawing samples from the same

posterior distribution. Similarly, in Figure 4.6, chains one (red) and two (blue) come together quickly after about 200 iterations. Chain three (green) does not merge with chains one and two until after about 1500 iterations. The history plots can be used to help determine when the chains may be starting to converge. However, it is important to run enough MCMC iterations to make sure that the chains stay together. Also, a large number of MCMC iterations are needed after the three chains have converged so that the samples drawn can be used to calculate statistics about the unknown parameters.

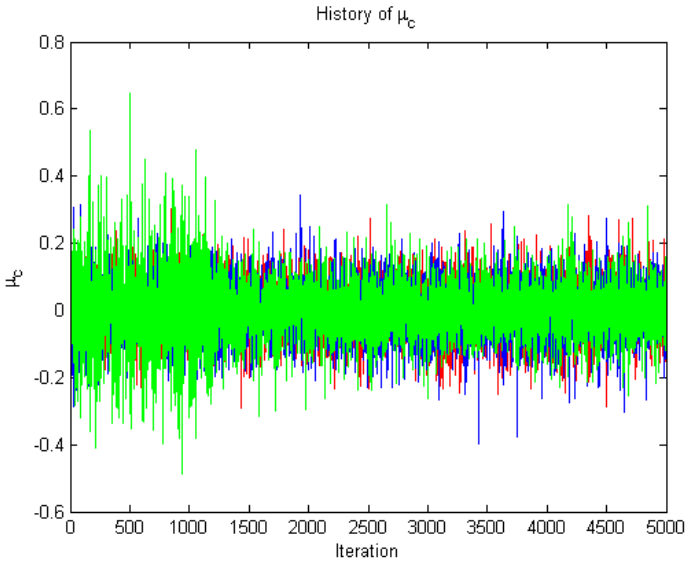

Figure 4.3 History Plot of the First 5000 Iterations for μ_c

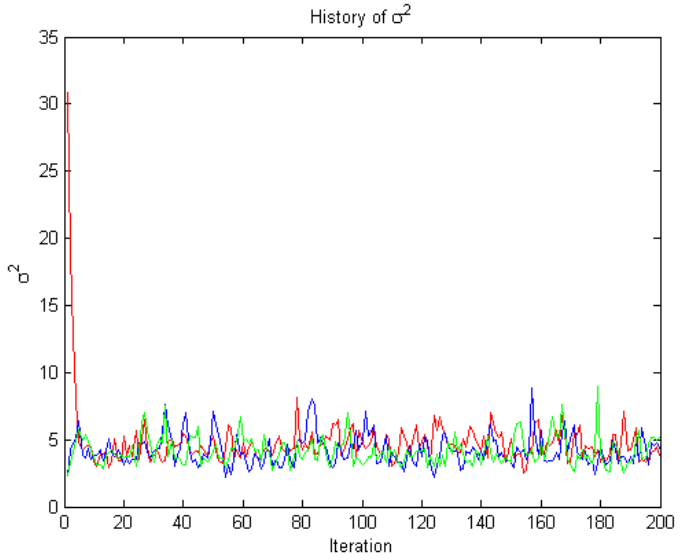

Figure 4.4 History Plot of the First 200 Iterations for σ^2

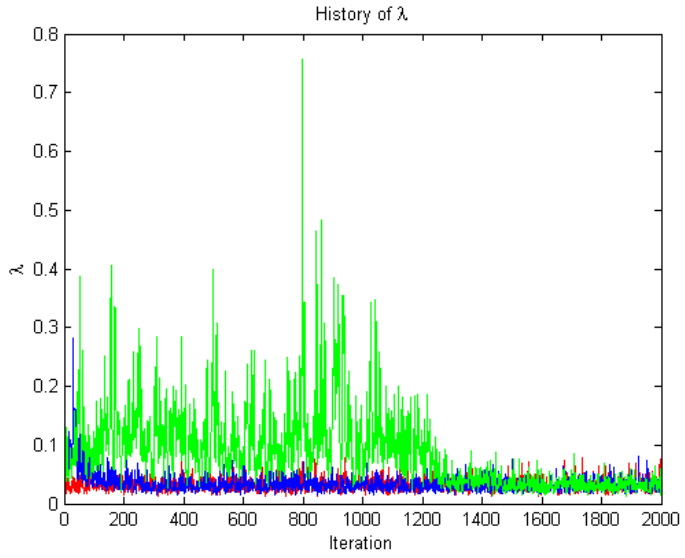

Figure 4.5 History Plot of the First 2000 Iterations for λ

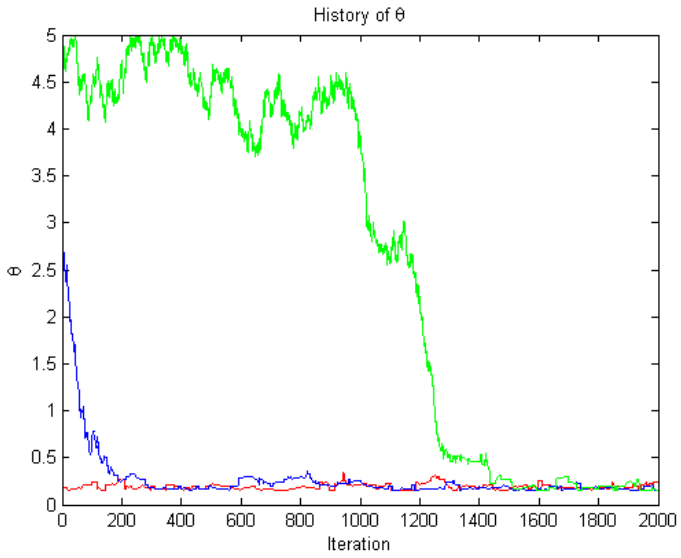

Figure 4.6 History Plot of the First 2000 Iterations for θ

Figures 4.7 – 4.10 are the history plots for all four unknown parameters and show the history for all 100,000 iterations.

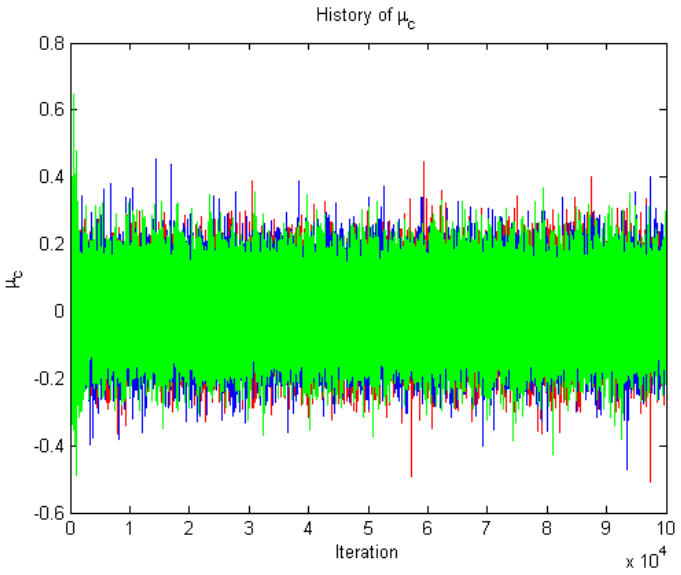

Figure 4.7 History Plot of All 100,000 Iterations for μ_c

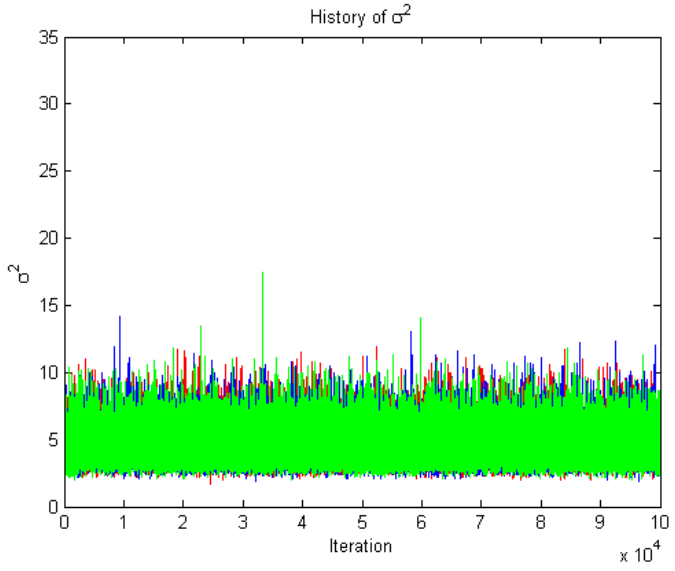

Figure 4.8 History Plot of All 100,000 Iterations for σ^2

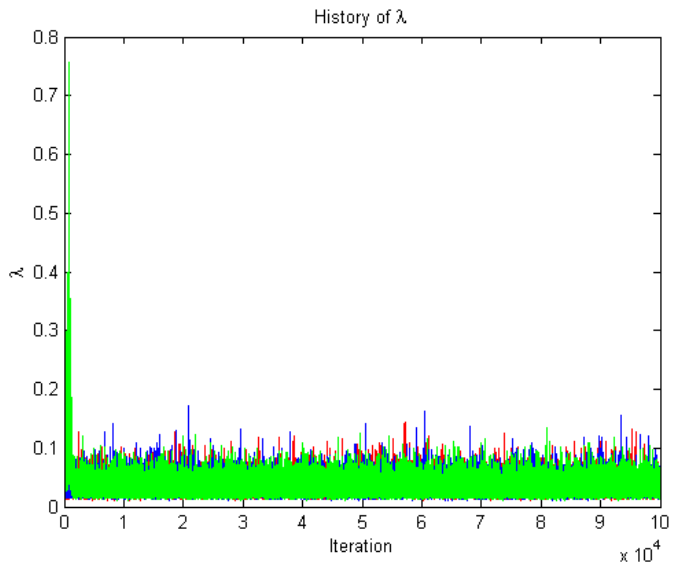

Figure 4.9 History Plot of All 100,000 Iterations for λ

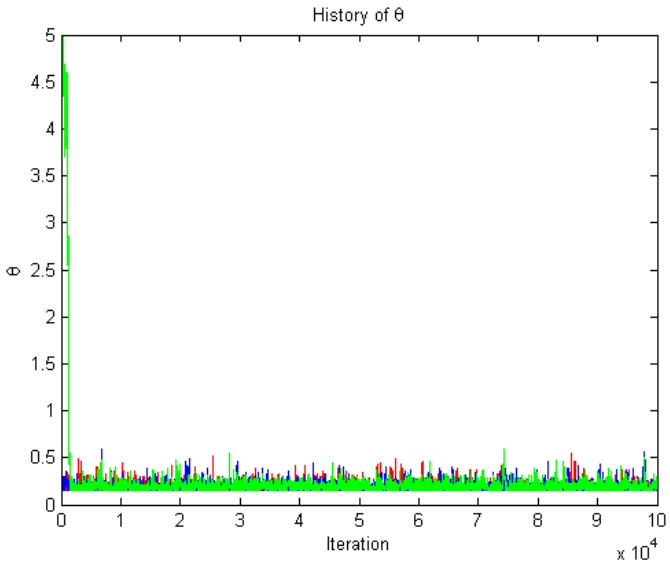

Figure 4.10 History Plot of All 100,000 Iterations for θ

As seen in all four figures, the three chains all appear to have converged to drawing samples from the same posterior distribution after about 1500 iterations. The history plots are helpful for trying to determine if the three chains look like they may have converged to drawing samples from the same posterior distribution. However, the history plots only provide a visual and do not provide an analytical way to assess if the three chains have converged to the same posterior distribution.

As discussed in Section 1.2, the use of the BGR diagnostic is a common and well-accepted method to help assess if the MCMC chains have converged. The BGR diagnostic provides both a visual and numerical way to help assess convergence of the MCMC chains. Figures 4.11 – 4.14 show the BGR plots for the μ_c, σ^2, λ, and θ parameters. To assess convergence using a BGR plot, the red line in the BGR plot should level off and should have a value of one once the three chains are drawing samples from the same posterior distribution. A rule of thumb for the red line is that if the value is less than 1.05 it is taken to be convergence to one. Recall from Section 1.2.3 that the BGR diagnostic is calculating the widths of credible sets two different ways. The width of each

MCMC chain is calculated. Using the width for each chain, an average width is calculated. The widths are also calculated after the parallel Markov chains are pooled together to form one sample set. As the Markov chains converge to drawing samples from the same posterior distribution, the value of the widths calculated by the two different methods should approach the same number. The ratio of these two width values is calculated. This ratio should approach one as the parallel Markov chains converge to drawing samples from the same posterior distribution. The red line in the BGR plot is a plot of this ratio as the number of MCMC iterations increases. The blue and green lines in the BGR plot should come together and level off to a constant value. The value at which the blue and green lines level off does not matter; it is only important that they come together and level off to a constant value as the number of iterations increases. The widths calculated using the two different methods are normalized and plotted in the BGR plot as the blue and green lines. Thus, the blue and green lines coming together and leveling off to a constant value in the plot show that the widths of the credible sets are stabilizing and that convergence to the posterior distribution may have been reached.

The BGR plots for all four parameters show that after about 20,000 iterations the red line has reached a value of one, and the blue and green lines have come together and leveled off to a constant value. Thus, for this example based on the history plots and BGR plots, it has been determined that the MCMC chains have converged to the posterior distribution by 20,000 iterations. That means the samples drawn after 20,000 iterations can be used to calculate statistics and predict response values using the surrogate model. The iterations before convergence in distribution has been achieved are discarded, and only the iterations after convergence are used for calculating statistics and plotting posterior distributions. The samples that are discarded are referred to as burn-in when performing a Bayesian analysis using MCMC.

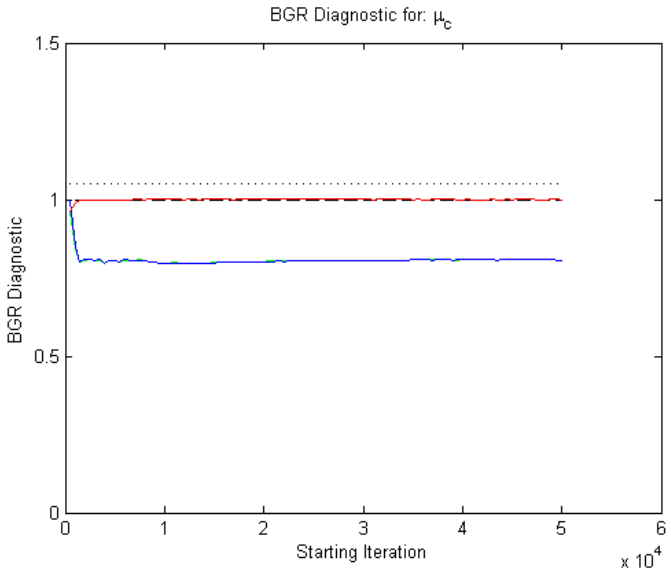

Figure 4.11 BGR Plot for μ_c

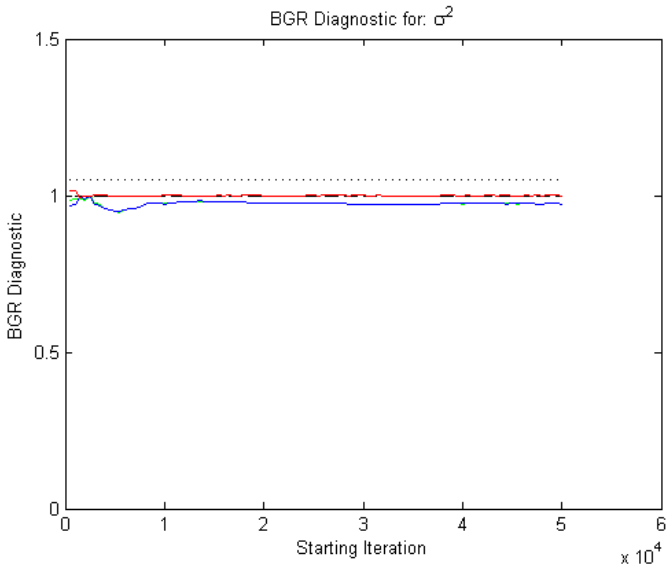

Figure 4.12 BGR Plot for σ^2

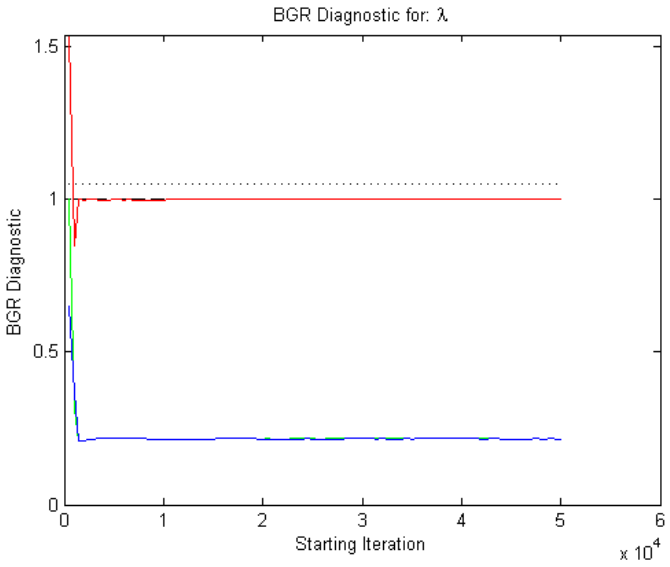

Figure 4.13 BGR Plot for λ

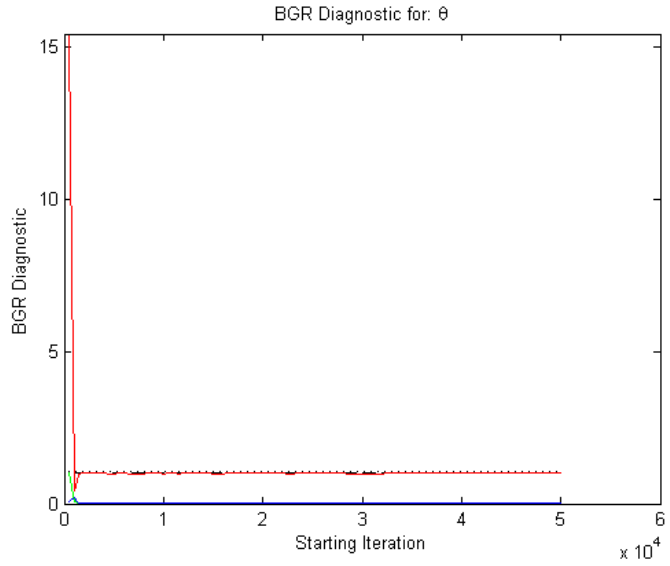

Figure 4.14 BGR Plot for θ

Recall that a total of 100,000 iterations were run and that the first 20,000 iterations were discarded as burn-in. This leaves 80,000 iterations for each of the three

parallel chains that can be used to calculate statistics and plot distributions. The remaining iterations from the three chains are pooled together to give a total of 240,000 samples from the posterior distribution to use for inference. Table 4.3 gives the posterior mean, standard deviation, and the 95% credible set using the 240,000 samples. If the λ parameter is considered, the posterior mean of λ is 0.0324 and the standard deviation of λ is 0.0107. The 95% credible set for λ is $[0.0176, 0.0588]$, which means that, for the priors and the 25 DoE samples used to fit the MBKG model, there is 95% posterior probability that the true λ is between 0.0176 and 0.0588. From Table 4.1, the 95% interval for the prior distribution of λ was $[0.05, 1.20]$. Thus, the posterior distribution is much narrower than the prior distribution as seen by the 95% credible set of the posterior distribution. This means that the 25 DoE samples provided a large amount of information about what the true value of λ should be and shrank the uncertainty of the value of λ considerably as is reflected in the posterior 95% credible set. Similarly, for the three parameters μ_c, σ^2, and θ, Table 4.3 shows that the 95% credible set is rather narrow compared to the 95% intervals for the prior distributions given in Table 4.1. Overall, it can be concluded that the 25 DoE samples provided a large amount of information about what the true value of the unknown parameters are and shrank the uncertainty in all the unknown parameters.

Table 4.3 Posterior Statistics for Unknown Parameters Using 25 DoE Samples

Unknown Parameter	Mean	Standard Deviation	95% Credible Set	
μ_c	–3.019e –5	0.0761	–0.1510	0.1507
σ^2	4.5542	1.0685	2.9146	7.0660
λ	0.0324	0.0107	0.0176	0.0588
θ	0.1968	0.0432	0.1516	0.3118

The mean predicted response value given by the MBKG surrogate model, the true function, and the 25 DoE samples are shown in Figure 4.15. The true function without noise is shown as the red curve, the 25 DoE samples are shown as the black asterisks, and the mean predicted response value is shown as the blue curve. The figure shows that the MBKG surrogate model gives a much better fit to the true function than the ordinary Kriging surrogate model did in Figure 4.2.

Figure 4.16 shows the 95% credible set without noise, the 95% credible set with noise in addition to the mean predicted response, the true function, and the 25 DoE samples. The 95% credible set without noise is shown by the green curves in Figure 4.16. The meaning of the 95% credible set without noise gives the band in which there is 95% probability that the true function lies. The 95% credible set with noise is shown by the pink dashed curves in Figure 4.16. The meaning of the 95% credible set with noise gives the band in which there is 95% probability that the noisy response values given by Eq. (4.3) lie.

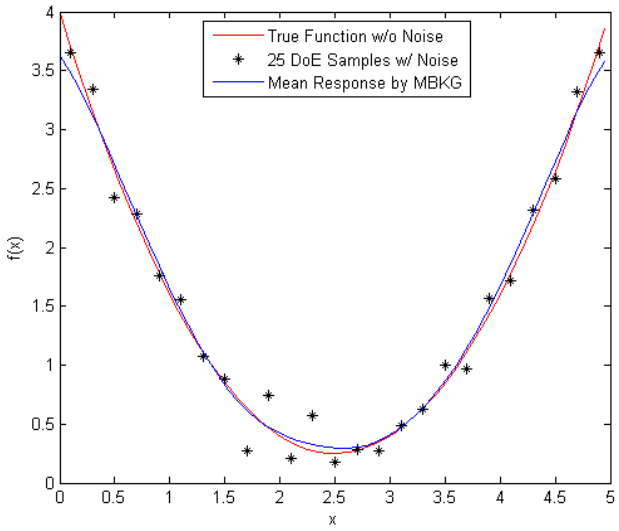

Figure 4.15 MBKG Fit of Eq. (4.1) Using 25 DoE

As shown in Figure 4.16, the 95% credible set without noise (green curves) does in fact contain the true function (red curve). Also, the 95% credible set with noise (pink dashed curves) does in fact contain both the true function and the 25 DoE sample points. However, it is noted that both the 95% credible sets with and without noise are rather wide. This is a direct measure and reflection of the uncertainty that exists in the MBKG surrogate model. In order to decrease the uncertainty in the MBKG surrogate model and therefore narrow the 95% credible sets, more DoE samples would need to be added.

To demonstrate how the 95% credible sets narrow with more DoE samples, 25 additional DoE samples were uniformly added to the initial 25 DoE samples, and the MBKG surrogate model was refitted using 50 DoE samples and the same priors as before. Figure 4.17 shows the fit of the new MBKG surrogate model: the red curve is the true function without noise, the black asterisks are the initial 25 DoE samples previously used, and the blue squares are the additional 25 DoE samples. The mean predicted response given by the MBKG surrogate model is the dark green curve. The 95% credible set without noise is shown as the light blue curves, and the 95% credible set with noise is shown as the black dashed curves.

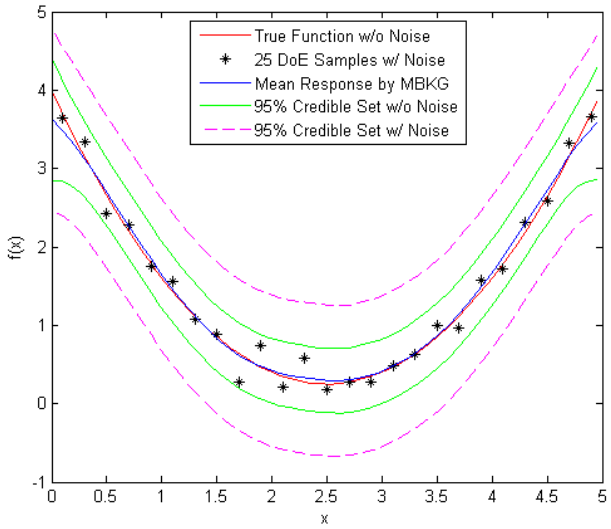

Figure 4.16 MBKG Fit of Eq. (4.1) and 95% Credible Sets Using 25 DoE

Figure 4.18 shows a plot of the 95% credible sets without noise when both 25 DoE samples and 50 DoE samples are used. The figure shows how the 95% credible set using 50 DoE samples is in fact narrower than the 95% credible set using 25 DoE samples. It also shows how the credible set is narrower in some regions and wider in other regions. This indicates where the uncertainty in the surrogate model is smaller and larger in those regions, respectively. Figure 4.19 shows a plot of the 95% credible sets with noise when both 25 DoE samples and 50 DoE samples are used. Once again, it is seen how for 50 DoE samples the credible set is narrower than when 25 DoE samples are used. Figures 4.18 and 4.19 both show how the information in the additional 25 DoE samples has decreased the amount of uncertainty we have in the surrogate model.

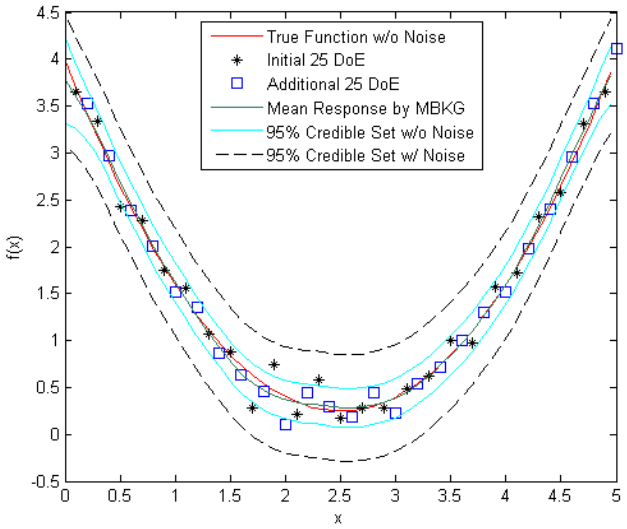

Figure 4.17 MBKG Fit of Eq. (4.1) and 95% Credible Sets Using 50 DoE

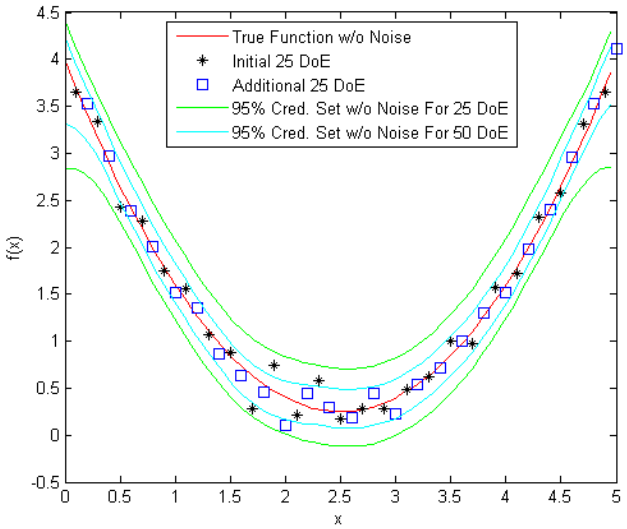

Figure 4.18 95% Credible Sets without Noise

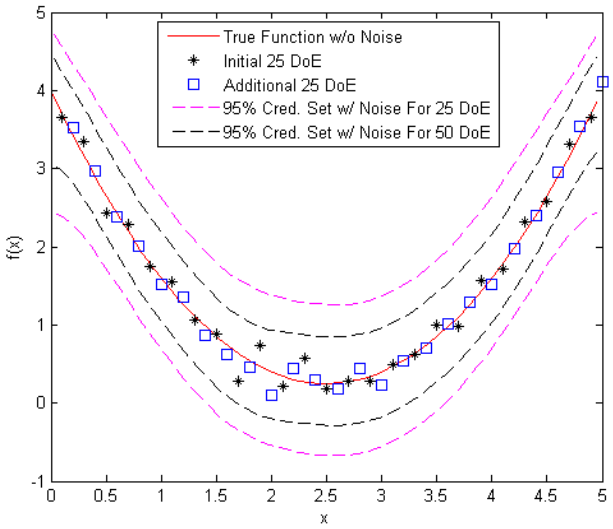

Figure 4.19 95% Credible Sets with Noise

Table 4.4 gives the posterior mean, the standard deviation, and the 95% credible set for the four unknown parameters of the MBKG surrogate model that was fitted using 50 DoE samples. Note that 240,000 MCMC samples were used to calculate the statistics. Comparing the standard deviations in Table 4.3 to those in Table 4.4, it can be seen that the posterior standard deviations are smaller when 50 DoE samples were used. Also, comparing the 95% credible sets in the two tables shows that the credible sets when 50 DoE samples were used are smaller, e.g., the lower and upper bounds of the 95% credible set for μ_c were both reduced by a little more than half. The smaller posterior standard deviations and narrower 95% credible sets both show how the additional 25 DoE samples provided more information and reduced the overall uncertainty in the MBKG surrogate model. Adding more DoE samples would continue to add more information to the MBKG surrogate model and thus continue to reduce the posterior standard deviation and shrink the 95% credible sets further.

Table 4.4 Posterior Statistics for Unknown Parameters Using 50 DoE Samples

Unknown Parameter	Mean	Standard Deviation	95% Credible Set	
μ_c	4.6847e –5	0.0326	–0.0642	0.0643
σ^2	3.1863	0.5890	2.2301	4.5341
λ	0.0169	0.0038	0.0110	0.0258
θ	0.1880	0.0328	0.1513	0.2779

4.3 Two-Dimensional Mathematical Examples

4.3.1 First 2-D Mathematical Example

This section demonstrates fitting two different mathematical functions [Lee et al. 2011] that are two-dimensional and contain noise using the MBKG surrogate modeling method. The first function without noise, referred to as $G_1(\mathbf{x})$, is defined as

$$G_1(\mathbf{x}) = 1 - \frac{X_1^2 X_2}{20} \tag{4.5}$$

where X_1 and X_2 are the two variables. For this example the two variables are correlated to each other by Clayton copula with a correlation coefficient of 0.5. Because of this correlation, the domain of interest for this example forms the raindrop shape seen in later figures. For this mathematical example, the noise is generated using a normal distribution with zero mean and a standard deviation of 0.1. The noise, ε, is generated as realizations from the distribution, which is expressed as

$$\varepsilon \sim N(0, 0.1^2) \tag{4.6}$$

The response values with noise are then calculated as a summation of the true response value without noise given by Eq. (4.5) and a realization of ε. This is expressed as

$$y_i = G_1(\mathbf{x}_i) + \varepsilon_i \tag{4.7}$$

where y_i is the i^{th} response value with noise, $\mathbf{x}_i$ is the i^{th} DoE sample point, and ε_i is the i^{th} realization of noise.

The red surface in Figure 4.20 shows a 3-D plot of Eq. (4.5) for the domain of interest for X_1 and X_2 for this example. Also shown in the figure as the light blue dots are the 25 DoE samples used for fitting the MBKG surrogate model. This example is going to focus on fitting an MBKG surrogate model to Eq. (4.7), keeping in mind the overall objective of carrying out confidence-based RBDO. Recall from Section 2.2 that the RBDO problem is formulated such that the failure is defined as response values greater than zero. The threshold value of zero of the function is referred to as the limit state. Figure 4.21 shows a contour plot of the true limit state, the red curve, for Eq. (4.5) in the domain of interest. The 25 DoE samples are shown as the red asterisks in the figure.

The prior parameter values used for fitting the MBKG surrogate model to Eq. (4.5) are shown in Table 4.5. Note that, because this is a two-dimensional problem, there are two θ parameters to be estimated. The prior parameter values for this example are actually the same as those used for the previous example. This is because these priors are relatively non-informative and thus are applicable for this example as well.

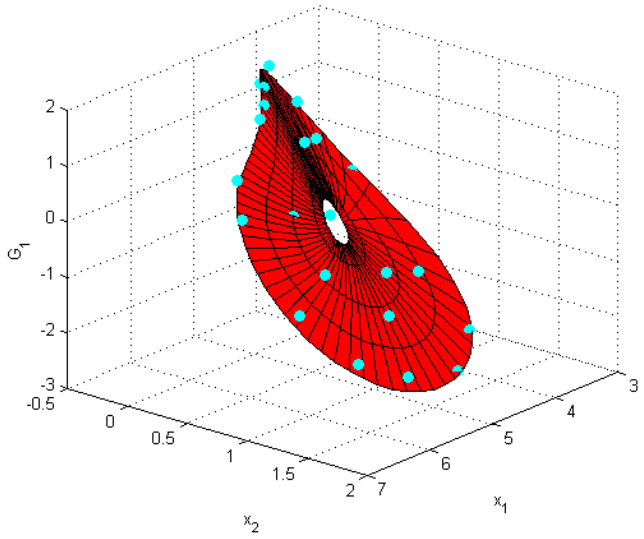

Figure 4.20 Surface Plot of Eq. (4.5) and 25 DoE Samples

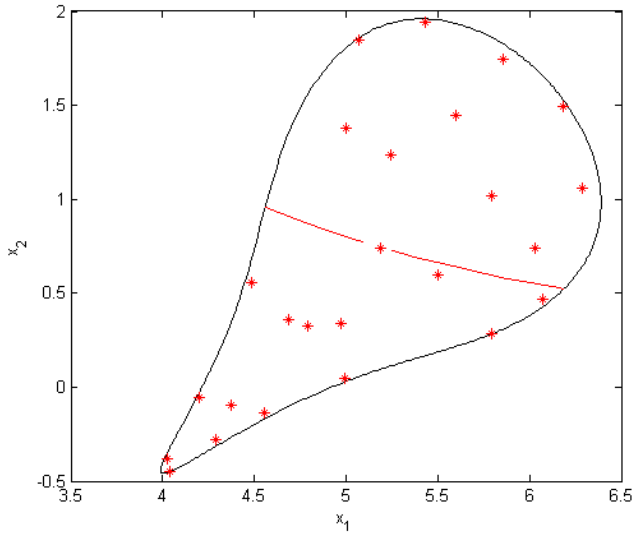

Figure 4.21 Contour Plot of True Limit State Function for Eq. (4.5) and 25 DoE Samples

Table 4.5 Prior Parameter Values for Fitting Eq. (4.5)

Unknown Parameter	Prior Distribution	Parameter 1	Parameter 2	95% Interval	
μ_c	Normal	$\mu_p = 0$	$\sigma_p^2 = 0.5$	–1.39	1.39
σ^2	Inverse-Gamma	$\alpha_\sigma = 8.25$	$\beta_\sigma = 72.5$	4.91	20.04
λ	Inverse-Gamma	$\alpha_\lambda = 1.9548$	$\beta_\lambda = 0.2747$	0.05	1.20
θ_1	Uniform	$a_{\theta_1} = 0.15$	$b_{\theta_1} = 5$	0.27	4.88
θ_2	Uniform	$a_{\theta_2} = 0.15$	$b_{\theta_2} = 5$	0.27	4.88

The initial values for the Markov chains used for this example are given in Table 4.6. Note that because the prior distributions are the same as those used in the example in the previous section, the initial values for the five unknown parameters in this example are the same as those in the previous section. This is because the initial values are based on the prior distributions.

Table 4.6 Initial Values for Markov Chains for Fitting Eq. (4.5)

Unknown Parameter	Chain 1 Initial Values	Chain 2 Initial Values	Chain 3 Initial Values
μ_c	0.00	–1.00	1.00
σ^2	4.00	9.00	17.0
λ	0.06	0.50	0.80
θ_1	0.20	2.50	5.00
θ_2	0.20	2.50	5.00

Using the prior parameters in Table 4.5 and the initial values in Table 4.6, MCMC was carried out to fit a MBKG surrogate model to the noisy responses given by Eq. (4.7). The contour plot of the limit state function, calculated using the mean of the predicted

response values, is shown in Figure 4.22. The true limit state function is shown as the red curve, and the predicted limit state function is shown as the blue curve in the figure. The 25 DoE samples are shown as the red asterisks.

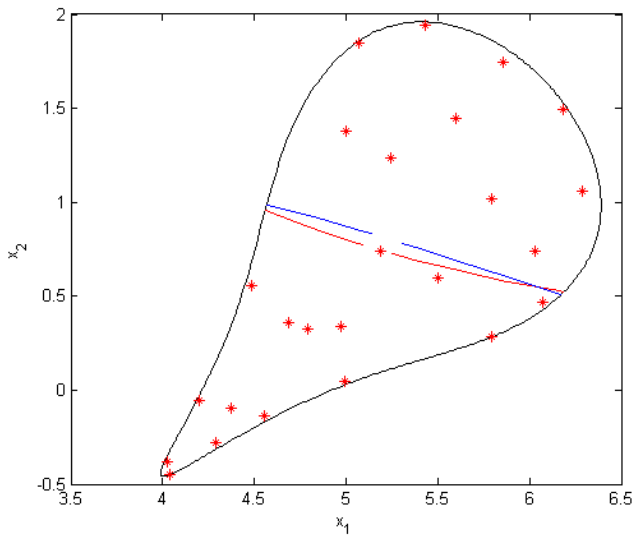

Figure 4.22 MBKG Predicted Contour Plot of Limit State Function for Eq. (4.5)

As seen in Figure 4.22, the predicted limit state by MBKG does not match extremely well to the true limit state function. This indicates that more DoE samples are required in order to improve the MBKG surrogate model and thus will improve the predicted limit state function given by the MBKG surrogate model. It is also easily seen in the figure that, out of the 25 DoE sample points, only a few of them are near the limit state function itself. This is another reason that the predicted limit state function is not very accurate. However, one thing that can be concluded from this is that adding additional samples near the limit state may be an efficient and effective way to improve the accuracy of the predicted limit state function given by the MBKG surrogate model. This will be discussed in more detail in the next chapter.

4.3.2 Second 2-D Mathematical Example

The second two-dimensional function considered in this section is referred to as $G_2(\mathbf{x})$ and is defined as

$$\begin{aligned} G_2(X) &= -1 + (0.9063X_1 + 0.4226X_2 - 6)^2 \\ &+ (0.9063X_1 + 0.4226X_2 - 6)^3 \\ &- 0.6(0.9063X_1 + 0.4226X_2 - 6)^4 \\ &- (-0.4226X_1 + 0.9063X_2) \end{aligned} \tag{4.8}$$

where X_1 and X_2 are the two variables. For this example the two variables are correlated to each other by Clayton copula with a correlation coefficient of 0.5. Because of this correlation, the domain of interest for this example forms the raindrop shape seen in later figures. For this mathematical example, the noise is once again generated using a normal distribution with zero mean and a standard deviation of 0.1. Thus, the noise, ε, is taken as realizations from the distribution in Eq. (4.6).

The response values with noise are then calculated as a summation of the true response value without noise give by Eq. (4.8) and a realization of ε. This is expressed as

$$y_i = G_2(\mathbf{x}_i) + \varepsilon_i \tag{4.9}$$

where y_i is the i^{th} response value with noise, $\mathbf{x}_i$ is the i^{th} DoE sample point, and ε_i is the i^{th} realization of noise

The red surface in Figure 4.23 shows a 3-D plot of Eq. (4.8) for the domain of interest for X_1 and X_2 for this example; note that this is the same domain as for $G_1(\mathbf{x})$ in Eq. (4.5). Also shown in the figure as the light blue dots are the 25 DoE samples used for fitting the MBKG surrogate model. Note these are the same 25 DoE samples that were used for fitting Eq.(4.7). Figure 4.24 shows the contour plot of the true limit state

function, the red curve, for Eq. (4.8) in the domain of interest, and the red asterisks are the 25 DoE.

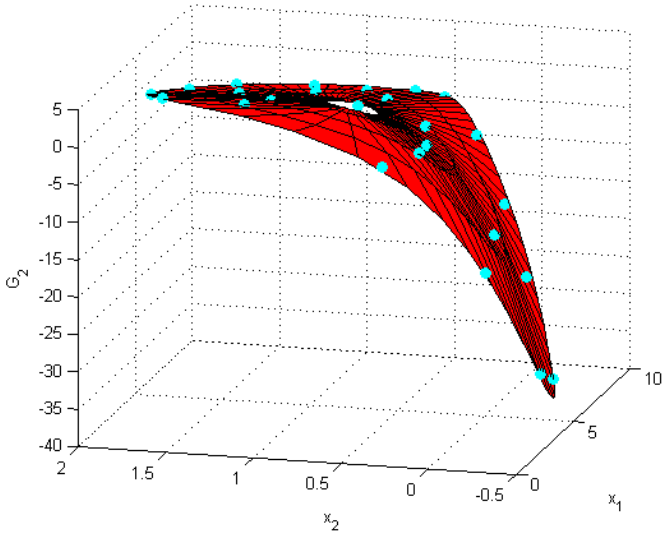

Figure 4.23 Surface Plot of Eq. (4.8) and 25 DoE Samples

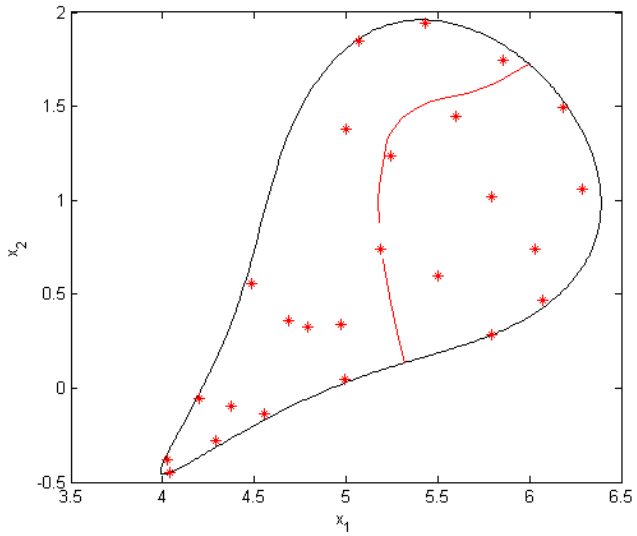

Figure 4.24 Contour Plot of True Limit State Function for Eq. (4.8) and 25 DoE Samples

The same prior distributions given in Table 4.5 are also used for fitting the MBKG surrogate model to Eq. (4.9). Because the same priors are being used, the same initial values for the Markov chains given in Table 4.6 are also used. Using these priors and initial values, MCMC was carried out to fit an MBKG surrogate model to Eq. (4.9). The contour plot of the limit state function, calculated using the mean of the predicted response values, is shown in Figure 4.25.

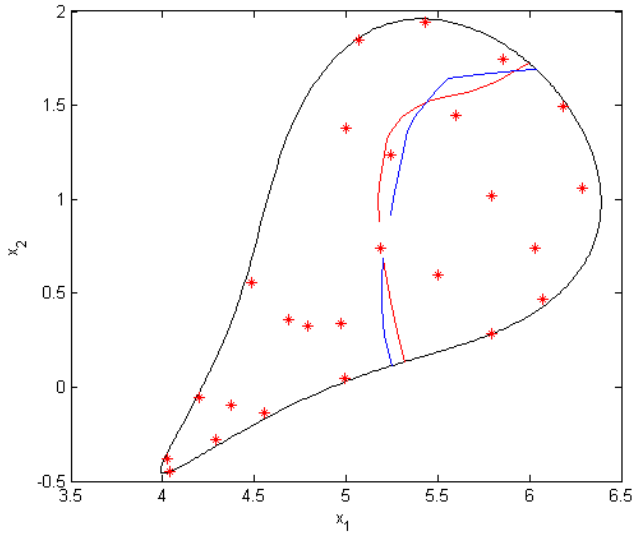

Figure 4.25 MBKG Predicted Contour Plot of Limit State Function for Eq. (4.8)

As seen in Figure 4.25, the predicted limit state by MBKG captures the general trend of the true limit state function; however, it does not accurately predict the true limit state function. This is an indication that more DoE samples need to be added, and the MBKG surrogate model refitted to give a more accurate prediction of the limit of the state function. Also, from the figure, it can be seen that only four of the 25 DoE samples are near the true limit state function. Thus, the same conclusion as for the $G_1(\mathbf{x})$ function has been reached: adding additional samples near the limit state may be an efficient and

effective way to improve the accuracy of the predicted limit state given by the MBKG surrogate model. The next chapter will discuss adding more DoE samples.

4.3.3 A Note about the Computational Time

For both examples presented in this section, three parallel Markov chains each containing 100,000 iterations were generated, giving a total of 300,000 samples being drawn. The concern then becomes what is the computational time for generating 300,000 iterations. For the two examples in this section, the computational time to fit the MBKG surrogate model for each example was about 10 minutes, i.e., it took about 10 minutes to draw the 300,000 samples. The hardware and software details used for fitting this model are as follows. A computer with Windows 7 Enterprise with Service Pack 1, 64-bit operating system was used. The processor in the computer used is an Intel® Core™ i7-2600 CPU @ 3.40GHz. The computer has 16.0 GB of RAM. Hyperthreading is enabled on the machine, giving a total of eight cores that can be used. All the code for the MBKG surrogate modeling method has been developed in MATLAB. For this example, MATLAB Version 8.1.0.604 (R2013a) was used.

Because the three Markov chains are independent, they were run in parallel using three cores. Thus, all three parallel Markov chains for both examples were run in about 10 minutes using three cores. The computational time for these examples seems to be reasonable and is expected to be similar for other examples.

CHAPTER 5

SEQUENTIAL SAMPLING VIA CREDIBLE SETS

5.1 Introduction

The two mathematical examples presented in Section 4.3 will be used and built upon in the next two sections. A sequential sampling method is developed for adding more design of experiment (DoE) samples to further improve the modified Bayesian Kriging (MBKG) surrogate model by utilizing the information in the posterior credible sets of the MBKG surrogate model.

Recall that the overall objective of this study is to carry out confidence-based RBDO using the MBKG surrogate model for problems whose responses contain noise. Also recall from Section 2.2 that the RBDO problem is formulated such that failure is defined as response values greater than zero. Thus, in order to use an MBKG surrogate model, the accuracy of the limit state function is of high importance because it will dictate the accuracy of the reliability analysis. Keep in mind that the MBKG surrogate model is not a deterministic surrogate model, but rather a surrogate model that produces posterior distributions for the MBKG parameters μ_c, σ^2, λ, $\boldsymbol{\theta}$, and $\boldsymbol{\varphi}$. Therefore, a predicted response value for a given point does not have one deterministic value but rather has a probability distribution that gives the probability of the predicted response value being in any interval. The Markov chain Monte Carlo (MCMC) samples drawn from the predictive distribution of the response variable can be used to estimate any desired characteristic of the distribution, e.g., the mean, standard deviation, and credible sets. The larger the standard deviation and the wider the credible set, the more uncertainty there is in the predicted value, i.e., more uncertainty in what the true value is. The probability distributions characterize this uncertainty and provide information that can be used to further improve the MBKG surrogate model. This section presents a sequential

sampling method for adding more DoE samples to the MBKG surrogate model using the posterior credible sets to improve the prediction of the limit state.

5.2 Sequential Sampling for the First 2-D Mathematical Example

Section 4.3.1 showed an example of fitting an MBKG surrogate model to a two-dimensional problem with noise. Figure 4.22 shows that the predicted limit state (blue curve in Figure 4.22) did not match well to the true limit state (red curve in Figure 4.22). Figure 5.1 shows the contour plot of the true limit state function as the red curve, the MBKG predicted limit state as the blue curve, and the 400 test points as the light blue circles. The 400 test points are uniformly and randomly distributed in the domain. The response values for the 400 test points are predicted using the MBKG surrogate model created in Section 4.3.1 using the 25 DoE samples in Figure 4.22. Each of the 400 test points has a distribution of the predicted response for that test point, i.e., there are 400 distributions of predicted responses, one for each of the 400 test points.

As described previously, the limit state function is the key for doing a reliability analysis. Thus, out of the 400 test points, the ones of interest are the test points whose distribution of the predicted response contains the limit state, i.e., the distribution of the predicted response contains the value zero. Out of the 400 test points, Figure 5.2 shows the 134 test points whose 95% credible sets capture the value zero. These 134 test points are shown as the yellow-filled circles; the light blue circles are the remaining test points. The red asterisks are the 25 DoE samples used to fit the MBKG surrogate model.

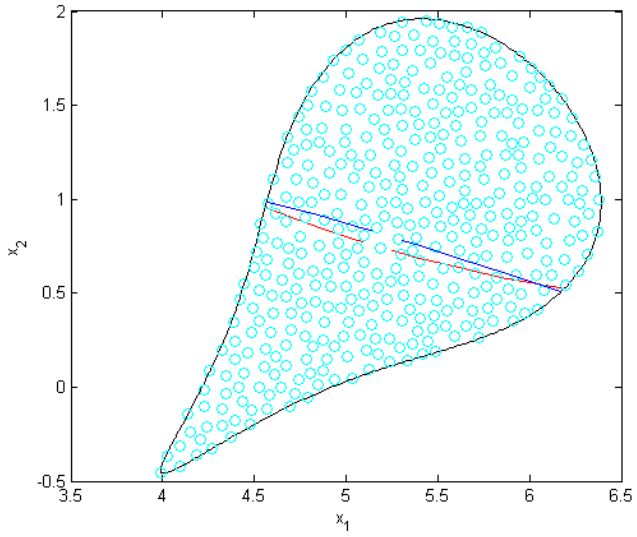

Figure 5.1 Contour Plot of Limit State and 400 Test Points

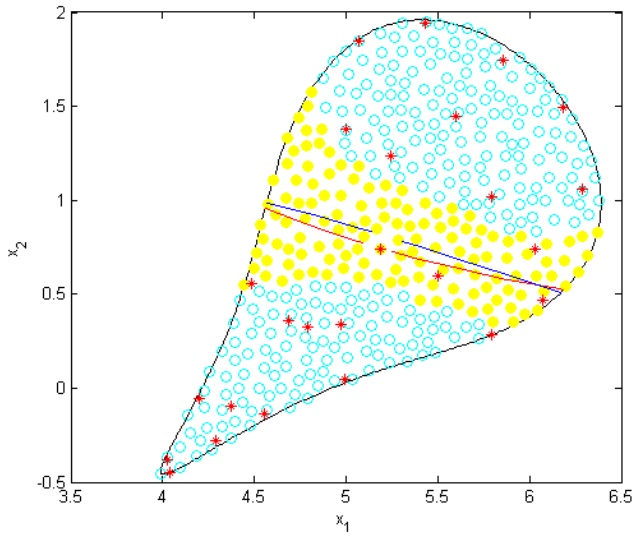

Figure 5.2 Test Points with 95% Credible Sets that Capture the Limit State Using 25 DoE

The yellow-filled circles in Figure 5.2 showing the 95% credible sets that contain zero form a rather wide band around both the true limit state and the predicted limit state.

This wide band is a visualization of the amount of uncertainty we have in the predicted limit state. When studying Figure 5.2, it is not surprising to see that the yellow band contains only six out of the 25 DoE samples used to fit the MBKG surrogate model. This explains why the yellow band is so wide, i.e., why the uncertainty in the predicted limit state is so large.

Naturally, this indicates that adding DoE samples chosen from the 134 test points that form the yellow band would reduce the amount of uncertainty in the limit state and effectively narrow the band. It is also desirable to select the additional DoE samples from the 134 test points such that they are as far from the existing 25 DoE samples as possible, as well as being far from the additional samples added. In addition to the new DoE samples being far from the existing samples, it is also desirable that they are at locations where the amount of uncertainty in the limit state is the largest. Placing the new DoE samples in this strategic way will provide the maximum amount of information to the MBKG surrogate model, requiring fewer DoE samples to be used to fit the MBKG surrogate model.

To determine the locations that meet the desired requirements just described, a weighting system is proposed to select the new DoE samples from the 134 candidate test points. The first step is to calculate the distance between the test points and the existing 25 DoE samples previously used, giving a matrix of distance values. This can be expressed mathematically as

$$D_{ij} = \left\| \mathbf{s}^j - \mathbf{t}^i \right\| \quad \text{for } i = 1, \ldots, n_{test} \text{ and } j = 1, \ldots, n_{DoE} \tag{5.1}$$

where D_{ij} is the distance between the i^{th} test point and j^{th} existing DoE sample, $\mathbf{s}^j$ is the j^{th} existing DoE sample, $\mathbf{t}^i$ is the i^{th} test point, n_{test} is the number of test points whose 95% credible set capture zero, and n_{DoE} is the number of existing DoE samples. For this example, $n_{test} = 134$ and $n_{DoE} = 25$. Once these distances are obtained, the next

step is to find the minimum distance from all the existing DoE samples for each test point. This then gives a vector of the distance between the test point and the closest existing DoE sample, which is expressed mathematically as

$$d_i = \min\left\{D_{ij}\right\} \quad \text{for } j = 1, \ldots, n_{DoE} \tag{5.2}$$

where d_i is the distance between the i^{th} test point and the closest existing DoE sample, D_{ij} is the distance between the i^{th} test point and the j^{th} existing DoE sample as calculated in Eq. (5.1), and n_{DoE} is the number of existing DoE samples.

The last value needed to calculate the weight is the width of the 95% credible sets that capture zero. This width is simply calculated as

$$c_i^w = \left|c_i^L\right| + c_i^U \quad \text{for } i = 1, \ldots, n_{test} \tag{5.3}$$

where c_i^w is the width of the 95% credible set for the i^{th} test point, c_i^L is the lower bound of the 95% credible set for the i^{th} test point, c_i^U is the upper bound of the 95% credible set for the i^{th} test point, and n_{test} is the number of test points. Note that only the 95% credible sets that contain zero are used. Therefore the lower bound, c_i^L, is always less than zero and upper bound, c_i^U, is always greater than zero. The weight for each test point is then calculated as

$$w_i = d_i * c_i^w \quad \text{for } i = 1, \ldots, n_{test} \tag{5.4}$$

where w_i is the weight for the i^{th} test point, d_i is the distance between the i^{th} test point and the closest existing DoE sample as calculated in Eq. (5.2), c_i^w is the width of the 95% credible set for the i^{th} test point as calculated in Eq. (5.3), and n_{test} is the number of test points. The test point that has the largest weight given by Eq. (5.4) is selected as the new DoE sample, expressed as

$$\mathbf{s}^{new} = \max\{w_i\} \quad \text{for } i = 1, \ldots, n_{test} \tag{5.5}$$

where $\mathbf{s}^{new}$ is the new DoE sample selected from the n_{test} test points, w_i is the weight for the i^{th} test point, and n_{test} is the number of test points. Then $\mathbf{s}^{new}$ is added to the list of existing DoE samples, and this selection process can be iterated until the desired number of new DoE samples are added.

For this example, 20 additional DoE samples were added to the 25 initial DoE samples, giving a total of 45 DoE samples used to fit the MBKG surrogate model. Figure 5.3 shows the 20 additional DoE samples added as the black squares in the figure. It can be seen in the figure that the samples selected are relatively uniformly spaced from each other as well as spaced from the existing 25 DoE samples shown as the red asterisks.

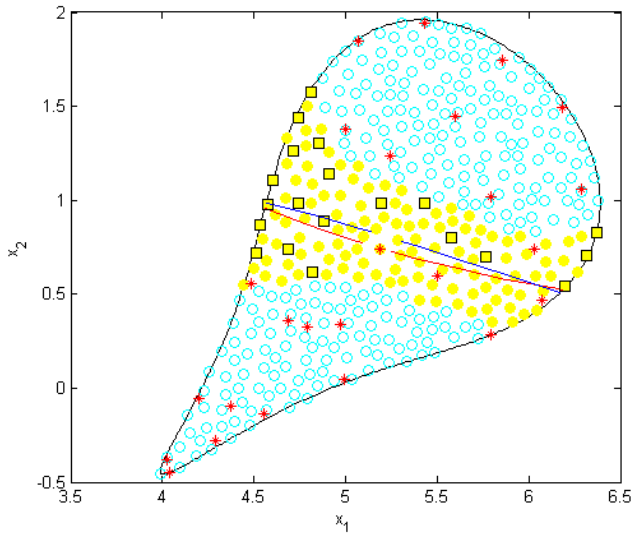

Figure 5.3 Additional 20 DoE Samples

The MBKG surrogate model was refitted using the 45 DoE samples and the same priors and initial values for the MCMC analysis that were used in Section 4.3.1. The contour plot of the limit state using the 45 DoE samples is shown in Figure 5.4 as the

green curve. The blue curve in the figure is the limit state given by the previous MBKG surrogate model that used 25 DoE samples. The red curve in the figure is the true limit state for Eq. (4.4). As seen in the figure, the left part of the green curve, i.e., the limit state using the 45 DoE samples, is a little closer to the true limit state (red curve). However, the right side of the green curve, i.e., the limit state using the 45 DoE samples, is a little farther away from the true limit state (red curve).

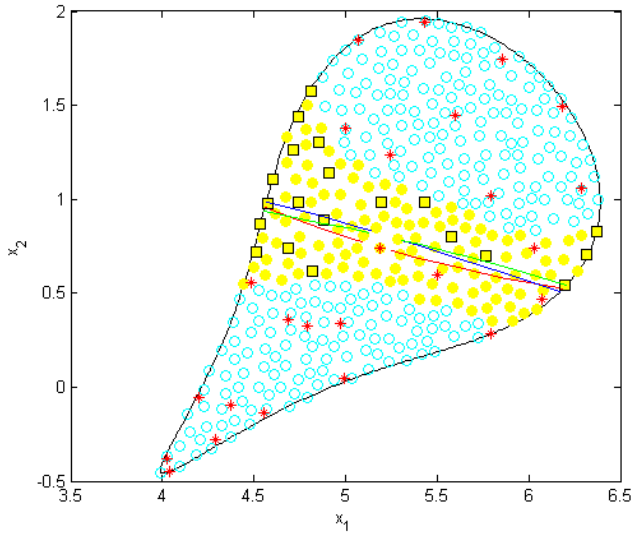

Figure 5.4 Contour Plot of Limit State of Eq. (4.4) Using 45 DoE Samples

Further investigation of the 45 DoE samples used to fit the MBKG surrogate model revealed that the noise added in Eq. (4.4) tended to be a little biased to the positive side. This is believed to be the reason that the predicted limit state for both the 25 DoE samples and 45 DoE samples seems to be consistently above the true limit state in Figure 5.4.

To see the overall improvement of the MBKG surrogate model using the additional 20 DoE samples, the response values for the same 400 test points shown in

Figure 5.1 were predicted using the MBKG surrogate model created using the 45 DoE samples. Then the 400 response values whose 95% credible sets that captured zero were selected again and are shown in Figure 5.5 as the small black diamonds. It is easily seen how the 95% credible set band formed by the small black diamonds is much narrower than the previous 95% credible set band formed by the yellow-filled circles. This shows how adding 20 DoE samples reduced the amount of uncertainty in the predicted limit state.

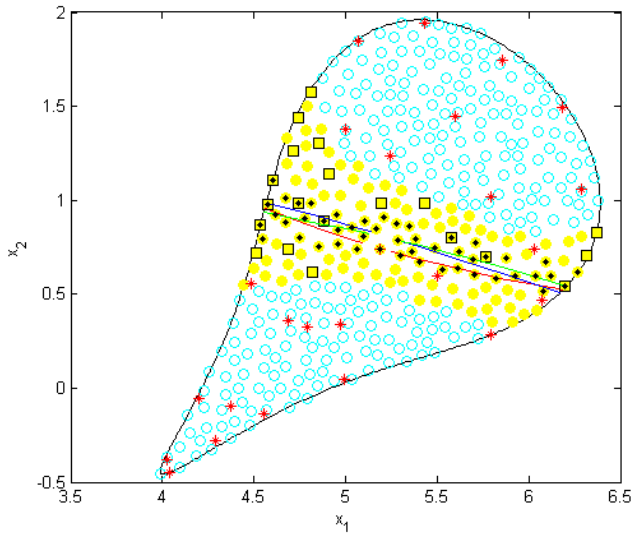

Figure 5.5 Test Points with 95% Credible Sets that Capture the Limit State Using 45 DoE

Despite the existence of bias in the DoE samples as explained earlier, the true limit state (red curve in Figure 5.5) is still within the 95% credible set of the limit state given by the MBKG surrogate model constructed using the 45 DoE samples. Further sequential sampling to add additional DoE samples using the same process will continue to improve the prediction of the limit state and reduce the width of the 95% credible set.

5.3 Sequential Sampling for the Second 2-D Mathematical Example

5.3.1 Prior Distributions and MCMC Initial Values

Section 4.3.2 showed an example of fitting an MBKG surrogate model to a two-dimensional problem with noise. Figure 4.25 showed that the predicted limit state (blue curve in Figure 4.25) did not match well to the true limit state (red curve in Figure 4.25). When fitting the MBKG surrogate model in Section 4.3.2 a constant mean structure was used and a wide prior for the σ^2 parameter was used. As seen in Figure 4.25 the true limit state is highly nonlinear. Therefore, in this section, a second-order mean structure for the MBKG surrogate model will be used. Recall from Section 3.3.1 that using a non-constant mean structure introduces the regression coefficients $\boldsymbol{\beta}$ that need to be fitted when fitting the MBKG surrogate model. Thus, different prior parameter values for the prior distributions are needed than were used in Section 4.3.2 for this example. Table 5.1 lists the prior parameters used for fitting the MBKG surrogate model in this section. The second-order mean structure is expected to absorb the trend of limit state function. Therefore, a slightly tighter prior for σ^2 is used than was used previously. The prior for the λ parameter remains the same as it was previously. The prior for the θ parameters also remains the same as previously used. Because the data used to fit the MBKG surrogate model is normalized as explained in Section 3.3.2, the $\boldsymbol{\beta}$ coefficients are expected to take on small values near zero. Therefore, a noninformative prior for the $\boldsymbol{\beta}$ coefficients that has a zero mean vector and a variance of 650 for each coefficient is used. This gives the 95% interval for the coefficients to be ± 49.97, which is expected to capture the $\boldsymbol{\beta}$ coefficients. The initial values needed to carry out the MCMC simulation to fit the MBKG surrogate model are given in Table 5.2.

Table 5.1 Prior Parameter Values

Unknown Parameter	Prior Distribution	Parameter 1	Parameter 2	95% Interval	
σ^2	Inverse-Gamma	$\alpha_\sigma = 2$	$\beta_\sigma = 2.5$	0.49	10.32
λ	Inverse-Gamma	$\alpha_\lambda = 1.9548$	$\beta_\lambda = 0.2747$	0.05	1.20
θ_i	Uniform	$a_\theta = 0.15$	$b_\theta = 5$	0.27	4.88
β	MVN	**0**	650**I**	–49.97	49.97

Table 5.2 Initial Values for Markov Chains

Unknown Parameter	Chain 1 Initial Values	Chain 2 Initial Values	Chain 3 Initial Values
σ^2	1.00	3.00	6.0
λ	0.06	0.50	0.80
θ_i	0.20	2.50	5.00
β_i	0.00	-50.0	50.0

5.3.2 Sequential Sampling

The sequential sampling process using the 95% credible set that was presented in Section 5.2 will be used for this example as well. Out of the 400 test points predicted using the MBKG surrogate model, Figure 5.6 shows the 298 test points whose 95% credible sets capture the value zero. These 298 test points are shown as the yellow-filled circles; the light blue circles are the remaining 102 test points whose 95% credible sets did not capture zero. The red asterisks are the 25 DoE samples used to fit the MBKG surrogate model. The red curve is the true limit state for Eq. (4.6). The blue curve is the predicted limit state given by the MBKG surrogate model. It is clearly seen that the mean of the predicted limit state does not match well to the true limit state in the top portion. It does, however, match well for the left part of the limit state.

The large number of test points that have a 95% credible set that captures zero indicates that there is a large amount of uncertainty in the MBKG surrogate model about where the true limit state is. Recalling from Figure 4.23, which showed the 3-D surface plot of Eq. (4.6), the function is relatively flat in the region near the limit state and then suddenly drops off. This explains why there is so much uncertainty in the MBKG surrogate model about where the limit is compared to the example in the previous section. Using the same sequential sampling procedure described in the previous section, 20 additional DoE samples are selected from the 298 test points to be used to improve the MBKG surrogate model. The 20 additional DoE samples are shown as the black squares in Figure 5.7.

The MBKG surrogate model was refitted using the 45 DoE samples, and the same priors and initial values listed in Section 5.3.1 were used for the MCMC simulation. The contour plot of the limit state using the 45 DoE samples is shown in Figure 5.8 as the green curve. The blue curve in the figure is the limit state given by the previous MBKG surrogate model that used 25 DoE samples. The red curve in the figure is the true limit state for Eq. (4.6). Figure 5.8 shows that the left part of the limit state (green curve) still matches as closely to the true limit state (red curve) as it did previously. It is also seen that the top part of the limit state (green curve) has improved and is closer to the true limit state than it was previously. The figure also shows that, even with the 45 DoE samples, only a few are actually near the true limit state.

To see the overall improvement of the MBKG surrogate model using the additional 20 DoE samples, the response values for the same 400 test points were predicted using the MBKG surrogate model created using the 45 DoE samples. Then the 400 response values whose 95% credible sets that captured zero were selected again and are shown in Figure 5.9 as the small black diamonds. It is easily seen how this 95% credible set band formed by the small black diamonds did shrink in on both sides of the true limit state compared to the previous 95% credible set band formed by the yellow-

filled circles. This shows how adding 20 DoE samples reduced the amount of uncertainty in the predicted limit state. From the figure it can be seen that the uncertainty in the lower left part of the limit state is rather small; this is reflected by the narrow 95% credible set in that area. However, the uncertainty in the upper top part of the limit state is still rather large, as reflected by the wider 95% credible set in that area. This indicates that additional DoE samples are needed to improve the MBKG surrogate model. Further sequential sampling using the same process is continued to improve the prediction of the limit state and reduce the width of the 95% credible set.

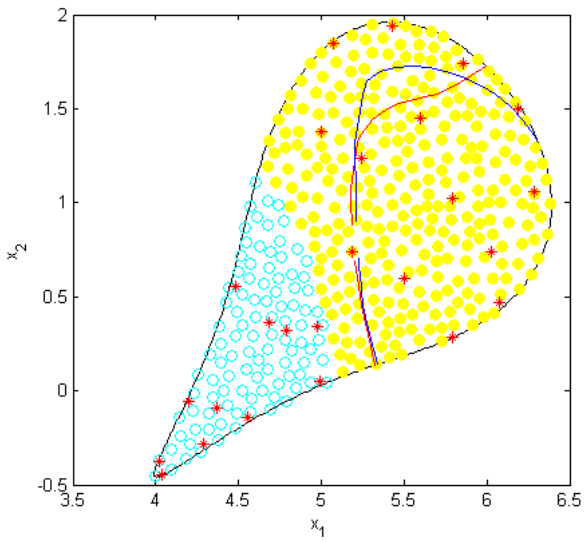

Figure 5.6 Test Points with 95% Credible Sets that Capture the Limit State Using 25 DoE

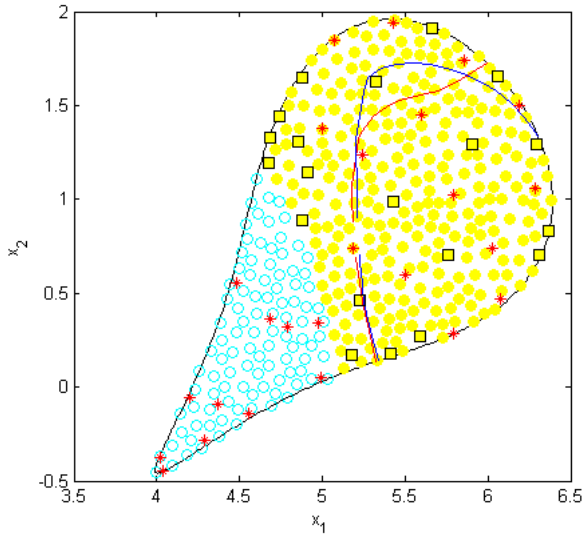

Figure 5.7 Additional 20 DoE Samples Shown as Black Squares

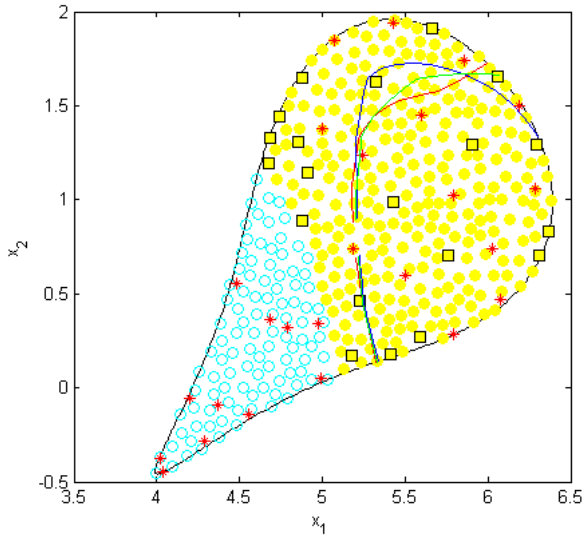

Figure 5.8 Contour Plot of Limit State Using 45 DoE Samples

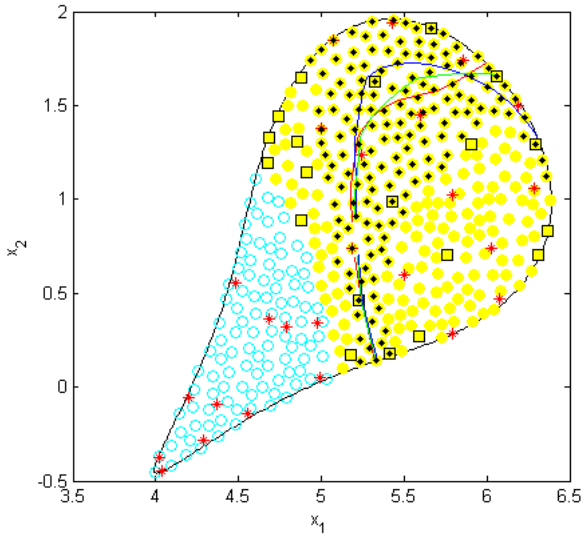

Figure 5.9 Test Points with 95% Credible Sets that Capture the Limit State Using 45 DoE

In Figure 5.9 there are 154 test points, shown as the black diamonds that have 95% credible sets that capture zero. Using these 154 test points, an additional 20 DoE samples are selected using the same sequential sampling method as before. The additional 20 DoE samples selected are shown in Figure 5.10 as the pink circles. The MBKG surrogate model was refitted using the 65 DoE samples. The contour plot of the predicted limit state using the 65 DoE samples is shown in Figure 5.11 as the blue curve. It is seen that the predicted limit (blue curve) is a close match to the true limit except for the top portion, which is a better fit than it was using the 45 DoE samples as shown by the green curve in Figure 5.8. The response values of the 400 test points were predicted using the MBKG surrogate model created using the 65 DoE samples. The 400 response values whose 95% credible sets captured zero were selected and are shown in Figure 5.12 as the light blue plus signs. It is seen that the 95% credible set did shrink, but not significantly.

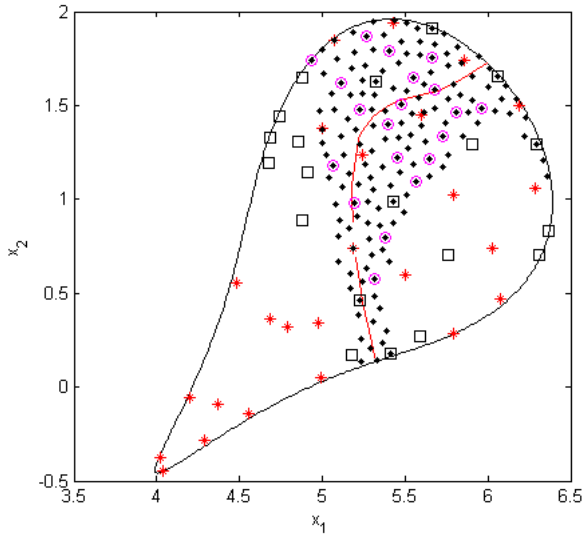

Figure 5.10 Additional 20 DoE Samples Shown as Pink Circles

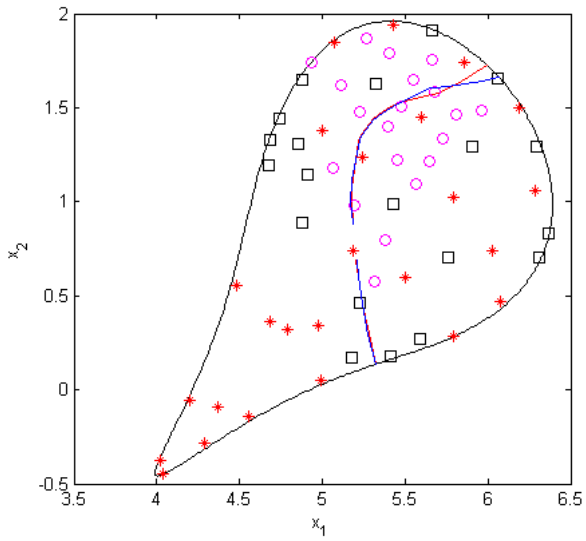

Figure 5.11 Contour of Limit State Using 65 DoE Samples

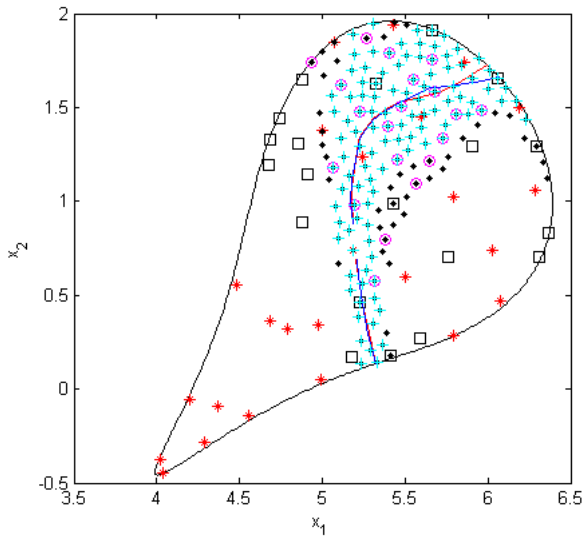

Figure 5.12 Test Points with 95% Credible Sets that Capture the Limit State Using 65 DoE Samples Shown as the Light Blue Plus Signs

Using the same sequential sampling procedure, another 20 DoE were selected from the 112 test points whose 95% credible sets captured zero. Figure 5.13 shows the additional 20 DoE as the blue diamonds. Figure 5.13 also shows the predicted limit state given by the MBKG surrogate created using the 85 DoE samples as the green curve. It is seen that limit state did improve a little but did not change significantly. Figure 5.14 shows the test points that have 95% credible sets capturing zero for the MBKG surrogate created using the 85 DoE samples; they are shown as the black crosses in the figure. It can be seen that the credible set has shrunk when compared to the credible set given by the 65 DoE samples shown in Figure 5.12. A final set of 20 DoE samples was added to give a total of 105 DoE samples using the same sequential sampling method. The additional 20 DoE are shown in Figure 5.15 as the green asterisks. It is seen that the 20 additional DoE closely surround the true limit state and also that 19 of them are in the top part where the credible set is wider. The predicted limit state given by the MBKG surrogate created using all 105 DoE is also shown in Figure 5.15 as the blue curve. It can

be seen that the curve appears to visually match the true limit state exactly except for at the top right tip. The test points with 95% credible sets capturing zero are shown in Figure 5.16 as the red circles. As seen in the figure, the credible set has shrunk more, thus reflecting the accurate prediction of the true limit state.

This section demonstrated how the developed sequential sampling method systemically reduces the uncertainty in the predicted limit and therefore reduces the 95% credible set of the limit state. The next section will show how the sequential sampling method reduces the uncertainty in the predicted probability of failure.

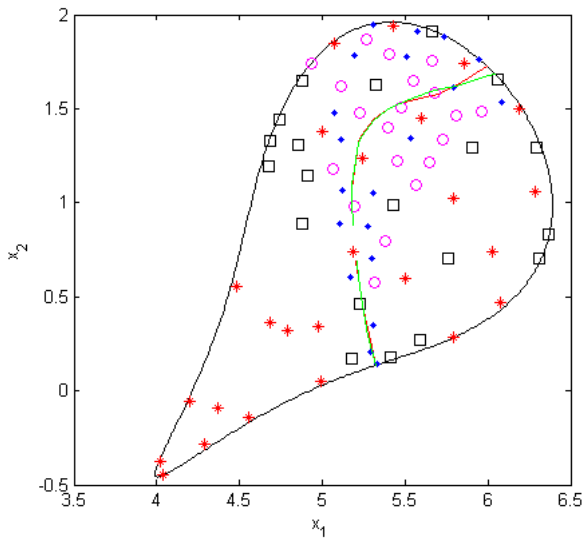

Figure 5.13 Additional 20 DoE Samples as Blue Diamonds and Contour of Limit State Using 85 DoE Samples as the Green Curve

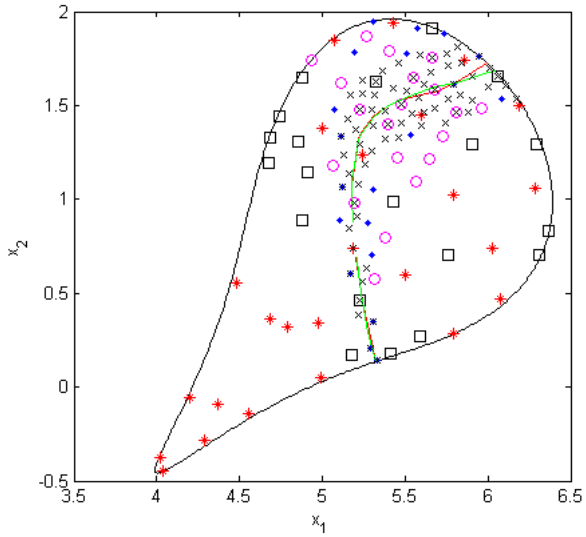

Figure 5.14 Test Points with 95% Credible Sets that Capture the Limit State Using 85 DoE Samples Shown as Black Crosses

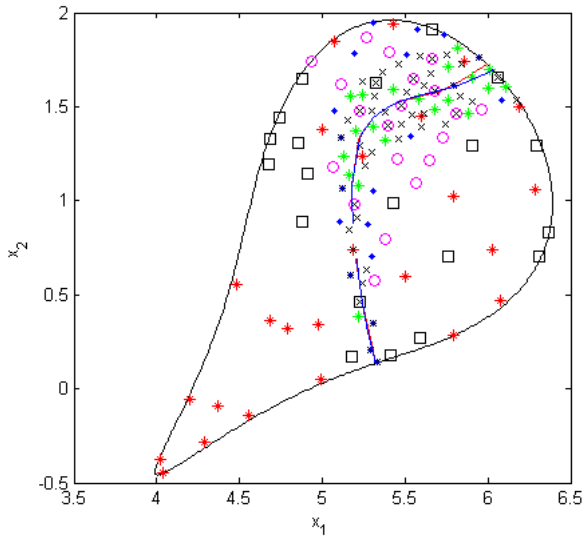

Figure 5.15 Additional 20 DoE Samples as Green Asterisk and Contour of Limit State Using 105 DoE Samples as the Blue Curve

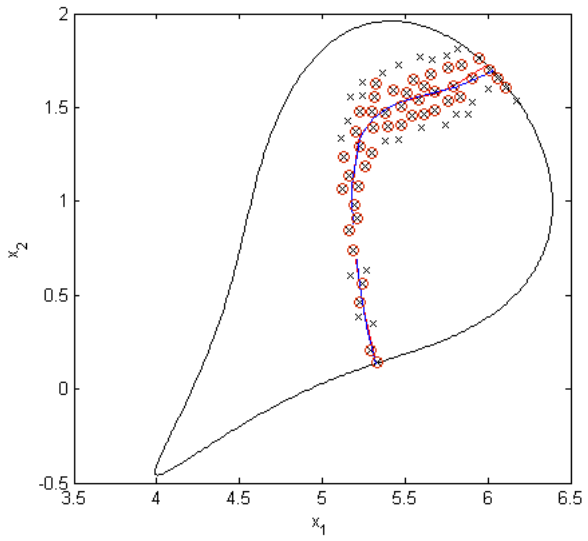

Figure 5.16 Test Points with 95% Credible Sets that Capture the Limit Statue Using 105 DoE Samples Shown as the Red Circles

5.3.3 Distribution of Probability of Failure for Small Noise

The previous section demonstrated how the predicted limit state converged to the true limit state using the developed sequential sampling method. It was also shown how the uncertainty of the predicted limit state was decreased by using the developed sequential sampling method. This was demonstrated by showing how the 95% credible set of the predicted limit state continued to shrink and become narrower as sequential sampling was performed. Recall that the limit state is used to predict the probability of failure. Because of the uncertainty in the predicted limit state, this naturally leads to uncertainty in the predicted probability of failure when using the predicted limit state.

When using Bayesian methods, all of the uncertainty about a predicted value is reflected in the posterior distribution of the predicted value. Thus, when predicting the probability of failure using the MBKG surrogate model, the probability of failure is not a deterministic value, but rather a distribution. The posterior distribution of the probability of failure captures all of the uncertainty in the probability of failure prediction. This

includes all of the uncertainty in the MBKG surrogate model, i.e., the uncertainty of the estimated parameters in the MBKG surrogate model and the uncertainty in the predicted limit state. This section will demonstrate how the distribution of the probability of failure converges as the sequential sampling method developed in Section 5.2 is used to improve the predicted limit state.

The MBKG surrogate models created using the sequential sampling method will be used in this section to predict the probability of failure. For the study of the distribution of the probability of failure, two different concepts will be shown. This section will show how the distribution of the probability of failure converges when using the developed sequential sampling method. The next section will show how the distribution converges for three different noise levels in the response function.

The response function used in this section is the same G_2 function given in Eq. (4.8) repeated below as Eq. (5.6).

$$\begin{aligned} G_2(X) &= -1 + (0.9063X_1 + 0.4226X_2 - 6)^2 \\ &+ (0.9063X_1 + 0.4226X_2 - 6)^3 \\ &- 0.6(0.9063X_1 + 0.4226X_2 - 6)^4 \\ &- (-0.4226X_1 + 0.9063X_2) \end{aligned} \tag{5.6}$$

Noise is added to the response function as shown in Eq. (5.7)

$$y_i = G_2(\mathbf{x}_i) + \varepsilon_i \tag{5.7}$$

where ε_i is generated as a realization from a normal distribution with zero mean and standard deviation of σ_ε as shown in Eq. (5.8).

$$\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right) \tag{5.8}$$

Three different noise levels will be considered and will be referred to as small, medium, and large noise. The noise levels, the standard deviation for each noise level, and the 95% interval for each noise distribution is given in Table 5.3.

Table 5.3 The Three Noise Levels Used

Noise Level	σ_e	95% Interval	
Small	0.01	–0.0196	0.0196
Medium	0.05	–0.098	0.098
Large	0.10	–0.196	0.196

For this example, the two random variables X_1 and X_2 are correlated to each other with a Clayton Copula with a correlation coefficient of 0.5. The marginal distributions of both random variables are given in Eq. (5.9). To calculate the probability of failure, Monte Carlo points are drawn from the joint distribution of X_1 and X_2.

$$\begin{aligned} \mathrm{X}_1 &\sim N\left(5.19, 0.3^2\right) \\ \mathrm{X}_2 &\sim N\left(0.74, 0.3^2\right) \end{aligned} \tag{5.9}$$

This section shows how using a different number of DoE samples affects the distribution of the probability of failure for the small noise case. Figures 5.17 – 5.21 show the posterior distributions of the probability of failure using 25, 45, 65, 85, and 105 DoE samples, respectively. To predict the probability of failure, 10,000 Monte Carlo points were used to predict the probability of failure at each MCMC iteration; a total of 300,000 MCMC iterations were used for this example. Recall from Figure 5.6 how the 95% credible set was really wide for the limit state, indicating that the uncertainty in the limit is large. This large uncertainty is carried through and is included in the uncertainty of the

probability of failure, as shown in Figure 5.17. The distribution of the probability of failure is extremely wide, ranging between roughly 0.0% to just under 90%; thus, it is clear that the amount of uncertainty in the probability of failure is rather large. After adding another 20 DoE samples and using a total of 45 DoE samples to fit the MBKG surrogate model, the amount of uncertainty in the limit state—and therefore the amount of uncertainty in the probability of failure—shrinks. This is shown by the distribution of the probability of failure in Figure 5.18. The distribution is not as wide and the left tail of the distribution is not nearly as fat as it was when using only 25 DoE samples. Continuing to add 20 DoE samples using the sequential sampling method continues to shrink the width of the distribution of the probability of failure. Figure 5.21 shows the distribution of the probability of failure using 105 DoE samples, and the distribution is significantly narrower than when only 25 DoE samples are used.

Table 5.4 shows a summary of the statistics for the probability of failure using the different numbers of DoE samples. The table lists the mean value, the standard deviation, and the 95% credible set for the probability of failure. It can be seen that the 95% credible set using 25 DoE samples is wide, with a lower bound of 3.94% and an upper bound of 71.82%. This means that there is 95% probability that the true probability of failure is in the interval [3.94%, 71.82%]. It is seen in the table how the 95% credible set continues to shrink as more DoE samples are added. Using the 105 DoE samples, the 95% credible set of the probability of failure now has a lower bound of 43.75% and an upper bound of 51.73%. This means that there is 95% probability that the true probability of failure is in the interval [43.75%, 51.73%]. The true probability of failure of Eq. (5.6) without noise is 48.4719%, as shown in Table 5.4. The 95% credible sets in the table capture the true probability of failure. Also, the mean value of the probability of failure approaches the true value.

The next section will look at how the level of noise in the response function affects the posterior distribution of the probability of failure.

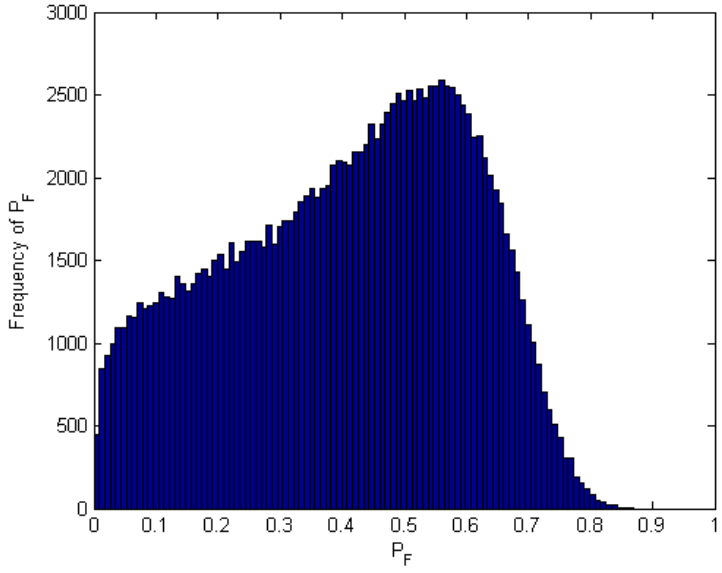

Figure 5.17 Posterior Distribution of Probability of Failure Using 25 DoE Samples

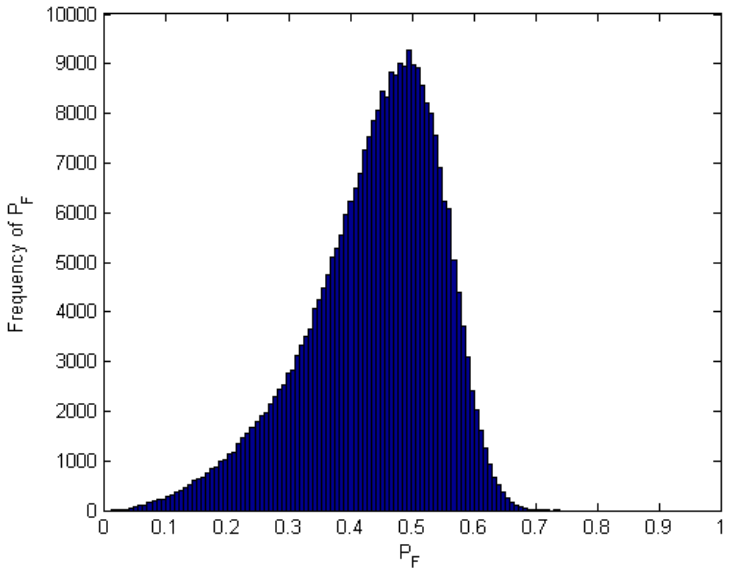

Figure 5.18 Posterior Distribution of Probability of Failure Using 45 DoE Samples

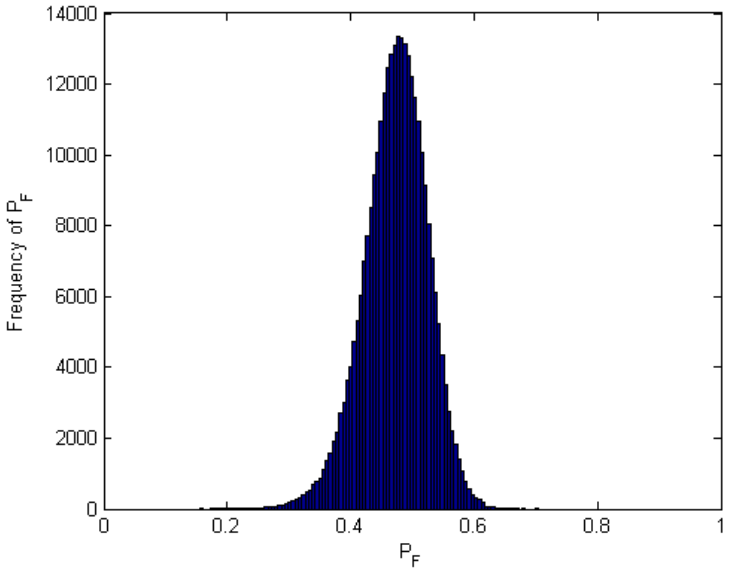

Figure 5.19 Posterior Distribution of Probability of Failure Using 65 DoE Samples

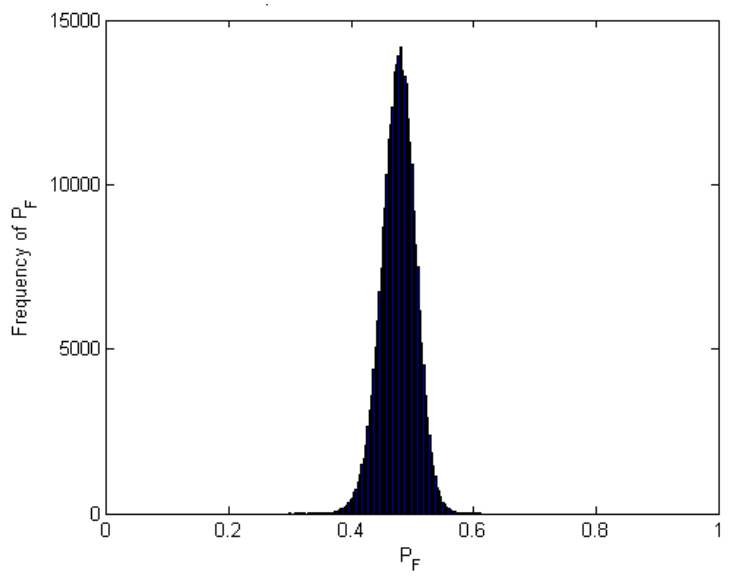

Figure 5.20 Posterior Distribution of Probability of Failure Using 85 DoE Samples

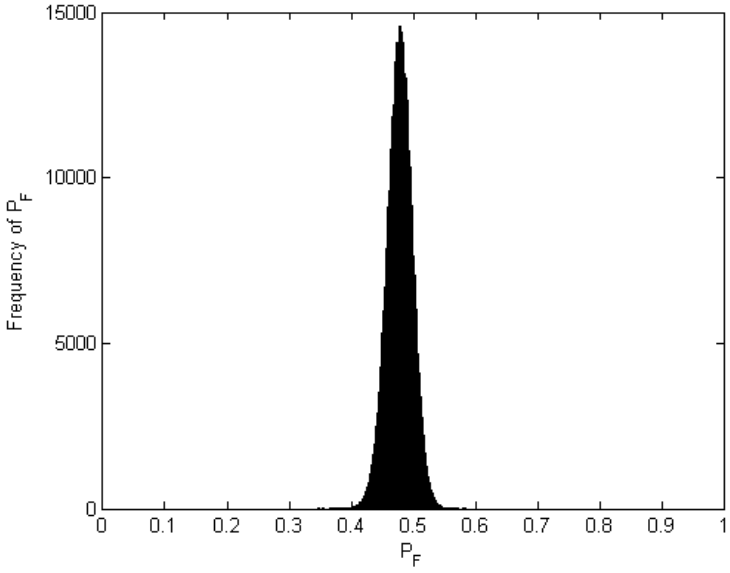

Figure 5.21 Posterior Distribution of Probability of Failure Using 105 DoE Samples

Table 5.4 Probability of Failure Statistics Using Different Numbers of DoE Samples

# DoE Samples	Mean	Std.	95% Credible Set	
25	40.6553 %	19.3437 %	3.94 %	71.82 %
45	43.7836 %	10.7319 %	18.41 %	60.24 %
65	47.2684 %	5.1559 %	36.26 %	56.63 %
85	47.6843 %	2.7809 %	41.97 %	52.92 %
105	47.8461 %	2.0249 %	43.75 %	51.73 %
True	48.4719 %	N/A	N/A	N/A

5.3.4 Distribution of the Probability of Failure for Small, Medium, and Large Noise

This section will show how the posterior distribution of the probability of failure changes for different levels of noise when using the developed sequential sampling

method. The plots of the posterior distribution of the probability of failure for medium and large noise look similar to those for the small noise. Thus, they are not shown here. Instead, tables showing the statistics that summarize the posterior distributions are presented. Tables 5.5 – 5.7 show the statistics of the posterior distribution of the probability of failure for the small, medium, and large noise levels, respectively. Note that Table 5.5 is the same as Table 5.4 and is just repeated for convenience. When studying the tables, it is evident that the credible sets shrink as more DoE samples are added for all three noise levels. It should be noted that the 25 initial DoE samples used for the small, medium, and large noise cases are different. This is because the initial uniform samples are randomly generated. When comparing the 95% credible sets for the medium and large noise cases, it is evident that they are similar. The credible sets for large noise are slightly larger than they are for medium noise, which is expected.

The 95% credible sets for the small noise are slightly larger than they are for the medium and large noise. It was found that the 25 initial DoE samples were not as uniform in the domain for the small noise as they were for the medium and large noise. Because of this, there is less information available when fitting the MBKG surrogate model, i.e., there is more uncertainty in the MBKG surrogate model. Therefore, the 95% credible sets are expected to be a little wider for the small noise for the same number of DoE samples than they are for the medium and large noise.

The Bayesian definition of the 95% credible sets for the probability of failure is that, given the information we have, i.e., the DoE samples and prior information, there is 95% probability that the true probability of failure is within the bounds of the 95% credible set. For this example, it is seen that the true probability of failure is in fact within the 95% credible sets for all noise levels considered.

The next chapter develops a method that uses the distribution of the probability of failure to carry out confidence-based reliability-based design optimization to obtain a reliable design with confidence.

Table 5.5 Probability of Failure Statistics for Small Noise

# DoE Samples	Mean	Std.	95% Credible Set	
25	40.6553 %	19.3437 %	3.94 %	71.82 %
45	43.7836 %	10.7319 %	18.41 %	60.24 %
65	47.2684 %	5.1559 %	36.26 %	56.63 %
85	47.6843 %	2.7809 %	41.97 %	52.92 %
105	47.8461 %	2.0249 %	43.75 %	51.73 %
True	48.4719 %	N/A	N/A	N/A

Table 5.6 Probability of Failure Statistics for Medium Noise

# DoE Samples	Mean	Std.	95% Credible Set	
25	42.1873 %	16.4255 %	9.26 %	69.69 %
45	45.3714 %	8.5311 %	25.91 %	59.11 %
65	47.0987 %	4.0376 %	38.37 %	54.19 %
85	48.1298 %	2.2939 %	43.31 %	52.35 %
105	48.0290 %	1.8029 %	44.3 %	51.37 %
True	48.4719 %	N/A	N/A	N/A

Table 5.7 Probability of Failure Statistics for Large Noise

# DoE Samples	Mean	Std.	95% Credible Set	
25	43.4495 %	16.1065 %	10.4 %	70.19 %
45	45.1517 %	8.5012 %	25.87 %	58.97 %
65	46.5778 %	4.1338 %	37.6 %	53.84 %
85	47.0866 %	2.6179 %	41.47 %	51.8 %
105	47.4154 %	2.3101 %	42.44 %	51.51 %
True	48.4719 %	N/A	N/A	N/A

CHAPTER 6
CONFIDENCE-BASED RELIABILITY-BASED DESIGN OPTIMIZATION VIA POSTERIOR DISTRIBUTIONS

6.1 Introduction

The previous chapter demonstrated how the posterior distribution of the probability of failure converged using the developed sequential sampling method. This chapter is going to develop a confidence-based reliability-based design optimization (RBDO) method that uses the posterior distribution of the probability of failure to carry out optimization. As explained in the previous chapter, the posterior distribution of the probability of failure contains all of the uncertainty about the probability of failure. The goal of developing a confidence-based RBDO method is to use the uncertainty in the optimization process in order to obtain a reliable design with a user-specified confidence level.

The next section will present the formulation of the confidence-based RBDO method. The following sections will then present examples using the confidence-based RBDO method.

6.2 Confidence-Based Reliability-Based Design Optimization Formulation

Recall from Section 2.4 that the constraint formulation for RBDO is that the probability of failure should be less than or equal to the target probability of failure as shown in Eq. (6.1), where P_F is the probability of failure for the current design and P_F^{TAR} is the target probability of failure. When using the MBKG surrogate model, we do not know the true probability of failure. However, we do have the posterior distribution of the probability of failure. The mean value of the probability of failure could be used as a point estimate of the true probability of failure. That point estimate could then potentially

be used to carry out RBDO. This, however, could lead to an optimum design that does not meet the target probability of failure because there is uncertainty about what the true probability of failure is.

$$P_F \leq P_F^{TAR} \tag{6.1}$$

It is desirable to take the uncertainty of the probability of failure into consideration for optimization. The posterior distribution of the probability of failure can be used to do exactly this. Figure 6.1 shows a simple representation of the posterior distribution of the probability of failure. Note that this is just for illustration proposes and the actual posterior distribution of the probability of failure may not be symmetric as was seen in Section 5.3. In Figure 6.1, the green line represents where the target probability of failure, P_F^{TAR}, is located on the $x-$axis. The gray area in the figure is the probability that the probability of failure is less than or equal to the target probability of failure as expressed in Eq. (6.2)

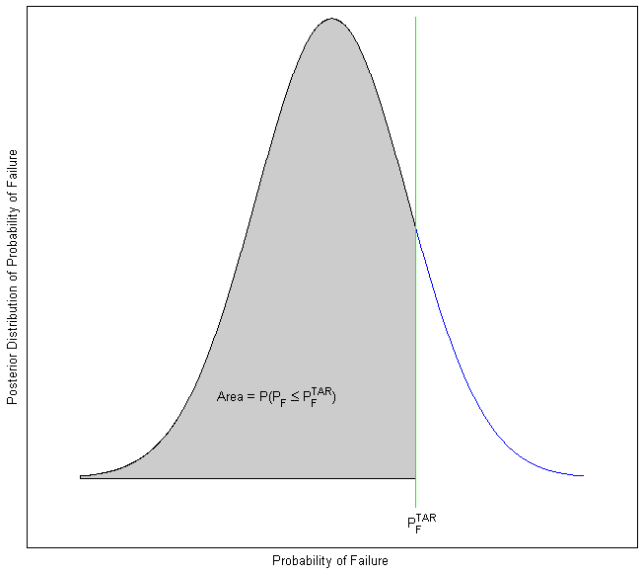

Figure 6.1 Posterior Distribution of the Probability of Failure

$$Area = P\left(P_F \le P_F^{TAR}\right) \tag{6.2}$$

where $Area$ is the gray shaded area in Figure 6.1, $P(\bullet)$ is a probability measure, P_F is the probability of failure, and P_F^{TAR} is the target probability of failure. For confidence-based RBDO, a target confidence level, denoted as $C.L.$, is defined. For example, if $C.L. = 90\%$, the goal is to find a design such that there is 90% probability that the probability of failure is less than or equal to the target probability of failure. If this condition is satisfied, then the grey shaded area in Figure 6.1 will be at least equal to $C.L.$, expressed mathematically as

$$Area = P\left(P_F \le P_F^{TAR}\right) \ge C.L. \tag{6.3}$$

where $C.L.$ is the target confidence level, e.g., 90%. Equation (6.3) states that the probability of the probability of failure being less than or equal to the target probability of failure is greater than or equal to the target confidence level $C.L.$

Now take for example the posterior distribution of the probability of failure for the current design shown in Figure 6.2. If the target confidence level is $C.L. = 80\%$, then the gray shaded area in Figure 6.2 is equal to 20%. The quantile value for $C.L. = 80\%$ is determined from the posterior distribution to be $P_F^{(C.L.)} = 7.3\%$ as shown in Figure 6.2. This means that there is 80% probability that the true probability of failure is less than or equal to 7.3%. Now take for the same current design and posterior distribution of the probability of failure. If the target confidence level is 90% as shown in Figure 6.3, then the gray shaded area is equal to 10%. Thus, the quantile value is $P_F^{(C.L.)} = 12.4\%$ as shown in Figure 6.3. This means that for the same current design there is 90% probability that the true probability of failure is less than or equal to 12.4%. The goal of the confidence-based RBDO method is to find a design point such that quantile value $P_F^{(C.L.)}$ is less than or equal to the target probability of failure P_F^{TAR}. This would mean that there is $C.L.$ probability that the true probability of failure is less than or equal to $P_F^{(C.L.)}$.

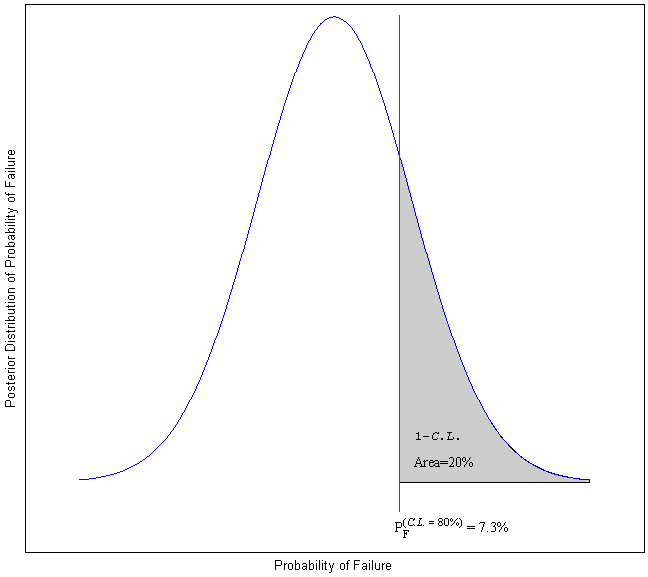

Figure 6.2 Posterior Distribution of the Probability of Failure with $C.L. = 80\%$

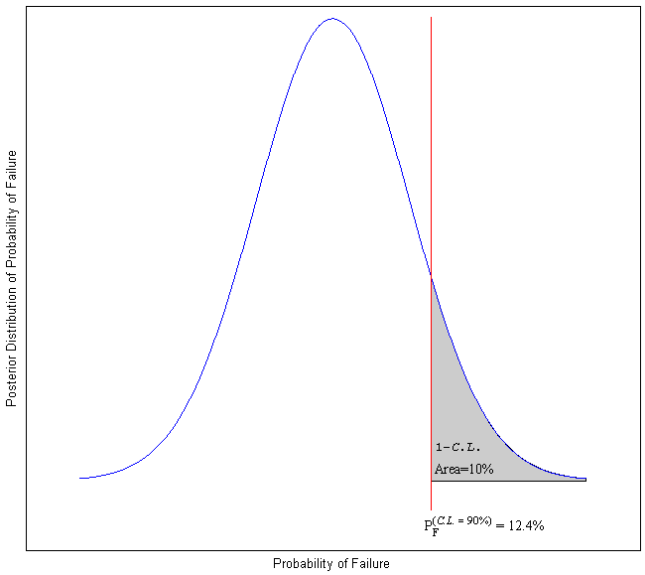

Figure 6.3 Posterior Distribution of the Probability of Failure with $C.L. = 90\%$

Using the posterior distribution of the probability of failure, the constraint used for confidence-based RBDO is expressed in Eq. (6.4)

$$P_F^{(C.L.)} \leq P_F^{TAR} \tag{6.4}$$

where $P_F^{(C.L.)}$ is the probability of failure quantile value at the target confidence level, $(C.L.)$ denotes the target confidence level, and P_F^{TAR} is the target probability of failure. The constraint states that the probability of failure quantile value for the target confidence level $C.L.$ should be less than or equal to the target probability of failure P_F^{TAR}. Figure 6.4 shows a diagram of the confidence-based RBDO constraint in Eq. (6.4). In the figure the blue curve is a representation of the posterior distribution of the probability of failure. The green line in the figure is the target probability of failure, denoted as P_F^{TAR}. The gray shaded area in the figure has an area equal to $1-C.L.$; e.g., if the target confidence level is 90%, then the gray area would be 10% of the area under the blue curve. The red line in the figure is the value of the probability of failure at which the gray shaded area is equal to $1-C.L.$, i.e., the quantile value for $C.L.$ denoted as $P_F^{(C.L.)}$ in the figure. As seen in the figure, the constraint given in Eq. (6.4) is violated because $P_F^{(C.L.)} > P_F^{TAR}$. During the confidence-based RBDO optimization process, the optimization algorithm would update the design so that the posterior distribution of the probability of failure would move to the left and the red line in Figure 6.4 meets the green line or is to the left of the green line, as shown by the arrow in the figure, in order to satisfy the constraint in Eq. (6.4).

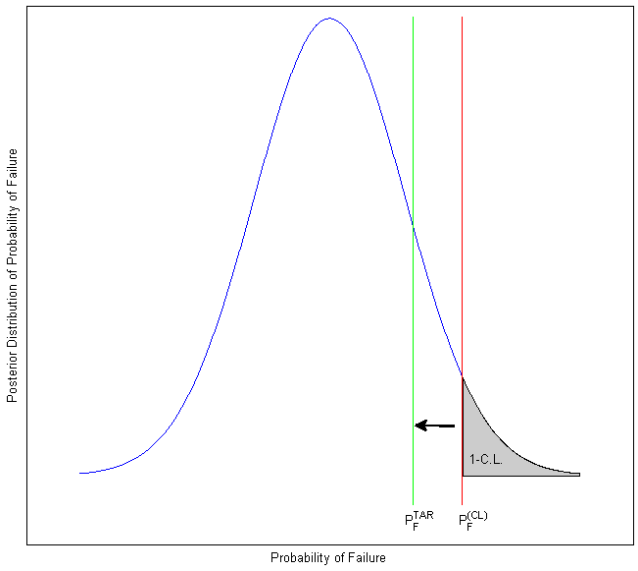

Figure 6.4 Constraint Diagram for Eq. (6.4)

The mathematical formulation for confidence-based RBDO is expressed as

$$\begin{aligned} &\text{minimize} \quad \text{Cost}(\mathbf{d}) \\ &\text{subject to} \quad \begin{cases} P_{F_i}^{(C.L.)} \le P_{F_i}^{TAR}, i = 1, \cdots, n_c \\ \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \mathbf{d} \in \mathbb{R}^{n_d} \text{ and } \mathbf{X} \in \mathbb{R}^{n_r} \end{cases} \end{aligned} \tag{6.5}$$

where $\mathbf{d} = \{d_i\}^T = \{E[X_i]\}^T$, $i = 1 \text{ to } n_d$ is the design vector, $E[\bullet]$ is the expectation operator, $\mathbf{X} = \{X_i\}^T$ is the vector of random variables, and n_c, n_d, and n_r are the number of constraints, design variables, and random variables, respectively. The constraint in Eq. (6.4) can be rewritten in normalized form to be used as an optimization constraint and is expressed as

$$h = \frac{P_F^{(C.L.)}}{P_F^{TAR}} - 1 \le 0 \tag{6.6}$$

To use a sensitivity-based optimization algorithm to find the solution to Eq. (6.5), the sensitivity of the constraint in Eq. (6.6) is needed. The sensitivity of this constraint can be written as

$$\frac{\partial h}{\partial d_j} = \frac{1}{P_F^{TAR}} \frac{\partial P_F^{(C.L.)}}{\partial d_j} \tag{6.7}$$

where d_j is the j^{th} design variable, $P_F^{(C.L.)}$ is the probability of failure quantile value at the target confidence level $C.L.$, and P_F^{TAR} is the target probability of failure. It can be seen that the partial derivative on the right-hand side of Eq. (6.7) is the partial derivative of the probability of failure with respect to the design variable when the probability of failure is equal to $P_F^{(C.L.)}$. Recall that the definition of the design variables in Eq. (6.5) are the mean values of the random variables $\mathbf{X}$. Therefore, the partial derivative on the right-hand side of Eq. (6.7) can be rewritten as

$$\left. \frac{\partial P_F}{\partial \mu_j} \right|_{P_F = P_F^{(C.L.)}} \tag{6.8}$$

where $\mu_j = E[X_j]$ is the mean value of the random variable X_j, P_F is the probability of failure, and $P_F^{(C.L.)}$ is the probability of failure quantile value at the target confidence level $C.L.$ Upon studying the partial derivative in Eq. (6.8), it can be seen that it takes on the same form as Eq. (2.21) and thus can be calculated using the same score function method presented in Section 2.5.2 [Rubinstein and Shapior 1993; Rahman 2009; Lee et. al. 2011; Zhao 2011]. By gathering all of these pieces together, the optimization problem in Eq. (6.5) can now be solved. The next section will present an example carrying out confidence-based RBDO.

6.3 A 2-D Mathematical Example

6.3.1 Problem Definition

A two-dimensional mathematical confidence-based RBDO problem is formulated as

$$
\begin{aligned}
&\text{minimize} \quad \text{Cost}(\mathbf{d}) \\
&\text{subject to} \quad \begin{cases} P_{F_i}^{(C.L.)} \le P_{F_i}^{TAR}, i = 1, 2, 3 \\ \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \mathbf{d} \in \mathbb{R}^2 \text{ and } \mathbf{X} \in \mathbb{R}^2 \end{cases} \\
&\text{where} \\
&Cost(\mathbf{d}) = -\frac{(d_1 + d_2 - 10)^2}{30} - \frac{(d_1 - d_2 + 10)^2}{120} \\
&G_1(\mathbf{X}) = 1 - \frac{X_1^2 X_2}{20} \\
&G_2(\mathbf{X}) = -1 + (0.9063X_1 + 0.4226X_2 - 6)^2 \\
&\qquad + (0.9063X_1 + 0.4226X_2 - 6)^3 \\
&\qquad - 0.6(0.9063X_1 + 0.4226X_2 - 6)^4 \\
&\qquad - (-0.4226X_1 + 0.9063X_2) \\
&G_3(\mathbf{X}) = 1 - \frac{80}{(X_1^2 + 8X_2 + 5)} \\
&h_i(\mathbf{X}) = G_i(\mathbf{X}) + \varepsilon, i = 1, 2, 3 \\
&\mathbf{d}^L = [0, 0]^T, \mathbf{d}^U = [10, 10]^T, \mathbf{d}_{initial} = [5.19, 0.74]^T \\
&X_j \sim N(d_j, 0.3^2), j = 1, 2 \\
&\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right) \\
&\mathrm{P}_{F_i} = P\left(h_i(\mathbf{X}) > 0\right), i = 1, 2, 3 \\
&P_{F_i}^{Tar} = 2.275\%, i = 1, 2, 3 \text{ and } C.L._i = 90\%, i = 1, 2, 3
\end{aligned}
\tag{6.9}
$$

As seen in Eq. (6.9), the target confidence level defined for this example is 90%, and the target probability of failure is defined to be 2.275%. The general problem definition for this example has random noise added to the true response value. Random noise is defined in the same way as it was in Section 5.3.3 and also uses the three different noise levels listed in Table (5.3). Table (5.3) is repeated below as Table 6.1.

Table 6.1 The Three Noise Levels Used

Noise Level	σ_e	95% Interval	
Small	0.01	–0.0196	0.0196
Medium	0.05	–0.098	0.098
Large	0.10	–0.196	0.196

6.3.2 Optimization Results for Small Noise Level

This section presents the optimization results for the optimization problem defined in Eq. (6.9) for using small noise, i.e., $\sigma_\varepsilon^2 = 0.01$. The MBKG surrogate models used in solving this problem performed sequential sampling to refine the limit state and reduce the uncertainty in the probability of failure. The MBKG surrogate models started out using 25 initial DoE samples and did four sequential sampling iterations so that a total of 105 DoE samples were used to create the MBKG surrogate models. The MBKG surrogate models were then used to generate the posterior distribution of the probability of failure, which was then used as described in Section 6.2 to calculate the constraint information needed for optimization.

The optimization history for the first constraint is shown in Table 6.2, and the optimization history for the second constraint is shown in Table 6.3. Note that for this example the third constraint is not active; thus, it was not fitted when creating the MBKG surrogate models. From Table 6.2 it is seen that optimization converged after six iterations. When studying the optimization history in both tables, it is seen that initially $P_F^{(C.L.)} = 49.8\%$ for constraint 1 and $P_F^{(C.L.)} = 50.3\%$ for constraint 2, both of which are much larger than $P_F^{TAR} = 2.275\%$. By iteration six, however, $P_F^{(C.L.)} = 2.295\%$ for constraint 1 and $P_F^{(C.L.)} = 2.294\%$ for constraint 2, both of which are close to the target probability of failure and were within the optimization tolerances, thus, optimization converged. From the formulation of the optimization problem, the optimum design gives

90% probability, given all available information (i.e., the DoE samples and prior information), that the true probability of failure is less than 2.295% and 2.294% for constraint 1 and constraint 2, respectively. Thus, there is 10% probability that the true probability of failure is larger than 2.295% and 2.294% for constraint 1 and constraint 2, respectively. This is due to limited information being available in this case a limited number of DoE samples being used, i.e., the true probability of failure cannot be calculated using the MBKG surrogate model. As more DoE samples are used, i.e., as more information is used, to create the MBKG surrogate model, the optimum result should converge to the true solution. As seen from both Tables 6.2 and 6.3, the true probability of failure is less than 2.295% and 2.294% for constraint 1 and constraint 2, respectively.

Also note that the mean value of the probability of failure should be close to the true probability of failure (i.e., probability of failure of constraints without noise) if the MBKG surrogate model is accurate. From Table 6.2, it is seen that the mean value is close to the true probability of failure for constraint 1. For constraint 2, from Table 6.3 it is seen that the mean probability of failure value is close to the true probability of failure but is not as close as constraint 1.

Table 6.2 Optimization History for Small Noise Constraint 1

Iteration	d_1	d_2	Cost	True P_{F_1}	Mean P_{F_1}	$P_{F_1}^{(C.L.)}$
1	5.1900	0.7400	–2.2922	48.891%	48.881%	49.814%
2	4.7314	1.3088	–2.0241	15.223%	15.419%	16.285%
3	4.8180	1.5269	–1.9174	5.466%	5.507%	5.933%
4	4.9635	1.5982	–1.8827	2.782%	2.853%	3.162%
5	5.0247	1.6258	–1.8701	2.082%	2.090%	2.318%
6	5.0338	1.6249	–1.8705	2.029%	2.041%	2.295%

Table 6.3 Optimization History for Small Noise Constraint 2

Iteration	d_1	d_2	Cost	True P_{F_2}	Mean P_{F_2}	$P_{F_2}^{(C.L.)}$
1	5.1900	0.7400	–2.2922	48.473%	47.977%	50.329%
2	4.7314	1.3088	–2.0241	1.908%	2.187%	3.119%
3	4.8180	1.5269	–1.9174	0.833%	1.009%	1.538%
4	4.9635	1.5982	–1.8827	1.187%	1.380%	1.972%
5	5.0247	1.6258	–1.8701	1.352%	1.499%	2.134%
6	5.0338	1.6249	–1.8705	1.451%	1.636%	2.294%

To demonstrate how the optimum solution converges to the true solution without noise when more DoE samples are used, i.e., more information is used, optimization was carried out using a different number of DoE samples. A total of five different optimization runs were done using 25, 45, 65, 85, and 105 DoE samples, respectively. Tables 6.4 and 6.5 show the optimization history using the different numbers of DoE samples to fit the MBKG surrogate model. Note that all MBKG surrogate models started with 25 initial DoE samples, and the remaining DoE samples were inserted 20 at a time using the developed sequential sampling method. When looking at the design history in the tables, it is seen that the optimum design appears to be converging to the true optimum design. For this example, the true optimum design means the optimum design for using the constraints with no noise. It is also seen for both constraints that the true probability of failure is less than $P_F^{(C.L.)}$ for all optimization runs. It is seen that when only 25 DoE samples are used the design is very conservative compared to when 105 DoE samples are used.

The next section will present optimization results for the medium and large noise levels as well as compare the results for the small, medium, and large noise levels.

Table 6.4 Optimization History for Different Numbers of DoE
Constraint 1

# DoE	d_1	d_2	Cost	True P_{F_1}	Mean P_{F_1}	$P_{F_1}^{(C.L.)}$
25	3.8538	2.6257	–1.4637	1.278%	1.376%	2.245%
45	4.7637	1.7715	–1.8068	1.909%	1.998%	2.297%
65	4.9260	1.6783	–1.8469	2.008%	1.988%	2.222%
85	4.9982	1.6416	–1.8630	2.020%	2.039%	2.295%
105	5.0338	1.6249	–1.8705	2.029%	2.041%	2.295%
True	5.05	1.59	–1.8860	2.291%	N/A	N/A

Table 6.5 Optimization History for Different Numbers of DoE
Constraint 2

# DoE	d_1	d_2	Cost	True P_{F_2}	Mean P_{F_2}	$P_{F_2}^{(C.L.)}$
25	3.8538	2.6257	–1.4637	0.000%	0.881%	2.266%
45	4.7637	1.7715	–1.8068	0.036%	0.838%	2.177%
65	4.9260	1.6783	–1.8469	0.391%	0.838%	1.746%
85	4.9982	1.6416	–1.8630	0.950%	1.285%	2.102%
105	5.0338	1.6249	–1.8705	1.451%	1.636%	2.294%
True	5.05	1.59	–1.8860	2.279%	N/A	N/A

6.3.3 Comparing Small, Medium, and Large Noise Optimization Results

The same optimization problem was solved using the medium and large noise values in Table 6.1. Tables 6.6 and 6.7 show the optimization history for the medium noise level for constraints 1 and 2, respectively. From the tables it is seen that optimization finished after eight iterations. Optimization stopped because the relative changes in the design variables were less than the tolerance. However, the relative maximum constraint violation at the eighth iteration was $1.46e-3$, which is slightly larger than the tolerance value of $1e-3$. For constraint 1 it is seen that $P_F^{(C.L.)} = 2.228\%$

is actually less than $P_F^{TAR} = 2.275\%$. Constraint 2, on the other hand, is the one with the constraint violation, $P_F^{(C.L.)} = 2.347\%$, which is slightly larger than $P_F^{TAR} = 2.275\%$. Even with the small constraint violation, it is seen that the true probability of failure is less than $P_F^{(C.L.)}$ for both constraints.

It can also be seen from the tables that, while the posterior mean value of the probability of failure does not match well to the true probability of failure, it is close to the true value for both constraints. This is actually why there is a need to use the developed confidence-based RBDO method: because there is uncertainty in what the probability of failure value is. Sometimes the mean value of the probability of failure underestimates the true value, and sometimes it overestimates the true value. However, using the confidence-based RBDO method gives a probability of 90% (90% for this example because $C.L. = 90\%$) that the true probability of failure is less than $P_F^{(C.L.)}$ at the optimum design.

Table 6.6 Optimization History for Medium Noise Constraint 1

Iteration	d_1	d_2	Cost	True P_{F_1}	Mean P_{F_1}	$P_{F_1}^{(C.L.)}$
1	5.1900	0.7400	–2.2922	48.902%	50.058%	51.648%
2	4.7479	1.3157	–2.0200	14.490%	13.686%	14.827%
3	4.8263	1.5291	–1.9163	5.312%	5.660%	6.345%
4	4.9713	1.6067	–1.8788	2.600%	2.612%	3.034%
5	5.0009	1.6272	–1.8694	2.181%	2.315%	2.668%
6	5.0158	1.6293	–1.8685	2.083%	2.051%	2.351%
7	5.0442	1.6208	–1.8723	2.062%	1.963%	2.273%
8	5.0347	1.6253	–1.8703	2.017%	1.961%	2.228%

Table 6.7 Optimization History for Medium Noise Constraint 2

Iteration	d_1	d_2	Cost	True P_{F_2}	Mean P_{F_2}	$P_{F_2}^{(C.L.)}$
1	5.1900	0.7400	–2.2922	48.431%	48.949%	51.465%
2	4.7479	1.3157	–2.0200	2.102%	2.306%	3.375%
3	4.8263	1.5291	–1.9163	0.876%	1.009%	1.607%
4	4.9713	1.6067	–1.8788	1.146%	1.529%	2.224%
5	5.0009	1.6272	–1.8694	1.144%	1.267%	1.861%
6	5.0158	1.6293	–1.8685	1.233%	1.261%	1.801%
7	5.0442	1.6208	–1.8723	1.618%	1.863%	2.643%
8	5.0347	1.6253	–1.8703	1.466%	1.493%	2.347%

Tables 6.8 and 6.9 show the optimization history for the large noise level for constraint 1 and constraint 2, respectively. From the tables it is seen that optimization finished after six iterations. However, similar to the medium noise level case, optimization stopped because the relative changes in the design variables were less than the tolerance. The relative constraint violation at iteration six was $2.64e-3$, which exceeds the tolerance value of $1e-3$. This can be seen in the tables for both constraints, as $P_F^{(C.L.)} = 2.321\%$ and $P_F^{(C.L.)} = 2.408\%$ for constraint 1 and constraint 2 respectively; both exceeded $P_F^{TAR} = 2.275\%$. Even with the constraint violation, the results in the tables make it seem as though optimization was converging towards a design that would satisfy the constraint tolerance if the tolerance for the relative change in design variables was tightened some. However, optimization would have to actually be done with this tighter tolerance to confirm that it does converge to a solution.

It is also interesting to note, that for constraint 1 at the sixth iteration, the true probability of failure, $P_F^{TRUE} = 2.362\%$, is actually larger than $P_F^{(C.L.)} = 2.321\%$. Recall that, by definition, given all information used, there is 90% probability that the true probability of failure is less than $P_F^{(C.L.)}$. This means that there is 10% probability that the true probability of failure is larger than $P_F^{(C.L.)}$. For constraint 1 this is an example of that

10%. This is an example where not having “perfect information” leads to a design that would not be acceptable. There may be two possibilities for why this occurred. The first is that maybe more DoE samples are needed to farther refine the MBKG surrogate model to improve the posterior distributions. From Table 6.8 is seen that the posterior mean of the probability of failure for constraint 1 does not compare well to the true probability of failure. This may indicate that using more DoE samples may help to further improve the MBKG surrogate model and posterior distributions.

The second possibility is that the large noise may be so large that it is “washing out” the true underlying response value completely. If this is true, then the problem cannot be solved using the MBKG surrogate model as formulated in this work. That is one limitation of the developed MBKG surrogate modeling method: if the noise is larger than the overall response variance, the developed method is not applicable. Further investigation using more DoE samples to carry out optimization would need to be done to try to determine if this situation is due to the lack of information, i.e., lack of DoE samples, or if indeed the large noise is too great for this method to be able to solve the problem.

Lastly, it is interesting to compare the optimization results for the three different noise levels, even though optimization did not fully converge for the medium and large noise levels. Tables 6.10 and 6.11 show a comparison of the optimization results for constraints 1 and 2 respectively. From the tables it can be seen that small and medium noise results are fairly similar to each other; they appear to have converged to a similar optimum design. The large noise, is not as similar to the others, but it is relatively close.

This chapter demonstrated that for the small and medium noise levels the developed confidence-based RBDO method was able to solve the problems to come up with a reliable optimum satisfying the desired confidence levels. Further investigation of the large noise is needed to see if using more DoE samples would give a satisfactory solution.

Table 6.8 Optimization History for Large Noise Constraint 1

Iteration	d_1	d_2	Cost	True P_{F_1}	Mean P_{F_1}	$P_{F_1}^{(C.L.)}$
1	5.1900	0.7400	–2.2922	48.884%	48.961%	51.132%
2	4.7189	1.3898	–1.9853	11.644%	11.338%	12.808%
3	4.8339	1.5610	–1.9013	4.547%	4.723%	5.520%
4	5.0297	1.5520	–1.9032	3.057%	2.626%	3.131%
5	5.0283	1.6120	–1.8762	2.202%	1.758%	2.214%
6	4.9888	1.6192	–1.8731	2.362%	1.950%	2.321%

Table 6.9 Optimization History for Large Noise Constraint 2

Iteration	d_1	d_2	Cost	True P_{F_2}	Mean P_{F_2}	$P_{F_2}^{(C.L.)}$
1	5.1900	0.7400	–2.2922	48.412%	49.642%	52.655%
2	4.7189	1.3898	–1.9853	1.105%	1.320%	2.252%
3	4.8339	1.5610	–1.9013	0.677%	0.553%	0.924%
4	5.0297	1.5520	–1.9032	2.803%	2.596%	3.891%
5	5.0283	1.6120	–1.8762	1.579%	1.992%	3.213%
6	4.9888	1.6192	–1.8731	1.141%	1.473%	2.408%

Table 6.10 Comparing Optimization Results for Different Noise Levels Constraint 1

Noise	# Iterations	d_1	d_2	Cost	True P_{F_1}	Mean P_{F_1}	$P_{F_1}^{(C.L.)}$
Small	6	5.0338	1.6249	–1.8705	2.029%	2.041%	2.295%
Medium	8	5.0347	1.6253	–1.8703	2.017%	1.961%	2.228%
Large	6	4.9888	1.6192	–1.8731	2.362%	1.950%	2.321%
No Noise	True	5.05	1.59	–1.8860	2.291%	N/A	N/A

Table 6.11 Comparing Optimization Results for Different Noise Levels Constraint 2

Noise	# Iterations	d_1	d_2	Cost	True P_{F_2}	Mean P_{F_2}	$P_{F_2}^{(C.L.)}$
Small	6	5.0338	1.6249	–1.8705	1.451%	1.636%	2.294%
Medium	8	5.0347	1.6253	–1.8703	1.466%	1.493%	2.347%
Large	6	4.9888	1.6192	–1.8731	1.141%	1.473%	2.408%
No Noise	True	5.05	1.59	–1.8860	2.279%	N/A	N/A

6.4 A 3-D Multibody Dynamics Block-Car Example

This section will present a 3-D multibody dynamics (MBD) block-car example that uses the developed methods to carry out confidence-based RBDO. The problem is a simple example to demonstrate using the method for an actual engineering problem that contains noise in the response. Figure 6.5 shows the multibody dynamics block-car used for this example. The car is modeled as a simple block with four wheels as shown in the figure.

Reliability-based design optimization of this example was attempted previously using standard Kriging methods as the surrogate modeling method. However, the standard Kriging methods failed when trying to create the surrogate model for the contact force due to the noise in the contact force response. The standard Kriging method failed in that the predicted response surface was not a smooth surface. The predicted response surface looked like white noise, i.e., it was not smooth and was jagged. This example was the motivation that led to the research carried out in this work and the development of the methods for handling noisy response problems.

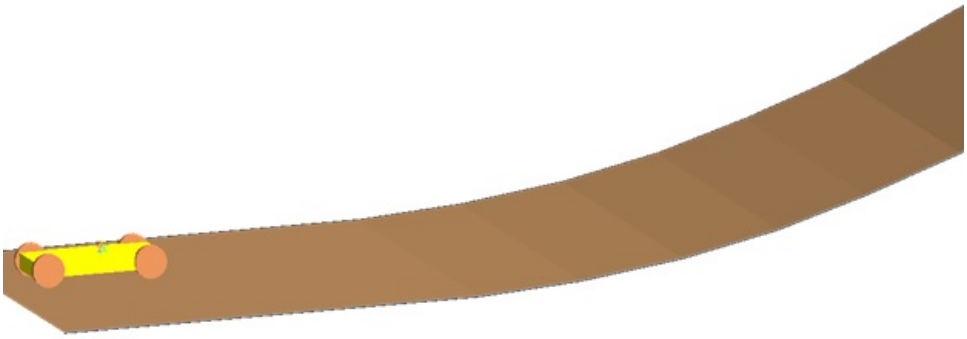

Figure 6.5 Multibody Dynamics Block-Car Example

The objective is to maximize the distance that the car travels up the incline before it loses traction. The three design variables are the mass of the car and the locations of the

center of mass in the x and y directions. The two constraints for this example are the contact force between the front wheels and the ground. The contact force is constrained so that it should not be less than 125 pounds. These constraints are imposed so that the car does not flip over backwards when going up the incline. Contact forces calculated by MBD simulation software are known to be inherently noisy responses. For this example the commercially available MBD software package used was RecurDyn. Deterministic design optimization (DDO) for this problem was carried out first. This was done in order to have a good starting point for confidence-based RBDO. In order to carry out DDO, the MBKG surrogate modeling method developed in this work was used. The MBKG surrogate model was fitted using 25 DoE for each constraint and the objective. After fitting the MBKG surrogate model, the 25 mean values for the responses were then available. The 25 DoE and 25 mean response values were then used with the standard Kriging methods to calculate the sensitivity of the response for the current design point. This was done for the two constraints and the objective function. The mean response values and sensitivity of the responses for both constraints and the objective were then provided to the optimization algorithm.

Table 6.12 shows the design bounds for the design variables and the initial starting design point, where M is the mass of the car, CM_x is the x coordinate location for the center of mass, and CM_y is the y coordinate location for the center of mass, denoted as d_1, d_2, and d_3, respectively. The objective function for optimization is to maximize the distance the car travels up the incline before it loses traction. The two constraints are that the contact force in the front wheels are greater than 125 pounds.

Table 6.12 Design Bounds and Initial Design Point for DDO

Design Variable	Initial	Lower Bound	Upper Bound
$d_1 = M$	55	45	65
$d_2 = CM_x$	0	–60	60
$d_3 = CM_y$	–40	–44	–36

The DDO problem is formulated as shown in Eq. (6.10). The initial design point for DDO is the center of the design domain. Table 6.13 shows the DDO optimization history results. As seen in the table, the distance the car traveled, i.e., the objective function, was maximized during the optimization process, and the constraints became active as the optimization algorithm progressed through the iterations. It is seen in the table that the optimization algorithm converged to a solution on the 11th iteration. The DDO optimum point was used as the initial design point for carrying out confidence-based RBDO.

$$
\begin{aligned}
&\text{maximize} \quad \text{Distance}(\mathbf{d}) \\
&\text{subject to} \quad \begin{cases} h_i(\mathbf{d}) \geq 125, i = 1, 2 \\ \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U, \mathbf{d} \in \mathbb{R}^3 \end{cases} \\
&\text{where} \\
&\text{Distance}(\mathbf{d}) = \text{Distance traveled before slip} \\
&h_1(\mathbf{d}) = \text{Contact Force Right Front Wheel} \\
&h_2(\mathbf{d}) = \text{Contact Force Left Front Wheel} \\
&\mathbf{d}^L = [45, -60, -44]^T, \mathbf{d}^U = [65, 60, -36]^T \\
&\mathbf{d}_{initial} = [55, 0, -40]^T
\end{aligned}
\tag{6.10}
$$

Table 6.13 DDO Optimization History for Block Car

Iteration	d_1	d_2	d_3	Objective	h_1	h_2
1	55.0000	0.0000	–40.0000	2381.773	139.414	139.069
2	54.9302	–3.9493	–38.9974	2399.437	132.918	132.733
3	56.9627	–20.8840	–36.0000	2507.059	120.636	120.289
4	62.0902	–23.8850	–36.6840	2521.744	127.108	128.648
5	65.0000	–28.6530	–39.1904	2550.072	130.828	129.992
6	65.0000	–33.6832	–43.5372	2587.930	125.868	125.280
7	65.0000	–34.2504	–44.0000	2593.061	125.775	124.736
8	65.0000	–34.2465	–43.6772	2594.082	125.207	124.516
9	65.0000	–33.8138	–43.8212	2588.483	125.608	125.984
10	65.0000	–33.8853	–43.7777	2589.537	125.797	125.087
11	65.0000	–33.9238	–43.7830	2589.827	125.313	125.218

Table 6.14 shows the design bounds for the design variables and the initial starting design point, where M is the mass of the car, CM_x is the x coordinate location for the center of mass, and CM_y is the y coordinate location for the center of mass. These design variables are denoted as d_1, d_2, and d_3, respectively, for confidence-based RBDO. The objective function for optimization is to maximize the distance the car travels up the incline before it loses traction. The two constraints are that the contact force in the front wheels are greater than 125 pounds.

Table 6.14 Design Bounds and Initial Design Point for Confidence-Based RBDO

Design Variable	Initial	Lower Bound	Upper Bound
$d_1 = M$	65	45	65
$d_2 = CM_x$	–33.92	–60	60
$d_3 = CM_y$	–43.78	–44	–36

The confidence-based RBDO problem is formulated as shown in Eq. (6.11). The initial design point for confidence-based RBDO is the DDO solution. When fitting the MBKG surrogate models 25 initial DoE samples were used. After fitting the MBKG surrogate models two sequential sampling iterations were performed using the developed sequential sampling method. For each sequential sampling iteration 25 DoE samples were added. Thus, giving a total of 75 DoE samples that were used to fit the MBKG surrogate models. The MBKG surrogate models were then used to generate the posterior distribution of the probability of failure that was used to carry out confidence-based RBDO.

Tables 6.15 and 6.16 show the optimization history for constraints 1 and 2. As seen in the tables, at the initial design $P_F^{(C.L.)}$ is 44.270% and 62.343% for constraint 1 and 2, respectively; both are much larger than the target 2.275%. It is seen that, in the first three iterations of optimization, $P_F^{(C.L.)}$ decreases quickly, and then for the remaining iterations it decreases at a slower rate for both constraints. The optimization algorithm converged to a solution after nine iterations, the optimization tolerances were satisfied, and there is no constraint violation at the optimum solution. As seen in Table 6.15, the objective, i.e., the maximum distance the car travels before losing traction, is decreased as the constraints are satisfied. At the optimum design there is 10% probability that the true probability of failure is larger than 0.439% and 1.751% for constraints 1 and 2 as seen in Tables 6.15 and 6.16, respectively. These values clearly satisfy the target value of 2.275%. It is interesting to note how the optimization algorithm did not change the mass of the car from 65, which was the mass value from the DDO solution. This is believed to be because the higher the mass the more traction the car will have to travel farther up the incline. Thus, the confidence-based RBDO method was successful in finding a solution that met both the target probability of failure and target confidence level.

$$
\begin{aligned}
&\text{maximize} \quad \text{Distance}(\mathbf{d}) \\
&\text{subject to} \quad \begin{cases} P_{F_i}^{(C.L.)} \le P_{F_i}^{TAR}, i = 1, 2 \\ \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \mathbf{d} \in \mathbb{R}^3 \text{ and } \mathbf{X} \in \mathbb{R}^3 \end{cases} \\
&\text{where} \\
&\text{Distance}(\mathbf{d}) = \text{Distance traveled before slip} \\
&h_1(\mathbf{X}) = \text{Contact Force Right Front Wheel} \\
&h_2(\mathbf{X}) = \text{Contact Force Left Front Wheel} \\
&\mathbf{d}^L = [45, -60, -44]^T, \mathbf{d}^U = [65, 60, -36]^T \\
&\mathbf{d}_{initial} = [65, -33.92, -43.78]^T \\
&X_1 \sim N\left(d_1, 0.5^2\right) \\
&X_2 \sim N\left(d_2, 0.4^2\right) \\
&X_3 \sim N\left(d_3, 0.2^2\right) \\
&\mathrm{P}_{F_i} = P\left(h_i(\mathbf{X}) < 125\right), i = 1, 2 \\
&P_{F_i}^{Tar} = 2.275\%, i = 1, 2 \text{ and } C.L._i = 90\%, i = 1, 2
\end{aligned}
\tag{6.11}
$$

Table 6.15 Confidence-Based RBDO Optimization History for Block-Car Constraint 1

Iteration	d_1	d_2	d_3	Objective	Mean P_{F_1}	$P_{F_1}^{(C.L.)}$
1	65	–33.9238	–43.7830	2589.806	38.657%	44.270%
2	65	–32.2557	–42.3679	2575.278	4.508%	5.878%
3	65	–31.7703	–41.3944	2572.964	2.061%	2.547%
4	65	–30.7138	–37.8149	2573.467	1.852%	2.573%
5	65	–30.8124	–38.8367	2571.488	1.566%	2.240%
6	65	–30.1286	–37.7002	2567.938	0.327%	0.603%
7	65	–30.3040	–37.5318	2569.594	0.412%	0.622%
8	65	–29.9892	–36.7301	2568.900	0.521%	0.746%
9	65	–29.8552	–36.3724	2568.612	0.307%	0.439%

Table 6.16 Confidence-Based RBDO Optimization History for Block-Car Constraint 2

Iteration	d_1	d_2	d_3	Objective	Mean P_{F_2}	$P_{F_2}^{(C.L.)}$
1	65	–33.9238	–43.7830	2589.806	56.693%	62.343%
2	65	–32.2557	–42.3679	2575.278	14.801%	18.107%
3	65	–31.7703	–41.3944	2572.964	7.821%	9.917%
4	65	–30.7138	–37.8149	2573.467	6.281%	8.765%
5	65	–30.8124	–38.8367	2571.488	4.933%	6.709%
6	65	–30.1286	–37.7002	2567.938	1.119%	1.682%
7	65	–30.3040	–37.5318	2569.594	2.189%	3.220%
8	65	–29.9892	–36.7301	2568.900	1.744%	2.625%
9	65	–29.8552	–36.3724	2568.612	0.990%	1.751%

CHAPTER 7

CONCLUSION, CURRENT RESEARCH, AND FUTURE RESEARCH

7.1 Conclusion

Numerous sensitivity-based reliability-based design optimization (RBDO) methods have been developed and applied to various engineering problems. However, it is common that the sensitivity is difficult or even impossible to calculate for highly nonlinear and coupled fluid structure interaction problems, e.g., crash and blast problems. Sampling-based RBDO methods that use surrogate models to approximate the simulation models and then carry out Monte Carlo simulation using the surrogate model to perform the reliability analysis have been developed. Current sampling-based RBDO applications have been used on problems in which the simulation models do not contain noise or it is assumed that there is no noise. However, it has been found that some simulation models contain noise in the responses, which makes carrying out RBDO difficult. This has brought about the need for a surrogate modeling method that can handle responses that contain noise. There is also a need for a surrogate modeling method that can provide a way to not only carry out RBDO but to carry out confidence-based RBDO to ensure a conservative reliable optimum design.

A modified Bayesian Kriging (MBKG) surrogate modeling method was developed for handling problems whose responses contain noise. The prior distributions to be used for fitting the MBKG surrogate model have been determined, and the full conditional distributions derived. The full conditional distributions were coded into a Gibbs sampling algorithm in order to use Markov chain Monte Carlo (MCMC) to fit the MBKG surrogate model. All the coding was done in MATLAB so that the methods developed in this study can be easily integrated with previously developed RBDO methods. It was shown that MBKG surrogate model does work for handling problems whose responses contain noise for which the standard Kriging approaches fail to work.

A sequential sampling method that uses the posterior credible sets of the MBKG surrogate model has been developed. The new method was tested with an example problem, and it was demonstrated how the uncertainty of the predicted limit state is reduced and the predicted limit state converges to the true limit state as more design of experiment (DoE) samples are used. It was also demonstrated that the posterior distribution of the probability of failure converges as more DoE samples are used.

A confidence-based RBDO method using the posterior distribution of the probability of failure was developed. The method was demonstrated using a mathematical example with different amounts of noise. It was shown that the method converged to a solution and was able to find an optimum reliable design that meets the desired confidence levels. It was shown that, for small and medium noise, the method found two solutions that are very similar, i.e., almost the same design point. For the large noise, the optimization did converge for the relative change in the design variable tolerance; however, the constraint tolerance was still violated. The optimization history did show that it was converging to a solution that would be similar to that of the small and medium noise.

Overall, an MBKG surrogate modeling method, a sequential sampling method for reducing the uncertainty in the posterior distributions, and a confidence-based RBDO method were successfully developed and demonstrated using a mathematical example. The developed methods were also successfully demonstrated using a 3-D multibody dynamics engineering example.

7.2 Future Research

Running the MCMC chains to fit the models tends to be the most computationally intensive part. Further investigation is needed so that the Metropolis-Hastings algorithm can be finely tuned; this would greatly improve the computational time for fitting the

MCMC chains. There are also some known bottle-necks in some of the functions used in fitting the MCMC chains, they can be rewritten to improve the computational time. A study on the use of different priors for different noise levels would also help with fine-tuning the model in regards to improving the computational efficiency.

Additional future research would be investigating how a dynamic Bayesian Kriging method can be developed in order to dynamically select the best mean structure and the correlation function that best fits the data for the problem being solved. The deviance information criterion (DIC) is one possible method that can be used to compare Bayesian regression models and select the one that best fits the data. There are also methods that use penalized loss functions to compare regression models. Research to study these methods and how they could be applied the developed MBKG surrogate modeling method is needed to determine if they can be used to develop a dynamic Bayesian Kriging method.

It has also been recognized that the use of the posterior distribution of the probability of failure from the Bayesian analysis provides for a convenient and natural way to carry out confidence-based RBDO. The same concept could be applied to a different Bayesian Kriging model for noise-free problems. Using the noise-free Bayesian Kriging model, confidence-based RBDO could be carried out using the same method as presented in Chapter 6.

REFERENCES

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis, 22*(5), 717-727. doi:10.1016/S0731-7085(99)00272-1

Almeida, J. S. (2002). Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology, 13*(1), 72-76. doi:10.1016/S0958-1669(02)00288-4

An, D., & Choi, J. H. (2012). Efficient reliability analysis based on bayesian framework under input variable and metamodel uncertainties. *Structural and Multidisciplinary Optimization, 46*(4), 533-547. doi:10.1007/s00158-012-0776-6

Bayes, T., & Price, R. (1763). An essay towards solving a problem in the doctrine of chance. by the late rev. mr. bayes, communicated by mr. price, in a letter to john canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London, 53*(0), 370-418.

Beers, W. C. M. v., & Kleijnen, J. P. C. (2003). Kriging for interpolation in random simulation. *The Journal of the Operational Research Society, 54*(3), 255-262.

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surface*. New York : Wiley.

Breitkopf, P., Naceur, H., Rassineux, A., & Villon, P. (2005). Moving least squares response surface approximation: Formulation and metal forming applications. *Computers & Structures, 83*(17–18), 1411-1428. doi:http://dx.doi.org/10.1016/j.compstruc.2004.07.011

Breitung, K. (1984). Asymptotic approximations for multinormal integrals. *Journal of Engineering Mechanics, 110*(3), 357-366. doi:10.1061/(ASCE)0733-9399(1984)110:3(357)

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics,7*(4), 434-455.

Buhmann, M. D. (2003). *Radial basis functions : Theory and implementations / M.D. buhmann*. Cambridge, U.K. ; New York, NY : Cambridge University Press.

Burges, C. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2*(2), 121-167. doi:10.1023/A:1009715923555

Choi, K. K., Tu, J., & Park, Y. H. (2001). Extensions of design potential concept for reliability-based design optimization to nonsmooth and extreme cases.*Structural and Multidisciplinary Optimization, 22*(5), 335-350. doi:10.1007/s00158-001-0154-2

Cowles, M. K. (2013). *Applied bayesian statistics [electronic resource] : With R and OpenBUGS examples*. New York, NY : Springer New York : Imprint: Springer.

Cowles, M. K., & Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association,91*(434), 883-904.

Cressie, N. A. C. (1991). *Statistics for spatial data*. New York : J. Wiley.

Currin, C., Mitchell, T., Morris, M., & Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association, 86*(416), 953-963.

Dyn, N., Levin, D., & Rippa, S. (1986). Numerical procedures for surface fitting of scattered data by radial functions. *SIAM Journal on Scientific and Statistical Computing, 7*(2), 639-659. doi:10.1137/0907043

Fang, H., Rais-Rohani, M., Liu, Z., & Horstemeyer, M. F. (2005). A comparative study of metamodeling methods for multiobjective crashworthiness optimization. *Computers & Structures, 83*(25–26), 2121-2136. doi:10.1016/j.compstruc.2005.02.025

Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed.). New York: Wiley.

Fonseca, D. J., Navaresse, D. O., & Moynihan, G. P. (2003). Simulation metamodeling through artificial neural networks. *Engineering Applications of Artificial Intelligence, 16*(3), 177-183. doi:10.1016/S0952-1976(03)00043-5

Forrester, A. I. J., Sóbester, A., & Keane, A. J. (2008). *Engineering design via surrogate modelling : A practical guide*. Chichester, U.K. : J. Wiley.

Forrester, A. I. J., & Keane, A. J. (2009). Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences, 45*(1–3), 50-79. doi:10.1016/j.paerosci.2008.11.001

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics, 19*(1), 1-67.

Friedman, J. H., & Roosen, C. B. (1995). An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research, 4*(3), 197-217. doi:10.1177/096228029500400303

Gallant, S. I. (1993). *Neural network learning and expert systems / stephen I. gallant*. Cambridge, Mass. : MIT Press.

Galushkin, A. I. (2007). *Neural networks theory / alexander I. galushkin*. Berlin ; New York : Springer.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*(410), 398-409. doi:10.1080/01621459.1990.10476213

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. Boca Raton, Fla. : Chapman & Hall/CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457-472.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1998). *Markov chain monte carlo in practice*. Boca Raton: Boca Raton : Chapman & Hall/CRC.

Haldar, A., & Mahadevan, S. (2000). *Probability, reliability, and statistical methods in engineering design*. New York : J. Wiley.

Hamada, M. S., Wilson, A. G., Reese, C. S., & Martz, H. F. (2008). *Bayesian reliability [electronic resource]*. New York, NY : Springer.

Hasofer, A. M., & Lind, N. C. (1974). Exact and invariant second-moment code format. *Journal of Engineering Mechanics Division, 100*(1), 111-121.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika, 57*(1), 97-109.

Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *Intelligent Systems and their Applications, IEEE, 13*(4), 18-28. doi:10.1109/5254.708428

Hogg, R. V., McKean, J. W., & Craig, A. T. (2005). *Introduction to mathematical statistics*. Upper Saddle River, N.J. : Pearson Education.

Hohenbichler, M., Gollwitzer, S., Kruse, W., & Rackwitz, R. (1987). New light on first- and second-order reliability methods. *Structural Safety, 4*(4), 267-284. doi:http://dx.doi.org/10.1016/0167-4730(87)90002-6

Hohenbichler, M., & Rackwitz, R. (1988). Improvement of Second-Order reliability estimates by importance sampling. *Journal of Engineering Mechanics,114*(12), 2195-2199. doi:10.1061/(ASCE)0733-9399(1988)114:12(2195)

Hu, C., & Youn, B. D. (2011). Adaptive-sparse polynomial chaos expansion for reliability analysis and design of complex engineering systems. *Structural and Multidisciplinary Optimization, 43*(3), 419-442. doi:10.1007/s00158-010-0568-9

Isukapalli, S. S. (1999). *Uncertainty analysis of transport-transformation models.* (Ph.D., Rutgers The State University of New Jersey - New Brunswick).*ProQuest Dissertations and Theses,* . (304542746).

Jin, R., Chen, W., & Simpson, T. W. (2001). Comparative studies of metamodelling techniques under multiple modelling criteria. *Structural and Multidisciplinary Optimization, 23*(1), 1-13. doi:10.1007/s00158-001-0160-4

Kass, R. E., Carlin, B. P., Gelman, A., & Neal, R. M. (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *The American Statistician, 52*(2), 93-100.

Kewlani, G., & Iagnemma, K. (2008). A stochastic response surface approach to statistical prediction of mobile robot mobility. *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on,* 2234-2239. doi:10.1109/IROS.2008.4651187

Lancaster, P., & Salkauskas, K. (1981). Surfaces generated by moving least squares methods. *Mathematics of Computation, 37*(155), 141-158.

Lee, I. (2008). *Reliability-based design optimization and robust design optimization using univariate dimension reduction method.* (Ph.D., The University of Iowa).*ProQuest Dissertations and Theses,* . (304609574).

Lee, I., Choi, K. K., Du, L., & Gorsich, D. (2008). Inverse analysis method using MPP-based dimension reduction for reliability-based design optimization of nonlinear and multi-dimensional systems. *Computer Methods in Applied Mechanics and Engineering, 198*(1), 14-27. doi:http://dx.doi.org/10.1016/j.cma.2008.03.004

Lee, I., Choi, K. K., Du, L., & Gorsich, D. (2008). Dimension reduction method for reliability-based robust design optimization. *Computers & Structures, 86*(13–14), 1550-1562. doi:http://dx.doi.org/10.1016/j.compstruc.2007.05.020

Lee, I., Choi, K. K., & Gorsich, D. (2010). System reliability-based design optimization using the MPP-based dimension reduction method. *Structural and Multidisciplinary Optimization, 41*(6), 823-839. doi:10.1007/s00158-009-0459-0

Lee, I., Gorsich, D., Choi, K. K., Noh, Y., & Zhao, L. (2011). Sampling-based stochastic sensitivity analysis using score functions for RBDO problems with correlated random variables. *Journal of Mechanical Design, 133*(2), 021003-021003. doi:10.1115/1.4003186

Levin, D. (1998). The approximation power of moving least-squares. *Mathematics of Computation, 67*(224), 1517-1531.

Lewis, P. A. W., & Stevens, J. G. (1991). Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). *Journal of the American Statistical Association, 86*(416), 864-877. doi:10.1080/01621459.1991.10475126

Liu, Y., & Fang, Y. (2009). Design of the nonlinear system predictor driven by the bayesian-gaussian neural network of sliding window data. *Computer and Information Science, 2*(2), 26.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine, 28*(25), 3049-3067. doi:10.1002/sim.3680

Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - A bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*(4), 325-337. doi:10.1023/A:1008929526011

Madsen, H. O., Krenk, S., & Lind, N. C. (1985). *Methods of structural safety*. Englewood Cliffs, NJ : Prentice-Hall.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics, 21*(6), 1087-1092. doi:http://dx.doi.org/10.1063/1.1699114

Mullur, A., & Messac, A. (2006). Metamodeling using extended radial basis functions: A comparative approach. *Engineering with Computers, 21*(3), 203-217. doi:10.1007/s00366-005-0005-7

Noh, Y., Choi, K. K., & Du, L. (2009). Reliability-based design optimization of problems with correlated input variables using a gaussian copula. *Structural and Multidisciplinary Optimization, 38*(1), 1-16. doi:10.1007/s00158-008-0277-9

Noh, Y., Choi, K. K., & Lee, I. (2010). Identification of marginal and joint CDFs using bayesian method for RBDO. *Structural and Multidisciplinary Optimization,40*(1-6), 35-51. doi:10.1007/s00158-009-0385-1

Omre, H., & Halvorsen, K. (1989). The bayesian bridge between simple and universal kriging. *Mathematical Geology, 21*(7), 767-786. doi:10.1007/BF00893321

Park, J., & Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural Computation, 3*(2), 246-257. doi:10.1162/neco.1991.3.2.246

Rahman, S., & Wei, D. (2008). Design sensitivity and reliability-based structural optimization by univariate decomposition. *Structural and Multidisciplinary Optimization, 35*(3), 245-261. doi:10.1007/s00158-007-0133-3

Rahman, S., & Wei, D. (2006). A univariate approximation at most probable point for higher-order reliability analysis. *International Journal of Solids and Structures, 43*(9), 2820-2839. doi:http://dx.doi.org/10.1016/j.ijsolstr.2005.05.053

Rahman, S. (2009). Stochastic sensitivity analysis by dimensional decomposition and score functions. *Probabilistic Engineering Mechanics, 24*(3), 278-287. doi:http://dx.doi.org/10.1016/j.probengmech.2008.07.004

Rajashekhar, M. R., & Ellingwood, B. R. (1993). A new look at the response surface approach for reliability analysis. *Structural Safety, 12*(3), 205-220. doi:http://dx.doi.org/10.1016/0167-4730(93)90003-J

Romero, D. A. (2008). *A multi-stage, multi-response bayesian methodology for surrogate modeling in engineering design.* (Ph.D., Carnegie Mellon University).*ProQuest Dissertations and Theses,* . (304667685).

Romero, D. A., Amon, C., & Finger, S. (2003). Computers and information in engineering. *Modeling Time-Dependent Systems using Multi-Stage Bayesian Surrogates,* Washington, DC, USA. 47-47-57. doi:10.1115/IMECE2003-55049

Romero, D. A., Amon, C. H., & Finger, S. (2012). Multiresponse metamodeling in simulation-based design applications. *Journal of Mechanical Design, 134*(9), 091001-091001. doi:10.1115/1.4006996

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics, 23*(3), 470-472.

Rubinstein, R. Y., & Shapiro, A. (1993). In Shapiro A. (Ed.), Discrete event systems : Sensitivity analysis and stochastic optimization by the score function method. Chichester England] ; New York: Chichester England ; New York : Wiley.

Sacks, J., Welch, W. J., Toby J. Mitchell, & Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science, 4*(4), 409-423.

Sakata, S., Ashida, F., & Zako, M. (2007). On applying kriging-based approximate optimization to inaccurate data. *Computer Methods in Applied Mechanics and Engineering, 196*(13–16), 2055-2069. doi:http://dx.doi.org/10.1016/j.cma.2006.11.004

Sakata, S., Ashida, F., & Zako, M. (2010). Comparative study on gradient and hessian estimation using the kriging method and neural network approximation. *Mathematical and Computer Modelling, 51*(3–4), 309-319. doi:http://dx.doi.org/10.1016/j.mcm.2009.08.016

Sakata, S., Ashida, F., & Zako, M. (2008). Approximate structural optimization using kriging method and digital modeling technique considering noise in sampling data. *Computers & Structures, 86*(13–14), 1477-1485. doi:http://dx.doi.org/10.1016/j.compstruc.2007.05.007

Shi, L., Yang, R. J., & Zhu, P. (2012). A method for selecting surrogate models in crashworthiness optimization. *Structural and Multidisciplinary Optimization,46*(2), 159-170. doi:10.1007/s00158-012-0760-1

Simpson, T. W., Poplinski, J. D., Koch, P. N., & Allen, J. K. (2001). Metamodels for computer-based engineering design: Survey and recommendations.*Engineering with Computers, 17*(2), 129-150. doi:10.1007/PL00007198

Song, H., Choi, K. K., Lee, I., Zhao, L., & Lamb, D. (2011). Adaptive virtual support vector machine for the reliability analysis of high-dimensional problems. Washington, DC, USA. *, 5*

Tierney, L. (1998). A note on metropolis-hastings kernels for general state spaces. *The Annals of Applied Probability, 8*(1), 1-9.

Tu, J., Choi, K. K., & Park, Y. H. (1999). A new study on reliability-based design optimization. *Journal of Mechanical Design, 121*(4), 557-564. doi:10.1115/1.2829499

Tu, J., Choi, K. K., & Park, Y. H. (2001). Design potential method for robust system parameter design. *AIAA Journal, 39*(4), 667-677. doi:10.2514/2.1360

Ukrainec, A., Haykin, S., & McGregor, J. (1989). A neural network nonlinear predictor. *Neural Networks, 1989. IJCNN., International Joint Conference on*, 622 vol.2.

van der Merwe, R., Leen, T. K., Lu, Z., Frolov, S., & Baptista, A. M. (2007). Fast neural network surrogates for very high dimensional physics-based models in computational oceanography. *Neural Networks, 20*(4), 462-478. doi:10.1016/j.neunet.2007.04.023

Wang, G. G., & Shan, S. (2007). Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical Design, 129*(4), 370-380.

Wei, D. L., Cui, Z. S., & Chen, J. (2008). Uncertainty quantification using polynomial chaos expansion with points of monomial cubature rules. *Computers & Structures, 86*(23–24), 2102-2108. doi:http://dx.doi.org/10.1016/j.compstruc.2008.07.001

Wei, D. (2006). *A univariate decomposition method for higher-order reliability analysis and design optimization.* (Ph.D., The University of Iowa). *ProQuest Dissertations and Theses,* . (305309521).

Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics, 60*(4), 897-936.

Wu, Y. -. (1994). Computational methods for efficient structural reliability and reliability sensitivity analysis. *AIAA Journal, 32*(8), 1717-1723. doi:10.2514/3.12164

Wu, Y. -., MILLWATER, H. R., & CRUSE, T. A. (1990). Advanced probabilistic structural analysis method for implicit performance functions. *AIAA Journal, 28*(9), 1663-1669. doi:10.2514/3.25266

Xu, H., & Rahman, S. (2003). A moment-based stochastic method for response moment and reliability analysis. *Proceedings of 2nd MIT Conference on Computational Fluid and Solid Mechanics,* Cambridge, MA.

Xu, H., & Rahman, S. (2004). A generalized dimension-reduction method for multidimensional integration in stochastic mechanics. *International Journal for Numerical Methods in Engineering, 61*(12), 1992-2019. doi:10.1002/nme.1135

Youn, B. D., & Choi, K. K. (2003). An investigation of nonlinearity of reliability-based design optimization approaches. *Journal of Mechanical Design, 126*(3), 403-411. doi:10.1115/1.1701880

Youn, B. D., Choi, K. K., & Du, L. (2005). Enriched performance measure approach for reliability-based design optimization. *AIAA Journal, 43*(4), 874-884. doi:10.2514/1.6648

Youn, B. D., Park, Y. H., & Choi, K. K. (2003). Hybrid analysis method for reliability-based design optimization. *Journal of Mechanical Design, 125*(2), 221-232. doi:10.1115/1.1561042

Zhao, L. (2011). *Reliability-based design optimization using surrogate model with assessment of confidence level.* (Ph.D., The University of Iowa). *ProQuest Dissertations and Theses,* . (894769536).

Zhao, L., Choi, K. K., & Lee, I. (2011). Metamodeling method using dynamic kriging for design optimization. *AIAA Journal, 49*(9), 2034-2046. doi:10.2514/1.J051017

Zobel, C. W., & Keeling, K. B. (2008). Neural network-based simulation metamodels for predicting probability distributions. *Computers & Industrial Engineering,54*(4), 879-888. doi:10.1016/j.cie.2007.08.012