January 2014

UNCERTAINTY QUANTIFICATION IN SCIENTIFIC MODELS

Xin Qi
Purdue University

Graduate School ETD Form 9
(Revised 01/14)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Xin Qi

Entitled
UNCERTAINTY QUANTIFICATION IN SCIENTIFIC MODELS

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Dongbin Xiu

Suchuan Dong

Greg Buzzard

Jianlin Xia

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Dongbin Xiu

Approved by Major Professor(s): _____

_____

Approved by: David Goldberg 08/25/2014

Head of the Department Graduate Program Date

UNCERTAINTY QUANTIFICATION IN SCIENTIFIC MODELS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Xin Qi

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2014

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

In the past five years, I have always been proud of being a Boilermaker at Purdue University. I have been grateful to work with so many intelligent, knowledgeable and hard-working people. I would like to acknowledge a few people who have made big impact on my life and study.

First of all, I would like to express my utmost gratitude to my advisor, Prof. Dr. Dongbin Xiu. It is always enjoyable to talk with Dongbin, since he not only guides me in my research, but also serves as a role model in my research path. His passion and creativity leads our research to today's success and motivates me to be a better scientist.

I would like to thank our group members Dr. Jing Li and Xiaoxiao Chen at Purdue. They shared their research experience and interest with me. These discussions motivate me in my research.

I wish to acknowledge the help provided by Dr. Xueyu Zhu for introducing me new research topics that broaden my insights in mathematics research.

I would like to express my deep gratitude to my committee members: Prof. Suchuan Dong, Prof. Greg Buzzard and Prof. Jianlin Xia, for their time, helpful comments and insightful questions. Their advice has guided me well to greatly improve the quality of my thesis and research.

Lastly I would like to thank my parents, Shuqin Liu and Baohe Qi for all their love and encouragement. Without their support, I could not complete my PhD research.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Qi, Xin Ph.D., Purdue University, December 2014. Uncertainty Quantification in Scientific Models . Major Professor: Dongbin Xiu.

Uncertainties widely exist in physical, finance, and many other areas. Some uncertainties are determined by the nature of the research subject, such as random variable and stochastic process. However, in many problems uncertainty is a result of lack of knowledge and may not be modeled as random variables/processes because of the lack of probability information. This is often referred to as epistemic uncertainty, and the traditional probabilistic approaches cannot be readily employed. First two parts of this work study epistemic uncertainties in the forward problems. A method to compute upper and lower bounds for the quantity of interest of problems whose uncertain inputs are of epistemic type is presented. Relative entropy is an important measure to study the distance between multiple probabilities. Its properties have motivated many important existing inequalities for quantifying epistemic uncertainties. Based on these works, we extend the inequalities to a large family of functions, the integrable functions, which play an important role in engineering and research. To be more specific, we provide upper and lower bounds for the statistics such as statistical moments of the quantities of our interest under the existence of epistemic uncertainty. We present the theoretical derivation of the bounds, along with numerical examples to illustrate their computations. Based on derived analytical lower and upper bounds, a procedure to compute numerical bounds of when the underlying system is subject to epistemic uncertainty is discussed. In particular, we consider the case where the uncertain inputs to the system take the form of parameters, physical and/or hyper parameters, and with unknown probability distributions. Our goal is to compute the lower and upper bounds of the statistical moments of quantity-of-interest of the sys-

tem response. We discuss exclusively the numerical algorithms for computing such bounds. More importantly, we established the properties of such numerical bounds and analyzed their accuracy compared to the analytical bounds.

Besides the uncertainties in forward problems, quantification of uncertainties in inverse problems is also discussed: Bayesian posterior estimation and model discrepancy. After a posterior is well studied for a selected prior, we proposed a method of estimating the posterior when a new prior is selected. . The method is based on the initial choice of prior distribution and a surrogate of the forward model under the initial prior. If a new prior is selected, instead of another complete circle for computing posterior distribution, we first study the relation between the two priors and then approximate the new forward model using the previous forward model surrogate. We also present an error analysis for the difference between our estimate and the true posterior distribution. This efficient numerical strategy is based on stochastic collocation methods and generalized polynomial chaos (gPC) to construct a polynomial approximation of the forward solution over the support of the initial prior distribution. The gPC strategy is also applied in model discrepancy to bring in more structure to the forward model. The problem is solved via an optimization procedure. The gPC based algorithms not only reduce computational cost but also bring more useful structure to the forward model.

1. INTRODUCTION

1.1 Overview

Uncertainty quantification (UQ) is the science to determine the effect of uncertainties in scientific models. The importance of understanding and modeling uncertainty has affected many disciplines. A large number of different numerical methods and techniques have been developed to quantify the propagation of uncertainty in scientific models. The methodologies we applied to study uncertainty quantification include Bayesian Inverse Analytics, Generalized Polynomial Chaos,Stochastic Collocation, Monte Carlo simulation and etc.

1.2 Uncertainty Quantification

Mathematical models and simulations are widely used for decision making in manufacturing and scientific research such as engineering, physics, biology, chemistry and environmental sciences. Uncertainties often exist in these models due to low volume of data or poor data quality such as missing/wrong data. Quantifying or reducing these uncertainties can help us well in improving these scientific models. For example, if we can quantify the uncertainties in the models, we can reduce the sample size in experimental design. It is also of great value in risk management and decision making.

Model uncertainty has two forms: aleatory and epistemic. Aleatory uncertainty is the randomness in a process and inevitable. Its physical existence is determined by the nature of the process. On the other hand, epistemic uncertainty is a result of lack of knowledge, data, information, so it can be reduced or avoided by collecting more data, introducing more structures and other advanced methods.Usually we can

study and quantify the aleatory uncertainties under a probabilistic framework. The randomness for discrete variables is parametrized by the probability of each possible value, while for continuous variable, the randomness is parametrized by the probability density function. Another important form of aleatory uncertainties is stochastic process which is often used to describe a time-dependent series of random events. One typical existence of aleatory uncertainty is in the coefficients of system when we know their true distribution. Most recent research has focused on aleatory uncertainty. Numerical methods based on probabilistic information including stochastic Galerkin, generalize Polynomial Chaos, KarhunenLoeve expansion, stochastic collocation are well developed to provide efficient quantifications of the aleatory uncertainties. However, in a lot of cases, we may not have enough information of the distribution of the uncertainties. For example, for many complex systems, they are not accessible or the simulation cost is extremely high. And without enough samples or further information, it's not possible to establish a legitimate distribution of the uncertainty. One way to solve this problem is to introduce more structures to the model so that we can "parametrize" the uncertainty and then reduce it. For example, if we have no idea of coefficient distribution in a system and we simplify the problem by assuming the uncertainty follows a normal distribution with unknown mean and variance, then we reduce the infinite dimensional problem to a two dimensional problem which greatly reduces the uncertainty and unpredictability in the system. However, these imposed assumptions may bring in great bias to the prediction. Thus how to add reasonable structure and parameterization is a long-term and challenging research direction. Another way is to provide a lower and upper bounds of quantities to our interest in the form of inequalities. The basic idea of this approach is to introduce an appropriate baseline probability, "nominal probability", which often comes from research experience and current observations. By relating the "nominal probability" and the true probability of the uncertainties, the quantities of our interest which are subject to epistemic uncertainty are studied under the "nominal" probability as an aleatory uncertainty. A confidence is provided as a penalty for the simplification.

However, to establish a helpful nominal distribution often requires some prior information of the uncertainty, otherwise the bounds may not be meaningful. Many efforts have been taken to establish tighter and tighter bounds.

1.2.1 Generalized Polynomial Chaos

Generalized Polynomial Chaos (gPC) is a very promising research method for uncertainty quantification in that it can provide an efficient surrogate for computational expensive system. With generalized polynomial chaos methods, we consider a polynomial approximation to one or more unknowns in the mathematical models as a finite series expansion. With the specific probability distributions, we will compute the coefficients for each orthogonal polynomial basis. The polynomial chaos was originally developed to expand homogeneous chaos that depends on white noise using Hermite polynomials. This scheme was generalized to various continuous and discrete distributions using corresponding orthogonal polynomials, including Hermite polynomials (with Gaussian distribution), Laguerre polynomials (with Gamma distribution), Jacobi polynomials (with Beta distribution), Legendre polynomials (with Uniform distribution)and etc. For aleatory uncertainty quantification, when probabilistic information is available, the optimal orthogonal polynomial basis is selected based on the distribution function. Thus, the expansion of a model is transferred to determining the coefficients.

1.2.2 Bayesian Analytics for Uncertainty Quantification

Uncertainties in model parameters drive uncertainties in prediction, thus it is vital to quantify uncertainties in a computational model and its propagation. The Bayesian statistical approach provides a natural framework to quantify uncertainties in a system when data is noisy or incomplete. The uncertainties in model inputs are treated as random variables and the prior distribution is often proposed by "experts". Conditioned on observation data, posterior distribution of random inputs

represents the updated knowledge about the uncertainties in the parameters. Efficient characterization of posterior such as simulation from this distribution, moments and other statistics plays an important role in the uncertainty quantification for the entire model.

1.3 Research Outline

The primary method of this work is presented in Chapter 2 where optimal lower and upper bounds were established for epistemic uncertainty. Both the numerical algorithms and derivation are demonstrated. In Chapter 3 more applicable numerical implementations are introduced and the error analysis is also presented. A methodology to quantify uncertainties in model inputs subject to multiple prior distribution functions is proposed in Chapter 4 and in Chapter 5 Bayesian framework is applied to modify scientific models. Final conclusions are drawn in Chapter 6.
The text of this dissertation includes the reprints of the following papers and projects,
J. Li, X. Qi and D. Xiu On Upper and Lower Bounds for Quantity of Interest in Problems Subject to Epistemic Uncertainty. Accepted

X. Qi, D. Xiu, Numerical Computation for Bounds of Quantity of Interest Subject to Epistemic Randomness and Error Analysis. Submitted

X. Qi, D. Xiu, Model Discrepancy Study under Bayesian Framework (In preparation)
X. Qi, D. Xiu, Bayesian Analysis subject to multiple choices of priors (Project)

2. ON UPPER AND LOWER BOUNDS FOR QUANTIFY OF INTEREST IN PROBLEMS SUBJECT TO EPISTEMIC UNCERTAINTY

2.1 Introduction

Uncertainty quantification accesses the impacts of uncertainties on the output of a certain system and plays an increasingly more important role in many scientific computing problems. Roughly speaking, there are two types of uncertainties. One is irreducible and inherent from the physical model under consideration. This is often referred to in the literature as aleatoric uncertainty. The most widely adopted methods to quantification of aleatory uncertainty are probabilistic, where the uncertain inputs are modeled as random variables and/or random processes. Well established techniques such as Monte Carlo sampling (MCS), generalized polynomial chaos (gPC), etc., can be employed. On the other hand, uncertainties may be due to lack of knowledge and are reducible via increasing understanding of the underlying systems. This is often referred to as epistemic uncertainty. The complete prescription of probability distribution of epistemic uncertainty is thus not possible, as a result of the lack of knowledge. Hence, the traditional probabilistic methods can not be readily applied to the quantification of epistemic uncertainty.

There exist a few studies for epistemic uncertainty, using different approaches and focusing on different aspects. Without resorting to probability theory, attempts have been made using interval analysis [1, 2], possibility theory [3], fuzzy set theory [4, 5],evidence theory [6] etc., or methods partially based on probability theory such as second-order probability [7], probability boxes or credal sets [8], etc. Direct computations of system responses without specifying input probability have also been investigated [9, 10].

In a recent work of [11], a method to compute upper bounds for certain system responses was developed. The method utilizes a variational inequality of relative entropy ([12]) and results in sharp upper bounds, i.e., bounds without asymptotic constants. A more recent extension was presented in [13], where computations of both upper and lower bounds of the system responses were derived using the similar variational inequality. A notable feature of the work of [11, 13] is that the variational inequality of relative entropy, from which the bounds are derived, requires the functions under consideration must be bounded either from below or from above. This poses a restriction on the applicability of the bounds, because many quantities of interest in practical problems are not defined by such kind of functions. For example, the mean response of the system, one of the most sought after quantity of interest, is defined as an integral over probability space and does not satisfy the condition.

The contribution of the present work lies in a further extension of the work of [11, 13]. In particular, we present another variational inequality of relative entropy, which resembles closely to the one presented in [12] and used by [11, 13]. The subtle, and yet major, difference is that the current inequality requires the function under consideration to be integrable. This results in more flexible applications of the inequality. For example, the mean response of a system naturally satisfies the requirement. By taking advantage of the more flexible inequality, we then derive both upper bound and lower bound for quantities of interest of systems subject to epistemic uncertainty. The quantities of interest are now integrable functions and thus include quantities such as the statistical moments of the system responses. From a practical point of view, the bounds can be viewed as estimates of "the best case scenario" and "the worst case scenario" of the system response.

This paper is organized as follows. After presenting the setup of the problem in Section 2.2, we introduce, in Section 2.3, the new inequality for relative entropy and then derive the upper bound and lower bound for quantities of interest subject to epistemic uncertainties. We then discuss the case of mixed aleatoric and epis-

temic uncertainty in Section 2.4. Numerical examples are presented in Section 2.5 to illustrate the computations of the bounds, before the conclusion in Section 2.6.

2.2 Problem Setup

Throughout this paper we adopt the standard notation $(\Omega, \mathcal{F}, P)$ to denote a probability space, where Ω is the sample space, $\mathcal{F}$ the σ-algebra, and P the probability measure. We will also denote $\mathcal{P}(\Omega)$ as the set of probability measures on $(\Omega, \mathcal{F})$.

2.2.1 Quantity of interest

We assume that there exists a physical/mathematical system whose solution is denoted as u. The solution depends on spatial and/or temporal variables (depending on the nature of the system), as well as a set of uncertain parameters $z = (z_1, \ldots, z_{n_z}) \in \mathbb{R}^{n_z}$, $n_z \geq 1$. A quantity of interest (QoI), $\mathcal{Q}$, is defined as a (potentially nonlinear) function of the solution u, i.e., $\mathcal{Q} = q(u)$. Focusing on the dependence on the input parameters, we write the QoI as

$$\mathcal{Q} = f(z) = (q \circ u)(z). \tag{2.1}$$

Throughout this paper, we will be concerned with the estimation of the mean of the QoI. That is, we focus on

$$\mathbb{E}[\mathcal{Q}] = \mathbb{E}[f] = \int f(z) d\rho(z), \tag{2.2}$$

where ρ is the probability distribution of z and $\mathbb{E}$ denotes the expectation operator. For example, if $\mathcal{Q} = q(u) = u^k$, then (3.2) is the k^{th}-moment of the solution u. We remark that f represents the direct dependence of the QoI on the uncertain input z. Its explicit form is usually unknown and can only be evaluated via the composite function $(q \circ u)$. That is, to evaluate $\mathcal{Q}$ one needs to evaluate the solution u of the system first. This usually involves numerical solution of a complex system and can be highly computational intensive.

2.2.2 Epistemic uncertainty

Evaluation of (3.2) requires the knowledge of the probability distribution function ρ. In many practical problems, however, the prescription of the probability distribution to the inputs requires extensive knowledge and measurement data to the underlying physical system and is often impossible. The leads to our current setup of the problem: we are concerned with the uncertainty due to lack of knowledge — epistemic uncertainty.

More precisely, the epistemic uncertainty considered here presents itself in term of the parameters $z \in \mathbb{R}^{n_z}$, $n_z \geq 1$. The parameters can be physical parameters of the underlying system and/or hyperparameters charactering the inputs of the system. For example, z can be a set of random variables used in Karhunen-Loeve expansion (or any other expansions) of certain input random processes. And we assume that a complete specification of the probability distribution functions of z, $\rho(s)$, is not available, due to our lack of knowledge. This is frequently encountered in practice, when one often does not possess sufficient information or measurement data to specify the distribution of all parameters. The situation is more obvious for hyperparameters. For most finite series approximation for random processes, e.g., the Karhunen-Loeve expansion, the complete probability distribution of the hyperparameters is almost never known (for example, the joint probability among the parameters), except for special cases with Gaussian processes.

On the other hand, we will also discuss the case of mixed aleatory and epistemic uncertainty. By aleatory uncertainty we refer to the parametric uncertainty (through either physical parameters or hyperparameters) whose probability distribution is known.

2.3 Bound estimation for epistemic uncertainty

In this section we establish both the lower bound and the upper bound for the statistical moments of quantity of interest, as illustrated in (3.2). The bounds are

derived using a variational inequality of relative entropy. The inequality is similar to an existing inequality but extends its applicability.

2.3.1 Relative entropy

The *relative entropy*, also known as *cross entropy* or *Kullback-Liebler* divergence, measures the difference between two probability measures. Again, let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{P}(\Omega)$ the set of probability measures on $(\Omega, \mathcal{F})$. For $\gamma \in \mathcal{P}(\Omega)$, the relative entropy $R(\cdot\|\gamma) : \mathcal{P}(\Omega) \to \mathbb{R} \cup \{\infty\}$ is defined as

$$R(\rho\|\gamma) \triangleq \int_\Omega \log\left(\frac{d\rho}{d\gamma}\right) d\rho, \tag{2.3}$$

whenever ρ is absolutely continuous with respect to γ; otherwise, we set $R(\rho\|\gamma) \triangleq \infty$. There are many properties associated with the relative entropy, and we will not attempt to itemize them here. Interested readers are referred to books, for example, [12].

An important property of the relative entropy is a variational duality between exponential integral and the relative entropy. Here we restate the property, in the form of Proposition 4.5.1 in [12], and leave its details and proof to [12].

Proposition 2.3.1 ([12]) *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $k : \Omega \to \mathbb{R}$ a measurable function that is either bounded from below or bounded from above, and $\gamma \in \mathcal{P}(\Omega)$. Then,*

$$-\log \int_\Omega e^{-k} d\gamma = \inf_{\rho \in \Delta(\Omega)} \left\{ R(\rho\|\gamma) + \int_\Omega k d\rho \right\}, \tag{2.4}$$

where $\Delta(\Omega) \triangleq \{\rho \in \mathcal{P}(\Omega) : R(\rho\|\gamma) < \infty\}$.

The condition of the function being bounded from below or from above may not be easy to satisfy in practice. In order to consider the statistical moments of quantity of interest, as in (3.2), we present a different result, where the function is required to be integrable.

Theorem 2.3.1 *Let $\gamma, \rho \in \mathcal{P}(\Omega)$ be probability measures with $R(\rho\|\gamma) < \infty$. Assume a function $f : \Omega \to \mathbb{R}$ is integrable with respect to ρ, that is, $\int_\Omega |f| d\rho < \infty$, then the following inequality holds*

$$-\log \int_\Omega e^{-f} d\gamma \leq R(\rho\|\gamma) + \int_\Omega f d\rho. \tag{2.5}$$

Proof For any $\alpha \in \mathbb{N}$, let us consider $f \vee (-\alpha)$, which is obviously bounded from below and measurable. Then by applying Proposition 2.3.1, we have

$$\begin{aligned} -\log \int_\Omega e^{-[f\vee(-\alpha)]} d\gamma &= \inf_{\theta\in\Delta(\Omega)} \left\{ R(\theta\|\gamma) + \int_\Omega [f \vee (-\alpha)] d\theta \right\} \\ &\leq R(\rho\|\gamma) + \int_\Omega [f \vee (-\alpha)] d\rho. \end{aligned} \tag{2.6}$$

Since $f\vee(-\alpha) \geq f\vee[-(\alpha+1)]$, then $e^{-[f\vee(-\alpha)]} \leq e^{-\{f\vee[-(\alpha+1)]\}}$ and $\lim_{\alpha\to\infty} e^{-[f\vee(-\alpha)]} = e^{-f}$. Thus by the Monotone Convergence Theorem, we obtain

$$\lim_{\alpha\to\infty} -\log \int_\Omega e^{-[f\vee(-\alpha)]} d\gamma = -\log \int_\Omega e^{-f} d\gamma. \tag{2.7}$$

Since $\lim_{\alpha\to\infty} f \vee (-\alpha) = f$ and $|f \vee (-\alpha)| < |f|$ point-wisely, and meanwhile $\int_\Omega |f| d\rho < \infty$ from the assumption, then by the Lebesgue Dominated Convergence Theorem, we obtain

$$\lim_{\alpha\to\infty} \int_\Omega f \vee (-\alpha) d\rho = \int_\Omega f d\rho. \tag{2.8}$$

Let α goes to ∞ in (2.6) and use (2.7) and (2.8), the main conclusion (2.5) holds. ■

2.3.2 Upper bound and lower bound

An immediate application of Theorem 2.3.1 follows by setting f to be cf for any $c > 0$, that is,

$$\int_\Omega f d\rho \geq -\frac{1}{c} R(\rho\|\gamma) - \frac{1}{c} \int_\Omega e^{-cf} d\gamma. \tag{2.9}$$

Since $\int_\Omega |-f| d\rho = \int_\Omega |f| d\rho < \infty$, by replacing f with $-f$ in (2.9), we obtain another inequality

$$\int_\Omega f d\rho \leq \frac{1}{c} R(\rho\|\gamma) + \frac{1}{c} \int_\Omega e^{cf} d\gamma. \tag{2.10}$$

For notational convenience, let us define, for any $c > 0$,

$$\begin{aligned} \Theta^-(c) &\triangleq -\frac{1}{c}\log\int_\Omega e^{-cf}d\gamma, \\ \Theta^+(c) &\triangleq \frac{1}{c}\log\int_\Omega e^{cf}d\gamma. \end{aligned} \tag{2.11}$$

Note that by l'Hospital's rule, we immediately have

$$\begin{aligned} \lim_{c\to 0^+}\Theta^-(c) &= \int_\Omega f d\gamma, \\ \lim_{c\to 0^+}\Theta^+(c) &= \int_\Omega f d\gamma, \end{aligned} \tag{2.12}$$

which naturally allow us to extend the definitions for $\Theta^-(c)$ and $\Theta^+(c)$ to $c \in [0,\infty)$.

The aforementioned results can be summarized as follows.

Theorem 2.3.2 *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\gamma \in \mathcal{P}(\Omega)$, $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ and define*

$$R^* = \sup_{\rho\in\mathcal{A}} R(\rho\|\gamma). \tag{2.13}$$

Assume $f : \Omega \to \mathbb{R}$ is a measurable function satisfying

$$\int_\Omega |f| d\rho < \infty, \qquad \text{for any } \rho \in \mathcal{A}.$$

With the definitions of Θ^+ and Θ^- in (3.5), define

$$\mathbb{E}[f]^u \triangleq \inf_{c\in(0,\infty)}\left[\Theta^+(c) + \frac{1}{c}R^*\right], \tag{2.14}$$

and

$$\mathbb{E}[f]^\ell \triangleq \sup_{c\in(0,\infty)}\left[\Theta^-(c) - \frac{1}{c}R^*\right]. \tag{2.15}$$

Then, for any probability measure $\rho \in \mathcal{A}$, the following bounds hold for $\mathbb{E}[f]$,

$$\mathbb{E}[f]^\ell \le \int_\Omega f d\rho \le \mathbb{E}[f]^u. \tag{2.16}$$

The following results ensure that the bounds (2.14) and (2.15) can be obtained as optimization problems over a single real parameter $c > 0$.

Proposition 2.3.2 *Let* $D = \{c : \Theta^+(c) < \infty\}$ *and* $E = \{c : \Theta^-(c) > -\infty\}$, *where* Θ^+ *and* Θ^- *are defined in* (3.5). *Denote* D° *the interior of* D *and* E° *the interior of* E, *and assume* $D^\circ \neq \emptyset$ *and* $E^\circ \neq \emptyset$. *Let* R^* *be a positive constant. Then, under the same assumptions as in Theorem 3.3.1, the following conclusions hold:*

(a) $\Theta^+(c)$ *is differentiable on* D° *and nondecreasing for* $c \geq 0$. *There is a unique* $c \in (0, \infty]$ *at which* $\Theta^+(c) + \frac{1}{c}R^*$ *attains a local minimum. The minimum occurs at* $c = \infty$ *means that* $\Theta^+(c) + \frac{1}{c}R^* > \lim_{c\to\infty} \Theta^+(c)$ *for a well defined* $\lim_{c\to\infty} \Theta^+(c)$;

(b) $\Theta^-(c)$ *is differentiable on* E° *and nonincreasing for* $c \geq 0$. *There exists a unique* $c \in (0, \infty]$ *at which* $\Theta^-(c) - \frac{1}{c}R^*$ *attains a local maximum. The maximum occurs at* $c = \infty$ *means that* $\Theta^-(c) - \frac{1}{c}R^* < \lim_{c\to\infty} \Theta^-(c)$ *for a well defined* $\lim_{c\to\infty} \Theta^-(c)$.

The proof is left to Appendix B.

2.4 Bound estimation for mixed aleatory and epistemic uncertainty

We now generalize the aforementioned results to the case of mixed aleatory and epistemic uncertainties. Once again, we use aleatory uncertainty to denote random parameters (either physical parameters or hyperparameters) with known probability distribution function. Hereafter we reserve the variable $x \in \mathbb{R}^{n_x}$ for aleatory variable with known probability measure $\mu(x)$ and the variable $y \in \mathbb{R}^{n_y}$ for epistemic variable whose probability distribution is not completely known. We also denote the domain of the aleatory variable x as Ω_1 and that of the epistemic variable y as Ω_2. The corresponding probability spaces are $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$, with $\Omega \triangleq \Omega_1 \otimes \Omega_2$. The quantity of interest (3.2) is then rewritten as

$$\mathbb{E}[f] = \int_{\Omega_2} \int_{\Omega_1} f(x, y)\mu(dx)\rho(dy), \tag{2.17}$$

where μ is known and ρ is unknown.

The definitions of Θ^+ and Θ^- in (3.5) can then be generalized in two different forms. The first form is a straightforward separation of the aleatory and epistemic variables in the exponential integral. That is, for $c > 0$, let

$$\begin{aligned}\Lambda^+(c) &\triangleq \frac{1}{c}\log\int_{\Omega_2}\int_{\Omega_1} e^{cf(x,y)}\mu(dx)\gamma(dy),\\ \Lambda^-(c) &\triangleq -\frac{1}{c}\log\int_{\Omega_2}\int_{\Omega_1} e^{-cf(x,y)}\mu(dx)\gamma(dy).\end{aligned} \tag{2.18}$$

The second form of generalization integrates the aleatory variable in the exponential function first and takes the following form, for $c > 0$,

$$\begin{aligned}\Lambda_1^+(c) &\triangleq \frac{1}{c}\log\int_{\Omega_2} e^{\int_{\Omega_1} cf(x,y)\mu(dx)}\gamma(dy),\\ \Lambda_1^-(c) &\triangleq -\frac{1}{c}\log\int_{\Omega_2} e^{\int_{\Omega_1} -cf(x,y)\mu(dx)}\gamma(dy).\end{aligned} \tag{2.19}$$

Since the limits of these quantities are well defined,

$$\lim_{c\to 0^+}\Lambda^-(c),\Lambda_1^-(c),\Lambda^+(c),\Lambda_1^+(c) = \int_{\Omega_2}\int_{\Omega_1} f(x,y)\mu(dx)\gamma(dy),$$

their definitions can be readily extended to $c \in [0,\infty)$ by setting $\Lambda^-(0),\Lambda^+(0),\Lambda_1^-(0)$ and $\Lambda_1^+(0)$ to be the aforementioned limit.

A straightforward application of the Jensen's inequality to the exponential function leads to

$$\Lambda^-(c) < \Lambda_1^-(c), \qquad \Lambda_1^+(c) < \Lambda^+(c).$$

Then, by setting $f(y) = \int_{\Omega_1} f(x,y)\mu(dx)$ and $\Omega = \Omega_2$ in (2.10), we obtain

$$\int_{\Omega_2}\int_{\Omega_1} f(x,y)\mu(dx)\rho(dy) \le \frac{1}{c}R(\rho(dy)\|\gamma(dy)) + \Lambda_1^+(c), \quad \text{for any } \rho\in\mathcal{P}(\Omega_2), \tag{2.20}$$

where $R(\rho\|\gamma)$ is the relative entropy of ρ with respect to γ, defined in (2.3). Similarly, by letting $f(y) = \int_{\Omega_1} f(x,y)\mu(dx)$ and $\Omega = \Omega_2$ in (2.9), we obtain

$$\int_{\Omega_2}\int_{\Omega_1} f(x,y)\mu(dx)\rho(dy) \ge -\frac{1}{c}R(\rho(dy)\|\gamma(dy)) + \Lambda_1^-(c), \quad \text{for any } \rho\in\mathcal{P}(\Omega_2). \tag{2.21}$$

These inequalities immediately result in a upper bound and a lower bound for (2.17), as summarized in the following statement.

Theorem 2.4.1 *Let $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ be the probability spaces as defined above. Let $\mu \in \mathcal{P}(\Omega_1)$ and $\gamma \in \mathcal{P}(\Omega_2)$ and define*

$$R^* = \sup_{\rho \in \mathcal{A}} R(\rho(dy) \| \gamma(dy)), \tag{2.22}$$

where $\mathcal{A} \subseteq \mathcal{P}(\Omega_2)$ is a set of probability measures. Assume

$$\int_{\Omega_2} \int_{\Omega_1} |f(x,y)| \mu(dx) \rho(dy) < \infty, \qquad \text{for any } \rho \in \mathcal{A}.$$

With the definitions of Λ_1^+ and Λ_1^- in (2.19), define

$$\begin{aligned} \mathbb{E}[f]^u &\triangleq \inf_{c \in (0,\infty)} \left[\frac{1}{c} R^* + \Lambda_1^+(c) \right], \\ \mathbb{E}[f]^\ell &\triangleq \sup_{c \in (0,\infty)} \left[-\frac{1}{c} R^* + \Lambda_1^-(c) \right]. \end{aligned} \tag{2.23}$$

Then the following bounds hold for $\mathbb{E}[f]$,

$$\mathbb{E}[f]^\ell \leq \int_{\Omega_2} \int_{\Omega_1} f(x,y) \mu(dx) \rho(dy) \leq \mathbb{E}[f]^u, \qquad \text{for any } \rho \in \mathcal{A}. \tag{2.24}$$

The following results ensure that the bounds (2.23) can be obtained by conducting optimization over a single real parameter $c > 0$.

Proposition 2.4.1 *Let $D^+ = \{c : \Lambda_1^+(c) < \infty\}$ and $D^- = \{c : \Lambda_1^-(c) > -\infty\}$, where Λ_1^+, and Λ_1^- are defined in (2.18). Denote $(D^+)^\circ$ as the interior of D^+ and $(D^-)^\circ$ the interior of D^-, and assume $(D^+)^\circ \neq \emptyset$ and $(D^-)^\circ \neq \emptyset$. Let R^* be a positive constant. Then, under the same assumptions as in Theorem 2.4.1, the following conclusions hold:*

(a) $\Lambda_1^+(c)$ is differentiable on $(D^+)^\circ$ and nondecreasing for $c \geq 0$. There is a unique $c \in (0, \infty]$ at which $\Lambda_1^+(c) + \frac{1}{c} R^*$ attains a local minimum. The minimum occurs at $c = \infty$ means that $\Lambda_1^+(c) + \frac{1}{c} R^* > \lim_{c \to \infty} \Lambda_1^+(c)$ for a well defined $\lim_{c \to \infty} \Lambda_1^+(c)$;

(b) $\Lambda_1^-(c)$ is differentiable on $(D^-)^\circ$ and nonincreasing for $c \geq 0$. There exists a unique $c \in (0, \infty]$ at which $\Lambda_1^-(c) - \frac{1}{c} R^*$ achieves a local maximum. The maximum occurs at $c = \infty$ means that $\Lambda_1^-(c) - \frac{1}{c} R^* < \lim_{c \to \infty} \Lambda_1^-(c)$ for a well defined $\lim_{c \to \infty} \Lambda_1^-(c)$.

Proof This proposition is a direct consequence of the Proposition 2.3.2 by setting $f(y)$ there to be $\int_{\Omega_1} f(x,y)\mu(dx)$ and Ω to be Ω_2. ■

Similar to the approach used in [11], we explore the behavior of $\Lambda_1^+(c)$ and $\Lambda_1^-(c)$ when c goes to ∞ and obtain the following results.

Theorem 2.4.2 *Under the same assumptions used in the definitions of* Λ_1^+ *and* Λ_1^- *in* (2.19)*, if we further assume* Ω_1 *and* Ω_2 *are finite dimensional spaces, then the following conclusions hold:*

(a) If f *is lower semi-continuous in* y *for each* $x \in \Omega_1$ *and bounded from below, then*

$$\operatorname{ess\,sup}_{y\in\Omega_2} \int_{\Omega_1} f(x,y)\mu(dx) = \lim_{c\to\infty} \Lambda_1^+(c); \tag{2.25}$$

(b) If f *is upper semi-continuous in* y *for each* $x \in \Omega_1$ *and bounded from above, then*

$$\operatorname{ess\,inf}_{y\in\Omega_2} \int_{\Omega_1} f(x,y)\mu(dx) = \lim_{c\to\infty} \Lambda_1^-(c). \tag{2.26}$$

The proof is a straightforward application of the Proposition 4 in [11] and will not be repeated here. This theorem ensures that $\lim_{c\to\infty} \Lambda_1^+(c)$ is the tightest upper bound and $\lim_{c\to\infty} \Lambda_1^-(c)$ is the tightest lower bound for $\int_{\Omega_1} f(x,y)\mu(dx)$ for any epistemic variable y. However, one should notice the additional assumptions made on the function f, i.e., the boundedness and semi-continuity.

2.5 Numerical Examples

In this section we provide a few examples to illustrate the estimation of the bounds. The quantities of interest in the examples are the mean solutions of some benchmark problems, which include an analytically known functions and partial differential equations subject to uncertain inputs. Estimating the bounds for statistical moments other than the mean will be exactly the same procedure.

To apply the bounds in Theorem 3.3.1 and 2.4.1, we first choose $\mathcal{A}$ as a set of probability measures from which the epistemic variables can take their distributions.

This will be used as a candidate set of all possible/allowable probability measures for the epistemic variables. In practice, the choice of the set $\mathcal{A}$ depends on the knowledge one has about the epistemic uncertainty. The more information one possesses, the smaller the set will be, and vice versa. We then choose a probability measure γ, a nominal measure that serves as a "best guess" for the epistemic variables. The distance R^* (3.4) can then be computed. And a forward stochastic computation is then carried out using the nominal distribution γ for the epistemic variables, along with the aleatory variable for the mixed case. With the distribution for the epistemic variable fixed as the nominal distribution, this step is the standard stochastic forward simulation. The estimation of the lower bound and upper bound is then carried out as an optimization problem over a single real variable $c \geq 0$, from (3.6) and (2.24). A more detailed description of the algorithm can be found in [13].

2.5.1 Example 1: analytically known function

We first consider an analytically known function

$$u(z_1, z_2) = \exp(2z_1)\cos(3\pi z_2), \tag{2.27}$$

where z_1 and z_2 are uncertain variables, one of which is epistemic and the other aleatoric. We consider the following four cases:

(a) z_1 is an epistemic variable with $\mathcal{A} = \{\text{Beta}(\alpha_1, \alpha_1), \alpha_1 \in [0, 0.5]\}$, z_2 is an aleatory variable with uniform distribution $U[-1, 1]$;

(b) z_1 is an epistemic variable with $\mathcal{A} = \{\text{Beta}(\alpha_1, \alpha_1), \alpha_1 \in [-\frac{1}{3}, 0]\}$, and z_2 is an aleatory variable with uniform distribution $U[-1, 1]$;

(c) z_1 is an aleatory variable with uniform distribution $U[-1, 1]$, z_2 is an epistemic variable with $\mathcal{A} = \{\text{Beta}(\alpha_2, \alpha_2), \alpha_2 \in [0, 0.5]\}$;

(d) z_1 is an aleatory variable with uniform distribution $U[-1, 1]$, z_2 is an epistemic variable with $\mathcal{A} = \{\text{Beta}(\alpha_2, \alpha_2), \alpha_2 \in [-\frac{1}{3}, 0]\}$;

Here Beta(α, β) stands for Beta distribution on $[-1, 1]$ with density function $\rho(x) \propto (1+x)^{\alpha}(1-x)^{\beta}$, $\alpha, \beta > -1$. Note that Beta$(0, 0)$ is the uniform distribution $U[-1, 1]$.

This is a case of mixed aleatory and epistemic uncertainty. We then apply the bounds derived in Theorem 2.4.1. The lower and upper bounds for the mean solution are tabulated in Table 3.2. We observe that both case (a) and (b) have very small bounds for the mean, whereas case (c) and (d) have bounds of order one separation. These indicate that the mean solution of this function is insensitive to the distribution of the first variable z_1 and sensitive only to the distribution of z_2. From a practical point of view, this implies that for this function the effect of epistemic uncertainty in z_1 is insignificant. Therefore, effort to reduce uncertainty should be devoted to z_2 first. The same function has been studied in [14], where distributional sensitivity–the sensitivity of solution with respect to the distribution of the inputs–was studied. It was found that the mean solution is insensitive to the change of probability distribution in variable z_1. The current estimation of the bounds on the mean solution is consistent with that finding.

Table 2.1
Example 1: bounds of $\mathbb{E}[u(z_1, z_2)]$ in each of the four cases.

Case number	(a)	(b)	(c)	(d)
R^*	0.0650	0.0530	0.0650	0.0530
Lower bound	-0.0006	-0.0005	-0.2970	-0.2687
Upper bound	0.0006	-0.0005	0.2970	0.2687

2.5.2 Example 2: Burger's equation

Consider viscous Burgers' equation

$$u_t + uu_x = \nu u_{xx}, \quad x \in [-1, 1],$$
$$u(-1) = 1 + \delta, \qquad u(1) = -1.$$

where δ is a small perturbation to the left boundary condition and $\nu > 0$ is the viscosity. The solution has a transition layer, whose location x^*, defined as the zero of the solution profile at steady state $u(x^*) = 0$, can be super-sensitive to the uncertainty δ at the left boundary condition. It can be found by solving a nonlinear system of equations ([15])

$$A \tanh\left[\frac{A}{2\nu}(1+x^*)\right] = 1+\delta, \qquad A \tanh\left[\frac{A}{2\nu}(1-x^*)\right] = 1,$$

for any fixed value of δ and ν. Detailed discussion and simulations of this problem can be found in [15].

We now model the uncertain input δ as epistemic and assume it is distributed on an interval $[0, e]$, $e \ll 1$, with unknown distribution function. The candidate set of probability measures $\mathcal{A}$, from which the epistemic variable can take its distribution, is set to be beta distribution Beta(α, β) with $0 \leq \alpha \leq 6$ and $0 \leq \beta \leq 3$. And the nominal distribution γ is set to be the uniform distribution in $[0, e]$. Then the R^* value, defined in (3.4), is $R^* = 0.2639$.

The quantity of interest is the mean location of the transition layer, i.e., $\mathbb{E}[x^*]$. The resulting upper and lower bounds are tabulated in Table 2.2 and 2.3, for $\nu = 0.05$ and $\nu = 0.1$, respectively. The notable separation between the lower bounds and the upper bounds, especially for the case of $\nu = 0.05$, indicates that the probability distribution of the uncertain input δ plays an important role in determining the location of the transition layer.

Table 2.2
Bounds for the mean location of the transition layer for steady Burger's equation, $\nu = 0.05$.

e	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Lower bound	0.6430	0.5266	0.4113	0.2962
Upper bound	0.8594	0.7354	0.6183	0.5029

Table 2.3
Bounds for the mean location of the transition layer for steady Burger's equation, $\nu = 0.1$.

e	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Lower bound	0.2892	0.1110	0.0203	0.0021
Upper bound	0.7188	0.4708	0.2375	0.0508

2.5.3 Example 3: steady state diffusion equation

We consider the following elliptic problem in one spatial dimension,

$$-\frac{\partial}{\partial x}\left[\kappa(x,z)\frac{\partial u}{\partial x}(x,z)\right] = f, \qquad x \in (0,1) \tag{2.28}$$

with a constant forcing term $f = 2$ and boundary conditions

$$u(0,z) = -1, \quad u(1,z) = 1.$$

The variable $z = (z_1, \ldots, z_{n_z})$ are hyperparameters in the diffusivity field κ. In particular, we assume the diffusivity field takes the following form

$$\kappa(x,z) = 1 + \sigma \sum_{k=1}^{n_z} \frac{1}{k^2\pi^2} \cos(2\pi k x) z_k, \tag{2.29}$$

where $\sigma \leq 1$ is a parameter. Note that many of the parameterization procedure for random fields result in similar expansion. For example, if one adopts the well known Karhunen-Loeve expansion for random processes, then the cosine functions will be replaced by the eigenfunctions of the covariance function of the underlying random process and the $1/(k^2\pi^2)$ coefficients replaced by the reciprocal of the square root of the corresponding eigenvalues. Here we adopt the particular form of (2.29) to isolate the issue of prescribing the distributions for z. In practice, one rarely, if not never, possesses sufficient information to fully prescribe the distribution of z, especially for non-Gaussian processes, regardless of the parameterization technique. And this leaves z as epistemic variables.

Here we set $\sigma = 1$ and $n_z = 5$. (The choice of n_z is rather arbitrary here.) The candidate set for probability measures $\mathcal{A}$ for z is taken as $\{\text{Beta}(\alpha,\beta), 0 \leq \alpha \leq 6, 0 \leq \beta \leq 3\}$ in $[-1,1]$ for each z_i, $i = 1, \ldots, n_z$, and with mutual independence among them. We then take $\gamma \sim U[-1,1]^{n_z}$, the independent uniform distribution in the hypercube $[-1,1]^{n_z}$, as the nominal distribution for z, and conduct a stochastic Galerkin computation of (2.28) using Legender polynomial chaos [16]. The upper bound and lower bound of the mean solution $\mathbb{E}[u]$ are then computed using Theorem 3.3.1. The bounds are plotted in Fig. 2.1. We observe that in this case, the bounds are very

tight around the mean solution obtained using the nominal distribution. This indicates that in this case the mean solution is not very sensitive to the distributions of the inputs, at least when they are drawn from beta distributions. Since the assumption of independent uniform distribution is widely used in most work concerning the stochastic diffusion problem, the current result in a way justifies the approach, as far as the mean solution is concerned. Note that the similar conclusion has been reached in [14] vis distributional sensitivity analysis.

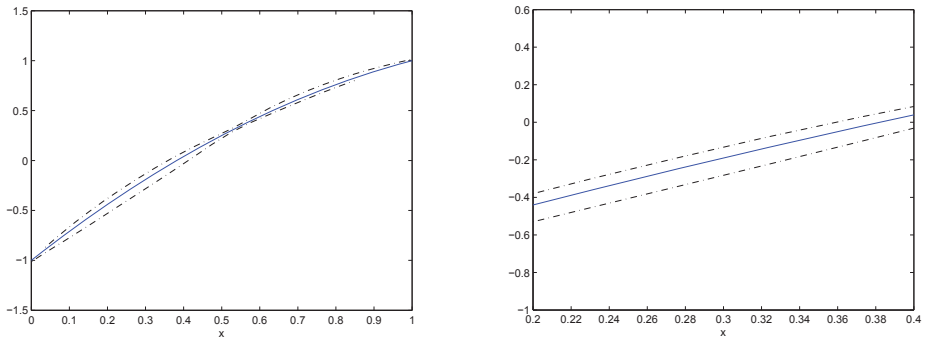

Figure 2.1. Upper bound and lower bound (dashed lines) of mean solution $\mathbb{E}[u]$, along with the mean solution computed using the nominal distribution (solid curve). Left: global view; Right: close-up view.

2.6 Summary

In this paper we derive rigorous upper bound and lower bound for the quantity of interest of problems subject to epistemic uncertain inputs. The bounds are based on a newer variational inequality, which applies to integrable functions. This allows us to derive upper bound and lower bound for statistical moments of the problems. In practical problems when one does not have sufficient information of the probability distribution of the inputs, the bounds can serve as reliable estimate of "the best case scenario" and "the worst case scenario of the quantity of interest.

3. NUMERICAL IMPLEMENTATION OF EPISTEMIC UNCERTAINTY QUANTIFICATION

3.1 Introduction

Uncertainty quantification (UQ) has become an essential component for prediction and simulation. It accesses the impacts of uncertainties on the output of a certain system and plays an increasingly important role in many scientific computing problems. One of the critical tasks in UQ is to propagate uncertainty throughout the system in an efficient manner. To this end, many numerical methods have been developed, especially in the last decade. Among them, it is widely acknowledged that the methods based on generalized polynomial chaos (gPC) ([16, 17]), either in stochastic Galerkin (SG) or stochastic collocation (SC) form, are advantageous at low and medium dimensions, and Monte Carlo (MC) and quasi-Monte Carlo (qMC) methods are usually preferred at very high dimensions. And many efforts have since been devoted to the development of efficient algorithms, especially in SC, using adaptive approach or sparsity, to further push the applicability of SC to higher dimensions. See, for example, [18, 19] for general descriptions and the references therein.

Most of these existing methods in UQ utilize probabilistic framework, where the uncertain inputs are modeled as random variables/processes. Consequently, the probability distributions of the inputs are required. For example, in MC, the distribution is required for conducting sampling; whereas in gPC methods it is required to construct orthogonal basis. In many practical problems, however, acquiring the probability distribution function of the inputs can be highly difficult, if not impossible. This is particularly true if the inputs involve multiple random variables. For Gaussian case, its multi-variate distribution function can be fully determined by its component-wise marginal (Gaussian) distributions and the covariance function; for non-Gaussian case,

this is not true and one has to acquire the complete information of the joint distributions among all components. This can be an insurmountable task to accomplish for many practical systems.

In this paper, we focus on a situation that often arises in practice. That is, we consider the UQ problems where the precise probability distribution function of the random inputs is not available and can belong to a family of distribution function. The common cause for this is the lack of knowledge, e.g., lack of observation data, inability to conduct measurements, insufficient inference procedure, etc. All of these contribute to a lack of sufficient information to fully prescribe the input probability distribution function. For example, the limited amount of information may suggest that the input distribution is bounded in an interval $[a, b]$ with $b > a$. However, it is not sufficient to determine the exact type of distribution. Rather, one may only be able to confirm that the real distribution belongs to the family of beta distribution in $[a, b]$ with parameters α and β in certain range. Or, the available information may confirm that the input distribution is indeed a Gaussian distribution, but is not able to determine the precise value of the mean and standard deviation.

Uncertainty due to lack of knowledge is commonly referred to as *epistemic uncertainty* in the literature, whereas uncertainty due to intrinsic stochasticity is often termed as *aleatory uncertainty*. (See, for example, [20]). Here, by slightly abusing the definitions we can also loosely consider the situation we focus on an *epistemic uncertainty*, because it is indeed caused by a lack of knowledge. Note, however, the real source of the uncertain inputs in our setting may as well be aleatoric – we simply do not possess sufficient knowledge to know its precise distribution function. In this situation, we are dealing with an epistemic type of uncertainty on top of aleatory uncertainty. (We remark that the classification of uncertainty is not the focus of this paper — the discussion here is to merely draw connection with some of the existing classifications in the literature.)

Given that the input uncertain belongs to a family of distributions, our goal is to compute sharp upper and lower bounds of the solution statistics. In practice, this

can serve as estimates of the best-case scenario and worst-case scenario responses. The method we present in the paper follows from a line of recent work. In [11], a method to compute sharp upper bounds for certain system responses was developed, by utilizing a variational inequality of relative entropy ([12]). This work was followed by [13], where the probability of failure is considered as the quantity of interest of the underlying system. In this special case, both upper and lower bounds can be computed. A notable extension of the methodology was presented in [21], where a new variational inequality applicable integrable functions is derived. (The inequality used in [11] applies to the functions bounded from either above or below). This allows the computation of both upper and lower bounds for general solution statistics and can be more flexible in practice.

This paper is built upon the work of [21]. Here we present a detailed discussion on how to use the theoretical bounds derived in [21] and compute the numerical upper and lower bounds for the solution statistics. In addition to the implementation procedure, we will analyze the errors introduced by the numerical approximations and provide numerical error bounds on top of the theoretical bounds. By combining both the numerical bounds and the theoretical bounds, we then obtain upper and lower bounds that can be directly used in practical problems.

This paper is organized as follows. After presenting the setup of the problem in Section 3.2, we introduce, in Section 3.3, the analytical bounds derived by [21]. In Section 3.4, we discuss the detailed numerical procedure, analyze the numerical errors and provide numerical bounds. Finally, numerical examples are presented in Section 3.5 to illustrate the computations of the bounds, before the conclusion in Section 3.6.

3.2 Problem Setup

Throughout this paper we adopt the standard notation $(\Omega, \mathcal{F}, P)$ to denote a probability space, where Ω is the sample space, $\mathcal{F}$ the σ-algebra, and P the probability measure. We will also denote $\mathcal{P}(\Omega)$ the set of probability measures on $(\Omega, \mathcal{F})$.

3.2.1 Quantity of interest

We assume that there exists a physical/mathematical system whose solution is denoted as u. Without specifying the actual system, we simply assume that it is governed by a set of (potentially highly complex) equations and can be accurately simulated by a numerical procedure. The solution depends on spatial and/or temporal variables (depending on the nature of the system), as well as a set of uncertain parameters $z = (z_1, \ldots, z_{n_z}) \in \mathbb{R}^{n_z}$, $n_z \geq 1$. These parameters are used to characterize the input uncertainty to the system. They can be physical parameters in the model (for example, reaction constants in a chemical reaction system) and/or hyperparameters used to parameterize certain input processes (for example, the random parameters in a Karhunen-Loeve expansion of an input random process). A quantity of interest (QoI), $\mathcal{Q}$, is defined as a function of the solution u, i.e., $\mathcal{Q} = q(u)$. Focusing on the dependence on the input parameters, we write the QoI as

$$\mathcal{Q} = f(z) = (q \circ u)(z). \tag{3.1}$$

Throughout this paper, we will be concerned with the estimation of the mean of the QoI. That is, we focus on

$$\mathbb{E}[\mathcal{Q}] = \mathbb{E}[f] = \int f(z) d\rho(z), \tag{3.2}$$

where ρ is the probability distribution of z and $\mathbb{E}$ denotes the expectation operator. For example, if $\mathcal{Q} = q(u) = u^k$, then (3.2) is the k^{th}-moment of the solution u. Or, if $\mathcal{Q} = \mathbb{I}_{\{u>0\}}(u)$, where $\mathbb{I}$ is the indicator function, then it is the probability of the occurrence of $u > 0$ and an usual definition for probability of failure. We remark that f represents the dependence of the QoI on the uncertain input z. Its explicit form is usually unknown and can only be evaluated via the composite function $(q \circ u)$. That is, to evaluate $\mathcal{Q}$ one needs to evaluate the solution u of the system first. This usually involves numerical solution of the underlying system and can be highly computationally intensive.

3.2.2 Uncertainty in Input Distribution

Evaluation of (3.2) requires the solution of the underlying system, which is usually handled by a properly chosen stochastic algorithm, e.g. example, gPC based stochastic collocation methods or Monte Carlo sampling, etc. Regardless of the chosen stochastic algorithm, the knowledge of the probability distribution function ρ is required. Mathematically, the distribution function is defined by $F_Z(s) = \text{Prob}(z \leq s)$, for $s \in \mathbb{R}^{n_z}$, from which the probability density function can be derived as $dF_z(s) = \rho(s)ds$. In many practical problems, the specification of this function requires a large amount of information and is often difficult, if not impossible, to accomplish.

In this paper, we loose this requirement and study the following problem. Given the available (limited amount) information, the probability distribution of the input variables Z lies in a family of distributions, then what is the upper and lower bounds of solution statistics (3.2). More specifically, we study

> Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a family of probability measures, from which the true distribution function ρ takes form, find P^u and P^ℓ such that
>
> $$P^\ell \leq \int f(z)d\rho(z) \leq P^u, \qquad \rho \in \mathcal{A}. \tag{3.3}$$

The family of probability measures $\mathcal{A}$ denotes all the possible/allowable probability distributions that the input variable z may take form. We assume that the true probability distribution ρ exists but is not known precisely due to lack of knowledge. Instead, with the available (limited) information, we can only assert that the true probability distribution ρ belongs to the family of distributions $\mathcal{A}$. Again, the nature of the true distribution ρ could be intrinsic (aleatoric) or due to lack of knowledge (epistemic). This is not the concern of this paper. Here we assume the true distribution ρ exists (even in the degenerate case of delta function, which implies that z is a deterministic variable), but its precise form is not known precisely and belongs to the family $\mathcal{A}$. To this end, we can consider this uncertainty in the form of the distribution function of z as an epistemic uncertainty, in addition to the original uncertainty (aleatory or epistemic) that induces the variable z in the first place.

3.3 Analytical Bounds by [21]

This problem has been studied in [21], where analytical bounds using a variational inequality of the relative entropy is derived.

Theorem 3.3.1 ([21]) *Let $(\Omega, \mathcal{F}, P)$ be a probability space. For a probability measure $\gamma \in \mathcal{P}(\Omega)$ and a set of measures $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, define*

$$R^* = \sup_{\rho \in \mathcal{A}} R(\rho \| \gamma), \tag{3.4}$$

where $R(\rho\|\gamma) = \int_\Omega \log(d\rho/d\gamma)d\rho$ is the relative entropy between ρ and γ. Assume $f : \Omega \to \mathbb{R}$ is a measurable function satisfying $\int_\Omega |f| d\rho < \infty$, for any $\rho \in \mathcal{A}$, and define, for any $c > 0$,

$$\Theta^+(c) = \frac{1}{c} \log \int_\Omega e^{cf} d\gamma, \qquad \Theta^-(c) = -\frac{1}{c} \log \int_\Omega e^{-cf} d\gamma. \tag{3.5}$$

Then, for any probability measure $\rho \in \mathcal{A}$, the following bounds hold

$$P^\ell \leq \int_\Omega f d\rho \leq P^u, \tag{3.6}$$

where

$$P^\ell = \sup_{c \in (0,\infty)} \left[\Theta^-(c) - \frac{1}{c} R^* \right] \qquad P^u = \inf_{c \in (0,\infty)} \left[\Theta^+(c) + \frac{1}{c} R^* \right]. \tag{3.7}$$

The bounds are sharp, in the sense that there is no asymptotic constants. Meanwhile, it is worth noting one can certainly derive other bounds using different methods.

3.4 Numerical Bounds

In this section we present a numerical procedures for computing the analytical bounds in Theorem 3.3.1. Naturally, this will produce numerical bounds that approximate the analytical bounds. We will then analyze the properties of the numerical bounds.

3.4.1 General Numerical Procedure

To keep consistent notations, we will use the following throughout this paper.

- ρ: the true probability distribution of the inputs. It is not available.
- $\mathcal{A}$: the family (or class) of probability distributions that the true distribution ρ may take form, $\rho \in \mathcal{A}$.
- γ: a nominal probability distribution chosen by the practitioner. Although there is no strong mathematical restriction on its choice, in practice it serves as the "best guess" of ρ and should satisfy $\gamma \in \mathcal{A}$. Note that the quality of this best-guess certainly depends on the amount of information one possesses.

Our general numerical procedure then takes the following steps.

1. Using the available information, choose the set of probability measures $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, from which the true distribution ρ may take.
2. Choose the nominal distribution γ and compute the quantity R^* in (3.4).
3. Using the nominal distribution γ for the variable z, conduct a standard stochastic forward simulation for the underlying system and approximate the QoI $f(z)$ in (3.1). With the chosen distribution γ in place, many of the standard stochastic algorithms can be applied. We require that the algorithms are able to approximate the integral of $f(z)$ with respect to γ such that, for any fixed $c > 0$, we obtain approximations to the quantities Θ^+ and Θ^- in (3.5). We denote $\widetilde{\Theta}^+(c) \approx \Theta^+(c)$ and $\widetilde{\Theta}^-(c) \approx \Theta^-(c)$ as the approximations.
4. Conduct a numerical optimization over the parameter $c > 0$ to obtain

$$\begin{aligned} \widetilde{P}^u &= \inf_{c\in(0,\infty)} \left[\widetilde{\Theta}^+(c) + \frac{1}{c}R^*\right], \\ \widetilde{P}^\ell &= \sup_{c\in(0,\infty)} \left[\widetilde{\Theta}^-(c) - \frac{1}{c}R^*\right]. \end{aligned} \tag{3.8}$$

And $\widetilde{P}^u$ and $\widetilde{P}^\ell$ are our numerical upper and lower bounds, respectively. They are numerical approximations to the analytical upper and lower bounds in Theorem 3.3.1.

3.4.2 Approximation Details

In this section we discuss the details in the aforementioned numerical procedure. Roughly speaking, we can categorize the steps into the following three phases.

- Pre-processing: This includes the step 1 and 2, for the choice of $\mathcal{A}$ and γ and the evaluation of R^*.
- Simulation: This is the step 3, which involves the computations of the QoI and the approximations of the integrals in Θ^+ and Θ^- (3.5).
- Post-processing: This is the step 4 and it computes the bounds by (3.8).

The pre-processing step is mostly a modeling procedure, as it requires one to gather all the available information to (1) determine the family of probability distributions $\mathcal{A}$, in which the true distribution ρ resides; and (2) choose a nominal distribution γ that serves as the "best guess" of the true distribution ρ. Obviously, these are problem dependent and shall not be further discussed in this paper. Here we will assume the choices have been made, that is, both $\mathcal{A}$ and γ are fixed, and then discuss how to compute the bounds. Once $\mathcal{A}$ and γ are chosen, the computation of R^* is straightforward by following its definition (3.4).

The major computations stem from Step 3, as it is here that the integrals in Θ^+ and Θ^- are computed and their approximations, $\widetilde{\Theta}^+$ and $\widetilde{\Theta}^-$ are obtained. This step requires the simulations of the QoI f, which involves the simulations of the underlying system and can be computationally intensive. Once $\widetilde{\Theta}^+$ and $\widetilde{\Theta}^-$ are obtained, the computations of the numerical bounds $\widetilde{P}^u$ and $\widetilde{P}^\ell$ in (3.8) require two separate optimization problems over a single real (and positive) parameter c. These can be carried out in a straightforward manner in the post-processing step.

One important issue to consider is that the functions Θ^+ and Θ^-, and respectively their approximations $\widetilde{\Theta}^+$ and $\widetilde{\Theta}^-$, are functions of the positive constant c. Their values at different values of c are needed in the step 4, via (3.8), to compute the numerical bounds. Consequently, it is vital importance that the QoI, f, does not need to repetitively evaluated at each value of c. Otherwise, the simulation cost can be prohibitively large. In the following, it will be made clear that this will not be the case.

There exist many methods to evaluate QoI in a stochastic system. Here we discuss two widely adopted approaches. One is sampling-based methods and the other is surrogate-based methods.

Sampling-based Methods for Θ^+ and Θ^-

In sampling methods, the goal is to estimate the statistical moments of the QoI (3.2), when the input probability distribution is given. To accomplish this, one generates a set of samples of the input random variable z, according to its probability distribution, then solves the underlying system to obtain an ensemble of output samples for the QoI, and the statistics of the QoI can then be approximated using the statistics of the output ensemble. This provides a natural means of approximating the Θ^+ and Θ^- in (3.5).

Let $z^{(1)}, \ldots, z^{(M)}$ be a set of samples of z, and $f^{(k)} = f(z^{(k)})$, $k = 1, \ldots, M$, be the corresponding solutions of the QoI. We then construct approximations of Θ^+ and Θ^- in the following way. For any fixed $c > 0$,

$$\widetilde{\Theta}^+(c) = \frac{1}{c} \log \left(\sum_{k=1}^{M} w^{(k)} e^{cf^{(k)}} \right), \qquad \widetilde{\Theta}^-(c) = -\frac{1}{c} \log \left(\sum_{k=1}^{M} w^{(k)} e^{-cf^{(k)}} \right), \tag{3.9}$$

where $w^{(k)}$, $k = 1, \ldots, M$, are the corresponding integration weights at the samples $z^{(k)}$. The weights are chosen in a way that $\{z^{(k)}, w^{(k)}\}_{k=1}^{M}$ construct an approximation to integrals of functions. That is, for a integrable function $g(z)$,

$$\int_{\Omega} g(z) d\gamma(z) \approx \sum_{k=1}^{M} w^{(k)} g(z^{(k)}). \tag{3.10}$$

An obvious example is Monte Carlo sampling, where $\{z^{(k)}\}$ are drawn randomly from the distribution γ and $w^{(k)} = 1/M, \forall k$, become equal weights. For most other integration rules, e.g. cubature rules, sparse grids, etc., the weights are in general not equal.

It is important to note that the samples $\{z^{(k)}, f^{(k)}\}$ can be repetitively used in the evaluation of (3.9) for any values of c. Hence, the sampling of the QoI is required only once.

Surrogate Methods

Another type of widely used methods for stochastic simulations are surrogate methods. In surrogate methods, one seeks to construct an approximation to the QoI directly, i.e., find $\widetilde{f}(z)$ such that $\|f(z) - \widetilde{f}(z)\| \ll 1$, in a properly defined norm. Typically, the norm is a weighted norm with respect to γ, the given (nominal) distribution. For example, the methods based on generalized polynomial chaos (gPC) often construct approximations in $L^2_\gamma(\Omega)$ norm.

There exist many methods to construct the surrogate. For example, stochastic Galerkin, least-square type regression, sparse grid interpolation, to name a few. We shall not discuss further on this and leave the construction of the surrogate as an user's choice. Once the surrogate is obtained, it is usually in an explicitly defined analytical form, e.g., a polynomial series. And the approximations of Θ^+ and Θ^- can be obtained by replacing f in the integrals by the surrogate $\widetilde{f}$, i.e.,

$$\widetilde{\Theta}^+(c) = \frac{1}{c}\log\int_\Omega e^{c\widetilde{f}}d\gamma, \qquad \widetilde{\Theta}^-(c) = -\frac{1}{c}\log\int_\Omega e^{-c\widetilde{f}}d\gamma. \tag{3.11}$$

If $\widetilde{f}$ is an accurate approximation to f, this will lead to an accurate approximation to $\Theta^+(c)$ and $\Theta^-(c)$, respectively. Again, we remark that, of the surrogate $\widetilde{f}$ needs to be constructed only once. It can then be used repetitively in the evaluations of $\widetilde{\Theta}^+(c)$ and $\widetilde{\Theta}^-(c)$ for any values of c.

Due to its flexibility in complex systems, from here on we will primarily focus on the sampling-based methods.

3.4.3 Properties of the Numerical Bounds

We now summarize the above discussion: For any given problem, we first utilize all the available information to choose a family of probability distributions $\mathcal{A}$, where the true (and unknown) distribution ρ resides, and a nominal distribution γ as our best-guess for ρ. We then conducts a one-time stochastic forward problem solver, using γ, for the underlying system to approximate the QoI. More specifically, we seek to find $\widetilde{\Theta}^+(c)$ and $\widetilde{\Theta}^-(c)$ approximate $\Theta^+(c)$ and $\Theta^-(c)$, respectively, either through a sampling method in the form of (3.9) or a surrogate method in the form of (3.11). Finally, a straightforward optimization problem over the single parameter c is carried out, in the form of (3.8), to find the approximate bounds $\widetilde{P}^u$ and $\widetilde{P}^\ell$.

It is important to note that, even though $\widetilde{P}^u$ and $\widetilde{P}^\ell$ are (accurate) approximations to the analytical bounds P^u and P^ℓ in (3.7), respectively. They are not bounds. The established analytical bound (3.6), $P^\ell \leq \int f d\rho \leq P^u$, does not lead to $\widetilde{P}^\ell \leq \int f d\rho \leq \widetilde{P}^u$, because of the numerical approximation errors involved. Let

$$\epsilon^u = \left| P^u - \widetilde{P}^u \right|, \qquad \epsilon^\ell = \left| P^\ell - \widetilde{P}^\ell \right| \tag{3.12}$$

be the numerical errors, then we expect the bounds to be

$$P^\ell - \epsilon^\ell \leq \int f d\rho \leq P^u + \epsilon^u. \tag{3.13}$$

These bounds, rather than the analytical ones from Theorem 3.3.1, should be the useful ones in practice.

Existence

We first establish the solvability of the optimization problem (3.8) via the following result.

Proposition 3.4.1 *Let $\widetilde{\Theta}^+$ and $\widetilde{\Theta}^-$ be defined as in* (3.9) *and R^* be a positive constant. Then, the following conclusions hold.*

(a) $\widetilde{\Theta}^+(c)$ *is differentiable and nondecreasing for* $c \geq 0$. *There is a unique* $c \in (0, \infty]$ *at which* $\widetilde{\Theta}^+(c) + \frac{1}{c}R^*$ *attains a local minimum. The minimum occurs at* $c = \infty$ *means that* $\widetilde{\Theta}^+(c) + \frac{1}{c}R^* > \lim_{c\to\infty} \widetilde{\Theta}^+(c)$ *for a well-defined* $\lim_{c\to\infty} \widetilde{\Theta}^+(c)$*;*

(b) $\widetilde{\Theta}^-(c)$ *is differentiable and nonincreasing for* $c \geq 0$. *There is a unique* $c \in (0, \infty]$ *at which* $\widetilde{\Theta}^-(c) - \frac{1}{c}R^*$ *attains a local maximum. The maximum occurs at* $c = \infty$ *means that* $\widetilde{\Theta}^-(c) - \frac{1}{c}R^* < \lim_{c\to\infty} \widetilde{\Theta}^-(c)$ *for a well-defined* $\lim_{c\to\infty} \widetilde{\Theta}^-(c)$*;*

Proof Let $H(c) = c\widetilde{\Theta}^+(c) = \log\left(\sum_{k=1}^M w^{(k)} e^{cf^{(k)}}\right)$, then $H(0) = \log(\sum_{k=1}^M w^{(k)}) = 0$, and

$$H'(c) = \frac{\sum_{k=1}^M w^{(k)} f(z^{(k)}) e^{cf(z^{(k)})}}{\sum_{k=1}^M w^{(k)} e^{cf(z^{(k)})}},$$

and

$$H'' = \frac{(\sum_{k=1}^M w^{(k)} f^2(z^{(k)}) e^{cf(z^{(k)})})(\sum_{k=1}^M w^{(k)} e^{cf(z^{(k)})}) - (\sum_{k=1}^M w^{(k)} f(z^{(k)}) e^{cf(z^{(k)})})^2}{(\sum_{k=1}^M w^{(k)} e^{cf(z^{(k)})})^2}.$$

By rewriting $w^{(k)} f(z^{(k)}) e^{cf(z^{(k)})}$ as $\left(e^{\frac{cf(z^{(k)})}{2}} \sqrt{w^{(k)}}\right)\left(e^{\frac{cf(z^{(k)})}{2}} f(z^{(k)}) \sqrt{w^{(k)}}\right)$ and using the discrete form of Hölder's Inequality, we obtain $H'' \geq 0$. By observing

$$(\widetilde{\Theta}^+(c))' = \frac{H'(c)c - H(c)}{c^2}$$

and considering $r(c) = H'(c)c - H(c)$, we obtain $r'(c) = H''(c)c \geq 0$. Since $H'(0) = \sum_{k=1}^M w^{(k)} f(z^{(k)}) < \infty$, we conclude that $r(c)$ is increasing and $r(0) = 0$, $r(c) \geq 0$. Thus $(\widetilde{\Theta}^+(c))' \geq 0$, and $\widetilde{\Theta}^+(c)$ is nondecreasing for $c \geq 0$.

Since

$$\frac{d}{dc}\left[\frac{1}{c}R^* + \frac{1}{c}H(c)\right] = \frac{1}{c^2}\left[-R^* + cH'(c) - H(c)\right] = \frac{1}{c^2}\left[-R^* + r(c)\right],$$

we let $K = \lim_{c\to\infty} r(c)$, which exists because of monotonicity. If $K = \infty$ or $0 \leq R^* < K$, then there is a unique solution to $r(c) = R^*$. If $R^* \geq K$, $\frac{R^*}{c} + \Theta_S^+$ is monotone and decreasing for $c \geq 0$, and since $\widetilde{\Theta}^+(c) \geq H'(0) > -\infty$, there is a well-defined limit $\lim_{c\to\infty} \widetilde{\Theta}^+(c)$ that is necessarily the minimum.

Part (b) is proved in a similar manner. ■

Error Analysis

Here we present an analysis on the numerical errors in the approximation of the upper and lower bounds (3.8). We present the details on the analysis of the upper bound P^u. The procedure is carried out on the lower bound P^ℓ in exactly the same way. For notational convenience, we define for the integral of e^{cf} and its numerical approximation as, respectively,

$$I_f(c) = \int e^{cf} d\gamma, \qquad \widetilde{I}_f(c) = \sum_{i=1}^{M} w^{(i)} e^{cf(z^{(i)})}. \tag{3.14}$$

The relatively error of the approximation is

$$e_f(c) = \frac{I_f(c) - \widetilde{I}_f(c)}{I_f(c)}. \tag{3.15}$$

We also introduce

$$\begin{aligned} &D^+ = \{c : \Theta^+(c) < \infty\}, && D^- = \{c : \Theta^-(c) < \infty\}, \\ &g^+(c) = c\Theta^+(c), && g^-(c) = c\Theta^-(c), \end{aligned} \tag{3.16}$$

where Θ^+ and Θ^- are defined in (3.5). We assume that the interior of D^+ and D^-, $(D^+)^o$ and $(D^-)^o$, respectively, are non-empty. Note that both $I_f(c)$ and $e_f(c)$ are well-defined in $(D^+)^o$, and $\widetilde{I}_f(c)$ is well defined for all $c \geq 0$ from Proposition 3.4.1 . Consider

$$I'_f(c) = \int f e^{cf} d\gamma, \qquad \widetilde{I}'_f(c) = \sum_{i=1}^{M} w^{(i)} f(z^{(i)}) e^{cf(z^{(i)})}, \tag{3.17}$$

where it is obvious that $\widetilde{I}'_f(c)$ is an approximation of $I'_f(c)$ using the same sampling rule. It is obvious that $I'_f(c)$ is also well-defined in $(D^+)^o$. (See, for example, [22].) We immediately have

$$g^+(c)' = I'_f(c)/I_f(c), \qquad g^-(c)' = I'_f(-c)/I_f(-c). \tag{3.18}$$

From the previous work in [21] (Appendix), it is known that $g^+(c)'$ is nondecreasing on D^+. We now make the following two assumptions:

(A) The relative slope of $I_f(c)$ is bounded below $+\infty$. More specifically,

$$g^+(c)' = \frac{I'_f(c)}{I_f(c)} = L^+ < +\infty, \qquad \text{as } c \to \sup(D^+). \tag{3.19}$$

(B) The relative error of the chosen sampling integration rule is less than 100%, that is, for any integrable function ψ satisfying $\int \psi d\gamma < \infty$, we have

$$\left| \frac{\int \psi d\gamma - \sum_{i=1}^{M} w^i \psi(z^{(i)})}{\int \psi d\gamma} \right| \leq E < 1. \tag{3.20}$$

Theorem 3.4.1 *Under the assumptions (A) and (B), we have*

$$\epsilon^u \triangleq \left| P^u - \widetilde{P}^u \right| \leq \max \left\{ E \left| \int f d\gamma \right|, L^+ \frac{2E}{1-E} \right\}, \tag{3.21}$$

where P^u is the analytical upper bound (3.7) *and $\widetilde{P}^u$ its approximation by* (3.8) *using a sampling integration rule with relative error E.*

Proof Under Assumption (A), we obtain $B^+ = \sup\{D^+\} = +\infty$. If otherwise $B^+ < \infty$, then by monotone convergence $g^+(c) \uparrow +\infty$, as $c \uparrow B^+$. This implies $g^+(c)' \to \infty$ as c grows and contradicts the Assumption (A) on the boundedness of $g^+(c)'$. Since D^+ is connected on $\mathbb{R}^+$ and $(D^+)^o$ is non-empty, we have for each $c \in [0, +\infty)$, $\int e^{cf} d\gamma < +\infty$ and $\int f e^{cf} d\gamma < \infty$.

Let us now consider

$$h^+(c) = \frac{1}{c} \left(\log(\sum_{i=1}^{M} w^{(i)} e^{cf(z^{(i)})}) - \log(\int e^{cf} d\gamma) \right) = \frac{1}{c} \log(1 - e_f(c)). \tag{3.22}$$

By the l'Hopital's rule, we have

$$\lim_{c \to 0} h^+(c) = \lim_{c \to 0} = \frac{e'(c)}{1 - e(c)} = \lim_{c \to 0} e'(c) = \int f d\gamma - \sum_{i=1}^{M} w^{(i)} f(z^{(i)}). \tag{3.23}$$

From assumption (B), we have $\lim_{c \to \infty} h^+(c) = 0$. Hence, $h^+(c)$ is well defined on $(0, +\infty)$, bounded and $\sup_{[0,\infty)}\{|h^+(c)|\}$ exits, and. By using the definitions for P^u (3.7), $\widetilde{P}^u$ (3.8), Θ^+ (3.5), and $\widetilde{\Theta}^+$ (3.9), we obtain

$$\begin{aligned}
P^u &= \inf_{c \in [0,\infty)} [\Theta^+(c) + \frac{1}{c} R^*] \\
&\leq \inf_{c \in [0,\infty)} [\widetilde{\Theta}^+(c) + \frac{1}{c} R^*] - \inf_{c \in [0,\infty)} [\widetilde{\Theta}^+(c) - \Theta^+(c)] \\
&= \widetilde{P}^u + \sup[\Theta^+(c) - \widetilde{\Theta}^+(c)] \\
&\leq \widetilde{P}^u + \sup_{c \in [0,\infty)} [|\Theta^+(c) - \widetilde{\Theta}^+(c)|] \\
&= \widetilde{P}^u + \sup_{c \in [0,\infty)} [|h^+(c)|].
\end{aligned}$$

Upon interchanging the places of P^u and $\widetilde{P}^u$, we obtain

$$\widetilde{P}^u \leq P^u + \sup_{c\in[0,\infty)} |h^+(c)|.$$

This immediately leads to

$$\left|P^u - \widetilde{P}^u\right| \leq \sup_{c\in[0,\infty)} [|h^+(c)|]. \tag{3.24}$$

Since $\lim_{c\to+\infty} h^+(c) = 0$, $h^+(c)$ can take its extremum either at $c_0 = 0$ or $0 < c_0 < +\infty$. We first consider the case of $c_0 \in (0, +\infty)$. Using the definition of $h^+(c)$ (3.22) and the fact that $h^+(c_0)' = 0$, we have

$$h^+(c_0)' = \frac{1}{c_0^2}\left[\frac{e_f'(c_0)}{1-e_f(c_0)}c_0 - \log(1-e_f(c_0))\right] = 0.$$

Consequently,

$$\begin{aligned}
|h(c_0)| &= \left|\frac{e'(c_0)}{1-e(c_0)}\right| = \left|\frac{I_f'(c_0)\widetilde{I}_f(c_0) - I_f(c_0)\widetilde{I}_f'(c_0)}{I_f(c_0)\widetilde{I}_f(c_0)}\right| \\
&\leq \left|\frac{I_f'(c_0)\widetilde{I}_f(c_0) - I_f(c_0)\widetilde{I}_f'(c_0)}{(1-E)I_f(c_0)^2}\right| \\
&\leq \frac{1}{1-E}\left|\frac{I_f'(c_0)}{I_f(c_0)}\right|\left(\left|\frac{I_f(c_0)-\widetilde{I}_f(c_0)}{I_f(c_0)}\right| + \left|\frac{I_f'(c_0)-\widetilde{I}_f'(c_0)}{I_f'(c_0)}\right|\right) \\
&\leq L^+\frac{2E}{1-E},
\end{aligned} \tag{3.25}$$

where the assumption (B) has been used. We now consider the case of $c_0 = 0$. Then by using (3.23), we have

$$\sup_{c\in[0,\infty)} |h^+(c)| = |h^+(0)| = \left|\int f d\gamma - \sum_{i=1}^{M} w^i f(z^{(i)})\right| \leq E\left|\int f d\gamma\right|. \tag{3.26}$$

The main conclusion then holds by combining the two cases. ■

The exactly same procedure can be repeated to the lower bound $\widetilde{P}^\ell$ after making an assumption similar to the assumption (A),

(A') The relative slope of $I_f(-c)$ is bounded above $-\infty$. More specifically,

$$g^-(c)' = \frac{I_f'(-c)}{I_f(-c)} > L^- > -\infty, \qquad \text{as } c \to \sup(D^-). \tag{3.27}$$

Theorem 3.4.2 *Under assumptions* (A') *and (B), we have*

$$\epsilon^{\ell} \triangleq \left| P^{\ell} - \widetilde{P}^{\ell} \right| \leq \max \left(E \left| \int f d\gamma \right|, L^{-} \frac{2E}{1-E} \right), \tag{3.28}$$

where P^{ℓ} *is the analytical lower bound* (3.7) *and* $\widetilde{P}^{\ell}$ *its approximation by* (3.8) *using a sampling integration rule with the relative error* E*.*

By combining these results together, Theorem 3.3.1, 3.4.1 and 3.4.2, we obtain estimates of the practical bounds in the form of (3.13).

3.5 Numerical Examples

In this section we provide examples to illustrate the computation of the numerical bounds. From the discussion above, it is obvious that the procedures presented here apply to quite general problems, as long as one has a well established deterministic code that allows sampling of the underlying system. Hence, we choose rather simple problems here for illustration purpose. For more complicated problems, one obviously needs to develop specific numerical solvers first.

3.5.1 $f(Z) = ln(Z)$

Consider the function $\mathcal{Q} = f(Z)$ where Z is a random variable with certain distribution $\rho(z)$ on $[a, b]$ where a and b are both known. The quantity of interest is the expectation of $\mathcal{Q}$, $\int f(z) d\rho$. After an initial study, we know ρ is in a family of Beta distributions on [0.6,1.5] with parameters $\alpha \in [1, 3.3]$, and $\beta \in [1, 3.3]$, denoted by $\mathcal{A}$. Then we choose the nominal distribution $\gamma(z)$ to be the uniform distribution on [0.6,1.5], and calculate R^*, the supremum of relative entropy between $\gamma(z)$ and any probability distribution in $\mathcal{A}$. In this case, R^* is 0.3. We move on to approximate Θ^+ and Θ^- with Monte Carlo and collocation methods using direct simulations from $\gamma(z)$. In this case, we can analytically write out the bounds of $\mathcal{Q} \circ I$, thus after obtaining the numerical bounds, we finally check the inequality in (4.21). Left

of Figure 3.1 demonstrates the curves of $\Theta_S^+ + \frac{1}{c}R^*$ and $\Theta_S^- - \frac{1}{c}R^*$ using N= 1,3,5 gauss quadrature points and we can see that each curve has one critical point that either the supremum or the infimum. Right of Figure 3.1 contains two curves: the error between analytical and numerical bounds and the right hand side of inequality (4.21), where both curve are in log scale. And the x-axis represents the number of gauss points used in numerical approximation. It shows the bound error is controlled by the right hand side of (4.21). Figure 3.2 shows results of $\Theta_M^+ + \frac{1}{c}R^*$ and $\Theta_M^- - \frac{1}{c}R^*$ using Monte Carlo simulation with 10,100,1000 sample points respectively, together with error plot on the right. Both axis are in log scale.

3.5.2 Stochastic Differential Equation

Epistemic Uncertainty in Diffusion

Let W_t be a standard Wiener process. We consider in this example the stochastic differential equation given by:

$$dX(t) = \lambda X(t)dt + (\mu + \psi)X(t)dW_t, \qquad X(0) = 1, \tag{3.29}$$

where ψ is a noise with epistemic distribution in the diffusion part may be resulted by derivation and observation, $\lambda = 1$ and $\mu = 0.2$ are constants. The quantity to our interest is the expectation of the solution at time 1,i.e. $E(X(1))$. In this case, our candidate family $\mathcal{A}$ consists of Beta distributions on [-1,1] with parameters $\alpha \in [1, 1.6]$ and $\beta \in [1, 1.6]$. And we choose the nominal distribution γ as the uniform distribution on [-1,1] and the corresponding R^* is calculated to be 0.065. Euler algorithm is applied to approximate the solution and Brownian motion is simulated with stepsize to be $\frac{1}{2^5}$, then the approximated solution $\hat{X}_1$ is a function of W and ψ, where W is a random vector of size 2^5. Then the numerical bounds of the $\mathcal{Q} \circ \mathcal{I}$ are evaluated by methods introduced in section (4)

Our implementation includes:

(1.1)First, we calculated the numerical bounds using Λ_1^+ and Λ_1^- by Monte Carlo

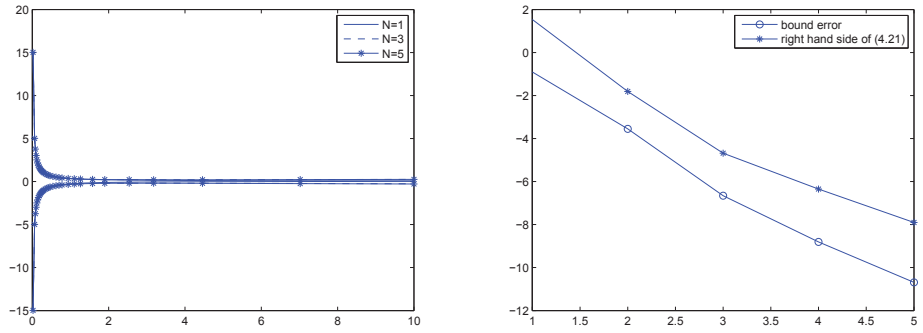

Figure 3.1. Upper bound and lower bound of expectation of $f(z)$ $\mathbb{E}[f]$ implemented by collocation method Left: Bound curve ; Right: Error between numerical bound and analytical bound.

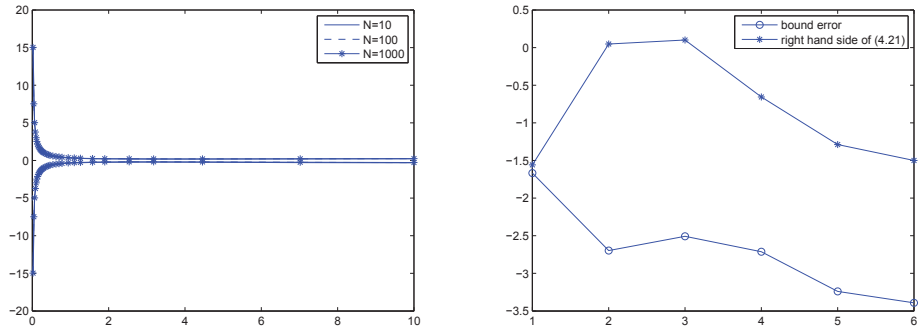

Figure 3.2. Upper bound and lower bound of expectation of $f(z)$ $\mathbb{E}[f]$ implemented by Monte Carlo simulation Left: Bound curve ; Right: Error between numerical bound and analytical bound.

method. We simulate different numbers of samples from Gaussian distribution for W, and nominal distribution $\gamma(z)$ for ψ The results are shown in (3.3) for different sample size.

(1.2)Second, we calculated the numerical bounds using Λ_1^+ and Λ_1^- with a mixed simulation method. We simulate different numbers of samples from Gaussian distribution for W, while fixing 10 collocation samples and weights for ψ from the nominal distribution $\gamma(z)$. The results are shown in (3.4) for different Monte Carlo sample size.

(1.3)Third, we calculated the numerical bounds using Λ^+ and Λ^- by Monte Carlo method. We simulate different numbers of samples from Gaussian distribution for W, and nominal distribution $\gamma(z)$ for ψ The results are shown in (3.5) for different sample size.

(1.4)Finally, we calculated the numerical bounds using Λ^+ and Λ^- with a mixed simulation method. We simulate different numbers of samples from Gaussian distribution for W, while fixing 10 collocation samples and weights for ψ from the nominal distribution $\gamma(z)$. The results are shown in (3.6) for different Monte Carlo sample size.

Figures from (1) and (2) are quite similar, and application of quadrature points dramatically reduces simulation cost without the loss of accuracy. The shapes of (3) and (4) are different from (1),(2) and the results are less sharper as shown in table (3.1) and (3.2).

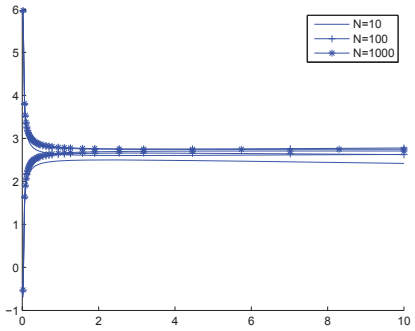

Figure 3.3. Numerical approximation of $\frac{R^*}{c} + \Lambda_1^+(c)$ and $-\frac{R^*}{c} + \Lambda_1^-(c)$ when same number of Monte Carlo samples are used for both W and ψ

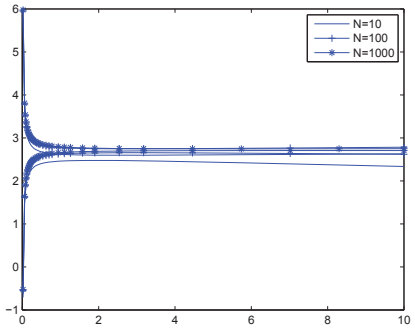

Figure 3.4. Numerical approximation of $\frac{R^*}{c} + \Lambda_1^+(c)$ and $-\frac{R^*}{c} + \Lambda_1^-(c)$ when Monte Carlo samples are used for W and 10 fixed collocation samples are used for ψ

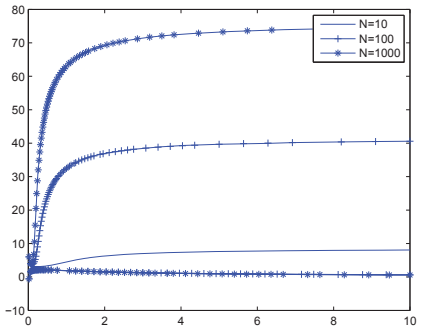

Figure 3.5. Numerical approximation of $\frac{R^*}{c} + \Lambda^+(c)$ and $-\frac{R^*}{c} + \Lambda^-(c)$ when same number of Monte Carlo samples are used for both W and ψ

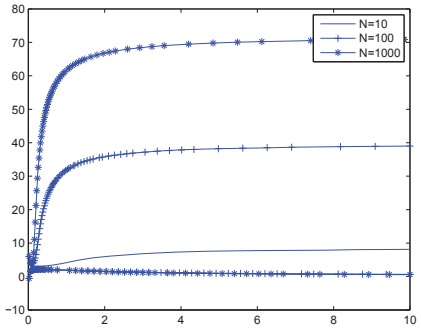

Figure 3.6. Numerical approximation of $\frac{R^*}{c}+\Lambda^+(c)$ and $-\frac{R^*}{c}+\Lambda^-(c)$ when Monte Carlo samples are used for W and 10 fixed collocation samples are used for ψ

Table 3.1
Comparison of $[Q_1^l, Q_1^u]$ and $[Q^l, Q^u]$ when Monte Carlo samples are used for W and 10 fixed collocation samples are used for ψ

Case	Λ_1, N=1000	Λ, N=1000	Λ_1, N=100	Λ, N=100	Λ_1, N=10	Λ, N=10
Lower bound	2.7067	2.1554	2.6563	2.1399	2.4773	2.1157
Upper bound	2.7459	3.8936	2.7469	3.8140	2.5973	3.0273

Table 3.2
Comparison of $[Q_1^l, Q_1^u]$ and $[Q^l, Q^u]$ when same number of Monte Carlo samples are used for both W and ψ

Case	Λ_1, N=1000	Λ, N=1000	Λ_1, N=100	Λ, N=100	Λ_1, N=10	Λ, N=10
Lower bound	2.7068	2.1685	2.6588	2.1497	2.4991	2.1087
Upper bound	2.7455	3.8764	2.7469	3.7794	2.6013	3.0749

Epistemic Uncertainty in Drift

We continue to consider the classical differential equation:

$$dX(t) = \lambda X(t)dt + \mu X(t)dW_t, \quad X(0) = 1 \tag{3.30}$$

with solution at time T:

$$X(T) = exp((\lambda - \frac{\mu^2}{2})T + \mu W_T) \tag{3.31}$$

However, the coefficient λ is no longer a constant but a variable with epistemic uncertainty. Again, the quantity to our interest is expectation of the solution at time $T = 1$. We assume after initial study, we get to know that λ follows a normal distribution with the mean $\tilde{\mu} \in [0.8, 1.2]$ and standard deviation $\sigma \in [0.05, 0.15]$. And we continue to select $\gamma(z) = \mathcal{N}(1, 0.1)$ as the nominal distribution since this is almost the best guess of the true probability. And R^* is obtained as 2.3181. With R^* and the nominal distribution, we implement the Monte Carlo method to approximate the sharper bounds $[Q_1^l, Q_1^u]$ with different number of sample points:100,1000,10000. The bounds result is shown in table 3.3 and we also plot the target function in 3.7 for sample size equals 1000. Figure 3.8 shows the plot of $\frac{R^*}{c} + \Lambda^+(c)$ and $-\frac{R^*}{c} + \Lambda^-(c)$. Although, both curves has a unique supremum or infimum (the infimum of the upper curve is obtained at a point close to 0), the bounds from them are too general to be useful.

Table 3.3
Monte Carlo Approximations for $[Q_1^l, Q_1^u]$ with sample size 10000,1000,100.

Case	N=10000	N=1000	N=100
Lower bound	2.2207	2.4451	2.9059
Upper bound	3.4508	3.8568	4.6622

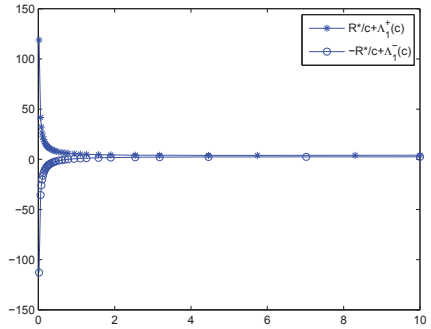

Figure 3.7. Numerical approximation of $\frac{R^*}{c} + \Lambda_1^+(c)$ and $-\frac{R^*}{c} + \Lambda_1^-(c)$ when using 1000 samples for both W and ψ

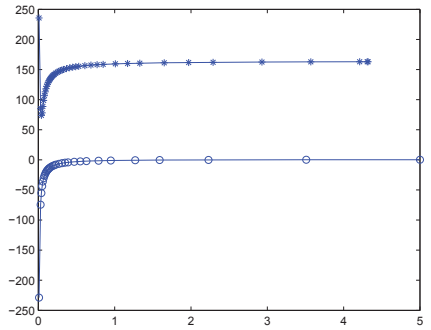

Figure 3.8. Numerical approximation of $\frac{R^*}{c} + \Lambda^+(c)$ and $-\frac{R^*}{c} + \Lambda^-(c)$ when using 1000 samples for both W and ψ

3.6 Conclusion

This paper develops efficient and accurate algorithms for bounds for quantity of interest in problems subject to epistemic uncertainty using generalized polynomial chaos (gPC) expansions, stochastic collocation, Monte Carlo simulation. Given the quantity of interest (Q$\circ$ I) and a nominal distribution, we use above methods to estimate the integration of Q$\circ$ I under the nominal distribution;the surrogate then replace the real $\int Q \circ I$ in the bounds and define an approximate bounds. We study the error between the approximate bounds and the true bounds, and adjust the numerical bounds. In particular, we show that the error of the numerical bounds are controlled by a function of the relative error from the estimation of $\int Q \circ I$, which is obtained by using gPC, stochastic collocation, Monte Carlo and etc.

Error properties of our algorithm are then demonstrated numerically:first on a log function with different nominal distributions, then on a stochastic differential equation. In both cases, the models contain epistemic uncertainties and the quantity of our interest is affected by the uncertainty. We follow the computing schemes to find numerical bounds for the quantities of our interest, calculate the theoretical bounds, compare the numerical bounds and theoretical bounds and the difference is within the range of our error estimation.

4. BAYESIAN ANALYSIS FOR UNCERTAINTY QUANTIFICATION

4.1 Introduction

Inverse problem is a framework to study the scientific model and its inputs using obtained observations, the model outputs. In the case of deterministic model such as seismic model the problem is often transformed to an optimization problem. In most cases, the model is stochastic either contains stochastic/random inputs or in a form that is not completely known. Besides studying the model itself, it's also important to study, quantify and reduce the random effects of model inputs on the outputs. It is an important topic of Uncertainty Quantification since it's vital for decision making, experimental design and model prediction. Bayesian framework [23–27] is a natural method to solve the problem especially when considering the fact that the observations inevitably contain noise.

A general procedure for quantifying the uncertainties under Bayesian framework will first impose the inputs with a prior which often comes from "experts". Based on the equipped prior and observations, a posterior distribution of the inputs is established using Baye's rule. Finally, a series of simulations according to the posterior are conducted to study the statistical properties/moments of the inputs and the outputs. The procedure is conceptionally natural and straightforward, however the implementation can be computationally expensive and challenging. First, the choice of priors brings more structure to the model which is subjective and can introduce bias to the model. Many researchers have been working on selecting priors automatically and robustly. Second, in most cases, the posterior distribution does not have an analytical form and often contains an integration part (the likelihood) which makes the simulation challenging. When there are multiple variables in the model, the pos-

terior is multi-variate distribution and can not be readily approximated. To solve this problem, many researchers have sought more efficient sampling algorithms from the posterior [28–30]. For example, Markov chain Monte Carlo (MCMC) [30–33] has greatly speed up the simulation of posterior. One key factor of MCMC is the form of the transition kernel which determines how long we should wait until our samples start to follow the posterior distribution. Third,the simulation of posterior can be overwhelmingly difficult in that the forward problem is usually in the form of PDEs and it can take long time to solve at one single sample point. Many efforts in reducing the computing cost have been taken. Most of them focused on finding a much less expensive surrogate [34–36] of true forward model. One main method is to use generalized polynomial chaos (gPC) [37] to approximate the forward model solution as a linear combination of the orthogonal polynomials under the prior distribution. The forward model is completely determined by the coefficients. Stochastic Galerkin method [38, 39] and stochastic collocation [40–43] methods have been widely used to determine the coefficients and study the propagation of uncertainties [44]. Forth, when new information about data or structure is available, it is beneficial to propose a new prior and simulate the system under the new prior. In this case, we will meet all above challenges again. Although many results for above three challenges have reduced the computing expense, when we want to simulate posterior for multiple priors,a more efficient algorithm is preferred.

This work is based on two more recent work of [10] and [45]. Work [10] proposes an approach for epistemic uncertainty analysis. The method consists of three components: (1) encapsulation of the epistemic uncertain inputs by a bounded or unbounded domain, i.e., a hypercube; (2) solution of the encapsulation problem, which is the original governing equations defined in the encapsulation domain; and (3) post-processing of the resulting numerical solution for its statistics, whenever the probability distribution of the epistemic inputs is known, or assumed, a posterior. Work [45] applied gPC stochastic collocation to construct posterior surrogates for efficient Bayesian inference in inverse problems and provided a rigorous convergence

proof of the approximate posterior distribution to the true posterior distribution is established and its asymptotic convergence rate obtained.

Our work extends the Uncertainty Quantification scheme in [10] and [9] to uncertainty quantification of inverse problem. Based on "expert's" prior , we first establish a surrogate of the forward model under the original prior distribution using gPC stochastic collocation. Then, after a new prior is equipped with the system, instead of simulating and computing the posterior again, an estimation based on new prior and previous forward model surrogate is proposed. Finally, a bound of the difference between our estimate and the true posterior is provided.

4.2 Problem Setup

We begin with our forward model and the Bayesian inference. Let's assume our forward model is in form of the following general stochastic partial differential equation

$$\begin{cases} u_t(x,t,Z) = \mathcal{L}(u) & D \times (0,T] \times I_Z, \\ \mathcal{B}(u) = 0 & \partial D \times (0,T] \times I_Z, \\ u = u_0 & D \times \{t=0\} \times I_Z, \end{cases} \tag{4.1}$$

First, in the random space $Z = (Z_1, ..., Z_d) \in I_Z \subseteq \mathbb{R}^d$,$d \geq 1$, represent the uncertainties in the model inputs and can be treated as coefficients of the system. In the deterministic space, $x = (x_1, ..., x_d)$ is the space coordinate in the physical space $D \subset \mathbb{R}^l$, $l = 1, 2, 3$. T>0 be a real number that represents upper bound of the time domain. $\mathcal{B}$ is the boundary condition operator of the (nonlinear) differential operator $\mathcal{L}$. u_0 is the initial condition. The solution of the system u is a function of the uncertainties Z

$$u(x,t,Z) : \bar{D} \times [0,T] \times I_Z \to \mathbb{R}^{n_u},$$

and is also a random variable. Assume a function of the solution $g(u), \mathbb{R} \to \mathbb{R}^{n_d}$ is observed, that is,

$$d^t = g(u) \in \mathbb{R}^{n_d}$$

The corresponding "forward model" $G : I_Z \to \mathbb{R}^{n_d}$ is then a function that maps the random inputs Z to the observation d^t:

$$d^t = G(Z) \triangleq g \circ u(Z). \tag{4.2}$$

One difficulty in the inverse problem is that our observation inevitably contains perturbation of noise, so instead, we should consider following true observations:

$$d = d^t + e = G(Z) + e, \tag{4.3}$$

where $e \in \mathbb{R}^{n_d}$ is the observation noise enters our observations. For simplicity, we make two assumptions about the observations: 1. the distribution of e is independent of the random inputs Z; 2 the noise for different observations are i.i.d. 3. For each e, it may be of multiple dimensions, and all the components are independent of each other.

Then according to the Baye's rule, posterior of the uncertainty Z conditioned on the observations d which is denoted by $\pi(z \mid d)$ can be written as:

$$\pi(z \mid d) = \frac{\pi(z \mid d)\pi(z)}{\int \pi(z \mid d)\pi(z)dz},$$

where $\pi(d \mid z)$ is called likelihood function; and the bottom is termed the marginal likelihood. For simplicity, we will denote the posterior density $\pi(z \mid d)$ with $\pi^d(z)$ and likelihood function $\pi(d \mid z))$ with L(z). Now the posterior becomes,

$$\pi^d(z) = \frac{L(z)\pi(z)}{\int L(z)\pi(z)dz} \tag{4.4}$$

Based on assumptions of observation error, the likelihood function is

$$L(z) \triangleq \pi(d \mid z) = \prod_{i=1}^{n_d} \pi_{e_i}(d_i - G_i(z)). \tag{4.5}$$

Most of the existing studies adopt one prior to the Bayesian framework.Then by using gPC based stochastic collocation method, the following is used to approximate posterior distribution in (2.3)

$$\pi_N^d(z) = \frac{L_N^\pi(z)\pi(z)}{\int L_N^\pi(z)\pi(z)dz} \tag{4.6}$$

$$L_N^{\pi}(z) \triangleq \pi_N(d \mid z) = \prod_{i=1}^{n_d} \pi_{e_i}(d_i - G_{N,i}^{\pi}(z)). \tag{4.7}$$

where $G_{N,i}^{\pi}$ is the i-th component of G_N^{π}, the gPC expansion of G under measure π.

But there are cases where the priors need to be updated as more information arrives or a sensitivity analysis of the prior is needed. Therefore, whenever a new prior is given, we need to approximate the posterior using gPC expansion under that prior. However, the explicit form of G which is a random function is unknown for most systems, and to establish the surrogate requires simulation under the prior. Posterior estimate is thus challenging due to highly time consuming simulation. In this paper, we introduce a new algorithm to approximate posterior in 4.6, and establish its efficiency.

4.3 Methodology

We now present our scheme for solving Bayesian problem subject to epistemic uncertain inputs, which is also based on generalized polynomial chaos expansions (gPC). The proposed methodology is a three-step procedure that involves using stochastic collocation to approximate forward model under initial prior, approximate the new posterior, and post processing the results.

4.3.1 Initial forward model approximation

At this point, we only have small amount of experience for the random input distribution which gives us the initial prior. The first task is to identify the support of the prior and form the forward model surrogate on the support. We denote the initial prior distribution by π^0 and its support by I_X. We now define the following forward problem:

$$\begin{cases} v_t(x,t,X) = \mathcal{L}(v) & D \times (0,T] \times I_X, \\ \mathcal{B}(v) = 0 & \partial D \times (0,T] \times I_X, \\ v = v_0 & D \times \{t=0\} \times I_X, \end{cases} \tag{4.8}$$

It is a governing system defined on the prior support I_X. We set up the corresponding "forward model" $F : I_X \to \mathbb{R}^{n_d}$ to describe the relation between the random parameters X and observable d^t:

$$d^t = F(X) \triangleq g \circ u(X). \tag{4.9}$$

Let $F_N^{\pi^0}$ be a numerical solution of the "forward model"(3.6) obtained using numerical scheme, such as in [45] , where N represents some parameters we adopt in the approximation process. For example, N can be the highest degree of a polynomial approximation, in which case, a higher N indicates a better approximation. An important property that our approximation needs to satisfy is that the L^p distance between the true forward and approximate model should converge to 0 as N increases, that is,

$$\epsilon_N \triangleq \|F - F_N^{\pi^0}\|_{L^p_{\pi^0}(I_X)} \to 0, \qquad N \to \infty.$$

Note that the L^p norm is defined by our prior in a functional space with following distance measure

$$(f, g)_{\pi^0} = \int_{I_X} f(s)g(s)\pi^0(s)ds,$$

4.3.2 Posterior Distribution approximate for new prior

Next we consider the procedure to approximate the posterior when a new prior $\pi(z)$ is given. We use I_Z to denote the support of $\pi(z)$, and we denote the super set and common set with

$$I^+ = I_Z \bigcup I_X, \qquad I^0 = I_Z \bigcap I_X,$$

and the symmetric difference is denoted by

$$I^- = I_Z \triangle I_X = I^+ \setminus I^0,$$

A critical requirement of the scheme is that:

$$P(z \in I^-) \leq \delta,$$

where δ is small enough, that is, to say the supports of the two priors do not differ much in a probability sense. Note that the governing system on the support of $\pi(z)$ is exactly the same with (4.1) and with the same forward model G(z) on I_Z. To approximate the new forward model G(z), instead of direct simulation and computation, we define following functions:

$$G_N^+(s) = \mathbb{I}_{I_X} F_N^{\pi^0}(s) \qquad s \in I^+, \tag{4.10}$$

$$G_N^{\pi^0}(s) = G_N^+(s)|_{I_Z} \qquad s \in I_Z, \tag{4.11}$$

(4.10) is the extension of initial surrogate onto the union set while (4.11) is its restriction onto the support of $\pi(z)$.

Under the new probability measure $\pi(z)$, we use the approximate gPC solution under π^0 to replace the true (but unknown) forward problem solution in Bayes' rule (4.4) and define the following approximate posterior probability,

$$\pi_N^{d,0}(Z) = \frac{L_N^{\pi^0}(z)\pi(z)}{\int L_N^{\pi^0}(z)\pi(z)dz,} \tag{4.12}$$

where $L_N^{\pi^0}$ is the approximate likelihood function defined as :

$$L_N^{\pi^0}(z) \triangleq \pi_N(d|z) = \prod_{i=1}^{n_d} \pi_{e_i}(d_i - G_{N,i}^{\pi^0}(z)). \tag{4.13}$$

where $G_{N,i}^{\pi^0}$ is the i-th component of $G_N^{\pi^0}$.

The key step is to find the surrogate under the new prior distribution. Two advantages of our surrogate include: 1. It usually takes simple forms such as polynomials. When simulation is performed. it goes into the likelihood and substitute the forward model, which can largely reduce the simulation cost. 2. Formulation of this surrogate does not introduce extra simulation cost and only a interval manipulation is need, since it's a generalized form of initial surrogate. For example, if the cost is C to find a single surrogate, then when we work on K different priors, we still need C instead of KC.

4.3.3 Convergence Study

To establish convergence of the algorithm, we first study convergence of $G_N^{\pi^0}$. Let us assume that the weight π^0(s) used in the encapsulation problem is non-vanishing in the interior of I^0 and $\pi(s)$ is the nominal prior distribution of Z on I_Z. We first introduce our denotations:

$$G^+(s) = \mathbb{I}_{I_Z} G(s) \qquad s \in I^+ \tag{4.14}$$

$$r(s) = \mathbb{I}_{I^0} \frac{\pi(s)}{\pi^0(s)} \qquad s \in I^+, \tag{4.15}$$

$$\pi^+(s) = \mathbb{I}_{I_Z}(s)\pi(s) \qquad s \in I^+ \tag{4.16}$$

$$\epsilon_N = \|F - F_N^{\pi^0}(X)\|_{L^p_{\pi^0}(I_X)}, \qquad p > 1. \tag{4.17}$$

together with definitions of $G_N^+(s)$ and $G_N^{\pi^0}(s)$ in (3.7)and (3.8),the following results can be established.

Lemma 4.3.1 *Assume the function r, defined in (3.12),satisfies*

$$C_{r,0} = \max_s r(s) < \infty.$$

Then

$$\|G_N^+ - G^+\|_{L^p_\pi}(I^+) \le C_{r,0}^{\frac{1}{p}} \epsilon_N + C_v \delta^{\frac{1}{p}}$$

Proof

$$\begin{aligned}
\|G_N^+ - G^+\|^p_{L^p_\pi(I^+)} &= \textstyle\int_{I^+} \sum_{i=1}^{n_d} |G_{N,i}^+ - G_i^+|^p \pi^+(s) ds \\
&= \textstyle\int_{I^0} \sum_{i=1}^{n_d} |G_{N,i}^+ - G_i^+|^p \pi^+(s) ds + \int_{I^-} \sum_{i=1}^{n_d} |G_{N,i}^+ - G_i^+|^p \pi^+(s) ds \\
&= \textstyle\int_{I^0} \sum_{i=1}^{n_d} |F_{N,i}^{\pi^0} - F_i|^p \pi^+(s) ds + \int_{I^- \bigcap I_Z} \sum_{i=1}^{n_d} |G_i|^p \pi(s) ds \\
&= \textstyle\int_{I^0} \sum_{i=1}^{n_d} |F_{N,i}^{\pi^0} - F_i|^p \pi^0(s) \frac{\pi(s)}{\pi^0(s)} ds + \int_{I^- \bigcap I_Z} \sum_{i=1}^{n_d} |G_i|^p \pi(s) ds \\
&\le \textstyle\int_{I^0} \sum_{i=1}^{n_d} |F_{N,i}^{\pi^0} - F_i|^p \pi^0(s) max_{s \in I^0} \frac{\pi(s)}{\pi^0(s)} ds + \int_{I^- \bigcap I_Z} \sum_{i=1}^{n_d} |G_i|^p \pi(s) ds \\
&\le C_{r,0} \epsilon_n^p + C_v^p \delta \\
&\le \left(C_{r,0}^{\frac{1}{p}} \epsilon_N + C_v \delta^{\frac{1}{p}} \right)^p
\end{aligned} \tag{4.18}$$

This completes the proof. ■

With this lemma, we can deduce the following theorem stating the convergence of $G_N^{\pi^0}$ to G:

Theorem 4.3.2 *Assume the function r, defined in (3.12),satisfies*

$$C_{r,0} = \max_s r(s) < \infty,$$

then

$$\|G_N^{\pi^0} - G\|_{L_\pi^p}(I_Z) \leq C_{r,0}^{\frac{1}{p}} \epsilon_N + C_v \delta^{\frac{1}{p}}.$$

Proof From lemma 1, we know:

$$\begin{aligned}
\|G_N^{\pi^0} - G\|^p_{L_\pi^p(I_Z)} &= \int_{I_Z} \sum_{i=1}^{n_d} |G_{N,i}^{\pi^0} - G_i|^p \pi(s) ds \\
&\leq \int_{I_Z} \sum_{i=1}^{n_d} |G_{N,i}^{+} - G_i^{+}|^p \pi^+(s) ds + \int_{I^+ \setminus I_Z} \sum_{i=1}^{n_d} |G_{N,i}^{+} - G_i^{+}|^p \pi^+(s) ds \\
&= \int_{I^+} \sum_{i=1}^{n_d} |G_{N,i}^{+} - G_i^{+}|^p \pi^+(s) ds \\
&= \|G_N^{+} - G^{+}\|^p_{L_\pi^p(I^+)} \\
&\leq \left(C_{r,0}^{\frac{1}{p}} \epsilon_N + C_v \delta^{\frac{1}{p}} \right)^p
\end{aligned} \tag{4.19}$$

■

To establish convergence of the Bayesian algorithm, we quantify the difference between the approximate posterior $\pi_N^{d,0}$ and the exact posterior π^d under measure π via Kullback-Leibler divergence. The Kullback-Leibler divergence (KLD) measures the difference between probability distributions and is defined, for probability $\pi^1(z)$ and $\pi^2(z)$, as

$$D(\pi^1 \| \pi^2) \triangleq \int \pi^1(z) log \frac{\pi^1(z)}{\pi^2(z)} dz.$$

It is always non-negative, and $D(\pi^1 \| \pi^2)$=0 when $\pi^1 = \pi^2$.

Theorem 4.3.3 *Assume that the observational error in (2.2) has an i.i.d. Gaussian distribution. If the nominal prior distribution of random parameter Z is $\pi(Z)$ and satisfies*

$$C_{r,0} = \max_s r(s) < \infty,$$

then the Kullback-Leibler divergence between the posterior probability $\pi_N^{d,0}(z)$ in (3.9) and the true posterior probability in (2.3) has the following bound,

$$\begin{aligned} D(\pi_N^{d,0}\|\pi^d) \ &\leq \frac{C_3 n_d}{2\sigma^2\gamma}\left(1+\frac{C_\gamma\left(C_{r,0}^{\frac{1}{2}}\epsilon_N+C_v\delta^{\frac{1}{2}}\right)^{n_d}}{1-C_\gamma\left(C_{r,0}^{\frac{1}{2}}\epsilon_N+C_v\delta^{\frac{1}{2}}\right)^{n_d}}\right)\left(C_{r,0}^{\frac{1}{2}}\epsilon_N+C_v\delta^{\frac{1}{2}}\right) \\ &+|log(1+C_\gamma\left(C_{r,0}^{\frac{1}{2}}\epsilon_N+C_v\delta^{\frac{1}{2}}\right)^{n_d})|. \end{aligned} \tag{4.20}$$

where C_γ and C_3 are constants.

Proof Let

$$\gamma=\int L(z)\pi(z)dz, \qquad \gamma_N=\int L_N^{\pi^0}(z)\pi(z)dz$$

Obviously, $\gamma>0$ and $\gamma_N>0$. By following the definitions of the likelihood functions L(z) in (2.9) and $L_N^{\pi^0}(z)$ in (3.10) and utilizing the fact that function e^{-x} is (uniformly) Lipschitz continuous for $x\geq 0$, i.e.,$|e^{-x}-e^{-y}|\leq\Lambda|x-y|$ for all $x,y>0$, where Λ is a positive constant, we have

$$\begin{aligned} |\gamma_N-\gamma| \ &=|\int\left(L(z)-L_N^{\pi^0}(z)\right)\pi(z)dz| \\ &\leq\prod_{i=1}^{n_d}\frac{1}{\sqrt{2\pi\sigma^2}}\int|e^{\frac{(d_i-G_{N,i}^{\pi^0}(z))^2}{2\sigma^2}}-e^{\frac{(d_i-G_i(z))^2}{2\sigma^2}}|\pi(z)dz \\ &\leq\prod_{i=1}^{n_d}\frac{\Lambda}{2\sigma^2\sqrt{2\pi\sigma^2}}\int|(d_i-G_{N,i}^{\pi^0}(z))^2-(d_i-G_i(z))^2|\pi(z)dz \\ &\leq\prod_{i=1}^{n_d}\frac{\Lambda}{2\sigma^2\sqrt{2\pi\sigma^2}}\|G_{N,i}^{\pi^0}-G_i\|_{L_\pi^2(I_Z)}\|2d_i-G_{N,i}^{\pi^0}-G_i\|_{L_\pi^2(I_Z)} \\ &\leq\prod_{i=1}^{n_d}C_1\|G_{N,i}^{\pi^0}-G_i\|_{L_\pi^2(I_Z)} \\ &\leq C_1\|G_N^{\pi^0}-G\|_{L_\pi^2(I_Z)}^{n_d} \\ &\leq C_1\left(C_{r,0}^{\frac{1}{2}}\epsilon_N+C_v\delta^{\frac{1}{2}}\right)^{n_d}, \end{aligned} \tag{4.21}$$

where Hölder's inequality has been used. Note the positive constant C_1 is independent of N. Next, we divide the above inequality by γ and require N to be sufficiently large, δ to be sufficient small, then we immediately have

$$0<1-C_\gamma\left(C_{r,0}^{\frac{1}{2}}\epsilon_N+C_v\delta^{\frac{1}{2}}\right)^{n_d}\leq\frac{\gamma_N}{\gamma}\leq 1+C_\gamma\left(C_{r,0}^{\frac{1}{2}}\epsilon_N+C_v\delta^{\frac{1}{2}}\right)^{n_d}, \tag{4.22}$$

and

$$|\frac{\gamma}{\gamma_N}-1|\leq\frac{C_\gamma\left(C_{r,0}^{\frac{1}{2}}\epsilon_N+C_v\delta^{\frac{1}{2}}\right)^{n_d}}{1-C_\gamma\left(C_{r,0}^{\frac{1}{2}}\epsilon_N+C_v\delta^{\frac{1}{2}}\right)^{n_d}}, \tag{4.23}$$

$$|log\frac{\gamma}{\gamma_N}| \leq |log(1 + C_\gamma \left(C_{r,0}^{\frac{1}{2}}\epsilon_N + C_v\delta^{\frac{1}{2}}\right)^{n_d})|, \tag{4.24}$$

where $C_\gamma = \frac{C_1}{\gamma}$ is independent of N and δ. Also,

$$\frac{\pi_N^{d,0}}{\pi^d} = \frac{L_N^{\pi^0}}{L}\frac{\gamma}{\gamma_N} = \frac{\gamma}{\gamma_N}\prod_{i=1}^{n_d}\frac{\pi_{e_i}(d_i - G_{N,i}^{\pi^0})}{\pi_{e_i}(d_i - G_i)} = \frac{\gamma}{\gamma_N}\prod_{i=1}^{n_d} exp\left(-\frac{(d_i - G_{N,i}^{\pi^0})^2 - (d_i - G_i)^2}{2\sigma^2}\right). \tag{4.25}$$

Therefore,

$$log\frac{\pi_N^{d,0}}{\pi^d} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n_d}[(d_i - G_{N,i}^{\pi^0})^2 - (d_i - G_i)^2] + log\left(\frac{\gamma}{\gamma_N}\right), \tag{4.26}$$

and

$$\begin{aligned} D(\pi_N^{d,0}\|\pi^d) &= \tfrac{1}{2\sigma^2\gamma_N}\textstyle\sum_{i=1}^{n_d}\int L_N^{\pi^0}(z)[(d_i - G_{N,i}^{\pi^0})^2 - (d_i - G_i)^2]\pi(z)dz + \tfrac{1}{\gamma_N}\int L_N^{\pi^0}(z)log\left(\tfrac{\gamma}{\gamma_N}\right)\pi(z \\ &= \tfrac{1}{2\sigma^2\gamma_N}\textstyle\sum_{i=1}^{n_d}\int L_N^{\pi^0}(z)[(d_i - G_{N,i}^{\pi^0})^2 - (d_i - G_i)^2]\pi(z)dz + log\left(\tfrac{\gamma}{\gamma_N}\right). \end{aligned} \tag{4.27}$$

Since both γ and γ_N are positive constants and $L_N^{\pi^0}(z)$ is bounded, i.e., $0 < L_N^{\pi^0}(z) \leq C_2$, we obtain immediately

$$\begin{aligned} D(\pi_N^{d,0}\|\pi^d) &\leq \tfrac{C_2}{2\sigma^2\gamma_N}\textstyle\sum_{i=1}^{n_d}\int|(d_i - G_{N,i}^{\pi^0})^2 - (d_i - G_i)^2|\pi(z)dz + |log\tfrac{\gamma}{\gamma_N}| \\ &\leq \tfrac{C_3}{2\sigma^2\gamma_N}\textstyle\sum_{i=1}^{n_d}\|G_{N,i}^{\pi^0} - G_i\|_{L_\pi^2(I_Z)} + |log\tfrac{\gamma}{\gamma_N}| \\ &\leq \tfrac{C_3 n_d}{2\sigma^2\gamma_N}\|G_N^{\pi^0} - G\|_{L_\pi^2(I_Z)} + |log\tfrac{\gamma}{\gamma_N}| \\ &\leq \tfrac{C_3 n_d}{2\sigma^2\gamma_N}\left(C_{r,0}^{\frac{1}{2}}\epsilon_N + C_v\delta^{\frac{1}{2}}\right) + |log\tfrac{\gamma}{\gamma_N}| \end{aligned} \tag{4.28}$$

Again, the Hölder's inequality has been used. By slightly rewriting above inequality, we have

$$\begin{aligned} D(\pi_N^{d,0}\|\pi^d) &\leq \tfrac{C_3 n_d}{2\sigma^2\gamma}\tfrac{\gamma}{\gamma_N}\left(C_{r,0}^{\frac{1}{2}}\epsilon_N + C_v\delta^{\frac{1}{2}}\right) + |log\tfrac{\gamma}{\gamma_N}| \\ &\leq \tfrac{C_3 n_d}{2\sigma^2\gamma}\left(1 + |\tfrac{\gamma}{\gamma_N} - 1|\right)\left(C_{r,0}^{\frac{1}{2}}\epsilon_N + C_v\delta^{\frac{1}{2}}\right) + |log\tfrac{\gamma}{\gamma_N}| \end{aligned} \tag{4.29}$$

By using inequality(3.19) and(3.20), we finally have

$$\begin{aligned} D(\pi_N^{d,0}\|\pi^d) &\leq \tfrac{C_3 n_d}{2\sigma^2\gamma}\left(1 + \frac{C_\gamma\left(C_{r,0}^{\frac{1}{2}}\epsilon_N + C_v\delta^{\frac{1}{2}}\right)^{n_d}}{1 - C_\gamma\left(C_{r,0}^{\frac{1}{2}}\epsilon_N + C_v\delta^{\frac{1}{2}}\right)^{n_d}}\right)\left(C_{r,0}^{\frac{1}{2}}\epsilon_N + C_v\delta^{\frac{1}{2}}\right) \\ &\quad + |log(1 + C_\gamma\left(C_{r,0}^{\frac{1}{2}}\epsilon_N + C_v\delta^{\frac{1}{2}}\right)^{n_d})|. \end{aligned} \tag{4.30}$$

■

From this theorem, we see that when the encapsulation range is quite close to the true prior distribution support, i.e, δ is small enough and N is large enough, we can use $\pi_N^{d,0}$ in (3.9) to approximate π^d, the true posterior probability since the Kullback-Leibler divergence between them is close to 0.

Now assume $n^d = 1$ and both "truncation error" and "numerical error" are well controlled , we have following result:

Corollary 4.3.4 *Suppose we have nominal probability π^1, and G_N is an approximation of G under π^0, such that:*

$$||G_N - G||_{L^2(\pi^1)} \leq \epsilon, \tag{4.31}$$

Then

$$D(\pi_{1,N}^d||\pi_1^d) \leq \eta \doteq \frac{C_3}{2\sigma^2\gamma_1(1 - C_\gamma^1\epsilon)}\epsilon + log(1 - \frac{C_1}{\gamma_1}\epsilon), \tag{4.32}$$

where $\pi_{1,N}^d = \frac{L_N(z)\pi^1(z)}{\int L_N(z)\pi^1(z)dz}$,$\pi_1^d = \frac{L(z)\pi^1(z)}{\int L(z)\pi^1(z)dz}$ and $\gamma_1 = \int L(z)\pi^1(z)dz$.

Next, we introduce some notations to do sensitivity analysis of KLD. Suppose we have two nominal probabilities π^1 and π^2, which have same range. Under π^1, we have $\gamma_1^N = \int L_N\pi^1 dz$, and $\gamma_1 = \int L\pi^1 dz$. Correspondingly, under π^2, we have γ_2^N and γ_2. We measure the difference between the probabilities with $d(\pi^1, \pi^2) = \int |\pi^1 - \pi^2| dz$. We would like to explore the dependence of KLD on distributional sensitivity:$DS_R[\pi^1, \pi^2](G, G_N) \doteq \frac{|D(\pi_{1,N}^d||\pi_1^d) - D(\pi_{2,N}^d||\pi_2^d)|}{d(\pi^1,\pi^2)}$.

Corollary 4.3.5 *Assume that π^1 and π^2 are equivalent measures, with $M_1 \leq M_2$ the constants of relating the norms and*

$$||G_N - G||_{L^2(\pi^1)} \leq \epsilon, \tag{4.33}$$

where ϵ is sufficiently small. Then

$$|DS_R[\pi^1, \pi^2](G, G_N)| \leq \frac{(\frac{2M_2^2}{M_1^2} + 1)\eta}{d(\pi^1, \pi^2)}$$

Proof

$$\begin{aligned} |\gamma_2^N - \gamma_2| &\leq \frac{\Lambda}{2\sigma^2\sqrt{2\pi\sigma^2}}||G_N - G||_{L^2(\pi^2)}||2d - G_N - G||_{L^2(\pi^2)} \\ &\leq \frac{M_2^2\Lambda}{2\sigma^2\sqrt{2\pi\sigma^2}}||G_N - G||_{L^2(\pi^1)}||2d - G_N - G||_{L^2(\pi^1)} \\ &\leq C_1 M_2^2 \epsilon \end{aligned} \tag{4.34}$$

Meanwhile, from of the equivalence of two measures, we have

$$\begin{aligned} \frac{\gamma_1^N}{\gamma_2^N} &= \frac{\int L_N \pi^1 dz}{\int L_N \pi^2 dz} \\ &= \frac{||\sqrt{L_N}||^2_{(L^2(\pi^1))}}{||\sqrt{L_N}||^2_{(L^2(\pi^2))}} \\ &\leq \frac{1}{M_1^2} \end{aligned} \tag{4.35}$$

Similarly,

$$\frac{\gamma_1}{\gamma_2} \leq \frac{1}{M_1^2} \tag{4.36}$$

From (3.30), we can get

$$\begin{aligned} log(\frac{\gamma_2}{\gamma_2^N}) &\leq |log(1 - \frac{M_2^2 C_1}{\gamma_1}\frac{\gamma_1}{\gamma_2}\epsilon)| \\ &\leq |log(1 - \frac{M_2^2 C_1}{\gamma_1}\frac{\gamma_1}{\gamma_2}\epsilon)| \\ &\leq |log(1 - \frac{M_2^2}{M_1^2}\frac{C_1}{\gamma_1}\epsilon)| \\ &\leq \frac{2M_2^2}{M_1^2}|log(1 - \frac{C_1}{\gamma_1}\epsilon)| \end{aligned} \tag{4.37}$$

$d(\pi^1, \pi^2)|DS_R[\pi^1, \pi^2](G, G_N)| \leq D(\pi^d_{1,N}||\pi^d_1) + D(\pi^d_{2,N}||\pi^d_2)$

For the second term:

$$\begin{aligned} D(\pi^d_{2,N}||\pi^d_2) &= \frac{1}{2\sigma^2\gamma_N^2}\int L_N[(d-G)^2 - (d-G_N)^2]\pi^2(z)dz + log(\frac{\gamma_2}{\gamma_2^N}) \\ &\leq \frac{C_3 M_2^2}{2\sigma^2\gamma_N^1}\frac{\gamma_N^1}{\gamma_N^2}\epsilon + log(\frac{\gamma_2}{\gamma_2^N}) \\ &\leq \frac{M_2^2}{M_1^2}\frac{C_3}{2\sigma^2\gamma_N^1}\epsilon + |log(\frac{\gamma_2}{\gamma_2^N})| \\ &\leq \frac{M_2^2}{M_1^2}\frac{C_3}{2\sigma^2\gamma_N^1}\epsilon + |log(\frac{\gamma_2}{\gamma_2^N})| \\ &\leq \frac{M_2^2}{M_1^2}\frac{C_3}{2\sigma^2\gamma_N^1}\epsilon + \frac{2M_2^2}{M_1^2}|log(1 - \frac{C_1}{\gamma_1}\epsilon)| \\ &\leq \frac{2M_2^2}{M_1^2}\eta \end{aligned} \tag{4.38}$$

Finally, we have:

$$|DS_R[\pi^1, \pi^2](G, G_N)| \leq \frac{(\frac{2M_2^2}{M_1^2} + 1)\eta}{d(\pi^1, \pi^2)} \tag{4.39}$$

■

4.4 Numerical Examples

In this section, we present numerical examples to support the theoretical analysis. The focus is on examination of the error behaviour.

4.4.1 Forward Model G(Z)=$exp(-Z)$

Suppose we have a system with forward model G(Z)=exp(-Z) defined on $(0, +\infty)$. We use a single observation $d = G(Z) + e$ to define a posterior density $\pi^d(z)$. e is Gaussian; e$\sim N(0, \sigma^2)$ with $\sigma = 0.1$. The original input $z_{true} = 0.2$.

We consider the following cases:

(a) The initial priors $\pi^0(z)$ are chosen to be uniform on [0,m] and the new prior is $\pi(z) = exp(-z)$ with support $[0, +\infty)$.

(b) The initial priors $\pi^0(z)$ are chosen to be uniform on support [-m,m] while the new prior $\pi(z)$ is uniform on [-1,1].

(c) The initial priors $\pi^0(z)$ are chosen to be uniform on [-m,m] and new prior $\pi(z)$ is a normal distribution, i.e. Z$\sim$ N(0,1).

(d) The initial prior $\pi^0(z)$ is uniform on [-1,1] and the new prior $\pi(z) \propto (1 - z)^\alpha (1 + z)^\beta$, with $\alpha = 3, \beta = 4$.

In case (a), assume we choose the prior for Z is uniform on [0,m] with different m, where convergence in posterior distribution under $\pi^1(z)$ is shown in Figure (4.1). Also, posterior density in a special case when m=0.3, N=5 is plotted in Figure (4.2). Figure (4.3) shows convergence of the forward model and posterior with respect to gPC order.

In case (b), we expand the forward model on support of the initial prior,i.e., $X \in [-m, m]$. The new prior is uniform on [-1,1], and we examine posterior density approximation to true posterior distribution with different initial priors. We show KLD approximation in Figure (4.4).Figure(4.5) shows posterior density approximation with initial prior to be uniform on[-5,5], and gPC expansion degree=5.

In case (c), we assume the initial prior is selected to uniform, $X \in [-m, m]$ with

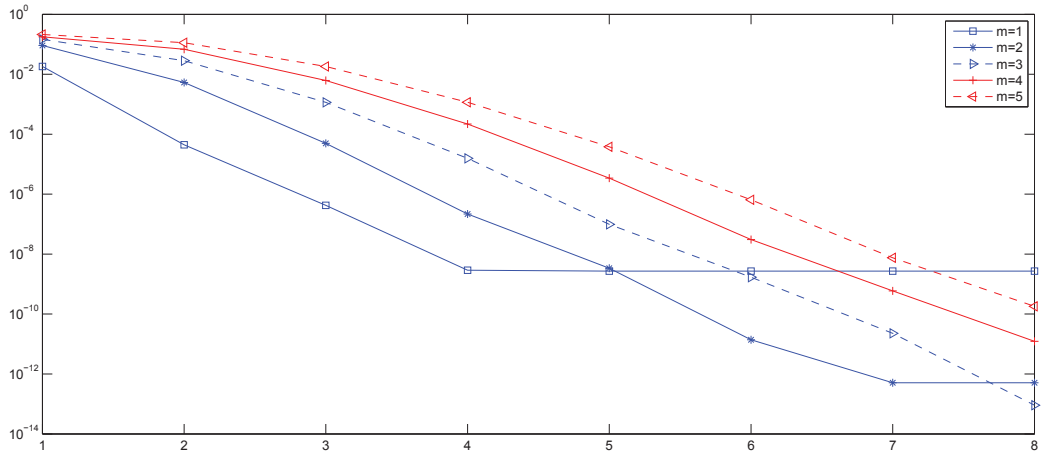

Figure 4.1. Error of new posterior, where original prior is chosen to be uniform on [0,m] and the new prior distribution is Z∼exp(-z)

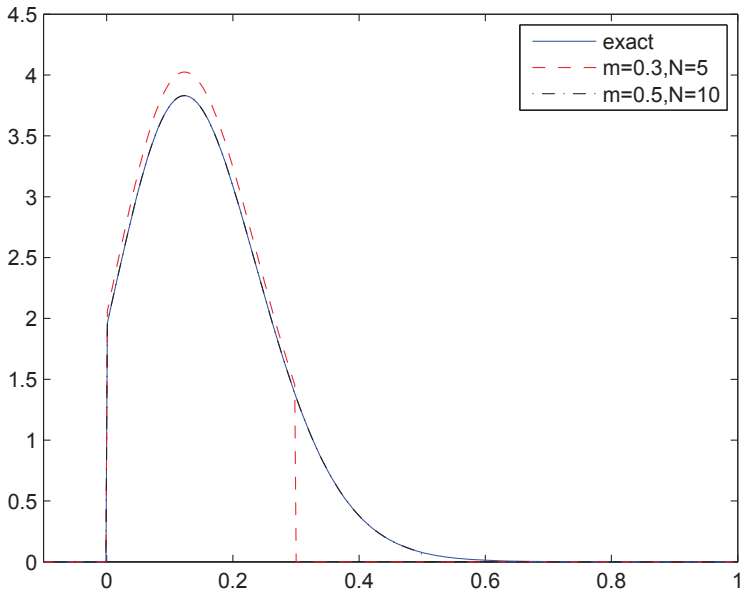

Figure 4.2. Exact and approximated posterior density for forward model G(Z)=exp(-Z) which is replaced by a 5th order gPC surrogate under the initial uniform prior on [0,0.3]; Exact and approximated posterior density for forward model G(Z)=exp(-Z) which is replaced by a 10th order gPC surrogate under the initial uniform prior on [0,0.5]

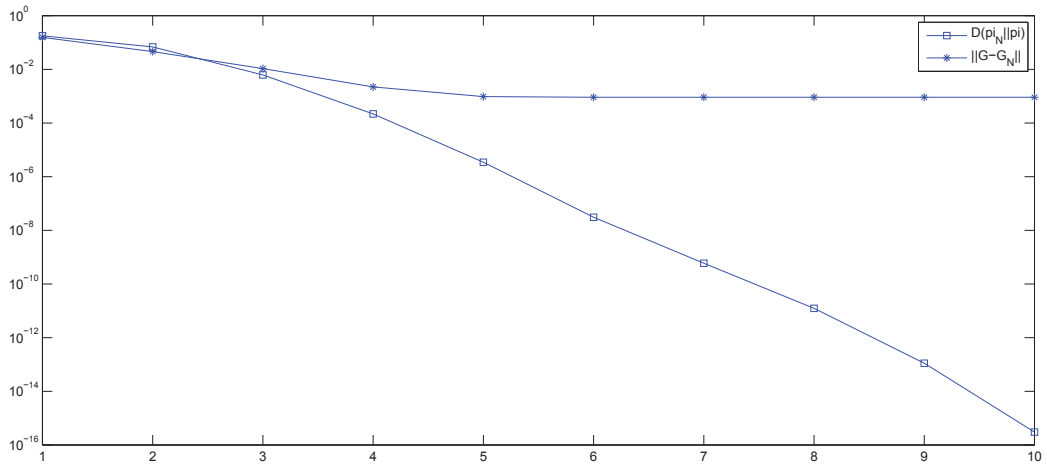

Figure 4.3. Convergence of the forward model and the posterior density with respect to gPC order N for the system with G(Z)=exp(-Z), initial prior to be uniform distribution on [0,4]

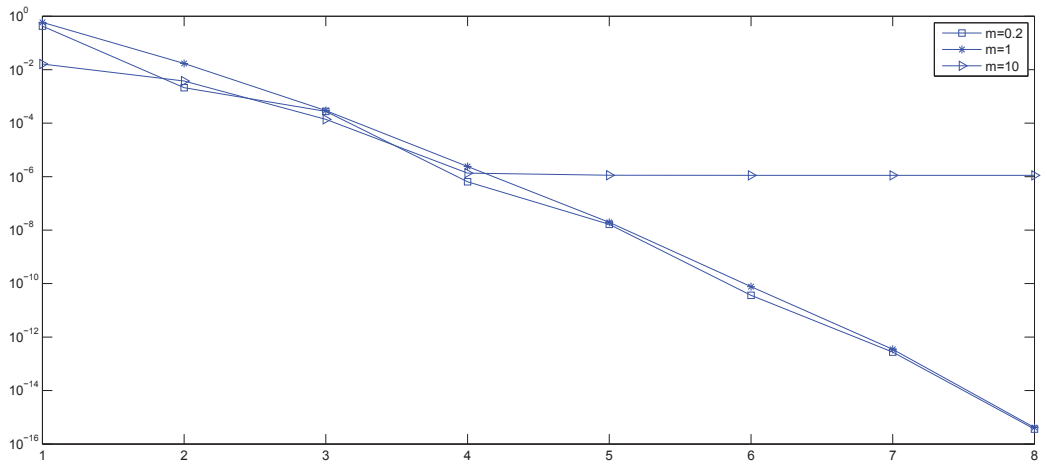

Figure 4.4. Error of posterior, where the new prior is uniform on [-1,1] and the initial priors are uniform distributions on [-m,m]

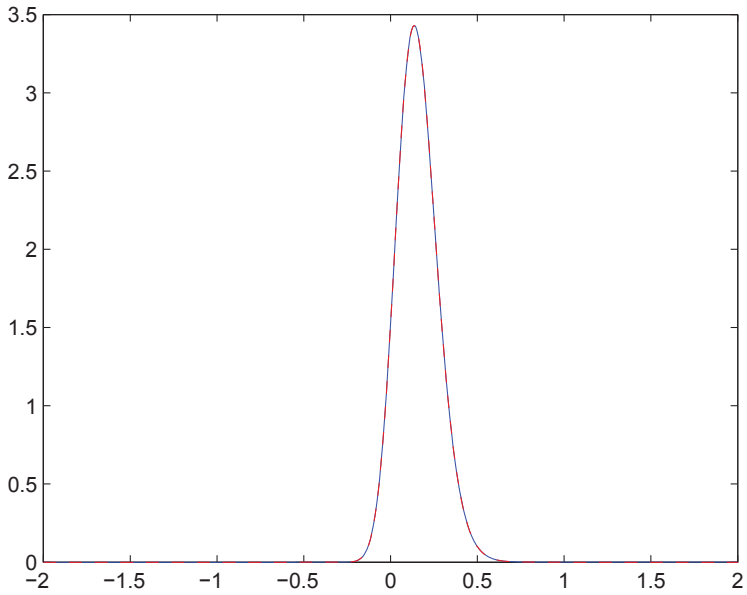

Figure 4.5. Approximation of posterior, where the new prior is uniform on [-1,1] and the initial prior is uniform distributions on [-5,5], with gPC degree=5

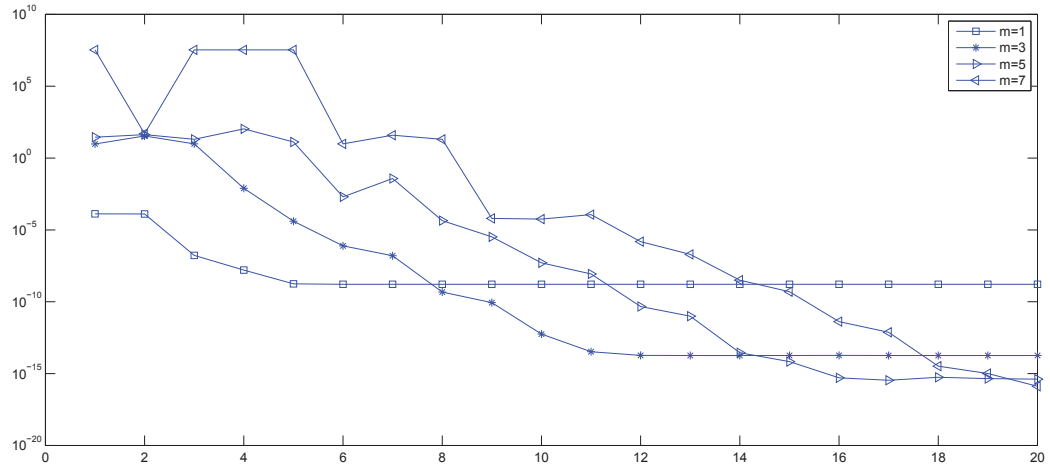

Figure 4.6. Error of new posterior, where initial prior is chosen to be uniform on [-m,m] and the new prior distribution is Z$\sim$ N(0,1)

various $m \geq 0.2$, since 0.2 is observed. Later, a new prior the normal distribution is used. Here we examine convergence of posterior in the sense of KLD and show result in Figure (4.6).

In case (d), we compare the KLD convergence rates between orthogonal gPC forward model expansion under the new and initial prior distributions. Figure(4.7) shows that stochastic collocation under new prior converges faster than that under other probabilities. But both them could be good approximations.

4.4.2 Two-dimension case with Forward model sin($Z_1 + Z_2$)

Suppose input of the system is of 2-dimension Z=(Z_1, Z_2), and foward model $G(Z) = sin(Z_1 + Z_2)$, so the output dimension is 1. Again, observation error e$\sim$ $N(0, \sigma^2)$ with $\sigma = 0.1$. We use one observation $d = sin(0.2) + e$ to define a posterior

Here we consider the following two new priors that are obtained after more data are collected:

(a) $Z_1 = Z_2$, and both of them are uniform on [-1,1]. (b) $Z_1, Z_2 \sim N(0, 1)$ with

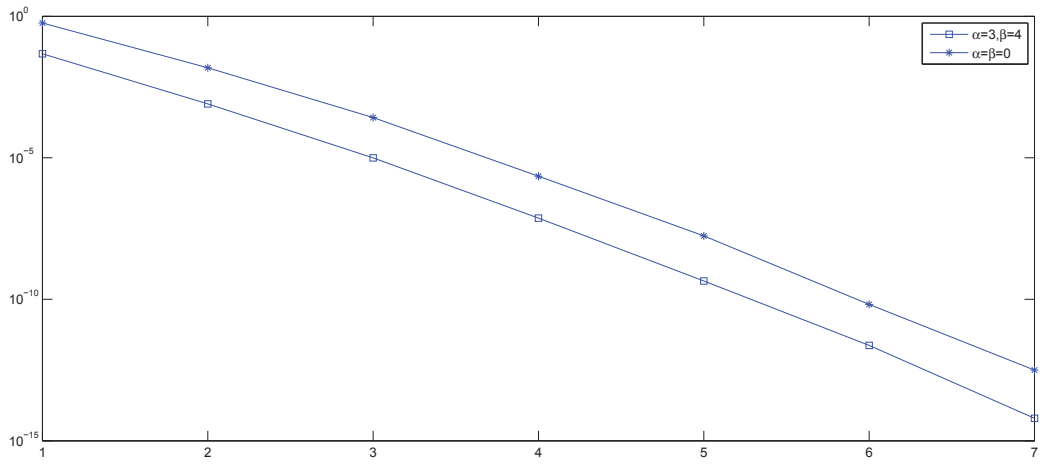

Figure 4.7. Error of posteriors, with forward model expanded under new and initial prior

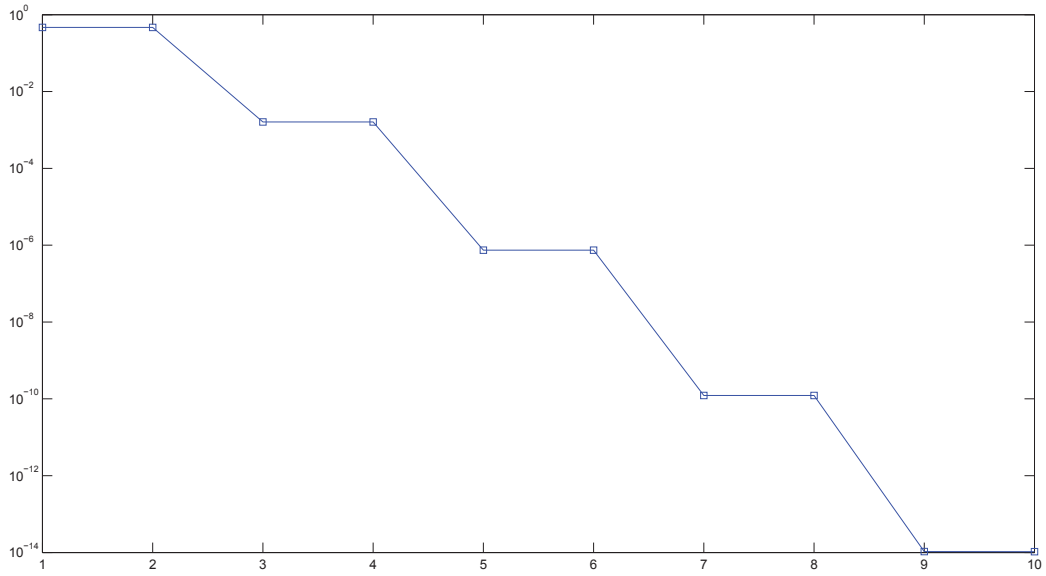

Figure 4.8. Convergence of posterior of case (a), where forward model G(Z)=$sin(Z_1 + Z_2)$

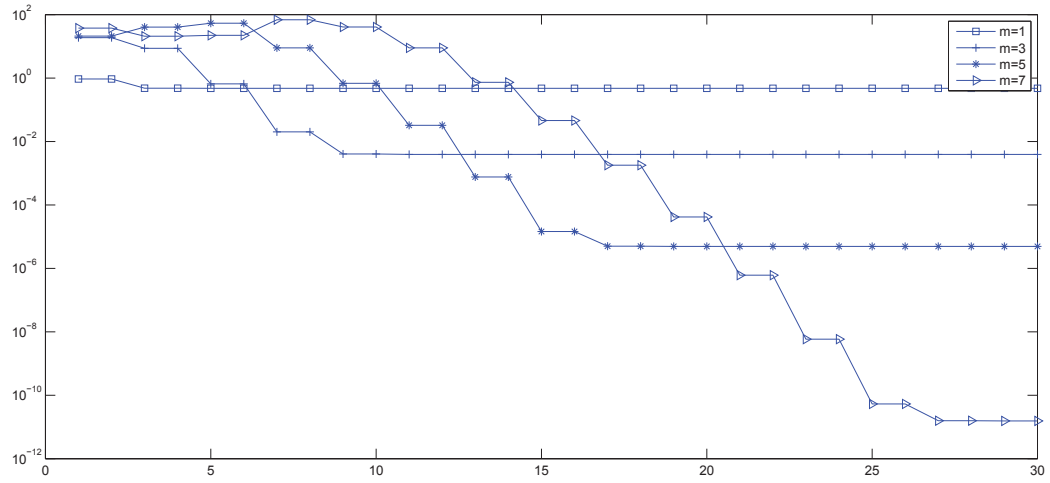

Figure 4.9. Convergence of posterior of case (b), where forward model G(Z)=$sin(Z_1 + Z_2)$

correlation $\rho = 0.3$.

In case (a), the initial prior is defined on $[-1, 1] \times [-1, 1]$, with probability i.i.d uniform distribution. We plot KLD with respect to gPC expansion degree in Figure (4.8). In case (b), since only one observation is known and no clue concerning dependence or distributions are given, we just assume both Z_1 and Z_2 uniform on [-m,m] with no correlation, where m is taken to be different values and they are the initial priors. The convergence result is shown in Figure (4.9).

In computation of (b), we first did a transformation to Z to eliminate correlation, which facilitates our computation. The transformation is:

$$Z_1 = X_1, Z_2 = \rho X_1 + \sqrt{(1 - \rho^2)} X_2.$$

Then X_1 and X_2 are i.i.d normal random variables.

4.5 Summary

This paper proposes a framework for estimating posterior distribution for the random inputs in scientific models. The validity and effectiveness of our approaches are illustrated through several examples.

The approach is based on "forward model" derived from an initial prior whose domain encloses the new prior with overwhelming probability. Once the support of the initial prior has been specified, the "forward model" is naturally established, which can be solved by generalized polynomial chaos (gPC) expansions and stochastic collocation. The polynomial approximation is chosen to converge point-wise throughout the Hilbert space defined by the initial prior. When the new prior is given, instead of a completely new forward model surrogate, a surrogate based on the current surrogate is constructed.

Based on the forward model approximation, we focus on the posterior evaluation of Bayesian inference. The surrogate model then replaces the true forward model in the likelihood function and defines an approximate posterior probability density. The posterior can be evaluated under many new prior distributions of the inputs/parameters and for an arbitrarily large number of samples, at minimal computational cost. We also prove the convergence of the approximate posterior to the true posterior, in terms of the Kullback-Leibler divergence (KLD).

5. MODEL DISCREPANCY EVALUATION

5.1 Introduction

Mathematical models are applied across many fields of physical science, engineering, environmental, social and health sciences to describe complex physical systems and predict their response when the phenomenon is not accessible or it is more expensive to test the system than numerical simulation. Based on current scientific understanding, the mathematical model is often highly complex and involves large systems of differential equations. As a result, the computer code is often large and may be expensive in terms of the computing time required for a single run.These models are used to predict and understand the underlying physical systems. They are widely applied in decision-making at all levels from individual companies through national governments to international agencies. However, the sophisticated mathematical models do not necessarily guarantee accuracy and predictability of physical systems, as a result there is a growing emphasis on understanding the uncertainties in the outputs of the models.

Let the inputs to a model be x and the output be $y = f(x)$, where f is the mathematical function that maps the inputs to the outputs and describes the phenomenon and f is usually referred as the forward problem. The majority of the uncertainty analysis techniques were developed for uncertainty propagation, i.e., forward propagation of uncertainty, to study the effect of given random inputs x on the response of a system $f(x)$. However, another important component of uncertainty is model discrepancy [46, 47], that is, the difference between the model output y and the true physical system value z that the mathematical model intends to predict. This is an inverse problem in uncertainty analysis, in which experimental data are used to learn about sources of modeling uncertainty such as calibration [48, 49] parameters

of a computer model and computer model discrepancy (inadequacy or bias). The inverse problem is receiving increased attention, because quantifying the uncertainty of a model and the resulting system response predictions is essential for predictive modeling and simulation-based decision-making.

Kennedy and OHagan [35] first introduced model discrepancy which was referred as model inadequacy formally as a source of uncertainty in model predictions. They considered the problem of using observations of the true physical system to learn about uncertain input parameters, and applied Bayesian approach to account for model discrepancy. Since then, their inferential ideas and modeling framework have been widely adopted and further developed. The Bayesian approach provides a foundation for model discrepancy inference from noisy and incomplete data, a natural framework for incorporating physical constraints and heterogeneous sources of information, and an effective quantification of uncertainty in the model discrepancy, see [47,50,51]. After Bayesian approach incorporates the observations into model discrepancy, the problem is transferred into a posterior distribution integration, which can be computationally expensive.

Several efforts at accelerating Bayesian inference in inverse problems have appeared in recent literature. Some of them are relying largely on reductions or surrogates for the forward model, while others have been seeking for more efficient sampling methods from the posterior. Recent work used (generalized) polynomial chaos (gPC)-based stochastic Galerkin methods to propagate prior uncertainty through the forward model,thus yielding a polynomial approximation of the forward solution over the support of the prior. This approximation then entered the likelihood function, resulting in a posterior density that was inexpensive to evaluate. This scheme was used to infer parameters appearing nonlinearly in a transient diffusion equation, demonstrating exponential convergence to the true posterior and multiple order-of-magnitude speed up in posterior exploration via Markov chain Monte Carlo (MCMC). The gPC stochastic Galerkin approach has also been extended to Bayesian inference of

spatially-distributed quantities, such as inhomogeneous material properties appearing as coefficients in a PDE.

This work uses gPC to expand the model discrepancy term thus the forward model, and construct the likelihood surrogates for quantifying the uncertainties in model discrepancy. After the likelihood function is established as a function of the coefficients of polynomials, the problem is translated to an optimization problem. Numerical examples are provided for a variety of nonlinear inverse problems to demonstrate the scheme and efficiency of our methodology.

5.2 Problem Setup

Let $D \subset \mathbb{R}^l$, $l = 1, 2, 3$, be a physical domain with coordinates $x = (x_1, ..., x_l)$ and let $T > 0$ be a real number. We consider the following general stochastic partial differential equation

$$\begin{cases} u_t(x,t,z) = \mathcal{L}(u) & D \times (0,T] \times I_Z, \\ \mathcal{B}(u) = 0 & \partial D \times (0,T] \times I_Z, \\ u = u_0 & D \times \{t = 0\} \times I_Z, \end{cases} \tag{5.1}$$

where $\mathcal{L}$ is a (nonlinear) differential operator, $\mathcal{B}$ is the boundary condition operator, u_0 is the initial condition, and $Z = (Z_1, ..., Z_d) \in I_Z \subseteq \mathbb{R}^d$,$d \geq 1$, are a set of random variables characterizing the random inputs to the governing equation. The solution is therefore a stochastic quantity,

$$u(x,t,z) : \bar{D} \times [0,T] \times I_Z \to \mathbb{R}^{n_v}. \tag{5.2}$$

Without loss of generality, hereafter we assume 5.2 is a scalar system with $n_v = 1$. We also make a fundamental assumption that the problem 5.1 is well posed in I_Z. If we focus on the random input Z by fixing x_0 and t_0, the solution u is function of Z, and we write it as $u_{x_0,t_0}(Z)$.

We assume each dimension of Z is independent and the joint distribution function of Z is:

$$F(z) = \Pi_{i=1}^{d} P(Z_i <= z_i).$$

In this paper we will focus on continuous random variables. Subsequently, each Z_i has a probability density function $\pi_i(z)$, then the density function of Z is:

$$\pi(z) = \Pi_{i=1}^{d} \pi_i(z_i).$$

The process in 5.1 can be very difficult to study or expensive to simulate, thus various sciences use mathematical models to describe the process, and these models are typically implemented in computer codes. For simplicity, we call them the computer models, and they are used as surrogate models for the original process during analysis. However, no model is perfect. The mathematical model and its corresponding computer model are often different from the real process. For example, we simulate both the real process and computer model at the same parameters and input, the result can be different. Then we say there is a discrepancy between the computer model and the true model (real process).

To make up for the discrepancy, we will make use of all the available observations from the real process. Here we only consider the direct observations from the real process. In real life, the observations can come from another process, called forward model, determined by 5.1, and our methods can be directly generalized. In practice, measurement error is inevitable and the observed data d may not match the true value of $d^t = u_{x_0,t_0}(z)$. Assuming additive observational errors, we have

$$d = d^t + e = u_{x_0,t_0}(z) + e,$$

where $e \in \mathbb{R}^{n_d}$ are mutually independent random variables with probability density functions $\pi(e) = \prod_{i=1}^{n_d} \pi(e_i)$. We make the usual assumption that e are also independent of Z. If we assume the discrepancy between $u(x,t,z)$ and $v(x,t,z)$ can be expressed as $u^D(x,t,z)$, then we have:

$$u(x,t,z) = v(x,t,z) + u^D(x,t,z) \tag{5.3}$$

and at a fixed point,

$$u_{x_0,t_0}(Z) = v_{x_0,t_0}(Z) + u^D_{x_0,t_0}(Z). \tag{5.4}$$

Now the system can be described as:

$$d(x,t) = v(x,t,z) + u^D(x,t,z) + e \tag{5.5}$$

where the computer model $v(x,t,z)$ is known and the observations $d(x,t)$ are collected at some parameter points (x,t). We make the usual assumption that the observation error that accounts for experimental and observational uncertainty follows a Gaussian distribution with mean 0 and variance λ, denoted $e \sim N(0,\lambda)$. The question is how to quantify the uncertainty in the discrepancy term $u^D(x,t,z)$. Most research favour a Bayesian approach , in which our judgements about $u^D(x,t,z)$ are updated by conditional on observations $d(x,t)$. If we treat the discrepancy u^D as a parameter, then the likelihood to obtain one observation d_j is

$$L(u^D|d_j) = P(d_j|u^D) \int \pi_e(d_j - (v(z) + u^D(z)))\pi(z)dz, \tag{5.6}$$

where we have omitted the fixed effect x and t. With independence assumption on the observations, the likelihood function for the observations $d = \{d_j\}_{j=i}^M$ is

$$L(u^D|d) \triangleq P(d|u^D) = \prod_{j=1}^{M} \int \pi_e(d_j - (v(z) + u^D(z)))\pi(z)dz. \tag{5.7}$$

However, taking account of possible non-linear system behaviour of d, it is likely that $u^D(x,t,z)$ is left under-determined. To solve the problem, we apply the generalized polynomial chaos methods to introduce more structure onto $u^D(x,t,z)$.

5.3 Methods

In this section we describe a scheme, based on generalized polynomial chaos (gPC) expansions, for the uncertainty quantification of model discrepancy.

5.3.1 Generalized polynomial chaos

The generalized polynomial chaos (gPC) is an orthogonal polynomial approximation to random functions. Without loss of generality, in this subsection we describe

the gPC approximation for function with random input dimension $n_d = 1$. When $n_d > 1$ the procedure will be applied to each component of G and is straightforward. Let $i = (i_1, ..., i_{nz}) \in N^{nz}$ be a multi-index with $|i| = i_1 + ... + i_{nz}$, and let $N \geq 0$ be an integer. The Nth-degree gPC expansion of G(Z) is defined as

$$G_N(Z) = \sum_{|i|=0}^{N} a_i \phi_i(Z), \tag{5.8}$$

where

$$a_i = E(G(Z)\phi_i(Z)) = \int G(z)\phi_i(z)\pi(z)dz, \tag{5.9}$$

are the expansion coefficients, E is the expectation operator, and $\phi_i(Z)$ are the basis functions defined as

$$\phi_i(Z) = \phi_{i1}(Z_1)\phi_{inz}(Z_{nz}). \tag{5.10}$$

Here $\phi_m(Z_k)$ is the mth-degree one-dimensional orthogonal polynomial in the Z_k direction satisfying, for all $k = 1, ..., n_z$,

$$E_k[\phi_m(Z_k)\phi_n(Z_k)] = \int \phi_m(z_k)\phi_n(z_k)\pi(z_k)dz_k = \delta_{m,n}, \tag{5.11}$$

where the expectation E_k is taken in terms of Z_k only and the basis polynomials have been normalized. Consequently $\phi_i(Z)$ are n_z-variate orthonormal polynomials of degree up to N satisfying

$$E[\phi_i(Z)\phi_j(Z)] = \int \phi_i(z)\phi_j(z)pi(z)dz_k = \delta_{i,j}, \tag{5.12}$$

where $\delta_{i,j} = \prod_{k=1}^{n_z} \delta_{i_k,j_k}$. The distribution of Z_k will determine the polynomial type. For example, Hermite polynomials are associated with the Gaussian distribution, Jacobi polynomials with the beta distribution, Laguerre polynomials with the gamma distribution, etc. For a detailed discussion of these correspondences and their resulting computational efficiency, see [16].

Following classical approximation theory, the gPC expansion converges when G(Z) is square integrable with respect to $\pi(z)$, that is,

$$||G(Z) - G_N(Z)||^2_{L^2_{\pi_Z}}, N \to \infty. \tag{5.13}$$

Furthermore, the rate of convergence depends on the regularity of G such that

$$||G(Z) - G_N(Z)||_{L^2_{\pi_Z}} \leq CN^{-\alpha}, \tag{5.14}$$

where C is a constant independent of N, and $\alpha > 0$ depends on the smoothness of G. When G is relatively smooth, the convergence rate can be large. This implies that a relatively low-degree expansion can achieve high accuracy and is advantageous in practical stochastic simulations. Many studies have been devoted to the convergence properties of gPC, numerically or analytically, and the computational efficiency of gPC methods, see example in [38, 39, 52, 53].

5.3.2 gPC-based likelihood function

In the gPC-based maximum likelihood method, we first find the Nth-degree gPC expansion of $u^D(Z)$:

$$u_N^D(z) = \sum_{|i|=0}^{N} a_i \Phi_i(z), \tag{5.15}$$

Then we use the approximate gPC discrepancy term to replace the exact discrepancy in likelihood function 5.7,and the approximate likelihood function is defined as:

$$L(\{a_i\}_{i=1}^N|d) = P(d|\{a_i\}_{i=1}^N) = \prod_{j=1}^{M} \int \pi_e(d_j - (v(z) + \sum_{|i|=0}^{N} a_i \Phi_i(z)))\pi(z)dz. \tag{5.16}$$

and all the information of $u^D(z)$ is contained in the coefficients $\{a_i\}_{i=1}^N$. In [45], we can see that $L(u_N^D|d)$ converges to $L(u^D|d)$ when $N \to \infty$, which means $L(\{a_i\}_{i=1}^N|d)$ converges to $L(\{a_i\}_{i=1}^\infty|d)$.

5.3.3 Maximum likelihood optimization

To best improve the computer model based on the observations, we employ the maximum likelihood method,see [54]. For a fixed N, we assume $u^D(z) = u_N^D(z)$, and

continue to maximize the corresponding likelihood in 5.16. The problem is translated into following optimization problem:

$$\begin{aligned}\{\tilde{a}_i\}_{i=0}^N &= argmax_{\{a_i\}_{i=1}^N} \quad L(\{a_i\}_{i=1}^N|d) \\ &= argmax_{\{a_i\}_{i=1}^N} \quad log(L(\{a_i\}_{i=1}^N|d)) \\ &= argmin_{\{a_i\}_{i=1}^N} \quad -log(L(\{a_i\}_{i=1}^N|d)) \\ &= argmin_{\{a_i\}_{i=1}^N} \quad -\sum_{j=1}^M log(L(\{a_i\}_{i=1}^N|d_j))\end{aligned}$$

In this work, we do not focus on optimization algorithm, but we propose the stochastic gradient optimization algorithm for above optimization problem. After finding the optimal coefficients in the deficiency term, we will have the updated model:

$$h(z) = v(z) + \sum_{i=0}^{N} \tilde{a}_i \Phi_i(z). \tag{5.17}$$

And the "distance" between $h(Z)$ and the true model $u(Z)$ will be used as a measure for the performance of our methods.

5.4 Numerical Implementation

Upon introducing and establishing the schemes for quantifying uncertainty in model discrepancy, we now investigate numerical algorithms for implementing the schemes. The likelihood function involve integration of posterior distribution, which is the key point. Another issue lies in how to reduce the uncertainties introduced by data variability. The entire problem is finally described as an optimization problem for optimal coefficients that maximize the likelihood.

5.4.1 Numerical Integration

There are many straightforward ways to computing the integration of the posterior integrations involved in the likelihood function when the distribution of the random effect is known, such as Monte Carlo, Monte Carlo Markov Chain (MCMC), stochastic collocation and etc. However, in some extreme cases when the distribution is clustered

at some point these sampling methods may not help a lot. When evaluating the posterior probability of an observation d_j, we will meet with an integration of the form $e^{Mf(x)}$. And when the noise variance is small, that is the cluster case, we will apply the Laplace approximation method.

$$\int_a^b e^{Mf(x)}dx \approx \sqrt{\frac{2\pi}{M|f''(x_0)|}}e^{Mf(x_0)} \quad as \quad M \to \infty,$$

where x_0 is the global maximum of $f(x)$ on $[a, b]$. During implementation, a general form of Laplace approximation [55] can be used, such as first dividing the whole interval into several subintervals and then apply the approximation on each subinterval.

While in multi-dimensional case, it's difficult to find a global maximum for a function, adaptive collocation or sparse grid can be good choices. Since these methods are already well-developed and not focuses of our work, we will avoid this problem by setting the observation noise at a larger level such as 0.1.

5.4.2 Efficient data utilization: Bootstrapping

Similar to Monte Carlo estimation, an estimator from one sample set is often affected by the variance of the variable. To reduce the effect, we can sample several sub-samples and average all the results from all the subsamples. Although the result from maximum likelihood can be good estimators of the true model, its asymptotic properties show that the performance is affected by data which is a random outcome. To reduce the uncertainty in the coefficients $\{a_i\}$, instead of using all the observations at one time, we propose to use bootstrapping method,see [56]. The first step of the procedure is to draw several subsets from the original data set with replacement, and then perform above algorithms on each sets, and finally average these results into the final result.

5.5 Splines in the deterministic space

Our final goal is to express the deficiency term $u^D(x,t,z)$ as a polynomial in Z:

$$u^D(x,t,z) = \sum_{i=1}^{N} a_i(x,t)\Phi(z).$$

Considering we have obtained the polynomial coefficients $\{a_(x_j,t_j)^i\}$ at the M discrete points in the deterministic space, we next try to find the global coefficients $\{a_i(x,t)\}$. There are many choices such as interpolation [57], l_2 minimization [58] and etc. In our work, to avoid oscillation, we propose spectral methods [59] or the spline interpolation [60, 61].

5.6 Numerical experiments

In this section we provide a few examples of computing the model discrepancy. For benchmark purpose, the examples here all have accessible forward model. One numerical result is obtained using Laplace approximation when the variance is small, and others, when the variance is greater, are by stochastic collocation with sufficiently large number of samples so that the numerical errors are negligible.

5.6.1 Ordinary differential equation

Let us consider

$$\frac{du}{dt} = Zu, u(0) = 1, t \in [0,1] \tag{5.18}$$

where the parameter Z is a random variable representing the uncertainty and is assumed to follow a standard normal distribution, that is, $Z \sim N(0,1)$. The exact solution

$$u(t,Z) = exp(Zt) \tag{5.19}$$

is a log-normal random variable at a fixed time t and can be expanded into a sum of Hermit polynomials[]:

$$u_N(t, Z) = e^{t^2/2} \sum_{k=0}^{N} \frac{t^k}{k!} H_k(Z), \tag{5.20}$$

where $H_k(z)$ is the Hermit polynomial of order k. And $u_N(Z)$ converges to $u(Z)$ in the L^2 sense when $N \to \infty$, if t is fixed. t is the deterministic time/space variable taking values from 0 to 1. Now we consider that, the computer model fails to capture all but only the linear information and is of form:

$$v(t, Z) = 1 + t * Z. \tag{5.21}$$

Then the inadequacy of the model is:

$$u^D(t, Z) = u(t, Z) - v(t, Z), \tag{5.22}$$

which will be studied by incorporating observations at predetermined time/space points t_i. We assume the model inadequacy at time t_0, $u^D_{t_0}(Z)$ is in the space spanned by Hermit polynomials with degree up to N, then it can be written as:

$$u^D_{t_0}(Z) = \sum_{i=0}^{N} a^i_{t_0} H_i(Z). \tag{5.23}$$

In practice, measurement error is inevitable and if assuming additive noise, the observations at t_0 is of the following form:

$$d = u(t_0, Z) + e = v(t_0, Z) + u^D(t_0, Z) + e = v(t_0, Z) + \sum_{i=0}^{N} a^i_{t_0} H_i(Z) + e$$

where e is the observation error and is assumed to follow a normal distribution with mean 0 and standard deviation 0.05. We move on to approximate the coefficients $\{a^i_{t_0}\}_{i=1}^N$ by maximum likelihood and the all the measurements at t_0. In this case, there are 1000 observations from all space points t_i. Then at t_0, we have the likelihood:

$$L(d|\{a_i(t_0)\}) = \Pi_{j=1}^M P(d_j|\{a_i(t_0)\}) = \Pi_{j=1}^M \int \pi_e(d_j - v(t_0, z) - \sum_{i=0}^{N} a^i_{t_0} H_i(z)) p(z) dz \tag{5.24}$$

where π is the distribution function of the noise e and our task is translated to maximize the last item in 5.24 or its logarithm. Since the density of a Gaussian is sparse at tail and traditional numerical integration such as Monte Carlo and Collocation does not work well for this case, we apply the Laplace Approximation to calculate the integration. During the optimization, we simply use the fminsearch command in Matlab to find the minimum of the minus log likelihood. After obtaining the coefficients at all the time points $\{t_k\}$, we continue to define our proxy model:

$$h(t,z) \doteq v(t,z) + \sum_{i=0}^{N} a^i(t) H_i(Z)$$

and at fixed time t_0

$$h(t_0,z) \doteq v(t_0,z) + \sum_{i=0}^{N} a^i_{t_0} H_i(Z)$$

and then verify our methods by checking the error between the proxy model $w(t,z)$ and the true model $u(t,z)$. Again, we first focus on the error at a fixed point t_0, and we will consider the l_2 error:

$$\begin{aligned} E_Z(t_0) \doteq \|u(t_0,Z) - h(t_0,Z))\|_2 &= \sqrt{\int (u(t_0,Z) - h(t_0,Z))^2 p(z) dz} \\ &\approx \sqrt{\textstyle\sum_j w_j (u(t_0,Z_j) - h(t_0,Z_j))^2}, \end{aligned} \tag{5.25}$$

where $p(z)$ is the density function of Z, and w_j and Z_j are the corresponding quadrature points and weights.

In this example, we collect 1000 independent samples at each of 11 time points: 0.0, 0.1,0.2,...,1. And the square of the l_2 errors at different time points are shown in figure 5.6.1, where we use polynomials with highest degree N up to 6 when approximate the deficiency term u^D.

From the figure, we see the error tends to increase as time increases this is because our true model is exponential and as time goes on the variance of our true model increases, and it's more difficult to recover. Another observation is as the degree of polynomial increases, the error at the same time point tends to decrease, this is because a higher degree of polynomial can gives a better approximation. At last but not the least, we can not eliminate the error because of the existence of noise.

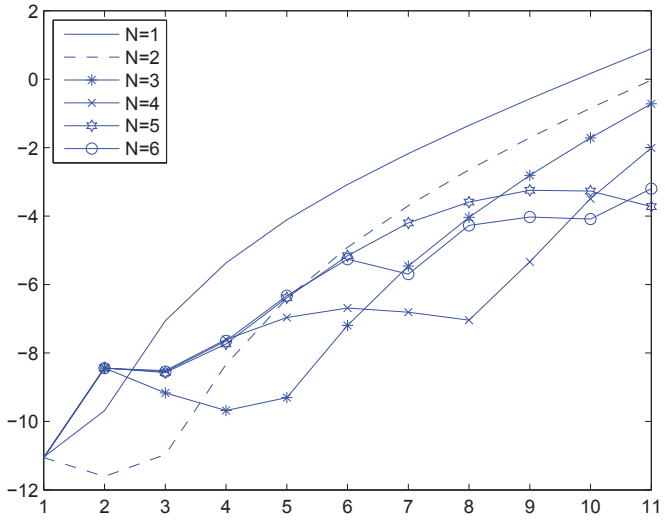

Figure 5.1. l_2 error square at each time point when highest degree of polynomial surrogate ranges from 1 to 6

Another experiment we did here is check whether an over-fitting could happen. What we did is to provide a true model in the form of polynomial with degree up to some D, and try to recover it using polynomials with higher degree. The result shows over-fitting is not a problem here, since the coefficients falls below a negligible level after the degree exceeds that of the true model. And this can be explained by the fact that the variable in our polynomials are all standardized which can reduce the multicollinearity to the maximum extent.

5.6.2 Stochastic Diffusion

In this example, we consider following SDE:

$$\frac{d}{dx}[a(x,z)\frac{du}{dx}(x,z)] = f(x,z), \quad (y,x) \in \Gamma \times [-1,1], \tag{5.26}$$

with boundary conditions:

$$u(-1,z) = 0, \qquad u(1,z) = 0, \qquad z \in \Gamma \tag{5.27}$$

Here the physical space is of one-dimension and let $f = 40$ be a constant. The true model is the solution to the differential equation in 5.26 with diffusivity field

$$a(x,z) = 4 + (\frac{cos(\pi x)}{100}z_1 + \frac{cos(\pi x)}{100}z_2 + z_3 + z_4), \tag{5.28}$$

where $z_i \in [-1,1], i = 1,2,3,4$ are independent uniformly distributed random variables. However, the computer model only captures the first two terms of $a(x,z)$, and its diffusivity field is of following form:

$$a_c(x,z) = 4 + (\frac{cos(\pi x)}{100}z_1 + \frac{cos(\pi x)}{100}z_2) \tag{5.29}$$

The computer model $v(x,z)$ is the solution to 5.26 with diffusivity field to be $a_c(x,z)$. To improve the computer model, we project the deficiency term:

$$u^D(x,z) = u(x,z) - v(x,z)$$

onto the space generated to Legendre polynomials. Then we can write:

$$u^D(x,z) = \sum_{i=0}^{\infty} a_i L_i(x,z),$$

where L_i is the Legendre polynomial of degree i. And our observations can be modeled as:

$$d = u(x,z) + e = v(x,z) + u^D(x,z) + e = v(x,z) + \sum_{i=0}^{\infty} a_i L_i(x,z) + e,$$

where e is the noise following Gaussian distribution $N(0, 0.1)$. On the deterministic space, we discretize the space $[-1,1]$ using ten chebyshev points, then there are a total of 11 points on $[-1,1]$ including the end points. And at each fixed point x_0, there are 10000 independent random points. After approximating the infinite sum with a truncation sum of N terms, the likelihood can be written as:

$$L(d|\{a_i(t_0)\}) = \Pi_{j=1}^{M} P(d_j|\{a_i(t_0)\}) = \Pi_{j=1}^{M} \int \pi_e(d_j - v(t_0,z) - \sum_{i=0}^{N} a_{t_0}^i L_i(z))p(z)dz \tag{5.30}$$

Our task now is to maximize the likelihood or minimize the $-log(L)$. The difficulty here is still the approximation of the integration. In multivariate cases, it's hard to

find global minimum or maximum of a function, thus we use Monte Carlo simulation with large enough number of sample points instead of Laplace approximation. In the experiment, we randomly select 10000 points on $[-1, 1]$ to approximate the integration then the likelihood. After finding the coefficients $\{a_i^{x_0}\}$ at each fixed x_0, we move on to check the l_2 errors. Figure shows the square of the l_2 errors at each x_j when using different polynomials with highest degree up to 4.

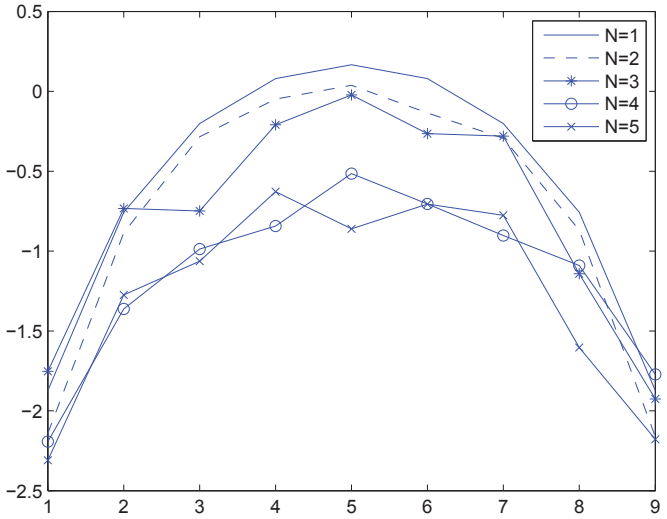

Figure 5.2. l_2 error square at each time point when highest degree of polynomial surrogate ranges from 1 to 5

After computing the coefficients at $t = 0, 0.1, ..., 1$ for $i = 1, 2, ..., N$, we can find a $C_i(t)$ which is a polynomial in t and of degree 9 for each i, such that $C_i(t_k) = a^i(t_k)$ for each $t_k \in 0, 0.1, ..., 1.0$. After rewriting the inadequacy as a function of t and Z:

$$u^D(t, Z) = \sum_{i=0}^{N} C_i(t) H_i(Z), \tag{5.31}$$

the computer model is modified as $h(t, Z)$:

After discretizing the time space, the error of the modified model can be finally calculated as:

$$E \doteq \frac{1}{K}\sqrt{\sum_{k=1}^{K} E_Z(t_k)^2}, \tag{5.32}$$

where t_k are the discretization points on the time space.

5.7 Conclusion

Mathematical models are widely used as simulators for expensive physical processes, but a good model can not grantee the accuracy for prediction, thus uncertainty quantification is necessary for model update and improvement. In this work, we develop efficient numerical implementation scheme for model discrepancy study, using generalized polynomial chaos (gPC) expansions and maximum likelihood method. Given observations at a known point in physical space, we construct gPC approximation of discrepancy part and then the forward model; gPC expansions then replace the full forward model in the likelihood function and define an approximate likelihood. We transferred the problem into an optimization problem of searching best polynomial coefficients.

In practical problems, the posterior distribution in the likelihood function can be difficult to evaluate, and we introduced several numerical methods to estimate the posterior. Also, we introduce a Bootstrapping strategy to efficiently use the data and reduce the uncertainty caused by data variability. Since all this optimization scheme is implemented locally in the physical space, we propose spline methods to construct a model discrepancy term over the whole deterministic space. The algorithms and performance are finally demonstrated numerically.

6. SUMMARY AND FUTURE WORK

In this work we seek to establish an efficient method to quantify the uncertainties in mathematical models: the model itself, its inputs (inverse problem), and its outputs ("forward problem") when epistemic or aleatory uncertainty exists. We have achieved these goals. On one hand, tight upper bound and lower bound are derived through the duality between the relative entropy and the exponential integral. Properties of the bounds are explored which give the guidance to construct efficient optimization algorithm to calculate these bounds. To obtain these bounds one need simulations only on the nominal distribution no matter how the partially unknown epistemic probability measure varies. In practical problems when we do not have sufficient information of the probability distribution of the inputs, the bounds can serve as reliable estimate of the best case scenario and the worst case scenario of the quantity of interest and this information is very helpful to improve the design in an effective way.

On the other hand, we establish an efficient scheme to approximate posteriors subject to multiple priors. We select an initial prior and construct an initial forward model surrogate and then approximate posterior distribution for different priors. These gPC based approximations facilitate further study on posterior simulation such as Monte Carlo Markov Chain (MCMC). A corresponding error analysis for using different prior distributions is presented. Meanwhile, under the Bayesian framework, an algorithm for model discrepancy study based on generalized polynomial chaos is established.

There are several areas that we can explore in the future work. First, in the model discrepancy work, after a polynomial structure is introduced, we can establish the distribution of the outputs based on forward model and distribution of noise. We intend to minimize the "distance" between the distribution of the output and the true output observations. Besides likelihood, there are other choices for the distance metric. A

good metric will result in better estimate and more efficient evaluation. The asymptotic convergence of our solution for the coefficients is a possible result. Since our problem is finally transformed to an optimization problem, many methodologies could be applied to improve the algorithm. For example, we could add some constraints to deal with the issue that our problem is under-determined. Lastly, when calculating the upper bound and lower bound of quantity of interest (QoI), the nominal probability measure plays an essential role. Different choices of nominal probability measure result in different bounds. To find an approach to compare the performance of two nominal measures that doesn't rely on the simulations on both probability measures helps determining the optimal nominal measure, and hence getting tighter bounds.

LIST OF REFERENCES

LIST OF REFERENCES

[1] J.C. Helton, J.D. Johnson, W.L. Oberkampf, and C.J. Sallaberry. Representation of analysis results involving aleatory and epistemic uncertainty. Technical Report 4379, Sandia National Laboratories, 2008.

[2] L. Swiler, T. Paez, R. Mayes, and M. Eldred. Epistemic uncertainty in the calculation of margins. In *AIAA Structures, Structural Dynamics, and Materials Conference*, Palm Springs, CA, May 2009.

[3] D. Dubois and H. Prade. *Possibility theory: an approach to computerized processing of uncertainty*. Plenum, New York, 1998.

[4] U.O. Akpan, T.S. Koko, I.R. Orisamolu, and B.K. Gallant. Practical fuzzy finite element analysis of structures. *Finite Elem. Anal.*, 38:93–111, 2001.

[5] M. Oberguggenberger and S. Pittschmann. Differential equations with fuzzy parameters. *Math. Comput. Model. Dyn. Sys.*, 5:181–202, 1999.

[6] J.C. Helton, J.D. Johnson, W.L. Oberkampf, and C.B. Storlie. A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory. Technical Report 5557, Sandia National Laboratories, 2006.

[7] I.R. Goodman and H.T. Nguyen. Probability updating using second order probabilities and conditional event algebra. *Infor. Sci.*, 121:295–347, 1999.

[8] P. Walley. Towards a unified theory of imprecise probability. *Int. J. Approx. Reasoning*, 24:125–148, 2000.

[9] J. Jakeman, M. Eldred, and D. Xiu. Numerical approach for quantification of epistemic uncertainty. *J. Comput. Phys.*, 229:4648–4663, 2010.

[10] X. Chen, E.-J. Park, and D. Xiu. A flexible numerical approach for quantification of epistemic uncertainty. *J. Comput. Phys.*, submitted, 2012.

[11] K. Chowdhary and P. Dupuis. Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification. *ESAIM Math. Model. Numer. Anal.*, in print, 2012.

[12] P. Dupuis and R.S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, New York, 1997.

[13] J. Li and D. Xiu. Computation of failure probability subject to epistemic uncertainty. *SIAM J. Sci. Comput.*, submitted, 2012.

[14] A. Narayan and D. Xiu. Distributional sensitivity for uncertainty quantification. *Commun. Comput. Phys.*, 10(1):140–160, July 2011.

[15] D. Xiu and G.E. Karniadakis. Supersensitivity due to uncertain boundary conditions. *Int. J. Numer. Meth. Engng.*, 61(12):2114–2138, 2004.

[16] D. Xiu and G.E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2):619–644, 2002.

[17] R.G. Ghanem and P. Spanos. *Stochastic Finite Elements: a Spectral Approach*. Springer-Verlag, 1991.

[18] D. Xiu. *Numerical methods for stochastic computations*. Princeton Univeristy Press, Princeton, New Jersey, 2010.

[19] O. Le Matre and O. Knio. Spectral methods for uncertainty quantiffication: with applications to computational fluid dynamics. *Springer*, 2010.

[20] Christopher J Roy William L Oberkampf. *Verification and validation in scientific computing*. Cambridge University Press, 2010.

[21] X. Qi J. Li and D. Xiu. On upper and lower bounds for quantity of interest in problems subject to epistemic uncertainty. *SIAM J. Sci. Comput.*, 2014.

[22] S.R.S. Varadhan. *Large deviations and applications*. Springer Berlin Heidelberg, 1988.

[23] S. N. Evans and P. B. Stark. Inverse problems as statistics. *Inverse Problems*, 2002.

[24] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2005.

[25] A. Mohammad-Djafari. Bayesian inference for inverse problems. *In Bayesian inference and Maximum Entropy Methods in Science and Engineering*, 2002.

[26] A. Tarantola. Inverse problem theory and methods fo rmodel parameter estimation. *SIAM*, 2005.

[27] J.Wang and N. Zabarasi. Hierarchical bayesianmodels for inverse problems in heat conduction. *Inverse Problems*, 2005.

[28] J. A. Christen and C. Foxi. Mcmc using an approximation. *J. Comput. Graph. Stat.*, 2005.

[29] W. Luo Y. Efendiev, T. Y. Hou. Preconditioning markov chain monte carlo simulations using coarse-scale models. *SIAMJ. Sci. Comput.*, 2006.

[30] H. Lee D. Higdon and C. Holloman. Markov chain monte carlo-based approaches for in- ference in computationally intensive inverse problems. *Bayesian Statistics*, 2003.

[31] S. Richardson W. R. Gilks and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 1996.

[32] D. Gamerman. *Markov Chain Monte Carlo : stochastic simulation for Bayesian inference*. Chapman & Hall, 1997.

[33] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov Chain Monte Carlo in practice*. Chapman & Hall, 1995.

[34] M. G. Ierapetritou G. P. Flach S. Balakrishnan, A. Roy and P. G. Georgopoulos. Uncertainty reduction and characterization for complex environmental fate and transport models: an empirical bayesian framework incorporating the stochastic response surfacemethod. *Water Resources Res*, 2003.

[35] M. C. Kennedy and A. OHagan. Bayesian calibration of computer models. *J. Royal Statist.*, 2001.

[36] H. N. Najm Y. M. Marzouk and L. A. Rahn. Stochastic spectral methods for efcient bayesian solution of inverse problems. *J. Comput. Phys*, 2007.

[37] D. Xiu. Fast numerical methods for stochastic computations: a review. *Comm. Comput. Phys.*, 5:242–272, 2009.

[38] R.G. Ghanem and P. Spanos. *Stochastic Finite Elements: a Spectral Approach*. Springer Verlag, 1991.

[39] D. Xiu and G.E. Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.*, 187:137–167, 2003.

[40] I. Babuška and R. Tempone. A stochastic collocation method for elliptic partial differential equations with randominput data. *SIAMJ.Numer. Anal*, 2007.

[41] M. Eldred. Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design. *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2009.

[42] D. Xiu. Efficient collocational approach for parametric uncertainty analysis. *Comm. Comput. Phys.*, 2(2):293–309, 2007.

[43] D. Xiu and J.S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.*, 27(3):1118–1139, 2005.

[44] Y. M. Marzouk and H. N. Najm. Dimensionality reduction and polynomial chaos accelera- tion of bayesian inference in inverse problems. *J. Comput. Phys*, 2009.

[45] D Xiu Y Marzouk. A stochastic collocation approach to bayesian inference in inverse problems. *Communications in Computational Physics*, 2009.

[46] Xiong Y. Tsui K. L. Chen, W. and S. Wang. A design-driven validation approach using bayesian prediction models. *Journal of Mechanical Design*, 2008.

[47] Tsui K. L. Wang, S. and W. Chen. Bayesian validation of computer models. *Technometrics*, 2009.

[48] Alberg H. Lindgren, L.-E. and K. Domkin. Constitutive modelling and parameter optimization. *7th International Conference on Computational Plasticity, Barcelona, Spain*, 2003.

[49] Mahadevan S. Romero V. McFarland, J. and L. Swiler. Calibration and uncertainty analysis for computer simulations with multivariate output. *AIAA Journal*, 2008.

[50] Bayarri M. J. Berger J. O. Paulo R. Liu, F. and J. Sacks. A bayesian analysis of the thermal challenge problem. *Comput. Meth. Appl. Mech. Eng.*, 2008.

[51] Bayarri M. J. Liu, F. and J. O. Berger. Modularization in bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 2009.

[52] I. Babuska, R. Tempone, and G.E. Zouraris. Galerkin finite element approximations of stochastic elliptic differential equations. *SIAM J. Numer. Anal.*, 42:800–825, 2004.

[53] O.M. Knio and O.P. Le Maitre. Uncertainty propagation in CFD using polynomial chaos decomposition. *Fluid Dyn. Res.*, 38(9):616–640, 2006.

[54] Hirotogu Akaike. *Information Theory and an Extension of the Maximum Likelihood Principle*. Springer, 1998.

[55] Sara Martino Havard Rue and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society*, 2009.

[56] GR Hancock J Nevitt. Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling*, 2001.

[57] Walter Gautschi. *Numerical analysis: an introduction*. Birkhuser, Boston, 1997.

[58] M. Zibulevsky, M. ; Elad. L1-l2 optimization in signal and image processing. *Signal Processing Magazine*, 2010.

[59] Li-Lian Wang Jie Shen, Tao Tang. *Spectral Methods: Algorithms, Analysis and Applications*. Springer, 2011.

[60] O. C. Zienkiewicz. *The finite element method in engineering science*. McGraw-Hill, 1971.

[61] Grace Wahba. Improper priors, spline smoothing and the problem of guarding against. model errors in regression. *Journal of the Royal Statistical Society*, 1978.

APPENDICES

A. A Brief Introduction to General Stochastic Collocation

In the pseudo-spectral stochastic collocation method, an approximate gPC expansion is sought, in the following form,

$$\tilde{G}_N(Z) = \sum_{|i|=0}^{N} a_i \phi_i(Z), \tag{A.1}$$

where the expansion coefficients are obtained by

$$\tilde{a}_i = \sum_{m=1}^{Q} G(Z^m)\phi_i(Z^m)\omega^m, \tag{A.2}$$

where $Z^m = (Z_1^m, ..., Z_{n_z}^m)$ are a set of nodes and ω^m are the corresponding weights for m=1,...,Q, of an integration rule (cubature) on R^{n_z} such that

$$\tilde{a}_i \approx \int G(z)\phi_i(z)\pi(z)dz = a_i. \tag{A.3}$$

The expansion A.1 thus becomes an approximation to the exact expansion ???; that is,

$$\tilde{G}_N(Z) \approx G_N(Z). \tag{A.4}$$

The difference between the two expansions is the so-called "aliasing error" and is induced by the error of using the integration rule. If a convergent integration rule is employed such that

$$\lim_{Q\to\infty} \tilde{a}_i = a_i, \ \forall i, \tag{A.5}$$

then

$$\lim_{Q\to\infty} \tilde{G}_N(Z) = G_N(Z), \ \forall Z, \tag{A.6}$$

and convergence of $\tilde{G}_N$ to the exact forward model G follows naturally,

$$||G(Z) - \tilde{G}_N(Z)||^2_{L^2_{\pi Z}} \to 0, N \to \infty, Q \to \infty \tag{A.7}$$

A prominent feature of the pseudo-spectral collocation method is that it only requires simulations of the forward model G(Z) at fixed nodes $Z^{(m)}$,m=1,... ,Q, which are uncoupled deterministic problems with different parameter settings. This significantly facilitates its application in practical simulations, as long as the aliasing error is under control.

B. Proof of Proposition 3.3.2

Here we provide the proof for Proposition 3.3.2:

Proof Part (a) is an extension of the Proposition 3 in [11]. Let $H^+(c) = c \cdot \Theta^+(c) = \log \int_\Omega e^{cf} \rho(dx)$. Its differentiation leads to

$$(H^+)'(c) = \frac{\int_\Omega f(x) e^{cf(x)} \rho(dx)}{\int_\Omega e^{cf(x)} \rho(dx)},$$
$$(H^+)''(c) = \frac{[\int_\Omega f^2(x) e^{cf(x)} \rho(dx)][\int_\Omega e^{cf(x)} \rho(dx)] - [\int_\Omega f(x) e^{cf(x)} \rho(dx)]^2}{[\int_\Omega e^{cf(x)} \rho(dx)]^2}.$$

By using Hölder's inequality, we have $(H^+)''(c) > 0$ for $c > 0$. By observing $(\Theta^+)'(c) = \frac{c(H^+)'(c) - H^+(c)}{c^2}$ and considering $G^+(c) \triangleq c(H^+)'(c) - H^+(c)$, we obtain $(G^+)'(c) = (H^+)''(c) > 0$, for $c > 0$. Since $H^+(0) = 0$ and $(H^+)'(0) = \int_\Omega f(x) \rho(dx) < \infty$, we conclude that $G^+(c)$ is monotone increasing and $G^+(0) = 0, G^+(c) \geq 0$ for $c > 0$, which indicates $(\Theta^+)'(c) \geq 0$ for $c > 0$. Thus $\Theta^+(c)$ is nondecreasing for $c > 0$.

Let $M \triangleq \sup D$. We first assume $M = \infty$. Since

$$\frac{d}{dc}\left[\frac{1}{c} R^* + \frac{1}{c} H^+(c)\right] = \frac{1}{c^2}\left[-R^* + c(H^+)'(c) - H^+(c)\right] = \frac{1}{c^2}\left[-R^* + G^+(c)\right],$$

we let $K^+ \triangleq \lim_{c \to \infty} G^+(c)$. If $K^+ = \infty$ or $0 \leq R^* < K^+$, then there is a unique solution to $G^+(c) = R^*$. When $R^* \geq K^+$, $R^* + \Theta^+$ is monotone decreasing for $c \geq 0$, and since $\Theta^+(c) \geq (H^+)'(0) > -\infty$, there is a well-defined limit $\lim_{c \to \infty} \Theta^+(c)$ that is necessarily the minimum.

Next we assume $M < \infty$. In this case, $\Theta^+ \uparrow \infty$, as $c \uparrow M$, therefore $H^+(c) \uparrow \infty$, as $c \uparrow M$. Using the fact that $(H^+)'$ is monotone increasing we have $(H^+)' \uparrow \infty$ as $c \uparrow M$. Then by applying the same strategy in [11], we complete the proof of part (a).

For part (b), let $H^-(c) = c \cdot \Theta^-(c) = -\log \int_\Omega e^{-cf(x)} \rho(dx)$. We then derive

$$(H^-)'(c) \quad = \quad \frac{\int_\Omega f(x) e^{-cf(x)} \rho(dx)}{\int_\Omega e^{-cf(x)} \rho(dx)},$$

and

$$(H^-)''(c) = -\frac{[\int_\Omega f(x)^2 e^{-cf(x)}\rho(dx)][\int_\Omega e^{-cf(x)}\rho(dx)] - [\int_\Omega f(x)e^{-cf(x)}\rho(dx)]^2}{[\int_\Omega e^{-cf(x)}\rho(dx)]^2}.$$

Again by using the Hölder's inequality, we have $(H^-)''(c) < 0$ for $c < 0$. Consequently $(H^-)'(c)$ is a nondecreasing function. By using

$$(\Theta^-)'(c) = \frac{c(H^-)'(c) - H^-(c)}{c^2}, \tag{B.1}$$

and considering $G^-(c) \triangleq c(H^-)'(c) - H^-(c)$, we have $(G^-)'(c) = c(H^-)''(c) < 0$ for $c > 0$. Using the similar reasoning as in part (a) we deduce $\Theta^-(c) = \frac{1}{c}H^-(c)$ is nonincreasing for $c \geq 0$.

To establish the local unique maximum, we define $M \triangleq \sup E$. We first assume $M = \infty$. Upon taking the derivative of $\Theta^-(c) - \frac{1}{c}R^*$ we have

$$\frac{d}{dc}\left[\frac{1}{c}H^-(c) - \frac{1}{c}R^*\right] = \frac{1}{c^2}\left[c(H^-)'(c) - H^-(c) + R^*\right] = \frac{G^-(c) - R^*}{c^2}.$$

If $\lim_{c\to\infty} G^-(c) < -R^*$, then there is a unique solution to $c(H^-)'(c) - H^-(c) + R^* = 0$ and the claim is satisfied. If $\lim_{c\to\infty} G^-(c) \geq -R^*$, then $\frac{1}{c}H^-(c) - \frac{1}{c}R^*$ is monotonically increasing on $[0, \infty)$. It follows from (B.1) that $\frac{1}{c}H^-(c) \leq (H^-)'(0) \leq \infty$. Then there exists a well-defined limit $\lim_{c\to\infty}\frac{1}{c}H^-(c) = \lim_{c\to\infty}[\frac{1}{c}H^-(c) - \frac{1}{c}R^*]$ which is the required maximum.

Next let us consider the case of $M \in (0, \infty)$. It is straightforward to see that $H^-(c) \downarrow -\infty$ and $(H^-)'(c) \downarrow -\infty$ as $c \uparrow M$. Let $0 < b < c < M$, then the relation

$$H^-(c) = \int_0^c (H^-)'(t)dt \geq b(H^-)'(b) + (c-b)(H^-)'(c)$$

implies $c(H^-)'(c) - H^-(c) \leq b((H^-)'(c) - (H^-)'(b))$. By fixing b and letting $c \uparrow M$, it follows that $c(H^-)'(c) - H^-(c) \downarrow -\infty$ as $c \uparrow M$. By applying the same argument in the case of $M = \infty$, the conclusion can be established. ■

VITA

VITA

Xin Qi was born in Hebei Province, P.R. China on October 15, 1986. She received the B.A. degree of mathematics in 2009 at Sichuan University. From 2009 to 2014, she enrolled the Ph.D. program in mathematics at Purdue University and received the M.S. degree of statistics and mathematics.