Hybrid Sampling for Uncertainty Quantification in Systems with High Dimensional Parameter Spaces

by

Fauziya Ado Yakasai

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 7, 2022

Keywords: Hybrid Sampling, Uncertainty Quantification, Sparse Grid Method, Monte Carlo Method

Approved by

Hans Werner van Wyk, Chair, Associate Professor of Mathematics and Statistics
Yanzhao Cao, Professor of Mathematics and Statistics
Mark Capenter, Professor of Mathematics and Statistics
Dmitry Glotov, Associate Professor of Mathematics and Statistics

Xiaoying (Maggie) Han, Professor of Mathematics and Statistics
Chandana Mitra, (University Reader) Associate Professor of Geosciences
George Flowers, Dean of the Graduate School and Professor in Mechanical Engineering

Abstract

Spatial models formulated as partial differential equations, often include space-dependent parameters that are not readily estimable and are therefore uncertain. There are two broad classes of methods for computing statistical quantities of interest related to the model solution: Spectral methods, such as the generalized polynomial chaos and stochastic collocation methods, are well-suited for systems with low parameter complexity. However, their convergence rates deteriorate as the dimension of the parameter space increases, and hence for systems with high parameter complexity, methods whose convergence rates are independent of the stochastic dimension, such as the Monte Carlo method, are more appropriate. In this work, we propose a hybrid sampling scheme which uses conditional sampling to combine sparse grid quadrature rules on a low-dimensional projection of the parameter space with a Monte Carlo scheme to compensate in the remaining dimensions. Using complexity arguments, we show that our method is more efficient than either of its constituents. We include some numerical examples to illustrate our results.

Acknowledgments

All praise be to Allah for allowing me to reach this milestone. I'd like to express my gratitude to my advisor, Dr. Hans Weiner van Wyk, for his invaluable assistance. He has always been kind, helpful, and patient with me. I couldn't have finished my dissertation without his assistance. I will be grateful for his contribution to my research studies for the rest of my life.

I am really grateful to my parents (Engr. Ado Umar Yakasai and Binta Kilishi), siblings (Dr. Umar, Engr. Shamsu, Naja' atu, and Sadiq), and friends for all of their support. In addition, I'd like to thank Drs. Yanzhao Cao, Mark Capenter, Dmitry Glotov, and Xiaoying (Maggie) Han for serving on my committee, as well as Dr. Chandana Mitra for joining the committee as a university reader.

Contents

List of Figures

List of Symbols

$\emptyset$	Empty Set
$\|\|\|$	Norm
$\mathbb{E}$	Expectation
$\mathbb{P}$	Probability space
$\mathcal{F}$	*Sigma*-algebra
Cov	Covariance
μ	mean
∇	Gradient
Ω	Sample Space
$\mathbb{R}$	Real numbers
ρ	Correlation
$\Sigma \in \mathbb{R}^{d \times d}$	Covariance Matrix
σ	Standard Deviation
θ	Random field
θ_0	Truncated random field
$\mathbb{V}$	Variance

e	exponential
$J(\theta)$	Quantity of interest
n	Number of random samples

Chapter 1

Introduction

Many complex physical processes are described in the deterministic language of partial differential equations (PDEs), yet operate in uncertain environments that can only be observed partially and/or indirectly. This aleatory uncertainty manifests itself in the underlying system parameters, which are often modeled as random quantities with known distributions: constants are modeled as random variables while spatially varying parameters are modeled as random fields. Material properties such as those of alloys, polymers, or composite materials in predictive materials modeling, structural properties of biological tissues in biomechanics [8], or properties of porous media in subsurface geology appear commonly as complex, multi-scale, and non-stationary random fields. For example, in a groundwater flow problem the uncertainties in the soil and sources or sinks can be model by describing the conductivity coefficient and forcing term as a random fields, as described by the following diffusion equation

$$-\nabla \cdot (\theta \nabla u) + v \cdot \nabla u = f, \qquad \text{for } x \in D, \tag{1.1}$$

subject to appropriate boundary conditions, where $D \subset \mathbb{R}^d$ is some bounded admissible region with boundary ∂D. The solution u is the hydraulic head; The random

field $\theta : D \times \Omega \to \mathbb{R}$ is the appropriate conductivity, v is the velocity and the right-hand side $f : D \times \Omega \to \mathbb{R}$ which may also be random are sources or sinks in the region D. Hence the solution $u(x, \omega)$ is also random. [18].

In application certain statistics and moments of $u(x, \omega)$ are of interest. For example given statistics such as the mean $\mathbb{E}_\theta$ and covariance Cov_θ of the data, we can compute statistics of the random solution u such as $\mathbb{E}_u$ or Cov_u [9].

Various numerical approaches are available for approximating statistical quantities of interest related to the associated random model output, the choice of which usually depends on (i) the complexity of the parameter space, (ii) the regularity and intrinsic dimension of the output, and (iii) the computational cost of solving the PDE. In particular polynomial-based methods, such as the generalized polynomial chaos [11], and stochastic collocation methods [21], are well-suited for systems with low parameter complexity, i.e. those whose uncertain inputs can be expressed in terms of a moderate number of random variables. These methods can be designed to exploit anisotropy in the parameter space, and yield fast convergence rates when the quantity of interest depends smoothly on the random inputs, while also allowing for adaptivity in the case of limited smoothness. However, their convergence rates deteriorate as the dimension of the parameter space increases, and hence for systems with high parameter complexity, methods whose convergence rates are only mildly dependent on the stochastic dimension, such as the quasi-Monte Carlo method [28, 15, 12], or even independent thereof, such as the Monte Carlo method, are more appropriate. The Monte Carlo approach is simple to use, permits deterministic codes to be reused, and can be implemented in parallel. Despite this, the root mean square

rate of convergence is often slow, decaying as $\frac{\sigma}{\sqrt{n}}$, where σ is the standard deviation of the quantity of interest and n is the number of random samples. To reach an estimate within an acceptable level of tolerance, a large number of evaluations of the quantity of interest are required, each of which involves a numerical simulation of the underlying partial differential equation.

Even when polynomial-based methods such as stochastic collocation are not feasible, e.g. when the parameter space has a meaningfully high dimension (in the case of rough fields), or when the statistical quantity of interest depends non-smoothly on the underlying random parameters, they can still be used as surrogate models to efficiently resolve its variations over *subsets* of the parameter space. In this dissertation, we propose a hybrid sampling scheme which uses conditional sampling to combine sparse grid quadrature rules on a low-dimensional projection of the parameter space with a Monte Carlo scheme to compensate in the remaining dimensions.

Specifically, let θ_0 be some low-dimensional approximation of the underlying parameter θ. If θ is a square integrable random field, θ_0 may represent its truncated Karhunen-Loève (KL) expansion [14, 16] or θ_0 might be a local spatial average or filtered version of θ. Using the law of total expectation, we write the statistical quantity of interest as the iterated integral

$$\mathbb{E}[J(\theta)] = \mathbb{E}_{\theta_0}\left[\mathbb{E}_{\theta|\theta_0}\left[J(\theta)|\theta_0\right]\right],$$

where $\mathbb{E}_{\theta_0}$ is the expectation over the low-dimensional parameter space and $\mathbb{E}_{\theta|\theta_0}$ is the conditional expectation given θ_0. Since the outer integral is computed over

a lower dimensional region, it can be approximated by polynomial-based interpolatory quadrature rules introduced in Chapter 3. However, evaluating its integrand $\mathbb{E}_{\theta|\theta_0}[J(\theta)|\theta_0]$ at a quadrature node requires the computation high-dimensional *conditional* expectation, which we will approximate by Monte Carlo sampling. Despite the fact that the computational cost of this approach (in terms of the number of function evaluations) amounts to the product of the number of polynomial-based quadrature points (for the outer integral) and the number of Monte Carlo samples per quadrature point (for the inner integral). There are a two reasons to believe that this can be done efficiently. Firstly, if the low-dimensional approximation θ_0 is responsible for most of the variance of $J(\theta)$ then the conditional variance of $J(\theta)$ given θ_0 will be relatively small and hence the number of Monte Carlo samples required at each θ_0-quadrature node will also be small. Secondly, the stochastic simulation at a given realization of the low-dimensional field θ_0 can be used to form an approximation of $J(\theta)$ for θ near θ_0, e.g. by means of linearization at θ_0, which can be used to further improve the efficiency of the Monte Carlo sample through variance reduction techniques (Chapter 5) or to lower the cost of function evaluations for the conditional sample.

Using complexity arguments, we show that our method is more efficient than either of its constituents. This is achieved through repeated evaluations of the forward model for different realizations of the random inputs.

1.1 Problem Setting and Notational Conventions

Throughout this dissertation we assume that the probability triple $(\Omega, \mathbb{P}, \mathcal{F})$ consisting of sample space Ω, probability measure $\mathbb{P}$, and σ-algebra $\mathcal{F}$, forms a complete probability space. We will use uppercase letters X, Y, and Z to denote random variables and lowercase Greek letters, such as η, θ for random fields. The expectation of a random variable X is written as $\mathbb{E}[X]$ while its variance is $\mathbb{V}(X)$. The covariance between random variables X and Y is denoted by $\mathrm{Cov}(X, Y)$ and their correlation is given by $\rho(X, Y)$. We will use subscripts on these mappings when we want to emphasize the underlying distributions. We will denote by $L^p(\Omega)$ the space of p-integrable random variables and by $L^p(\Omega, W)$ the space of W-valued p-integrable random fields, i.e.

$$L^p(\Omega, W) = \{\theta : \Omega \to W | \left(\int_\Omega \|\theta\|_W^p \, d\omega \right)^{\frac{1}{p}} < \infty\},$$

where W is a normed function space with norm $\|\cdot\|_W$. The function space $C^k(D)$ consists of k continuously differentiable functions on D, with $C^\infty(D)$ being infinitely differentiable functions on D. $W^{p,s}$ denotes the Sobolev space of functions with p-integrable weak derivatives up to order s. Specifically, $H^s(D)$ is the Sobolev space of functions with square-integrable weak derivatives up to order s. We will use boldface font to indicate vectors, e.g. $\mathbf{X} = [X_1, \ldots X_k]^T \in \mathbb{R}^n$, or multi-indices, e.g. $\mathbf{i} = (i_1, ..., i_d) \subset \mathbb{N}^d$.

The problem discussed in this dissertation is that of estimating statistical quantities of interest of the form

$$Q = \mathbb{E}[J(\theta)] = \int_{\Omega} J(\theta(\omega))d\omega,$$

where J is a nonlinear functional related to the spatially varying random parameter θ, possibly via the partial differential equation (1.1), and the parameter θ is contained in the appropriate parameter space

$$\Theta := L^2(\Omega, W(D)),$$

where $W(D)$ is an appropriately defined function space determined by the parameter's underlying covariance kernel (see Section 2.2).

1.2 Outline of the Dissertation

The major goal of this dissertation is to propose a more efficient method than Monte Carlo or sparse grid methods for estimating statistical quantities of interest associated with a stochastic partial differential equation's random model output. First, Chapter 2 will cover the properties, representation, and numerical approximation of spatially varying Gaussian random fields. Stochastic sampling methods like Monte Carlo and sparse grid methods will be explored in Chapter 3. The proposed hybrid sampling method will be introduced in Chapter 4. The algorithm and the error estimates will be discussed, and I will compare the hybrid sampling method and the Monte Carlo method, and the sparse grid method. In Chapter 5, I will go

over variance reduction methods used in improving the convergence rate of Monte Carlo method. Finally, to demonstrate the new method, I will conduct numerical experiments in Chapter 6.

Chapter 2

Spatially Varying Random Fields

In this chapter, we present some background information on spatially variable random fields, which will be utilized to define the random inputs to the system. We shall talk about their representation, approximation, and properties that will be relevant in the coming chapters of this book. Further, there is a spectral form of the square integrable random field that allows us to write it as an infinite sum that we may truncate, which is a more particular representation. A more in-depth discussion of Gaussian random fields is presented since they are frequently employed in modeling and because they lend themselves well to marginalization, conditioning, and simulation.

2.1 Preliminary Definitions and Results

We begin by introducing some standard mathematical definitions that will be useful in the subsequent chapters. Constants are described as random variables, whereas spatially varying parameters are modeled as random fields to represent the uncertainty in the models underlying system parameters. We assume in this work that the random variables and random fields are explicitly defined.

Definition 2.1 (Probability Triple [29])**.** A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a sample space (set of outcomes); $\mathcal{F}$ is a collection of events that include

the sample space Ω and the empty set $\emptyset$ with the property that $\mathcal{F}$ is closed under countable intersections and unions, complements; $\mathbb{P}$ is a probability function that assigns probability to events in $\mathcal{F}$, i.e. $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\emptyset) = 0$.

Definition 2.2 (Random Variable [29])**.** A real valued random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is an $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable mapping $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Certain statistics and moments of the random variables X, such as expectations, or higher order moments are of interest in the context of applications.

Definition 2.3 (Expectation and Variance [33])**.** The expectation and variance of a random variable X are defined by,

$$\mathbb{E}[X] := \int_{\Omega} X(\omega) d\omega, \quad \text{and } \mathbb{V}(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right],$$

respectively.

Definition 2.4 (Correlation and Covariance)**.** Suppose we have two random variables X_1 and X_2 then, the covariance and correlation of X_1 and X_2 are defined as

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$$

and

$$\rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathbb{V}(X_1)}\sqrt{\mathbb{V}(X_2)}}$$

respectively.

As we will see in Chapter 4, the hybrid sampling method is dependent on specific conditioning arguments, and the rate at which the conditional variance decays plays

a significant role in the process. As a result, we provide the following definitions and go into greater detail about the conditioning of random fields in Section 2.4.

Definition 2.5 (Conditional Density)**.** Given two random variables X_1 and X_2 with conditional probability density function $f(x_1|X_2)$. The conditional expectation of X_1 given X_2 is

$$\mathbb{E}[X_1|X_2] = \int_\Omega x_1 f(x_1|X_2) dx_1 \tag{2.1}$$

and the conditional variance of X_1 given X_2 is given by:

$$\mathbb{V}[X_1|X_2] = \int_\Omega (x_1 - \mathbb{E}[X_1|X_2])^2 f(x_1|X_2) dx_1 \tag{2.2}$$

Notice that the above conditional statistics are also random variables since they depends on X_2.

The Strong Law of Large Numbers, stated below, is commonly used to show convergence of the Monte Carlo sampling method, presented in Chapter 3.

Theorem 2.1 (Strong Law of Large Numbers [34])**.** *let $\{X_i\}_{i=1}^\infty$ be a sequence of independent and identically distributed random variables with values in $\mathbb{R}^d$. Assume that*

$$\mathbb{E}[X_i] < \infty, i = 1, 2, \ldots, \qquad \sum_{i=1}^{\infty} \frac{\mathbb{V}(X_i)}{i^2} < \infty.$$

For $n \geq 1$, denote the empirical mean of $X_1, \ldots, X_n$ by

$$\hat{S}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$$

Then,

$$\lim_{n\to\infty} \hat{S}_n = \mathbb{E}[X_1], \qquad \mathbb{P}\text{-}\textit{almost surely.}$$

Because Gaussian random fields and their properties are frequently used to characterize the uncertainties in stochastic modeling, we will define them and their properties below. Also, Marginalization, conditioning, and simulation work effectively with them.

Definition 2.6 (Gaussian random vector [24])**.** A real-valued random vector $X : \Omega \to \mathbb{R}^d$ is said to be Gaussian (or normally distributed) if it has the density function

$$\rho(x) = \left(\frac{1}{2\pi\sigma}\right)^{\frac{d}{2}} \exp^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)},$$

where $\mu \in \mathbb{R}^d$ is the mean of X and is a positive definite covariance matrix. We abbreviate this by writing $X \sim N(\mu, \Sigma)$.

Definition 2.7 (Random Fields [33])**.** Let $D \subset \mathbb{R}^d$, a random field is a mapping $\theta : D \times \Omega \to \mathbb{R}$, such that $\theta(x, .)$ is measurable for every $x \in D$. A random filed is called centered if $\mathbb{E}[\theta(x)] = 0$ for all $x \in D$ and is mean square continuous if

$$\lim_{y\to x} \mathbb{E}[\theta(y) - \theta(x)] = 0$$

for any $x \in D$.

Definition 2.8. Let $\theta : D \times \Omega \to \mathbb{R}$ be a Random Field. For a fixed $\omega \in \Omega$, we define $\theta(\cdot, \omega) : D \to \mathbb{R}$ by $x \mapsto \theta(x, \omega)$ to be a realization of the random field.

Definition 2.9. The covariance function $R(x, y)$ of a random field θ is defined as

$$R(x, y) = \mathbb{E}[\theta(x) - \mathbb{E}\theta(x)][\theta(y) - \mathbb{E}\theta(y)]$$

. [5]

Definition 2.10. Consider a random field θ. If the joint distribution

$$F_{x_1,...,x_m}(z_1, ..., z_m) = P[\theta(x_1) \leq z_1, ..., \theta(x_m) \leq z_m]$$

is invariant under translation

$$(x_1, ..., x_m) \to (x_1 + \tau, ..., x_m + \tau),$$

θ is said to be stationary or homogeneous. For a stationary random field θ, we can show

$$\mathbb{E}\left[\theta(x)\right] = \mathbb{E}\left[\theta(0)\right]$$

and subsequently

$$R(x, y) = f(x - y)$$

for some function f. [5]

Definition 2.11. *A random vector* $\mathbf{X} = (X_1, ..., X_m)$ is multivariate normal if $\sum_{i=1}^{m} c_i X_i$ is Gaussian for every possible $c_i \in \mathbb{R}$ [5].

Definition 2.12. *A random field* θ is a Gaussian random field if $\theta(x_1), ..., \theta(x_m)$ are multivariate normal for every $(x_1, ..., x_m) \in \mathbb{R}^m$ [5].

Definition 2.13. θ is a mean zero Gaussian field if $\mathbb{E}\theta(x) = 0$ for all x.

Definition 2.14. The integration of Gaussian fields is also Gaussian (see[5]).

2.2 The Karhunen-Loève Expansion

The Karhunen-Loève (KL) expansion is a spectral expansion that is frequently used to provide a computationally feasible representation of a random field. Through the separation of deterministic and stochastic variables in the coefficient $\theta(x, \omega)$, it is possible to reduce the problem (1.1) to a deterministic (albeit infinite dimensional) one [9]. Specifically, let $\theta \in \Theta$ be a second order random field on D with mean $\mu(x) := \mathbb{E}[\theta(x)]$ and continuous, symmetric, positive semi-definite covariance kernel $k : D \times D \to \mathbb{R}$ defined by $k(x, y) = \mathbb{E}\left[(\theta(x) - \mu(x))(\theta(y) - \mu(y))\right]$. Mercer's theorem [19] stated below, asserts that the integral operator defined via k is a compact Hilbert-Schmidt operator on $L^2(D)$ and therefore has a countable complete orthonormal spectral basis.

Theorem 2.2 (Mercer's Theorem [19])**.** *The operator* $\mathcal{K} : L^2(D) \to L^2(D)$ *defined by*

$$\mathcal{K}u(x) := \int_D k(x, y)u(y)dy, \tag{2.3}$$

where k *is given as above, is a Hilbert-Schmidt operator and hence there exists a complete set of* $L^2(D)$*-orthonormal eigenfunctions* $\{\phi_i\}_{i=1}^{\infty}$ *and a decreasing set of*

eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \to 0$ *as* $n \to \infty$*, so that*

$$k(x, y) = \sum_{n=1}^{\infty} \lambda_n \phi_n(x) \phi_n(y).$$

The eigenfunctions guaranteed by Mercer's theorem can be used to construct the Karhunen-Loève expansion, as stated in the well-known theorem below.

Theorem 2.3 (Karhunen-Loève expansion of Random Field [9, 16])**.** *Let* $\theta(x, \omega) \in \Theta$ *be a square-integrable random field with mean* $\mu(x)$ *continuous, symmetric, and positive semi-definite covariance kernel* $k(x, y)$ *defined on* $D \times D$*. Then it can be expanded as*

$$\theta(x, \omega) = \mu(x) + \sum_{n=1}^{\infty} \sqrt{\lambda_n} \phi_n(x) Y_n(\omega) \tag{2.4}$$

where

$$Y_n(\omega) = \int_D \theta(x, \omega) \phi_n(x) dx, \qquad n = 1, 2, ...$$

are centered, pairwise uncorrelated random variables on the probability space $(\Omega, \mathbb{P}, \mathcal{F})$ *and* $\{\lambda_n\}_{n=1}^{\infty}$ *and* $\{\phi_n\}_{n=1}^{\infty}$ *are the eigenvalues and eigenvectors of the covariance kernel* $\mathcal{K}$*, guaranteed by Mercer's theorem. The convergence is in the* $L^2(\Omega)$*-sense in probability and uniform over* D*. Moreover, the variance satisfies*

$$\mathbb{V}(\theta(x)) := \mathbb{E}\left[(\theta(x) - \mu(x))^2\right] = \sum_{n=1}^{\infty} \lambda_n \left(\phi_n(x)\right)^2. \tag{2.5}$$

As a direct result of Equation (2.5) and the orthonormality of $\{\phi_n\}_{n=1}^{\infty}$, we have

$$\mathbb{E}\left[\|\theta - \mu\|_{L^2(D)}^2\right] = \sum_{n=1}^{\infty} \lambda_n.$$

When the underlying random field is Gaussian, then the random variables $\{Y_n\}_{n=1}^{\infty}$ are independent standard normal. This renders Gaussian fields easy to simulate.

Corollary 2.1 (KL Expansion of Gaussian fields)**.** *The random variables in the Karhunen-Loève expansion of a Gaussian random field are independent standard normal random variables, i.e.* $Y_n \sim N(0,1)$ *for* $n = 1, 2, \ldots$.

Remark 2.1. In general, the joint distribution of the random variables $\{Y_n\}_{n=1}^{\infty}$ is difficult to determine. Nevertheless, researchers often use the KL expansion as an Ansatz, considering random fields of the form

$$\theta(x,\omega) = \mu(x) + \sum_{n=1}^{\infty} \sqrt{\lambda_n}\phi_n(x)Y_n(\omega),$$

where $Y_n \sim U(-1,1)$ are independent, identically distributed uniform random variables.

2.2.1 Karhunen-Loève Eigenvalue Decay

For purpose of computation, the Karhunen-Lève expansion given by Equation (2.4) is impractical and is usually replaced with a truncated KL expansion θ_0 of the form

$$\theta_0(x,\omega) = \mu(x) + \sum_{i=1}^{k} \sqrt{\lambda_i}\phi_i(x)Y_i(\omega), \tag{2.6}$$

where k is some fixed truncation level. In light of Equation (2.5), the mean-squared truncation error is given by

$$\mathbb{E}\left[\|\theta - \theta_0\|^2\right] = \sum_{i=k+1}^{\infty} \lambda_i, \tag{2.7}$$

which in depends on the decay rate of the covariance operator's eigenvalues. This decay rate also influences the stochastic regularity of the random field. Here we will discuss the decay criteria for the KL eigenvalue sequence $\{\lambda_j\}_{j=1}^{\infty}$ since the stochastic regularity and complexity of the stochastic methods used in solving the model are determined by the KL eigenvalue decay (see Chapter 3). The smoother the covariance kernel of the coefficient, the faster the KL eigenvalue decay, with analyticity giving exponential decay and finite Sobolev regularity implying algebraic decay [9].

Definition 2.15 (Smoothness of covariance kernels [9, 26])**.** If D is a bounded domain in $\mathbb{R}^d$, a covariance function $k : D \times D \to \mathbb{R}$ is said to be piecewise analytic/smooth/$H^{p,q}$ on $D \times D$ if there exists a finite family $(D_j)_{1\leq j\leq J} \subset \mathbb{R}^d$ of open hypercubes such that

$$\bar{D} \subset \bigcup_{j=1}^{J} \bar{D}_j \tag{2.8}$$

$D_j \cap D_{j'} = \emptyset \quad \forall j \neq j'$ and $k|_{D_j \times D_{j'}}$ has an analytic/smooth/$H^p \otimes H^q$ continuation in a neighborhood of $\bar{D_j} \times \bar{D_{j'}}$ for any pair (j, j'). Here $\bar{D}$ denotes the closure of D.

Proposition 2.1 (see [9]). *If the covariance kernel k is piecewise analytic on $D \times D$ with $D \subset \mathbb{R}^d$ and if $(\lambda_j)_{j\geq 1}$ is the eigenvalue sequence associated with 2.3, then there exist constant $c_1, c_2 > 0$ such that*

$$0 \leq \lambda_j \leq c_1 e^{-c_2 j^{1/d}} \quad \forall j \geq 1. \tag{2.9}$$

Proposition 2.2 (see [9]). *If $k \in L^2(D \times D)$ is symmetric and piecewise $H^{p,0}(D)$ with $p \geq 0$ and $D \subset \mathbb{R}^d$, then for the associated symmetric, non-negative and compact covariance operator defined by (2.3) it holds*

$$0 \leq \lambda_j \leq c_3 j^{-p/d} \quad \forall j \geq 1, \tag{2.10}$$

for some constant $c_3 > 0$ independent of j or p.

The eigenvalue decay rate outlined in the previous two propositions, implies the following decay rate in the truncation error of the Karhunen-Loève expansion.

Proposition 2.3 (Truncation Error (see [9])). *If $k(x, y)$ is piecewise analytic/smooth on $D \times D$, then the truncated KL expansion θ_0 of θ converges in $L^2(D \times \Omega)$ as*

$$\|\theta - \theta_0\|_{L^2(D\times\Omega)} \lessapprox \begin{cases} c_4 e^{-c_2(1/2-s)k^{1/d}} & \text{if } k \text{ piecewise analytic} \\ c_5 k^{1-p(1-s)/d} & \text{if } k \text{ piecewise smooth} \end{cases} \tag{2.11}$$

for all $k \in \mathbb{N}$, for any $s > 0, p \geq 0$ and with constants c_4 and c_5 depending on d, s, c_1, and c_2.

2.2.2 Numerical Approximation of KL expansions Using Finite Element Methods

An efficient numerical approximation of the eigenpairs of the operator connected to the covariance kernel via (2.3) is needed for the efficient numerical approximation of Problem (1.1) [9]. The variational form of the eigenvalue problem is:

$$\int_{D\times D} k(x,y)\phi_j(y)v(x)dydx = \lambda_j \int_D \phi_j(x)v(x)dx \quad \forall v \in L^2(D). \tag{2.12}$$

To compute KL eigenpairs, finite elemenent (FE) discretization of (2.12) is done with piecewise polynomial functions on a regular, quasi-uniform triangulation of D with meshwidth h. [18], [9]. Let S_h^0 denote the finite element space of discontinous, piecewise constant functions. The Galerkin approximation of (2.12) with the finite element space $S_h^0 \subset L^2(D)$ reads: Find $0 \neq \lambda_j^h, \phi_j^h \in S_h^0$ such that

$$\int_{D\times D} k(x,y)\phi_j(y)^h v(x)dydx = \lambda_j^h \int_D \phi_j(x)^h v(x)dx \quad \forall v \in S_h^0. \tag{2.13}$$

For the eigenpair approximation error we have the following pointwise bound

Proposition 2.4 (see [9])**.**

$$\begin{cases} \|\phi_j - \phi_j^h\| = O(h), \ \ as\ h \to 0, \\ |\lambda_j - \lambda_j^h| = O(h), \ \ as\ h \to 0. \end{cases} \tag{2.14}$$

The calculation of KL eigenpairs involves the solution of the dense matrix eigenproblem corresponding to proposition (2.4), i.e of

$$Pv = \lambda Mv, \tag{2.15}$$

where both matrices P and M are symmetric and positive definite, with the mass matrix M being diagonal if we choose as basis of S_h^0 the characteristic functions of elements in the triangulation.

2.3 Approximation of Random fields by spatial averaging method

The spatial averaging approach is another method for approximating the random field, which approximates the random field θ by a set of random variables $\hat{\theta}_i, i = 1 \ldots, d$, where each random variable represents the local average of θ over the domain Ω_i defined by the following integral. [27] [33]

$$\hat{\theta}_i = \frac{R_i}{V_i} = \frac{1}{V_i} \int_{\Omega_i} \theta(\omega) d\omega \tag{2.16}$$

where V_i is the volume of the spatial domain Ω_i. R_i is the local integral of $\theta(x)$ over the domain Ω_i.

For Gaussian random fields, the averaging random variables are fully define by a mean vector and a covariance matrix that have been determined through spatial integration of the Random field moment functions. The mean of the random variable $\hat{\theta}_i$ is determined by integrating the mean $\mathbb{E}_\theta(x)$ over the averaging domain Ω_i and is

define by:

$$\mathbb{E}(\hat{\theta}_i) = \frac{1}{V_i} \int_{\Omega_i} \mathbb{E}_\theta(\omega) d\omega \tag{2.17}$$

and the variance is found by integration over the spatial autocovariance $K_\theta(x_1, x_2)$:

$$\mathbb{V}(\hat{\theta}_i) = \frac{1}{V_i^2} \int_{\Omega_i} \int_{\Omega_i} K_\theta(\omega_1, \omega_2) d\omega_1 d\omega_2 \tag{2.18}$$

[27][33].

The spatial averaging methods' approximation error often propagates through the model response, and the error in the model response may be higher or smaller than the error in the random field model, depending on the quantity of interest. If $\mathcal{Q}$ and $\hat{\mathcal{Q}}(x)$ represent the quantity of interest and its approximation, then the point-wise approximation error is defined as $\epsilon(x) = \mathcal{Q}(x) - \hat{\mathcal{Q}}(x)$. The normalized mean bias and normalized variance bias are defined using the bias error measure as follows [27], [31]:

$$\epsilon_\mu(X) = \frac{\mathbb{E}[\mathcal{Q}(x)] - \mathbb{E}[\hat{\mathcal{Q}(x)}]}{\mathbb{E}[\mathcal{Q}(x)]} \tag{2.19}$$

$$\epsilon_\mathbb{V}(X) = \frac{\mathbb{V}[\mathcal{Q}(x)] - \mathbb{V}[\hat{\mathcal{Q}(x)}]}{\mathbb{V}[\mathcal{Q}(x)]} \tag{2.20}$$

Taking the weighted integral of (2.19) and (2.20) over the domain Ω gives the corresponding global error measures [27], [23]:

$$\bar{\epsilon}_\mu = \frac{1}{V}\int_\Omega \epsilon_\mu(X)dx \tag{2.21}$$

$$\bar{\epsilon}_{\mathbb{V}} = \frac{1}{V}\int_\Omega \epsilon_{\mathbb{V}}(X)dx \tag{2.22}$$

2.4 Conditional Random Fields

In Chapter 4, we decompose the random field into a low dimensional approximation θ_0 and a high dimensional residual. The KL expansion lends itself readily to such a decomposition, i.e. the approximation is given by

$$\theta_0 = \sum_{j=1}^{k} \sqrt{\lambda_j}\phi_j(x)Y_j(\omega),$$

while the residual is

$$\Delta\theta = \theta - \theta_0 = \sum_{j=k+1}^{\infty} \sqrt{\lambda_j}\phi_j(x)Y_j(\omega).$$

Note that for Gaussian fields, the residual is independent of the approximation, and hence the conditional random field θ given samples of θ_0 is simply

$$\theta_0 + \Delta\theta.$$

This decomposition, while convenient may not be how the random field is usually approximated in practice. Sometimes some prior information about the random field may be known observations or coarse-scale simulation. The observation may take several forms such as direct measurements of the field in a region i.e $\boldsymbol{\theta}_{iB} = b$ with $i_b \subset \{1, \ldots, n\}$, or its image under a linear mapping define by $A\boldsymbol{\theta} = \mathbf{b}$ with $A \in \mathbb{R}^k$ or a process $\mathbf{D}$ with $\mathbf{D}|\boldsymbol{\theta} \sim N(A\boldsymbol{\theta}, \Sigma_{\mathbf{e}})$ [25]. A conditioning of random field can be on the following:

- **Conditioning on a subset of points** [25]: for $\Sigma \succ 0$, $i_A \subset \{1, \ldots, n\}$ and $i_b := i_A^c$ be its complement and if we partition $\boldsymbol{\theta} \sim N(\mu, Q^{-1})$ such that $\boldsymbol{\theta} = [\boldsymbol{\theta}_A, \boldsymbol{\theta}_B]^T$ where

$$Q = \begin{bmatrix} Q_{AA} & Q_{AB} \\ Q_{BA} & Q_{BB} \end{bmatrix}$$

  then $\boldsymbol{\theta}_{A|B} := \boldsymbol{\theta}_A | \{\boldsymbol{\theta}_B = \theta_B\} \sim N(\mu_{A|B}, Q_{A|B}^{-1})$ where

$$\mu_{A|B} = \mu_A - Q_{AA}^{-1} Q_{AB} (\theta_B - \mu_B), \tag{2.23}$$

  and

$$Q_{A|B} = Q_{AA} \tag{2.24}$$

- **Conditioning on Hard Constraints** [25]: For $\boldsymbol{\theta}$ define as above and $A \in \mathbb{R}^{k \times n}$ a full rank matrix with $0 < k < n$, the conditional distribution of $\boldsymbol{\theta}$ given

the linear constraint $A\boldsymbol{\theta} = \mathbf{e}$ for some vector $\mathbf{e} \in \mathbb{R}^k$ is :

$$\begin{bmatrix} \boldsymbol{\theta} \\ A\boldsymbol{\theta} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ A\mu \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma A^T \\ A\Sigma & A\Sigma A^T \end{bmatrix} \right)$$

with

$$\mu|\{A\boldsymbol{\theta} = \mathbf{e}\} = \mu - \Sigma A^T \left(A\Sigma A^T\right)^{-1} (A\mu - \mathbf{e}) \tag{2.25}$$

and

$$\Sigma|\{A\boldsymbol{\theta} = \mathbf{e}\} = \Sigma - \Sigma A^T \left(A\Sigma A^T\right)^{-1} (A\Sigma) \tag{2.26}$$

- **Conditioning on the truncated Karhunen-Loève Expansion** Here we consider the conditioning of a Gaussian random field on known measurements or samples of its truncated KL expansion as a specific instance. Because the terms in the series are independent, the conditional mean and covariance are calculated as follows:

$$\begin{aligned}\mathbb{E}(\theta|\theta_{k_0}) &= \mathbb{E}(\theta_{k_0}) + \sum_{j=k_0+1}^{\infty} \mathbb{E}(\sqrt{\lambda_j}\phi_j(x)Y_j) \\ &= \mathbb{E}(\mathbb{E}_\theta(x) + \sum_{j=1}^{k_0} \sqrt{\lambda_j}\phi_j(x)Y_j) + \sum_{j=k_0+1}^{\infty} \mathbb{E}(\sqrt{\lambda_j}\phi_j(x)Y_j) \\ &= \mathbb{E}_\theta(x)\end{aligned} \tag{2.27}$$

and

$$\begin{aligned}\mathbb{V}(\theta|\theta_{k_0}) &= \mathbb{E}[(\theta|\theta_{k_0} - \mathbb{E}(\theta|\theta_{k_0}))^2] \\ &= \mathbb{E}[(\theta_{k_0} + \sum_{j=k_0+1}^{\infty} \sqrt{\lambda_j}\phi_j(x)Y_j - \mathbb{E}_\theta(x))^2] \\ &= \mathbb{E}[(\mathbb{E}_\theta(x) + \sum_{j=1}^{k_0} \sqrt{\lambda_j}\phi_j(x)Y_j + \sum_{j=k_0+1}^{\infty} \sqrt{\lambda_j}\phi_j(x)Y_j - \mathbb{E}_\theta(x))^2] \\ &= \mathbb{E}[(\sum_{j=1}^{k_0} \sqrt{\lambda_j}\phi_j(x)Y_j + \sum_{j=k_0+1}^{\infty} \sqrt{\lambda_j}\phi_j(x)Y_j)^2] \\ &= \mathbb{E}[(\sum_{j=1}^{\infty} \sqrt{\lambda_j}\phi_j(x)Y_j)^2] \\ &= \sum_{j=1}^{\infty} \mathbb{E}\left(\lambda_j\phi_j^2(x)\right)\end{aligned} \tag{2.28}$$

- **Conditioning on Soft Constraints**[25] : Suppose the linear transformation $A\theta$ of θ is observed with some additional noise i.e $e|\theta \sim N(A\theta, \Sigma_\epsilon)$, then $\theta|e \sim$

$N_C(b_{\theta|e}Q_{\theta|e})$, where

$$b_{\theta|e} = Q\mu + A^T\Sigma_\epsilon^{-1}e \tag{2.29}$$

$$Q_{\theta|e} = Q + A^T\Sigma_\epsilon^{-1}A \tag{2.30}$$

Chapter 3

Stochastic Sampling Methods

Let $J : \Theta \to \mathbb{R}$ be a smooth function depending on the parameter $\theta \in \Theta$, possibly via the solution u of the associated partial differential equation (1.1). Suppose we are interested in computing a quantity of interest $Q := \mathbb{E}[J(\theta)]$ defined by:

$$Q := \mathbb{E}\left[J(\theta)\right] = \int_{\Omega} J(\theta) d\omega.$$

This chapter will discuss two different sampling methods that are used in approximating the quantity of interest Q, specifically the Monte Carlo method and the stochastic collocation method. Both approximate the integral Q by a weighted sum,

$$\int_{\Omega} J(\theta) d\omega \approx \sum_{i=1}^{n} w_i J(\theta_i),$$

where $w_1, ..., w_n$ are a set of quadrature weights and $\theta_1, ..., \theta_n$ the corresponding quadrature nodes. Both methods can be implemented fully in parallel, since evaluations of J are computationally independent. Differences in efficiency are therefore directly attributable to the sample size required to reach a given level of accuracy. In Section 3.1 we will outline the Monte Carlo sampling method, based on independent random sampling of the underlying parameter and in Section we will introduce

stochastic collocation, an interpolation-based quadrature scheme that can be adopted to moderate dimensional problems.

This is of the form above, were the $w_i = \int_\Gamma L_i(y)\rho(y)dy$. For one-dimensional interpolatory quadrature rules, Gauss nodes are often preferred, i.e. the nodes corresponding to the roots of polynomials that are orthogonal in the ρ- weighted L2 space. This guarantees a precision of $2n+1$, i.e. polynomials of this degree can be integrated exactly using these quadrature rules. If we assume a Gaussian random field, then Hermite basis is natural (orthogonal w.r.t. Gauss weights). Other useful quadrature points are the Chebyshev points, which are applicable on bounded domains. In the multi-dimensional setting, sparse grid methods give us a way to produce efficient high accuracy rules by taking tensor products of appropriate one-dimensional rules.

3.1 The Monte Carlo Method

The Monte Carlo method is the most widely used method for approximating high-dimensional integrals. Implementation of the Monte Carlo approach is straightforward, as it makes use of already existing deterministic codes. The rate of convergence, on the other hand, is quite slow, depending on the reciprocal of the square root of the sample size as well as the variance of the integrand. It would thus take a hundredfold increase in the sample size to achieve every additional digit of accuracy. On the other hand, the dimension of the underlying parameter space has only a mild influence on the error, namely via its influence on the variance [21].

In the classical Monte Carlo method, we estimate the statistical quantity of interest Q by

$$Q \approx \mathcal{Q}^{MC}[J] := \frac{1}{n}\sum_{i=1}^{n} J(\theta_i), \tag{3.1}$$

where $\theta_1, \theta_2, ..., \theta_n$ is a sequence of identically distributed samples of the parameter θ, generated according to the underlying distribution $\mathbb{P}$.

Since the parameter sample is independent and identically distributed (i.i.d), the estimator $\mathcal{Q}^{MC}$ is an unbiased estimator of Q since

$$\mathbb{E}\left[\mathcal{Q}^{MC}[J]\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[J(\theta_i)\right] = Q$$

The convergence of this estimate is guaranteed by Theorem 2.1, which states that $\mathcal{Q}^{MC}$ converges almost surely to the true expected value, i.e.

$$\lim_{n\to\infty} Q^{MC}[J] = Q, \qquad \mathbb{P}\text{-a.s.},$$

provided the variance of the individual terms is bounded i.e $\mathbb{V}[\theta]$ is bounded [35]. Suppose $\mathbb{V}[\theta] < \infty$, we have

$$\mathbb{V}[\mathcal{Q}^{MC}] = \mathbb{E}[(\mathcal{Q} - \mathcal{Q}^{MC})^2] = \frac{\mathbb{V}(J)}{N}$$

and the root mean square error of $\mathcal{Q}^{MC}$ is $\sqrt{C}$, which means the error estimates of the estimator is of order $n^{-1/2}$ written as $RMSE = \mathcal{O}(n^{-1/2})$. The error estimate

may be inverted to show the number of samples needed to yield a desired error ϵ [17]:

$$N = \frac{\mathbb{V}(J)}{\epsilon^2}$$

One of the strength of Monte Carlo Method is that the sample values can be used for error estimation [35]. The most common used estimators of $\mathbb{V}(J)$ are

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\theta_i - \mathcal{Q}^{MC})^2, \tag{3.2}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(\theta_i - \mathcal{Q}^{MC})^2, \tag{3.3}$$

and the estimator in (3.2) is unbiased since $\mathbb{E}[\sigma^2] = \mathbb{V}(J)$ for $N \geq 2$. From the central limit theorem, the error $\mathcal{Q}^{MC} - \mathcal{Q}$ has approximately a normal distribution with mean 0 and variance $\frac{\mathbb{V}(J)}{N}$ [35]. we summarise the Monte Carlo Method in the folllowing algorithm:

Algorithm 1 Monte Carlo Sampling Algorithm

1: Identify the random Field θ form the model

2: Identify the variablity of θ by assuming a probability density function ρ_θ

3: Generate $n \in N$ i.i.d samples from $\rho_t heta$, $\{\theta_i, \ldots, \theta_N\} \sim \rho_t heta$

4: **for** $i = 1, ...n$ **do**

5: compute $J(\theta_i)$

6: **end for**

7: Compute $E[J(\theta)] \approx Q^{\mathrm{MC}} = \sum_{i=1}^{n} \frac{1}{N} J(\theta_i)$

.

Some variance reduction approaches are used to increase the convergence rate of the Monte Carlo method, where the knowledge about the sensitivity of the function J is exploited [3] (see chapter 5). Correlation between subsequent samples can also be used to one's advantage. Control variates employ an easy integrand to reduce variance, and if successive random variables are negatively associated, the variance will be less than if they were independent. The antithetic variates technique takes use of the reduction in variance that occurs when negatively correlated samples are purposefully formed and grouped together [17].

3.2 The Stochastic Collocation Method

Before we get onto the other component of the hybrid sampling (Sparse Grid Methods). We shall define the various sorts of polynomials that are used in approximating functions based on the type of random field we are utilizing. Polynomials

such as the Lagrange polynomial are employed in uniform random fields, whereas Hermite polynomials are utilized in Gaussian random fields.

3.2.1 One-Dimensional Interpolation

It is known that given n points in a plane (y_k, f_k), $k = 1, \ldots, n$ with distinct y_k's, there is a unique polynomial in y of degree less than n whose graph passes through the points. Suppose we fit a polynomial of degree n-1,

$$P(y) = a_{n_1} y^{n-1} + a_{n-2} y^{n-2} + \dot{a}_1 y + a_0$$

through the n known points (y_k, f_k), $k = 1, \ldots, n$. This polynomial is called the interpolating polynomial because it exactly reproduces the given data: $P(y_k) = f_k \quad k = 1 \ldots n$. There are many different formulas for the polynomial, but they all define the same function [10].

Global Polynomial Interpolation

One way to construct the polynomial $P(y)$ is to use the Lagrange polynomials of degree $n - 1$:

Definition 3.1. Let

$$L_k = \prod_{i=1, i \neq k} \left(\frac{y - y_i}{y_k - y_i} \right), \quad k = 1, \ldots, n. \tag{3.4}$$

The Lagrange form of the polynomial of degree n-1 interpolating the set of points $\{(y_k, f_k) | k = 1, \ldots, n\}$ and satisfying $P(y_k) = f_k$ is defined by,

$$P(y) = \sum_{k=1}^{n} f_k L_k(y) = \sum_{k=1}^{n} f_k \prod_{i=1, i\neq k} \left(\frac{y - y_i}{y_k - y_i} \right) \tag{3.5}$$

[10].

Theorem 3.1. ***(The error term in global polynomial interpolation)*** *If f is a function that is n times differentiable in [a,b] with continuous derivatives then $\forall y \in [a, b]$, there exist $\eta \in (a, b)$, that depends on y, such that*

$$f(y) - P(y) = \frac{f^{(n)}(\eta)}{n!} \prod_{k=1}^{n} (y - y_k). \tag{3.6}$$

Hermite Polynomial

The Hermite interpolation matches the function f, and the observed value of its first n derivatives. The Hermite polynomial of degree n is defined as

$$H_n = (-1)^n e^{y^2} \frac{d^n}{dy^n} e^{y^2} \tag{3.7}$$

physicist's Hermite polynomials and

$$H_n = (-1)^n e^{y^2/2} \frac{d^n}{dy^n} e^{y^2/2} \tag{3.8}$$

probabilist's Hermite polynomials [6].

The Hermite polynomials are orthogonal i.e

$$\int_{-\infty}^{\infty} w(y)H_n(y)H_m(y)dy = 0 \quad if n \neq m. \tag{3.9}$$

for $w(y) = e^{-y^2}$ in case of physicist's Hermite polynomials or $w(y) = e^{-(y/2)^2}$ for probabilist's Hermite polynomials. The nth Hermite polynomial satisfies the following differential equation:

$$\frac{d^2}{dy^2}\left[H_{n+1}(y)\right] - 2y\frac{d}{dy}\left[H_n(y)\right] + 2nH_n(y) = 0 \tag{3.10}$$

and have the following recurrence relation:

$$H_{n+1}(y) = 2yH_n(y) - 2nH_{n-1}(y) \tag{3.11}$$

In addition we have:

$$H_n^{'} = 2nH_{n-1}(y) \tag{3.12}$$

Local Piecewise Polynomial Interpolants

The most common local piecewise polynomial interpolant are the hat functions obtained by dividing the interval into n subinterval of equal lengths h. [20]

Definition 3.2. The family of first $(n+1)$ 1D-hat basis functions on the interval $[0, T]$ are defined as

$$\phi_0(x) := \begin{cases} \frac{h-x}{h} & 0 \geq x \geq h \\ 0, & \text{otherwise} \end{cases}$$

$$\phi_i(x) := \begin{cases} \frac{x-(i-1)h}{h} & (i-1)h \geq x \geq ih \\ \frac{(i+1)h-x}{h} & ih \geq x \geq (i+1)h \\ 0, & \text{otherwise} \end{cases}$$

and

$$\phi_n(x) := \begin{cases} \frac{x-(T-h)}{h} & T-h \geq x \geq T \\ 0, & \text{otherwise} \end{cases}$$

3.2.2 Sparse Grid Tensorization

Here we will describe two different types of collocation techniques used in solving numerical integrals namely the full grids methods and the sparse grid methods.

Collocation Techniques on Uniform Random Fields

Suppose $J : [0,1]^d \to \mathbb{R}$, To study J and to solve problems involving J, the idea in the sparse grid methods is to replace J with another function $f : [0,1]^d \to \mathbb{R}$ that approximates J well and is much cheaper to evaluate[32]. f is been constructed by interpolation of J i.e evaluate J at a small number of points $\mathbf{x_k}$ and use the values $\mathbf{J}(\mathbf{x_k})$ to define f. We use the following definitions to show how the sparse grids method on uniform random fileds is obtained.

Definition 3.3 (See [32])**.** Suppose we discretized the domain of J by splitting it into 2^n segments of equal size. The d dimensional full grid of level n is given by the following cartesian product:

$$\Omega_n := \Omega_{n_1} \times ... \times \Omega_{n_d}$$

of the one dimensional grids Ω_{n_i} and $d-$dimensional grid points are defined as

$$\mathbf{x_{n,i}} = (x_{n_1 i_1}, ..., x_{n_d i_d})$$

.

Definition 3.4. The multivariate basis function is constructed from the univariate basis functions $\phi_{n_j,i_j} : [0,1] \to \mathbb{R}$ as a tensor product:

$$\phi_{\mathbf{n},\mathbf{i}}(\mathbf{x_{n,i}}) := \phi_{n_1,i_1}(x_{n_1,i_1})...\phi_{n_d,i_d}(x_{n_d,i_d}).$$

Each grid point $\mathbf{x_{n,i}}$ corresponds to one basis function $\phi_{\mathbf{n},\mathbf{i}}$. The most common choice of $\phi_{\mathbf{n},\mathbf{i}}$ are the hat functions defined by

$$\phi_{n_j,i_j}(x) := \max(1 - |2^{n_j}x - i_j|, 0).$$

[32].

Definition 3.5. The nodal Space V_n of level n is defined as the linear span of all basis functions $\phi_{\mathbf{n},\mathbf{i}}$:

$$V_n = \text{span}\{\phi_{\mathbf{n},\mathbf{i}} | i_j = 0, \ldots, 2^n \text{ for } j = 1, ..., d\}.$$

[32]

Therefore, the corresponding interpolant function is simply:

$$f(\mathbf{x}) := \sum_{i_1=0}^{2^{n_1}} ... \sum_{i_d=0}^{2^{n_d}} \mathbf{c}_{\mathbf{n},\mathbf{i}} \phi_{\mathbf{n},\mathbf{i}}(\mathbf{x}), \qquad \mathbf{c}_{\mathbf{n},\mathbf{i}} \in \mathbb{R},$$

where $c_{\mathbf{n},\mathbf{i}} = J(\mathbf{x}_{\mathbf{n},\mathbf{i}})$. The size of Ω grows rapidly in higher dimensions (curse of dimensionality). To address this a change of basis described below is performed so that basis functions and their grid points have different levels of importance, but the same functions that is continuous and piecewise linear [32].

Definition 3.6. Let

$$\mathbf{I_n} := \begin{cases} \{1, 3, 5, ..., 2^n - 1\} & for \quad n \geq 1 \\ \{0, 1\} & for \quad n = 0. \end{cases}$$

Then,

$$\Omega_n = \cup_{l=0}^{n} \tilde{\Omega}_l, \quad \Omega_n := \{\mathbf{x}_{\mathbf{n},\mathbf{i}} | \mathbf{i} \in \mathbf{I_n}\}$$

The corresponding interpolating function is

$$f(x) := \sum_{l=0}^{n} \sum_{i \in I_n} \alpha_{n,i} \phi_{n,i}(x), \quad \alpha_{n,i} \in \mathbb{R}.$$

Where $\{\phi_{n,i} | 0 \leq l \leq n, i \in I_n\}$ are called the hierarchical basis and $\{\alpha_{n,i} | 0 \leq n \leq n, i \in I_n\}$ are called hierarchical surpluses.

Definition 3.7. (Full Grids) The d-dimensional full grid

$$\mathbf{\Omega_n} := \bigcup_{n_1=0}^{n_1} \ldots \bigcup_{n_d=0}^{n_d} \tilde{\Omega}_{n_1} \times \ldots \times \tilde{\Omega}_{n_d}$$

and the interpolant function is given by:

$$f(x) := \sum_{\mathbf{l}=0}^{\mathbf{n}} \sum_{\mathbf{i} \in \mathbf{I_n}} \alpha_{\mathbf{n},\mathbf{i}} \phi_{\mathbf{n},\mathbf{i}}(x), \quad \alpha_{\mathbf{n},\mathbf{i}} \in \mathbb{R}.$$

Regular Sparse Grids

The size of the support of $\phi_{n,i} = 2^{-n+1}$ is smaller in high levels which means that these functions only have little influence on the resulting linear combination f whereas basis function of a low level $n \leq 1$ even have a large support. In d dimension, the area of the support of $\phi_{\mathbf{n},\mathbf{i}}$ is

$$(2h_{n_1})...(2h_{n_d}) = 2^{\|n\|_1 + d}, \quad \|n\|_1 := n_1 + ... + n_d,$$

where $||n||_1$ is the norm of n, i.e the area decreases with increasing $||n||_1$. Sparse grid supposed that functions with high $\|n\|_1$ only contribute little to the solution

and omits it. Hence we only use the levels $\mathbf{n}$ whose 1-norm doesn't exceed a specific threshold $n \geq 0$:

$$f(\mathbf{x}) := \sum\nolimits_{\|\mathbf{l}\|_1 \leq \mathbf{n}} \sum\nolimits_{\mathbf{i} \in \mathbf{I_n}} \alpha_{\mathbf{n},\mathbf{i}} \phi_{\mathbf{n},\mathbf{i}}(\mathbf{x}), \quad \alpha_{\mathbf{n},\mathbf{i}} \in \mathbb{R}.$$

The corresponding grid points from the regular sparse grid $\Omega^s_{n,d}$ of level n is:

$$\mathbf{\Omega^s_{n,d}} := \{\mathbf{x_{n,i}} \| \mathbf{l} \|_1 \leq \mathbf{n}, \quad \mathbf{i} \in \mathbf{I_n}\}.$$

Smolyak Approximation

Given an ensembles $\{J(y_i)\}_{i=1}^m$, the interpolation approach finds an approximation $\mathscr{Q}(J)$ of J that fulfils the condition $\mathscr{Q}(J)(y_i) = J(y_i) \quad j = 1, \ldots, m$. Such that

$$\mathscr{Q}(J)(y) = \sum_{j=1}^{m} J(y_i) L_i(y) \tag{3.13}$$

where $L_i(y)$ are the Lagrange polynomials.

To get the approximation in multiple dimensions (k>1), the following full tensor product rule can be use

$$\mathscr{Q}^{m_1} \otimes \cdots \otimes \mathscr{Q}^{m_k}(J) = \sum_{i_1=1}^{m_1} \cdots \sum_{i_1=1}^{m_k} J(y_{i1}, \ldots, y_{ik}).(L_{j1} \otimes \cdots \otimes L_{jk}) \tag{3.14}$$

.

A more efficient method is the **Smolyak grids interpolation** It takes a subset of the full tensor define by

$$\mathscr{A}(w,k) = \sum_{w-k+1\leq|i|\leq w} (-1)^{q-|i|} \begin{pmatrix} k-1 \\ w-|i| \end{pmatrix} (\mathscr{Q}^{m_1} \otimes \cdots \otimes \mathscr{Q}^{m_k}) \tag{3.15}$$

where $w \geq k$ is an integer denoting the level of construction. To compute the operator $\mathscr{A}(w,k)$, one needs to evaluate J on the set of points

$$\mathscr{H}(w,d) = \bigcup_{w-k+1\leq|i|\leq w} (\tau^{i1} \times \cdots \times \tau^{ik}) \tag{3.16}$$

where $\tau^i = \{y_1^i, \ldots, y_{m_i}^i\} \subset [-1.1]$ is the collection of nodes used by univariate interpolating operator $\mathscr{Q}^i$, where τ^i denotes the set of abscissas used by $\mathscr{Q}^i$,the choice of which can be the Clenshaw-Curtis abscissas or the Gaussian abscissas. The set $\mathscr{H}(w,d)$ is a much smaller subset of those required by the full tensor product rule.

Collocation Techniques on Gaussian Random Fields

Suppose $J : \mathbb{R}^d \to \mathbb{R}$. The procedure for getting the sparse quadrature for $\mathbb{E}(J(\theta))$ When θ is a Gaussian random field is similar to what we discussed above only that in this case the so called Gauss-Hermite quadrature rule is used [4]. The density ρ is used as the weight function, where $y_0 = 0$ and $w_0 = 1$ for $n = 0$, and for $n \geq 1$, $y_i, i = 0, \ldots, m_n - 1$ are the roots of the orthonormal (with respect to ρ)

Hermite polynomial H_n for $n = m_l$, where

$$H_n(y) = \frac{(-1)^n \rho^{(n)}(y)}{\sqrt{n!}\rho(y)}$$

and the weights $w_i, k = 0, 1, \ldots, m_n - 1$ are given by

$$w_i = \frac{1}{m_n^2 (H_{m_{l-1}}(y_i))^2} \tag{3.17}$$

Therefore the full tensorization of J takes the form

$$J(y) = \sum_{m \in M} J_m H_m \tag{3.18}$$

where the multivariate Hermite polynomials H_s and J_s are define as:

$$H_m(y) = H_{m_i}(y_i), \text{ and } J_m = \int J(y) H_m(y) d\Omega(y) \tag{3.19}$$

and the sparse quadrature is define as:

$$\mathscr{Q}_\Lambda(J) = \sum_{m \in \Lambda} \bigotimes_k (\mathscr{Q}^{mk} - \mathscr{Q}^{mk-1}(J)) \tag{3.20}$$

3.2.3 Error Estimates

In this section we cite some well-known convergence results [21, 2] for the sparse grid stochastic collocation method. The cited estimates were derived from interpolation error estimates for the constituent one-dimensional rules, which were then combined within the Smolyak grid. For the purpose of this section, we suppose that

the random parameter θ has been approximated by its finite noise approximation $\theta_0 : \Gamma \times D \to \mathbb{R}$, where $\Gamma = \prod_{i=1}^{k} \Gamma_i \subset \mathbb{R}^k$ is a hyper-rectangle denoting the range of the random variables $Y_1, ..., Y_k$ with joint probability density function ρ. Further, let

$$\mathcal{Q}[J(\theta_0)] \approx \mathbb{E}[J(\theta_0)]$$

be the numerical approximation of the statistical quantity of interest using the sparse grid stochastic collocation method. Since these are based on polynomial interpolation, their convergence requires some degree of smoothness for $J(\theta_0)$ as a function of the random variables $Y_1, ..., Y_k$, as summarized in the following two assumptions, which can generally be established on a case-by-case basis.

Assumption 3.1 (Bounded Mixed Derivatives). Assume that the mapping $J : \Gamma \to \mathbb{R}$ given by $y_1, ..., y_k \mapsto J(\theta_0(y_1, ..., y_k))$ is continuously differentiable in each variable with bounded partial derivatives in each direction.

A stronger assumption is given by the following.

Assumption 3.2 (Analytic Extension). Assume that in each component y_j, the function $J(y_1, .., y_j, .., y_k)$ has an analytic extension to a region in the complex plane for every fixed points $y_1, .., y_{j-1}, y_{j+1}, ..., y_k$.

The error estimates for the low-dimensional quadrature rules depend primarily on the dimension k of the parameter space and the smoothness of the integrand and is stated in the following theorem.

Theorem 3.2 (Smolyak Convergence [21]). *Suppose that the mapping J satisfies Assumption 3.1 or Assumption 3.2. Then there are constants independent of the*

sample size n so that

$$\|\mathcal{Q}[J(\theta_0)] - \mathbb{E}[\theta_0]\|_{L^2} \leq Cn^{-\frac{\gamma}{\zeta+\log(k)}}.$$

Chapter 4

The Hybrid Sampling Approach

In this chapter, we will derive the hybrid sampling method, obtain an estimate of its error, and compare it to the error estimates obtained from each of its constituents (Monte Carlo and Sparse Grid). We describe the conditions in which the hybrid approach performs much better.

Suppose θ_0 is a lower scale, k_0-dimensional finite-noise approximation of the random field θ, obtained for example by means of a truncated Karhunen-Loève (KL) expansion or some other low-resolution representation. By the law of total expectation,

$$Q = \mathbb{E}\left[J(\theta)\right] = \mathbb{E}_{\theta_0}\left[\mathbb{E}_{\theta|\theta_0}\left[J\right]\right], \tag{4.1}$$

where $\mathbb{E}_{\theta_0}$ is the marginal expectation with respect to θ_0 and $\mathbb{E}_{\theta|\theta_0}$ is the conditional expectation of J with respect to θ, given θ_0.

The hybrid sampling approach we proposed uses an efficient quadrature rule $\mathcal{Q}_{\theta_0}$ suited to low dimensional parameter spaces to approximate the outer integral, while employing a sampling method $\mathcal{Q}_{\theta|\theta_0}$ that is more robust to the parameter space dimensionality, such as the Monte Carlo method, to resolve the inner integral. Specifically, let $\theta_0^1, \ldots, \theta_0^{n_0}$ and $w_1, \ldots, w_{n_0}$ be the quadrature nodes and weights that determine the low-dimensional quadrature rule $\mathcal{Q}_{\theta_0}$ and let $\{J_j^i\}_{j=1}^{n_i}, i = 1, \ldots, n_0$ be independent and identically distributed samples of the quantity $J(\theta)$ given $\theta_0 = \theta_0^i$.

Identity (4.1) now suggests the series of approximations

$$\begin{aligned}\mathbb{E}_{\theta_0}[\mathbb{E}_{\theta|\theta_0}[J(\theta)|\theta_0]] &\approx \mathcal{Q}_{\theta_0}\left[\mathbb{E}_{\theta|\theta_0}\left[J(\theta)|\theta_0\right]\right] = \sum_{i=1}^{n_0} w_i \mathbb{E}_{\theta|\theta_0}\left[J(\theta)|\theta_0 = \theta_0^i\right] \\ &\approx \mathcal{Q}_{\theta_0}\left[\mathcal{Q}_{\theta|\theta_0}\left[J(\theta)|\theta_0\right]\right] = \sum_{i=1}^{n_0} w_i \left(\frac{1}{n_i}\sum_{j=1}^{n_i} J_j^i\right).\end{aligned}$$

Note that the 'finite noise' truncation parameter k moderates the contributions of the sparse grid stochastic collocation and the conditional Monte Carlo sampling, with $k = 0$ corresponding to a full Monte Carlo sample, and the contribution of the sparse grid collocation growing as $k \to \infty$. As shown in Chapter 2 on the truncated Karhunen-Loève (KL) expansion, the decay rate of the covariance operator's eigenvalues suggests a level of k above which the contributions of successive KL terms are negligible, i.e. a practical upper bound for the truncation level. When the field is smooth and J is a smoothly varying function in θ, then The resulting full tensor product quadrature rule requires n_i Monte Carlo simulations for each low-scale quadrature point θ_0^i, $i = 1, \ldots, n_0$, amounting to a total of $\sum_{i=1}^{n_0} n_i$ function evaluations.

4.1 Hybrid Sampling Algorithm

The approach above is summarized in Algorithm 2 below.

Algorithm 2 Hybrid sampling algorithm

Set truncation levels $k_0 \ll k$

for $i = 1, ...n_0$ **do**

  Sample field θ_0^i on sparse grid

  for $j = 1, ..., n_i$ **do**

    Sample θ^j, given that $\theta_0 = \theta_0^i$

    Compute $J_j^i = J(\theta^j)$.

  end for

  $\mathcal{Q}_{\theta|\theta_0}[J|\theta_0^i] = \frac{1}{n_i} \sum_{j=1}^{n_i} J_j^i$

end for

$\mathcal{Q} = \sum_{i=1}^{n} w_i \mathcal{Q}_{\theta|\theta_0}[J|\theta_0^i]$.

4.2 Convergence Analysis of The Hybrid Sampling Method

In this section we establish the convergence of the hybrid sampling method for any truncation level k, as the sample sizes n_0, and $\{n_i\}_{i=1}^{n_0}$ approach infinity.

Theorem 4.1 (Convergence)**.** *For any fixed truncation level k and tolerance level $\varepsilon > 0$, we can find n_0 and n_i, $i = 1, ..., n_0$ large enough so that*

$$\left\| \mathcal{Q}_{\theta_0} \left[\mathcal{Q}_{\theta|\theta_0} \left[J(\theta)|\theta_0 \right] \right] - \mathbb{E}\left[J(\theta) \right] \right\|_{L^2(\Omega)} < \varepsilon.$$

Proof. Using the law of total expectation and adding and subtracting $Q_{\theta_0}\left[\mathbb{E}_{\theta|\theta_0}[J(\theta)|\theta_0]\right]$, we decompose the error

$$\begin{aligned}
&\mathbb{E}\left[J(\theta)\right] - \mathcal{Q}_{\theta_0}\left[\mathcal{Q}_{\theta|\theta_0}\left[J(\theta)|\theta_0\right]\right] \\
&\qquad = \left(\mathbb{E}_{\theta_0} - \mathcal{Q}_{\theta_0}\right)\left[\mathbb{E}_{\theta|\theta_0}\left[J(\theta)|\theta_0\right]\right] + \mathcal{Q}_{\theta_0}\left[\mathbb{E}_{\theta|\theta_0}[J(\theta)|\theta_0] - \mathcal{Q}_{\theta|\theta_{k_0}}[J(\theta)|\theta_0]\right]. \quad (4.2)
\end{aligned}$$

For simplicity we do not include spatial discretization. To be more specific, we shall require that both error contributions be of size ε.

The first term describes the quadrature error for the sparse grid method in integrating the conditional expectation $f(\theta_0) := \mathbb{E}_{\theta|\theta_0}\left[J(\theta)|\theta_0\right]$ over the range of θ_0. Under the appropriate smoothness assumptions, of f on θ_0 it can be shown using the results outlined in Chapter 3. That is, [21]

$$\left\|\left(\mathbb{E}_{\theta_{k_0}} - \mathcal{Q}_{\theta_{k_0}}\right)\left[\mathbb{E}_{\theta|\theta_{k_0}}\left[J\right]\right]\right\|_{L^2}^2 \sim C^{k_0} n_0^{-\frac{\gamma}{\zeta+\log(k_0)}}. \quad (4.3)$$

The sample size n_0 needed to achieve an error of ε must therefore satisfy

$$n_0 \geq \left(\frac{C^{k_0}}{\varepsilon}\right)^{\frac{\zeta+\log(k_0)}{\gamma}} \quad (4.4)$$

The second term is the weighted sum of quadrature errors associated with computing various conditional expectations by means of the dimension-independent Monte Carlo quadrature rule. To control the second error term, let $\sigma^i_{\theta|\theta_{k_0}} = \sqrt{\mathbb{V}_{\theta|\theta_{k_0}}(J)}$,

where $\mathbb{V}_{\theta|\theta_{k_0}}(J)$ is the conditional variance of J given θ_{k_0}, i.e.

$$\mathbb{V}_{\theta|\theta_{k_0}}(J) = \mathbb{E}_{\theta|\theta_{k_0}}\left[\left(J - \mathbb{E}_{\theta|\theta_{k_0}}[J|\theta_{k_0}]\right)^2 |\theta_{k_0}\right].$$

If we choose $n_i \geq \left(\frac{\sigma^i_{\theta|\theta_{k_0}}}{\varepsilon}\right)^2$ for $i = 1, \ldots, n_0$ we have, by independence,

$$\left\|\mathcal{Q}_{\theta_{k_0}}\left[\mathbb{E}_{\theta|\theta_{k_0}}[J] - \mathcal{Q}_{\theta|\theta_{k_0}}[J]\right]\right\|^2_{L^2(\Omega)} = \mathbb{E}_\theta\left[\left(\mathcal{Q}_{\theta_{k_0}}\left[\mathbb{E}_{\theta|\theta_{k_0}}[J] - \mathcal{Q}_{\theta|\theta_{k_0}}[J]\right]\right)^2\right] \quad (4.5)$$

$$= \sum_{i=1}^{n_0} w_i^2 \mathbb{E}_\theta\left[\left(\mathbb{E}_{\theta|\theta_{k_0}}[J] - \mathcal{Q}_{\theta|\theta_{k_0}}[J]\right)^2\right] = \sum_{i=1}^{n_0} w_i^2 \frac{\left(\sigma^i_{\theta|\theta_{k_0}}\right)^2}{n_i} \leq \sum_{i=1}^{n_0} w_i^2 \varepsilon^2 = \varepsilon^2 \|w\|^2_{l^2}, \quad (4.6)$$

where $w = [w_1, \ldots, w_{n_0}]^T$ is the vector of quadrature weights for $\mathcal{Q}_{\theta_{k_0}}$. It can be shown [22] that the sparse grid quadrature weights, while not non-negative in general, are bounded in norm by

$$\|w\|_{l^2} = O((\log(n_0^{k_0}))^{k_0-1}) = O(\log(\varepsilon)^{k_0-1}), \quad (4.7)$$

if n_0 is chosen as above.

Substituting (4.7), (4.5) and (4.3) into (4.2), we have

$$\left\|\mathbb{E}_{\theta_{k_0}}\left[\mathbb{E}_{\theta|\theta_{k_0}}[J]\right] - \mathcal{Q}_{\theta_{k_0}}\left[\mathcal{Q}_{\theta|\theta_{k_0}}[J(\theta)|\theta_{k_0}]\right]\right\|_{L^2(\Omega)}$$

$$\leq C^{k_0} n_0^{-\frac{\gamma}{\zeta + \log(k_0)}} + \sqrt{\sum_{i=1}^{n_0} w_i^2 \frac{\left(\sigma^i_{\theta|\theta_0}\right)^2}{n_i}} \leq 2\varepsilon. \quad (4.8)$$

□

4.3 Efficiency of The Hybrid Sampling Method

With the error estimates on hand, we can analyze the efficiency of the hybrid sampling algorithm.

Definition 4.1. The ε-cost of a sampling algorithm is the computational work (number of samples) required to achieve an error below a given $\varepsilon > 0$.

For example, The ε cost of the Monte Carlo sample is

$$\sqrt{\frac{\mathbb{V}(J)}{n}} < \varepsilon \Rightarrow n > \frac{\mathbb{V}(J)}{\varepsilon^2}.$$

Letting $\sigma^2_{\theta|\theta_{k_0}} = \max_{i=1,\dots,n_0} \{(\sigma^i_{\theta|\theta_{k_0}})^2\}$, the ε-cost of the hybrid approach is approximately

$$n_0 \left(\frac{\sigma_{\theta|\theta_{k_0}}}{\varepsilon}\right)^2 \sim \sigma^2_{\theta|\theta_{k_0}} \left(\frac{1}{\varepsilon}\right)^{2+\frac{\zeta+\log(k_0)}{\gamma}}.$$

Hence, efficiency of the hybrid sampling method hinges both on:

- The efficiency of the low-dimensional quadrature rule in resolving the low-dimensional approximation of Q,
- The efficiency of high-dimensional rule in computing the conditional expectations to sufficient accuracy.

4.3.1 Comparison of The Hybrid Sampling Method and Monte Carlo Method

To justify using the hybrid approach over Monte Carlo sampling in terms of computational complexity, we compare the ε-costs of both methods and require

$$\sigma^2_{\theta|\theta_{k_0}} \left(\frac{1}{\varepsilon}\right)^{2+\frac{\zeta+\log(k_0)}{\gamma}} < \sigma^2_\theta \left(\frac{1}{\varepsilon}\right)^2 \qquad \Rightarrow \qquad \frac{\sigma^2_{\theta|\theta_{k_0}}}{\sigma^2_\theta} < \varepsilon^{\frac{\zeta+\log(k_0)}{\gamma}} \tag{4.9}$$

In comparison with the Monte Carlo method, the hybrid sampling scheme is thus favored when (i) the low-dimensional quadrature rule has a high convergence rate and (ii) when the conditional variance given θ_{k_0} is low relative to the quantity of interest's total variance.

Lemma 4.1 (Yakasai and Van Wyk)**.** *If v_θ is piecewise analytic/smooth on $D \times D$, and the family $Y = (Y_j)_j \geq 1$, then the KL expansion of θ converges uniformly on $D \times \Omega$ and*

$$\mathbb{V}(J(\theta)|\theta_{k_0}) \lesssim \|J'(\theta_{k_0})\|^2 \begin{cases} e^{-2bk_0} & \text{if } v_\theta \text{ pw analytic} \\ k_0^{-2r} & \text{if } v_\theta \text{ pw smooth} \end{cases}$$

for k_0 large enough and any any b, $r \geq 0$ depending on the decay of the eigenpairs of v_θ

Proof. Let θ be a random field and $d\theta = \theta - \theta_{k_0} = \sum_{j=k_0+1}^{\infty} \sqrt{\lambda_j}\phi_j Y_j(\omega)$

Then,

$$\begin{aligned}
&\mathbb{V}(J(\theta)|\theta_{k_0}) \qquad (4.10)\\
&\quad = \mathbb{V}(J(\theta_{k_0} + d\theta))\\
&\quad = \mathbb{V}\left(J(\theta_{k_0}) + J'(\theta_{k_0})d\theta + \frac{J''(\theta_{k_0})}{2}d\theta + ...\right)\\
&\quad = J'(\theta_{k_0})\mathrm{Cov}(d\theta)(J'(\theta_{k_0}))^T\\
&\quad = J'(\theta_{k_0}) \sum_{j=k_0+1}^{\infty} \mathbb{E}\left(\lambda_j \phi_j^2(x)\right)(J'(\theta_{k_0}))^T
\end{aligned}$$

Recall from Proposition(2.3) we have,

$$\|\theta - \theta_{k_0}\|_{\mathbf{L}^2(D\times\Omega)} \lessapprox \begin{cases} e^{-bk_0} & \textit{if } v_\theta \textit{ pw analytic} \\ k_0^{-r} & \textit{if } v_\theta \textit{ pw smooth} \end{cases} \quad \forall k_0 \in \mathbb{N} \qquad (4.11)$$

for any b, $r \geq 0$ depending on the decay of the eigenpairs of v_θ [21].

$$\mathrm{Cov}(d\theta) = \mathbb{E}((d\theta)^2) \lesssim \begin{cases} e^{-2bk_0} & \text{if } v_\theta \text{ pw analytic} \\ k_0^{-2r} & \text{if } v_\theta \text{ pw smooth} \end{cases}$$

for k_0 large enough and any $s > 0$, $p \geq 0$.

Hence substituting in equation(4.10) we have,

$$\mathbb{V}(J(\theta)|\theta_{k_0}) \lesssim \|J'(\theta_{k_0})\|^2 \begin{cases} e^{-2bk_0} & \text{if } v_\theta \text{ pw analytic} \\ k_0^{-2r} & \text{if } v_\theta \text{ pw smooth} \end{cases}$$

for k_0 large enough and any any b, $r \geq 0$ depending on the decay of the eigenpairs of v_θ □

Therefore for v_θ pw analytic and using equation(4.9), we have

$$\frac{\sigma^2_{\theta|\theta_{k_0}}(J)}{\sigma^2_\theta(J)} < \varepsilon^{\frac{\zeta+\log(k)}{\gamma}} \implies \frac{\|J'(\theta_{k_0})\|^2 e^{-2bk_0}}{\sigma^2_\theta} < \varepsilon^{\frac{\zeta+\log(k_0)}{\gamma}} \tag{4.12}$$

4.3.2 Comparison of The Hybrid Sampling Method and Sparse grid Method

Now consider a comparison of the ε-cost of the hybrid method with that of applying the low-dimensional quadrature rule to the finite noise approximation of J_0 of J. To ensure an overall error of order $O(\varepsilon)$, we must bound both the truncation error and the quadrature error of the resulting finite noise problem. In particular, we decompose the approximation,

$$\mathbb{E}_\theta[J] - \mathcal{Q}_{\theta_{k_0}}[J_0] = (\mathbb{E}_\theta[J] - \mathbb{E}_{\theta_{k_0}}[J_0]) + (\mathbb{E}_{\theta_{k_0}}[J_0] - \mathcal{Q}_{\theta_{k_0}}[J_0]) \approx \varepsilon, \tag{4.13}$$

where the first term on the right hand side represents the truncation error and the second represents the quadrature error. Now suppose θ is truncated then, the

truncation error is:

$$\mathbb{E}_{\theta}[J] - \mathbb{E}_{\theta_{k_0}}[J_0] \leq \delta(k_0) \tag{4.14}$$

for some monotonic decreasing function $\delta(k) \to 0$ as $k \to \infty$ such as in (2.3) and the quadrature error is of the form:

$$(\mathbb{E}_{\theta_{k_0}}[J_0] - \mathcal{Q}_{\theta_{k_0}}[J_0]) \leq C^{k_0} n_s^{-\frac{\gamma}{\zeta + \log(k_0)}} \approx \varepsilon$$

and given the value of $k_0(\varepsilon)$, the number of collocation points

$$n_s(\varepsilon) \approx \left(\frac{C^{k_0}}{\varepsilon}\right)^{\frac{\zeta_s + \log(k_0)}{\gamma_s}} \tag{4.15}$$

can be found and compare with the number of samples needed to achieve accuracy ε with hybrid sampling, which is

$$n(\varepsilon) \approx \sigma^2_{\theta|\theta_{k_0}}(J) \left(\frac{1}{\varepsilon}\right)^{2 + \frac{\zeta + \log(k_0)}{\gamma}}.$$

For exponential truncation error with $\delta(k_0) = ae^{-bk_0}$, where a and b are constants,

$$k_0(\varepsilon) = \frac{1}{b} \log\left(\frac{a}{\varepsilon}\right) \tag{4.16}$$

and

$$n_s(\varepsilon) \approx \left(\frac{C^{\frac{1}{b}\log\left(\frac{a}{\varepsilon}\right)}}{\varepsilon} \right)^{\frac{\zeta_s + \log(\frac{1}{b}\log(\frac{a}{\varepsilon}))}{\gamma_s}} \tag{4.17}$$

$$n(\varepsilon) \approx \sigma^2_{\theta|\theta_{k_0}}(J) \left(\frac{1}{\varepsilon} \right)^{2+\frac{\zeta + \log(\frac{1}{b}\log(\frac{a}{\varepsilon}))}{\gamma}}.$$

Hence for exponential truncation error case the Hybrid sampling method would be more efficient than sparse grid if

$$2 + \frac{\zeta + \log(\frac{1}{b}\log\left(\frac{a}{\varepsilon}\right))}{\gamma} < \frac{\zeta_s + \log(\frac{1}{b}\log\left(\frac{a}{\varepsilon}\right))}{\gamma_s}. \tag{4.18}$$

Similarly, for Algebraic truncation error with $ak_0{}^{-r}$, where a and r are positive constants. Then,

$$k_0(\varepsilon) = \left(\frac{\varepsilon}{a} \right)^{-r} \tag{4.19}$$

and

$$n_s(\varepsilon) \approx \left(\frac{C^{\left(\frac{\varepsilon}{a}\right)^{-r}}}{\varepsilon} \right)^{\frac{\zeta_s + \log((\frac{\varepsilon}{a})^{-r})}{\gamma_s}} \tag{4.20}$$

$$n(\varepsilon) \approx \sigma^2_{\theta|\theta_{k_0}}(J) \left(\frac{1}{\varepsilon}\right)^{2+\frac{\zeta+\log((\frac{\varepsilon}{a})^{-r})}{\gamma}}.$$

Hence for algebraic truncation error case wen can partially say the Hybrid sampling method would be more efficient than sparse grid if

$$2 + \frac{\zeta + \log((\frac{\varepsilon}{a})^{-r})}{\gamma} < \frac{\zeta_s + \log((\frac{\varepsilon}{a})^{-r})}{\gamma_s} \tag{4.21}$$

Chapter 5

Sensitivity Enhanced Hybrid Methods

First-order Taylor method is used in place of a traditional Monte Carlo simulation to avoid the prohibitive computational expense. Indeed, because of its slow convergence rate and the costly generation of correlated input variables in the multivariate case, Monte Carlo methods are usually reserved to establish a reference against which other methods are compared. However, by identifying and generating stochastic versions of only those parameters to which the output variables are most sensitive and at the same time improving the convergence characteristics of traditional Monte Carlo methods, it is possible to perform simulations utilizing the original model to capture the more intricate behavior that a low-order approximation, such as a first-order Taylor expansion, might otherwise sacrifice.The modified Monte Carlo method(s) we describe below is a step towards this goal [7]. A common approach to improve the convergence rate of the classical Monte Carlo Method, is known as variance-reduction methods. In this section, we review the theory underlying the methods and compare the convergent rate of the two methods.

5.1 Sensitivity Derivative Enhanced Monte Carlo Method (SDMC)

A variance-reduction method was developed that exploits information regarding the sensitivity of J with respect to the random variable θ (measured via derivatives

of J with respect to θ) to improve the convergence characteristics of the Monte Carlo method. The result of their efforts [3] was the sensitivity derivative enhanced sampling method (SDMC). The first-order SDES method is described below. Under the tacit assumption that the appropriate derivatives of the function J exist, consider the first-order Taylor expansion of J about μ_θ :

$$J_1(\theta) = J[u(\mu_\theta)] + J_u[u(\mu_\theta)]u_\theta(\mu_\theta)(\theta - \mu_\theta) \tag{5.1}$$

where u_θ is the sensitivity u with respect to θ and $\mathbb{E}(u) = \mu_\theta$.

Using

$\int \rho(\theta) = 1$ and $\int(\theta - \mu_\theta)p(\theta)d\theta = 0$

We have:

$$\int (J[u(\theta)] - J_1[u(\theta)])\rho(\theta)d\theta = \int (J[u(\theta)])p(\theta)d\theta - J[u(\mu_\theta)] \tag{5.2}$$

which yield the SDMC approximation of $\mathbb{E}(J)$:

$$\mathbb{E}(J) := \mathcal{Q}^{SDMC} \approx J[u(\mu_\theta)] + \frac{1}{N}\sum_{i=1}^{N}\{J[u(\theta_i)] - J_1(\theta_i)\} \tag{5.3}$$

The following inequalities illustrate the extent to which the variance of (5.3) is reduced compared with the variance of the traditional Monte Carlo Method (3.1).

Theorem 5.1. *Let* $m = \max|(d/d\theta)J(\theta)|$ *and* $M = \max|(d^2/d\theta^2)J(\theta)|$*, then:*

$$\mathbb{V}(J) \leq 2m^2\mathbb{V}(\theta) \tag{5.4}$$

and

$$\mathbb{V}(J - J_1) \leq \frac{M^2}{2}(\mathbb{V}(\theta)^2 + \mathbb{E}[(\theta - \mu_\theta)^4]) \tag{5.5}$$

[3]

Equations (5.4) and (5.5) indicate that the SDMC method is more efficient when $\mathbb{V}(\theta)$ is small [3]. It must be emphasized that whereas the convergence rate of the Monte Carlo method depends on $\sqrt{\mathbb{V}(J)/N}$, the SDMC convergence rate for the first-order case depends on $\sqrt{\mathbb{V}(J - J_1)/N}$, where the quantity $\mathbb{V}(J - J_1)$ is simply the variance of the first-order Taylor remainder of the objective function. If the objective function satisfies certain smoothness properties, the variance of the Taylor remainder $\mathbb{V}(J - J_1)$ is always less than that of the objective function $\mathbb{V}(J)$ itself, leading to an order of magnitude faster convergence of SDMC relative to that of the Monte Carlo method. [3]

5.2 Control Variate Sensitivity Enhanced Monte Carlo Method

This method reduces the standard error of Monte Carlo Method by adding to the approximation of $\mathcal{Q} = \mathbb{E}(J)$ in (3.1) a control variate.[30]

Let $C : \mathbb{R}^n \to \mathbb{R}$ be the control variate so that $\mu_C = \mathbb{E}[C(u(\theta))]$ is known. Consider the random variable

$$J^c(u(\theta)) = J(\theta) - \alpha[C(u(\theta)) - \mu_C] \tag{5.6}$$

for some constant α and $J^c(u(\theta))$ is an unbiased estimator for $\mathcal{Q}$.

We can estimate $\mathcal{Q}$ with the following Monte carlo estimator

$$\mathcal{Q}^C = \mathbb{E}(J^C) \approx \frac{1}{N}\sum_{i=1}^{N}\left(J^C(u(\theta_i))\right) = \frac{1}{N}\sum_{i=1}^{N}\left(J(\theta_i) - \alpha[C(u(\theta_i)) - \mu_C]\right) \tag{5.7}$$

(5.7) will have a lower standard error than (3.1) if $\mathbb{V}(J^C)$ is smaller than $\mathbb{V}(J) = \sigma^2$. This will occur if C has a high correlation $\rho_{J,C}$ with J. [13]

$$\mathbb{V}(J^C) = \mathbb{V}[J(\theta_i) - \alpha[C(u(\theta_i)) - \mu_C]] = \sigma^2 + \alpha^2\sigma_C^2 - 2\alpha\sigma\sigma_C\rho_{J,C} \tag{5.8}$$

Where $\mathbb{V}(C) = \sigma_C^2$. Hence, $\mathbb{V}(J^C)$ will be smaller than σ^2 if

$$\rho_{J,C} > \frac{\alpha\sigma_C}{2\sigma} \tag{5.9}$$

$\mathbb{V}(J^C)$ is minimized by setting

$$\alpha = \frac{\sigma\rho_{J,C}}{\sigma_C} \tag{5.10}$$

Therefore, from (5.8)

$$\mathbb{V}(J^C) = \sigma^2(1 - \rho^2_{J,C}). \tag{5.11}$$

So the convergence rate of the Covariate Sensitivity Enhanced Monte Carlo Method depends on $\sigma^2(1 - \rho^2_{J,C})$. [13].

Now, if we choose $C^j = \Delta J^T(\theta - \bar{\theta})$, where $\bar{\theta} = \mathbb{E}(\theta)$,
then $\mathbb{E}(C) = 0$ and $\mathbb{V}(c) = \mathbb{E}[(J'(\bar{\theta})(\theta - \bar{\theta}))^2] = J'(\bar{\theta})^2\mathbb{V}(\theta_j)$ and we have,

$$\rho^2_{J,C} = \frac{\mathrm{Cov}^2(J, C)}{\mathbb{V}(C)\mathbb{V}(J)} \tag{5.12}$$

$$= \frac{\left(J'(\bar{\theta})^2\mathbb{V}(\theta_j) + J(\bar{\theta})\mathbb{E}\left(\frac{J''(\xi)}{2}(\theta_i - \bar{\theta})^3 J'(\theta)\right)\right)^2}{J'(\bar{\theta})^2\mathbb{V}(\theta_j)\mathbb{V}(J)} \tag{5.13}$$

$$\geq \frac{2J'(\bar{\theta})^2\mathbb{V}(\theta_j)J(\bar{\theta})\mathbb{E}\left(\frac{J''(\xi)}{2}(\theta_i - \bar{\theta})^3 J'(\theta)\right)}{J'(\bar{\theta})^2\mathbb{V}(\theta_j)\mathbb{V}(J)} \tag{5.14}$$

$$= \frac{2J(\bar{\theta})\mathbb{E}\left(\frac{J''(\xi)}{2}(\theta_i - \bar{\theta})^3 J'(\theta)\right)}{\mathbb{V}(J)} \tag{5.15}$$

Hence,

$$\mathbb{V}(J)(1 - \rho^2_{J,C}) \leq \mathbb{V}(J)\left(1 - \frac{2J(\bar{\theta})\mathbb{E}\left(\frac{J''(\xi)}{2}(\theta_i - \bar{\theta})^3 J'(\theta)\right)}{\mathbb{V}(J)}\right) \tag{5.16}$$

$$= \mathbb{V}(J) - 2J(\bar{\theta})\mathbb{E}\left(\frac{J''(\xi)}{2}(\theta_i - \bar{\theta})^3 J'(\theta)\right) \tag{5.17}$$

If we want to compare the SDMC method and the control variate method we need to compare equation(5.11) and equation(5.4).

Now, using equation(5.4),

$$\mathbb{V}(J)(1-\rho_{J,C}^2) \leq 2m^2\mathbb{V}(\theta) - \underbrace{2J(\bar{\theta})\mathbb{E}\left(\frac{J''(\xi)}{2}(\theta_i-\bar{\theta})^3 J'(\theta)\right)}_{K} \tag{5.18}$$

Let

$$f(\mathbb{V}(\theta)) = \frac{M^2}{2}(\mathbb{V}(\theta)^2 + \mathbb{E}[(\theta-\mu_\theta)^4]) - 2m^2\mathbb{V}(\theta) + K \tag{5.19}$$

$$= \frac{M^2}{2}\mathbb{V}(\theta)^2 - 2m^2\mathbb{V}(\theta) + \frac{M^2}{2}(\mathbb{E}[(\theta-\mu_\theta)^4]) + K \tag{5.20}$$

Notice that equation(5.19) is quadratic in $\mathbb{V}(\theta)$ with $a = \frac{M^2}{2} > 0$, $b = -2m^2$ and $c = \frac{M^2}{2}(\mathbb{E}[(\theta-\mu_\theta)^4]) + K$.
Let $\Delta = b^2 - 4ac$ then,

- if $\Delta < 0$ we have no real solutions and $f(\mathbb{V}(\theta)) \geq 0$ so, $\mathbb{V}(J-J_1) \geq \mathbb{V}(J)(1-\rho_{J,C}^2)$
- if $\Delta > 0$ and $\mathbb{V}(\theta)$ is between the x-intercepts, then $f(v(\theta)) \leq 0$, so, $\mathbb{V}(J-J_1) \leq \mathbb{V}(J)(1-\rho_{J,C}^2)$. Otherwise,
- $f(v(\theta)) \leq 0$ and $\mathbb{V}(J-J_1) \geq \mathbb{V}(J)(1-\rho_{J,C}^2)$

5.2.1 Computing Sensitivity Using Adjoint Method

Let $x \in \mathbb{R}$ and $p \in \mathbb{R}^{n_p}$.Suppose we have the function $f(x,p) : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}$ and the relationship $g(x,p) = 0$ for a function $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}$ whose partial derivative g_x is eveywhere nonsingular.

one method to approximate the gradient $d_p f$ is to compute n_p finite differences over the element of p. Each finite difference computation requirea solving $g(x,p) = 0$. For moderate large n_p, this can be quite costly. [1]

In the program to solve $g(x,p) = 0$, it is likely that the Jacobian matrix $\partial_x g$ is calculated. The adjoint method uses the transpose of this matrix, g_x^T, to compute the gradient $d_p f$. The computational cost is usually no greater than solving $g(x,p) = 0$ once and sometimes even less costly.

Now,

$$d_p f = d_p f(x(p)) = \partial_x f d_p x (= f_x x_p) \tag{5.21}$$

then,

$$g(x,p) = 0 \quad \textit{everywhere implies}$$

$$d_p g = 0.$$

Expanding the derivative,

$$0 = d_p g = g - p = g_x x_p + g_p p_p$$

which implies;

$$g_x x_p + g_p = 0 \implies x_p = -g_x^{-1} g_p$$

Subtituting this relationship into the above equation yields

$$d_p f = -f_x g_x^{-1} g_p$$

Notice that the expression $-f_x g_x^{-1}$ can now be seen as the solution to the linear equation

$$g_x^T \lambda = -f_x^T \tag{5.22}$$

The matrix conjugate transpose is also called the matrix adjoint and for this reason the vector λ is called the vector of adjoint variables and the linear equation is called the **adjoint equation**. In terms of λ, $d_p f = \lambda^T g_p$. [1]

5.2.2 Adjoint Method For Elliptic Equation

Suppose we want to know the sensitivity of the flux on the right boundary with respect to the variations in the parameter in an elliptic equation i.e we want to compute

$$\frac{d}{d\theta} J(\theta) = \frac{d}{d\theta} f(x, \theta) \tag{5.23}$$

where u solves the equation:

$$g(u, \theta) = 0 \tag{5.24}$$

Then,

$$\frac{df}{d\theta} = \frac{\partial f}{\partial \theta} + \frac{\partial f}{\partial u} u_\theta \tag{5.25}$$

where $u_\theta = \frac{\partial u}{\partial \theta}$ solves the sensitivity equation:

$$\frac{d}{d\theta} g(u, \theta) = \frac{d}{d\theta} 0 \tag{5.26}$$

which gives the sensitivity equation:

$$\frac{\partial g}{\partial u} . u_\theta + \frac{\partial g}{\partial \theta} = 0 \tag{5.27}$$

Solving for u_θ and pluging it back into equation(5.25) we have,

$$\frac{df}{d\theta} = \frac{\partial f}{\partial \theta} - \frac{\partial f}{\partial u} \left(\frac{\partial g}{\partial u} \right)^{-1} \frac{\partial g}{\partial \theta} \tag{5.28}$$

To use the Adjoint method, we let $\lambda^T = \frac{df}{du} \left(\frac{dg}{du} \right)^{-1}$ which gives the following adjoint equation:

$$\left(\frac{\partial g}{\partial u} \right)^T \lambda = \left(\frac{\partial f}{\partial u} \right)^T \tag{5.29}$$

and use the following steps:

- Solve $g(u, \theta) = 0$ for u.
- Solve adjoint equation(5.29) for λ.
- Compute $\frac{df}{d\theta} = \frac{\partial f}{\partial \theta} - \frac{\partial f}{\partial u} \lambda^T \frac{dg}{d\theta}$

Example 5.1. Suppose $f(u,\theta) = \int_{0.8}^{1} u du$, then,

$$\frac{\partial f}{\partial u} = \int_{0.8}^{1} I du \text{ and } \frac{\partial f}{\partial \theta} = 0$$

Hence,

$$\frac{\partial f}{\partial \theta} = \int_{0.8}^{1} u_\theta du.$$

Now, $g(x,\theta) = 0$

$$\implies g(u,\theta) = -\frac{d}{dx}\left(\theta \frac{d}{dx} u\right) - 1 = 0$$

$$\frac{\partial g}{\partial \theta} = -\frac{d}{dx}\left(I \frac{du}{dx}\right)$$

and

$$\frac{\partial g}{\partial u} = -\frac{d}{dx}\left(\theta \frac{d}{dx} I\right)$$

and we get the following adjoint equation:

$$-\frac{d}{dx}\left(\theta \frac{d\lambda}{dx}\right) = \int_{0.8}^{1} dx \tag{5.30}$$

$$\lambda(0) = \lambda(1) = 0$$

with corresponding weak form

$$\int_0^1 \theta \lambda'(x) v'(x) dx = \int_{0.8}^{1} u(x) dx \qquad \text{for all test functions } v(x) \tag{5.31}$$

$$\lambda(0) = \lambda(1) = 0$$

Once we have λ, we compute

$$\frac{df}{d\theta} = \frac{\partial f}{\partial \theta} - \frac{\partial f}{\partial u}\lambda^T \frac{dg}{d\theta} \tag{5.32}$$

$$= -\int_0^1 -\frac{d}{dx}\left(\frac{du}{dx}\right)\lambda dx \tag{5.33}$$

$$= -\int_0^1 Iu'(x)\lambda'(x)dx \tag{5.34}$$

Chapter 6

Numerical Experiment

In this chapter we present some numerical results we obtained by implementing the Hybrid sampling Algorithm 2 on an explicit functional of a random field, and on an elliptic partial differential equations.

6.1 Hybrid Sampling on a Functional

Here, our goal is to compute the statistical quantity of interest $Q = \mathbb{E}[J]$, where

$$J(\omega) = \int_{\frac{3}{4}}^{1} \theta(x,\omega)^2 dx,$$

and $\theta(x,\omega)$ is a Gaussian random field with a given covariance structure. Here, we use the trapezoidal rule to approximate all spatial integrals, i.e.

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n} w_i f(x_i),$$

where $x_0, ..., x_n$ are $n+1$ equally spaced points in the interval $[a,b]$, and $w_0, ..., w_n$ are trapezoidal quadrature weights, i.e. $w_0 = w_n = \frac{h}{2}$ and $w_1 = ... = w_{n-1} = h$ with $h = \frac{b-a}{n}$.

The random field θ is given by a Karhunen-Loeve expansion in the form

$$\theta(x,\omega) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} Y_i \phi_i(x),$$

where $Y_i \sim N(0,1)$ are iid Gaussians and $\{\lambda_i, \phi_i(x)\}_{i=1}^{\infty}$ are the eigenpairs of the covariance operator $\mathcal{K} : L^2[a,b] \to L^2[a,b]$ that maps $u \in L^2[a,b]$ to $\mathcal{K}u(x) = \int_a^b k(x,y)u(y)dy$.

Hence, we have $\theta(x,\omega) = \theta_0(x, Y_1, ..., Y_k) + \theta_r(x, Y_{k+1}, ...)$ and we want to compute

$$J(\theta) = \int_{0.75}^{1} (\theta_0 + \theta_r)^2 dx$$

First, we approximate the eigenfunctions and eigenvalues of the covariance operator $\mathcal{K}$ by

- Using numerical quadrature to approximate $\mathcal{K}$, i.e.

$$\mathcal{K}u(x) \approx \hat{\mathcal{K}}u(x) = \sum_{l=0}^{n} w_l k(x, x_l) u(x_l),$$

- Estimating the ith eigenfunction $\phi_i(x)$ by the vector $v_i = [\phi_i(x_0), ...\phi_i(x_n)]$.

This gives rise to the eigenvalue problem

$$KWv_i = \lambda_i v_i,$$

where

$$K_{ij} = k(x_i, x_j), \quad i, j = 0, ..., n,$$

and

$$W = \text{diag}(w_0, ..., w_n).$$

Letting $u_i = \sqrt{W}v_i$ and simplifying, we can turn this into a symmetric eigenvalue problem

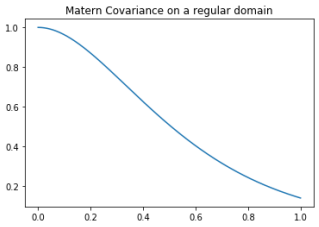
$$\sqrt{W}K\sqrt{W}u_i = \lambda_i u_i, \quad i = 0, ..., n.$$

We used the "karhunen_loeve" module in python, which has a convenience function for computing the eigendecomposition which supports the following covariance functions:

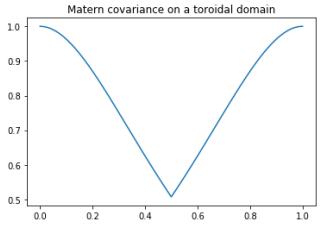
- Gaussian: $k(x, y) = \sigma^2 \exp\left(-\frac{|x-y|^2}{2l^2}\right)$
- Exponential: $k(x, y) = \sigma^2 \exp\left(-\frac{|x-y|}{l}\right)$
- Matern:
- Rational: $k(x, y) = \frac{1}{(1+|x-y|^2)^a}$

The semilog plot of the eigenvalues of the matern covariance function is shown in figure 6.1c.

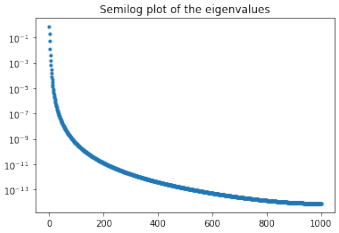
The true mean is $Q = 0.249$ and the true mean on truncated field ($k = 1$) is $Q \approx 0.144$. We used sparse grid method on the truncated filed and got approximately thesame value with true mean of the truncated field i.e $Q \approx 0.144$.



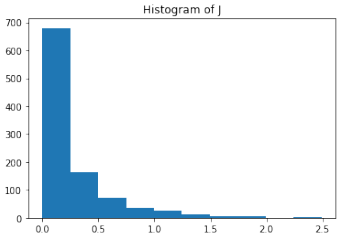
(a) Matern Covariance on a Regular Domain



(b) Matern Covariance on a Regular Domain



(c) Semilog Plot of Eigenvalues

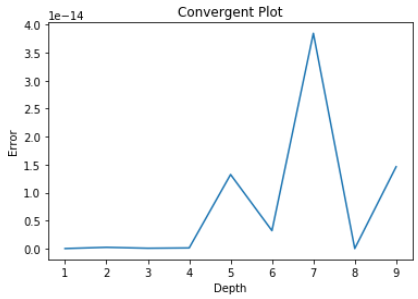


(d) Distribution of Quantity of Interest

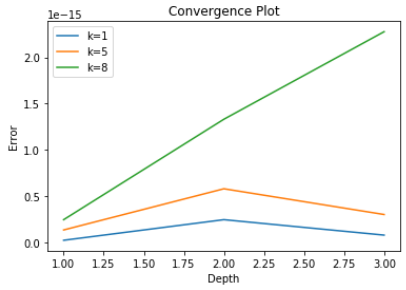
Figure 6.2 shows the convergence rate of the sparse grid method for a fixed k, different depth values and for different k and different depth values, respectively. From the graph we can see that for high value of k and depth the error is increasing. it also shows that for different value of k the sparse grid error is basically zero because polynomial interpolation-based quadrature rules are exact for quadratics and so it captures the integral for truncation levels 1,5,8.

Using the hybrid sampling to calculate the quantity of interest we got $Q \approx 0.24789$ which is approximately thesame with the true mean. Figure 6.3a shows the maximum Monte Carlo error obtained as k is increasing. As expected for high value of k we have fewer MC runs.

Since the sparse grid is capturing the integral on truncated field exactly, we don't need a high level of sparse grid so we fix the truncated field. i.e no sparse



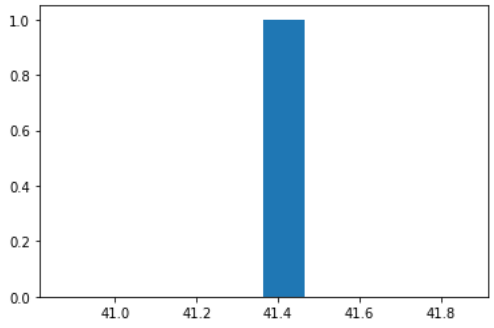
(a) Convergence Plot



(b) Convergent Plot For Different k

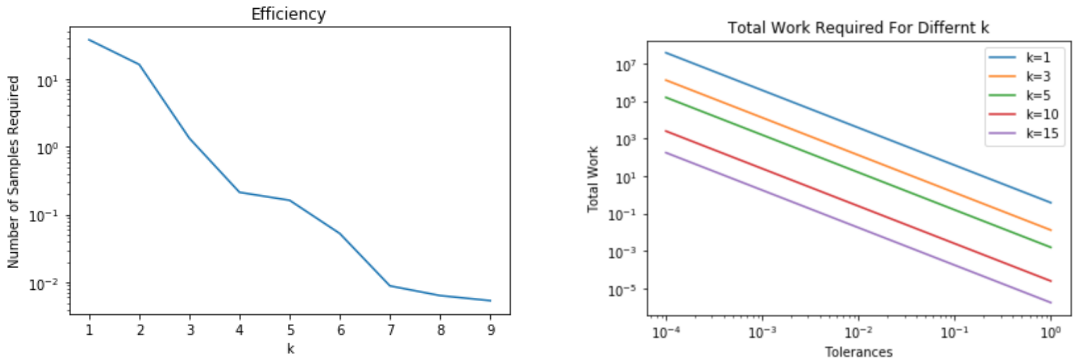
Figure 6.2: Convergence rate of the sparse grid approximation

grid error. Therefore, In figure 6.4a for $tolerance = 0.1$, we look at the number of Monte Carlo runs required as k is increasing, which shows that for high value of k the conditional variance get smaller and therefore we require fewer MC runs. Similarly in figure 6.4b the less you worry about getting things right the less you have to work which is true for any truncation level and if we look at the truncation levels we can see that if k increases the total work decreases. Hence, in this case the sparse grid is very efficient to get the integral.



(a) Histogram of the number of Monte Carlo runs to get certain tolerance

(b) Maximum Mc runs with respect to k



(a) Efficiency (b) Total Work Required For Different k

Figure 6.4: Number of Sample Required for Different k

6.2 Hybrid Sampling on 1D PDE

Consider the problem of determining the expectation of the spatially-averaged flux over a small region adjacent to the right endpoint, i.e.

$$J(q,u) = -\frac{1}{h}\int_{1-h}^{h} q(x,\omega)u'(x,\omega)dx, \tag{6.1}$$

where u satisfies the elliptic problem

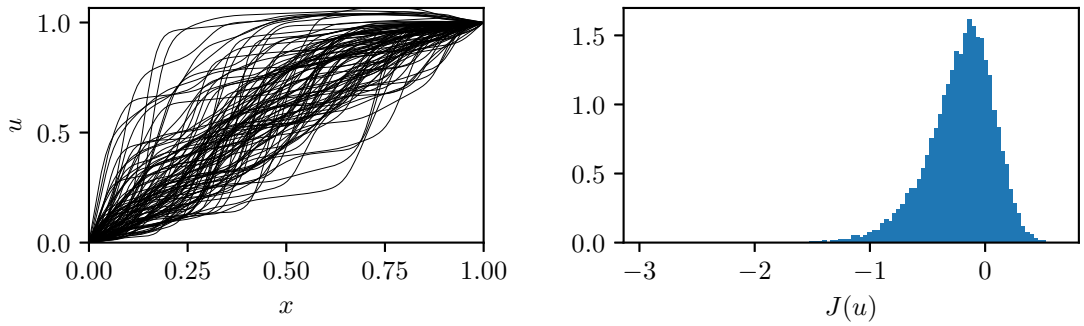
$$-\frac{d}{dx}\left(q(x,\omega)\frac{d}{dx}u(x)\right) = 1, \qquad \text{in } (0,1)$$
$$u(0) = 0, \quad u(1) = 1.$$

$$q(x,\omega) = e^{\theta(x,\omega)},$$

where θ is a Gaussian random field on $(0, 1)$ with Gaussian covariance, specifically

$$v_\theta(x, y) = e^{-\frac{|x-y|^2}{2l^2}},$$

with correlation length $l = 0.05$.



(a) Sample paths of solution.

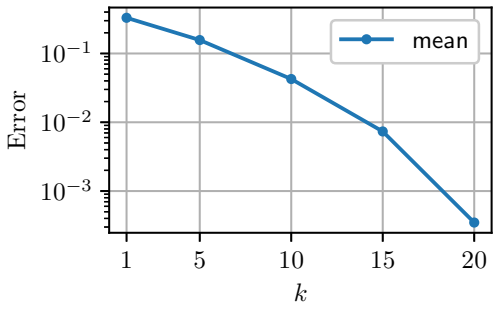
(b) Distribution of quantity of interest.

Figure 6.5: Statistical properties of the statistic computed via Monte Carlo sample of size 10^6.

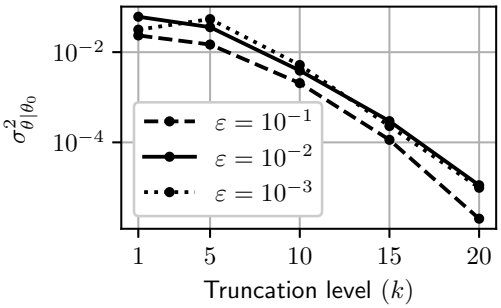
The Statistical properties of the J is computed via Monte Carlo sample of size 10^6 in figure 6.5. Where, $\mathbb{E}[J] \approx -2.19206 \times 10^{-1}$ and variance $\hat{\sigma}_J^2 = 8.9817 \times 10^{-2}$. The root mean square error is thus 2.9970×10^{-4}.

Figure 6.6 shows the effect of truncation on both the Monte Carlo estimates of $\mathbb{E}[J]$ and on the expected conditional variance of J, given the truncated field θ_0. The error estimates in Figure 6.6a were obtained by comparing Monte Carlo sample averages of J using truncated KL expansions for θ with that of the full expansion. The sample size used for each average was 10^6. The expected conditional variances $\mathbb{E}_\theta[\mathbb{V}_{\theta|\theta_0}[J]]$ in Figure 6.6b were computed using sparse grid approximations in the

truncated component θ_0 whose quadrature size was determined to be within ε units of the appropriate truncated Monte Carlo estimates of size 10^6. At each quadrature point θ_0^i, the conditional variance $\mathbb{V}_{\theta|\theta_0}[J|\theta_0^i]$ was computed by means of a small sample of size 100. Note that both the truncation error and the conditional variances exhibit similar decay rate as k increases.



(a) Estimated truncation error for $\mathbb{E}[J]$ using Monte Carlo samples of size 10^6.



(b) Estimated expected conditional variances of J, given θ_0. ε is the estimated accuracy.

Figure 6.6: Conditional dependence

We investigate the accuracy of the sparse grid quadrature rule in Figure 6.7, compared with a 10^6 sample Monte Carlo estimate. Note that, due to sampling error, the error estimates are reliable only above the level 3×10^{-4}. As predicted earlier, the convergence rate decays with an increase in k, requiring a greater number of quadrature nodes for the same tolerance level ε for more complex parameter approximations.

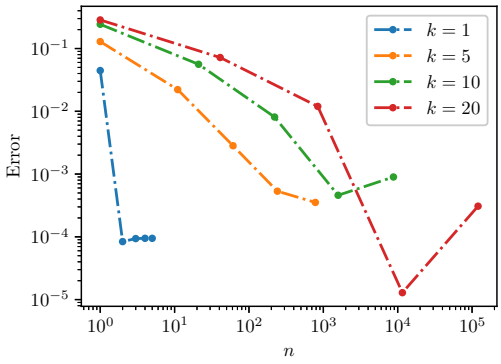


Figure 6.7: Log-log plot of the sparse grid quadrature error for various truncation levels.

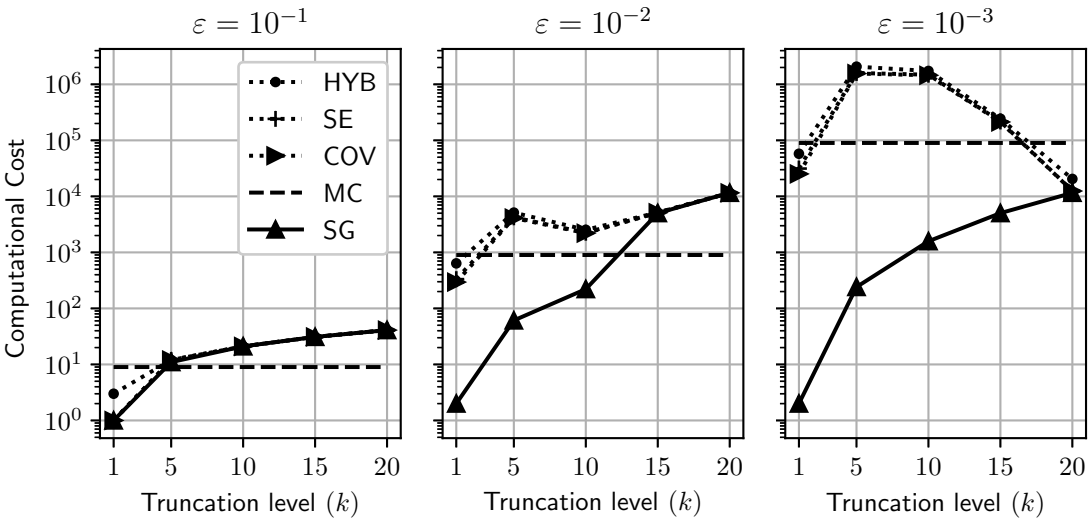


Figure 6.8: Comparison of ε-costs of Monte Carlo, Sparse grids, and Hybrid sampling.

In figure 6.8 we compared the ε-costs Monte Carlo (MC), Sparse grids (SG), Hybrid sampling (HYB), Sensitivity derivative enhanced monte carlo (SE) and the Control variate enhanced Monte Carlo method (COV) for tolerance $\varepsilon = 10^{-1}$, $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-3}$. From the graph it seems that combining Mc and sparse grid

even if just using sparse grid on the first term or dominant term in the KL expansion and the rest Mc does a better job than Mc.So the Hybrid sampling provides an improvement of the Monte Carlo Method.

Chapter 7

Conclusion and Future Work

We have developed the Hybrid Sampling Method which is an efficient sampling method for estimating a statistical quantity of interest depending on a random field in a partial differential method $\mathbb{E}[J(\theta)]$. This method uses conditional sampling to combine sparse grid quadrature rules on a low-dimensional projection of the parameter space with a Monte Carlo scheme to compensate in the remaining dimensions. Conditioning on coarse scale field approximations allows for the use of efficient sparse grid methods in lower dimensions and reduces the conditional variance of the conditioned quantity of interest.We analyzed the error of our method and performed a complexity analysis where we provided conditions in which the Hybrid Sampling is more efficient than either of it's constituent. We extended our algorithm to incorporate sensitivity-based variance reduction techniques such as the sensitivity derivative enhance Monte Carlo method and the control variate sensitivity enhanced Monte Carlo method to further improve the efficiency of our algorithm. We then demonstrated that our algorithm works by using two numerical examples.

The conditional random field $\theta|\theta_0$ typically becomes less smooth and less correlated as k increases (requires a finer spatial grid). Perhaps upscaling methods would be a more appropriate way to "enrich" the sparse grid sample. It is also plausible that some quantities of interest don't require the random parameter to have the same

high resolution throughout the domain. So, given a low resolution sample, can we adaptively choose a refinement of the field in regions that determine the variance of J? We would also like to consider more complex physical problems.

Bibliography

[1] Pde-constrained optimization and the adjoint method, 2012.

[2] Ivo Babuška, Fabio Nobile, and Raúl Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Review*, 52(2):317–355, 2010.

[3] Yanzhao Cao, M.Y. Hussaini, T. Zang, and A. Zatezalo. A variance reduction method based on sensitivity derivatives. *Applied Numerical Mathematics*, 56(6):800–813, June 2006.

[4] Peng Chen. Sparse quadrature for high-dimensional integration with gaussian measure. *AMS*, 2017.

[5] Moo K. Chung. Introduction to random fields. 2020.

[6] R. Courant and D. Hilbert. *Methods of Mathematical Physics*. Wiley, April 1989.

[7] Scott L Goodrick Edwin Jimenez, M. Yousuff Hussaini. Uncertainty quantification in rothermelś model using an efficient sampling method. *USDA Forest Service Proceedings.*, 20007.

[8] Peter Fratzl and Richard Weinkamer. Nature's hierarchical materials. *Progress in Materials Science*, 52(8):1263?1334, 2007.

[9] Philipp Frauenfelder, Christoph Schwab, and Radu Alexandru Todor. Finite elements for elliptic problems with stochastic coefficients. *Computer Methods in Applied Mechanics and Engineering*, 194(2-5):205–228, February 2005.

[10] F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, April 1980.

[11] Roger Ghanem, David Higdon, and Houman Owhadi. *Introduction to Uncertainty Quantification*, pages 1–4. Springer International Publishing, Cham, 2016.

[12] Ivan G Graham, Frances Y Kuo, Dirk Nuyens, Robert Scheichl, and Ian H Sloan. Quasi-monte carlo methods for elliptic pdes with random coefficients and applications. *Journal of Computational Physics*, 230(10):3668–3694, 2011.

[13] Glyn Holton. *Value-at-risk : theory and practice*. Academic Press, Amsterdam Boston, 2003.

[14] Kari Karhunen. *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, volume 37. Universitat Helsinki, 1947.

[15] Frances Y Kuo, Christoph Schwab, and Ian H Sloan. Quasi-monte carlo finite element methods for a class of elliptic partial differential equations with random coefficients. *SIAM Journal on Numerical Analysis*, 50(6):3351–3374, 2012.

[16] Michel Loeve. *Probability Theory, Vol. II*, volume 46. 1978.

[17] Paula A. Whitlock Malvin H. Kalos. Monte carlo methods. *Wiley & Sons Inc*, 1986.

[18] Hermann G. Matthies and Andreas Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 194(12-16):1295–1331, April 2005.

[19] James Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 83(559):69–70, nov 1909.

[20] Farshid Mirzaee and Nasrin Samadyar. Application of hat basis functions for solving two-dimensional stochastic fractional integral equations. *Computational and Applied Mathematics*, 37(4):4899–4916, March 2018.

[21] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, January 2008.

[22] Erich Novak and Klaus Ritter. The curse of dimension and a universal method for numerical integration. In *Multivariate Approximation and Splines*, pages 177–187. Birkhäuser Basel, 1997.

[23] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[24] Jonathan E. Taylor Robert J. Adler. Random fields and geometry. *Springer New York*, 2007.

[25] Haavard Rue and Leonhard Held. *Gaussian Markov random fields*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2005. Theory and applications.

[26] Christoph Schwab and Radu Alexandru Todor. Karhunen–loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122, September 2006.

[27] Iason Papaioannoua Sebastian Geyera. The spatial averaging method for non-homogeneous random felds with application to reliability analysis. 2022.

[28] Ian H Sloan and Henryk Woźniakowski. When are quasi-monte carlo algorithms efficient for high dimensional integrals? *Journal of Complexity*, 14(1):1–33, 1998.

[29] T.J. Sullivan. *Introduction to Uncertainty Quantification*. Springer International Publishing, 2015.

[30] Roberto Szechtman. Control variates techniques for monte carlo simulation. *Proceedings of the 2003 Winter Simulation Conference*, 2003.

[31] Felipe Uribe, Iason Papaioannou, Wolfgang Betz, and Daniel Straub. Bayesian inference of random fields represented with the karhunen–loève expansion. *Computer Methods in Applied Mechanics and Engineering*, 358:112632, 2020.

[32] Julian Valentin. B-splines for sparse grids:algorithms and application tohigher-dimensional optimization. *University of Stuttgart Germany*, 2019.

[33] E. Vanmarcke. Random fields: analysis and synthesis, revised and expanded new edition. *World Scientific*, 2010.

[34] David Williams. Probability with martingales. *Cambridge University Press*, 1991.

[35] Jiaxin Zhang. Modern monte carlo methods for efficient uncertainty quantification and propagation: A survey, 2020.