RICE UNIVERSITY

**Discontinuous Galerkin Methods for Elliptic
Partial Differential Equations with Random
Coefficients**

by

Kun Liu

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Master of Arts

APPROVED, THESIS COMMITTEE:

Béatrice M. Rivière, Chair
Associate Professor of Computational and
Applied Mathematics

William W. Symes
Noah G. Harding Professor of
Computational and Applied Mathematics

Matthias Heinkenschloss
Professor of Computational and Applied
Mathematics

Houston, Texas

May, 2010

ABSTRACT

Discontinuous Galerkin Methods for Elliptic Partial Differential Equations with Random Coefficients

by

Kun Liu

This thesis proposes and analyses two numerical methods for solving elliptic partial differential equations with random coefficients. The stochastic problem is first transformed into a parametrized one by the use of the Karhunen–Loève expansion. This new problem is then discretized by the discontinuous Galerkin (DG) method. A priori error estimate in the energy norm for the stochastic discontinuous Galerkin solution is derived. In addition, the expected value of the numerical error is theoretically bounded in the energy norm and the L^2 norm. In the second approach, the Monte Carlo method is used to generate independent identically distributed realizations of the stochastic coefficients. The resulting deterministic problems are solved by the DG method. Next, estimates are obtained for the error between the average of these approximate solutions and the expected value of the exact solution. The Monte Carlo discontinuous Galerkin method is tested numerically on several examples. Results show that the nonsymmetric DG method is stable independently of meshes and the value of penalty parameter. Symmetric and incomplete DG methods are stable only when the penalty parameter is large enough. Finally, comparisons with the Monte Carlo finite element method and the Monte Carlo discontinuous Galerkin method are presented for several cases.

Acknowledgements

In the first place, I heartily thank God for offering me the talent, giving me the wisdom and blessing me all the time. I would love to record my gratitude to my families for supporting me when I go abroad for further study in U.S.. Many thanks go in particular to my advisor Dr. Rivière, who inspires me and encourages me during these two years. Her invaluable and precious suggestion and advice enable me to understand the subject more clearly and improve my skills on research tremendously. I also benefit by the advice and guidance from Dr. Symes and Dr. Heinkenschloss. Their rich experience and constructive comments improve my skills on presentation sufficiently. Their attention and patience are highly appreciated.

Besides, I gratefully thank Lily Lam, who teaches me lots of basic life principles which strengthen me when I encounter the setback and difficulties in my life. I would like to express my appreciation to Xin Wang and Drew Kouri. Thank them for taking time to discuss with me about my research.

Contents

Abstract	ii
Acknowledgements	iii
1 Introduction	1
1.1 Discontinuous Galerkin Methods	2
1.2 Literature Review	2
1.3 Thesis Outline	5
2 Stochastic Partial Differential Equations	6
2.1 Introduction	6
2.2 Function Spaces	6
2.2.1 Banach Space	6
2.2.2 Hilbert Space	7
2.2.3 Measure Theory	9
2.2.4 Sobolev Space	11
2.2.5 Stochastic Sobolev Space	13
2.3 Important Inequalities	16
2.3.1 Trace Inequalities	16
2.3.2 Cauchy-Schwarz's and Young's Inequalities	17
2.4 Model and Problem	18
2.4.1 Probability Theory	18
2.4.2 Model	20
2.5 Existence and Uniqueness for the Solution	20

2.6	Karhunen–Loève Expansion and Finite Dimensional Noise	22
2.6.1	Karhunen–Loève Expansion	22
2.6.2	Finite Dimensional Noise	24
3	Stochastic Finite Element Method	27
3.1	Introduction	27
3.2	Finite Element Spaces and Approximation Results	27
3.3	$k \times h$ –version Stochastic Finite Element Method	29
4	Stochastic Discontinuous Galerkin Method	32
4.1	Introduction	32
4.2	Primal Discontinuous Galerkin Methods	32
4.2.1	Model Problem	32
4.2.2	Discontinuous Galerkin Space	33
4.2.3	Variational Formulation	35
4.2.4	Existence and Uniqueness of DG solution	36
4.2.5	Error Estimates	36
4.3	Stochastic Discontinuous Galerkin Formulation	37
4.4	Error Estimate in the Energy Norm	39
4.5	Error Estimate in the L^2 Norm	44
4.6	$k \times h$ –version SFEM versus $k \times h$ –version SDG	48
5	The Monte Carlo Discontinuous Galerkin Method	51
5.1	Introduction	51
5.2	Statistical Error	52
5.3	Space Discretization Error	54
6	Numerical Examples	59
6.1	Introduction	59

	vi
6.2 K-L Expansion of Standard Wiener Process	59
6.3 Monte Carlo Discontinuous Galerkin Method	63
6.3.1 Nonsymmetric Interior Penalty Galerkin Method	63
6.3.2 Symmetric Interior Penalty Galerkin Method	68
6.3.3 Incomplete Interior Penalty Galerkin Method	71
6.4 Discontinuous Diffusion and Load Functions	74
6.5 Comparison of Monte Carlo Finite Element Method and Monte Carlo Discontinuous Galerkin Method	75
6.5.1 Case 1: Continuous Coefficients	75
6.5.2 Case 2: Discontinuous Coefficients	79
7 Conclusions	82
Bibliography	85

Chapter 1

Introduction

This thesis establishes the stochastic discontinuous Galerkin (SDG) method and the Monte Carlo discontinuous Galerkin (MCDG) method for elliptic partial differential equations in which both of the diffusion and the load functions are stochastic processes.

Modelling uncertainty is a promising research area which has rapidly developed in the recent decades. Stochastic partial differential equations (SPDEs) inevitably have become one of the most powerful tools to model stochastic problems arising from the practical world. For example, Dawson-Watanabe processes and Fleming-Voit processes have been widely used in genetics and the stochastic Navier-Stokes equations plays an important role in fluid mechanics [1]. However, finding the joint probability distribution of the solution to SPDEs can be challenging. In practice, we compute the expected value of the solution.

In my thesis, I make the finite dimensional noise assumption. Truncated Karhunen-Loève (KL) expansion is used to transform the SPDEs into a parametrized problem. Compared with other methods, such as spectral representation which is used to expand a stochastic process in a sum of trigonometric functions with random phase angles and amplitudes, K-L expansion is less computationally expensive [2].

1.1 Discontinuous Galerkin Methods

The discontinuous Galerkin (DG) method is a class of variational methods for solving partial differential equations by approximating the solution by discontinuous piecewise polynomials. In 1973, the first discontinuous Galerkin method for hyperbolic equations was established (Reed and Hill [3]). In the late 1970s, the symmetric interior penalty Galerkin (SIPG) method was introduced by Wheeler [4] and Arnold [5]. In 1999, the nonsymmetric interior penalty Galerkin (NIPG) method was introduced by Rivière, Wheeler, and Girault [6]. Finally, in 2004, the incomplete interior penalty Galerkin (IIPG) method was introduced by Dawson, Sun and Wheeler [7]. Since the late nineties, the DG method has been widely used to solve deterministic PDEs because of its appealing properties: stable, locally mass conservative, effectively deals with complicated geometries and irregular meshes with hanging nodes and handles discontinuous coefficients. In my thesis, I discretize the parametrized SPDEs by the SIPG, NIPG and IIPG methods.

One motivation for using DG discretization is that for deterministic elliptic problems with discontinuous coefficients, the DG solution is more accurate on a fixed mesh than the classical finite element solution [8]. Another motivation is the great flexibility of DG methods. For flow and transport problems arising in porous media, the DG method is well suited as it easily allows for adaptive meshes and it is locally mass conservative.

1.2 Literature Review

The problem that this thesis addresses is a linear elliptic partial differential equation with random coefficients. There are many numerical methods to approximate the

solution to this problem, which depend on the structure of the noise in the stochastic domain. For example, Taylor expansion up to second order can be used (Kleiber and Hien [9]). In the paper [10], Babuška, Tempone and Zouraris formulate $k \times h$ -version and $p \times h$ -version stochastic finite element method and give the H^1 and L^2 error estimates for the expected value of the solution. They point out that the $p \times h$ -version method yields an exponential rate of convergence rate with respect to the degree of the polynomials used in the Galerkin approximation. Besides, they establish the Monte Carlo finite element method and conclude that the stochastic error tends to zero a.s. when the number of the realization increases and the spatial discretization error is of second order with respect to the mesh size. In my thesis, I formulate the $k \times h$ -version stochastic DG method and the Monte Carlo DG method and discuss their convergence properties.

Furthermore, when the dimension of the outcome set is large, Babuška, Tempone and Zouraris [11] point out that the Monte Carlo finite element method is the most effective when the K-L expansion of the diffusion coefficient needs many terms, regardless of exponential rate of $p \times h$ -version stochastic finite element method. In my thesis, I implement the Monte Carlo DG method. However, if we have a small number of random variables, i.e., the dimension of the outcome set is small, such as 20, Babuška, Nobile and Tempone [12] conclude that the stochastic collocation method offers many appealing advantages. It leads to completely uncoupled deterministic problems and treats this problem efficiently when the stochastic coefficients depend non-linearly on the random variables. In [13], Ganis, Yotov and Zhong present an efficient multi-scale stochastic framework for uncertainty quantification in porous media model with multiple rock types. They use the stochastic collocation method to approximate the solution to this problem which are based on Darcy's law with

nonstationary stochastic permeability which is represented by K–L expansion. I will consider the stochastic collocation method in the future.

If we involve the white noise in our problem, it becomes much more complicated. However, if the coefficients are deterministic, Allen, Novosel and Zhang [14] study finite difference and finite element schemes on a one-dimensional spatial domain. Additionally, they propose the finite element method for both parabolic and elliptic stochastic partial differential equations. Recently, Yao and Bo [15] study two and three dimensional cases by using discontinuous Galerkin method. They present the local discontinuous Galerkin method and prove that the convergence rates of the L^2 errors in the two and three dimensional space are lower than the convergence rates in the one dimensional case. However, if the coefficients are stochastic processes, the problem is challenging. I hope to do further research on this case.

As I mention, this thesis implements Monte Carlo method [16] to numerically approximate the solutions to the elliptic partial differential equations with stochastic coefficients. The Monte Carlo method can be coded concisely. It generates independent identically distributed approximations of the solution by sampling the diffusion function and the load function. In my thesis, DG methods are applied to discretize the spatial domain of the partial differential equations. Although the convergence rate of Monte Carlo method is considered slow, in my thesis its computational work is only determined by the number of the simulation, the degree of the polynomial and the mesh size. In [17, 18], we find that Quasi-Monte Carlo methods have a better convergence rate than Monte Carlo method if the number of the random variables is small.

1.3 Thesis Outline

In chapter 2, I introduce the function spaces and the parametrized SPDE problem which is discussed in my thesis. In Chapter 3, I recall a priori error estimate in the energy norm by Babuška, Tempone and Zouraris [10] and give a detailed proof. In Chapter 4, I formulate the stochastic discontinuous Galerkin method and derive two a priori error estimates of the expected value in the energy norm and L^2 norm for this method. In Chapter 5, I construct the Monte Carlo discontinuous Galerkin method and develop a priori error estimates for the expected value of the solution. In Chapter 6, I give numerical examples of using truncated Karhunen–Loève expansion to approximate the standard Wiener process and I use Monte Carlo discontinuous Galerkin method to simulate the expected value and variance of the NIPG, SIPG and IIPG solution to the parametrized problem introduced in Chapter 2.

Chapter 2

Stochastic Partial Differential Equations

2.1 Introduction

In this chapter, I first introduce the definitions of the function spaces in which the problem in my thesis are discussed. The isomorphism between some of these spaces is particularly mentioned in this part. Then, I introduce the stochastic model which this thesis addresses. It is a second order linear elliptic problem with homogeneous Dirichlet boundary conditions and random coefficients. The existence and uniqueness of the solution to this problem is studied. In the last section of this chapter, I introduce the Karhunen–Loève expansion, which is used to transform a stochastic problem into a parametrized problem. Then, I consider the case of finite dimensional noise, which allows the use of finite element methods, such as continuous Galerkin method and discontinuous Galerkin method, which are discussed in Chapter 3 and Chapter 4.

2.2 Function Spaces

In this section, I introduce the function spaces in which the problems in my thesis are addressed. Additionally, isomorphism between these spaces is used when establishing the definition of the stochastic Sobolev Space.

2.2.1 Banach Space

Let X denote a real linear space.

Definition 2.1 A mapping $\| \cdot \| : X \rightarrow [0, \infty)$ is called a *norm* if

- $\|u + v\| \leq \|u\| + \|v\| \quad \forall u, v \in X.$
- $\|\lambda u\| = |\lambda| \|u\| \quad \forall u \in X, \lambda \in \mathbb{R}.$
- $\|u\| = 0$ if and only if $u = 0.$

In the following, I will assume X is a normed linear space.

Definition 2.2 A sequence $\{u_k\}_{k=1}^{\infty} \subset X$ *converges* to $u \in X$, which is written as $u_k \rightarrow u$, if

$$\lim_{k \rightarrow \infty} \|u_k - u\| = 0.$$

Definition 2.3 (Cauchy Sequence) A sequence $\{u_k\}_{k=1}^{\infty} \subset X$ is called a *Cauchy sequence* if for any $\epsilon > 0$ there exists $N > 0$ such that

$$\|u_m - u_n\| < \epsilon \quad \forall m, n \geq N.$$

The space X is *complete* if each Cauchy sequence in X converges; that is, whenever $\{u_k\}_{k=1}^{\infty}$ is a Cauchy sequence, there exists $u \in X$ such that $\{u_k\}_{k=1}^{\infty}$ converges to u .

Definition 2.4 (Banach Space) A *Banach Space* is a complete, normed linear space.

2.2.2 Hilbert Space

Let H be a real linear space.

Definition 2.5 (Inner Product) A mapping $(\cdot, \cdot) : H \times H \rightarrow \mathbb{R}$ is called an *inner product* if

- $(u, v) = (v, u) \quad \forall u, v \in H.$

- the mapping $u \mapsto (u, v)$ is linear for each $v \in H$.
- $(u, u) \geq 0 \quad \forall u \in H$.
- $(u, u) = 0$ if and only if $u = 0$.

Remark: If $(\cdot, \cdot)$ is an inner product, the associated *norm* is

$$\|u\| = \sqrt{(u, u)}, \quad \forall u \in H.$$

Definition 2.6 (Hilbert Space) A *Hilbert Space* is a Banach space endowed with an inner product and associated norm.

Definition 2.7 A metric space is called *separable* if it contains a countable dense subset; that is, there exists a sequence of elements of this space such that every nonempty open subset of the space contains at least one element of the sequence.

Definition 2.8 (Orthogonal) Let H be a separable Hilbert Space.

- Two elements $u, v \in H$ are *orthogonal* if $(u, v) = 0$.
- We say a countable basis $\{\phi_k\}_{k=1}^{\infty} \subset H$ is *orthogonal* if

$$\begin{cases} (\phi_m, \phi_n) = 0, & \forall m, n \geq 1, m \neq n \\ \|\phi_m\| = 1, & \forall m \geq 1. \end{cases}$$

Remark: If $u \in H$ and $\{\phi_k\}_{k=1}^{\infty} \subset H$ is an orthogonal basis, we can write

$$u = \sum_{k=1}^{\infty} (u, \phi_k) \phi_k,$$

and the series converges in H . In addition

$$\|u\|^2 = \sum_{k=1}^{\infty} (u, \phi_k)^2.$$

2.2.3 Measure Theory

Definition 2.9 Let $\mathcal{F}$ be a collection of subsets of a set Ω . Then $\mathcal{F}$ is called a σ -field if and only if it satisfies the following properties:

- The empty set $\emptyset$ belongs to $\mathcal{F}$;
- If A belongs to $\mathcal{F}$, then the complement A^c belongs to $\mathcal{F}$;
- If $A_1, A_2, \dots$, is a sequence of elements of $\mathcal{F}$, then their union $\bigcup_{i=1}^{\infty} A_i$ belongs to $\mathcal{F}$.

Definition 2.10 If $\mathcal{C}$ is any collection of subsets of Ω , then the σ -field generated by $\mathcal{C}$, denoted $\sigma(\mathcal{C})$, is the smallest σ -field containing $\mathcal{C}$. (Here, “smallest” means that if $\mathcal{F}$ is any σ -field containing $\mathcal{C}$, then $\sigma(\mathcal{C}) \subset \mathcal{F}$.)

A pair $(\Omega, \mathcal{F})$ consisting of a set Ω and a σ -field $\mathcal{F}$ is called a measurable space. The elements of $\mathcal{F}$ are called measurable sets or events.

We now define the Borel σ -field on the real line $\mathbb{R}$.

Definition 2.11 The Borel σ -field $\mathcal{B}$ on $\mathbb{R}$ is the σ -field generated by the collection of all finite open intervals. We can write

$$\mathcal{B} = \sigma(\{(a, b) : -\infty < a < b < \infty\}).$$

The elements of $\mathcal{B}$ are called Borel sets.

Definition 2.12 A measure space $(\Omega, \mathcal{F}, \mu)$ is a triple, consisting of an underlying set Ω , a σ -field $\mathcal{F}$, a function μ called the measure with domain $\mathcal{F}$ and satisfying the following:

- $0 \leq \mu(A) \leq \infty$ for all $A \in \mathcal{F}$;

- $\mu(\emptyset) = 0$;
- If $A_1, A_2, \dots$ is a sequence of disjoint elements of $\mathcal{F}$ (i.e. $A_i \cap A_j = \emptyset$ for all $i \neq j$), then

$$\mu \left(\bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu(A_i).$$

A *probability space* is a measure space $(\Omega, \mathcal{F}, P)$ for which $P(\Omega) = 1$. It is called a *complete* probability space if $\mathcal{F}$ contains all subsets G of Ω with P -outer measure zero, i.e. with

$$P^*(G) := \inf\{P(F) \mid F \in \mathcal{F}, G \subset F\} = 0.$$

Now, I introduce an important theorem about Lebesgue measure.

Theorem 2.1 (Existence of Lebesgue measure and Lebesgue measurable sets)

There exists a σ -field $\mathcal{M}$ of subsets of $\mathbb{R}^N$ and a mapping:

$$|\cdot| : \mathcal{M} \rightarrow [0, +\infty]$$

with the following properties:

- Every open subset of $\mathbb{R}^N$, and thus every closed subset of $\mathbb{R}^N$, belong to $\mathcal{M}$.
- If B is any ball in $\mathbb{R}^N$, then $|B|$ equals the N -dimensional volume of B .
- If $\{A_k\}_{k=1}^{\infty} \subset \mathcal{M}$ and the sets $\{A_k\}_{k=1}^{\infty}$ are pairwise disjoint, then

$$\left| \bigcup_{k=1}^{\infty} A_k \right| = \sum_{k=1}^{\infty} |A_k|.$$

- If $A \subseteq B$, where $B \in \mathcal{M}$ and $|B| = 0$, then $A \in \mathcal{M}$ and $|A| = 0$.

Then, the sets in $\mathcal{M}$ are called *Lebesgue measurable sets* and $|\cdot|$ is the N -dimensional *Lebesgue measure*.

Remark: If some property holds everywhere on $\mathbb{R}^N$, except on a measurable set with

Lebesgue measure zero, we say the property holds *almost everywhere*, abbreviated “a.e.”.

Definition 2.13 Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be measurable spaces and $f : \Omega \rightarrow \Lambda$ a function. Then f is a *measurable function* iff $f^{-1}(\mathcal{G}) \subset \mathcal{F}$, in which case we shall write $f : (\Omega, \mathcal{F}) \rightarrow (\Lambda, \mathcal{G})$. If $\Lambda = \mathbb{R}$ and $\mathcal{G}$ is the Borel σ -field, then we say f is Borel measurable or a (real valued) Borel function.

In probability theory, a measurable function is called a random object or random element and usually denoted by $X, Y, \dots$, and a real valued Borel function is called a *random variable*.

2.2.4 Sobolev Space

Definition 2.14 A measurable function $f : \Omega \rightarrow \mathcal{R}$, where $(\Omega, \mathcal{F}, \mu)$ is a measure space, is said to be *summable* if the Lebesgue integral of the absolute value of f exists and is finite,

$$\int_{\Omega} |f| dx < \infty.$$

Definition 2.15 Let $(\Omega, \mathcal{F}, \mu)$ be measure space and $f : \Omega \rightarrow \mathbb{R}$ be measurable. For $1 \leq p < \infty$, let

$$\|f\|_{L^p(\Omega)} = \left(\int_{\Omega} |f|^p dx \right)^{\frac{1}{p}} \quad 1 \leq p < \infty.$$

and for the case $p = \infty$, set

$$\|f\|_{L^\infty(\Omega)} = \text{ess sup}\{|f(x)|_{L^p(\Omega)} : x \in \Omega\}.$$

where $\text{ess sup } f = \inf\{a \in \mathbb{R} : \mu(\{x : f(x) > a\}) = 0\}$.

In either case, we define the *Lebesgue Space*

$$L^p(\Omega) = \{f : \Omega \rightarrow \mathbb{R} \mid \|f\|_{L^p(\Omega)} < \infty\}.$$

Definition 2.16 For a domain Ω , the set of *locally integrable* functions is denoted by

$$L^p_{loc}(\Omega) = \{f : f \in L^p(U) \quad \forall \text{ compact } U \subset \Omega\}$$

We define:

$$C^\infty(\Omega) = \{u : \Omega \rightarrow \mathbb{R} \mid u \text{ is infinitely differentiable}\}. \quad (2.1)$$

$$C_c^\infty(\Omega) = \{u \in C^\infty(\Omega) \text{ with compact support}\}. \quad (2.2)$$

Definition 2.17 (Weak Derivative) Suppose $u, v \in L^1_{loc}(\Omega)$, and α is a multi-index.

We say that v is the α^{th} -weak partial derivative of u , written

$$D^\alpha u = v,$$

provided

$$\int_{\Omega} u D^\alpha \phi dx = (-1)^{|\alpha|} \int_{\Omega} v \phi dx, \quad \forall \phi \in C_c^\infty(\Omega).$$

where $|\alpha| = \alpha_1 + \dots + \alpha_n$ for $\alpha = (\alpha_1, \dots, \alpha_n)$.

Definition 2.18 (Sobolev Space) The *Sobolev Space* $W^{k,p}(\Omega)$ consists of all locally summable function $u : \Omega \rightarrow \mathbb{R}$ such that for each multi-index α with $|\alpha| \leq k$, $D^\alpha u$ exists in the weak sense and belongs to $L^p(\Omega)$.

Definition 2.19 (Sobolev Norm) If $u \in W^{k,p}(\Omega)$, we define its norm to be

$$\|u\|_{W^{k,p}(\Omega)} = \begin{cases} \left(\sum_{|\alpha| \leq k} \int_{\Omega} |D^\alpha u|^p dx \right)^{\frac{1}{p}} & 1 \leq p < \infty, \\ \sum_{|\alpha| \leq k} \operatorname{ess\,sup}_{\Omega} |D^\alpha u| & p = \infty. \end{cases}$$

Remark: if $p = 2$, we usually write

$$H^k(\Omega) = W^{k,2}(\Omega), \quad k = 0, 1, \dots$$

Note that the space $W^{k,2}(\Omega)$ is a Hilbert space, with inner product $(\cdot, \cdot)_{W^{k,2}(\Omega)}$ defined by

$$(u, v)_{W^{k,2}(\Omega)} = \sum_{|\alpha| \leq k} (D^\alpha u, D^\alpha v)_{L^2(\Omega)}.$$

We also define

$$H_0^1(\Omega) = \{u \in H^1(\Omega), u = 0 \text{ on } \partial\Omega\}.$$

equipped with the norm

$$\|u\|_{H_0^1(\Omega)} = \left(\int_{\Omega} |\nabla u|^2 dx \right)^{\frac{1}{2}}.$$

2.2.5 Stochastic Sobolev Space

In this part, I recall the definition of stochastic Sobolev space as given in [11]. I introduce some basic concepts which are the foundation for building the stochastic Sobolev space.

Definition 2.20 (Indicator Function) The *indicator function* of a subset A of a set X is a function

$$I_A : X \rightarrow \{0, 1\}$$

defined as

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Definition 2.21 (Simple Function) Let $(\Omega, \mathcal{F})$ be measurable space. Let $A_1, \dots, A_n$ be a sequence of measurable sets in $\mathcal{F}$, and let $a_1, \dots, a_n$ be a sequence of real or complex numbers. A *simple function* is a function of the form

$$f(x) = \sum_{k=1}^n a_k I_{A_k}(x),$$

where I_{A_k} is an indicator function.

Now, I give an important definition in the process of establishing the stochastic Sobolev space.

Definition 2.22 (Strongly Measurable) A function $f : B \rightarrow X$ is strongly measurable if there exists a sequence of simple functions $I_k : B \rightarrow X$ such that

$$I_k(x) \rightarrow f(x) \quad \text{for a.e. } x \in B.$$

Definition 2.23 A measure space $(\Omega, \mathcal{F}, \mu)$ is complete if $A \subset B, B \in \mathcal{F}$ and $\mu(B) = 0$ together imply that $A \in \mathcal{F}$ (and hence that $P(A) = 0$).

Definition 2.24 (Completion of a Measure Space) Given a measure space $(\Omega, \mathcal{F}, \mu)$, there is an extension $(\Omega, \mathcal{F}_0, \mu_0)$ of this measure space that is complete. The smallest such extension (i.e. the smallest σ -field $\mathcal{F}_0$) is called the completion of the measure space.

Then, we define the tensor product space of two Hilbert spaces.

Definition 2.25 Let H_1, H_2 be two Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_{H_1}$ and $\langle \cdot, \cdot \rangle_{H_2}$, respectively. The *tensor product space* $H_1 \otimes H_2$ is the completion of the inner product space by defining

$$\langle u_1 \otimes u_2, v_1 \otimes v_2 \rangle = \langle u_1, v_1 \rangle_{H_1} \langle u_2, v_2 \rangle_{H_2} \quad \forall u_1, v_1 \in H_1 \text{ and } u_2, v_2 \in H_2.$$

Remark: Note that $L^2(\Omega_1)$ and $H^k(\Omega_2)$ are Hilbert spaces, where $\Omega_i \subset \mathbb{R}^N, i = 1, 2$. Consider the tensor product space $L^2(\Omega_1) \otimes H^k(\Omega_2)$. Its inner product is defined as

$$\langle u, \hat{u} \rangle_{L^2(\Omega_1) \otimes H^k(\Omega_2)} = \int_{\Omega_1} \langle u(y, \cdot), \hat{u}(y, \cdot) \rangle_{H^k(\Omega_2)} dy.$$

Therefore, if $u \in L^2(\Omega_1) \otimes H^1(\Omega_2)$, then $u(y, \cdot) \in H^k(\Omega_2)$ a.e. on Ω_1 and $u(\cdot, x) \in L^2(\Omega_1)$ a.e. on Ω_2 .

In order to find the relationship between tensor product spaces, we need the definition of Hilbert space isomorphism.

Definition 2.26 (Isomorphism) An isomorphism of the Hilbert space H_1 onto the Hilbert space H_2 is a one to one mapping T of H_1 onto H_2 such that for all $f_1, f_2 \in H_1$,

- $T(af_1 + bf_2) = aTf_1 + bTf_2$ for all scalars a and b .
- $(Tf_1, Tf_2)_{H_2} = (f_1, f_2)_{H_1}$.

Remark: If there is an isomorphism of the spaces H_1 and H_2 , we denote by

$$H_1 \simeq H_2.$$

Note that if $H_1 \simeq H_2$ and $H_1 \simeq H_3$, then $H_2 \simeq H_3$.

Recall that $\Omega_i \subset \mathbb{R}^N, i = 1, 2$. We define

$$L^2(\Omega_1; H^k(\Omega_2)) = \left\{ v : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R} \mid v \text{ is strongly measurable and } \int_{\Omega_1} \|v(y, \cdot)\|_{H^k(\Omega_2)}^2 dy < \infty \right\},$$

and

$$H^k(\Omega_2; L^2(\Omega_1)) = \left\{ v : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R} \mid v \text{ is strongly measurable and } \forall \alpha \in \mathbb{N}^k, \text{ with } |\alpha| \leq k, \right. \\ \left. \text{there exists } \partial_\alpha v \in L^2(\Omega_2) \otimes L^2(\Omega_1) \text{ such that } \forall \phi \in C_0^\infty(\Omega_1 \times \Omega_2) : \right. \\ \left. \int_{\Omega_1} (\partial_\alpha v(y, \cdot), \phi(y, \cdot))_{L^2(\Omega_2)} dy = (-1)^{|\alpha|} \int_{\Omega_1} (v(y, \cdot), \partial_\alpha \phi(y, \cdot))_{L^2(\Omega_2)} dy. \right\}$$

Since $L^2(\Omega_1) \otimes H^k(\Omega_2) \simeq L^2(\Omega_1; H^k(\Omega_2))$, $H^k(\Omega_2) \otimes L^2(\Omega_1) \simeq H^k(\Omega_2; L^2(\Omega_1))$ and $L^2(\Omega_1) \otimes H^k(\Omega_2) = H^k(\Omega_2) \otimes L^2(\Omega_1)$, thus we have $L^2(\Omega_1) \otimes H^k(\Omega_2) \simeq L^2(\Omega_1; H^k(\Omega_2)) \simeq H^k(\Omega_2; L^2(\Omega_1))$.

Definition 2.27 (stochastic Sobolev spaces) For a given complete probability space

$(\Omega, \mathcal{F}, P)$, $D \subset \mathbb{R}^N$, the stochastic Sobolev spaces are defined by:

$$\begin{aligned} L_P^q(\Omega; W^{k,q}(D)) &= \left\{ v : \Omega \times D \rightarrow \mathbb{R} \mid v \text{ is strongly measurable} \right. \\ &\quad \left. \text{and } \int_{\Omega} \|v(w, \cdot)\|_{W^{k,q}(D)}^q dP(w) < \infty \right\}, \\ L_P^\infty(\Omega; W^{k,q}(D)) &= \left\{ v : \Omega \times D \rightarrow \mathbb{R} \mid v \text{ is strongly measurable} \right. \\ &\quad \left. \text{and } P - \text{ess sup}_{w \in \Omega} \|v(w, \cdot)\|_{W^{k,q}(D)}^q < \infty \right\} \end{aligned}$$

where $\int_A dP = \int I_A(w) dP(w) = P(A)$ for all $A \in \mathcal{F}$.

2.3 Important Inequalities

In this section, I introduce some important inequalities which are frequently used in my thesis.

2.3.1 Trace Inequalities

In this subsection, I first give the trace theorem, then the trace inequalities are recalled.

Theorem 2.2 (Trace Theorem)

Let D be a bounded domain with polygonal boundary ∂D and outward normal vector $\mathbf{n}$. There exist trace operators $\gamma_0 : H^s(D) \rightarrow H^{s-1/2}(\partial D)$ for $s > 1/2$ and $\gamma_1 : H^s(D) \rightarrow H^{s-3/2}(\partial D)$ for $s > 3/2$ that are extensions of the boundary values and boundary normal derivatives, respectively. The operators γ_j are surjective. Furthermore, if $v \in C^1(\bar{D})$, then

$$\gamma_0 v = v|_{\partial D}, \quad \gamma_1 v = \nabla v \cdot \mathbf{n}|_{\partial D}.$$

We recall some important trace inequalities which are frequently used in the analysis of the DG methods. Let D be a bounded polygonal domain with diameter h_D :

$$h_D = \sup_{x,y \in D} \|x - y\|,$$

where $\|x\|$ is the Euclidean norm ($\|x\| = (x \cdot x)^{1/2}$). Let $|D|$ denote the length of D in one dimension, the area of D in two dimensions, and the volume of D in three dimensions. Similarly, we use the length or area $|e|$ for an edge or a face of D . Then, there is a constant C independent of h_D, v such that for any $v \in H^s(D)$, we have

$$s \geq 1 \quad \forall e \subset \partial D, \|\gamma_0 v\|_{L^2(e)} \leq C|e|^{1/2}|D|^{-1/2}(\|v\|_{L^2(D)} + h_D\|\nabla v\|_{L^2(D)}), \quad (2.3)$$

$$s \geq 2 \quad \forall e \subset \partial D, \|\gamma_1 v\|_{L^2(e)} \leq C|e|^{1/2}|D|^{-1/2}(\|\nabla v\|_{L^2(D)} + h_D\|\nabla^2 v\|_{L^2(D)}). \quad (2.4)$$

When I use these two trace inequalities in my thesis, I will abuse the notation and replace the trace $\gamma_0 v$ and $\gamma_1 v$ by v and $\nabla v \cdot \mathbf{n}$, respectively.

Denote by $\mathbb{P}_k(D)$ the space of polynomials of degree less than or equal to k :

$$\mathbb{P}_k(D) = \text{span}\{x_1^{i_1} x_2^{i_2} \cdots x_d^{i_d} : i_1 + i_2 + \cdots + i_d \leq k, \mathbf{x} \in D\}.$$

The trace inequalities now become

$$\forall v \in \mathbb{P}_k(D), \forall e \subset \partial D, \|v\|_{L^2(e)} \leq C_0|e|^{1/2}|D|^{-1/2}\|v\|_{L^2(D)}, \quad (2.5)$$

$$\forall v \in \mathbb{P}_k(D), \forall e \subset \partial D, \|v\|_{L^2(e)} \leq C_1 h_D^{-1/2}\|v\|_{L^2(D)}, \quad (2.6)$$

$$\forall v \in \mathbb{P}_k(D), \forall e \subset \partial D, \|\nabla v \cdot \mathbf{n}\|_{L^2(e)} \leq C_0|e|^{1/2}|D|^{-1/2}\|\nabla v\|_{L^2(D)}, \quad (2.7)$$

$$\forall v \in \mathbb{P}_k(D), \forall e \subset \partial D, \|\nabla v \cdot \mathbf{n}\|_{L^2(e)} \leq C_1 h_D^{-1/2}\|\nabla v\|_{L^2(D)}, \quad (2.8)$$

2.3.2 Cauchy-Schwarz's and Young's Inequalities

Cauchy-Schwarz's and Young's Inequality are used in several places in my thesis.

Cauchy-Schwarz's Inequality:

$$\forall f, g \in L^2(D), \quad |(f, g)_D| \leq \|f\|_{L^2(D)}\|g\|_{L^2(D)}. \quad (2.9)$$

Young's Inequality:

$$\forall \epsilon > 0, \quad \forall a, b \in \mathbb{R}, \quad ab \leq \frac{\epsilon}{2}a^2 + \frac{1}{2\epsilon}b^2. \quad (2.10)$$

2.4 Model and Problem

Before I introduce the original stochastic problem in my thesis, I give some basic definitions in probability theory.

2.4.1 Probability Theory

First, we define the expected value (mean) of a random variable.

Definition 2.28 The *expected value* of a random variable X on the probability space $(\Omega, \mathcal{F}, P)$ is the integral of X respect to the measure P :

$$E[X] = \int_{\Omega} X(w) dP(w).$$

In practice, we usually compute the expected value over the real line rather than an integral over the original probability space. Thus, we need to define the distribution of a random variable.

Definition 2.29 Let $(\Omega, \mathcal{F}, P)$ be a measure space, $(\mathbb{R}, \mathcal{B})$ is a measurable space (where $\mathcal{B}$ is a Borel σ -field), $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$ is a random variable. Define a *distribution* of X on $\mathcal{B}$ by

$$(P \circ X^{-1})(C) = P(X^{-1}(C)), \quad C \in \mathcal{B}.$$

$P \circ X^{-1}$ is sometimes denoted by P_X .

Theorem 2.3 (Change of Variable)

Suppose $f : (\Omega, \mathcal{F}, \mu) \rightarrow (\Lambda, \mathcal{G})$ and $g : (\Lambda, \mathcal{G}) \rightarrow (\mathbb{R}, \mathcal{B})$. Then

$$\int_{\Omega} (g \circ f)(w) d\mu(w) = \int_{\Lambda} g(\lambda) d(\mu \circ f^{-1})(\lambda)$$

If $(\Omega, \mathcal{F}, P)$ is a probability space, X is a random variable defined thereon and if $E[X] = \int_{\Omega} X(\omega) dP(\omega)$ exists, then by Theorem 2.3 (with $g(x) = x, f = X, \mu = P$), we have

$$E[X] = \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbb{R}^N} x dP_X(x).$$

If P_X is absolutely continuous with respect to the Lebesgue measure, then there exists a density function $\rho : \mathbb{R} \rightarrow [0, +\infty)$ associated with X , such that

$$E[X] = \int_{\mathbb{R}} x \rho(x) dx.$$

Definition 2.30 If a random variable X has the expected value $E[X]$, then the variance of X is given by

$$Var(X) = E[(X - E[X])^2].$$

Next, we need the definition of stochastic processes.

Definition 2.31 A stochastic process is a parametrized collection of random variables

$$\{X_t\}_{t \in T}$$

defined on a probability space $(\Omega, \mathcal{F}, P)$ and assuming values in $\mathbb{R}^N$. The parameter space T is usually the halfline $[0, +\infty)$, but it may also be an interval $[a, b]$, the non-negative integers and even subsets of $\mathbb{R}^N$ for $N \geq 1$.

Besides, suppose $a(\omega, x)$ is a stochastic process with $\omega \in \Omega, x \in \bar{D}$ (where $D \in \mathbb{R}^n$ is a convex bounded domain), its covariance function is defined by

$$cov[a](x_1, x_2) = E[(a(\cdot, x_1) - E[a(\cdot, x_1)])(a(\cdot, x_2) - E[a(\cdot, x_2)])] \text{ for } x_1, x_2 \in \bar{D}.$$

In addition, $a(\omega, x)$ is called centered stochastic process if $E[a(\cdot, x)] = 0, \forall x \in D$.

2.4.2 Model

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, where Ω is the outcome space, $\mathcal{F}$ is the σ -algebra of events, and $P : \mathcal{F} \rightarrow [0, 1]$ is a probability measure. Assume $D \in \mathbb{R}^d$ is a convex bounded domain. My thesis addresses the linear elliptic stochastic boundary value problem: find a stochastic function, $u : \Omega \times \bar{D} \rightarrow \mathbb{R}$, such that the following equation is P -a.e. satisfied:

$$\begin{aligned} \nabla \cdot (a(w, \cdot) \nabla u(w, \cdot)) &= f(w, \cdot) \quad \text{in } D, \\ u(w, \cdot) &= 0 \quad \text{on } \partial D, \end{aligned} \tag{2.11}$$

where $a, f : \Omega \times D \rightarrow \mathbb{R}$ are stochastic functions with continuous and bounded covariance functions. In addition, I need to assume that a is bounded and uniformly coercive, i.e.,

$$\exists a_{\min}, a_{\max} \in (0, \infty) : \quad P(\omega \in \Omega : a(\omega, x) \in [a_{\min}, a_{\max}] \quad a.e. \ x \in \bar{D}) = 1. \tag{2.12}$$

The regularity of the solution u can be ensured by the assumption that the first derivative of a is uniformly bounded and continuous, i.e., there exists a constant C such that

$$P(\omega \in \Omega : a(\omega, \cdot) \in C^1(\bar{D}) \quad \text{and} \quad \max_{\bar{D}} |\nabla_x a(\omega, \cdot)| < C) = 1. \tag{2.13}$$

Additionally, the load function f is assumed to satisfy

$$\int_{\Omega} \int_D f^2(\omega, x) dx dP(\Omega) < \infty, \quad \text{which implies} \quad \int_D f^2(\omega, x) dx < \infty \quad a.s. \tag{2.14}$$

2.5 Existence and Uniqueness for the Solution

Following the work [10] of Babuška, Tempone and Zouraris, we apply the Lax-Milgram Lemma [19] to show the existence and uniqueness of the solution to (2.11).

Define $H = L_P^2(\Omega; H_0^1(D))$, the stochastic Sobolev space equipped with the inner product

$$(v, w)_H = E[(\nabla v, \nabla w)_{L^2(D)}] \quad \forall v, w \in H. \quad (2.15)$$

Define the bilinear form $\mathcal{B} : H \times H \rightarrow \mathbb{R}$:

$$\mathcal{B}(v, w) = E[(a \nabla v, \nabla w)_{L^2(D)}] \quad \forall v, w \in H.$$

By assumption (2.12), we have

$$\begin{aligned} |\mathcal{B}(v, w)| &\leq a_{max} \|v\|_H \|w\|_H \quad v, w \in H, \\ a_{min} \|v\|_H^2 &\leq \mathcal{B}(v, v) \quad v \in H. \end{aligned} \quad (2.16)$$

where $\|\cdot\|_H$ is the norm induced by (2.15). We thus have the continuity and coercivity of $\mathcal{B}$. Define $\mathcal{L}(v) = E[(f, v)_{L^2(D)}]$ for all $v \in H$. Since the load function f satisfies assumption (2.14), $\mathcal{L}$ is bounded and linear on H . Applying the Lax-Milgram Lemma, we obtain existence and uniqueness for the solution to the variational formulation: find $u \in H$ such that

$$\mathcal{B}(u, v) = \mathcal{L}(v) \quad \forall v \in H. \quad (2.17)$$

Since the diffusion function a satisfies (2.12) and (2.13), and the domain D is convex and bounded, by the theory of elliptic regularity, we know the solution of (2.11) satisfies $u \in H^2(D) \cap H_0^1(D)$ almost surely [20]. Additionally, we can show that the solution to (2.11) also solves (2.17).

The formulation (2.17) is the basis for the derivation of the Stochastic Finite Element Method and the Monte Carlo Finite Element Method, which I review in Chapter 3 and Chapter 4.

2.6 Karhunen–Loève Expansion and Finite Dimensional Noise

In this section, I first give a detailed introduction to the Karhunen–Loève (K–L) expansion. Then, I give the important assumption about finite dimensional noise, which is motivated by the K–L expansion. Therefore, I can transform the stochastic model in my thesis to a “deterministic” problem which can be approximated by the use of finite element and discontinuous Galerkin methods [10, 11].

2.6.1 Karhunen–Loève Expansion

The K–L expansion is an appropriate method to approximate stochastic processes [21]. In this section, I give a detailed description of K–L expansion. I define the integral operator T_K for $K : D \times D \rightarrow \mathbb{R}^d$:

$$(T_K f)(x) = \int_D K(x, y) f(y) dy \quad \forall f \in L^2(D) \quad \forall x \in D, \quad (2.18)$$

where D is a convex and bounded domain in $\mathbb{R}^d$.

Theorem 2.4 (Mercer’s Theorem[22])

Let K be a real-valued symmetric, continuous function of x and y . Assume in addition that the operator T_K is non-negative in the usual sense:

$$(T_K f, f) \geq 0 \quad \forall f \in L^2(D).$$

Then the kernel K can be expanded in a uniformly convergent series

$$K(x, y) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(y),$$

where λ_i and ϕ_i are the non-negative eigenvalues and orthonormal eigenfunctions of T_K respectively.

By Mercer’s Theorem, we obtain the following important result which is known as Karhunen–Loève Theorem.

Theorem 2.5 (Karhunen–Loève Theorem[23])

Consider a centered stochastic process $a(\omega, x)$ with continuous covariance function $cov[a]$. Then the corresponding integral operator $T_a : L^2(D) \rightarrow L^2(D)$ defined by

$$(T_a f)(\cdot) = \int_D cov[a](x, \cdot) f(x) dx \quad \forall f \in L^2(D), \quad (2.19)$$

has an orthonormal basis of eigenfunctions $\{\phi_i\}_i \in L^2(D)$. If $Z_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D a(\omega, x) \phi_i(x) dx$, then Z_i are centered random variables that are orthogonal, namely $E[Z_i Z_j] = 0, i \neq j$ and

$$a(\omega, x) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \phi_i(x) Z_i(\omega).$$

Let $\{\phi_i\}_{i=1}^{\infty}$ be the sequence of the corresponding orthonormal eigenfunctions, i.e.,

$$(\phi_i, \phi_j)_{L^2(D)} = \delta_{ij}, i, j \in \mathbb{N}_+.$$

By Mercer's Theorem, we know that the eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ of the operator T_a are non-negative. We order these eigenvalues so that the eigenvalues of T_a are decreasing, i.e., $\lambda_i \geq \lambda_{i+1}, i \in \mathbb{N}_+$.

Proposition 1 Let $\{\lambda_i\}_{i=1}^{\infty}$ be the set of eigenvalues of T_a . Then

$$\lambda_i \leq \left(\int_{D \times D} (cov[a])^2(x_1, x_2) dx_1 dx_2 \right)^{\frac{1}{2}}.$$

Proof 2.1 We have

$$\begin{aligned} (T_a \phi_i)(x_2) &= \int_D cov[a](x_1, x_2) \phi_i(x_1) dx_1, \\ (T_a \phi_i)(x_2) &= \lambda_i \phi_i(x_2). \end{aligned}$$

By Cauchy-Schwarz's inequality, this implies

$$\lambda_i \|\phi_i\|_{L^2(D)} \leq \left(\int_{D \times D} (cov[a])^2(x_1, x_2) dx_1 dx_2 \right)^{\frac{1}{2}} \|\phi_i\|_{L^2(D)}.$$

Since, $\|\phi_i\|_{L^2(D)} = 1$, we obtain

$$\lambda_i \leq \left(\int_{D \times D} (\text{cov}[a])^2(x_1, x_2) dx_1 dx_2 \right)^{\frac{1}{2}}, \quad \forall i \in \mathbb{N}_+.$$

This is the conclusion we desire. $\square$

Note that $a(\omega, x) - E[a(\cdot, x)]$ is a centered stochastic process. We apply Karhunen–Loève theorem to this new stochastic process. Define

$$Y_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D (a(\omega, x) - E[a(\cdot, x)]) \phi_i(x) dx, \quad \text{for } \lambda_i > 0.$$

Therefore, Y_i 's have the properties

- $E[Y_i] = 0$,
- $E[Y_i Y_j] = 0, i \neq j$,
- $E[Y_i^2] = 1$.

By Karhunen–Loève Theorem, the stochastic process $a(\omega, x)$ can be written as

$$a(\omega, x) = E[a(\cdot, x)] + \sum_{i=1}^{\infty} \sqrt{\lambda_i} \phi_i(x) Y_i(\omega).$$

Then we can approximate the stochastic process $a(\omega, x)$ by the truncated Karhunen–Loève expansion $a_N(\omega, x)$ which is defined by

$$a_N(\omega, x) = E[a(\cdot, x)] + \sum_{i=1}^N \sqrt{\lambda_i} \phi_i(x) Y_i(\omega), \quad N \in \mathbb{N}_+.$$

2.6.2 Finite Dimensional Noise

In my thesis, I assume the diffusion function a and the load function f have the form:

$$a(\omega, x) = a(Y_1(\omega), \dots, Y_N(\omega), x) \quad \text{on } \Omega \times \bar{D}, \quad (2.20)$$

$$f(\omega, x) = f(Y_1(\omega), \dots, Y_N(\omega), x) \quad \text{on } \Omega \times \bar{D}, \quad (2.21)$$

and satisfy the truncated finite Karhunen–Loève expansion, i.e.

$$a(\omega, x) = E[a(\cdot, x)] + \sum_{n=1}^N \sqrt{\lambda_n} b_n(x) Y_n(\omega), \quad (2.22)$$

$$f(\omega, x) = E[f(\cdot, x)] + \sum_{n=1}^N \sqrt{\hat{\lambda}_n} \hat{b}_n(x) \hat{Y}_n(\omega). \quad (2.23)$$

where $N \in \mathbb{N}_+$, $\{Y\}_{i=1}^N$ are real-valued random variables with mean value zero and unit variance. Moreover, I assume that for $i = 1, \dots, N$:

- the image $\Gamma_i = Y_i(\Omega)$ is a bounded interval in $\mathbb{R}$.
- each Y_i has a density function $\rho_i : \Gamma_i \rightarrow \mathbb{R}^+$.

In the following, I denote $\rho = \rho(y) : \Gamma \rightarrow \mathbb{R}^+$ for the joint probability density of the random vector $Y = (Y_1, \dots, Y_N)$, where $\Gamma = \prod_{i=1}^N \Gamma_i \subset \mathbb{R}^N$ is the support of the joint probability. I also assume that the random variables $\{Y_i\}_{i=1}^N$ are independent, thus we have

$$\rho(y) = \prod_{i=1}^N \rho_i(y_i) \quad \forall y \in \Gamma.$$

If we apply the Doob-Dynkin lemma [24] with the finite dimensional noise assumption, we know that the solution u corresponding to (2.11) can be expressed by a finite number of random variables, i.e. $u(\omega, x) = u(Y_1(\omega), \dots, Y_N(\omega), x)$. Now, instead of approximating $u(\omega, \cdot)$, our task is to approximate $u(y, x)$, where $y \in \Gamma$ and $x \in D$. Notice that (2.11) is rewritten as: find $u \in L_P^2(\Gamma) \otimes H_0^1(D)$ such that

$$\begin{aligned} & \int_{\Gamma} \rho(y) \int_D a(y, x) \nabla u(y, x) \cdot \nabla v(y, x) dx dy \\ &= \int_{\Gamma} \rho(y) \int_D f(y, x) v(y, x) dx dy \quad \forall v \in L_P^2(\Gamma) \otimes H_0^1(D). \end{aligned} \quad (2.24)$$

Note: Unless specifically mentioned, the gradient notation ∇ in my thesis always means differentiation with respect to $x \in D$.

Therefore, the stochastic boundary value problem (2.11) is transformed into an elliptic partial differential equation with an N -dimensional parameter, i.e.,

$$\begin{aligned}\nabla \cdot (a(y, x) \nabla u(y, x)) &= f(y, x) \quad \forall (y, x) \in \Gamma \times D, \\ u(y, x) &= 0 \quad \forall (y, x) \in \Gamma \times \partial D.\end{aligned}\tag{2.25}$$

Thus, the finite dimensional noise assumption helps us transform the stochastic elliptic problem into a parametrized elliptic problem. At this point, if I sample the $a(y, \cdot), f(y, \cdot)$, then I can use the finite element or discontinuous Galerkin method with Monte Carlo method to approximate the solution to (2.25).

Chapter 3

Stochastic Finite Element Method

3.1 Introduction

In this chapter, I recall the work by Babuška, Tempone and Zouraris (Section 3 in [10]) about the stochastic finite element method. First, I recall the h -version on the spatial space D , followed by the k -version on the outcome space Γ . Then, using tensor product finite element spaces, I recall the $k \times h$ -version of stochastic finite element method. Besides, I give a detailed proof for a priori error estimate in the energy norm.

3.2 Finite Element Spaces and Approximation Results

On a convex polygonal domain $D \subset \mathbb{R}^d$, suppose $\{\mathcal{E}_h\}_{h>0}$ is a non-degenerate family of triangulations. Considering the finite element space $V_h \subset H_0^1(D)$ which consists of the continuous piecewise linear functions, we have the approximation property:

$$\min_{\phi \in V_h} \|u - \phi\|_{H_0^1(D)} \leq Ch \|u\|_{H^2(D)} \quad \forall u \in H^2(D) \cap H_0^1(D),$$

where h is the mesh size, C is a constant independent of u and h , but dependent on D . Throughout my thesis, C is a constant that takes different values at different places.

Recall $\Gamma = \prod_{n=1}^N \Gamma_i$ as defined in section 2.6.2. Consider the disjoint $\mathbb{R}^N$ -boxes

$$\beta = \prod_{n=1}^N (a_n^\beta, b_n^\beta) \quad \text{where } (a_n^\beta, b_n^\beta) \subset \Gamma_n$$

as a partition of Γ . The mesh size is defined by $k_n = \max_{\beta} |b_n^{\beta} - a_n^{\beta}|, n = 1, \dots, N$. Consider the finite element space $Y_k^q \subset L^2(\Gamma)$ consisting of discontinuous piecewise polynomials with degree at most q_n on each direction y_n , where $q = (q_1, \dots, q_N)$ and $k = (k_1, \dots, k_N)$ are non-negative N-uples. Clearly, if $\phi \in Y_k^q \subset L^2(\Gamma)$, its restriction to each box β of the partition of Γ satisfies:

$$\phi|_{\beta} \in \text{span} \left(\prod_{n=1}^N y_n^{\alpha_n} : \alpha_n \in \mathbb{N} \text{ and } \alpha_n \leq q_n, n = 1, \dots, N \right).$$

The finite dimensional space Y_k^q has the approximation property [25]:

$$\min_{\phi \in Y_k^q} \|u - \phi\|_{L^2(\Gamma)} \leq \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} u\|_{L^2(\Gamma)}}{(q_n + 1)!} \quad \forall u \in H^{q+1}(\Gamma). \quad (3.1)$$

In the section 3.3 of [10], the authors obtain the following approximation property for the tensor product finite element space $Y_k^q \otimes V_h$.

Proposition 2 : There exists a constant $C > 0$ independent of h, N, q and k such that

$$\begin{aligned} \inf_{\psi \in Y_k^q \otimes V_h} \|u - \psi\|_{L^2(\Gamma; H_0^1(D))} &\leq C \left(h \|u\|_{L^2(\Gamma; H^2(D))} \right. \\ &\quad \left. + \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} u\|_{L^2(\Gamma; H_0^1(D))}}{(q_n + 1)!} \right) \end{aligned} \quad (3.2)$$

for all $u \in C^{q+1}(\Gamma; H^2(D) \cap H_0^1(D))$.

The elliptic regularity of problem (2.25) yields that

$$\|u(y, \cdot)\|_{H^2(D)} \leq C \|f(y, \cdot)\|_{L^2(D)} \quad \forall y \in \Gamma, \quad (3.3)$$

where C is a constant depending on D and $\|a\|_{L^\infty(\Gamma; W^{1,\infty}(D))}$. In addition, we have

$$\frac{\|\partial_{y_n}^{q_n+1} u\|_{H_0^1(D)}}{(q_n + 1)!} \leq (r_n)^{q_n} \frac{C}{a_{\min}} (\|\partial_{y_n} f\|_{L^2(D)} + r_n \|f\|_{L^2(D)}), \quad (3.4)$$

where $r_n = \sqrt{\lambda_n} \|\frac{\phi_n}{a}\|_{L^\infty(\Gamma \times D)}, n = 1, \dots, N$, a is the coefficient of problem (2.25), λ_n and ϕ_n are the eigenvalue and eigenfunction associated with the operator (2.19)

and C is a constant only depending on the domain D . Substituting (3.3) and (3.4) into (3.2), we obtain

$$\begin{aligned}
\inf_{\psi \in Y_k^q \otimes V_h} \|u - \psi\|_{L^2(\Gamma; H_0^1(D))} &\leq C \left(h \|u\|_{L^2(\Gamma; H^2(D))} + \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} u\|_{L^2(\Gamma; H_0^1(D))}}{(q_n+1)!} \right) \\
&\leq Ch \|f\|_{L^2(\Gamma; L^2(D))} \\
&\quad + C \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{(r_n)^{q_n}}{a_{min}} \left(\|\partial_{y_n} f\|_{L^2(\Gamma; L^2(D))} \right. \\
&\quad \left. + r_n \|f\|_{L^2(\Gamma; L^2(D))} \right).
\end{aligned} \tag{3.5}$$

3.3 $k \times h$ -version Stochastic Finite Element Method

The finite dimensional space is the tensor product space $Y_k^q \otimes V_h$:

$$Y_k^q \otimes V_h = \{\psi \in L^2(\Gamma \times D) : \psi \in \text{span}(\varphi(y)\chi(x) : \varphi \in Y_k^q, \chi \in V_h)\}$$

where V_h and Y_k^q are defined in section 3.2.

Multiply by a test function $\psi \in Y_k^q \otimes V_h$ both sides of the first equation in problem (2.25). Integrate by parts and integrate again on the statistic domain Γ :

$$\int_{\Gamma} \rho(a \nabla u, \nabla \psi)_{L^2(D)} dy = \int_{\Gamma} \rho(f, \psi)_{L^2(D)} dy, \quad \forall \psi \in Y_k^q \otimes V_h. \tag{3.6}$$

We rewrite (3.6) as

$$\mathcal{A}(u, \psi) = L(\psi), \quad \forall \psi \in Y_k^q \otimes V_h. \tag{3.7}$$

Our goal is to find the approximate solution $u_{kh} \in Y_k^q \otimes V_h$ such that

$$\mathcal{A}(u_{kh}, \psi) = L(\psi), \quad \forall \psi \in Y_k^q \otimes V_h. \tag{3.8}$$

By (3.7) and (3.8), we have

$$\mathcal{A}(u_{kh} - u, \psi) = 0, \quad \forall \psi \in Y_k^q \otimes V_h.$$

Substitute $(u_{kh} - u + u - \psi)$ for ψ in the above formula, we get

$$\mathcal{A}(u_{kh} - u, u_{kh} - u) = \mathcal{A}(u_{kh} - u, \psi - u), \quad \forall \psi \in Y_k^q \otimes V_h.$$

Define the energy norm

$$\|v\|_{\mathcal{A}} = (\mathcal{A}(v, v))^{\frac{1}{2}}.$$

Cauchy-Schwarz's inequality implies that

$$\|u_{kh} - u\|_{\mathcal{A}}^2 \leq \|u_{kh} - u\|_{\mathcal{A}} \|u - \psi\|_{\mathcal{A}}.$$

Hence,

$$\|u_{kh} - u\|_{\mathcal{A}} \leq \|u - \psi\|_{\mathcal{A}}.$$

Therefore

$$\|u_{kh} - u\|_{\mathcal{A}} \leq \inf_{\psi \in Y_k^q \otimes V_h} \|u - \psi\|_{\mathcal{A}}.$$

Since

$$\begin{aligned} \|u - \psi\|_{\mathcal{A}}^2 &= \int_{\Gamma} \rho(a \nabla(u - \psi), \nabla(u - \psi))_{L^2(D)} dy \\ &\leq \|a\rho\|_{L^\infty(\Gamma \times D)} \int_{\Gamma} (\nabla(u - \psi), \nabla(u - \psi))_{L^2(D)} dy \\ &= \|a\rho\|_{L^\infty(\Gamma \times D)} \|u - \psi\|_{L^2(\Gamma; H_0^1(D))}^2. \end{aligned}$$

This implies

$$\|u_{kh} - u\|_{\mathcal{A}} \leq \sqrt{\|a\rho\|_{L^\infty(\Gamma \times D)}} \inf_{\psi \in Y_k^q \otimes V_h} \|u - \psi\|_{L^2(\Gamma; H_0^1(D))}. \quad (3.9)$$

As a consequence of (3.9) and (3.5), we obtain an a priori error estimate for the $k \times h$ -version stochastic finite element method:

$$\begin{aligned} \frac{\|u_{kh} - u\|_{\mathcal{A}}}{\sqrt{\|a\rho\|_{L^\infty(\Gamma \times D)}}} &\leq Ch \|f\|_{L^2(\Gamma; L^2(D))} \\ &\quad + C \sum_{n=1}^N \left(\frac{k_n}{2}\right)^{q_n+1} \frac{(r_n)^{q_n}}{a_{min}} \left(\|\partial_{y_n} f\|_{L^2(\Gamma; L^2(D))} \right. \\ &\quad \left. + r_n \|f\|_{L^2(\Gamma; L^2(D))} \right). \end{aligned} \quad (3.10)$$

We also recall the $H_0^1(D)$ and $L^2(D)$ error estimates of the expected value of the solution to the problem (2.25).

Theorem 3.1 ([10])

Let u be the solution of the problem (2.25) and $u_{kh} \in Y_k^q \otimes V_h$ be the $k \times h$ -version stochastic finite element method approximation of u defined in (3.8). If $\rho \in L^\infty(\Gamma)$, then for $l = 0, 1$, we have

$$\frac{\|E[u(\cdot, x)] - E[u_{kh}(\cdot, x)]\|_{H^l(D)}}{\|a\rho\|_{L^\infty(\Gamma \times D)}} \leq C \left(h^{2-l} + \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{2-l} \left(\frac{k_n r_n}{2} \right)^{(2-l)q_n} \right). \quad (3.11)$$

where the constant C depends on D, f , and a , but is independent of k, q , and h .

Chapter 4

Stochastic Discontinuous Galerkin Method

4.1 Introduction

In this chapter, I establish the framework for the stochastic discontinuous Galerkin (SDG) method. I first introduce the primal discontinuous Galerkin (DG) methods by defining the DG space, introducing the variational formulation for the DG scheme, discussing the existence and uniqueness of the DG solution and stating a priori error estimates in the energy norm and L^2 norm. In section 3, I establish the variational formulation for the SDG method. In section 4 and section 5, I give two a priori error estimates for the solution and its expected value in the energy norm and L^2 norm for the SDG methods.

4.2 Primal Discontinuous Galerkin Methods

In this section, I recall the definition and properties of the primal discontinuous Galerkin methods [26].

4.2.1 Model Problem

Let D be a polygonal domain in $\mathbb{R}^d$, $d = 2$ or 3 . The sides of the boundary ∂D are grouped into two disjoint sets Γ_M and Γ_N . Let n be the unit normal vector to the boundary exterior to D . For f given in $L^2(D)$, g_M given in $H^{\frac{1}{2}}(\Gamma_M)$ and g_N given in

$L^2(\Gamma_N)$, we consider the following elliptic problem:

$$-\nabla \cdot (a \nabla u) = f \quad \text{in } D, \quad (4.1)$$

$$u = g_M \quad \text{on } \Gamma_M, \quad (4.2)$$

$$a \nabla u \cdot n = g_N \quad \text{on } \Gamma_N. \quad (4.3)$$

The coefficient a is a matrix-valued function $a = (a_{ij})_{1 \leq i, j \leq d}$ that is symmetric positive definite and bounded below and above uniformly, which means there exist two positive constants a_{min} and a_{max} such that

$$\forall v \in \mathbb{R}^d, \quad a_{min} v^T v \leq v^T a v \leq a_{max} v^T v.$$

The equation (4.2) is called a Dirichlet boundary condition. The equation (4.3) is called a Neumann boundary condition. We discuss the framework of primal discontinuous Galerkin method in this section for the problem (4.1)–(4.3).

4.2.2 Discontinuous Galerkin Space

Before I give the definition of discontinuous Galerkin space, I need to introduce the broken Sobolev spaces, which are the natural spaces working with the DG methods. Let D be a polygonal domain which is divided into elements E . In 2D space, E is triangular or a quadrilateral and in 3D space, E is a tetrahedron or a hexahedron. In my thesis, I only consider conforming meshes, i.e. the intersection of two elements in the mesh is either empty, a vertex, an edge, or a face. I now give some notation:

- $\mathcal{E}_h$: the subdivision (or mesh) of the domain D ,
- h : the maximum element diameter,
- h_E : the diameter of element E ,

- ρ_E : the maximum diameter of a ball inscribed in element E .

I assume that the subdivision is *regular*, which means there is a constant $\rho > 0$ such that

$$\frac{h_E}{\rho_E} \leq \rho \quad \forall E \in \mathcal{E}_h.$$

Now, we are ready to introduce the *broken Sobolev spaces*.

Definition 4.1 The *broken Sobolev spaces* for any real number s are defined :

$$H^s(\mathcal{E}_h) = \{u \in L^2(D) : \forall E \in \mathcal{E}_h, u|_E \in H^s(E)\},$$

with the broken Sobolev norm:

$$\|u\|_{H^s(\mathcal{E}_h)} = \left(\sum_{E \in \mathcal{E}_h} \|u\|_{H^s(E)}^2 \right)^{1/2}.$$

Then, we introduce the DG spaces that are finite dimensional subspaces of the broken Sobolev space $H^s(\mathcal{E}_h)$. Let r be a positive integer. The discontinuous Galerkin spaces are defined as:

$$D_r(\mathcal{E}_h) = \{u \in L^2(D) : \forall E \in \mathcal{E}_h, u|_E \in \mathbb{P}_r(E)\}, \quad (4.4)$$

where $\mathbb{P}_r(E)$ denotes the space of polynomials of total degree less than or equal to r on E .

I now give the concepts of the jumps and averages related to the elements in the subdivision $\mathcal{E}_h$ to prepare for the introduction of the variational formulation in the following section. Denote by Γ_h the set of interior edges (or faces) of the subdivision $\mathcal{E}_h$. With each edge (or face) e , we associate a unit normal vector n_e . If e is on the boundary ∂D , then n_e is the outward vector normal to ∂D . We assume that $u \in H^1(\mathcal{E}_h)$. If two elements E_1^e and E_2^e are neighbours with one common side e ,

there are two traces of u along e . We also assume that the normal vector n_e is oriented from E_1^e to E_2^e , and define the average and jump as:

$$\{u\} = \frac{1}{2}(u|_{E_1^e}) + \frac{1}{2}(u|_{E_2^e}), \quad [u] = (u|_{E_1^e}) - (u|_{E_2^e}) \quad \forall e = \partial E_1^e \cap \partial E_2^e.$$

We extend the definition to sides that belong to the boundary ∂D :

$$\{u\} = [u] = u|_{E_1^e}, \quad \forall e = \partial E_1^e \cap \partial D.$$

4.2.3 Variational Formulation

We assume $s > \frac{1}{2}$ and introduce the bilinear form

$$J_0^{\sigma, \beta} : H^s(\mathcal{E}_h) \times H^s(\mathcal{E}_h) \rightarrow \mathbb{R}$$

that penalizes the jump of the function values:

$$J_0^{\sigma, \beta}(v, w) = \sum_{e \in \Gamma_h \cup \Gamma_M} \frac{\sigma_e}{|e|^\beta} \int_e [v][w]. \quad (4.5)$$

The parameter σ_e are called penalty parameters. They are nonnegative real numbers.

The power β is a positive number that depends on the dimension d .

We now define the DG bilinear form $A_\epsilon : H^s(\mathcal{E}_h) \times H^s(\mathcal{E}_h) \rightarrow \mathbb{R}$, for $s > \frac{3}{2}$

$$\begin{aligned} A_\epsilon(v, w) &= \sum_{E \in \mathcal{E}_h} \int_E a \nabla v \cdot \nabla w - \sum_{e \in \Gamma_h \cup \Gamma_M} \int_e \{a \nabla v \cdot n_e\} [w] \\ &\quad + \epsilon \sum_{e \in \Gamma_h \cup \Gamma_M} \int_e \{a \nabla w \cdot n_e\} [v] + J_0^{\sigma, \beta}(v, w). \end{aligned} \quad (4.6)$$

- If $\epsilon = -1$, and σ_e is bounded below by a large enough constant, the resulting method is called the symmetric interior penalty Galerkin (SIPG) method.
- If $\epsilon = +1$, and $\sigma_e = 1$, the resulting method is called the nonsymmetric interior penalty Galerkin (NIPG) method.

- If $\epsilon = 0$ and σ_e is bounded below by a large enough constant, we obtain the incomplete interior penalty Galerkin (IIPG) method.

We also define the following linear form:

$$L(v) = \int_D f v + \epsilon \sum_{e \in \partial D} \int_e \left(a \nabla v \cdot n_e + \frac{\sigma_e}{|e|^\beta} v \right) g_M + \sum_{e \in \Gamma_N} \int_e v g_N.$$

The general DG variational formulation of problem (4.1)-(4.3) is as follows: Find u_h in $D_r(\mathcal{E}_h)$, such that

$$\forall v \in D_r(\mathcal{E}_h), \quad A_\epsilon(u_h, v) = L(v). \quad (4.7)$$

4.2.4 Existence and Uniqueness of DG solution

Assume that the following assumptions hold true:

Hyp 1. in the NIPG case, $r \geq 1$ and $\sigma_e > 0$ for all e ;

Hyp 2. in the SIPG or IIPG case, $r \geq 1$ and σ_e is bounded below by a large constant for all e ;

Hyp 3. in the NIPG case, $r \geq 2$ and $\sigma_e = 0$ for all e .

Then, the bilinear form A_ϵ is coercive and the DG solution u_h exists and is unique [26].

4.2.5 Error Estimates

In this section, I recall two basic theorems about the error estimates between the exact and the DG solution. One is with respect to the energy norm and the other is with respect to the L^2 norm. Define the energy norm on $\mathcal{D}_r(\mathcal{E}_h)$:

$$\|v\|_{\mathcal{E}} = \left(\sum_{E \in \mathcal{E}_h} \int_E a(\nabla v)^2 + J_0^{\sigma, \beta}(v, v) \right)^{\frac{1}{2}}.$$

We have the following theorem:

Theorem 4.1

Assume that the exact solution to (4.1)-(4.3) belongs to $H^s(\mathcal{E}_h)$ for $s > \frac{3}{2}$. Assume that Hyp 1, Hyp 2 and Hyp 3 hold true. Then, there is a constant C independent of h such that the following optimal a priori error estimate holds:

$$\|u - u_h\|_{\mathcal{E}} \leq Ch^{\min(r+1, s)-1} \|u\|_{H^s(\mathcal{E}_h)}. \quad (4.8)$$

I also recall the L^2 error estimate for the SIPG method.

Theorem 4.2

Assume that the exact solution to (4.1)-(4.3) belongs to $H^s(\mathcal{E}_h)$ for $s > \frac{3}{2}$. Assume that the domain D is convex and Hyp 2 holds true. Then, there is a constant C independent of h such that

$$\|u - u_h\|_{L^2(D)} \leq Ch^{\min(r+1, s)} \|u\|_{H^s(\mathcal{E}_h)}. \quad (4.9)$$

4.3 Stochastic Discontinuous Galerkin Formulation

In this section, I formulate the stochastic discontinuous Galerkin formulation for the linear elliptic problem defined in section 2.6.2. We recall the problem below:

$$\begin{aligned} -\nabla \cdot (a(y, x) \nabla u(y, x)) &= f(y, x) \quad \forall (y, x) \in \Gamma \times D, \\ u(y, x) &= 0 \quad \forall (y, x) \in \Gamma \times \partial D. \end{aligned} \quad (4.10)$$

I construct the tensor product discontinuous Galerkin space on $\Gamma \times D$:

$$Y_k^q \otimes \mathcal{D}_r(\mathcal{E}_h) = \{\psi \in L^2(\Gamma \times D) : \psi \in \text{span}(\varphi(y)\chi(x) : \varphi \in Y_k^q, \chi \in \mathcal{D}_r(\mathcal{E}_h))\}, \quad (4.11)$$

where Y_k^q is the finite dimensional space on Γ , defined in section 3.2 and $\mathcal{D}_r(\mathcal{E}_h)$ is the DG space on the spacial domain D , defined in section 4.2.2.

Following (4.6), we define the bilinear form:

$$\begin{aligned} A_\epsilon(y; u, v) = & \sum_{E \in \mathcal{E}_h} \int_E a(y, \cdot) \nabla u \cdot \nabla v - \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a(y, \cdot) \nabla u \cdot n_e\} [v] \\ & + \epsilon \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a(y, \cdot) \nabla v \cdot n_e\} [u] + J_0^{\sigma, \beta}(v, u), \end{aligned} \quad (4.12)$$

where $J_0^{\sigma, \beta}(v, u)$ is given by (4.5) with $\Gamma_M = \partial D$ and Γ_h is the set of interior edges (or faces) of the subdivision $\mathcal{E}_h$; ∂D is the boundary of the domain D . The SDG bilinear form can be defined as:

$$\mathcal{A}_\epsilon(u, v) = \int_\Gamma \rho A_\epsilon(y; u, v) dy. \quad (4.13)$$

Define the DG norm on D :

$$\|u(y, \cdot)\|_{DG} = \left(\sum_{E \in \mathcal{E}_h} \int_E a(y, x) (\nabla u)^2 + \sum_{e \in \Gamma_h \cup \partial D} \int_e [u]^2 \right)^{\frac{1}{2}}.$$

Then we define the energy norm on $\Gamma \times D$:

$$\|u\|_{\mathcal{E}} = \left(\int_\Gamma \rho \|u(y, \cdot)\|_{DG}^2 dy \right)^{\frac{1}{2}}.$$

The coercivity of A_ϵ implies the coercivity of $\mathcal{A}_\epsilon$:

$$\mathcal{A}_\epsilon(v, v) \geq \alpha \|v\|_{\mathcal{E}}^2 \quad v \in Y_k^q \otimes \mathcal{D}_r(\mathcal{E}_h) \quad (4.14)$$

where the constant α is independent of h and a , but depends on $a_{\min}$.

I also define the following linear form:

$$\mathcal{L}(v) = \int_\Gamma \rho(f, v)_{L^2(D)} dy. \quad (4.15)$$

Thus, the SDG variational formulation of (4.10) is as follows: Find the stochastic solution $u_{kh} \in Y_k^q \otimes \mathcal{D}_r(\mathcal{E}_h)$ such that

$$\forall v \in Y_k^q \otimes \mathcal{D}_r(\mathcal{E}_h), \quad \mathcal{A}_\epsilon(u_{kh}, v) = \mathcal{L}(v). \quad (4.16)$$

4.4 Error Estimate in the Energy Norm

In this section, we derive a priori error estimates for the SDG method with the parameter $\epsilon = -1, 1, 0$. We want to estimate $\|u - u_{kh}\|_{\mathcal{E}}$. By triangular inequality, we split the estimate into two parts:

$$\|u - u_{kh}\|_{\mathcal{E}} \leq \|u - \tilde{u}\|_{\mathcal{E}} + \|u_{kh} - \tilde{u}\|_{\mathcal{E}},$$

where $\tilde{u} \in Y_k^q \otimes \mathcal{D}_r(\mathcal{E}_h)$ is a good approximation of the exact solution u , which satisfies the following theorem.

Theorem 4.3

Let $v \in L^2(\Gamma; H^s(D)) \cap H^{q+1}(\Gamma; H_0^1(D))$ for $s \geq 2$ and let $r \geq 0$ be an integer. There exists a constant C independent of v, h and k_n , and a function $\tilde{v} \in Y_k^q \otimes \mathcal{D}_r(\mathcal{E}_h)$ such that

$$\begin{aligned} \|\nabla(v - \tilde{v})\|_{L^2(\Gamma; L^2(D))} &\leq Ch^{\min(r+1, s)-1} \|v\|_{L^2(\Gamma; H^s(D))} \\ &\quad + \sum_{n=1}^N \left(\frac{k_n}{2}\right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} v\|_{L^2(\Gamma; H_0^1(D))}}{(q_n + 1)!}, \end{aligned} \quad (4.17)$$

$$\begin{aligned} \|\nabla^2(v - \tilde{v})\|_{L^2(\Gamma; L^2(D))} &\leq Ch^{\min(r+1, s)-2} \|v\|_{L^2(\Gamma; H^s(D))} \\ &\quad + Ch^{-1} \sum_{n=1}^N \left(\frac{k_n}{2}\right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} v\|_{L^2(\Gamma; H_0^1(D))}}{(q_n + 1)!}, \end{aligned} \quad (4.18)$$

where q, q_n, k , and k_n are defined in Section 3.2. In addition, we can require that $\tilde{v}$ is continuous and vanishes on the boundary of domain D .

Proof 4.1 First, we prove the inequality (4.17).

Recall the L^2 -projection operator $\Pi_k^q : L^2(\Gamma) \rightarrow Y_k^q$ by

$$(\Pi_k^q w - w, \varphi)_{L^2(\Gamma)} = 0 \quad \forall \varphi \in Y_k^q, \quad \forall w \in L^2(\Gamma).$$

and the H_0^1 -projection operator $\mathcal{R}_h : H_0^1(D) \rightarrow D_r(\mathcal{E}_h) \cap H_0^1(D)$ by

$$(\nabla(\mathcal{R}_h w - w), \nabla \chi)_{L^2(D)} = 0 \quad \forall \chi \in D_r(\mathcal{E}_h), \quad \forall w \in H_0^1(D). \quad (4.19)$$

Fix $v \in L^2(\Gamma; H^s(D)) \cap H^{q+1}(\Gamma; H_0^1(D))$. The equation (3.1) implies that

$$\|v - \Pi_k^q v\|_{L^2(\Gamma; H_0^1(D))} \leq \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} v\|_{L^2(\Gamma; H_0^1(D))}}{(q_n + 1)!}. \quad (4.20)$$

By the Theorem 2.6 in [26], we have

$$\|v - \mathcal{R}_h v\|_{L^2(\Gamma; H_0^1(D))} \leq C h^{\min(r+1, s)-1} \|v\|_{L^2(\Gamma; H^s(D))}.$$

Choose $\tilde{v} = \Pi_k^q(\mathcal{R}_h v)$, thus

$$\begin{aligned} \|v - \tilde{v}\|_{L^2(\Gamma; H_0^1(D))} &= \|v - \Pi_k^q(\mathcal{R}_h v)\|_{L^2(\Gamma; H_0^1(D))} \\ &\leq \|v - \mathcal{R}_h v\|_{L^2(\Gamma; H_0^1(D))} + \|\mathcal{R}_h v - \Pi_k^q(\mathcal{R}_h v)\|_{L^2(\Gamma; H_0^1(D))} \\ &= \|v - \mathcal{R}_h v\|_{L^2(\Gamma; H_0^1(D))} + \|\mathcal{R}_h v - \mathcal{R}_h(\Pi_k^q v)\|_{L^2(\Gamma; H_0^1(D))}. \end{aligned}$$

Take $\chi = \mathcal{R}_h v$ in (4.19). This implies

$$(\nabla(\mathcal{R}_h v), \nabla(\mathcal{R}_h v))_{L^2(D)} = (\nabla v, \nabla(\mathcal{R}_h v))_{L^2(D)}.$$

Therefore, by Cauchy-Schwarz's inequality

$$\|\nabla(\mathcal{R}_h v)\|_{L^2(D)} \leq \|\nabla v\|_{L^2(D)}.$$

So

$$\|v - \tilde{v}\|_{L^2(\Gamma; H_0^1(D))} \leq \|v - \mathcal{R}_h v\|_{L^2(\Gamma; H_0^1(D))} + C \|v - \Pi_k^q v\|_{L^2(\Gamma; H_0^1(D))}.$$

By (4.20), we have

$$\|v - \tilde{v}\|_{L^2(\Gamma; H_0^1(D))} \leq C h^{\min(r+1, s)-1} \|v\|_{L^2(\Gamma; H^s(D))} + \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} v\|_{L^2(\Gamma; H_0^1(D))}}{(q_n + 1)!}.$$

This implies the inequality (4.17).

Next, we prove the inequality (4.18).

As above, we can write:

$$\begin{aligned}\|\nabla^2(v - \tilde{v})\|_{L^2(\Gamma; L^2(D))} &= \|\nabla^2(v - \Pi_k^q(\mathcal{R}_h v))\|_{L^2(\Gamma; L^2(D))} \\ &\leq \|\nabla^2(v - \mathcal{R}_h v)\|_{L^2(\Gamma; L^2(D))} + \|\nabla^2(\mathcal{R}_h v - \mathcal{R}_h(\Pi_k^q v))\|_{L^2(\Gamma; L^2(D))}.\end{aligned}$$

Since $\mathcal{R}_h v$ is a polynomial, by the inverse inequality, we obtain

$$\begin{aligned}\|\nabla^2(\mathcal{R}_h v)\|_{L^2(D)} &\leq Ch^{-1}\|\nabla(\mathcal{R}_h v)\|_{L^2(D)} \\ &\leq Ch^{-1}\|\nabla v\|_{L^2(D)}.\end{aligned}$$

By the Theorem 2.6 in [26] and (4.20), we have

$$\begin{aligned}\|\nabla^2(v - \tilde{v})\|_{L^2(\Gamma; L^2(D))} &\leq \|\nabla^2(v - \mathcal{R}_h v)\|_{L^2(\Gamma; L^2(D))} + Ch^{-1}\|\nabla(v - \Pi_k^q v)\|_{L^2(\Gamma; L^2(D))} \\ &\leq C_1 h^{\min(r+1, s)-2} \|v\|_{L^2(\Gamma; H^s(D))} \\ &\quad + C_2 h^{-1} \sum_{n=1}^N \left(\frac{k_n}{2}\right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} v\|_{L^2(\Gamma; H_0^1(D))}}{(q_n+1)!}.\end{aligned}$$

Therefore, we obtain the conclusion. $\square$

In order to approximate the error in the energy norm, I first consider $\|u_{kh} - \tilde{u}\|_{\mathcal{E}}$. By the coercivity of the bilinear form (4.14) and the error orthogonality, we have

$$\begin{aligned}\alpha \|u_{kh} - \tilde{u}\|_{\mathcal{E}}^2 &\leq \mathcal{A}_\epsilon(u_{kh} - \tilde{u}, u_{kh} - \tilde{u}) + \underbrace{\mathcal{A}_\epsilon(u - u_{kh}, u_{kh} - \tilde{u})}_{=0} \\ &= \mathcal{A}_\epsilon(u - \tilde{u}, \underbrace{u_{kh} - \tilde{u}}_{\chi}) \\ &= \int_{\Gamma} \rho \left(\sum_{E \in \mathcal{E}_h} \int_E a \nabla(u - \tilde{u}) \nabla \chi - \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla(u - \tilde{u}) \cdot n_e\} [\chi] \right. \\ &\quad \left. + \epsilon \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla \chi \cdot n_e\} [u - \tilde{u}] + \sum_{e \in \Gamma_h \cup \partial D} \frac{\sigma_e}{|e|^\beta} \int_e [u - \tilde{u}] [\chi] \right) dy \\ &\leq |T_1 + T_2 + T_3 + T_4|.\end{aligned}$$

Using the bound (2.12), Cauchy-Schwarz's inequality and Young's inequality, we have

$$\begin{aligned}
|T_1| &\leq \int_{\Gamma} \rho a_{max}^{\frac{1}{2}} \left(\sum_{E \in \mathcal{E}_h} \|a^{\frac{1}{2}} \nabla \chi\|_{L^2(E)}^2 \right)^{\frac{1}{2}} \left(\sum_{E \in \mathcal{E}_h} \|\nabla(u - \tilde{u})\|_{L^2(E)}^2 \right)^{\frac{1}{2}} dy \\
&\leq \int_{\Gamma} \rho \left(\frac{3}{2\alpha} a_{max} \sum_{E \in \mathcal{E}_h} \|\nabla(u - \tilde{u})\|_{L^2(E)}^2 + \frac{\alpha}{6} \|\chi\|_{DG}^2 \right) dy \\
&\leq \frac{3}{2\alpha} a_{max} \int_{\Gamma} \rho \sum_{E \in \mathcal{E}_h} \|\nabla(u - \tilde{u})\|_{L^2(E)}^2 dy + \frac{\alpha}{6} \int_{\Gamma} \rho \|\chi\|_{DG}^2 dy \\
&\leq C \sum_{E \in \mathcal{E}_h} \|u - \tilde{u}\|_{L^2(\Gamma; H_0^1(E))}^2 + \frac{\alpha}{6} \|\chi\|_{\mathcal{E}}^2 \\
&= C \|u - \tilde{u}\|_{L^2(\Gamma; H_0^1(D))}^2 + \frac{\alpha}{6} \|\chi\|_{\mathcal{E}}^2.
\end{aligned}$$

By Theorem 4.3, I obtain

$$\begin{aligned}
|T_1| &\leq C \left(h^{\min(r+1, s)-1} \|u\|_{L^2(\Gamma; H^s(D))} \right. \\
&\quad \left. + \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} u\|_{L^2(\Gamma; H_0^1(D))}}{(q_n+1)!} \right)^2 + \frac{\alpha}{6} \|\chi\|_{\mathcal{E}}^2.
\end{aligned} \tag{4.21}$$

Let us now bound T_2 . I assume that $\sigma_e > 0$ for all e . A given edge e is shared by two elements E_1^e and E_2^e . By Cauchy-Schwarz's inequality, Young's inequality and trace inequality (2.4), we have

$$\begin{aligned}
|T_2| &\leq \int_{\Gamma} \rho \sum_{e \in \Gamma_h \cup \partial D} \left(\left(\frac{|e|^\beta}{\sigma_e} \right)^{\frac{1}{2}} \|\{a \nabla(u - \tilde{u}) \cdot n_e\}\|_{L^2(e)} \left(\frac{\sigma_e}{|e|^\beta} \right)^{\frac{1}{2}} \|\chi\|_{L^2(e)} \right) dy \\
&\leq \int_{\Gamma} \rho \left(\frac{\alpha}{6} J_0(\chi, \chi) + C \sum_{e \in \Gamma_h \cup \partial D} \frac{|e|^\beta}{\sigma_e} \|\{a \nabla(u - \tilde{u}) \cdot n_e\}\|_{L^2(e)}^2 \right) dy \\
&\leq \frac{\alpha}{6} \|\chi\|_{\mathcal{E}}^2 + C \int_{\Gamma} \rho \sum_{e \in \Gamma_h \cup \partial D} \frac{|e|^\beta}{\sigma_e} |e| |E_1^e|^{-1} \left(\|\nabla(u - \tilde{u})\|_{L^2(E_1^e)} + h \|\nabla^2(u - \tilde{u})\|_{L^2(E_1^e)} \right)^2 dy \\
&\quad + C \int_{\Gamma} \rho \sum_{e \in \Gamma_h \cup \partial D} \frac{|e|^\beta}{\sigma_e} |e| |E_2^e|^{-1} \left(\|\nabla(u - \tilde{u})\|_{L^2(E_2^e)} + h \|\nabla^2(u - \tilde{u})\|_{L^2(E_2^e)} \right)^2 dy.
\end{aligned}$$

Assume $\beta \geq (d-1)^{-1}$. Theorem 4.3 implies that

$$|T_2| \leq C \left(h^{\min(r+1,s)-1} \|u\|_{L^2(\Gamma; H^s(D))} + \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} u\|_{L^2(\Gamma; H_0^1(D))}}{(q_n+1)!} \right)^2 + \frac{\alpha}{6} \|\chi\|_{\mathcal{E}}^2. \quad (4.22)$$

Assume $\tilde{u}$ is continuous and $\tilde{u} = 0$ on ∂D , then T_4, T_3 vanish. Combining (4.21) and (4.22), we have the following result:

$$\|u_{kh} - \tilde{u}\|_{\mathcal{E}} \leq C \left(h^{\min(r+1,s)-1} \|u\|_{L^2(\Gamma; H^s(D))} + \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} u\|_{L^2(\Gamma; H_0^1(D))}}{(q_n+1)!} \right). \quad (4.23)$$

where C is a constant dependent on a_{max} but independent of u, h and k .

In the following, I discuss $\|u - \tilde{u}\|_{\mathcal{E}}$. By the definition of energy norm and assuming that $\tilde{u}$ is continuous, we have:

$$\begin{aligned} \|u - \tilde{u}\|_{\mathcal{E}}^2 &= \int_{\Gamma} \rho \|u - \tilde{u}\|_{DG}^2 dy = \int_{\Gamma} \rho \left(\sum_{E \in \mathcal{E}_h} \int_E a \nabla(u - \tilde{u})^2 + 0 \right) dy \\ &= \int_{\Gamma} \rho \int_D a \nabla(u - \tilde{u})^2 dy = \|u - \tilde{u}\|_{L^2(\Gamma; H_0^1(D))}^2. \end{aligned} \quad (4.24)$$

Using Theorem 4.3, we conclude the following theorem:

Theorem 4.4

Assume $u \in L^2(\Gamma; H^s(D)) \cap H^{q+1}(\Gamma; H_0^1(D))$, then there is a constant C independent of h, k_n and u , s.t.

$$\|u - u_{kh}\|_{\mathcal{E}} \leq C \left(h^{\min(r+1,s)-1} \|u\|_{L^2(\Gamma; H^s(D))} + \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} u\|_{L^2(\Gamma; H_0^1(D))}}{(q_n+1)!} \right). \quad (4.25)$$

where k_n, q_n are defined in Section 3.2.

4.5 Error Estimate in the L^2 Norm

In this section, I prove an L^2 error estimate for stochastic discontinuous Galerkin methods in the symmetric case, i.e., SIPG case. The method I use is the Aubin-Nitsche lift technique.

I first give an error estimate for the computation of the expected value of the solution in DG norm. Let $\tilde{\chi} = u - u_{kh}$. By the definition of the DG norm, we have

$$\begin{aligned}
\|E[\tilde{\chi}]\|_{DG}^2 &= \sum_{E \in \mathcal{E}_h} \int_E a \nabla E(\tilde{\chi}(z, \cdot)) \nabla E(\tilde{\chi}(y, \cdot)) + \sum_{e \in \Gamma_h \cup \partial D} \frac{\sigma_e}{|e|^\beta} \int_e [E(\tilde{\chi}(z, \cdot))] [E(\tilde{\chi}(y, \cdot))] \\
&= \int_\Gamma \rho \sum_{E \in \mathcal{E}_h} \int_E a \nabla E(\tilde{\chi}(z, \cdot)) \nabla \tilde{\chi}(y, \cdot) dy + \int_\Gamma \rho \sum_{e \in \Gamma_h \cup \partial D} \frac{\sigma_e}{|e|^\beta} \int_e [E(\tilde{\chi}(z, \cdot))] [\tilde{\chi}(y, \cdot)] dy \\
&= \int_\Gamma \rho \left(\sum_{E \in \mathcal{E}_h} \int_E a \nabla E(\tilde{\chi}(z, \cdot)) \nabla \tilde{\chi}(y, \cdot) + \sum_{e \in \Gamma_h \cup \partial D} \frac{\sigma_e}{|e|^\beta} \int_e [E(\tilde{\chi}(z, \cdot))] [\tilde{\chi}(y, \cdot)] \right) dy \\
&= \int_\Gamma \rho \left(\int_\Gamma \rho \left(\sum_{E \in \mathcal{E}_h} \int_E a \nabla \tilde{\chi}(z, \cdot) \nabla \tilde{\chi}(y, \cdot) + \sum_{e \in \Gamma_h \cup \partial D} \frac{\sigma_e}{|e|^\beta} \int_e [\tilde{\chi}(z, \cdot)] [\tilde{\chi}(y, \cdot)] \right) dz \right) dy \\
&\leq \int_\Gamma \rho \left(\int_\Gamma \rho \|\tilde{\chi}(z, \cdot)\|_{DG} \|\tilde{\chi}(y, \cdot)\|_{DG} dz \right) dy \\
&= \int_\Gamma \rho \|\tilde{\chi}(y, \cdot)\|_{DG} dy \int_\Gamma \rho \|\tilde{\chi}(z, \cdot)\|_{DG} dz \\
&\leq \left(\int_\Gamma \rho dy \right)^{\frac{1}{2}} \left(\int_\Gamma \rho \|\tilde{\chi}(y, \cdot)\|_{DG}^2 dy \right)^{\frac{1}{2}} \left(\int_\Gamma \rho dy \right)^{\frac{1}{2}} \left(\int_\Gamma \rho \|\tilde{\chi}(z, \cdot)\|_{DG}^2 dz \right)^{\frac{1}{2}} \\
&\leq \left(\int_\Gamma \rho \|\tilde{\chi}(y, \cdot)\|_{DG}^2 dy \right) \\
&= C \|u - u_{kh}\|_{\mathcal{E}}^2.
\end{aligned} \tag{4.26}$$

This implies that

$$\|E[u - u_{kh}]\|_{DG} \leq C \|u - u_{kh}\|_{\mathcal{E}}.$$

Thus,

$$\begin{aligned}
\|E[u - u_{kh}]\|_{DG} &\leq C \left(h^{\min(r+1, s)-1} \|u\|_{L^2(\Gamma; H^s(D))} + \right. \\
&\quad \left. \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} u\|_{L^2(\Gamma; H_0^1(D))}}{(q_n + 1)!} \right).
\end{aligned} \tag{4.27}$$

I assume that the auxiliary function $\hat{u}$ is the solution to the dual problem

$$\begin{aligned} -\nabla \cdot (a \nabla \hat{u}) &= E[\tilde{\chi}] \quad \text{in } D, \\ \hat{u} &= 0 \quad \text{on } \partial D. \end{aligned}$$

This implies that

$$\|\hat{u}(z, \cdot)\|_{H^2(D)} \leq C \|E[u - u_{kh}]\|_{L^2(D)} \quad \forall z \in \Gamma. \quad (4.28)$$

Then, we have

$$\|E[\tilde{\chi}]\|_{L^2(D)}^2 = \int_D E[\tilde{\chi}]^2 dx = \int_D (-\nabla \cdot (a \nabla \hat{u})) E[\tilde{\chi}].$$

Integrating by parts on each element yields

$$\begin{aligned} \|E[\tilde{\chi}]\|_{L^2(D)}^2 &= \sum_{E \in \mathcal{E}_h} \int_E a \nabla \hat{u} \cdot \nabla E[\tilde{\chi}] - \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla \hat{u} \cdot n_e\} [E[\tilde{\chi}]] \\ &= \int_{\Gamma} \rho \left(\sum_{E \in \mathcal{E}_h} \int_E a \nabla \hat{u}(z) \cdot \nabla \tilde{\chi}(y) - \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla \hat{u} \cdot n_e\} [\tilde{\chi}] \right) dy. \end{aligned} \quad (4.29)$$

The error orthogonality implies that

$$\int_{\Gamma} \rho A_{\epsilon}(\tilde{\chi}, v) dy = 0 \quad v \in Y_k^q \otimes D_r(\mathcal{E}_h).$$

Subtracting this orthogonality equation from (4.29), we have

$$\begin{aligned} \|E[\tilde{\chi}]\|_{L^2(D)}^2 &= \int_{\Gamma} \rho \left(\sum_{E \in \mathcal{E}_h} \int_E a \nabla (\hat{u} - v) \cdot \nabla \tilde{\chi} - \epsilon \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla v \cdot n_e\} [\tilde{\chi}] \right. \\ &\quad \left. - \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla \hat{u} \cdot n_e\} [\tilde{\chi}] + \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla \tilde{\chi} \cdot n_e\} [v] - J_0^{\sigma, \beta}(\tilde{\chi}, v) \right) dy \\ &= A_1 + A_2 + A_3 + A_4 + A_5. \end{aligned} \quad (4.30)$$

Let $\tilde{u}$ be a good approximation of $\hat{u}$ that satisfies Theorem 4.3. We assume that $\tilde{u}$ is continuous and vanishes on ∂D . Choose $v = \tilde{u}$. We note that $\tilde{u} = u = 0$ on the

boundary ∂D . This implies that the last two terms on the right-hand of (4.30) vanish, then we have

$$\begin{aligned}
A_1 &\leq \int_{\Gamma} \rho \sum_{E \in \mathcal{E}_h} \left(\int_E a(\nabla \tilde{\chi})^2 dx \right)^{\frac{1}{2}} \left(\int_E a(\nabla(\hat{u} - \tilde{u}))^2 dx \right)^{\frac{1}{2}} dy \\
&\leq \int_{\Gamma} \rho \left(\sum_{E \in \mathcal{E}_h} \int_E a(\nabla \tilde{\chi})^2 dx \right)^{\frac{1}{2}} \left(\sum_{E \in \mathcal{E}_h} \int_E a(\nabla(\hat{u} - \tilde{u}))^2 dx \right)^{\frac{1}{2}} dy \\
&\leq \int_{\Gamma} \rho \|\tilde{\chi}\|_{DG} \|\hat{u} - \tilde{u}\|_{DG} dy \\
&\leq \left(\int_{\Gamma} \rho \|\tilde{\chi}\|_{DG}^2 dy \right)^{\frac{1}{2}} \left(\int_{\Gamma} \rho \|\hat{u} - \tilde{u}\|_{DG}^2 dy \right)^{\frac{1}{2}} \\
&= \|\tilde{\chi}\|_{\mathcal{E}} \|\hat{u} - \tilde{u}\|_{\mathcal{E}}.
\end{aligned} \tag{4.31}$$

Thus, we have

$$\|E[\tilde{\chi}]\|_{L^2(D)}^2 \leq \|\tilde{\chi}\|_{\mathcal{E}} \|\hat{u} - \tilde{u}\|_{\mathcal{E}} + S,$$

where

$$S = \left| \int_{\Gamma} \rho \left(\sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla(\hat{u} + \epsilon \tilde{u}) \cdot n_e\} [\tilde{\chi}] \right) dy \right|.$$

If the method of SIPG is employed, the term $\hat{u} + \epsilon \tilde{u} = \hat{u} - \tilde{u}$ is the approximation error. Then, we take advantage of the penalty parameters to derive that

$$\begin{aligned}
S &\leq \int_{\Gamma} \rho \left(\sum_{e \in \Gamma_h \cup \partial D} \left(\frac{|e|^\beta}{\sigma_e} \right)^{\frac{1}{2} - \frac{1}{2}} \|\{a \nabla(\hat{u} - \tilde{u}) \cdot n_e\}\|_{L^2(e)} \|\tilde{\chi}\|_{L^2(e)} \right) dy \\
&\leq \int_{\Gamma} \rho \left(\sum_{e \in \Gamma_h \cup \partial D} \frac{\sigma_e}{|e|^\beta} \|\tilde{\chi}\|_{L^2(e)}^2 \right)^{\frac{1}{2}} \left(\sum_{e \in \Gamma_h \cup \partial D} \|\{a \nabla(\hat{u} - \tilde{u}) \cdot n_e\}\|_{L^2(e)}^2 \cdot \frac{|e|^\beta}{\sigma_e} \right)^{\frac{1}{2}} dy \\
&\leq \int_{\Gamma} \rho J_0^{\sigma, \beta}(\tilde{\chi}, \tilde{\chi})^{\frac{1}{2}} \left(\sum_{e \in \Gamma_h \cup \partial D} \|\{a \nabla(\hat{u} - \tilde{u}) \cdot n_e\}\|_{L^2(e)}^2 \cdot \frac{|e|^\beta}{\sigma_e} \right)^{\frac{1}{2}} dy \\
&\leq \left(\int_{\Gamma} \rho J_0^{\sigma, \beta}(\tilde{\chi}, \tilde{\chi}) dy \right)^{\frac{1}{2}} \left(\int_{\Gamma} \rho \sum_{e \in \Gamma_h \cup \partial D} \|\{a \nabla(\hat{u} - \tilde{u}) \cdot n_e\}\|_{L^2(e)}^2 \cdot \frac{|e|^\beta}{\sigma_e} dy \right)^{\frac{1}{2}} \\
&\leq \|\tilde{\chi}\|_{\mathcal{E}} \left(\int_{\Gamma} \rho \sum_{e \in \Gamma_h \cup \partial D} \|\{a \nabla(\hat{u} - \tilde{u}) \cdot n_e\}\|_{L^2(e)}^2 \cdot \frac{|e|^\beta}{\sigma_e} dy \right)^{\frac{1}{2}}.
\end{aligned} \tag{4.32}$$

Now, I consider the term $\|\{a \nabla(\hat{u} - \tilde{u}) \cdot n_e\}\|_{L^2(e)}^2$. By the definition of the average and the trace inequality (2.4), we have

$$\begin{aligned}
\|\{a \nabla(\hat{u} - \tilde{u}) \cdot n_e\}\|_{L^2(e)} &\leq C \left(h^{-1/2} \|\nabla(\hat{u} - \tilde{u})\|_{L^2(E_1^e)} + h^{1/2} \|\nabla^2(\hat{u} - \tilde{u})\|_{L^2(E_1^e)} \right) \\
&\quad + C \left(h^{-1/2} \|\nabla(\hat{u} - \tilde{u})\|_{L^2(E_2^e)} + h^{1/2} \|\nabla^2(\hat{u} - \tilde{u})\|_{L^2(E_2^e)} \right).
\end{aligned} \tag{4.33}$$

Assume $\beta \geq (d-1)^{-1}$. From (4.32) and (4.33), we obtain

$$S \leq C \|\tilde{\chi}\|_{\mathcal{E}} \left(\|\nabla(\hat{u} - \tilde{u})\|_{L^2(\Gamma; L^2(D))} + h \|\nabla^2(\hat{u} - \tilde{u})\|_{L^2(\Gamma; L^2(D))} \right).$$

Since $h < 1$, we have

$$\begin{aligned}
\|E[\tilde{\chi}]\|_{L^2(D)}^2 &\leq \|\tilde{\chi}\|_{\mathcal{E}} \|\hat{u} - \tilde{u}\|_{\mathcal{E}} \\
&\quad + C \|\tilde{\chi}\|_{\mathcal{E}} \left(h \|\hat{u}\|_{L^2(\Gamma; H^2(D))} + \sum_{n=1}^N \left(\frac{k_n}{2} \right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1} u\|_{L^2(\Gamma; H_0^1(D))}}{(q_n+1)!} \right)
\end{aligned} \tag{4.34}$$

Using (4.17) and (4.24), we obtain

$$\begin{aligned} \|E[\tilde{\chi}]\|_{L^2(D)}^2 &\leq C\|\tilde{\chi}\|_{\mathcal{E}} \left(h\|\hat{u}\|_{L^2(\Gamma;H^2(D))} + \right. \\ &\quad \left. \sum_{n=1}^N \left(\frac{k_n}{2}\right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1}\hat{u}\|_{L^2(\Gamma;H_0^1(D))}}{(q_n+1)!} \right). \end{aligned}$$

Recall the result in (3.4)

$$\frac{\|\partial_{y_n}^{q_n+1}\hat{u}\|_{L^2(\Gamma;H_0^1(D))}}{(q_n+1)!} \leq (r_n)^{q_n+1} \frac{C}{a_{min}} \|E[\tilde{\chi}]\|_{L^2(D)}, \quad (4.35)$$

and (4.28), we have

$$\|E[\tilde{\chi}]\|_{L^2(D)} \leq C\|\tilde{\chi}\|_{\mathcal{E}} \left(h + \sum_{n=1}^N \left(\frac{k_n r_n}{2}\right)^{q_n+1} \right). \quad (4.36)$$

We summarize our result in the following theorem:

Theorem 4.5

Assume $u \in L^2(\Gamma; H^s(D)) \cap H^{q+1}(\Gamma; H_0^1(D))$, then there is a constant C independent of h, k_n and u s.t.

$$\begin{aligned} \|E[u - u_{kh}]\|_{L^2(D)} &\leq C \left(h^{\min(r+1,s)-1} \|u\|_{L^2(\Gamma;H^s(D))} + \right. \\ &\quad \left. \sum_{n=1}^N \left(\frac{k_n}{2}\right)^{q_n+1} \frac{\|\partial_{y_n}^{q_n+1}u\|_{L^2(\Gamma;H_0^1(D))}}{(q_n+1)!} \right) \left(h + \sum_{n=1}^N \left(\frac{k_n r_n}{2}\right)^{q_n+1} \right). \end{aligned} \quad (4.37)$$

where k_n, y_n, q_n, r_n are defined in Section 3.2.

Assume $s = 2, r \geq 1$, combining with (4.25) and (4.37), we find that the L^2 error is of the order $h^2 + \left(\sum_{n=1}^N (k_n)^{q_n+1}\right)^2$.

4.6 $k \times h$ -version SFEM versus $k \times h$ -version SDG

In this section, I compare the H^1 and L^2 error estimates and the computational work for the $k \times h$ -version SFEM recalled in Chapter 3 with the $k \times h$ -version SDG method

introduced in Chapter 4. Recall the result in Chapter 8 in [10], the computational work of $k \times h$ -version SFEM is:

$$Work_{k \times h-SFEM} = \frac{(1+q)^N}{h^d k^N}. \quad (4.38)$$

For the $k \times h$ -version SDG method, the global dimension $\mathcal{N}$ of $\mathcal{D}_r(\mathcal{E}_h)$ is proportional to the product of the number of triangular elements and the square of the polynomial degree r . Consider the d dimensional space domain, the number of triangular elements is proportional to $\frac{1}{h^d}$, where h is the mesh size. Thus, we obtain

$$\mathcal{N} \propto \frac{r^2}{h^d}. \quad (4.39)$$

Now, I consider the outcome set $\Gamma = \prod_{n=1}^n \Gamma_n$. In each direction y_n , since the degree of the discontinuous piecewise polynomial is at most q_n , the number of unknowns in this direction is:

$$(q_n + 1)(N_n + 1) - 2, \quad (4.40)$$

where N_n is the number of intervals in the direction y_n . The leading term in (4.40) is $(q_n + 1)N_n$. Since N_n is proportional to $\frac{1}{k_n}$, where k_n is the mesh size in the direction y_n , and the dimension of the outcome set is N , thus the computational work of $k \times h$ -version of SDG is

$$Work_{k \times h-SDG} = \prod_{n=1}^N \left(\frac{q_n + 1}{k_n} \right) \frac{r^2}{h^d}. \quad (4.41)$$

Table 4.1 gives the computational work for SFEM and SDG. It also contains the theoretical convergence for the case of $u \in L_P^2(\Gamma; H^2(D))$.

Table 4.1 : $k \times h$ -version SFEM versus $k \times h$ -version SDG

	$k \times h$ -version SFEM	$k \times h$ -version SDG
Work	$\frac{1}{h^d} \prod_{n=1}^N \left(\frac{q_n+1}{k_n} \right)$	$\frac{r^2}{h^d} \prod_{n=1}^N \left(\frac{q_n+1}{k_n} \right)$
$H_0^1(D)$ Error	$h + \sum_{n=1}^N (k_n)^{q_n+1}$	$h + \sum_{n=1}^N (k_n)^{q_n+1}$
$L^2(D)$ Error	$h^2 + \left(\sum_{n=1}^N (k_n)^{q_n+1} \right)^2$	$h^2 + \left(\sum_{n=1}^N (k_n)^{q_n+1} \right)^2$

Chapter 5

The Monte Carlo Discontinuous Galerkin Method

5.1 Introduction

In this chapter, I define the Monte Carlo discontinuous Galerkin Method (MCDG) and derive a priori error estimates for the expected value of the solution.

The MCDG method consists of the following steps:

- Choose a number of realizations, $M \in \mathbb{N}_+$, and a DG space $D_r(\mathcal{E}_h)$.
- For each $j = 1, \dots, M$, sample independent, identically distributed (iid) realizations of the diffusion $a(y_j, \cdot)$, the load $f(y_j, \cdot)$ and find an approximation $u_h(y_j, \cdot) \in D_r(\mathcal{E}_h)$ such that

$$A_\epsilon(y_j; u_h(y_j, \cdot), v) = (f(y_j, \cdot), v)_{L^2(D)} \quad \forall v \in D_r(\mathcal{E}_h), \quad (5.1)$$

where A_ϵ is defined in (4.13).

- Approximate $E[u(\cdot, \cdot)]$ by the sample average:

$$B(u_h; M) = \frac{1}{M} \sum_{j=1}^M u_h(y_j; \cdot). \quad (5.2)$$

Here, I assume that $D_r(\mathcal{E}_h)$ is same for all realizations; i.e., the triangulation of D is deterministic. The computational error is naturally split into two parts:

$$E[u] - B(u_h; M) = (E[u] - E[u_h]) + (E[u_h] - B(u_h; M)) \equiv \eta_h + \eta_S.$$

The space discretization error η_h is controlled by the size of the spatial triangulation, while the statistical error η_S is controlled by the number of realizations.

5.2 Statistical Error

In order to study the properties of the statistical error, I first consider the random variable $\|\eta_S\|$.

Lemma 5.1

Let $\langle \cdot, \cdot \rangle$ be an inner product for the space $D_r(\mathcal{E}_h)$, then $E[\langle u_h(y_i, \cdot) - E[u_h], u_h(y_j, \cdot) - E[u_h] \rangle] = 0$, for $i \neq j$.

Proof 5.1 For simplicity, I denote $u_h(\omega_j, \cdot)$ by u_j , $E[u_h]$ by v . Let $\{\phi_n\}_{n=1}^Q$ be a basis for $\mathcal{D}_r(\mathcal{E}_h)$. Since $u_i - v = \sum_{n=1}^Q \langle u_i - v, \phi_n \rangle \phi_n$, we have

$$\begin{aligned} E[\langle u_i - v, u_j - v \rangle] &= E\left[\sum_{n=1}^Q \sum_{m=1}^Q \langle u_i - v, \phi_n \rangle \langle u_j - v, \phi_m \rangle\right] \\ &= \sum_{n=1}^Q \sum_{m=1}^Q E[\langle u_i - v, \phi_n \rangle \langle u_j - v, \phi_m \rangle] \\ &= \sum_{n=1}^Q \sum_{m=1}^Q \langle E[u_i] - v, \phi_n \rangle \langle E[u_j] - v, \phi_m \rangle \\ &= 0. \end{aligned}$$

Thus, we obtain the conclusion. □

Theorem 5.1

We define the inner product $\langle \cdot, \cdot \rangle_{DG}$ by

$$\langle f, g \rangle_{DG} = \sum_{E \in \mathcal{E}_h} (a \nabla f, \nabla g)_{L^2(E)} + \sum_{e \in \Gamma_h \cup \partial D} \frac{\sigma_e}{|e|^\beta} ([f], [g])_{L^2(e)}, \quad (5.3)$$

then we have

$$ME[\|\eta_S\|_{DG}^2] = E[\|u_h\|_{DG}^2] - \|E[u_h]\|_{DG}^2.$$

Proof 5.2 We use Lemma 5.1 with the inner product (5.3). we have

$$\begin{aligned}
E[\|\eta_S\|_{DG}^2] &= E[\|v - B(u_h; M)\|_{DG}^2] \\
&= E[\|\frac{1}{M} \sum_{j=1}^M (u_j - v)\|_{DG}^2] \\
&= \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M E[\langle u_i - v, u_j - v \rangle] \\
&= \frac{1}{M^2} \sum_{i=1}^M E[\|u_i - v\|_{DG}^2] = \frac{1}{M} E[\|u_i - v\|_{DG}^2] \\
&= \frac{1}{M} (E[\|u_i\|_{DG}^2] - 2E[\langle u_i, v \rangle] + \|v\|_{DG}^2) \\
&= \frac{1}{M} (E[\|u_i\|_{DG}^2] - \|E[u_i]\|_{DG}^2).
\end{aligned}$$

That is the conclusion. $\square$

Theorem 5.2

Let $D_r(\mathcal{E}_h)$ be a choice of DG space in the MCDG. Also, suppose that there exists a constant $C > 0$ independent of M and h such that the statistical error in DG norm satisfies

$$ME[\|\eta_S\|_{DG}^2] < C. \quad (5.4)$$

Taking an increasing sequence of number of realizations $\{M_k\}_{k=1}^\infty$ which is the subset of $\{2^k : k \in \mathbb{N}_+\}$, then for any $\alpha \in (0, \frac{1}{2})$ and any choice of mesh size h , we have

$$\lim_{k \rightarrow \infty} M_k^\alpha \|\eta_S\|_{DG} = 0 \quad a.s.$$

and for any $\epsilon > 0$ we have

$$P\left(\|\eta_S\|_{DG} > \frac{\epsilon}{\sqrt{M}}\right) \leq \frac{C}{\epsilon^2}.$$

Proof 5.3 Markov's inequality [27] and (5.4) imply that

$$P((M_k)^\alpha \|\eta_S\|_{DG} > \epsilon) = P((M_k)^{2\alpha} \|\eta_S\|_{DG}^2 > \epsilon^2) \leq \frac{E[(M_k)^{2\alpha} \|\eta_S\|_{DG}^2]}{\epsilon^2} \leq \frac{C}{\epsilon^2 (M_k)^{1-2\alpha}}.$$

Since $\alpha \in (0, \frac{1}{2})$, we have

$$\sum_{k=1}^{\infty} P((M_k)^\alpha \|\eta_S\|_{DG} > \epsilon) \leq \frac{C}{\epsilon^2} \sum_{k=1}^{\infty} \frac{1}{(M_k)^{1-2\alpha}} \leq \frac{C}{\epsilon^2} \sum_{k=1}^{\infty} \frac{1}{(2^{1-2\alpha})^k} < \infty.$$

Then, by Borel-Cantelli lemma [27], we obtain that

$$\lim_{k \rightarrow \infty} M_k^\alpha \|\eta_S\|_{DG} = 0 \quad a.s.$$

Still using Markov's inequality, we have

$$P\left(\|\eta_S\|_{DG} > \frac{\epsilon}{\sqrt{M}}\right) \leq \frac{ME[\|\eta_S\|_{DG}^2]}{\epsilon^2} \leq \frac{C}{\epsilon^2}.$$

5.3 Space Discretization Error

In this section, I discuss the space discretization error in the $L^2(D)$ and DG norms.

Lemma 5.2

If u is the solution of problem (2.25) and $u_h \in D_r(\mathcal{E}_h)$ is the approximation solution in MCDG defined in (5.1). Then we have

$$\|u(y, \cdot) - u_h(y, \cdot)\|_{DG} \leq Ch^{\min(r+1, s)-1} \|u(y, \cdot)\|_{H^s(D)} \quad a.s. \quad (5.5)$$

$$\|u(y, \cdot) - u_h(y, \cdot)\|_{L^2(D)} \leq Ch^{\min(r+1, s)} \|u(y, \cdot)\|_{H^s(D)} \quad a.s. \quad (5.6)$$

where C is independent of h and u .

Proof 5.4 We discuss $\|u(y, \cdot) - u_h(y, \cdot)\|_{DG}$. By Cauchy-Schwarz's inequality, we split the estimate into two parts:

$$\|u(y, \cdot) - u_h(y, \cdot)\|_{DG} \leq \|u(y, \cdot) - \tilde{u}(y, \cdot)\|_{DG} + \|u_h(y, \cdot) - \tilde{u}(y, \cdot)\|_{DG}.$$

where $\tilde{u}(y, \cdot) \in \mathbb{P}_r(D) \cap H_1^0(D)$ is a good approximation of $u(y, \cdot)$.

Then, we consider $\|u_h(y, \cdot) - \tilde{u}(y, \cdot)\|_{DG}$. By the coercivity of the bilinear form in DG

and the error orthogonality, we have

$$\begin{aligned}
& \alpha \|u_h(y, \cdot) - \tilde{u}(y, \cdot)\|_{DG} \\
& \leq A_\epsilon(u_h(y, \cdot) - \tilde{u}(y, \cdot), u_h(y, \cdot) - \tilde{u}(y, \cdot)) + \underbrace{A_\epsilon(u(y, \cdot) - u_h(y, \cdot), u_h(y, \cdot) - \tilde{u}(y, \cdot))}_{=0} \\
& = A_\epsilon(u(y, \cdot) - \tilde{u}(y, \cdot), \underbrace{u_h(y, \cdot) - \tilde{u}(y, \cdot)}_\chi) \\
& = \sum_{E \in \mathcal{E}_h} \int_E a \nabla(u(y, \cdot) - \tilde{u}(y, \cdot)) \nabla \chi - \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla(u(y, \cdot) - \tilde{u}(y, \cdot)) \cdot n_e\} [\chi] \\
& \quad + \epsilon \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla \chi \cdot n_e\} [u(y, \cdot) - \tilde{u}(y, \cdot)] + \sum_{e \in \Gamma_h \cup \partial D} \frac{\sigma_e}{|e|^\beta} \int_e [u(y, \cdot) - \tilde{u}(y, \cdot)] [\chi] \\
& \leq |\tilde{T}_1 + \tilde{T}_2 + \tilde{T}_3 + \tilde{T}_4|,
\end{aligned}$$

where

$$\begin{aligned}
|\tilde{T}_1| & \leq a_{max}^{\frac{1}{2}} \left(\sum_{E \in \mathcal{E}_h} \|a^{\frac{1}{2}} \nabla \chi\|_{L^2(E)}^2 \right)^{\frac{1}{2}} \left(\sum_{E \in \mathcal{E}_h} \|\nabla(u(y, \cdot) - \tilde{u}(y, \cdot))\|_{L^2(E)}^2 \right)^{\frac{1}{2}} \\
& \leq \frac{3}{2\alpha} a_{max} \sum_{E \in \mathcal{E}_h} \|\nabla(u(y, \cdot) - \tilde{u}(y, \cdot))\|_{L^2(E)}^2 + \frac{\alpha}{6} \|\chi\|_{DG}^2 \\
& \leq C \sum_{E \in \mathcal{E}_h} \|u(y, \cdot) - \tilde{u}(y, \cdot)\|_{H_0^1(E)}^2 + \frac{\alpha}{6} \|\chi\|_{DG}^2 \\
& = C \|u(y, \cdot) - \tilde{u}(y, \cdot)\|_{H_0^1(D)}^2 + \frac{\alpha}{6} \|\chi\|_{DG}^2 \\
& \leq Ch^{2\min(r+1, s)-2} \|u(y, \cdot)\|_{H^s(D)} + \frac{\alpha}{6} \|\chi\|_{DG}^2
\end{aligned}$$

With the similar methods used in section 4.4, we obtain the similar results for $\tilde{T}_2, \tilde{T}_3$ and $\tilde{T}_4$. Hence, we have

$$\|u_h(y, \cdot) - \tilde{u}(y, \cdot)\|_{DG} \leq Ch^{\min(r+1, s)-1} \|u(y, \cdot)\|_{H^s(D)}. \quad (5.7)$$

Next, we discuss $\|u(y, \cdot) - \tilde{u}(y, \cdot)\|_{DG}$. We assume $\tilde{u}(y, \cdot)$ is continuous. By the definition of DG norm, we have

$$\begin{aligned}
\|u(y, \cdot) - \tilde{u}(y, \cdot)\|_{DG}^2 & = \sum_{E \in \mathcal{E}_h} \int_E a \nabla(u(y, \cdot) - \tilde{u}(y, \cdot))^2 + 0 \\
& = \int_D a \nabla(u(y, \cdot) - \tilde{u}(y, \cdot))^2 \leq C \|u(y, \cdot) - \tilde{u}(y, \cdot)\|_{H_0^1(D)}^2.
\end{aligned}$$

where C is a constant independent of u .

This implies that

$$\|u(y, \cdot) - \tilde{u}(y, \cdot)\|_{DG} \leq C \|u(y, \cdot) - \tilde{u}(y, \cdot)\|_{H_0^1(D)}. \quad (5.8)$$

Combining (5.7) and (5.8), we obtain that

$$\begin{aligned} \|u(y, \cdot) - u_h(y, \cdot)\|_{DG} &\leq \|u(y, \cdot) - \tilde{u}(y, \cdot)\|_{DG} + \|u_h(y, \cdot) - \tilde{u}(y, \cdot)\|_{DG} \\ &\leq Ch^{\min(r+1, s)-1} \|u(y, \cdot)\|_{H^s(D)} \quad a.s. \end{aligned}$$

To prove (5.6), we apply the Aubin-Nitsche lift technique. Construct the dual problem with the auxiliary function ϕ :

$$\begin{aligned} -\nabla \cdot (a \nabla \phi) &= u(y, \cdot) - u_h(y, \cdot) \quad \text{in } D, \\ \phi &= 0 \quad \text{on } \partial D. \end{aligned}$$

We assume the solution to this problem belongs to $H^2(D)$ and satisfies:

$$\|\phi\|_{H^2(D)} \leq C \|u(y, \cdot) - u_h(y, \cdot)\|_{L^2(D)}. \quad (5.9)$$

Denote $\theta = u(y, \cdot) - u_h(y, \cdot)$. Then, we have

$$\|\theta\|_{L^2(D)}^2 = \int_D \theta^2 dx = \int_D (-\nabla \cdot (a \nabla \phi)) \theta.$$

Integrating by parts on each element yields

$$\|\theta\|_{L^2(D)}^2 = \sum_{E \in \mathcal{E}_h} \int_E a \nabla \phi \cdot \nabla \theta - \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla \phi \cdot n_e\} [\theta]. \quad (5.10)$$

Subtracting the error orthogonality

$$A_\epsilon(\theta, v) = 0 \quad \forall v \in \mathcal{D}_r(\mathcal{E}_h),$$

from (5.10), we have

$$\begin{aligned} \|\theta\|_{L^2(D)}^2 &= \sum_{E \in \mathcal{E}_h} \int_E a \nabla(\phi - v) \nabla \theta - \epsilon \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla v \cdot n_e\}[\theta] - \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla \phi \cdot n_e\}[\theta] \\ &\quad + \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla \theta \cdot n_e\}[v] - J_0^{\sigma, \beta}(\theta, v). \end{aligned}$$

Choose $v = \tilde{\phi}$, a continuous interpolant of u of degree r . We assume such an interpolant exists. In this case, we note that $\tilde{\phi} = \phi = 0$ on the boundary ∂D . The last two terms on the right-hand vanish. Then for the first term:

$$A_1 \leq Ch \|\phi\|_{H^2(D)} \|\theta\|_{DG}.$$

Hence, we obtain that

$$\|\theta\|_{L^2(D)}^2 \leq Ch \|\phi\|_{H^2(D)} \|\theta\|_{DG} + S,$$

where

$$S = \left| \sum_{e \in \Gamma_h \cup \partial D} \int_e \{a \nabla(\phi + \epsilon \tilde{\phi}) \cdot n_e\}[\theta] \right|.$$

If SIPG is employed, then $\phi + \epsilon \tilde{\phi} = \phi - \tilde{\phi}$ is the approximation error. We take the advantage of the penalty parameters to derive that

$$\begin{aligned} S &\leq \sum_{e \in \Gamma_h \cup \partial D} \left(\frac{|e|^\beta}{\sigma_e} \right)^{\frac{1}{2} - \frac{1}{2}} \|\{a \nabla(\phi - \tilde{\phi}) \cdot n_e\}\|_{L^2(e)} \|\theta\|_{L^2(e)} \\ &\leq \left(\sum_{e \in \Gamma_h \cup \partial D} \frac{\sigma_e}{|e|^\beta} \|\theta\|_{L^2(e)}^2 \right)^{\frac{1}{2}} \left(\sum_{e \in \Gamma_h \cup \partial D} \|\{a \nabla(\phi - \tilde{\phi}) \cdot n_e\}\|_{L^2(e)}^2 \cdot \frac{|e|^\beta}{\sigma_e} \right)^{\frac{1}{2}} \\ &\leq C J_0^{\sigma, \beta}(\theta, \theta)^{\frac{1}{2}} \left(\sum_{e \in \Gamma_h \cup \partial D} \|\{a \nabla(\phi - \tilde{\phi}) \cdot n_e\}\|_{L^2(e)}^2 \cdot \frac{|e|^\beta}{\sigma_e} \right)^{\frac{1}{2}} \\ &\leq Ch^{\frac{1}{2}(d-1) + \frac{1}{2}} \|\phi\|_{H^2(D)} \|\theta\|_{DG}. \end{aligned}$$

Using (5.9), we obtain

$$\|\theta\|_{L^2(D)}^2 \leq C(h + h^{\frac{1}{2}(d-1) + \frac{1}{2}}) \|\theta\|_{DG} \|\theta\|_{L^2(D)}.$$

Table 5.1 : Computational work of MCFEM and MCDG

	MCFEM	MCDG
Work	$\frac{M}{h^d}$	$M \left(\frac{r^2}{h^d} \right)$

By (5.5), we obtain

$$\|u(y, \cdot) - u_h(y, \cdot)\|_{L^2(D)} \leq Ch^{\min(r+1, s)} \|u(y, \cdot)\|_{H^s(D)} \quad a.s.$$

Therefore, the conclusion is obtained. $\square$

Theorem 5.3

Let u be the solution of problem (2.25) and $u_h \in D_r(\mathcal{E}_h)$ be the approximation solution by MCDG defined in (5.1). Then we have

$$\|E[u] - E[u_h]\|_{DG} \leq Ch^{\min(r+1, s)-1} \|u\|_{L_P^2(\Gamma; H^s(D))}, \quad (5.11)$$

$$\|E[u] - E[u_h]\|_{L^2(D)} \leq Ch^{\min(r+1, s)} \|u\|_{L_P^2(\Gamma; H^s(D))}. \quad (5.12)$$

where C is a constant independent of h and u .

Proof 5.5 Using the same argument as in (4.26), it is easy to obtain

$$\|E[u - u_h]\|_{DG} \leq E[\|u - u_h\|_{DG}^2]^{\frac{1}{2}}.$$

Then, by (5.5) and (5.6) in Lemma 5.2, we obtain the conclusion. $\square$

Finally, I compare the computational cost of MCFEM and MCDG. Recall the result in chapter 8 in [10], we know that the computational work of MCFEM is $\frac{M}{h^d}$, where d is the dimension of the space domain D . Consider the MCDG, for each simulation, (4.39) also holds. Thus, we have

$$Work_{MCDG} = M \left(\frac{r^2}{h^d} \right). \quad (5.13)$$

We summarize the result in Table 5.1.

Chapter 6

Numerical Examples

6.1 Introduction

In this chapter, I first give a numerical example for the K-L expansion of the standard Wiener process. I plot the expansion of the Wiener process with first 5, 50, 500 and 1000 terms respectively. For each case, I randomly generate 1000 samples and plot the expected value and the variance. Then, I implement the Monte Carlo discontinuous Galerkin method in cases of nonsymmetric interior penalty Galerkin (NIPG) method, symmetric interior penalty Galerkin (SIPG) method and incomplete interior penalty Galerkin (IIPG) method. In each case, I plot the expected value and the variance on two different meshes with penalty parameters $\sigma_e = 0.1, 1, 10$ and 100 . Additionally, I study the effect of the penalty magnitude on the expected value and variance for NIPG. In the last part, I compare the Monte Carlo Finite Element Methods and Monte Carlo Discontinuous Galerkin Method.

6.2 K-L Expansion of Standard Wiener Process

In this section, I construct a numerical example about the standard Wiener process by using truncated K-L expansion. Let me first give an introduction to the standard Wiener process.

Definition 6.1 (Standard Wiener Process) It is a random variable W_t on the interval $[0, T]$ that depends continuously on $t \in [0, T]$ and satisfies the following:

- $W_0 = 0$.
- For $0 \leq s < t \leq T$,

$$W_t - W_s \text{ satisfies } N(0, t - s). \quad (6.1)$$

where $N(0, t - s)$ is the normal distribution with expected value 0 and variance $t - s$.

- For $0 \leq s < t < u < v \leq T$, $W_t - W_s$ and $W_v - W_u$ are independent.

By the Karhunen–Loève expansion [21],

$$W_t(\omega) = \sum_{n=0}^{\infty} Z_n(\omega) \phi_n(t) \quad \text{for } 0 \leq t \leq T.$$

where the $Z_0, Z_1, \dots, Z_n, \dots$ are independent standard Gaussian random variables; $\phi_0, \phi_1, \dots, \phi_n, \dots$ are the nonrandom functions

$$\phi_n(t) = \frac{2\sqrt{2T}}{(2n+1)\pi} \sin\left(\frac{(2n+1)\pi t}{2T}\right) \quad \text{for } n \geq 0.$$

Also, we have

$$Z_n(t) = \frac{2}{T} \left(\frac{(2n+1)\pi}{2\sqrt{2T}} \right)^2 \int_0^T W_t(\omega) \phi_n(t) dt \quad \text{for } n \geq 0,$$

and the Z_n 's are independent standard Gaussian random variables.

In this example, I take $T = 1$. For the truncated K-L expansion of the Wiener process,

$$W_{t,N}(\omega) = \sum_{n=0}^N Z_n(\omega) \phi_n(t) \quad \text{for } 0 \leq t \leq 1$$

By taking $N = 10, 50, 500, 1000$ and generating Z_n 's in MATLAB using the function “randn”, I plot 1000 truncated Wiener processes in Fig. 6.1.

In Fig. 6.1, I find that as the number of the terms in K-L expansion increases, the curves become less smooth and more oscillatory. This change is especially apparent

Figure 6.1 : 1000 samples of truncated K-L expansion of the standard Wiener process.

from the curve with first 10 terms of K-L expansion to the curve with first 50 terms. When truncating with the first 50 terms in K-L expansion, the simulation is much more accurate compared with the real standard Wiener process [21].

In order to verify the second property of the standard Wiener process (6.1), I plot the expected value and the variance of the truncated K-L expansion of this process.

In Fig. 6.2, we see that as the number of the terms in the truncated K-L expansion increases, the curve becomes more oscillatory. Also, I observe that the accuracy of the expected value obtained with 1000 simulations does not change when increasing the number of the terms in K-L expansion. Indeed the numerical expected value is of the order $\mathcal{O}(10^{-2})$ independently of N . This is reasonable because in Monte Carlo simulation, the accuracy is bounded by a constant times $\frac{1}{\sqrt{M}}$ where M is the times of simulation. In my example, we obtain $\frac{1}{\sqrt{1000}} = \mathcal{O}(10^{-2})$, which matches the result of the numerical experiment in Fig. 6.2. In Fig. 6.3, the value of the variance linearly increases and the slope is nearly 1. This exactly describes the second property 6.1.

Figure 6.2 : Expected value of the 1000 samples.

Figure 6.3 : Variance of the 1000 samples.

Through this numerical example, we conclude that the K-L expansion is a reliable and accurate way of approximating a stochastic process.

6.3 Monte Carlo Discontinuous Galerkin Method

In this section, I consider three methods: NIPG, SIPG and IIPG. In each case, I plot the expected value and the variance of the solution on a coarse mesh and a refined mesh with four different penalty parameters $\sigma_e = 0.1, 1, 10$ and 100 . In these experiments, the unstructured coarse mesh consists of 228 triangular elements and the unstructured refined mesh consists of 1824 triangular elements. Both meshes were generated by Gmsh [28]. The space domain is $D = [0, 1] \times [0, 1]$ and the outcome set is $\Gamma = [-1, 1]^{10}$. The stochastic diffusion function and load function are

$$\begin{aligned} a(\mathbf{y}, \mathbf{x}) &= \pi^2 + 10 \sum_{i=1}^{10} \sin(2\pi i x_1) \cos(2\pi i x_2) y_i / (i\pi)^2, \\ f(\mathbf{y}, \mathbf{x}) &= 10\pi^2 \sin(2\pi x_1) \cos(2\pi x_2) \\ &\quad + 6 \sum_{i=1}^{10} \sin(2\pi(x_1^2 + x_2^2)) \exp(-i(x_1^2 + x_2^2)) y_i. \end{aligned}$$

6.3.1 Nonsymmetric Interior Penalty Galerkin Method

I first sample 10 simulations of the solution by MCDG using the coarse mesh and plot the expected value and variance in Fig. 6.4 and Fig. 6.5. The penalty parameter is $\sigma_e = 0.1, 1, 10$ and 100 for all e and the polynomial degree is $r = 2$.

We see that the NIPG solution is stable for different values of penalty parameters on the coarse mesh. In order to see whether NIPG is also stable on the refined mesh, I sample 10 realizations of the solution by MCDG using the refined mesh. In Fig. 6.6 and Fig. 6.7, we see that NIPG is also stable on the refined mesh.

In order to obtain a more accurate solution by MCDG, I simulate 100 solutions

Figure 6.4 : Expected value of 10 NIPG solutions on the coarse mesh.

Figure 6.5 : Variance of 10 NIPG solutions on the coarse mesh.

Figure 6.6 : Expected value of 10 NIPG solutions on the refined mesh.

Figure 6.7 : Variance of 10 NIPG solutions on the refined mesh.

Figure 6.8 : Expected value of 100 NIPG solutions on the coarse mesh.

and plot the expected value and variance in Fig. 6.8 and Fig. 6.9. To find out whether the values of the penalty parameters affect the expected value and the variance by MCDG, I plot the difference of the expected value and variance of the solution with penalty parameters between 0.1 and 1, 10 and 1, 100 and 1, 1000 and 1 in Fig. 6.10 and Fig. 6.11. From the circular areas in the middle of Fig. 6.10 and Fig. 6.11, we find that the magnitude of the expected value and the variance are increasing with the value of penalty parameters increases. However, the difference remains overall small ($\mathcal{O}(10^{-3})$ for expected value and $\mathcal{O}(10^{-4})$ for variance). We conclude that the size of the penalty parameter for the NIPG method has negligible effects on the solution.

Figure 6.9 : Variance of 100 NIPG solutions on the coarse mesh.

Figure 6.10 : Difference of the expected values obtained with NIPG with varying penalty parameters.

Figure 6.11 : Difference of the variance obtained with NIPG with varying penalty parameters.

6.3.2 Symmetric Interior Penalty Galerkin Method

In primal discontinuous Galerkin method, the stability of SIPG method depends on the magnitude of value of the penalty parameter. I show here that a similar phenomenon takes place in MCDG. In Fig. 6.12 and Fig. 6.13, I plot the expected value and the variance of 10 simulations of the solution using the coarse mesh and the refined mesh with penalty parameters $\sigma_e = 0.1, 1, 10$ and 100 .

I first observe that meshes have an effect on the performance of SIPG in the MCDG method. Generally speaking, SIPG is more stable on the refined mesh than on the coarse mesh. However, when the penalty parameter is equal to 10 , the expected value and the variance of the solutions on the refined mesh is of poor accuracy compared with the solutions generated with penalty parameter $\sigma_e = 0.1, 1$ and 100 . For the

Figure 6.12 : Expected value of the SIPG solution by MCDG with different penalty parameters on coarse (left) and refined (right) meshes.

Figure 6.13 : Variance of the SIPG solution by MCDG with different penalty parameters on coarse (left) and refined (right) meshes.

case of the coarse mesh, when the penalty parameters are less than or equal to 10, the expected value and the variance are not accurate. The SIPG method is stable in this example when the penalty parameter is equal to 100. We conclude that the penalty parameter has to be chosen large enough to obtain a reasonable MCDG solution.

By the results in [29], the estimate of the threshold value is:

$$\sigma^* = \frac{3a_u^2}{a_l} p(p+1) \cot(\theta_T).$$

where a_u and a_l are the upper and lower bound of the diffusion function a ; θ_T is the smallest angle over all triangles in the mesh; p is the polynomial degree in DG method. In the numerical experiment, I take

$$\begin{aligned} a_u &= \pi^2 + 10 \sum_{i=1}^{10} \frac{1}{(i\pi)^2}, \\ a_l &= \pi^2 - 10 \sum_{i=1}^{10} \frac{1}{(i\pi)^2} \end{aligned}$$

In the coarse mesh, $\theta_T = 39.4851^\circ$, $\sigma^* = 344.5033$; in the refined mesh, $\theta_T = 29.0184^\circ$, $\sigma^* = 511.6674$. This implies that if the value of penalty parameter is larger than 344.5033, SIPG is stable on the coarse mesh for sure; if it is larger than 511.6674, SIPG is stable on the refined mesh for sure.

6.3.3 Incomplete Interior Penalty Galerkin Method

For the IIPG method, I repeat the experiments done for the SIPG method. I plot the expected value and the variance of 10 solutions on the coarse mesh and the refined mesh with penalty parameter $\sigma_e = 0.1, 1, 10$ and 100.

From Fig. 6.14 and Fig. 6.15, I first find that compared to SIPG, IIPG is not stable when the penalty parameter is equal to 0.1 on both meshes, while SIPG is stable with the same parameter 0.1 on the refined mesh. However, IIPG is of higher

Figure 6.14 : Expected value of the IIPG solution by MCDG with different penalty parameters on coarse (left) and refined (right) meshes.

Figure 6.15 : Variance of the IIPG solution by MCDG with different penalty parameters on coarse (left) and refined (right) meshes.

accuracy when the penalty parameter is equal to 10 on both meshes, while SIPG is not. When the penalty parameter is greater than or equal to 100, IIPG is also stable, as is the case for SIPG. From these numerical results, we conclude that the choice of the penalty parameters is crucial for the IIPG method. It has to be large enough.

6.4 Discontinuous Diffusion and Load Functions

In this section, I discuss NIPG method with the discontinuous diffusion and load functions. I generate an unstructured refined mesh consisting 1344 triangular elements by Gmsh. The space domain is $[-1, 1] \times [0, 1]$ and the outcome set is $[-1, 1]^{10}$. The discontinuous stochastic diffusion function and load function are:

When $x_1 \geq 0$:

$$\begin{aligned} a(\mathbf{y}, \mathbf{x}) &= \pi^2 + 10 \sum_{i=1}^{10} \sin(2\pi i x_1) \cos(2\pi i x_2) y_i^1 / (i\pi)^2, \\ f(\mathbf{y}, \mathbf{x}) &= 10\pi^2 \sin(2\pi x_1) \cos(2\pi x_2) \\ &\quad + 6 \sum_{i=1}^{10} \sin(2\pi(x_1^2 + x_2^2)) \exp(-i(x_1^2 + x_2^2)) y_i^1. \end{aligned}$$

When $x_1 < 0$:

$$\begin{aligned} a(\mathbf{y}, \mathbf{x}) &= \pi^2 + 10 \sum_{i=1}^{10} \cos(2\pi i x_1) \cos(2\pi i x_2) y_i^2 / (i\pi)^2, \\ f(\mathbf{y}, \mathbf{x}) &= 10\pi^2 \cos(2\pi x_1) \cos(2\pi x_2) \\ &\quad + 6 \sum_{i=1}^{10} \cos(2\pi(x_1^2 + x_2^2)) \exp(-i(x_1^2 + x_2^2)) y_i^2. \end{aligned}$$

I sample 10 NIPG solutions and plot the expected value and the variance in Fig. 6.16.

From Fig. 6.16, we conclude that NIPG solution is stable and continuous.

Figure 6.16 : Expected value (left) and Variance (right) of the NIPG solution by MCDG with discontinuous diffusion function and load function.

6.5 Comparison of Monte Carlo Finite Element Method and Monte Carlo Discontinuous Galerkin Method

In this section, the performance of the Monte Carlo Finite Element Method (MCFEM) and the Monte Carlo Discontinuous Galerkin (MCDG) method is compared. I consider two cases: one is with continuous diffusion and load functions, the other is with discontinuous coefficient functions. The formulation of MCFEM can be found in section 4 in [10]. It is very similar to the MCDG method defined in Chapter 5. In Section 5.1, instead of solving (5.1), I now solve

$$(a(y_j, \cdot) \nabla u_h(y_j, \cdot), \nabla v)_{L^2(D)} = (f(y_j, \cdot), v)_{L^2(D)}, \quad \forall v \in X_h^d.$$

where X_h^d is the finite element space of continuous piecewise linear polynomials defined on D .

6.5.1 Case 1: Continuous Coefficients

In this part, I take the same diffusion and load functions to the ones defined in Section 6.3, but discuss the difference of the two methods on three different meshes. I generate the coarse mesh and the refined ones using Gmesh to obtain the other two.

Table 6.1 : Error of MCFEM and MCDG

Mesh	MCFEM Error	MCDG Error
Coarse	2.7×10^{-3}	3.7996×10^{-4}
Refine1	1×10^{-3}	3.1386×10^{-4}
Refine2	4.6623×10^{-4}	3.5857×10^{-4}

Firstly, I consider the case where the DG method is used with polynomials of degree 2. In Fig. 6.17, I take 1000 simulations on the first two cases and 100 for the last case.

Then, I compare the MCFEM with the MCDG, where DG is used with polynomial of degree 1. Fig. 6.18 shows the results. I take 1000 simulations on the first two cases and 100 for the last case.

From Fig. 6.18, we cannot tell which method is more accurate. In order to compare the accuracy of these two methods, I consider the expected value of the solution using MCDG with polynomials of degree 2 on the finest mesh in Fig. 6.17 as the exact solution u_{ext} and compute the error of MCFEM and MCDG, where DG is used with polynomial of degree 1. The error is computed using the formulation:

$$\text{Error} = \|u_{app} - u_{ext}\|/N$$

where N is the number of the common nodes in the meshes on which the approximate solution and the exact solution are generated. Table 6.1 implies that MCDG is more accurate than MCFEM.

Then I compare the MCDG and MCFEM on two meshes of roughly the same degrees of freedom, 837 and 858 respectively. I take 1000 simulations using these two methods. Fig 6.19 shows the results.

Figure 6.17 : Approximation of expected value by MCDG method (left column) and MCFEM method (right column) on a coarse mesh and two finer meshes with continuous coefficients. Polynomial of degree 2 are used for MCDG.

Figure 6.18 : Approximation of expected value by MCDG method (left column) and MCFEM method (right column) on a coarse mesh and two finer meshes with continuous coefficients. Polynomial of degree 1 are used for MCDG.

Figure 6.19 : Approximation of expected value by MCDG method (left) and MCFEM method (right) obtained on the meshes with roughly the same degrees of freedom. Degree of freedom is 837 for MCDG and 858 for MCFEM.

6.5.2 Case 2: Discontinuous Coefficients

In this part, I compare the MCFEM and MCDG with discontinuous diffusion and load functions. I use the same coefficient functions as the ones in Section 6.4. I generate the coarse mesh using Gmsh and refine it twice by Gmsh to obtain the other two. I take 1000 simulations on the first two meshes and 100 simulations on the third one. In the same way as in Section 6.5.1, I first compare the MCFEM with MCDG, where DG is with polynomial of degree 2.

From Fig. 6.20, we conclude that the MCDG is more accurate than the MCFEM after we refine the mesh. However, in the coarse mesh, the discontinuity of the coefficient functions leads to the instability of the MCDG solution. Fig. 6.21 is the comparison of the MCDG with polynomial of degree 1 and the MCFEM on the same three meshes.

In Fig. 6.21 and Fig. 6.20, we find that as we refine mesh, MCDG is stable and performs accurately. If we take the polynomial of degree 2 in DG method, MCDG approximates the solution much accurately than MCFEM, although it is more expensive.

Figure 6.20 : Approximation of expected value by MCDG method (left column) and MCFEM method (right column) on a coarse mesh and two finer meshes with discontinuous coefficients. Polynomial of degree 2 are used for MCDG.

Figure 6.21 : Approximation of expected value by MCDG method (left column) and MCFEM method (right column) on a coarse mesh and two finer meshes with discontinuous coefficients. Polynomial of degree 1 are used for MCDG.

Chapter 7

Conclusions

In the last few decades, stochastic partial differential equations have become one of the most dynamic research areas and many numerical methods have been applied to those equations. In my thesis, I consider elliptic partial differential equations with stochastic coefficients. I establish and analyse the $k \times h$ -version stochastic discontinuous Galerkin method (SDG) and provide the a priori error estimates in the energy norm and L^2 norm. In the energy norm, the error between the exact solution and the SDG solution is of order $h + \sum_{n=1}^N (k_n)^{q_n+1}$. In addition, I compute the expected value of the error in the DG norm and conclude that the order is the same with the one in the energy norm. In the L^2 norm, I obtain the error estimates for the computation of the expected value of the solution. The order of the error is $h^2 + \sum_{n=1}^N (k_n)^{q_n+1}$, which is higher in space than the ones in the energy norm and DG norm. Furthermore, I discuss the computational work of the stochastic discontinuous Galerkin method and find that the corresponding computational cost is $\prod_{n=1}^N \left(\frac{q_n+1}{k_n} \right) \frac{r^2}{h^d}$. It is challenging to compare it with the cost of the stochastic finite element method in order to conclude which method is cheaper. Even though in Table 4.1 we see the cost of the stochastic finite element method is only $\prod_{n=1}^N \left(\frac{q_n+1}{k_n} \right) \frac{1}{h^d}$, we cannot say stochastic finite element method is cheaper, since the constant in front of the cost depends on the polynomial degree r , while the constant in front of the cost of stochastic discontinuous Galerkin method is independent of r .

Besides, I establish the Monte Carlo discontinuous Galerkin method and analyze

its stochastic error and space discretization error. I find that as the number of the realizations increases, the stochastic error in the DG norm tends to zero. For the space discretization error, I obtain two a priori error estimates in the DG norm and L^2 norm of order $h^{\min(k+1,s)-1}$ and $h^{\min(k+1,s)}$ respectively, which are the same order with the ones of the a priori error estimates for the Monte Carlo discontinuous Galerkin method by computing the expected value of the error. And the computational work of this method is $M \left(\frac{r^2}{h^d} \right)$.

In the numerical experiments in Chapter 6, we conclude that in the MCDG method, NIPG is stable independently of meshes and the values of the penalty parameters. SIPG and IIPG are stable when the penalty parameter is large enough. SIPG and IIPG perform better on the refined mesh than on the coarse mesh, i.e., the SIPG and IIPG solution are more stable on the refined mesh. These results are similar to these obtained for deterministic PDEs. Also, I compare the MCDG and MCFEM for the cases where the coefficients are discontinuous and continuous. To find which method is more accurate, further investigation should be taken.

As future work, I plan

1. to find a faster linear system solver. In my thesis, I use Lapack to solve the system $Ax = b$. In the numerical examples, one simulation on the refined mesh takes around 75 seconds for NIPG, SIPG and IIPG. If I take 1000 simulations, the solver is not effective. Since the times of simulation in Monte Carlo method determine the accuracy of the solution, I need to find a faster solver leading to higher accuracy. I plan to use the Petsc solvers.
2. to implement the stochastic collocation method. For a problem with a low dimensional outcome set, the stochastic collocation method has many advantages.

It leads to the solution of uncoupled deterministic problems and deals with the case where the input data depending non-linearly on the random variables efficiently. In the numerical experiment in my thesis, the dimension of the outcome set is 10, so applying stochastic collocation method will be promising.

3. to apply the developed methods to a time-dependent non-linear problem, such as the two-phase flow problem in porous media.

Bibliography

- [1] P. E. Kloeden, *Stochastic Ordinary and Stochastic Partial Differential Equations*. Springer-Verlag, New York, 2008.
- [2] G. Stefanou and M. Papadrakakis, “Assessment of spectral representation and Karhunen–Loève expansion methods for the simulation of Gaussian stochastic fields,” *Computer Methods in Applied Mechanics and Engineering*, vol. 196, pp. 2465–2477, 2007.
- [3] W. Reed and T. Hill, “Triangular mesh methods for the neutron transport equation,” tech. rep., Los Alamos Scientific Laboratory, 1973.
- [4] M. Wheeler, “An elliptic collocation-finite element method with interior penalties,” *SIAM Journal on Numerical Analysis*, vol. 15, pp. 152–161, 1978.
- [5] D. Arnold, “An interior penalty finite element method with discontinuous elements,” *SIAM Journal on Numerical Analysis*, vol. 19, pp. 742–760, 1982.
- [6] B. Rivière, M. Wheeler, and V. Girault, “Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems,” *Part I, Computational Geosciences*, vol. 3, pp. 337–360, 1999.
- [7] C. Dawson, S. Sun, and M. Wheeler, “Compatible algorithms for coupled flow and transport,” *Computer Methods in Applied Mechanics and Engineering*, vol. 193, pp. 2565–2580, 2004.

- [8] B. Rivière, M. Wheeler, and K. Banas, “Part II. Discontinuous Galerkin Method Applied to Single Phase Flow in Porous Media,” *Computational Geosciences*, vol. 4(4), pp. 337–341, 2000.
- [9] M. Kleiber and T. D. Hien, *The Stochastic Finite Element Method*. Wiley, 1992.
- [10] I. Babuška, R. Tempone, and G. E. Zouraris, “Galerkin finite element approximations of stochastic elliptic partial differential equations,” *SIAM Journal on Numerical Analysis*, vol. 42(2), pp. 800–825, 2004.
- [11] I. Babuška, R. Tempone, and G. E. Zouraris, “Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation,” *Computer Methods in Applied Mechanics and Engineering*, vol. 194(12-16), pp. 1251–1294, 2005.
- [12] I. Babuška, F. Nobile, and R. Tempone, “A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data,” *SIAM Journal on Numerical Analysis*, vol. 45(3), pp. 1005–1034, 2007.
- [13] B. Ganis, I. Yotov, and M. Zhong, “A stochastic mortar mixed finite element method for flow in porous media with multiple rock types,” tech. rep., University of Pittsburgh, 2010.
- [14] E. Allen, S. Novosel, and Z. Zhang, “Finite Element and difference approximation of some linear stochastic partial differential equations,” *Stochastics and Stochastics Reports*, vol. 64(1-2), pp. 117–142, 1998.
- [15] R. Yao and L. Bo, “Discontinuous Galerkin method for elliptic stochastic partial differential equations on two and three dimensional spaces,” *Science in China Series A: Mathematics*, vol. 50(11), pp. 1661–1672, 2007.

- [16] G. Fishman, *Monte Carlo: Concepts, Algorithms, Applications*. Springer-Verlag, 1996.
- [17] R. Caflisch, *Monte Carlo and Quasi-Monte Carlo Methods*. Cambridge University Press, Cambridge, UK, 1998.
- [18] I. Sobol, “On quasi-Monte Carlo integrations,” *Mathematics and Computers in Simulation*, vol. 47, pp. 103–112, 1998.
- [19] M. S. Gockenbach, *Understanding and Implementing the Finite Element Method*. Society for Industrial and Applied Mathematics, 1998.
- [20] L. C. Evans, *Partial Differential Equations*. Graduate Studies in Mathematics, American Mathematical Society, 1998.
- [21] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equation*. Springer-Verlag, New York, 1992.
- [22] P. D. Lax, *Functional Analysis*. Pure and Applied Mathematics: A Wiley-Interscience Series of Texts, Monographs, and Tracts, 2002.
- [23] M. Loève, *Probability Theory. I, fourth edition*, in: *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1977.
- [24] B. Øksendal, *Stochastic Differential Equations. An Introduction with Applications*. Springer-Verlag, 1998.
- [25] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, 1994.

- [26] B. Rivière, *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations : Theory and Implementation*. Society for Industrial and Applied Mathematics, 2008.
- [27] P. Billingsley, *Probability and measure*. New York : J. Wiley and Sons, 1995.
- [28] C. Geuzaine and J.-F. Remacle, “Gmsh: A Three-dimensional Finite Element Mesh Generator with Built-in Pre- and Post-processing Facilities,” *International Journal for Numerical Methods in Engineering*, vol. 79(11), pp. 1309–1331, 2009.
- [29] Y. Epshteyn and B. Rivire, “Estimation of penalty parameters for symmetric interior penalty Galerkin methods,” *Journal of Computational and Applied Mathematics*, vol. 206, pp. 843–872, 2007.