THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

SPARSE GRID STOCHASTIC COLLOCATION TECHNIQUES FOR THE NUMERICAL SOLUTION OF PARTIAL DIFFERENTIAL EQUATIONS WITH RANDOM INPUT DATA

By

CLAYTON G. WEBSTER

A Dissertation submitted to the
Department of Mathematics and School of Computational Science
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Spring Semester, 2007

The members of the Committee approve the Dissertation of Clayton G. Webster defended on Feb. 15, 2007.

Max D. Gunzburger
Professor Directing Dissertation

Kyle Gallivan
Outside Committee Member

Janet Peterson
Committee Member

Raul Tempone
Committee Member

The Office of Graduate Studies has verified and approved the above named committee members.

To Sarah, Mom and Dad
for your love and encouragement

ACKNOWLEDGEMENTS

With great satisfaction I wish to thank the many people who made this thesis possible. I would first like to acknowledge my gratitude to my committee members, who waited patiently for the finished version of this work.

It is difficult to overstate my gratitude to my Ph. D. supervisor, Dr. Max Gunzburger. With his enthusiasm, his inspiration, and his great efforts to explain things clearly and simply, he helped me discover an area of mathematics that I am very passionate about. Throughout my years at FSU, he provided encouragement, sound advice, excellent teaching, great company, and a significant amount of new ideas. I would have been lost without him. I must especially thank Max for having such confidence in my abilities and always having my best interest in mind. Max, you are an exceptional person and extremely talented mathematician and I am in constant awe of your tremendous breadth, of not only mathematics but more importantly, *baseball*. It has been an "honour" to work for you and I look forward to spending time together in the future discussing both topics.

I also wish to thank a wonderful teaching and great listener, Dr. Janet Peterson. I had a variety of interesting teachers while completing my Ph. D. and I wanted to take this opportunity to express my appreciation for your dedication to providing your students with excellent sets of notes, helpful assignments and useful feedback. I would also like to thank both you and Max for providing me with a "home away from home" and for regularly inviting and including me several activities outside of work. You both always expressed a genuine interest in my well-being, which constantly reminded me of my reasons for choosing FSU over any other institute.

I will be forever indebted to my great friend, Dr. John Burkardt. None of this work would have ever completely finished without your assistance and friendship. You will never understand how much I appreciated our talks, both intellectually and non-intellectually (there were a few). You helped me make it through some extremely tough days. Your

patience in supporting my many computational questions made my days infinitely better. Not to mention the 5-10 daily cups of coffee we shared. I never could have imagined that coming to FSU would allow me to meet such a great person.

I wish to thank my "best man" Dr. Catalin Trenchea for all the emotional support, comraderie, entertainment, and caring you provided. I am grateful that you expressed an interest in my research and for your mathematical assistance in understanding some exceptionally challenging problems. I am extremely optimistic that our future collaborations will be fruitful and very exciting. You may be "cold" in your new climate but you will always hold a "warm" place in my thoughts.

I would also like to take this opportunity to thank both Dr. Raul Tempone and Dr. Fabio Nobile. What started as an simple question posed through an email has turned into some excellent mathematics and great friendships. I thank-you both for always being available to answer questions, discuss proofs, talks, codes, etc. I would also like to express my sincere gratitude towards Raul and Fabio for introducing me to many people in this field and exposing me to a variety of new and awesome experiences. Your support has been greatly appreciated and will always be remembered.

I most also express my gratitude to the the School of Computational Science (SCS) at Florida State University and my advisor, Max Gunzburger, for generously supporting my research.

I wish to thank my entire extended family for providing a loving environment and an incredible childhood. My grandparents , Jim and Ethel Duncan, my sister, Harmony Webster-Payea, and my Aunt, Allison Pellick were particularly supportive. I wish to also acknowledge my beautiful and precious nieces, Mekayla and Kennedy Webster-Payea, who often remind me how magnificent life can be.

Lastly, and most importantly, I wish to thank my wonderful parents, Gerald and Jay-Dee Webster and the love of my life, Sarah Duncan. Mom and Dad, the completion of this work would not have been possible without your encouragement, love and support. Thanks for provided me with so many great memories, opportunities and experiences. I have always considered myself extremely lucky to have such incredible people for parents. It is because of hard work and dedication that I reached my dreams and it is because of you that I realized there isn't any other way to get there. Sarah, you are my best friend, the greatest thing to ever happen to me and I will forever appreciate how understanding you have been throughout

this tremendous undertaking. You are my foundation and you have held me up when no one else could. Thanks for tolerating those long nights and weekends of work and I promise to spend the rest of our lives together making it up to you. To the three of you, I dedicate this thesis.

— Clayton

TABLE OF CONTENTS

List of Tables ix

List of Figures xi

Abstract xiv

INTRODUCTION 1

1. PROBLEM SETTING 10
- 1.1 Notation and function spaces 10
- 1.2 On finite dimensional noise 13
- 1.3 Regularity 16

2. LINEAR ELLIPTIC PDES WITH RANDOM INPUT DATA 18

3. NONLINEAR ELLIPTIC PDES WITH RANDOM INPUT DATA 21
- 3.1 Coercivity, lower semicontinuity 23
- 3.2 Existence and uniqueness of minimizers 26
- 3.3 Weak solutions of a nonlinear stochastic PDE 27
- 3.4 Analyticity and the deterministic problem 31

4. AN OVERVIEW OF NUMERICAL METHODS 43
- 4.1 Approximation spaces 46
- 4.2 Monte Carlo 48
- 4.3 Stochastic Galerkin 49
- 4.4 Stochastic Collocation techniques 51

5. SELECTION OF THE WEIGHTS FOR ANISOTROPIC SMOLYAK 59
- 5.1 A priori selection 61
- 5.2 A posteriori selection 61

6. ERROR ANALYSIS 65
- 6.1 Analysis of the sparse interpolation error: The isotropic case 67
- 6.2 Analysis of the sparse interpolation error: The anisotropic case 81

7. NUMERICAL EXAMPLES 94

8. SUMMARY 111

9. A REDUCED-ORDER MODEL FOR SOLVING SPDES: A PROMISING RESEARCH DIRECTION 114
9.1 Some notation 117
9.2 The model nonlinear SPDE and its finite element approximation 121
9.3 A Reduced-order model for the nonlinear SPDE 124
9.4 Computational experiments 130
9.5 Summary 139

10. CONCLUDING REMARKS AND FUTURE RESEARCH 141

REFERENCES 146

BIOGRAPHICAL SKETCH 160

LIST OF TABLES

5.1 The $N = 11$ values of the function $g(n) = \alpha_n$ constructed from *a posteriori* information. The values $g(n)$, $n = 1, 2, \ldots, N$ are the slopes of N linear least squares fits to the error of an univariate anisotropic method when solving problem (7.0.1) with correlation lengths $L_c = 1/2$ and $1/64$. 64

7.1 The $N = 11$ values of the function $g(n)$ constructed from *a priori* information given by (7.0.9) for correlation lengths $L_c = 1/2, 1/4, 1/16, 1/64$. 97

7.2 The $N = 11$ components of the maximum indices $\boldsymbol{\kappa}(\boldsymbol{\alpha}, w)$ and $\widehat{\boldsymbol{\kappa}} = \boldsymbol{\kappa}(\widehat{\boldsymbol{\alpha}}, \overline{w} + 1)$, defined by (7.0.10) and (7.0.11), respectively, used for solving problem (7.0.1) with a correlation length $L_c = 1/2$. 99

7.3 The $N = 11$ components of the maximum indices $\boldsymbol{\kappa}(\boldsymbol{\alpha}, w)$ and $\widehat{\boldsymbol{\kappa}} = \boldsymbol{\kappa}(\widehat{\boldsymbol{\alpha}}, \overline{w})$, defined by (7.0.10) and (7.0.11), respectively, used for solving problem (7.0.1) with a correlation length $L_c = 1/64$. 100

7.4 The $N = 11$ components of the multi index $\mathbf{p}$ computed by he anisotropic full tensor product algorithm when solving problem (7.0.1) with a correlation length $L_c = 1/2$. 101

7.5 The $N = 11$ components of the multi index $\mathbf{p}$ computed by the anisotropic full tensor product algorithm when solving problem (7.0.1) with a correlation length $L_c = 1/64$. 101

7.6 For $N = 11$, we compare the number of function evaluations required by the Anisotropic Smolyak (AS) using Clenshaw-Curtis abscissas, Anisotropic Full Tensor product method (AF) using Gaussian abscissas, Isotropic Smolyak (IS) using Clenshaw-Curtis abscissas and the Monte Carlo (MC) method using random abscissas, to reduce the original error of problem (7.0.1), in expectation, by a factor of 10^4. 103

7.7 The $N = 11$ values of the function $g(n) = \alpha_n$ constructed from *a posteriori* information. The values $g(n)$, $n = 1, 2, \ldots, N$ are the slopes of N linear least squares fits to the error of an univariate anisotropic method when solving problem (7.0.2) with correlation lengths $L_c = 1/2$, $1/4$, $1/16$ and $1/64$. . . . 104

9.1 The first 16 singular values of the snapshot matrix for the example problem (9.4.1). 132

9.2 $\mathfrak{E}_T$ for the POD-based reduced-order models vs. the dimension K of the reduced basis space for the six test cases. 136

LIST OF FIGURES

4.1 For a finite dimensional Γ^N with $N = 2$ we plot the full tensor product grid using the Clenshaw-Curtis abscissas and isotropic Smolyak sparse grids utilizing the Clenshaw-Curtis abscissas and the Gaussian abscissas for $w = 5$. 54

4.2 For a finite dimensional Γ^N with $N = 2$ and $w = 7$ we plot: on the top, the isotropic Smolyak grid and the anisotropic Smolyak grids with $\alpha_2/\alpha_1 = 3/2$ and $\alpha_2/\alpha_1 = 2$, utilizing the Clenshaw-Curtis abscissas and the bottom, the corresponding indices $(i_1, i_2) \in X_{\boldsymbol{\alpha}}(7, 2)$. 56

5.1 A linear least square approximation to fit $\log_{10}(\|E[\varepsilon_n]\|_{L^2(D)})$ versus p_n with ε_n defined by (5.2.2). For $n = 1, 2, \ldots, N = 11$ we plot: on the left, the highly anisotropic case $L_c = 1/2$ and on the right, the isotropic case $L_c = 1/64$. . . 63

7.1 The rate of convergence of the isotropic Smolyak algorithm for solving problem (7.0.1) with given correlation lengths $L_c = 1/64$ using both the Gaussian and Clenshaw-Curtis abscissas. For a finite dimensional probability space Γ^N with $N = 5$ and $N = 11$ we plot $\log(\epsilon)$ versus the number of collocation points. The $L^2(D)$ approximation error in the expected value for the anisotropic sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[\mathscr{A}(w, N)\pi_h u_N - \mathscr{A}(w + 1, N)\pi_h u_N]\|_{L^2(D)}$ where $w = 0, 1, \ldots, \overline{w}$. 98

7.2 The convergence of the isotropic Smolyak algorithm for solving problem (7.0.1) with given correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$ using both the Gaussian and Clenshaw-Curtis abscissas. For a finite dimensional probability space Γ^N with $N = 5$ and $N = 11$ we plot $\log(\epsilon)$ versus the logarithm of the number of collocation points. The $L^2(D)$ approximation error in the expected value for the isotropic sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[\mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N - \mathscr{A}_{\boldsymbol{\alpha}}(w+1, N)\pi_h u_N]\|_{L^2(D)}$ where $w = 0, 1, \ldots, \overline{w}$. 105

7.3 The rates of convergence of the anisotropic Smolyak algorithm for solving problem (7.0.1) with given correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$ using both the Gaussian and Clenshaw-Curtis abscissas. For a finite dimensional probability space Γ^N with $N = 5$ and $N = 11$ we plot $\log(\epsilon)$ versus the number of collocation points. The $L^2(D)$ approximation error in the expected value for the anisotropic sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[\mathscr{A}_{\boldsymbol{\alpha}}(w,N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w},N)\pi_h u_N]\|_{L^2(D)}$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (7.0.11). 106

7.4 The convergence of the anisotropic Smolyak algorithm for solving problem (7.0.1) with given correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$ using both the Gaussian and Clenshaw-Curtis abscissas. For a finite dimensional probability space Γ^N with $N = 5$ and $N = 11$ we plot $\log(\epsilon)$ versus the logarithm of the number of collocation points. The $L^2(D)$ approximation error in the expected value for the anisotropic sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[\mathscr{A}_{\boldsymbol{\alpha}}(w,N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w},N)\pi_h u_N]\|_{L^2(D)}$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (7.0.11). 107

7.5 A 11-dimensional comparison of the anisotropic Smolyak method, the isotropic Smolyak method, the anisotropic full tensor product algorithm and Monte Carlo approach for solving problem (7.0.1) with correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$. The $L^2(D)$ approximation error in the expected value for the sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx E[\mathscr{A}_{\boldsymbol{\alpha}}(w,N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w},N)\pi_h u_N]$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (7.0.11). The $L^2(D)$ approximation error in the expected value for the anisotropic full tensor product method is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[u^N_{h,\mathbf{p}} - u^N_{h,\widetilde{\mathbf{p}}}]\|_{L^2(D)}$, where $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ and $\widetilde{\mathbf{p}} = (p_1 + 1, p_2 + 1, \ldots, p_N + 1)$. The $L^2(D)$ approximation error in the expected value for the Monte Carlo method is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|\overline{E}[u^N_{h,k}; M] - E[\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w},N)\pi_h u_N]\|_{L^2(D)}$, where $M = 2^i$, $i = 0, 1, 2, \ldots, 10$. 108

7.6 A linear least square approximation to fit $\log_{10}(\|E[\varepsilon_n]\|_{L^2(D)})$ versus p_n with ε_n defined by (5.2.2). For $n = 1, 2, \ldots, N = 11$ we plot the cases corresponding to $L_c = 1/2$, $1/4$, $1/16$ and $1/64$ when solving the nonlinear problem given by (7.0.2). 109

7.7 A 17-dimensional comparison of the anisotropic Smolyak method with Monte Carlo approach for solving problem (7.0.2) with correlation lengths $L_c = 1/4$ using the *a posterirori* approach described in Section 5.2. The $L^2(D)$ approximation error in the expected value for the anisotropic sparse collocation method is given by: $\|E[\epsilon]\|_{L^2(D)} \approx E[\mathscr{A}_{\boldsymbol{\alpha}}(w,N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w},N)\pi_h u_N]$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (7.0.11). The $L^2(D)$ approximation error in the expected value for the Monte Carlo method is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|\overline{E}[u^N_{h,k}; M] - E[\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w},N)\pi_h u_N]\|_{L^2(D)}$, where $M = 2^i$, $i = 0, 1, 2, \ldots, 10$. . . 110

9.1 Piecewise constant approximation for the noise $\Delta\widehat{\mathbf{W}}(t,\mathbf{x})$, defined by 9.1.5, at the final time $T = 1$ over a uniform finite element triangulation. The number of elements ranges from $\widetilde{M} = 2$ to $\widetilde{M} = 512$. 119

9.2 The first 100 singular values of the snapshot matrix for the example problem (9.4.1). 133

9.3 The POD reduced basis of cardinality 16 for the nonlinear stochastic problem 9.4.1. 134

9.4 $\mathfrak{E}(t)$ for the POD-based reduced-order models with $\epsilon = 0$ and $\epsilon = 0.1$. The plots are on the same scale indicated the effects on the error by introducing the perturbation parameter ϵ into the ROM. 135

9.5 The approximate expected value $\overline{E}(u^h)$, using the full FEM versus the approximate expected value $\overline{E}(u^K)$ using the ROM with 2, 8 and 16 POD vectors at times $t = 0, 0.25, 0.5, 0.75$ and 1. All plots are generated for the case $\epsilon = 0.1$. 137

9.6 $\mathfrak{E}(t)$ for the POD-based reduced-order models for Cases 1-6. As the perturbation parameter ϵ increases from 0 to 1.0, the scale of the plots is adjusted to clearly display the error behavior. 138

ABSTRACT

The objective of this work is the development of novel, efficient and reliable sparse grid stochastic collocation methods for solving linear and nonlinear partial differential equations (PDEs) with random coefficients and forcing terms (input data of the model). These techniques consist of a Galerkin approximation in the physical domain and a collocation, in probability space, on sparse tensor product grids utilizing either Clenshaw-Curtis or Gaussian abscissas. Even in the presence of nonlinearities, the collocation approach leads to the solution of uncoupled deterministic problems, just as in the Monte Carlo method.

The full tensor product spaces suffer from the *curse of dimensionality* since the dimension of the approximating space grows exponentially in the number of random variables. When this number is moderately large, we combine the advantages of isotropic sparse collocation with those of anisotropic full tensor product collocation: the first approach is effective for problems depending on random variables which weigh equally in the solution; the latter approach is ideal when solving highly anisotropic problems depending on a relatively small number of random variables. We also include *a priori* and *a posteriori* procedures to adapt the anisotropy of the sparse grids to each problem. These procedures are very effective for the problems under study.

This work also provides a rigorous convergence analysis of the fully discrete problem and demonstrates: (sub)-exponential convergence in the asymptotic regime and algebraic convergence in the pre-asymptotic regime, with respect to the total number of collocation points. Numerical examples illustrate the theoretical results and compare this approach with several others, including the standard Monte Carlo. For moderately large dimensional problems, the sparse grid approach with a properly chosen anisotropy is very efficient and superior to all examined methods.

Due to the high cost of effecting each realization of the PDE this work also proposes the use of reduced-order models (ROMs) that assist in minimizing the cost of determining accurate statistical information about outputs from ensembles of realizations. We explore the use of ROMs, that greatly reduce the cost of determining approximate solutions, for determining outputs that depend on solutions of stochastic PDEs. One is then able to cheaply determine much larger ensembles, but this increase in sample size is countered by the lower fidelity of the ROM used to approximate the state. In the contexts of proper orthogonal decomposition-based ROMs, we explore these counteracting effects on the accuracy of statistical information about outputs determined from ensembles of solutions.

INTRODUCTION

Uncertainty quantification

Mathematical modeling and computer simulations are nowadays widely used tools to predict the behavior of physical and engineering problems. All such predictions are obtained by formulating mathematical models and solving corresponding problems. Therefore, mathematical models can be regarded as a transformation of available input information onto desired quantities of interest, and the predictions are the images of specific inputs. The problem that arises with these models is that, to date, most scientists ignore certain critical issues that should be addressed.

Some sources of errors arising in computer simulations can be controlled and reduced, by now, using sophisticated techniques such as *a posteriori* error estimation (Ainsworth and Oden, 2000; Babuška and Strouboulis, 2001; Eriksson et al., 1995; Verfürth, 1996; Johnson, 2000), mesh adaptivity Moon et al. (2006) and the more recent *modeling error* analysis (Oden and Vemaganti, 2000; Oden et al., 2001; Oden and Prudhomme, 2002; Braack and Ern, 2003; Romkes and Oden, 2004; Oden et al., 2005a,b). All this has increased the accuracy of numerical predictions as well as our confidence in them.

Yet, many engineering applications are affected by a relatively large amount of uncertainty in the input data such as model coefficients, forcing terms, boundary conditions, geometry, etc (Babuška and Chleboun, 2002, 2003). This can be due to an intrinsic variability in the physical system as, for instance, in the mechanical properties of many bio-materials, polymeric fluids, or composite materials, the action of wind or seismic vibrations on civil structures, etc. These types of uncertainties are known as *epistemic* since they can be reduced via additional experimentation, improvements in measuring devices, etc. For example, one can reduce uncertainties associated with characterization of material properties by carrying out a more exhaustive experimental program and/or using better measuring devices.

In other situations, uncertainty may come from our difficulty in characterizing accurately

the physical system under investigation as in the study of groundwater flows, where the subsurface properties such as porosity and permeability in an aquifer have to be extrapolated from measurements taken only in few spatial locations. In this case the uncertainties are known as *aleatoric* since one cannot reduce the associated uncertainties. In practice, it is necessary to address both types of uncertainties. This problem has been addressed in a general setting in Cullen and Frey (1999) and some applications to solid mechanics in Ben-Haim (1996); Mrczyk (1997); Elishakoff (1999); Melchers (1999); Elisakoff and Ren (2003).

The goal of this work is to advance key enabling mathematical and computational methods for *Uncertainty Quantification*, focusing on stochastic models and their numerical simulations.

Although such methods are essential for dealing with realistic experimental data and assessing reliability of predictions based on numerical simulations, their current understanding and use in the engineering community is quite limited. Additionally, their development within the computational mathematics community is very much ongoing. See for instance the cited works below. Uncertainties can be quantified in several ways, for instance:

Worst-case-scenario (or anti-optimization) are useful in cases where we only know little information on the uncertainty in the input data, namely that the input data lie in a functional set (that might well be infinite dimensional). These methods have been treated comprehensively in Hlaváček et al. (2004). It covers mathematical fundamentals, provides a survey of the pertinent literature, and discusses many applications. See also the recent work Babuška et al. (2005).

Probabilistic methods characterize uncertainties using stochastic differential equations. Significant literature is devoted to this subject in both mathematics and engineering (Ghanem and Spanos, 1991; Kleiber and Hien, 1992; Benth and Gjerde, 1998a,b; Ghanem and Red-Horse, 1999; Glimm et al., 2003; Xiu and Karniadakis, 2002a; Schwab and Todor, 2003b,a; Xiu and Karniadakis, 2003; Soize, 2003; Lucor et al., 2003; Lucor and Karniadakis, 2004; Le Maître et al., 2004a,c; Soize and Ghanem, 2004; Babuška et al., 2004b; Badri Narayanan and Zabaras, 2004; Zabaras and Samanta, 2004; Lu and Zhang, 2004; Xiu and Tartakovsky, 2004; Regan et al., 2004; Babuška et al., 2005d; Schueller, 2004; Keese and Matthies, 2005; Matthies and Keese, 2005; Frauenfelder et al., 2005a; Soize, 2005; Rubinstein and Choudhari, 2005; Velamur Asokan and

Zabaras, 2005; Asokan Badri Narayanan and Zabaras, 2005; Mathelin et al., 2005; Roman and Sarkis, 2006).

Knowledge-based methods characterize uncertainties using fuzzy sets (Bernardini, 1999; Dubois and Prade, 2000), evidence theory (Dempster-Shafer theory) (Oberkampf et al., 2001; Kramosil, 2001; Ferson et al., 2003), subjective probability (Vick, 2002; Helton, 1997), Bayesian inference (Box, 1973; Lemm, 2003; Beck and Au, 2002; Yuen and Beck, 2003; Beck and Ching, 2004; Wang and Zabaras, 2005), and other means of including expert opinions.

It is widely recognized that new methods for treating uncertainty will become important in virtually all branches of computational mechanics, see Oden et al. (2003); Babuška and Oden (2005). Above, we mentioned a few approaches that can be applied to solve problems with uncertainties. All of them can be applied directly and indirectly to a partial differential equation problem with uncertain input data. The major problem in all these approaches is the description of the uncertainty in the input set. Attending to this trend, we concentrate on the probabilistic characterization of uncertainties in mathematical models described by both linear and nonlinear Stochastic Partial Differential Equations (SPDEs). In particular, we will focus on the development of mathematical theory and efficient numerical algorithms to produce reliable numerical approximations for a wide variety of applied problems.

Such uncertainties can be included in the mathematical model adopting a probabilistic setting, provided enough information is available for a complete statistical characterization of the physical system. In this framework, the input data are modeled as *random variables*, like in the case where the input coefficients are piecewise constant and random over fixed subdomains, or more generally, as *random fields* with a given spatial (or temporal) correlation structure.

Therefore, the goal of the mathematical and computational analysis becomes the prediction of statistical moments (mean value, variance, covariance, etc.) or even the whole probability distribution of some responses of the system (quantities of physical interest), given the probability distribution of the input random data.

The stochastic formulation of uncertainty

In many applications the mathematical model depends on a set of parameters, which may be affected by uncertainty and be represented as random variables with a given joint probability distribution. For instance, we might think of the deformation of an elastic homogeneous material in which the Young's modulus and the Poisson's ratio (parameters that characterize the material properties) are random variables, either independent or not.

In other situations, the input data may vary randomly from one point of the physical domain D to another and their uncertainty should rather be described in terms of random fields with a given covariance structure (i.e. each point of the domain is a random variable and the correlation between two distinct points in the domain is known and non-zero, in general; this case is sometimes referred to as *colored noise*). Examples of this situation are, for instance, the deformation of inhomogeneous materials such as wood, foams, or bio-materials such arteries, bones, etc.; groundwater flow problems where the permeability in each layer of sediments (rocks, sand, etc.) should not be assumed constant; the action of wind (direction and point intensity) on structures; etc.

A possible way to describe such random fields consists of using a Karhunen-Loève expansion (Loève, 1977), a Polynomial Chaos (PC) expansion (Wiener, 1938; Ghanem and Spanos, 1991) or its generalized version (Xiu and Karniadakis, 2002b). The former represents the random field as a linear combination of an infinite number of uncorrelated random variables, while the latter uses polynomial expansions in terms of independent random variables. Both expansions exist provided that the given random field $a : \Omega \to W$, as a mapping from the probability space into a functional space W, has a bounded second moment. Other nonlinear expansions can be considered. See, e.g. Grigoriu (2002), for a technique to express a stationary random field with given covariance structure and marginal distribution as a function of (infinite) independent random variables; nonlinear transformations have been used also in Matthies and Keese (2005); Winter and Tartakovsky (2002). The use of non-polynomial expansions may be advantageous in some situations, for instance, in groundwater flow problems, the permeability coefficient within each layer of sediments can feature huge variability, which is often expressed in a logarithmic scale. In this case, one might want to use a Karhunen-Loève (or Polynomial Chaos) expansion for the logarithm of the permeability, instead of the permeability field itself. This leads to

an exponential dependence of the permeability on the random variables and the resulting random field might even have unbounded second moments. An advantage of such a nonlinear expansion is that it guarantees a positive permeability almost surely (a condition which is difficult to enforce with a standard truncated Karhunen-Loève or PC expansion).

In general, such random fields are properly described only by means of an infinite number of random variables, whenever the realizations are slowly varying in space, with a correlation length comparable to the size of the domain, only a few terms in the above mentioned expansion are typically needed to describe the random field with sufficient accuracy. Therefore, for these applications, it is reasonable to limit the analysis to just a few random variables in the expansion, see e.g. Babuška et al. (2003); Frauenfelder et al. (2005b).

In this work we focus on elliptic partial differential equations whose coefficients and forcing terms are described by a finite dimensional random vector (*finite dimensional noise assumption*, cf. Section 1.2), either because the problem itself can be described by a finite number of random variables or because the input coefficients are modeled as truncated random fields. We especially address the situation where the input data are assumed to depend on a moderately large number of random variables.

The method proposed here, namely an anisotropic Sparse Grid Stochastic Collocation, extends the isotropic method (Nobile et al., 2006) described in Section 4.4.2 and analyzed in Section 6.1, which consists of a Galerkin approximation in the space variables and a collocation, in probability space, on sparse tensor product grids utilizing either Clenshaw-Curtis or Gaussian abscissas. As a consequence of the collocation approach our technique naturally lead to the solution of uncoupled deterministic problems, just as in the Monte Carlo method.

The work Nobile et al. (2006) revealed that the isotropic sparse collocation algorithm is very effective for problems whose input data depend on a moderate number of random variables, which "weigh equally" in the solution. For such isotropic situations the displayed convergence is faster than standard collocation techniques built upon full tensor product spaces. On the other hand, the convergence rate of the isotropic sparse collocation algorithm (see Nobile et al. (2006) and Section 6.1) deteriorates for highly anisotropic problems, such as those appearing when the input random variables come e.g. from Karhunen-Loève -type truncations of "smooth" random fields. In such cases, a full anisotropic tensor product

approximation may still be more effective for a small or modest number of random variables. However, if the number of random variables is large, the construction of the full tensor product spaces becomes infeasible, since the dimension of the approximating space grows exponentially fast with respect to the number of random variables in the problem.

The main contribution of this work is to propose and analyze the use of anisotropic sparse tensor product spaces constructed from a weighted Smolyak interpolant with suitable abscissas. This approach is particularly attractive in the case of truncated expansions of random fields, since the anisotropy can be tuned to the decay properties of the expansion. We will present *a priori* and *a posteriori* procedures for choosing the anisotropy of the sparse grids which are extremely effective for the problems under study.

This work provides a rigorous convergence analysis of the fully discrete problem and demonstrates: (sub)-exponential convergence in the asymptotic regime and algebraic convergence in the pre-asymptotic regime, with respect to the total number of collocation points. Numerical examples illustrate the theoretical results and are used to compare this approach with several others, including the standard Monte Carlo. In particular, for moderately large dimensional problems, the sparse grid approach with properly chosen anisotropy seems to be very efficient and superior to all examined methods.

The Monte Carlo method described in Fishman (1996a) is the classical and most popular approach for approximating expected values of quantities of interest depending on the solution of a partial differential equations with random inputs. The algorithm approximates the desired expectation by a sample average of independent identically distributed (iid) realizations. When solving partial differential equations with random inputs, this method implies the solution on one deterministic differential equation for each realization of the input parameters. This makes the method simple to implement, allows for maximal code reusability and it is straightforward to parallelize. Its numerical error is approximately $\mathcal{O}(1/\sqrt{M})$, where M is the number of realizations. The advantage of utilizing this approach is that the rate does not deteriorate with respect to the number of random variables in the problem, making the method very attractive for problems with large dimensional random inputs. On the other hand, when solving large-scale applications, the exponent $1/2$ in the rate of convergence generates a tremendous amount of computational work required to achieve accurate solutions. Other ensemble based methods like Quasi Monte Carlo, Latin Hypercube Sampling, etc. (see e.g. Niederreiter (1992); Helton and Davis (2003) and references therein),

have been devised to produce faster convergence, $\mathcal{O}(\log(M)^r/M)$, where the coefficient $r > 0$ becomes larger with the dimension of the random input. We explore alternative methods that obtain faster convergence rates, exploiting the high regularity that the solution of elliptic PDEs may have with respect to the random input, while preserving the implementation advantages of ensemble-based methods.

In the last few years, other approaches have been proposed, which in certain situations feature a much faster convergence rate. We mention, among others, the Spectral Galerkin method (Ghanem and Spanos, 1991; Xiu and Karniadakis, 2002a; Babuška et al., 2004b, 2005d; Frauenfelder et al., 2005b; Matthies and Keese, 2005; Le Maître et al., 2004b; Roman and Sarkis, 2006), Stochastic Collocation (Babuška et al., 2005c; Tatang, 1995; Mathelin et al., 2005; Xiu and Hesthaven, 2005), perturbation methods or Neumann expansions (Gaudagnini and Neumann, 1999; Winter and Tartakovsky, 2002; Babuška and Chatzipantelidis, 2002a; Karniadakis et al., 2005; Todor, 2005).

A non-ensemble-based substitute for the Monte Carlo method is the so called Spectral Galerkin method, see e.g. Ghanem and Spanos (1991). It employs standard approximations in space (finite elements, finite volumes, spectral or h-p finite elements, etc.) and polynomial approximation in the probability domain, either by full polynomial spaces, see Xiu and Karniadakis (2002a); Matthies and Keese (2005); Ghanem (1999), tensor product polynomial spaces, see Babuška et al. (2004b); Frauenfelder et al. (2005b); Roman and Sarkis (2006) or piecewise polynomial spaces, see Babuška et al. (2004b); Le Maître et al. (2004b). This family of methods exploit the regularity of the solution to acquire faster convergence rates. However, in general, this technique requires solving a system of equations that couples all degrees of freedom in the approximation to the stochastic solution.

A new numerical technique, which has gained much attention recently by the computational community is Stochastic Collocation, which can be based either on full or sparse tensor product approximation spaces (Babuška et al., 2005c; Ganapathysubramanian and Zabaras, 2006; Nobile et al., 2006; Mathelin et al., 2005; Xiu and Hesthaven, 2005). As shown in Babuška et al. (2005c), Stochastic Collocation essentially preserves the fast convergence of the Spectral Galerkin method, even coinciding in particular cases, while maintaining an ensemble-based approach, just as Monte Carlo. In particular, in the work Babuška et al. (2005c) proposed a Stochastic Collocation/Finite Element method based on standard finite element approximations in space and a collocation on a tensor grid built upon the zeros of

orthogonal polynomials with respect to the joint probability density function of the input random variables. It was shown that for an elliptic PDE the error converges exponentially fast with respect to the number of points employed in each probability direction. Furthermore, even in presence of input data which depend nonlinearly on the driving random variables, the Stochastic Collocation can be easily implemented and leads naturally to the solution of uncoupled deterministic problems as for the Monte Carlo method. It can also treat efficiently the case of non independent random variables with the introduction of an auxiliary density and handle for instance cases with lognormal diffusivity coefficient, which is not bounded in $\Omega \times D$ but that has bounded realizations.

Hence, Stochastic collocation seems to be an ideal method for computing realistic statistics from solutions of PDEs with random input data, but the challenge comes when solving problems with a relatively large number of input random variables. Therefore, our goal is investigating variants of this procedure to attain highly accurate solutions while reducing the *curse of dimensionality*. To extend the applicability of the collocation method to cases with large dimensional random inputs, several sparse-approximation-based methods have been recently explored (Xiu and Hesthaven, 2005; Nobile et al., 2006, 2007; Ganapathysubramanian and Zabaras, 2006). Aligned with this research effort, the main contribution of this work is to propose and analyze a novel anisotropic stochastic collocation method that is based on a weighted version of the Smolyak algorithm. As our numerical and theoretical results indicate that this method seems very promising and worth exploring further, both from the implementation and the theoretical point of view.

The outline of this work is the following: in Chapter 1 we introduce the mathematical problem and the main notation used throughout. Here we also state assumptions on the parameterization of the random inputs, which is useful when later transforming the original stochastic problem into a deterministic parametric one, and on the problem solution's regularity, which is used to later prove the error estimates in Chapter 6.

In Chapters 2 and 3 we focus on applications to both linear and nonlinear elliptic PDEs with random input data respectively. The main ideas here are to exhibit the well-posedness of the PDEs under study and to verify the assumptions from Chapter 1 in a particular setting, showing that they are justified and that one can extend them to a variety of problems.

In Chapter 4 we introduce the approximation spaces and provide an overview of the before mentioned numerical procedures used to solve such problems. If the number of random

variables is moderately large, we motivate the consideration of sparse tensor product spaces as first proposed in Smolyak (1963) and further investigated by, e.g. Gerstner and Griebel (1998a); Barthelmann et al. (2000a); Frauenfelder et al. (2005b); Xiu and Hesthaven (2005), which will be the primary focus of this chapter. The use of sparse grids allows one to reduce dramatically the number of collocation points, while preserving a high level of accuracy. We also describe the anisotropic sparse approximation method to be considered as well as the different interpolation techniques to be employed.

The Karhunen-Loève representation of the input random fields naturally requires anisotropic refinements in the probability direction. The most important random variables in the expansion have the largest influence on the solutions, hence demanding higher polynomial degrees in those directions. In Chapter 5, we provide *a priori* and *a posteriori* procedures for tuning the anisotropy of our sparse grid method to the problem at hand.

In Chapter 6 we provide a complete error analysis for both the isotropic and anisotropic sparse grid Stochastic Collocation methods considered, including cases where the sparse interpolant uses both Clenshaw-Curtis and Gaussian abscissas. This analysis relies on the regularity of the solution and exploits the behavior of sparse approximations from the anisotropic Smolyak method. We also address the case where the input random variables come from suitably truncated expansions of random fields and discuss how the size of the sparse grid can be algebraically related to the number of random variables retained in the expansion in order to have a discretization error of the same order as that of the error due to the truncation of the input random fields.

To demonstrate the effectiveness of the proposed methods in Chapter 7 we present some numerical results validating our theoretical results. We also include a comparison of our proposed methods with other ensemble-based approaches, including Monte Carlo. In Chapter 8 we summarize the significance and achievements of our work.

The complexity of such problems may be decreased further by means of *reduced order modeling* (ROM) in physical space. In Chapter 9 we introduce the concept of a reduced order model for solving related nonlinear SPDEs. It is in this area that we hope to devote much future research. In particular, the coupling of this technique with beforementioned sparse approximation techniques is an important approach which is expected to also be a breakthrough in the area. Finally, in Chapter 10 we offer some concluding remarks and immediate and future directions we intend to investigate.

CHAPTER 1

PROBLEM SETTING

In this chapter we introduce the mathematical problem and the main notation used throughout. Here we also state assumptions on the parameterization of the random inputs, which is useful when later transforming the original stochastic problem into a deterministic parametric one, and on the problem solution's regularity, which is used to later prove the error estimates in Chapter 6.

1.1 Notation and function spaces

We begin by letting $\mathbf{Y} = (Y_1(\omega), Y_2(\omega), \ldots, Y_N(\omega))$ be an $\mathbb{R}^N$-valued random variable in a complete probability space $(\Omega, \mathcal{F}, P)$. Here Ω is the set of outcomes, $\mathcal{F} \subset 2^\Omega$ is the σ-algebra of events and $P : \mathcal{F} \to [0,1]$ is a probability measure. For $q \in [1, \infty)$, let $\left(L^q_P(\Omega)\right)^N$ be the set comprising those random variables $\mathbf{Y}$ with $\sum_{i=1}^{N} \int_\Omega |Y_i(\omega)|^q \, \mathrm{d}P(\omega) < +\infty$. If $Y \in L^1_P(\Omega)$ we denote its expected value by

$$E[Y] = \int_\Omega Y(\omega) \, \mathrm{d}P(\omega) = \int_{\mathbb{R}^N} y \, \mathrm{d}\mu_Y(y),$$

where $\mu_Y(y)$ is the distribution measure for Y, defined for the Borel sets $\widetilde{b} \in B(\mathbb{R}^N)$, by $\mu_Y(\widetilde{b}) \equiv P\left(Y^{-1}(\widetilde{b})\right)$. If μ_Y is absolutely continuous with respect to the Lebesgue measure then there exists a density function $\rho : \mathbb{R} \to [0, \infty)$ such that

$$\int_{\mathbb{R}^N} y\rho(y) \, \mathrm{d}y.$$

Analogously, whenever $\mathbf{Y} \in \left(L^2_P(\Omega)\right)^N$, the positive semi-definite covariance matrix of $\mathbf{Y}$, $\mathrm{cov}[\mathbf{Y}] \in \mathbb{R}^{N \times N}$, is defined by $\mathrm{cov}[\mathbf{Y}](i,j) = \mathrm{cov}(Y_i, Y_j) = E\left[(Y_i - E[Y_i])\,(Y_j - E[Y_j])\right]$ for $i, j = 1, 2, \ldots, N$. Moreover, whenever $u = u(\omega, x)$ is a stochastic function with $\omega \in \Omega$

and $x \in \overline{D}$, its covariance function is defined by $\mathrm{cov}[u] \times (x_1, x_2) = \mathrm{cov}\,(u(\cdot, x_1), u(\cdot, x_2))$ for $x_1, x_2 \in \overline{D}$.

Here, $\mathbb{N}$ will denote the set of the nonnegative integer numbers and $\mathbb{N}_+$ will be used for the set of the positive integer numbers. Let $d \in \mathbb{N}_+$ and D be a bounded domain of $\mathbb{R}^d$. with polygonal boundary ∂D. Denote by $|D| = \int_D 1 \, \mathrm{d}x$ the volume of D, i.e. the Lebesgue measure of D. For $s \in \mathbb{N}$ and $q \in [1, \infty]$, let $\mathscr{W}^{s,q}(D)$ be the Sobolev space of functions having generalized derivatives up to order s in the space $L^q(D)$. Using the standard multi-index notation, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)$ is a d-tuple of non-negative integers, i.e. $\boldsymbol{\gamma} \in \mathbb{N}^d$, and its length is given by $|\boldsymbol{\gamma}| := \sum_{i=1}^d \gamma_i$. The Sobolev norm of $v \in \mathscr{W}^{s,q}(D)$ will be denoted by

$$\|v\|_{\mathscr{W}^{s,q}(D)} \equiv \left\{ \sum_{|\boldsymbol{\gamma}| \leq s} \int_D |\partial^{\boldsymbol{\gamma}} v|^q \, \mathrm{d}x \right\}^{1/q}, \quad 1 \leq q < \infty$$

and

$$\|v\|_{\mathscr{W}^{s,\infty}(D)} \equiv \max_{|\boldsymbol{\gamma}| \leq s} \left(\operatorname*{ess\,sup}_D |\partial^{\boldsymbol{\gamma}} v| \right).$$

We note that the space $\mathscr{W}^{s,2}(D)$ is a Hilbert space, with inner product $(\cdot, \cdot)_{\mathscr{W}^{s,2}(D)}$ defined by

$$(v, w)_{\mathscr{W}^{s,2}(D)} \equiv \sum_{|\boldsymbol{\gamma}| \leq s} (\partial^{\boldsymbol{\gamma}} v, \partial^{\boldsymbol{\gamma}} w)_{L^{2(D)}}.$$

We will write $H^s(D) \equiv \mathscr{W}^{s,2}(D)$, and hence $\|\cdot\|_{H^s(D)} = \|\cdot\|_{\mathscr{W}^{s,2}(D)}$ and $(\cdot, \cdot)_{H^s(D)} = (\cdot, \cdot)_{\mathscr{W}^{s,2}(D)}$. We will use the function space $H_0^1(D)$ which is a subspace of $H^1(D)$, consisting of functions which vanish at the boundary of D in the trace sense, and equipped with the norm $\|v\|_{H_0^1(D)} = \left\{ \int_D |\nabla v|^2 \, \mathrm{d}x \right\}^{1/2}$. Whenever $s = 0$ we will keep the notation $L^q(D)$ instead of $\mathscr{W}^{0,q}(D)$.

Since stochastic functions have intrinsically different structure with respect to ω and with respect to x, the analysis of numerical approximations requires the definition of new functional spaces. For $s \in \mathbb{N}$ and $q \in [1, \infty]$ we define the stochastic Sobolev spaces $\mathscr{L}_P^q \equiv L_P^q(\Omega; \mathscr{W}^{s,q}(D))$ and $\mathscr{L}_P^\infty \equiv L_P^\infty(\Omega; \mathscr{W}^{s,q}(D))$ by:

$$\mathscr{L}_P^q := \left\{ v : \Omega \times D \to \mathbb{R} \mid v \text{ is strongly measurable and } \int_\Omega \|v(\omega, \cdot)\|_{\mathscr{W}^{s,q}(D)}^q \, \mathrm{d}P(\omega) < +\infty \right\}$$

and

$$\mathscr{L}_P^\infty := \left\{ v : \Omega \times D \to \mathbb{R} \mid v \text{ is strongly measurable and } P - \operatorname*{ess\,sup}_{\omega \in \Omega} \|v(\omega, \cdot)\|_{\mathscr{W}^{s,q}(D)}^2 < +\infty \right\}.$$

We now focus our attention on an elliptic operator $\mathcal{L}$, linear or nonlinear, on a domain $D \subset \mathbb{R}^d$, which depends on some coefficients $a = a(\omega, x)$ with $x \in D$, $\omega \in \Omega$. Similarly, the forcing term $f = f(\omega, x)$ can be assumed random as well.

Consider the stochastic elliptic boundary value problem: find a random function, $u : \Omega \times \overline{D} \to \mathbb{R}$, such that P-almost everywhere in Ω, or in other words almost surely (a.s.), the following equation holds:

$$\mathcal{L}(a)(u) = f \quad \text{in } D \tag{1.1.1}$$

equipped with suitable boundary conditions. Before introducing some assumptions we denote by $W(D)$ a Banach space of functions $v : D \to \mathbb{R}$ and analogously to the above construction we define, for $q \in [1, \infty]$, the stochastic Banach spaces $L^q_P \equiv L^q_P(\Omega; W(D))$ and $L^\infty_P \equiv L^\infty_P(\Omega; W(D))$ by:

$$L^q_P := \left\{ v : \Omega \to W(D) \mid v \text{ is strongly measurable and } \int_\Omega \|v(\omega, \cdot)\|^q_{W(D)} dP(\omega) < +\infty \right\}$$

and

$$L^\infty_P := \left\{ v : \Omega \to W(D) \mid v \text{ is strongly measurable and } P - \operatorname{ess\,sup}_{\omega \in \Omega} \|v(\omega, \cdot)\|^2_{W(D)} < +\infty \right\}.$$

Of particular interest is the space $L^2_P(\Omega; W(D))$, consisting of Banach valued functions that have finite second moments.

We will now make the following assumptions:

A_1) the solution to (1.1.1) has realizations in the Banach space $W(D)$, i.e. $u(\cdot, \omega) \in W(D)$ almost surely and $\forall \omega \in \Omega$

$$\|u(\cdot, \omega)\|_{W(D)} \leq C \|f(\cdot, \omega)\|_{W^*(D)},$$

where we denote $W^*(D)$ to be the dual space of $W(D)$, and C is a constant independent of the realization $\omega \in \Omega$.

A_2) the forcing term $f \in L^2_P(\Omega; W^*(D))$ is such that the solution u is unique and bounded in $L^2_P(\Omega; W(D))$.

Here we give two example problems that are posed in this setting:

Example 1.1.1 *The linear problem*

$$\begin{cases} -\nabla \cdot (a(\omega,\cdot)\nabla u(\omega,\cdot)) &= f(\omega,\cdot) \quad \text{in } \Omega \times D, \\ u(\omega,\cdot) &= 0 \quad \text{on } \Omega \times \partial D, \end{cases} \tag{1.1.2}$$

with $a(\omega,\cdot)$ uniformly bounded and coercive, i.e.

$$\text{there exists } a_{min}, a_{max} \in (0,+\infty) \text{ such that } P(\omega \in \Omega : \ a(\omega,x) \in [a_{min}, a_{max}]\, \forall x \in \overline{D}) = 1$$

and $f(\omega,\cdot)$ square integrable with respect to P, i.e.

$$\int_D E\left[f^2\right] \ dx < +\infty,$$

satisfies assumptions A_1 and A_2 with $W(D) = H_0^1(D)$ (see Chapter 2 and Babuška et al. (2005c)).

Example 1.1.2 *Similarly, for $k \in \mathbb{N}^+$, the nonlinear problem*

$$\begin{cases} -\nabla \cdot (a(\omega,\cdot)\nabla u(\omega,\cdot)) + u(\omega,\cdot)|u(\omega,\cdot)|^k &= f(\omega,\cdot) \quad \text{in } \Omega \times D, \\ u(\omega,\cdot) &= 0 \quad \text{on } \Omega \times \partial D, \end{cases} \tag{1.1.3}$$

with $a(\omega,\cdot)$ uniformly bounded and coercive and $f(\omega,\cdot)$ square integrable with respect to P, satisfies assumptions A_1 and A_2 with $W(D) = H_0^1(D) \cap L^{k+2}(D)$ (see Chapter 3 and Babuška et al. (2007)).

1.2 On finite dimensional noise

In some applications, the coefficient a and the forcing term f appearing in (1.1.1) can be described by a random vector $\mathbf{Y} = [Y_1, \ldots, Y_N] : \Omega \to \mathbb{R}^N$, as in the following examples. In such cases, we will emphasize such dependence by writing a_N and f_N.

Example 1.2.1 (Piecewise constant random fields) *Let us consider again problem (1.1.2) where the physical domain D is the union of non-overlapping subdomains D_i, $i = 1, \ldots, N$. We consider a diffusion coefficient that is piecewise constant and random on each subdomain, i.e.*

$$a_N(\omega, x) = a_{min} + \sum_{i=1}^{N} \sigma_i \, Y_i(\omega) 1_{D_i}(x).$$

Here 1_{D_i} is the indicator function of the set D_i, σ_i, a_{min} are positive constants, and the random variables Y_i are nonnegative with unit variance.

In other applications the coefficients and forcing terms in (1.1.1) may have other types of spatial variation that is amenable to describe by an expansion. Depending on the decay of such expansion and the desired accuracy in our computations we may retain just the first N terms.

Example 1.2.2 (Karhunen-Loève expansion) *We recall that any second order random field $g(\omega, x)$, with continuous covariance function $cov[g] : \overline{D \times D} \to \mathbb{R}$, can be represented as an infinite sum of random variables, by means, for instance, of a Karhunen-Loève expansion (Loève, 1977). To this end, introduce the compact and self-adjoint operator $T_g : L^2(D) \to L^2(D)$, which is defined by*

$$T_g v(\cdot) := \int_D cov[g](x, \cdot)\, v(x)\ dx \quad \forall v \in L^2(D).$$

Then, for mutually uncorrelated real random variables $\{Y_i(\omega)\}_{i=1}^{\infty}$ with zero mean and unit variance, i.e. $E[Y_i] = 0$ and $E[Y_i Y_j] = \delta_{ij}$ for $i, j \in \mathbb{N}_+$, the expansion is given by

$$g(\omega, x) = E[g](x) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\, b_i(x)\, Y_i(\omega)$$

where $\{\lambda_i\}_{i=1}^{\infty}$ is a sequence of non-negative decreasing eigenvalues of T_g and $\{b_i\}_{i=1}^{\infty}$ the corresponding sequence of orthonormal eigenfunctions satisfying

$$T_g b_i = \lambda_i b_i, \quad (b_i, b_j)_{L^2(D)} = \delta_{ij} \ \text{ for } \ i, j \in \mathbb{N}_+.$$

In addition, for $i = 1, \ldots,$ the mutually uncorrelated real random variables are uniquely determined by

$$Y_i(\omega) := \frac{1}{\sqrt{\lambda_i}} \int_D (g(\omega, x) - E[g](x))\, b_i(x) dx.$$

The truncated Karhunen-Loève expansion g_N, of the stochastic function g, is defined by

$$g_N(\omega, x) = E[g](x) + \sum_{i=1}^{N} \sqrt{\lambda_i}\, b_i(x)\, Y_i(\omega) \quad \forall N \in \mathbb{N}_+.$$

Then by Mercer's theorem (cf. (Riesz and Sz-Nagy, 1990, p. 245)) it follows that

$$\lim_{N\to\infty} \left\{ \sup_D E\left[(g - g_N)^2\right] \right\} = \lim_{N\to\infty} \left\{ \sup_D \left(\sum_{i=N+1}^{\infty} \lambda_i b_i^2 \right) \right\} = 0.$$

Observe that the N random variables in (1.2.2), describing the random data, are then weighted differently due to the decay of the eigen-pairs of the Karhunen-Loève expansion.

The decay of eigenvalues and eigenvectors has been investigated e.g. in the works Frauenfelder et al. (2005b) and Todor (2005).

The above examples motivate us to consider problems whose coefficients are described by finitely many random variables. Thus, we will seek a random field $u_N : \Omega \times \overline{D} \to \mathbb{R}$, such that a.s., the following equation holds:

$$\mathcal{L}(a_N)(u_N) = f_N \quad \text{in } D, \tag{1.2.1}$$

We assume that equation (1.2.1) admits a unique solution $u_N \in L^2_P(\Omega; W(D))$. We then have, by the Doob–Dynkin's lemma (cf. Øksendal (2003)), that the solution u_N of the stochastic elliptic boundary value problem (1.2.1) can be described by $u_N = u_N(\omega, x) = u_N(Y_1(\omega), \ldots, Y_N(\omega), x)$. We underline that the coefficients a_N and f_N in (1.2.1) may be an exact representation of the input data as in Example 1.2.1 or a suitable truncation of the input data as in Example 1.2.2. In the latter case, the solution u_N will also be an approximation of the exact solution u in (1.1.1) and the truncation error $u - u_N$ has to be properly estimated, see for instance section 6.1.3 and the work (Nobile et al., 2006, Section 4.2).

Remark 1.2.3 (Nonlinear coefficients) *In certain cases, one may need to ensure qualitative properties on the coefficients a_N and f_N and it may be worth while to describe them as nonlinear functions of $\mathbf{Y}$. For instance, in Example 1.1.1 one is required to enforce positiveness on the coefficient $a_N(\omega, x)$, say $a_N(\omega, x) \geq a_{min}$ for all $x \in D$, a.s. in Ω. Then a better choice is to expand $\log(a_N - a_{min})$. The following standard transformation guarantees that the diffusivity coefficient is bounded away from zero almost surely*

$$\log(a_N - a_{min})(\omega, x) = b_0(x) + \sum_{1 \leq n \leq N} \sqrt{\lambda_n} b_n(x) Y_n(\omega), \tag{1.2.2}$$

i.e. one performs a Karhunen-Loève expansion for $\log(a_N - a_{min})$, assuming that $a_N > a_{min}$ almost surely. On the other hand, the right hand side of (1.2.1) can be represented as a truncated Karhunen-Loève expansion

$$f_N(\omega, x) = c_0(x) + \sum_{1 \leq n \leq N} \sqrt{\mu_n} c_n(x) Y_n(\omega).$$

Remark 1.2.4 *It is usual to have f_N and a_N independent, because the forcing terms and the parameters in the operator $\mathcal{L}$ are seldom related. In such a situation we have $a_N(\mathbf{Y}(\omega), x) = a_N(\mathbf{Y}_a(\omega), x)$ and $f_N(\mathbf{Y}(\omega), x) = f_N(\mathbf{Y}_f(\omega), x)$, with $\mathbf{Y} = [\mathbf{Y}_a, \mathbf{Y}_f]$ and the vectors $\mathbf{Y}_a$, $\mathbf{Y}_f$ are independent.*

For this work we denote $\Gamma_n \equiv Y_n(\Omega)$ the image of Y_n, where we assume $Y_n(\omega)$ to be bounded. Without loss of generality we can assume $\Gamma_n = [-1, 1]$. We also let $\Gamma^N = \prod_{n=1}^N \Gamma_n$ and assume that the random variables $[Y_1, Y_2, \ldots, Y_n]$ have a joint probability density function $\rho : \Gamma^N \to \mathbb{R}_+$, with $\rho \in L^\infty(\Gamma^N)$. Thus, the goal is to approximate the function $u_N = u_N(y, x)$, for any $y \in \Gamma^N$ and $x \in \overline{D}$. See also the works Babuška et al. (2005c) and Babuška et al. (2004b).

Remark 1.2.5 (Unbounded Random Variables) *By using a similar approach to the work Babuška et al. (2005c) we can easily deal with unbounded random variables, such as Gaussian or exponential ones. For the sake of simplicity in the presentation we focus our study on bounded random variables only.*

1.3 Regularity

Before discussing various collocation techniques and going through the convergence analysis of such methods, we need to state some regularity assumptions on the data of the problem and consequent regularity results for the exact solution u_N. We will perform a one-dimensional analysis in each direction y_n, $n = 1, \ldots, N$. For this, we introduce the following notation: $\Gamma_n^* = \prod_{\substack{j=1 \\ j \neq n}}^N \Gamma_j$, $\mathbf{y}_n^*$ will denote an arbitrary element of Γ_n^*. We require the solution to problem (1.1.1) to satisfy the following estimate:

Assumption 1.3.1 *For each $y_n \in \Gamma_n$, there exists $\tau_n > 0$ such that the function $u_N(y_n, \mathbf{y}_n^*, x)$ as a function of y_n, $u_N : \Gamma_n \to C^0(\Gamma_n^*; W(D))$ admits an analytic extension $u(z, \mathbf{y}_n^*, x)$, $z \in \mathbb{C}$, in the region of the complex plane*

$$\Sigma(\Gamma_n; \tau_n) \equiv \{z \in \mathbb{C}, \quad dist(z, \Gamma_n) \leq \tau_n\}. \tag{1.3.1}$$

Moreover, $\forall z \in \Sigma(\Gamma_n; \tau_n)$,

$$\|u_N(z)\|_{C^0(\Gamma_n^*; W(D))} \leq \lambda \tag{1.3.2}$$

with λ a constant independent of n.

The previous assumption should be verified for each particular application. In particular, this has implications on the allowed regularity of the input data, e.g. coefficients, loads, etc., with respect to $y_n, n = 1, 2, \ldots, N$, of the SPDE under study. In the next Chapter we recall some theoretical results and validate Assumption 1.3.1, which was shown in (Babuška et al., 2005c, Section 3), for the linear problem introduced in Example 1.1.1. The more difficult SPDE is analyzed in Chapter 3 where we describe similar theoretical results for the nonlinear problem described in Example 1.1.3.

CHAPTER 2

LINEAR ELLIPTIC PDES WITH RANDOM INPUT DATA

In this chapter we give more details concerning the linear problem described in Example 1.1.1. The main ideas here are to exhibit the well-posedness of the PDEs under study and to verify the assumptions from Chapter 1 in a particular setting, showing that they are justified and that one can extend them to a variety of problems. To begin, we describe the following stochastic Hilbert spaces which were defined in Chapter 1:

- $V_P = L^2_P(\Omega; H^1_0(D))$ equipped with the norm

$$\|v\|_P^2 = \int_D E\left[|\nabla v|^2\right] dx = E\left[\|\nabla v\|^2_{L^2(D)}\right] = E\left[\|v\|^2_{H^1_0(D)}\right]$$

- $V_{P,a} := \left\{v \in V_P \,:\, \int_D E\left[a|\nabla v|^2\right] dx < +\infty\right\}$, equipped with the norm

$$\|v\|^2_{P,a} = \int_D E\left[a|\nabla v|^2\right] dx.$$

Observe that under the assumptions of Example 1.1.1., the space $V_{P,a}$ is continuously embedded in V_P and

$$\|v\|_P \leq \frac{1}{\sqrt{a_{min}}}\|v\|_{P,a}.$$

Moreover, this embedding allows us to control the solution u of problem (1.1.2) in the space V_P.

Problem (1.1.2) can be written in a weak form as: find $u \in V_{P,a}$ such that

$$\int_D E[a\nabla u \cdot \nabla v]\, dx = \int_D E[fv]\, dx \quad \forall\, v \in V_{P,a}. \tag{2.0.1}$$

A straightforward application of the Lax-Milgram theorem allows one to state the well posedness of problem (2.0.1); precisely:

Lemma 2.0.2 *Under the assumptions defined in Example 1.1.1, problem (2.0.1) admits an unique solution $u \in V_{P,a}$. Moreover, the following a priori estimates hold*

$$\|u\|_{H_0^1(D)} \leq \frac{C_P}{a_{min}} \|f(\omega, \cdot)\|_{L^2(D)} \qquad a.s. \tag{2.0.2}$$

and

$$\|u\|_{L^2_P(\Omega; H_0^1(D))} \leq \frac{C_P}{a_{min}} \left(\int_D E[f^2]\, dx \right)^{1/2}. \tag{2.0.3}$$

In the previous Lemma C_P denotes the constant appearing in the Poincaré inequality:

$$\|w\|_{L^2(D)} \leq C_P \|\nabla w\|_{L^2(D)} \qquad \forall w \in H_0^1(D).$$

Once we have the input random fields described by a finite set of random variables, i.e. $a(\omega, x) = a_N(Y_1(\omega), \ldots, Y_N(\omega), x)$, and similarly for $f(\omega, x)$, the "finite dimensional" version of the stochastic variational formulation (2.0.1) has a "deterministic" equivalent which is the following: find $u_N \in L^2_\rho(\Gamma^N; H_0^1(D))$ such that

$$\int_{\Gamma^N} \rho\, (a_N \nabla u_N, \nabla v)_{L^2(D)}\, dy = \int_{\Gamma^N} \rho\, (f_N, v)_{L^2(D)}\, dy, \quad \forall\, v \in L^2_\rho(\Gamma^N; H_0^1(D)). \tag{2.0.4}$$

Observe that in this work the gradient notation, ∇, always means differentiation with respect to $x \in D$ only, unless otherwise stated. The stochastic boundary value problem (2.0.1) now becomes a deterministic Dirichlet boundary value problem for an elliptic partial differential equation with an $N-$dimensional parameter. For convenience, we consider the solution u_N as a function $u_N : \Gamma^N \to H_0^1(D)$ and we use the notation $u_N(y)$ whenever we want to highlight the dependence on the parameter y. We use similar notations for the coefficient a_N and the forcing term f_N. Then, it can be shown that problem (2.0.1) is equivalent to

$$\int_D a_N(y) \nabla u_N(y) \cdot \nabla \phi\, dx = \int_D f_N(y) \phi\, dx, \quad \forall \phi \in H_0^1(D), \;\; \rho\text{-a.e. in } \Gamma^N. \tag{2.0.5}$$

For our convenience, we will suppose that the coefficient a_N and the forcing term f_N admit a smooth extension on the ρ-zero measure sets. Then, equation (2.0.5) can be extended a.e. in Γ^N with respect to the Lebesgue measure (instead of the measure ρdy).

It has been proved in Babuška et al. (2005c) that problem (2.0.5) satisfies the analyticity result stated in Assumption 1.3.1. For instance, if we take the diffusivity coefficient as in Example 1.2.1 and a deterministic load, the size of the analyticity region is given by

$$\tau_n = \frac{a_{min}}{4\sigma_n}. \tag{2.0.6}$$

On the other hand, if we take the diffusivity coefficient as a truncated expansion like in Remark 1.2.3, then the analyticity region $\Sigma(\Gamma_n; \tau_n)$ is given by

$$\tau_n = \frac{1}{4\sqrt{\lambda_n}\|b_n\|_{L^\infty(D)}} \tag{2.0.7}$$

Observe that, in the latter case, as $\sqrt{\lambda_n}\|b_n\|_{L^\infty(D)} \to 0$ for a regular enough covariance function, see Frauenfelder et al. (2005b), the analyticity region increases as n increases. This fact introduces, naturally, an anisotropic behavior with respect to the "direction" n. This effect will be exploited in the numerical methods proposed in the next sections and is the subject of ongoing research. In the next Chapter we will extend the theoretical results for the nonlinear elliptic SPDE described by Example 1.1.2 and explore the complications arising when analyzing such problems.

CHAPTER 3

NONLINEAR ELLIPTIC PDES WITH RANDOM INPUT DATA

In this Chapter we focus our attention to the nonlinear problem described in Example 1.1.2. Similar to the previous Chapter, the goal here is to first thoroughly examine the theoretical aspects of the continuous problem, where we exhibit the well-posedness of the nonlinear SPDE, and to second, verify the assumptions from Chapter 1. Similar to the linear problem described by Example 1.1.1 and analyzed in Chapter 2 we extend otherwise deterministic techniques to assist in analyzing these complicated SPDEs.

We open this chapter introducing the following Banach spaces whose definitions can be found in Chapter 1, where we recall the stochastic Hilbert spaces defined in Chapter 2:

- $\widetilde{V}_P = V_P \bigcap L_P^{k+2}\left(\Omega;\, L^{k+2}(D)\right)$ equipped with the norm

$$\begin{aligned}\|v\|_{\widetilde{P}} &= \left(\int_D E\left[|\nabla v|^2\right] dx\right)^{1/2} + \left(\int_D E\left[|v|^{k+2}\right] dx\right)^{1/(k+2)} \\ &= \|v\|_P + \left(E\left[\|v\|^{k+2}_{L^{k+2}(D)}\right]\right)^{1/(k+2)}\end{aligned}$$

- $\widetilde{V}_{P,a} := \left\{v \in \widetilde{V}_P \,:\, \|v\|_{P,a} + E\left[\|v\|^{k+2}_{L^{k+2}(D)}\right] < +\infty\right\}$, equipped with the norm

$$\|v\|_{\widetilde{P},a} = \|v\|_{P,a} + \left(E\left[\|v\|^{k+2}_{L^{k+2}(D)}\right]\right)^{1/(k+2)}.$$

Similarly to Chapter 2, observe that under the assumptions of Example 1.1.3 the space $\widetilde{V}_{P,a}$ is continuously embedded in $\widetilde{V}_P$. Finally, we also acquire the embedding $\widetilde{V}_P$ in V_P. Again, this allows us to control the solution u of problem (1.1.3) in the space V_P.

We begin the analysis with a simple lemma deduced from Young's inequality:

Lemma 3.0.3 *For some constant $C(k) > 0$ and $v \in \widetilde{V}_P$, the norm $\|v\|_{\widetilde{P}}$ satisfies the following inequality:*

$$\|v\|_{\widetilde{P}}^2 \leq 2\left[E\left(\|\nabla v\|_{L^2(D)}^2\right) + E\left(\| v\|_{L^{k+2}(D)}^{k+2}\right) + C(k)\right].$$

Proof. The proof follows as a direct application of Young's inequality:

$$\begin{aligned}
\|v\|_{\widetilde{P}}^2 &= \left[\left(\int_D E\left[|\nabla v|^2\right] dx\right)^{1/2} + \left(\int_D E\left[|v|^{k+2}\right] dx\right)^{1/(k+2)}\right]^2 \\
&\leq 2\left[E\left[\|\nabla v\|_{L^2(D)}^2\right] + \left(E\left[\| v\|_{L^{k+2}(D)}^{k+2}\right]\right)^{2/(k+1)}\right] \\
&\leq 2\left[E\left[\|\nabla v\|_{L^2(D)}^2\right] + E\left[\| v\|_{L^{k+2}(D)}^{k+2}\right] + C(k)\right]
\end{aligned}$$

where we define

$$C(k) = \frac{k}{k+2}\left(\frac{2}{k+2}\right)^{1/k} \tag{3.0.1}$$

for $k \in \mathbb{N}^+$. □

Similarly to Evans (1998) we establish the global existence and uniqueness of a weak solution to the nonlinear elliptic stochastic partial differential equation (1.1.3). To begin, we consider the *Lagrangian* function

$$\mathscr{L} : \mathbb{R}^d \times \mathbb{R} \times \overline{D} \to \mathbb{R},$$

defined by

$$\mathscr{L}(\nabla u, u, x) = E\left[L(\nabla u(\omega, \cdot), u(\omega, \cdot), x, \omega)\right], \tag{3.0.2}$$

where

$$L(\nabla u(\omega, x), u(\omega, x), x, \omega) = \frac{1}{2}\, a|\nabla u(\omega, x)|^2 + \frac{1}{k+2}\,|u(\omega, x)|^{k+2} - f(\omega, x)\, u(\omega, x), \tag{3.0.3}$$

and $\mathscr{F} : \widetilde{V}_{P,a} \to \mathbb{R}$ by

$$\mathscr{F}(u) = \int_D \mathscr{L}(\nabla u, u, x)\, dx. \tag{3.0.4}$$

In what follows we identify conditions on the Lagrangian $\mathscr{L}$ which ensure that the functional $\mathscr{F}$ does indeed have a minimizer, at least within the appropriate Banach space $\widetilde{V}_{P,a}$.

3.1 Coercivity, lower semicontinuity

In general we will require some hypothesis controlling $\mathscr{F}(u)$ for "large" functions u. Certainly the most effective way to ensure this would be to hypothesize that $\mathscr{F}(u)$ "grows rapidly as $|u| \to \infty$". More specifically we have the following lemma:

Lemma 3.1.1 (Coercivity Condition) *Assume that $\mathscr{L}$ is defined by (3.0.2). Then there exists constants $\epsilon, \delta > 0$ such that*

$$\mathscr{F}(u) \geq \epsilon \, \|u\|_{\widetilde{P}}^2 - \delta. \tag{3.1.1}$$

Futhermore, $\mathscr{F}(u) \to \infty$ as $\|u\|_{\widetilde{P}} \to \infty$. It is customary to call (3.1.1) a **coercivity condition** *on $\mathscr{F}(u)$.*

Proof. The proof of the coercivity condition (3.1.1) follows from Lemma 3.0.3 and the Cauchy-Schwarz, Poincaré and Young inequalities. First we define the constant

$$\widetilde{C}(k) = \min\left\{\frac{a_{min}}{4}, \frac{1}{k+2}\right\}$$

then from (3.0.4) we get

$$\begin{aligned}
\mathscr{F}(u) \geq\ & a_{min}\, E\left(\|\nabla u\|_{L^2(D)}^2\right) + \frac{1}{k+2}\, E\left(\|\ u\|_{L^{k+2}(D)}^{k+2}\right) - E\left(\|u\|_{L^2(D)}\ \|f\|_{L^2(D)}\right) \\
\geq\ & \frac{a_{min}}{2}\, E\left(\|\nabla u\|_{L^2(D)}^2\right) + \frac{1}{k+2}\, E\left(\|\ u\|_{L^{k+2}(D)}^{k+2}\right) - \frac{C_P}{a_{min}}\, E\left(\|f\|_{L^2(D)}^2\right) \\
& - \frac{a_{min}}{4C_P}\, E\left(\|u\|_{L^2(D)}^2\right) \\
\geq\ & \frac{a_{min}}{2}\, E\left(\|\nabla u\|_{L^2(D)}^2\right) + \frac{1}{k+2}\, E\left(\|\ u\|_{L^{k+2}(D)}^{k+2}\right) - \frac{C_P}{a_{min}}\, E\left(\|f\|_{L^2(D)}^2\right) \\
& - \frac{a_{min}}{4}\, E\left(\|\nabla u\|_{L^2(D)}^2\right) \\
=\ & \frac{a_{min}}{4}\, E\left(\|\nabla u\|_{L^2(D)}^2\right) + \frac{1}{k+2}\, E\left(\|\ u\|_{L^{k+2}(D)}^{k+2}\right) - \frac{C_P}{a_{min}}\, E\left(\|f\|_{L^2(D)}^2\right) \\
\geq\ & \widetilde{C}(k)\left[E\left(\|\nabla u\|_{L^2(D)}^2\right) + E\left(\|\ u\|_{L^{k+2}(D)}^{k+2}\right)\right] - \frac{C_P}{a_{min}}\, E\left(\|f\|_{L^2(D)}^2\right) \\
\geq\ & \frac{\widetilde{C}(k)}{2}\, \|u\|_{\widetilde{P}}^2 - \frac{C_P}{a_{min}}\, E\left(\|f\|_{L^2(D)}^2\right) + 2\, C(k) \\
\geq\ & \epsilon\, \|u\|_{\widetilde{P}}^2 - \delta,
\end{aligned}$$

where we take

$$\epsilon = \frac{\widetilde{C}(k)}{2} \text{ and } \delta = \frac{C_P}{a_{min}}\, E\left(\|f\|_{L^2(D)}^2\right) - 2\, C(k)$$

and $C(k)$ is defined by (3.0.1). $\square$

We next turn our attention to the basic task of finding minimizers for the functional $\mathscr{F}$. First let us remind the reader of some useful definitions and quote a few theorems that will prove to be very beneficial to our discussion. We begin with the notion of a function being *weakly lower semicontinuous.*

Definition 3.1.2 *We say that a function $\mathscr{F}$ is (sequentially)* **weakly lower semicontinuous** *on $\widetilde{V}_{P,a}$, provided*

$$\mathscr{F}(u) \leq \liminf_{k\to\infty} \mathscr{F}(u_k)$$

whenever

$$u_k \rightharpoonup u \text{ weakly in } \widetilde{V}_{P,a}$$

Our goal therefore is to now identify reasonable conditions on the nonlinearity $\mathscr{L}$ that ensure $\mathscr{F}(\cdot)$ is weakly lower semicontinuous. It is important to have a large supply of weakly lower semicontinuous functions. The following is often convenient:

Definition 3.1.3 *A function $\mathscr{F}$ on $\widetilde{V}_{P,a}$ is called* **convex** *if*

$$\mathscr{F}(\lambda x + (1-\lambda)y) \leq \lambda\mathscr{F}(x) + (1-\lambda)\mathscr{F}(y)$$

for all $x, y \in \widetilde{V}_{P,a}$ and $\lambda \in [0,1]$.

We then have the following Lemma which will aid in showing $\mathscr{F}(\cdot)$ is weakly lower semicontinuous on $\widetilde{V}_{P,a}$.

Lemma 3.1.4 *The Lagrangian $\mathscr{L}$, defined by (3.0.2), is smooth, bounded below, and in addition*

$$\text{the mapping } (p,u) \mapsto \mathscr{L}(p,u,x) \text{ is strictly convex,}$$

for each $x \in D$. Therefore, $\mathscr{F}$ is strictly convex on $\widetilde{V}_{P,a}$.

Proof. The fact the $\mathscr{L}$ is smooth is trivial to see and the boundedness follows directly from Lemma 3.1.1. To see the mapping $(p,u) \mapsto \mathscr{L}(p,u,x)$ is strictly convex recall that the sum of convex functions is also convex and thus, only one term is required to be strictly convex to enforce the same in the resulting function. It follows immediately from (3.0.2) and (3.0.3) that the second and final terms of $\mathscr{L}$ are convex, but not "strictly so". Therefore, one is only

required to show that the mapping $p \mapsto \mathscr{L}(p, u, x)$ is strictly convex to achieve the desired result. To do so simply define the function $f : \mathbb{R} \to \mathbb{R}$ by $f(p) = p^2$. Taking $\lambda \in [0, 1]$ and following Definition (3.1.3) we get that

$$\begin{aligned} f(\lambda x + (1-\lambda)y) - \lambda\, f(x) - (1-\lambda)\, f(y) &= (\lambda^2 - \lambda)x^2 + 2\lambda(1-\lambda)xy \\ &\quad + ((1-\lambda)^2 - (1-\lambda))y^2 \\ &= \lambda(\lambda-1)(x^2 + 2xy + y^2) \\ &= \lambda(\lambda-1)(x+y)^2 > 0. \end{aligned}$$

Then replacing p with ∇u and taking expectations the result follows. □

The convexity in the first argument is fundamental when trying to exhibit the existence of a minimizer. First we will require $\mathscr{F}$ to be weakly lower semicontinuous on $\widetilde{V}_{P,a}$ which leads us to the following general result, whose proof can be found in (Reed and Simon, 1990, p.356), pertaining to weakly lower semicontinuous functions on Banach spaces:

Proposition 3.1.5 *Let $\mathcal{F}$ be a function from a Banach space B to $\mathbb{R}$. Suppose that $\mathcal{F}$ is convex on B, and that $\mathcal{F}$ is norm lower semicontinuous. Then $\mathcal{F}$ is weakly lower semicontinuous.*

Since our goal is to demonstrate that there exists a minimizer of $\mathscr{F}$ in $\widetilde{V}_{P,a}$, then the prerequisite will be that $\mathscr{F}$ is weakly lower semicontinuous. With Proposition 3.1.5 we proceed and arrive at the following theorem.

Theorem 3.1.6 *Assuming that $\mathscr{L}$ satisfies the conditions of Lemma 3.1.4, then the function $\mathscr{F}$ defined by (3.0.4) is weakly lower semicontinuous on $\widetilde{V}_{P,a}$.*

Proof. The last term is obviously weakly continuous and the second term is weakly lower semicontinuous by Proposition 3.1.5. As for the first term, let $u_n \rightharpoonup u$ weakly; then

$$\begin{aligned} \int_D a|\nabla u|^2\, dx &= \sup\left\{ \left|(f, a\,\nabla u)_{L^2(D)}\right| : f \in C_0^\infty(D),\ \|f\|_{L^2(D)} = 1 \right\} \\ &= \sup\left\{ \left|(\nabla\cdot(af), u)_{L^2(D)}\right| : f \in C_0^\infty(D),\ \|f\|_{L^2(D)} = 1 \right\} \\ &\le \sup\left\{ \lim_{n\to\infty} \left|(\nabla\cdot(af), u_n)_{L^2(D)}\right| : f \in C_0^\infty(D),\ \|f\|_{L^2(D)} = 1 \right\} \\ &= \sup\left\{ \lim_{n\to\infty} \left|(f, a\nabla u_n)_{L^2(D)}\right| : f \in C_0^\infty(D),\ \|f\|_{L^2(D)} = 1 \right\} \\ &\le \liminf_{n\to\infty} \int_D a|\nabla u_n|^2\, dx. \end{aligned}$$

Then taking the expected value yields

$$\int_D E\left[a|\nabla u|^2\right]\,dx \leq \liminf_{n\to\infty} \int_D E\left[a|\nabla u_n|^2\right]\,dx.$$

We can now conclude that $\mathscr{F}(\cdot)$ is weakly lower semicontinuous. □

3.2 Existence and uniqueness of minimizers

We can at last establish that $\mathscr{F}$ has a minimizer among all possible functions in $\widetilde{V}_{P,a}$.

Theorem 3.2.1 (Existence of a minimizer) *Assume that $\mathscr{L}$ satisfies the conditions of Lemma 3.1.4 and the coercivity condition (3.1.1), then there exists at least one function $u \in \widetilde{V}_{P,a}$ satisfying*

$$\mathscr{F}(u) = \min_{w\in\widetilde{V}_{P,a}} \mathscr{F}(w).$$

Proof. Set $m := \inf_{w\in\widetilde{V}_{P,a}} \mathscr{F}(w)$. If $m = +\infty$ we are done, and so we henceforth assume m is finite. Select a minimizing sequence $\{u_k\}_{k=1}^{\infty}$. Then

$$\mathscr{F}(u_k) \to m \text{ as } k \to \infty. \tag{3.2.1}$$

We would like to show that some subsequence of $\{u_k\}_{k=1}^{\infty}$ converges to an actual minimizer. From (3.2.1) we know that, for all k

$$m \leq \mathscr{F}(u_k) < m + \frac{1}{k} \leq m + 1.$$

The coercivity condition (3.1.1) guarantees that

$$\mathscr{F}(u_k) \geq \epsilon \left\|u_k\right\|_{\widetilde{P}}^2 - \delta,$$

and hence, we arrive at the following inequality

$$\left\|u_k\right\|_{\widetilde{P}}^2 \leq \frac{1}{\epsilon}\left(m + 1 + \delta\right). \tag{3.2.2}$$

Since m is finite, we conclude from (3.2.1) and (3.2.2) that

$$\sup_k \left\|u_k\right\|_{\widetilde{P}} < \infty.$$

This estimate ensures that $\{u_k\}_{k=1}^{\infty}$ is bounded in $\widetilde{V}_P$. Consequently there exists a subsequence $\{u_{k_j}\}_{k=1}^{\infty} \subset \{u_k\}_{k=1}^{\infty}$ and a function $u \in \widetilde{V}_P$ such that

$$u_{k_j} \rightharpoonup u \text{ weakly in } \widetilde{V}_P.$$

We assert next that $u \in \widetilde{V}_{P,a}$. In view of Theorem 3.1.6 and the coercivity condition (3.1.1) we have for some constant $C_a > 0$,

$$m = \liminf_{j\to\infty} \mathscr{F}(u_{k_j}) \geq \mathscr{F}(u) \geq C_a \|u\|_{\widetilde{P},a} - E\left(\|u\|_{L^2(D)}\ \|f\|_{L^2(D)}\right)$$
$$\geq C_a \|u\|_{\widetilde{P},a} - \frac{C_P}{a_{min}} E\left(\|f\|^2_{L^2(D)}\right) - \frac{a_{min}}{4C_P} E\left(\|u\|^2_{L^2(D)}\right).$$

Therfore, it is easy to see that

$$\|u\|_{\widetilde{P},a} \leq \frac{1}{C_a}\left(m + \frac{C_P}{a_{min}} E\left(\|f\|^2_{L^2(D)}\right) + \frac{a_{min}}{4C_P} E\left(\|u\|^2_{L^2(D)}\right)\right) < +\infty.$$

Since $u \in \widetilde{V}_{P,a}$ then it follows that

$$\mathscr{F}(u) = m = \min_{w\in\widetilde{V}_{P,a}} \mathscr{F}(w).$$

□

We turn next to the problem of uniqueness. In general there can be many minimizers, but Lemma 3.1.4 ensures that this does not happen here, as will be shown next.

Theorem 3.2.2 (Uniqueness of minimizer) *Assume that $\mathscr{L}$ satisfies the conditions of Lemma 3.1.4. Then a minimizer $u \in \widetilde{V}_{P,a}$ of $\mathscr{F}(\cdot)$ is unique.*

Proof. Assume $u, \widetilde{u} \in \widetilde{V}_{P,a}$ are both minimizers of $\mathscr{F}(\cdot)$ over $\widetilde{V}_{P,a}$. Then $v := \frac{u+\widetilde{u}}{2} \in \widetilde{V}_{P,a}$. Following from the strict convexity proved in Lemma 3.1.4

$$\mathscr{F}(v) = \mathscr{F}\left(\frac{u+\widetilde{u}}{2}\right) < \frac{1}{2}\mathscr{F}(u) + \frac{1}{2}\mathscr{F}(\widetilde{u}).$$

Hence, from Theorem 3.2.1 we get that $m = \inf_{w\in\widetilde{V}_{P,a}} \mathscr{F}(w) = \mathscr{F}(u) = \mathscr{F}(\widetilde{u})$ which implies

$$\mathscr{F}(v) < \frac{m}{2} + \frac{m}{2} = m.$$

Therefore, $u = \widetilde{u}$ for a.e. $x \in D$ and for a.s. $\omega \in \Omega$. □

3.3 Weak solutions of a nonlinear stochastic PDE

We wish next to demonstrate that any minimizer $u \in \widetilde{V}_{P,a}$ of $\mathscr{F}$ solves the nonlinear SPDE (1.1.3) in some suitable sense. Conversely, since the joint map $(p, u) \mapsto \mathscr{L}(p, u, x)$ is convex for each x then each such solution of (1.1.3) is in fact a minimizer. This does not follow from

direct calculations since we only know that $u \in \widetilde{V}_{P,a}$. We do know that $\mathscr{L}$ and its derivatives satisfy the following growth conditions:

$$|\mathscr{L}(\nabla u, u, x)| \leq C\left(a_{max}\, E\left[|\nabla u|^2\right] + E\left[|u|^{k+2}\right] + 1\right) \tag{3.3.1}$$

and also

$$\begin{cases} |D_{\nabla u}\mathscr{L}(\nabla u, u, x)| \leq C\left(a_{max}\, E\left[|\nabla u|\right]\right) \\ |D_u\mathscr{L}(\nabla u, u, x)| \leq C\left(E\left[|u|^{k+1}\right] + 1\right) \end{cases} \tag{3.3.2}$$

for some constant $C = \max\left\{\frac{1}{k+2},\, E[f]\right\}$. This motivates the following.

Definition 3.3.1 *We say $u \in \widetilde{V}_{P,a}$ is a weak solution of the boundary-value problem (1.1.3) for the nonlinear stochastic PDE provided*

$$\int_D E\left[a\nabla u \cdot \nabla v\right]\, dx + \int_D E\left[u\,|u|^k v\right]\, dx - \int_D E\left[fv\right]\, dx = 0 \tag{3.3.3}$$

for all $v \in \widetilde{V}_{P,a}$.

Theorem 3.3.2 (Solution of the nonlinear stochastic PDE) *A solution $u \in \widetilde{V}_{P,a}$ satisfies*

$$\mathscr{F}(u) = \min_{w \in \widetilde{V}_{P,a}} \mathscr{F}(w)$$

if and only if u is a weak solution of the nonlinear stochastic PDE (1.1.3).

Proof. ($\Rightarrow$). First we assume that $u \in \widetilde{V}_{P,a}$ satisfies

$$\mathscr{F}(u) = \min_{w \in \widetilde{V}_{P,a}} \mathscr{F}(w)$$

and show that u is a weak solution (1.1.3). We proceed by first differentiating inside the integrals. Fix $v \in \widetilde{V}_{P,a}$ and set

$$i(\tau) := \mathscr{F}(u + \tau v), \quad \tau \in \mathbb{R}.$$

In view of (3.3.1) we see that $i(\tau)$ is finite for all τ. Let $\tau \neq 0$ and write the difference quotient

$$\begin{aligned} \frac{i(\tau) - i(0)}{\tau} &= \int_D \frac{\mathscr{L}(\nabla u + \tau\nabla v, u + \tau v, x) - \mathscr{L}(\nabla u, u, x)}{\tau}\, dx \\ &= \int_D \frac{E\left[L(\nabla u(\omega, x) + \tau\nabla v(\omega, x), u(\omega, x) + \tau v(\omega, x), x, \omega)\right]}{\tau}\, dx \\ &\quad - \int_D \frac{E\left[L(\nabla u(\omega, x), u(\omega, x), x, \omega)\right]}{\tau}\, dx \\ &= \int_D \mathcal{L}^\tau(\cdot, x)\, dx \end{aligned} \tag{3.3.4}$$

where

$$\mathcal{L}^\tau(\cdot,x) = \frac{1}{\tau}E\left[L(\nabla u(\cdot,x)+\tau\nabla v(\cdot,x), u(\cdot,x)+\tau v(\cdot,x), x, \omega)\right] - E\left[L(\nabla u(\cdot,x), u(\cdot,x), x, \omega)\right]$$

for a.e. $x \in D$. Clearly, when $\tau \to 0$

$$\mathcal{L}^\tau(\cdot,x) \to E\left[a\nabla u(\cdot,x)\cdot\nabla v(\cdot,x)\right] + E\left[u(\cdot,x)\,|u(\omega,\cdot)|^k v(\cdot,x)\right] - E\left[f(\cdot,x)v(\cdot,x)\right] \tag{3.3.5}$$

for a.e. $x \in D$. Furthermore,

$$\begin{aligned}\mathcal{L}^\tau(\cdot,x) &= \frac{1}{\tau}\int_0^\tau \frac{\mathrm{d}}{\mathrm{d}s}E\left[L(\nabla u(\cdot,x)+s\nabla v(\cdot,x), u(\cdot,x)+sv(\cdot,x), x, \omega)\right]\ \mathrm{d}s \\ &= \frac{1}{\tau}\int_0^\tau \left\{E\left[a\nabla u(\cdot,x)\cdot\nabla v(\cdot,x)\right] + E\left[u(\cdot,x)\,|u(\cdot,x)|^k v(\cdot,x)\right] - E\left[f(\cdot,x)v(\cdot,x)\right]\right\}\ \mathrm{d}s.\end{aligned}$$

Then since $u, v \in \widetilde{V}_{P,a}$, inequalities (3.3.1) and Young's inequality imply after some elementary calculations that

$$|\mathcal{L}^\tau(\cdot,x)| \le C\left(a_{max}\left(E\left[|\nabla u|^2\right] + E\left[|\nabla v|^2\right]\right) + E\left[|u|^{k+2}\right] + E\left[|v|^{k+2}\right] + 1\right) \in L^1(D)$$

for each $\tau \neq 0$. Consequently we may invoke the Dominated Convergence Theorem to conclude from (3.3.4), (3.3.5) that $i'(0)$ exists and equals

$$\int_D E\left[a\nabla u\cdot\nabla v\right]\ dx + \int_D E\left[u\,|u|^k v\right]\ dx - \int_D E\left[fv\right]\ dx.$$

But then since $i(\cdot)$ has a minimum for $\tau = 0$, we know that $i'(0) = 0$; and thus u is a weak solution.

($\Leftarrow$). Now suppose that $u \in \widetilde{V}_{P,a}$ solves (1.1.3) in the weak sense, i.e.,

$$\int_D E\left[a\nabla u\cdot\nabla v\right]\ dx + \int_D E\left[u\,|u|^k v\right]\ dx - \int_D E\left[fv\right]\ dx = 0 \tag{3.3.6}$$

for $k \in \mathbb{N}^+$. Select any $w \in \widetilde{V}_{P,a}$. Utilizing the convexity of the mapping $(p,u) \mapsto \mathscr{L}(p,u,x)$, we have

$$\mathscr{L}(\nabla u, u, x) + E\left[D_pL(p,u,x)(q-p)\right] + E\left[D_uL(p,u,x)(w-u)\right] \le \mathscr{L}(q,w,x).$$

Let $p = \nabla u$, $q = \nabla w$ and integrate over D:

$$\mathscr{F}(u) + \int_D \left\{E\left[D_pL(p,u,x)(\nabla w - \nabla u)\right] + E\left[D_uL(p,u,x)(w-u)\right]\right\}\ dx \le \mathscr{F}(w)$$

which implies

$$\mathscr{F}(u) + \int_D \left\{ E\left[a\nabla u \cdot \nabla(w-u)\right] + E\left[(u|u|^k - f)(w-u)\right] \right\} dx \leq \mathscr{F}(w).$$

Therefore, in view of (1.1.3) the second term on the left is zero, and therefore $\mathscr{F}(u) \leq \mathscr{F}(w)$ for each $w \in \widetilde{V}_{P,a}$. □

We finally have enough tools to state the well-posedness of (3.3.3); precisely:

Corollary 3.3.3 *The weak form of* (1.1.3) *defined by (3.3.3) admits a unique solution* $u \in \widetilde{V}_{P,a}$*, which satisfies the estimates:*

$$\|u\|_P^2 \leq \frac{4C_P}{a_{min}^2} E\left[\|f\|_{L^2(D)}^2\right] \tag{3.3.7}$$

and

$$E\left[\|v\|_{L^{k+2}(D)}^{k+2}\right] \leq (k+2)\frac{2C_P}{a_{min}} E\left[\|f\|_{L^2(D)}^2\right]. \tag{3.3.8}$$

Furthermore, as before, define

$$\widetilde{C}(k) = \min\left\{\frac{a_{min}}{4}, \frac{1}{k+2}\right\}$$

then $u \in \widetilde{V}_{P,a}$ *satisfies the following:*

$$\|u\|_{\widetilde{P}}^2 \leq \frac{2}{\widetilde{C}(k)}\left(\frac{C_P}{a_{min}} E\left[\|f\|_{L^2(D)}^2\right]\right). \tag{3.3.9}$$

Proof. The proof of the existence and uniqueness of a solution $u \in \widetilde{V}_{P,a}$ follows directly from Theorems (3.2.1), (3.2.2) and (3.3.2). The estimates (3.3.7), (3.3.8) and (3.3.9) follow from the coercivity condition (3.1.1) derived in Section 3.1. To see this first notice that $0 \in \widetilde{V}_{P,a}$ and therefore since $u \in \widetilde{V}_{P,a}$ is a unique minimizer then

$$\begin{aligned} 0 = \mathscr{F}(0) \geq \mathscr{F}(u) &\geq \frac{\widetilde{C}(k)}{2}\|u\|_{\widetilde{P}}^2 - \frac{C_P}{a_{min}} E\left[\|f\|_{L^2(D)}^2\right] + 2\,C(k) \\ &\geq \frac{\widetilde{C}(k)}{2}\|u\|_{\widetilde{P}}^2 - \frac{C_P}{a_{min}} E\left[\|f\|_{L^2(D)}^2\right] \end{aligned}$$

and (3.3.9) follows directly. Also from the coercivity condition (3.1.1) we get that

$$\begin{aligned} \mathscr{F}(u) &\geq \frac{a_{min}}{4} E\left[\|\nabla u\|_{L^2(D)}^2\right] + \frac{1}{k+2} E\left[\| u\|_{L^{k+2}(D)}^{k+2}\right] - \frac{C_P}{a_{min}} E\left[\|f\|_{L^2(D)}^2\right] \\ &\geq \frac{a_{min}}{4} E\left[\|\nabla u\|_{L^2(D)}^2\right] - \frac{C_P}{a_{min}} E\left[\|f\|_{L^2(D)}^2\right] \end{aligned}$$

and then (3.3.7) follows. Similarly, from (3.1.1) and using the estimate (3.3.7) we observe that

$$\begin{aligned}\mathscr{F}(u) &\geq \frac{a_{min}}{2}\, E\left[\|\nabla u\|^2_{L^2(D)}\right] + \frac{1}{k+2}\, E\left[\|\, u\|^{k+2}_{L^{k+2}(D)}\right] - \frac{C_P}{a_{min}}\, E\left[\|f\|^2_{L^2(D)}\right] - \frac{a_{min}}{4C_P}\, E\left[\|u\|^2_{L^2(D)}\right] \\ &\geq \frac{1}{k+2}\, E\left[\|\, u\|^{k+2}_{L^{k+2}(D)}\right] - \frac{C_P}{a_{min}}\, E\left[\|f\|^2_{L^2(D)}\right] - \frac{a_{min}}{4C_P}\, E\left[\|u\|^2_{L^2(D)}\right] \\ &\geq \frac{1}{k+2}\, E\left[\|\, u\|^{k+2}_{L^{k+2}(D)}\right] - 2\,\frac{C_P}{a_{min}}\, E\left[\|f\|^2_{L^2(D)}\right]\end{aligned}$$

which implies the estimate (3.3.8) and completes the proof. □

3.4 Analyticity and the deterministic problem

Similarly to Chapter 2, once we have the input random fields described by a finite set of random variables, i.e. $a(\omega, x) = a_N(Y_1(\omega), \ldots, Y_N(\omega), x)$, and similarly for $f(\omega, x)$, the "finite dimensional" version of the stochastic variational formulation (3.3.3) has a "deterministic" equivalent which is the following: find $u_N \in L^2_\rho(\Gamma^N; L^{k+2}(D) \cap H^1_0(D))$ such that

$$\int_{\Gamma^N} \rho\, (a_N \nabla u_N, \nabla v)_{L^2(D)}\, dy + \int_{\Gamma^N} \rho\, (u_N |u_N|^k, v)_{L^2(D)}\, dy = \int_{\Gamma^N} \rho\, (f_N, v)_{L^2(D)}\, dy, \quad (3.4.1)$$

for all $v \in L^2_\rho(\Gamma^N; L^{k+2}(D) \cap H^1_0(D))$. Furthermore, as in Chapter 2 the nonlinear stochastic boundary value problem (3.3.3) now becomes a deterministic Dirichlet boundary value problem for a nonlinear elliptic partial differential equation with an $N-$dimensional parameter. Also, we use the notation $u_N(y)$ whenever we want to highlight the dependence on the parameter y and define $u_N : \Gamma^N \to L^{k+2}(D) \cap H^1_0(D)$. Then, problem (3.3.3) is equivalent to

$$\int_D a_N(y) \nabla u_N(y) \cdot \nabla \phi\, dx + \int_D u_N(y)\, |u_N(y)|^k \phi\, dx = \int_D f_N(y) \phi\, dx, \quad (3.4.2)$$

for all $\phi \in L^{k+2}(D) \cap H^1_0(D)$, ρ-a.e. in Γ^N. Recall that we assume that the coefficient a_N and the forcing term f_N admit a smooth extension on the ρ-zero measure sets. Then, equation (2.0.5) can be extended a.e. in Γ^N with respect to the Lebesgue measure (instead of the measure ρdy).

As discussed in Section 1.3 Assumption 1.3.1 must be verified for each application. In particular Assumption 1.3.1 was verified for the linear problem (2.0.5) by Babuška et al. (2005c) and described in Chapter 2.

It remains to be shown that the nonlinear problem (3.4.2) also satisfies the regularity restriction defined by Assumtion 1.3.1. As an initial approach we take $k = 1$ and restrict our analysis to the region $u \geq u_{min} > 0$ and $a_{min}^2 \geq 4C_P^4 \left(\frac{\epsilon-1}{\epsilon-5}\right)$, where $\epsilon > 5$ and C_P is defined by (3.3.7).

With this in mind, problem (3.4.2) is equivalent to the the following: find $u_N : \Gamma^N \to W(D)$ such that

$$\int_D a_N(y)\nabla u_N(y) \cdot \nabla\phi \, dx + \int_D (u_N(y))^2 \phi \, dx = \int_D f_N(y)\phi \, dx, \tag{3.4.3}$$

for all $\phi \in W(D)$, ρ-a.e. in Γ^N, where we now take $W(D) = L^3(D) \cap H_0^1(D)$. The next result concerns the analyticity of the solution $u = u_N$ of (3.4.3) whenever the diffusivity coefficient $a = a_N$ and the forcing term $f = f_N$ are infinitely differentiable w.r.t. y, under mild assumptions on the growth of their derivatives in y. With this in mind, we require several derivatives of the nonlinear function $\mathscr{N}(u) := u^2$. The Faá di Bruno's formula (di Bruno, 1855) gives an explicit equation for the ℓ-th derivative of the composition $\mathscr{N}(u)$. Namely:

$$\partial_{y_n}^{\ell} \mathscr{N}(u) = \sum \frac{\ell!}{j_1! \dots j_\ell!} \mathscr{N}^{(k)}(u) \left(\frac{\partial_{y_n} u}{1!}\right)^{j_1} \dots \left(\frac{\partial_{y_n}^{\ell} u}{\ell!}\right)^{j_\ell} \tag{3.4.4}$$

where

$$j = j_1 + \ldots + j_\ell$$

and the sum is over all partitions of ℓ, i.e., values of $k_1, \ldots, k_\ell$ such that

$$j_1 + 2j_2 + \ldots + \ell j_\ell = \ell.$$

We will perform a one-dimensional analysis in each direction y_n, $n = 1, \ldots, N$. For this, we recall the following notation: $\Gamma_n^* = \prod_{\substack{j=1 \\ j \neq n}}^N \Gamma_j$ and $\mathbf{y}_n^*$ will denote an arbitrary element of Γ_n^*. The analyticity result is presented next:

Theorem 3.4.1 *Assume that, for every $y = (y_n, \mathbf{y}_n^*) \in \Gamma^N$, there exists $\gamma_n < +\infty$ such that*

$$\left\| \frac{\partial_{y_n}^\ell a(y)}{a(y)} \right\|_{L^\infty(D)} \leq \gamma_n^\ell \ell! \qquad \text{and} \qquad \frac{\|\partial_{y_n}^\ell f(y)\|_{L^2(D)}}{1 + \|f(y)\|_{L^2(D)}} \leq r_\ell \, \gamma_n^\ell \ell!, \tag{3.4.5}$$

where the constant r_ℓ is defined by:

$$r_\ell = \begin{cases} 1 & \text{if } 0 \leq \ell \leq 2, \\ M & \text{if } \ell \geq 3, \end{cases}$$

and $\frac{C_P^2}{a_{min}} \leq M \leq \left(\frac{\epsilon-5}{\epsilon-1}\right) \frac{a_{min}}{2C_P^2}$, with $\epsilon > 5$ and the constant C_p as in (3.3.7). Then the solution $u(y_n, \mathbf{y}_n^*, x)$ to problem (3.4.3), as a function of y_n, $u : \Gamma_n \to C^0(\Gamma_n^*; W(D))$, admits an analytic extension $u(z, \mathbf{y}_n^*, x)$, $z \in \mathbb{C}$, in the region of the complex plane

$$\Sigma(\Gamma_n; \tau_n) \equiv \{z \in \mathbb{C}, \quad dist(z, \Gamma_n) \leq \tau_n\}. \tag{3.4.6}$$

with $0 < \tau_n < 1/(\epsilon\gamma_n)$. Moreover, $\forall z \in \Sigma(\Gamma_n; \tau_n)$,

$$\|Re(z)\, u(z)\|_{C^0(\Gamma_n^*; W(D))} \leq \widetilde{\kappa} \frac{M^{1/2}}{\sqrt{a_{min}}(1 - \epsilon\tau_n\gamma_n)} \left(1 + \|f\|_{C^0(\Gamma^N; L^2(D))}\right). \tag{3.4.7}$$

where the constant $\widetilde{\kappa} > 1$.

Proof. From (3.4.3), in every point $y \in \Gamma^N$, the ℓ-th derivative of u w.r.t y_n satisfies the following:

$$\partial_{y_n}^\ell B(y; u, v) + \langle \partial_{y_n}^\ell \mathscr{N}(u), v \rangle = \langle \partial_{y_n}^\ell f(y), v \rangle, \quad \forall v \in L^3(D) \cap H_0^1(D) \tag{3.4.8}$$

where B is the parametric bilinear form defined by $B(y; u, v) = \displaystyle\int_D a(y) \nabla u \cdot \nabla v \, dx$ and $\langle \cdot, \cdot \rangle \equiv (\cdot, \cdot)_{L^2(D)}$. Then, for $\ell \geq 2$, we define

$$p = \begin{cases} \frac{\ell}{2} & \text{for } \ell \text{ even}, \\ \frac{\ell-1}{2} & \text{for } \ell \text{ odd}, \end{cases} \qquad \beta_k = \begin{cases} \frac{1}{2} & \text{for } k = \frac{\ell}{2}, \\ 1 & \text{otherwise}, \end{cases}$$

and from (3.4.4) we get that

$$\begin{aligned} \langle \partial_{y_n}^\ell f, v \rangle = & \sum_{m=0}^{\ell} \binom{\ell}{m} \int_D \partial_{y_n}^m a(y) \partial_{y_n}^{\ell-m} \nabla u \cdot \nabla v \, dx \\ & + 2\ell! \int_D \left(u \frac{\partial_{y_n}^\ell u}{\ell!} + \sum_{j=1}^{p} \beta_j \frac{\partial_{y_n}^j u \, \partial_{y_n}^{(\ell-j)} u}{j!(\ell-j)!} \right) v \, dx. \end{aligned}$$

This implies that

$$\int_D a(y)\partial_{y_n}^{\ell}\nabla u\cdot\nabla v\,dx+2\ell!\int_D u\,\partial_{y_n}^{\ell}u\,v\,dx=-\sum_{m=1}^{\ell}\begin{pmatrix}\ell\\ m\end{pmatrix}\int_D\partial_{y_n}^{m}a(y)\partial_{y_n}^{\ell-m}\nabla u\cdot\nabla v\,dx$$
$$-2\ell!\int_D\left(\frac{\partial_{y_n}^{1}u}{1!}\frac{\partial_{y_n}^{\ell-1}u}{(\ell-1)!}+\frac{\partial_{y_n}^{2}u}{2!}\frac{\partial_{y_n}^{\ell-2}u}{(\ell-2)!}+\dots\right)v\,dx+\int_D\partial_{y_n}^{\ell}f(y)v\,dx,$$

for all $v\in L^3(D)\cap H_0^1(D)$. Therefore, choosing $v:=\partial_{y_n}^{\ell}u$ and utilizing the assumption (3.4.5) we obtain

$$\begin{aligned}
&\int_D a(y)\left(\partial_{y_n}^{\ell}\nabla u\right)^2\,dx+2\ell!\int_D u\left(\partial_{y_n}^{\ell}u\right)^2\,dx\\
&=-\sum_{m=1}^{\ell}\begin{pmatrix}\ell\\ m\end{pmatrix}\int_D\frac{\partial_{y_n}^{m}a(y)}{a(y)}\partial_{y_n}^{\ell-m}\nabla u\cdot\partial_{y_n}^{\ell}\nabla u\,a(y)\,dx\\
&\quad-2\ell!\int_D\left(\frac{\partial_{y_n}^{1}u}{1!}\frac{\partial_{y_n}^{\ell-1}u}{(\ell-1)!}+\frac{\partial_{y_n}^{2}u}{2!}\frac{\partial_{y_n}^{\ell-2}u}{(\ell-2)!}+\dots\right)\partial_{y_n}^{\ell}u\,dx+\int_D\partial_{y_n}^{\ell}f(y)\partial_{y_n}^{\ell}u\,dx\\
&\le\sum_{m=1}^{\ell}\begin{pmatrix}\ell\\ m\end{pmatrix}\left\|\frac{\partial_{y_n}^{m}a(y)}{a(y)}\right\|_{L^\infty(D)}\int_D\sqrt{a(y)}\partial_{y_n}^{\ell-m}\nabla u\,\sqrt{a(y)}\partial_{y_n}^{\ell}\nabla u\,dx\\
&\quad+2\ell!\int_D\left(\frac{\partial_{y_n}^{1}u}{1!}\frac{\partial_{y_n}^{\ell-1}u}{(\ell-1)!}+\frac{\partial_{y_n}^{2}u}{2!}\frac{\partial_{y_n}^{\ell-2}u}{(\ell-2)!}+\dots\right)\partial_{y_n}^{\ell}u\,dx+\left\|\partial_{y_n}^{\ell}f(y)\right\|_{L^2(D)}\left\|\partial_{y_n}^{\ell}u\right\|_{L^2(D)}\\
&\le\sum_{m=1}^{\ell}\begin{pmatrix}\ell\\ m\end{pmatrix}\gamma_n^m m!\left\|\sqrt{a(y)}\partial_{y_n}^{\ell-m}\nabla u\right\|_{L^2(D)}\left\|\sqrt{a(y)}\partial_{y_n}^{\ell}\nabla u\right\|_{L^2(D)}\\
&\quad+\frac{2\ell!}{a_{min}u_{min}^{1/2}}\left\|\sqrt{u_{min}}\partial_{y_n}^{\ell}u\right\|_{L^2(D)}\left(\frac{1}{1!(\ell-1)!}\left\|\sqrt{a_{min}}\partial_{y_n}^{1}u\right\|_{L^4(D)}\left\|\sqrt{a_{min}}\partial_{y_n}^{\ell-1}u\right\|_{L^4(D)}\right.\\
&\quad\left.+\frac{1}{2!(\ell-2)!}\left\|\sqrt{a_{min}}\partial_{y_n}^{2}u\right\|_{L^4(D)}\left\|\sqrt{a_{min}}\partial_{y_n}^{\ell-2}u\right\|_{L^4(D)}+\dots\right)\\
&\quad+\frac{C_P}{\sqrt{a_{min}}}\left\|\partial_{y_n}^{\ell}f(y)\right\|_{L^2(D)}\left\|\sqrt{a(y)}\partial_{y_n}^{\ell}\nabla u\right\|_{L^2(D)}\\
&\le\ell!\sum_{m=1}^{\ell}\gamma_n^m\frac{1}{(\ell-m)!}\left\|\sqrt{a(y)}\partial_{y_n}^{\ell-m}\nabla u\right\|_{L^2(D)}\left\|\sqrt{a(y)}\partial_{y_n}^{\ell}\nabla u\right\|_{L^2(D)}\\
&\quad+\frac{2\ell!}{a_{min}u_{min}^{1/2}}C_P^2\left\|\sqrt{u}\partial_{y_n}^{\ell}\nabla u\right\|_{L^2(D)}\left(\frac{c}{1!(\ell-1)!}\left\|\sqrt{a(y)}\partial_{y_n}^{1}\nabla u\right\|_{L^2(D)}\left\|\sqrt{a(y)}\partial_{y_n}^{\ell-1}\nabla u\right\|_{L^2(D)}\right.\\
&\quad\left.+\frac{c}{2!(\ell-2)!}\left\|\sqrt{a(y)}\partial_{y_n}^{2}\nabla u\right\|_{L^2(D)}\left\|\sqrt{a(y)}\partial_{y_n}^{\ell-2}\nabla u\right\|_{L^2(D)}+\dots\right)\\
&\quad+\frac{C_P}{\sqrt{a_{min}}}\left\|\partial_{y_n}^{\ell}f(y)\right\|_{L^2(D)}\left\|\sqrt{a(y)}\partial_{y_n}^{\ell}\nabla u\right\|_{L^2(D)},
\end{aligned}$$

and therefore, we conclude that

$$
\begin{aligned}
\|\sqrt{a(y)}\partial_{y_n}^{\ell}\nabla u\|^2 + 2\ell!\left\|\sqrt{u}\partial_{y_n}^{\ell}u\right\|^2 &\leq 4(\ell!)^2\sum_{m=1}^{\ell}\gamma_n^{2m}\frac{\left\|\sqrt{a(y)}\partial_{y_n}^{\ell-m}\nabla u\right\|_{L^2(D)}^2}{[(\ell-m)!]^2} \\
&+ 4\ell!\frac{C_P^2}{a_{min}^2 u_{min}}\left(\frac{1}{1!\{(\ell-1)!\}^2}\left\|\sqrt{a(y)}\partial_{y_n}^{1}\nabla u\right\|_{L^2(D)}^2\left\|\sqrt{a(y)}\partial_{y_n}^{\ell-1}\nabla u\right\|_{L^2(D)}^2\right. \\
&+ \left.\frac{1}{\{2!\}^2\{(\ell-2)!\}^2}\left\|\sqrt{a(y)}\partial_{y_n}^{2}\nabla u\right\|_{L^2(D)}^2\left\|\sqrt{a(y)}\partial_{y_n}^{\ell-2}\nabla u\right\|_{L^2(D)}^2+\dots\right) \\
&+ 2\frac{C_P^2}{a_{min}}\left\|\partial_{y_n}^{\ell}f(y)\right\|_{L^2(D)}^2.
\end{aligned}
\tag{3.4.9}
$$

By setting the general term $R_\ell = \left\|\sqrt{a(y)}\partial_{y_n}^{\ell}\nabla u(y)\right\|_{L^2(D)}^2 + 2\ell!\left\|\sqrt{u}\partial_{y_n}^{\ell}u(y)\right\|_{L^2(D)}^2$ and using the bounds on the derivatives a and f, we will prove that

$$
R_\ell \leq M(\ell!)^2(\epsilon\gamma_n)^{2\ell}(1+\|f(y)\|_{L^2(D)})^2, \qquad \forall \ell \geq 0. \tag{3.4.10}
$$

First, we will verify the bound (3.4.10) for $\ell = 0$ and $\ell = 1$ since (3.4.9) is only valid for $\ell \geq 2$ and, hence, these cases must been shown directly. We will prove (3.4.10) for all $\ell \geq 2$ by induction.

For the case $\ell = 0$, take $v := u$ in (3.4.8) and we arrive at the following:

$$
\begin{aligned}
\left\|\sqrt{a(y)}\nabla u\right\|_{L^2(D)}^2 + \left\|\sqrt{u}u\right\|_{L^2(D)}^2 &\leq C_P\|f(y)\|_{L^2(D)}\|\nabla u\|_{L^2(D)} \\
&\leq C_P\frac{\|f(y)\|_{L^2(D)}}{\sqrt{a_{min}}}\left\|\sqrt{a(y)}\nabla u\right\|_{L^2(D)} \\
&\leq \frac{1}{2}\left(\left\|\sqrt{a(y)}\nabla u\right\|_{L^2(D)}^2 + C_P^2\frac{\|f(y)\|_{L^2(D)}^2}{a_{min}}\right),
\end{aligned}
$$

i.e.,

$$
\begin{aligned}
\left\|\sqrt{a(y)}\nabla u\right\|_{L^2(D)}^2 + 2\left\|\sqrt{u}u\right\|_{L^2(D)}^2 \leq C_P^2\frac{\|f(y)\|_{L^2(D)}^2}{a_{min}} &\leq \frac{C_P^2}{a_{min}}(0!)(\epsilon\gamma_n)^0\left(1+\|f(y)\|_{L^2(D)}\right)^2 \\
&\leq M\left(1+\|f(y)\|_{L^2(D)}\right)^2.
\end{aligned}
\tag{3.4.11}
$$

For the case $\ell = 1$, we utilize (3.4.8) which yields:

$$
\langle\partial_{y_n}a(y)\nabla u, \nabla v\rangle + \langle a(y)\nabla\partial_{y_n}u, \nabla v\rangle + 2\langle u\,\partial_{y_n}u, v\rangle = \langle\partial_{y_n}f(y), v\rangle, \tag{3.4.12}
$$

and by setting $v := \partial_{y_n} u$ in (3.4.12) we conclude that

$$\begin{aligned}
&\left\|\sqrt{a(y)}\nabla\partial_{y_n}u\right\|^2_{L^2(D)} + 2\left\|\sqrt{u}\partial_{y_n}u\right\|^2_{L^2(D)} = \int_D \partial_{y_n} f(y)\partial_{y_n}u dx - \int_D \partial_{y_n} a(y)\nabla u\nabla\partial_{y_n}u dx \\
&\leq \frac{C_P}{\sqrt{a_{min}}}\left\|\partial_{y_n} f(y)\right\|_{L^2(D)}\left\|\sqrt{a(y)}\nabla\partial_{y_n}u\right\|_{L^2(D)} \\
&\quad + \left\|\frac{\partial_{y_n}a(y)}{a(y)}\right\|_{L^\infty(D)}\left\|\sqrt{a(y)}\nabla u\right\|_{L^2(D)}\left\|\sqrt{a(y)}\nabla\partial_{y_n}u\right\|_{L^2(D)} \\
&= \frac{C_P}{\sqrt{a_{min}}}\left\|\partial_{y_n} f(y)\right\|_{L^2(D)}\left\|\sqrt{a(y)}\nabla\partial_{y_n}u\right\|_{L^2(D)} \\
&\quad + \left\|\frac{\partial_{y_n}a(y)}{a(y)}\right\|_{L^\infty(D)}\left\|\sqrt{a(y)}\nabla u\right\|_{L^2(D)}\left\|\sqrt{a(y)}\nabla\partial_{y_n}u\right\|_{L^2(D)} \\
&\leq \frac{1}{4}\left\|\sqrt{a(y)}\nabla\partial_{y_n}u\right\|^2_{L^2(D)} + C_P^2\frac{\left\|\partial_{y_n} f(y)\right\|^2_{L^2(D)}}{a_{min}} \\
&\quad + \frac{1}{4}\left\|\sqrt{a(y)}\nabla\partial_{y_n}u\right\|^2_{L^2(D)} + \left(\frac{\partial_{y_n}a(y)}{a(y)}\right)^2\left\|\sqrt{a(y)}\nabla u\right\|^2_{L^2(D)} \\
&= \frac{1}{2}\left\|\sqrt{a(y)}\nabla\partial_{y_n}u\right\|^2_{L^2(D)} + C_P^2\frac{\left\|\partial_{y_n} f(y)\right\|^2_{L^2(D)}}{a_{min}} + \left(\frac{\partial_{y_n}a(y)}{a(y)}\right)^2\left\|\sqrt{a(y)}\nabla u\right\|^2_{L^2(D)}.
\end{aligned}$$

Equivalently, using (3.4.11) and the assumption (3.4.5) yields

$$\begin{aligned}
&\left\|\sqrt{a(y)}\nabla\partial_{y_n}u\right\|^2_{L^2(D)} + 4\left\|\sqrt{u}\partial_{y_n}u\right\|^2_{L^2(D)} \leq \frac{2C^2}{a_{min}}\left\|\partial_{y_n} f(y)\right\|^2_{L^2(D)} \\
&\qquad + 2\left\|\frac{\partial_{y_n}a(y)}{a(y)}\right\|^2_{L^\infty(D)}\left\|\sqrt{a(y)}\nabla u\right\|^2_{L^2(D)} \\
&\leq \frac{2C_P^2}{a_{min}}(1+\|f(y)\|_{L^2(D)})^2\gamma_n^2 + \frac{2C_P^2}{a_{min}}\|f(y)\|^2_{L^2(D)}\gamma_n^2 \\
&\leq \frac{4C_P^2}{a_{min}}\gamma_n^2(1+\|f(y)\|_{L^2(D)})^2
\end{aligned}$$

and hence,

$$\|\sqrt{a(y)}\nabla\partial_{y_n}u\|^2 + 2\|\sqrt{u}\partial_{y_n}u\|^2 \leq \frac{6C_P^2}{a_{min}}\gamma_n^2(1+\|f(y)\|_{L^2(D)})^2 \tag{3.4.13}$$

$$\leq M(\epsilon\gamma_n)^2(1+\|f(y)\|_{L^2(D)})^2 \tag{3.4.14}$$

To begin the induction we must first show directly that

$$R_2 \leq M(2!)^2(\epsilon\gamma_n)^4(1+\|f(y)\|_{L^2(D)})^2. \tag{3.4.15}$$

Indeed, taking another derivative with respect to y in (3.4.12) we have

$$\begin{aligned}
&\partial^2_{y_n}a(y)\nabla u\nabla\partial^2_{y_n}u + 2\,\partial_{y_n}a\nabla\partial_{y_n}u\nabla\partial^2_{y_n}u + a(y)\left(\nabla\partial^2_{y_n}u\right)^2 \\
&\qquad\qquad + 2u(\partial^2_{y_n}u)^2 + 2\partial_{y_n}u^2\partial^2_{y_n}u = \partial^2_{y_n}f(y)\partial^2_{y_n}u,
\end{aligned}$$

which can be equivalently written as

$$
\begin{aligned}
&a(y)\left(\nabla\partial_{y_n}^2 u\right)^2 + 2u\left(\partial_{y_n}^2 u\right)^2 \\
&= \partial_{y_n}^2 f(y)\partial_{y_n}^2 u - \partial_{y_n}^2 a(y)\nabla u\nabla\partial_{y_n}^2 u - 2\partial_{y_n} a(y)\nabla\partial_{y_n} u\nabla\partial_{y_n}^2 u - 2\partial_{y_n} u^2\partial_{y_n}^2 u \\
&= -\nabla\partial_{y_n}^2 u\partial_{y_n}^2 a(y)\nabla u - 2\partial_{y_n} a(y)\nabla\partial_{y_n} u\nabla\partial_{y_n}^2 u - 2\partial_{y_n} u^2\partial_{y_n}^2 u + \partial_{y_n}^2 f(y)\partial_{y_n}^2 u \\
&= -\sqrt{a(y)}\nabla\partial_{y_n}^2 u\frac{\partial_{y_n}^2 a(y)}{a(y)}\sqrt{a(y)}\nabla u - 2\frac{\partial_{y_n} a(y)}{a(y)}\sqrt{a(y)}\nabla\partial_{y_n} u\sqrt{a(y)}\nabla\partial_{y_n}^2 u \\
&\qquad - 2(\sqrt{a(y)}\partial_{y_n} u)^2\sqrt{a(y)}\partial_{y_n}^2 u\frac{1}{(a(y))^{3/2}} + \frac{\partial_{y_n}^2 f(y)}{\sqrt{a(y)}}\sqrt{a(y)}\partial_{y_n}^2 u \\
&\le \frac{1}{8}\left\|\sqrt{a(y)}\nabla\partial_{y_n}^2 u\right\|_{L^2(D)}^2 + 2\left(\frac{\partial_{y_n}^2 a(y)}{a(y)}\right)^2\left(\sqrt{a(y)}\nabla u\right)^2 + \frac{1}{8}\left\|\sqrt{a(y)}\nabla\partial_{y_n}^2 u\right\|_{L^2(D)}^2 \\
&\qquad + 8\left(\frac{\partial_{y_n} a(y)}{a(y)}\right)^2\left\|\sqrt{a(y)}\nabla\partial_{y_n} u\right\|_{L^2(D)}^2 + \frac{1}{8}\left\|\sqrt{a(y)}\nabla\partial_{y_n}^2 u\right\|_{L^2(D)}^2 \\
&\qquad + \frac{8}{(a(y))^3 C_P^4}\left\|\sqrt{a(y)}\nabla\partial_{y_n} u\right\|_{L^2(D)}^2 + \frac{1}{8}\left\|\sqrt{a(y)}\nabla\partial_{y_n}^2 u\right\|^2 + 2\frac{\|\partial_{y_n}^2 f(y)\|_{L^2(D)}^2}{a(y)} \\
&= \frac{1}{2}\left\|\sqrt{a(y)}\nabla\partial_{y_n}^2 u\right\|_{L^2(D)}^2 + 2\left(\frac{\partial_{y_n}^2 a(y)}{a(y)}\right)^2\left(\sqrt{a(y)}\nabla u\right)^2 \\
&\qquad + 8\left(\left(\frac{\partial_{y_n} a(y)}{a(y)}\right)^2 + \frac{1}{(a(y))^3 C_P^4}\right)\left\|\sqrt{a(y)}\nabla\partial_{y_n} u\right\|_{L^2(D)}^2 + 2\frac{\|\partial_{y_n}^2 f(y)\|_{L^2(D)}^2}{a(y)}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&\left\|\sqrt{a(y)}(\nabla\partial_{y_n}^2 u)\right\|_{L^2(D)}^2 + 4\left\|\sqrt{u}(\partial_{y_n}^2 u)\right\|_{L^2(D)}^2 \\
&\le 4\left(\frac{\partial_{y_n}^2 a}{a(y)}\right)^2\left\|\sqrt{a(y)}\nabla u\right\|_{L^2(D)}^2 + 16\left(\left(\frac{\partial_{y_n} a}{a(y)}\right)^2 + \frac{1}{a^3 C_P^4}\right)\left\|\sqrt{a(y)}\nabla\partial_{y_n} u\right\|_{L^2(D)}^2 + 4\frac{\|\partial_{y_n}^2 f\|^2}{a(y)} \\
&\le 4\left(\frac{\partial_{y_n}^2 a}{a(y)}\right)^2\frac{C_P^2}{a_{min}}\left(1 + \|f\|_{L^2(D)}\right) + \frac{4}{a(y)}\gamma_n^2\left(1 + \|f(y)\|_{L^2(D)}\right)^2 \\
&\qquad + 16\left(\left(\frac{\partial_{y_n} a}{a(y)}\right)^2 + \frac{1}{(a(y))^3 C_P^4}\right)\frac{6C_P^2}{a_{min}}\gamma_n^2\left(1 + \|f(y)\|_{L^2(D)}\right)^2 \\
&\le 112\gamma_n^4\frac{C_P^2}{a_{min}}\left(1 + \|f(y)\|_{L^2(D)}\right)^2 + 96\gamma_n^2\left(1 + \|f(y)\|_{L^2(D)}\right)^2\frac{1}{a_{min}^4 C_P^2} \\
&\qquad + \frac{4}{a_{min}}\gamma_n^2\left(1 + \|f(y)\|_{L^2(D)}\right)^2 \\
&\le 4(\epsilon\gamma_n)^4(1 + \|f(y)\|_{L^2(D)})^2\left[\frac{C_P^2}{a_{min}} + \frac{1}{\gamma_n^2 a_{min}^4 C_P^2} + \frac{1}{\gamma_n^2 a_{min}}\right].
\end{aligned}
$$

Hence (3.4.15) holds under that assumption that $M \geq \frac{C_P^2}{a_{min}}$ and provided that we choose the constant γ_n such that

$$\gamma_n^2 \geq \frac{\frac{1}{a_{min}^4 C_P^2} + \frac{1}{a_{min}}}{M - \frac{C_P^2}{a_{min}}}. \tag{3.4.16}$$

Finally, assuming that (3.4.10) holds for all $\ell = 1, 2, \ldots, L-1$ we will now show the induction step by verifying (3.4.10) for the case $\ell = L$. Indeed, from (3.4.9), for all $\ell = 1, \ldots, L-1$, we obtain

$$\left\| \sqrt{a(y)} \partial_{y_n}^{\ell} \nabla u \right\|_{L^2(D)} \leq M^{1/2} \ell! (\epsilon \gamma_n)^{\ell} \left(1 + \|f(y)\|_{L^2(D)}\right)^2,$$

which indicates the following:

$$\begin{aligned}
&\left\| \sqrt{a(y)} \partial_{y_n}^{L} \nabla u \right\|_{L^2(D)}^2 + 2L! \left\| \sqrt{u} \partial_{y_n}^{L} u \right\|_{L^2(D)}^2 \\
&\leq 4(L!)^2 M \sum_{m=1}^{L} \gamma_n^{2m} (\epsilon \gamma_n)^{2(L-m)} \left(1 + \|f(y)\|_{L^2(D)}\right)^2 \\
&\qquad + 4L!\, M^2 \frac{L+1}{2} \frac{C_P^2}{a_{min}^2 u_{min}} (\epsilon \gamma_n)^{2L} \left(1 + \|f(y)\|_{L^2(D)}\right)^4 \\
&\qquad\qquad + M^2 \frac{C_P^2}{2a_{min}} (L!)^2 \gamma_n^{2L} \left(1 + \|f(y)\|_{L^2(D)}\right)^2 \\
&\leq 4(L!)^2 M (\epsilon \gamma_n)^{2L} \left(1 + \|f(y)\|_{L^2(D)}\right)^2 \sum_{m=1}^{L} \epsilon^{-2m} \\
&\qquad + 4L! \frac{L+1}{2} M^2 \frac{C_P^2}{a_{min}^2 u_{min}} (\epsilon \gamma_n)^{2L} \left(1 + \|f(y)\|_{L^2(D)}\right)^4 \\
&\qquad\qquad + 2(L!)^2 \gamma_n^{2L} M^2 \frac{C_P^2}{a_{min}} \left(1 + \|f(y)\|_{L^2(D)}\right)^2 \\
&\leq 4(L!)^2 M (\epsilon \gamma)^{2L} \left(1 + \|f(y)\|_{L^2(D)}\right)^2 \frac{1}{\epsilon - 1} \\
&\qquad + 4L!\, L\, M^2 \frac{C_P^2}{a_{min}^2 u_{min}} (\epsilon \gamma)^{2L} \left(1 + \|f(y)\|_{L^2(D)}\right)^4 + 2 \frac{C_P^2}{a_{min}} M^2 (L!)^2 \gamma_n^{2L} \left(1 + \|f(y)\|_{L^2(D)}\right)^2 \\
&\leq (L!)^2 (\epsilon \gamma_n)^{2L} \left(1 + \|f(y)\|_{L^2(D)}\right)^2 M \\
&\qquad \times \left(4 \frac{1}{\epsilon - 1} + 4 \frac{M}{(L-1)!} \frac{C_P^2}{a_{min}^2 u_{min}} \left(1 + \|f(y)\|_{L^2(D)}\right)^2 + 2M \frac{C_P^2}{a_{min}} \right)
\end{aligned}$$

Consequently, $R_L \leq M (L!)^2 (\epsilon \gamma_n)^{2L} (1 + \|f\|_{L^2(D)})^2$ provided that

$$\left(4 \frac{1}{\epsilon - 1} + 4M \frac{C_P^2}{a_{min}^2 u_{min}} \left(1 + \|f(y)\|_{L^2(D)}\right)^2 + 2M \frac{C_P^2}{a_{min}} \right) \leq 1,$$

or equivalently,

$$M \leq \left(\frac{\epsilon-5}{\epsilon-1}\right) \frac{1}{2C_P^2\left(\frac{2}{a_{min}^2 u_{min}}\left(1+\|f(y)\|_{L^2(D)}\right)^2+\frac{1}{a_{min}}\right)} \leq \left(\frac{\epsilon-5}{\epsilon-1}\right) \frac{1}{2C_P^2\left(\frac{2}{\kappa a_{min}^2}+\frac{1}{a_{min}}\right)}$$
$$\leq \left(\frac{\epsilon-5}{\epsilon-1}\right) \frac{a_{min}}{C_P^2}$$

as described in the assumptions. Above we utilize (3.4.3) to provide the following bounds:

$$\|u\|_{L^3(D)}^3 \leq \|f\|_{L^{3/2}(D)} \|u\|_{L^3(D)}, \quad \|u\|_{L^3(D)}^2 \leq \widetilde{C}\|f\|_{L^2(D)}$$

and

$$u_{min} \leq \kappa \|f\|_{L^2(D)}^{1/2}, \quad \text{with } \kappa = \frac{\widetilde{C}^{1/2}}{(\mathrm{vol}(D))^{1/3}},$$

with constant $\widetilde{C} > 0$. This completes the induction step and therefore, the generic term R_ℓ admits the bound given by (3.4.10). We now get the final estimate on the growth of the derivatives of u:

$$\left\|\nabla \partial_{y_n}^\ell u\right\|_{L^2(D)} \leq \left(\frac{M}{a_{min}}\right)^{1/2} \ell!\,(\epsilon\gamma_n)^\ell \left(1+\|f(y)\|_{L^2(D)}\right)$$

We now define for every $y_n \in \Gamma_n$ the power series $u : \mathbb{C} \to C^0(\Gamma_n^*; W(D))$ as

$$u(z, \mathbf{y}_n^*, x) = \sum_{\ell=0}^{\infty} \frac{(z-y_n)^\ell}{\ell!} \partial_{y_n}^\ell u(y_n, \mathbf{y}_n^*, x).$$

where, for $\widetilde{\kappa} > 1$, we have

$$\|u(z)\|_{C^0(\Gamma_n^*;W(D))} = \|\nabla u(z)\|_{C^0(\Gamma_n^*;L^2(D))} + \|u(z)\|_{C^0(\Gamma_n^*;L^3(D))}$$
$$\leq \widetilde{\kappa}\, \|\nabla u(z)\|_{C^0(\Gamma_n^*,L^2(D))}$$

Hence,

$$\|u(z)\|_{C^0(\Gamma_n^*;W(D))} \leq \widetilde{\kappa} \sum_{\ell=0}^{\infty} \frac{|z-y_n|^\ell}{\ell!} \|\partial_{y_n}^\ell u(y_n)\|_{C^0(\Gamma_n^*;H_0^1(D))}$$
$$\leq \widetilde{\kappa} \left(\frac{M}{a_{min}}\right)^{1/2} \max_{y_n \in \Gamma_n} \left\{\left(\|f(y_n)\|_{C^0(\Gamma_n^*;L^2(D))}+1\right)\right\} \sum_{\ell=0}^{\infty} \left(|z-y_n|\epsilon\gamma_n\right)^\ell$$
$$\leq \widetilde{\kappa} \left(\frac{M}{a_{min}}\right)^{1/2} \left(\|f(y)\|_{C^0(\Gamma^N;L^2(D))}+1\right) \sum_{\ell=0}^{\infty} \left(|z-y_n|\epsilon\gamma_n\right)^\ell$$

where the series converges for all $z \in \mathbb{C}$ such that $|z - y_n| \leq \tau_n < 1/(\epsilon\gamma_n)$. Moreover, in the ball $|z - y_n| \leq \tau_n$, we have

$$\begin{aligned}\mathrm{Re}(z)\|u(z)\|_{C^0(\Gamma_n^*, W(D))} &\leq \widetilde{\kappa}\left(\frac{M}{a_{min}}\right)^{1/2}(\|f(y)\|_{C^0(\Gamma^N;L^2(D))}+1)\sum_{\ell=0}^{\infty}(|z-y_n|\epsilon\gamma_n)^{\ell} \\ &\leq \widetilde{\kappa}\,\frac{M^{1/2}}{\sqrt{a_{min}}(1-\epsilon\tau_n\gamma_n)}\left(1+\|f(y)\|_{C^0(\Gamma^N;L^2(D))}\right).\end{aligned}$$

The power series converges for every $y_n \in \Gamma_n$, hence, by a continuation argument, the function u can be extended analytically on the whole region $\Sigma(\Gamma_n;\tau_n)$ and estimate (3.4.7) follows. □

The following example illustrates the validity of assumption (3.4.5) (Babuška et al., 2005c).

Example 3.4.2 *Let us consider the case where the diffusivity coefficient a is expanded in a linear truncated Karhunen-Loève expansion*

$$a(\omega, x) = b_0(x) + \sum_{n=1}^{N}\sqrt{\lambda_n}b_n(x)Y_n(\omega),$$

provided that such expansion guarantees $a(\omega, x) \geq a_{min}$ for almost every $\omega \in \Omega$ and $x \in D$ (Schwab and Todor, 2007). In this case we have

$$\left\|\frac{\partial_{y_n}^k a}{a}\right\|_{L^\infty(\Gamma\times D)} \leq \begin{cases}\sqrt{\lambda_n}\|b_n\|_{L^\infty(D)}/a_{min} & \text{for } k=1\\ 0 & \text{for } k>1\end{cases}$$

and we can safely take $\gamma_n = \sqrt{\lambda_n}\|b_n\|_{L^\infty(D)}/a_{min}$ in (3.4.5).

If we consider, instead, a truncated exponential expansion

$$a(\omega, x) = a_{min} + e^{b_0(x)+\sum_{n=1}^{N}\sqrt{\lambda_n}b_n(x)Y_n(\omega)}$$

we have

$$\left\|\frac{\partial_{y_n}^k a}{a}\right\|_{L^\infty(\Gamma\times D)} \leq \left(\sqrt{\lambda_n}\|b_n\|_{L^\infty(D)}\right)^k$$

and we can take $\gamma_n = \sqrt{\lambda_n}\|b_n\|_{L^\infty(D)}$. Hence, both choices fulfill the assumption in Lemma 3.4.1.

Example 3.4.3 *Similarly to the previous case, let us consider a forcing term f of the form*

$$f(\omega, x) = c_0(x) + \sum_{n=1}^{N} c_n(x) Y_n(\omega)$$

where the random variables Y_n are Gaussian (either independent or not), and the functions $c_n(x)$ are square integrable for any $n = 1, \ldots, N$. Then, the function f belongs to the space $C^0_\sigma(\Gamma^N; L^2(D))$, with weight σ defined in (Babuška et al., 2005c, Section 3). Moreover,

$$\frac{\|\partial^k_{y_n} f(y)\|_{L^2(D)}}{1 + \|f(y)\|_{L^2(D)}} \leq \begin{cases} \|c_n\|_{L^2(D)} & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases}$$

and we can safely take $\gamma_n = \|c_n\|_{L^2(D)}$ in (3.4.5). Hence, such a forcing term satisfies the assumptions of Lemma 3.4.1. In this case, though, the solution u is linear with respect to the random variables Y_n (hence, clearly analytic) and our theory is not needed.

As we previously mentioned, Theorem 3.4.1 only discusses the nonlinear case when $k = 1$. Moreover, assuming that $u \geq u_{min} > 0$ may be difficult to show in general. This is an area of extreme interest to us and we intend to pursue several varying approaches in the future, to prove this result in general. In particular, we are optimistic that borrowing techniques from abstract nonlinear parabolic theory will assist in validating this result. For example, the works Lunardi (1987, 2004); Escher and Simonett (2003) prove that the solution to a fully nonlinear parabolic problem admits an analytic extension in some region of the complex plane with respect to the parameter t (time). In this setting we can think of the parameter t as the noise parameter $y_n \in \Gamma_n = [-1, 1]$ and we can attempt to extend these analyticity results to our particular problem, described by 3.4.2.

Broadening these abstract theoretical results will open the door to many new applications. In particular the work of Takáč et al. (1996) proved the analyticity of essentially bounded solutions to semilinear parabolic systems with applications to the Ginzburg-Landau equations. It is this direction that we predict this area of study to be headed. This thesis is just a beginning to many realistic engineering applications which we intend to devote much future research.

In the next chapter we provide an overview of several numerical methods used to solve the problems described in this chapter and Chapter 2 and introduce the approximation spaces used by such techniques. If the number of random variables is moderately large, we

motivate the consideration of sparse tensor product spaces as first proposed by Smolyak Smolyak (1963) and further investigated by e.g. Gerstner and Griebel (1998a); Barthelmann et al. (2000a); Frauenfelder et al. (2005b); Xiu and Hesthaven (2005), which will be the primary focus of this chapter. The use of sparse grids allows one to reduce dramatically the number of collocation points, while preserving a high level of accuracy. We also describe the anisotropic sparse approximation method to be considered as well as the different interpolation techniques to be employed.

CHAPTER 4

AN OVERVIEW OF NUMERICAL METHODS

Depending on the structure of the noise that drives an elliptic SPDE, there are different numerical approximations. For example, when the size of the noise is relatively small, a Neumann expansion around the mean value of the operator of the equation is a popular approach. It requires only the solution of standard deterministic partial differential equations, whose number is equal to the number of terms in the expansion. Equivalently, a Taylor expansion of the solution around its mean value with respect to the noise yields the same result. Similarly, the work Kleiber and Hien (1992) uses formal Taylor expansions up to second order of the solution but does not study their convergence properties. Recently, the work Babuška and Chatzipantelidis (2002b) proposed a perturbation method with successive approximations. They show that uniform coercivity of the diffusion coefficient is sufficient for the convergence of the method. Other perturbation approaches are discussed in Winter et al. (2002); Lu and Zhang (2004). The drawback of perturbation methods is their necessary conditions for convergence. In most of the cases the expansions are formal or, as in Babuška and Chatzipantelidis (2002b), they rely on rather stringent conditions on the allowed noise size.

Monte Carlo methods (Fishman, 1996b; Sobol′, 1994) are not restricted to small noise size. They are both general and simple to code and naturally suited for parallelization. They generate a set of independent identically distributed (iid) approximations of the solution by sampling the coefficients of the equation, using a spatial discretization of the partial differential equation, e.g. by a Galerkin finite element method. Then, using these approximations we can compute corresponding sample averages of the desired statistics. Monte Carlo methods have a rate of convergence that may be considered slow, but the required computational effort grows only like a polynomial with respect to the number of

random variables present in the problem. It is worth mentioning that in particular cases their convergence can be accelerated by variance reduction techniques (Jouini et al., 2001). The convergence rate of the Monte Carlo method is interpreted in the probabilistic sense and a practical estimate of its error needs an a posteriori estimate of the variance of the sampled random variable, which in turn requires an a priori bound on higher statistical moments, cf. the Berry-Esseen Theorem in Durrett (1996). Moreover, if the probability density of a random variable is smooth, the convergence rate of the Monte Carlo method for the approximation of its expected value can be improved (Novak, 1988; Traub and Werschulz, 1998). Quasi Monte Carlo methods (Caflisch, 1-49, Cambridge Univ. Press, 1998; Sobol′, 1998) offer a way to get a better convergence rate than the Monte Carlo method, although this advantage seems to deteriorate in general when the number of random variables present in the problem becomes too large.

When only the load f_N is stochastic, it is also possible to derive deterministic equations for the statistical moments of the solution. This linear elliptic case was analyzed in Babuška (1961); Larsen (1986) and more recently in the work Schwab and Todor (2002), where a new method to solve these equations with optimal complexity is presented.

On the other hand, the works of Deb (2000); Deb et al. (2001b), Ghanem and Red-Horse (1999); Ghanem and Spanos (1991) and (Xiu and Karniadakis, 2002a) address the general case where all the coefficients are stochastic. These approaches transform the original stochastic problem into a deterministic one with a large dimensional parameter and they differ in the choice of the approximating functional spaces. We refer to them here as Stochastic Galerkin approaches. The works Deb (2000); Deb et al. (2001b) use finite elements to approximate the noise dependence of the solution, while Ghanem and Red-Horse (1999); Ghanem and Spanos (1991) use a formal expansion in terms of Hermite polynomials. A similar approach with a generalized basis is followed by Xiu and Karniadakis (2002a).

The approximation error in the approach of Deb (2000); Deb et al. (2001b) can be then bounded in terms of deterministic quantities, as described in Babuška et al. (2004b); Frauenfelder et al. (2005a). Moreover, this approach combined with the use of double orthogonal polynomials yields a set of uncoupled "deterministic" problems for its solution, just as in the Monte Carlo method.

The work Benth and Gjerde (1998a), which developed a related error analysis for elliptic stochastic differential equations, gives approximation error estimates for functionals of the

solution. On the other hand, the analysis in Benth and Gjerde (1998a) uses the regularity of the computed functional together with estimates in negative spaces for the approximation error in the solution of the SPDE. These negative estimates can, in principle, accommodate rough solutions; however, they require H^2 spatial regularity, an assumption that is not clearly fulfilled by rough solutions.

Recently, the work Matthies and Keese (2005) proposed Wiener Chaos approximations for linear and nonlinear elliptic SPDEs. In particular, the convergence of the Wiener Chaos approximations for linear elliptic SPDEs is addressed. Thanks to Galerkin orthogonality, such convergence can be achieved by taking sufficiently many terms in the expansion of the diffusion coefficient. Moreover, once this is done the convergence results in Benth and Gjerde (1998a) apply to the linear case in Matthies and Keese (2005), because the bilinear form in (2.0.1) is computed exactly. However, the nonlinear case remains without convergence analysis. Moreover, the proposed numerical methodology by Matthies and Keese (2005); Xiu and Karniadakis (2002a); Ghanem and Red-Horse (1999); Ghanem and Spanos (1991) leads to a large system of coupled equations. For such nonlinear problems a Stochastic Collocation approach will outperform these methods; in terms of both the analytic convergence results and their numerical implementation. The new resulting methods preserve the convergence properties of Stochastic Galerkin and at the same time be computationally efficient and nonintrusive: they should allow maximum code reusability by relying on appropriate "samples" of the solution, similarly as in the Monte Carlo method.

The works Babuška et al. (2004b, 2005d); Frauenfelder et al. (2005a) base their analysis on the Karhunen-Loève expansion of the data. In addition, they assume that the mentioned expansion for the diffusivity coefficient has independent random variables with uniformly bounded support, an assumption that is not valid in general. In other words, if we proceed to fit a diffusion coefficient by means of a truncated Karhunen-Loève expansion there is no guarantee that the resulting fit will be coercive. On the other hand, if one fits the log of the coefficient then there is no such difficulty.

Even in the case of linear equations, when the diffusion coefficient $a_N(Y_1(\omega), \ldots, Y_N(\omega), x)$ is not linear with respect to Y_n, for $n = 1, \ldots, N$, the system of linear equations that defines the stochastic Galerkin approximate solution cannot be decoupled by means of double orthogonal polynomials (Babuška et al., 2004b). This same phenomena occurs in the presence of nonlinearities in the elliptic PDE such as problem 1.1.3. It is clear that

nonlinear Y_n-dependence offers good control over the coercivity of $a_N(Y_1(\omega), \ldots, Y_N(\omega), x)$ and also there is a clear motivation for addressing nonlinear problems. At the same time, we would like to avoid solving large coupled systems as much as possible. This motivates us to consider stochastic collocation as proposed and analyzed in the work Babuška et al. (2005a). This method has many advantages, such as:

- leads to uncoupled deterministic problems (also in case of input data which depend nonlinearly on the random variables),
- treats efficiently the case of non-independent random variables by introducing an auxiliary density,
- easily includes the case of random variables with unbounded support (such as Gaussian or exponential ones), and
- handles without difficulty a diffusivity coefficient a_N with unbounded second moment.

With this in mind, it is very natural to use this method as a starting point for our research. The rest of this chapter provides a deeper mathematical background on the numerical approximation techniques that motivated this work. We also discuss some implementation issues for each method.

4.1 Approximation spaces

We seek a numerical approximation to the exact solution of (1.2.1) in a suitable finite dimensional subspace. To describe such a subspace properly, we introduce some standard approximation subspaces, namely:

- $W_h(D) \subset W(D)$ is a standard finite element space of dimension N_h, which contains continuous piecewise polynomials defined on regular triangulations $\mathcal{T}_h$ that have a maximum mesh-spacing parameter $h > 0$. We suppose that W_h has the following deterministic approximability property: for a given function $\varphi \in W(D)$,

$$\min_{v \in W_h(D)} \|\varphi - v\|_{W(D)} \leq C(s; \varphi)\, h^s, \tag{4.1.1}$$

  where s is a positive integer determined by the smoothness of φ and the degree of the approximating finite element subspace and $C(s; \varphi)$ is independent of h.

Example 4.1.1 *Let D be a convex polygonal domain and $W(D) = H_0^1(D)$. For piecewise linear finite element subspaces we have, for $\varphi \in H^2(D)$,*

$$\min_{v \in W_h(D)} \|\varphi - v\|_{H_0^1(D)} \leq c\, h\, \|\varphi\|_{H^2(D)}.$$

That is, $s = 1$ and $C(s;\varphi) = \|\varphi\|_{H^2(D)}$, see for example Brenner and Scott (1994).

We will also assume that there exists a finite element operator $\pi_h : W(D) \to W_h(D)$ with the optimality condition

$$\|\varphi - \pi_h \varphi\|_{W(D)} \leq C_\pi \min_{v \in W_h(D)} \|\varphi - v\|_{W(D)}, \quad \forall \varphi \in W(D), \tag{4.1.2}$$

where the constant C_π is independent of the mesh size h.

- $\mathcal{P}_{\mathbf{p}}(\Gamma^N) \subset L^2_\rho(\Gamma^N)$ is the span of tensor product polynomials with degree at most $\mathbf{p} = (p_1, \ldots, p_N)$ i.e. $\mathcal{P}_{\mathbf{p}}(\Gamma^N) = \bigotimes_{n=1}^N \mathcal{P}_{p_n}(\Gamma_n)$, with

$$\mathcal{P}_{p_n}(\Gamma_n) = \text{span}(y_n^k,\ k = 0, \ldots, p_n), \quad n = 1, \ldots, N.$$

  Hence the dimension of $\mathcal{P}_{\mathbf{p}}(\Gamma^N)$ is $N_{\mathbf{p}} = \prod_{n=1}^N (p_n + 1)$.

In the case of the Monte Carlo method applied to the general problem (1.2.1) (see Section 4.2), a suitable approximating space for each sample is just $W_h(D) \subset W(D)$. Here we introduce the notion of a semi-discrete approximation u_h^N to the solution u_N of (1.2.1). When considering stochastic Galerkin or Stochastic Collocation, the approximating spaces may be based on a tensor product space defined by $V_{\mathbf{p},h}^N = \mathcal{P}_{\mathbf{p}}(\Gamma^N) \otimes W_h(D)$.

Another suitable choice is the decomposition of the random domain. See the $k-h$ method introduced in Babuška et al. (2002, 2004b) or use full polynomial approximation or, more interestingly, sparse tensor product approximations (see Section 4.4.2 and 4.4.3).

The goal of this research is to extend the efforts of Babuška et al. (2005a) where they consider a full tensor product method for solving linear elliptic PDEs with random input data to the more computationally effective sparse tensor products. From Chapter 3 it is obvious that we also want to include the case of a nonlinear elliptic PDE with random input data in this more practical setting. Throughout this work, special emphasis will be placed on the choice of approximating spaces and basis so the resulting computational technique is highly efficient and is "embarrassingly" parallelizable.

4.2 Monte Carlo

The standard Monte Carlo Finite Element Method (MCFEM) approximates the expected value function $E[u_N(Y_1(\omega), Y_2(\omega), \ldots, Y_N(\omega), \cdot)]$ by sample averages of independent, identically distributed (iid) realizations corresponding to sample coefficient functions, where u_N is the solution of a given PDE, as in, for instance, the problem (1.2.1) . For each realization of the coefficients a_N and f_N, a realization of the approximate solution is computed using the standard Galerkin finite element method. The stochastic coefficients depend on the noise as described in Section 1.2 and for simplicity, we use the notation $a_N(\omega, \cdot)$ and $f_N(\omega, \cdot)$, meaning $a_N(Y_1(\omega), Y_2(\omega), \ldots, Y_N(\omega), \cdot)$ and $f_N(Y_1(\omega), Y_2(\omega), \ldots, Y_N(\omega), \cdot)$, respectively.

Formulation of the Monte Carlo Finite Element Method (MCFEM):

Step 1. Choose a number of realizations, $M \in \mathbb{N}_+$, and a finite element space on D, $W_h(D)$, as defined in Section 4.1.

Step 2. For each $j = 1, \ldots, M$, sample iid realizations of the diffusion $a_N(\omega_j, \cdot)$, the load $f_N(\omega_j, \cdot)$ and find an approximation $u_h^N(\omega_j; \cdot) \in W_h(D)$ such that

$$B(\omega_j; u_h^N(\omega_j; \cdot), \phi_h) + \mathscr{N}(\omega_j; u_h^N(\omega_j; \cdot), \phi_h) = (f_N(\omega_j, \cdot), \phi_h)_{L^2(D)}, \quad \forall\, \phi_h \in W_h(D), \quad (4.2.1)$$

where B and $\mathscr{N}$ are the stochastic bilinear and nonlinear forms.

Example 4.2.1 *For the linear problem described in Example 1.1.1 the stochastic forms are defined by:*

$$\forall\, \omega \in \Omega : \quad B(\omega; \phi_1, \phi_2) \equiv (a(\omega, \cdot)\, \nabla\phi_1, \nabla\phi_2)_{L^2(D)}, \quad \forall\, \phi_1, \phi_2 \in H_0^1(D).$$

and

$$\forall\, \omega \in \Omega : \quad \mathscr{N}(\omega; \phi_1, \phi_2) \equiv 0, \quad \forall\, \phi_1, \phi_2 \in H_0^1(D).$$

Example 4.2.2 *For the nonlinear problem described in Example 1.1.2 the stochastic forms are defined by:*

$$\forall\, \omega \in \Omega : \quad B(\omega; \phi_1, \phi_2) \equiv (a(\omega, \cdot)\, \nabla\phi_1, \nabla\phi_2)_{L^2(D)}, \quad \forall\, \phi_1, \phi_2 \in L^{k+2}(D) \cap H_0^1(D).$$

and

$$\forall\, \omega \in \Omega : \quad \mathscr{N}(\omega; \phi_1, \phi_2) \equiv (\phi_1 |\phi_1|^k, \phi_2)_{L^2(D)}, \quad \forall\, \phi_1, \phi_2 \in L^{k+2}(D) \cap H_0^1(D).$$

Step 3. Approximate $E[u_N](\cdot)$ by the sample average:

$$\overline{E}(u_h^N; M) \equiv \frac{1}{M} \sum_{j=1}^{M} u_h^N(\omega_j; \cdot).$$

Here, we consider only the case where $W_h(D)$ is the same for all realizations, i.e. the triangulation of D is deterministic. The computational error naturally separates into the two parts

$$E[u_N] - \overline{E}(u_h^N; M) = \big(E\big[u_N\big] - E\big[u_h^N\big]\big) + \big(E[u_h^N] - \overline{E}(u_h^N; M)\big) \equiv \mathcal{E}_h^N + \mathcal{E}_{h,S}^N. \tag{4.2.2}$$

Therefore, the size of the space triangulation controls the space discretization error $\mathcal{E}_h^N$, while the number of realizations, M of u_h^N, controls the statistical error $\mathcal{E}_{h,S}^N$. Since the convergence rate with respect to the number of samples is roughly $\mathcal{O}(1/\sqrt{M})$ (Babuška et al., 2005d), we will not pursue this method further in this project. Instead, we will focus on other methods for approximation.

4.3 Stochastic Galerkin

This method approximates problem (1.2.1) with $u_{h,\mathbf{p}}^{N,G} \in V_{\mathbf{p},h}^N$ such that

$$E\left[B(\omega; u_{h,\mathbf{p}}^{N,G}, v)\right] + E\left[\mathscr{N}(\omega; u_{h,\mathbf{p}}^{N,G}, v)\right] = E\left[(f_N, v)_{L^2(D)}\right], \quad \forall\, v \in V_{\mathbf{p},h}^N. \tag{4.3.1}$$

Example 4.3.1 *If we examine the linear problem described in Example 1.1.1 then the approximate solution $u = u_{h,\mathbf{p}}^{N,G}$ satisfies*

$$\int_{\Gamma^N} \rho\, (a_N \nabla u, \nabla v)_{L^2(D)}\, dy = \int_{\Gamma^N} \rho\, (f_N, v)_{L^2(D)}\, dy, \quad \forall\, v \in V_{\mathbf{p},h}^N. \tag{4.3.2}$$

Example 4.3.2 *Similarly for the nonlinear problem described in Example 1.1.2, the approximate solution $u = u_{h,\mathbf{p}}^{N,G}$ satisfies*

$$\int_{\Gamma^N} \rho\, (a_N \nabla u, \nabla v)_{L^2(D)}\, dy + \int_{\Gamma^N} \rho\, (u|u|^k, v)_{L^2(D)}, = \int_{\Gamma^N} \rho\, (f_N, v)_{L^2(D)}\, dy, \tag{4.3.3}$$

for all $v \in V_{\mathbf{p},h}^N$.

For the linear problem (4.3.2), this approach has been considered by several authors (Babuška et al., 2002, 2004b; Deb et al., 2001b; Frauenfelder et al., 2005a; Xiu and

Karniadakis, 2002a; Ghanem and Spanos, 1991; Matthies and Keese, 2005). Observe that, in general, problem (4.3.2) leads to a fully coupled system of linear equations, whose dimension is $N_h \times N_{\mathbf{p}}$ and demands highly efficient strategies and parallel computations for its numerical solution (Le Maître et al., 2003; Elman et al., 2005). Conversely, the Stochastic Collocation techniques (see e.g. Section 4.4) developed in this work only require the solution of $N_{\mathbf{p}}$ uncoupled linear systems of dimension N_h, and is fully parallelizable. A similar observation can be made even in the presence of polynomial nonlinearities, as in problem (1.1.3), where we now solve $N_{\mathbf{p}}$ uncoupled nonlinear systems of dimension N_h.

In Babuška et al. (2004b, 2005d) a particular choice of basis functions, *double orthogonal polynomials*, for the space $P_{\mathbf{p}}(\Gamma^N)$ is proposed. This choice decouples the linear system given by (4.3.2) in the special case where the diffusion coefficient and the forcing term are multilinear combinations of the random variables $Y_n(\omega)$, for $n = 1, 2, \ldots, N$ (as is the case if one performs a truncated linear Karhunen-Loève expansion) and the random variables are independent, i.e. $\rho(y) = \prod_{n=1}^{N} \rho_n(y_n)$. The proposed basis is then obtained by solving the following eigenvalue problems, for each $n = 1, \ldots, N$,

$$\int_{\Gamma_n} z\psi_{kn}(z)v(z)\rho_n(z)\,dz = c_{kn}\int_{\Gamma_n} \psi_{kn}(z)v(z)\rho_n(z)\,dz, \qquad k = 1, \ldots, p_n + 1.$$

The eigenvectors ψ_{kn} are normalized so as to satisfy the property

$$\int_{\Gamma_n} \psi_{kn}(z)\psi_{jn}(z)\rho_n(z)\,dz = \delta_{kj}, \qquad \int_{\Gamma_n} z\psi_{kn}(z)\psi_{jn}(z)\rho_n(z)\,dz = c_{kn}\delta_{kj}.$$

See Babuška et al. (2004b, 2005d) for further details on the double orthogonal basis.

When the diffusion coefficient $a_N(\mathbf{Y}, x)$ is not linear with respect to Y_n, for $n = 1, 2, \ldots, N$, the system of linear equations that defines the stochastic Galerkin approximate solution cannot be decoupled by means of double orthogonal polynomials (Babuška et al., 2004b). This is also the case when studying the system of nonlinear equations given by (4.3.3), since this method depends on the linearity of the PDE. It is clear that nonlinear Y_n-dependence offers good control over the coercivity of $a_N(\mathbf{Y}, x)$. At the same time, we would like to avoid solving large coupled systems as much as possible. This motivates us to consider various Stochastic Collocation techniques, described next.

4.4 Stochastic Collocation techniques

Many authors have independently considered related different collocation methods for differential equations with random coefficients (Tatang, 1995; Mathelin et al., 2005; Xiu and Hesthaven, 2005). The paper that is the main inspiration for this work (Babuška et al., 2005a) is the only one that produced convergence rates for the approximate solutions, accommodating easily non-independent random variables with possibly unbounded support (such as Gaussian or exponential ones) and easily treats the case where the diffusivity coefficient a_N has unbounded second moment. These efforts have been extended to higher dimensional problems in our latest works (Nobile et al., 2006, 2007).

Thus far the analysis has been carried out for linear problems only and for applications where the input data depend on a relatively small number of random variables. One of the goals of this work is the extension of the methodology and corresponding convergence analysis to the class of nonlinear problems presented in Chapter 3. More importantly, our primary concern is those applications whose input data carry a large amount of uncertainty, and therefore, depend on a large number of random variables.

With this in mind, the Stochastic collocation method entails the sampling of approximate values $\pi_h u_N(y_k) = u_h^N(y_k) \in W_h(D)$, to the solution u_N of (1.2.1) on a suitable set of abscissas $y_k \in \Gamma^N$.

Example 4.4.1 *If we examine the linear PDE for example, then we introduce the semi-discrete approximation $u_h^N : \Gamma^N \to W_h(D)$, obtained by projecting equation (2.0.5) onto the subspace $W_h(D)$, for each $y \in \Gamma^N$, i.e.*

$$\int_D a_N(y)\nabla u_h^N(y) \cdot \nabla\phi_h \, dx = \int_D f_N(y)\phi_h \, dx, \quad \forall \phi_h \in W_h(D), \ \ \textit{for a.e. } y \in \Gamma^N. \tag{4.4.1}$$

Example 4.4.2 *Similarly for the nonlinear PDE defined by Example 1.1.2, the semi-discrete approximation u_h^N satisfies:*

$$\int_D a_N(y)\nabla u_h^N(y) \cdot \nabla\phi_h \, dx + \int_D u_h^N(y) \, |u_N^h(y)|^k \phi_h \, dx \int_D f_N(y)\phi_h \, dx, \tag{4.4.2}$$

for all $\phi_h \in W_h(D)$, for a.e. $y \in \Gamma^N$.

Notice that the finite element functions $u_h^N(y)$ described in Examples 4.4.1 and 4.4.2 satisfy the optimality condition (4.1.2), for all $y \in \Gamma^N$.

Then the construction of a fully discrete approximation, $u^N_{h,\mathbf{p}} \in C^0(\Gamma^N; W_h(D))$, is based on a suitable interpolation of the sampled values. That is

$$u^N_{h,\mathbf{p}}(y,\cdot) = \sum_k u^N_h(y_k,\cdot) l^{\mathbf{p}}_k(y), \tag{4.4.3}$$

where, for instance, the functions $l^{\mathbf{p}}_k$ can be taken as the Lagrange polynomials (see Section 4.4.1, 4.4.2 and 4.4.3). This formulation can be used to compute the mean value or variance of u, as:

$$E[u] \approx \overline{u}^N_h \equiv \sum_k u^N_h(y_k,\cdot) \int_{\Gamma^N} l^{\mathbf{p}}_k(y)\rho(y)dy$$

and

$$\mathrm{Var}[u] \approx \sum_k \left(u^N_h(y_k,\cdot)\right)^2 \int_{\Gamma^N} l^{\mathbf{p}}_k(y)\rho(y)dy - \left(\overline{u}^N_h\right)^2.$$

Several choices are possible for the interpolation points. We will discuss two of them, namely Clenshaw-Curtis and Gaussian in Sections 4.4.4 and 4.4.5 respectively. See the work Trefethen (2006) for an insightful comparison between these two choices. Regardless of the choice of interpolating knots, the interpolation can be constructed by using either full tensor product polynomials, see Section 4.4.1, or the space of sparse polynomials, see Sections 4.4.2 and 4.4.3.

4.4.1 Full tensor product interpolation

In this Section we briefly recall interpolation based on Lagrange polynomials. Let $i \in \mathbb{N}_+$ and $\{y^i_1, \dots, y^i_{m_i}\} \subset [-1,1]$ be a sequence of abscissas for Lagrange interpolation on $[-1,1]$.

For $u \in C^0(\Gamma^1; W(D))$ and $N = 1$ we introduce a sequence of one-dimensional Lagrange interpolation operators $\mathscr{U}^i : C^0(\Gamma^1; W(D)) \to V_{m_i}(\Gamma^1; W(D))$

$$\mathscr{U}^i(u)(y) = \sum_{j=1}^{m_i} u(y^i_j) \cdot l^i_j(y), \quad \forall u \in C^0(\Gamma^1; W(D)), \tag{4.4.4}$$

where $l^i_j \in \mathcal{P}_{m_i-1}(\Gamma^1)$ are Lagrange polynomials of degree $p_i = m_i - 1$ and

$$V_m(\Gamma^1; W(D)) = \left\{ v \in C^0(\Gamma^1; W(D)) \,:\, v(y,x) = \sum_{k=1}^{m} \widetilde{v}_k(x) l_k(y),\ \{\widetilde{v}_k\}_{k=1}^m \in W(D) \right\}.$$

Here of course we have, for $i \in \mathbb{N}_+$,

$$l^i_j(y) = \prod_{\substack{k=1 \\ k\neq j}}^{m_i} \frac{(y - y^i_k)}{(y^i_j - y^i_k)}$$

and formula (4.4.4) reproduces exactly all polynomials of degree less than m_i. Now, in the multivariate case $N > 1$, for each $u \in C^0(\Gamma^N; W(D))$ and the multi-index $\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{N}_+^N$ we define the full tensor product interpolation formulas

$$\mathcal{I}_{\mathbf{i}}^N u(y) := \left(\mathscr{U}^{i_1} \otimes \cdots \otimes \mathscr{U}^{i_N}\right)(u)(y) = \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_N=1}^{m_{i_N}} u\left(y_{j_1}^{i_1}, \ldots, y_{j_N}^{i_N}\right) \cdot \left(l_{j_1}^{i_1} \otimes \cdots \otimes l_{j_N}^{i_N}\right). \tag{4.4.5}$$

Clearly, the above product needs $(m_{i_1} \cdots m_{i_N})$ function values, sampled on a grid. These formulas will also be used as the building blocks for the Smolyak method, described next.

4.4.2 The isotropic Smolyak method

Here we follow closely the works Novak and Ritter (1996); Gerstner and Griebel (1998b); Novak and Ritter (1999); Novak et al. (1999); Barthelmann et al. (2000a) and describe the Smolyak *isotropic* formulas $\mathscr{A}(w, N)$. The Smolyak formulas are just linear combinations of product formulas (4.4.5) with the following key properties: only products with a relatively small number of knots are used and the linear combination is chosen in such a way that an interpolation property for $N = 1$ is preserved for $N > 1$. With $\mathscr{U}^0 = 0$ and for $i \in \mathbb{N}_+$ define

$$\Delta^i := \mathscr{U}^i - \mathscr{U}^{i-1}. \tag{4.4.6}$$

Moreover, for integers $w \in \mathbb{N}$, we define the sets

$$X(w, N) := \left\{ \mathbf{i} \in \mathbb{N}_+^N, \mathbf{i} \geq \mathbf{1} \,:\, \sum_{n=1}^{N} (i_n - 1) \leq w \right\}$$

and

$$Y(w, N) := \left\{ \mathbf{i} \in \mathbb{N}_+^N, \mathbf{i} \geq \mathbf{1} \,:\, w - N + 1 \leq \sum_{n=1}^{N} (i_n - 1) \leq w \right\}$$

and for $\mathbf{i} \in X(w, N)$ or $\mathbf{i} \in Y(w, N)$ we put $|\mathbf{i}| = i_1 + \cdots + i_N$. Then the isotropic Smolyak algorithm is given by

$$\mathscr{A}(w, N) = \sum_{\mathbf{i} \in X(w,N)} \left(\Delta^{i_1} \otimes \cdots \otimes \Delta^{i_N}\right). \tag{4.4.7}$$

Equivalently, formula (4.4.7) can be written as (see Wasilkowski and Woźniakowski (1995))

$$\mathscr{A}(w, N) = \sum_{\mathbf{i} \in Y(w,N)} (-1)^{w+N-|\mathbf{i}|} \binom{N-1}{w+N-|\mathbf{i}|} \cdot \left(\mathscr{U}^{i_1} \otimes \cdots \otimes \mathscr{U}^{i_N}\right). \tag{4.4.8}$$

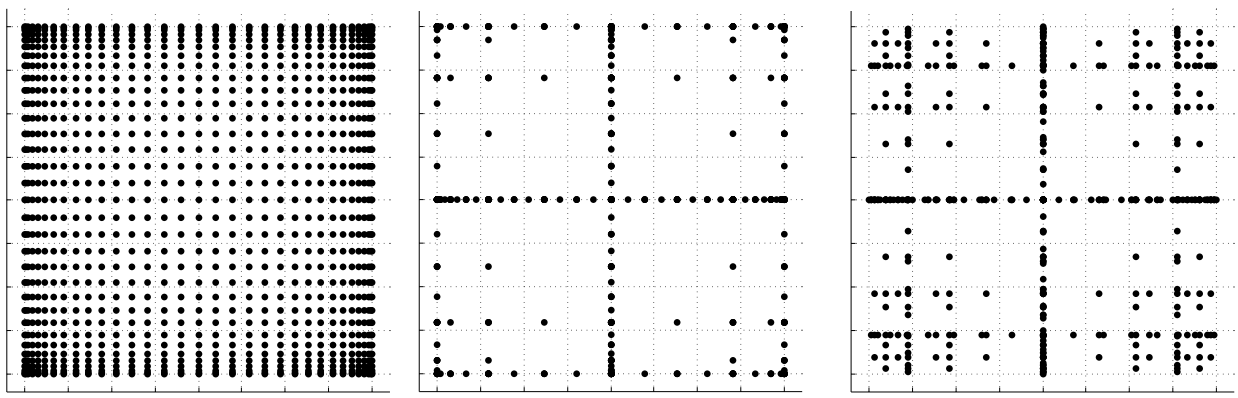

Figure 4.1: For a finite dimensional Γ^N with $N = 2$ we plot the full tensor product grid using the Clenshaw-Curtis abscissas and isotropic Smolyak sparse grids utilizing the Clenshaw-Curtis abscissas and the Gaussian abscissas for $w = 5$.

To compute $\mathscr{A}(w, N)(u)$, one only needs to know function values on the "sparse grid"

$$\mathscr{H}(w, N) = \bigcup_{\mathbf{i} \in Y(w,N)} \left(\vartheta^{i_1} \times \cdots \times \vartheta^{i_N}\right) \tag{4.4.9}$$

where $\vartheta^i = \left\{y_1^i, \ldots, y_{m_i}^i\right\} \subset [-1, 1]$ denotes the set of points used by $\mathscr{U}^i$. Note that the Smolyak algorithm, as presented in this Section, is isotropic, since all directions are treated equally. This can be seen from (4.4.7) where the multi-index $\mathbf{i} \in \mathbb{N}_+^N$ that determines the number of sample points in each dimension, i_n, $n = 1, 2, \ldots, N$ is sampled from the set $X(w, N)$. This ensures that if $(i_1, i_2, \ldots, i_N)$ is a valid index, then any permutation of it is also a valid index.

Examples of isotropic sparse grids, for $N = 2$, constructed from the nested Clenshaw-Curtis abscissas, described in Section 4.4.4, and the non-nested Gaussian abscissas, described in Section 4.4.5, are shown in Figure 4.1. To see the reduction in function evaluations we also include a plot of the corresponding full tensor product grid computed from the Clenshaw-Curtis abscissas. We note that since the Gaussian abscissas and the Clenshaw-Curtis abscissas tend to accumulate towards the boundary, plots of full tensor product grids will look superficially the same.

We will next discuss improvements that can be made to further reduce the number of points used to compute $\mathscr{U}^i$ by considering an anisotropic version of the method.

4.4.3 The anisotropic Smolyak method

It was shown in Nobile et al. (2006) that the conventional Smolyak construction is very effective for problems whose input data depend on a moderate number of random variables, which "weigh equally" in the solution. Upon the assumption that the solution is analytic with respect to each stochastic direction we show that for such isotropic situations the displayed convergence is faster than standard collocation techniques built upon full tensor product spaces.

On the other hand, the convergence rate deteriorates when we attempt to solve highly anisotropic problems, such as those appearing when the input random variables come e.g., from Karhunen-Loève -type truncations of "smooth" random fields described by Example 1.2.2. In this particular case we can rely on a priori information related to the decay of the noise coefficients to develop an anisotropic Smolyak algorithm.

To generalize the traditional sparse grid method with respect to different stochastic dimensions we propose an algorithm that considers a different index set rather than the unit simplex $|\mathbf{i}| \leq w + N$. Similar to Garcke and Griebel (2002); Gerstner and Griebel (2003) we let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N) \in \mathbb{R}_+^N$ be N-dimensional weight vector for the different stochastic dimensions. Furthermore, we define $\underline{\alpha} := \min_{1 \leq n \leq N} \alpha_n$ and consider the follow general class of simplicies defined by the index set

$$X_{\boldsymbol{\alpha}}(w, N) = \left\{ \mathbf{i} \in \mathbb{N}_+^N, \mathbf{i} \geq \mathbf{1} \, : \, \sum_{n=1}^{N} (i_n - 1)\alpha_n \leq w\underline{\alpha} \right\}.$$

The strategy to this approach relies on constructing the weight vector $\boldsymbol{\alpha} \in \mathbb{R}_+^N$ from either *a priori* knowledge or *a posteriori* information. See Chapter 5 respectively for a more detailed description.

Then, using (4.4.6) we described the anisotropic Smolyak formulas, $\mathscr{A}_{\boldsymbol{\alpha}}(w, N)$, given by

$$\mathscr{A}_{\boldsymbol{\alpha}}(w, N) = \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} \left(\Delta^{i_1} \otimes \cdots \otimes \Delta^{i_N} \right) \tag{4.4.10}$$

for $w \in \mathbb{N}$. Equivalently, (4.4.10) can be written as

$$\mathscr{A}_{\boldsymbol{\alpha}}(w, N) = \sum_{\mathbf{i} \in Y_{\boldsymbol{\alpha}}(w,N)} c_{\boldsymbol{\alpha}}(\mathbf{i}) \left(\mathscr{U}^{i_1} \otimes \cdots \otimes \mathscr{U}^{i_N} \right) \tag{4.4.11}$$

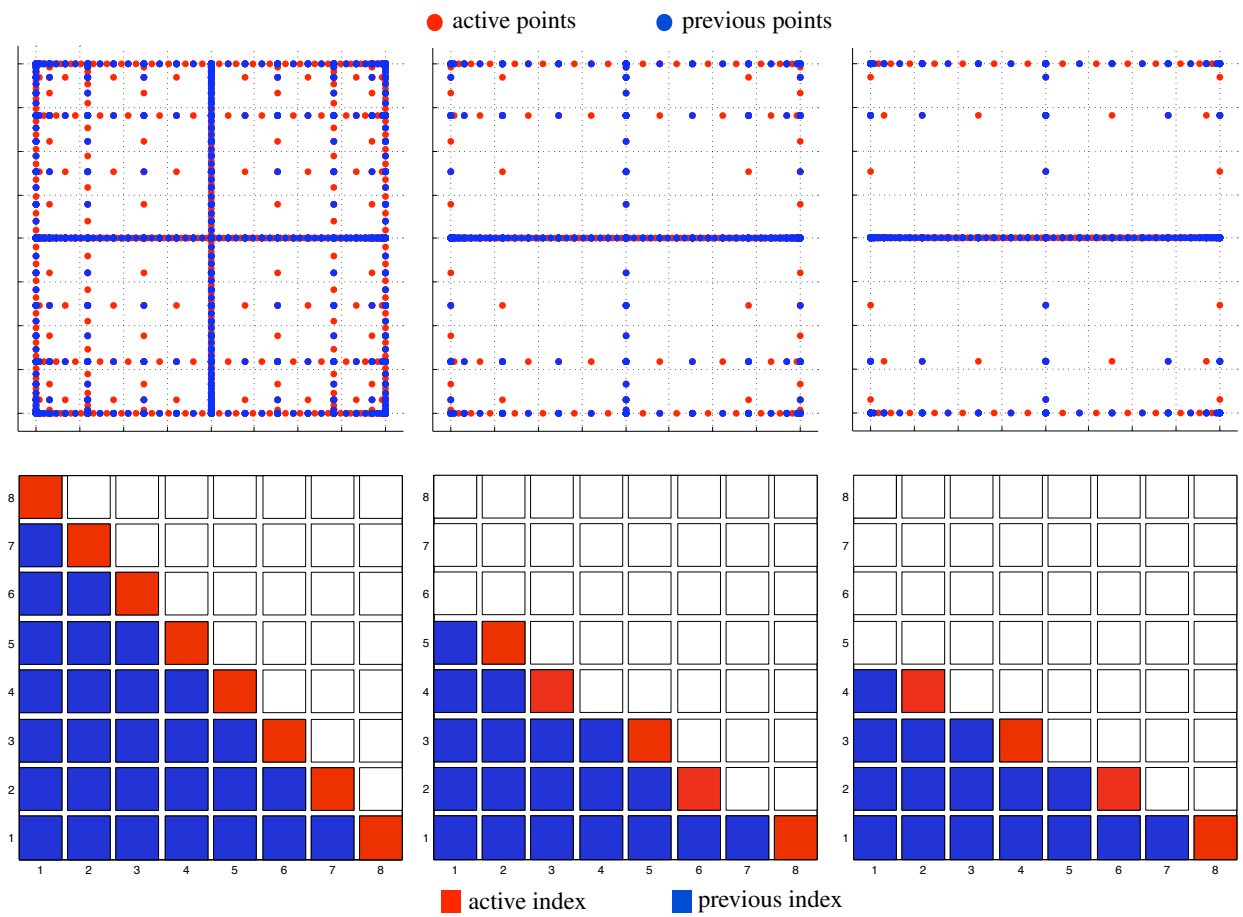

Figure 4.2: For a finite dimensional Γ^N with $N = 2$ and $w = 7$ we plot: on the top, the isotropic Smolyak grid and the anisotropic Smolyak grids with $\alpha_2/\alpha_1 = 3/2$ and $\alpha_2/\alpha_1 = 2$, utilizing the Clenshaw-Curtis abscissas and the bottom, the corresponding indices $(i_1, i_2) \in X_{\boldsymbol{\alpha}}(7, 2)$.

with

$$c_{\boldsymbol{\alpha}}(\mathbf{i}) := \sum_{\substack{\mathbf{j}\in\{0,1\}^N \\ \mathbf{i}+\mathbf{j}\in X_{\boldsymbol{\alpha}}(w,N)}} (-1)^{|\mathbf{j}|}$$

and

$$Y_{\boldsymbol{\alpha}}(w, N) := X_{\boldsymbol{\alpha}}(w, N) \setminus X_{\boldsymbol{\alpha}}\left(w - \frac{|\boldsymbol{\alpha}|}{\underline{\alpha}}, N\right).$$

Similarly to the isotropic case, to compute $\mathscr{A}_{\boldsymbol{\alpha}}(w, N)(u)$, one only needs to know function values on the "sparse grid"

$$\mathscr{H}_{\boldsymbol{\alpha}}(w, N) = \bigcup_{\mathbf{i}\in Y_{\boldsymbol{\alpha}}(w,N)} \left(\vartheta^{i_1} \times \cdots \times \vartheta^{i_N}\right). \tag{4.4.12}$$

We note that the isotropic Smolyak method presented in Section 4.4.2 is a special case of the anisotropic algorithm. This can be observed by simply taking the components of weight vector to be equal, i.e. $\alpha_1 = \alpha_2 = \cdots = \alpha_N$. In this case formula (4.4.10) is equivalent to (4.4.7) and it can be easily shown that (4.4.11) reduces to (4.4.8). Finally, in Figure 4.2, for $N = 2$ and $w = 7$, we show anisotropic Smolyak sparse grids utilizing the Clenshaw-Curtis points described in Section 4.4.4, corresponding to the anisotropy ratio $\alpha_2/\alpha_1 = 1$, $\alpha_2/\alpha_1 = 3/2$ and $\alpha_2/\alpha_1 = 2$ respectively. We also show the indices $\mathbf{i} \in X_{\boldsymbol{\alpha}}(7, 2)$ that were used to construct the anisotropic sparse interpolant $\mathscr{A}_{\boldsymbol{\alpha}}(7, 2)$. In accordance with the nested structure of the Clenshaw-Curtis abscissas, in Figure 4.2 we point out the difference between active points/indices and previously computed points/indices.

4.4.4 Clenshaw-Curtis formulas

We first suggest to use the Smolyak algorithm based on polynomial interpolation at the extrema of Chebyshev polynomials. For any choice of $m_i > 1$ these knots are given by

$$y_j^i = -\cos\left(\frac{\pi(j-1)}{m_i - 1}\right), \quad j = 1, \ldots, m_i. \tag{4.4.13}$$

In addition, we define $y_1^i = 0$ if $m_i = 1$. It remains to specify the numbers m_i of knots that are used in formulas $\mathscr{U}^i$. In order to obtain nested sets of points, i.e., $\vartheta^i \subset \vartheta^{i+1}$ and thereby $\mathscr{H}_{\mathscr{A}}(w, N) \subset \mathscr{H}_{\mathscr{A}}(w+1, N)$, we choose

$$m_1 = 1 \text{ and } m_i = 2^{i-1} + 1, \text{ for } i > 1. \tag{4.4.14}$$

For such a choice of m_i we arrive at Clenshaw-Curtis formulas, see Clenshaw and Curtis (1960). It is important to choose $m_1 = 1$ if we are interested in optimal approximation in relatively large N, because in all other cases the number of points used by $\mathscr{A}(w, N)$ and $\mathscr{A}_{\boldsymbol{\alpha}}(w, N)$ increases too fast with N.

A variant of the Clenshaw-Curtis formulas are the Filippi formulas in which the abscissas at the boundary of the interval are omitted (Gerstner and Griebel, 1998a). In this case only the smaller degree $m_i - 1$ of exactness is obtained.

4.4.5 Gaussian formulas

We also propose to apply the Smolyak formulas based on polynomial interpolation at the zeros of the orthogonal polynomials with respect to a weight ρ. This naturally leads to

the Gauss formulas that have a maximum degree of exactness of $2m_i - 1$. However, these Gauss-Legendre formulas are in general not nested. Regardless, as in the Clenshaw-Curtis case, we choose

$$m_1 = 1 \text{ and } m_i = 2^{i-1} + 1, \text{ for } i > 1.$$

The natural choice of the weight ρ should be the probability density of the random variables $Y_i(\omega)$ for all i. Yet, in the general multivariate case, if the random variables Y_i are not independent, the density ρ does not factor across the dimensions, i.e.

$$\rho(y_1, \ldots, y_n) \neq \prod_{n=1}^{N} \rho_n(y_n).$$

To this end, we first introduce an auxiliary probability density function $\hat{\rho} : \Gamma^N \to \mathbb{R}^+$ that can be seen as the joint probability of N independent random variables, i.e. it factorizes as

$$\hat{\rho}(y_1, \ldots, y_n) = \prod_{n=1}^{N} \hat{\rho}_n(y_n), \quad \forall y \in \Gamma^N, \qquad \text{and is such that} \quad \left\| \frac{\rho}{\hat{\rho}} \right\|_{L^\infty(\Gamma^N)} < \infty. \tag{4.4.15}$$

For each dimension $n = 1, \ldots, N$ let the m_n Gaussian abscissas be the roots of the m_n degree polynomial that is $\hat{\rho}_n$-orthogonal to all polynomials of degree $m_n - 1$ on the interval $[-1, 1]$.

The use of sparse approximation techniques seems to be an adequate tool for breaking the complexity of higher dimensional problems (Nobile et al., 2006; Babuška et al., 2007; Nobile et al., 2007). In particular, the methods presented in Sections 4.4.2 and 4.4.3 are very flexible computational tools that are at the same time relatively simple to implement and have a strong theoretical foundation. It is extremely important that we understand the computational errors associated with applying our sparse collocation techniques when solving problem (1.1.1), while utilizing the formulas presented in Sections 4.4.4 and 4.4.5. The goal of the Chapter 6 will be to settle such matters by providing a detailed error analysis of the techniques under study. In particular, we will describe how to estimate the rate of convergence of the isotropic Smolyak method in Section 6.1 and the anisotropic Smolyak method in Section 6.2. First, in the next chapter we provide *a priori* and *a posteriori* procedures for tuning the anisotropy of our sparse grid method to the problems of interest.

CHAPTER 5

SELECTION OF THE WEIGHTS FOR ANISOTROPIC SMOLYAK

The ability to evaluate the stochastic dimensions differently is a necessity since many practical problems exhibit highly anisotropic behavior. The rationale behind our anisotropic sparse grid approach is based on an examination of the total error

$$\varepsilon = \|u_N - \mathcal{I}_{\mathbf{p}}^N u_N\|_{L^2_\rho(\Gamma^N;W(D))}, \tag{5.0.1}$$

produced by anisotropic full tensor product polynomial interpolation on Gaussian abscissas.

Each stochastic dimension contributes to this total error. When the total error is divided equally among the random variables our earlier work Nobile et al. (2006) revealed that the isotropic Smolyak method, described in Section 4.4.2, displays a fast convergence rate and is very effective in considerably reducing the *curse of dimensionality*. We are also able to treat effectively problems that depend on a moderately large number of random variables, while keeping a high level of accuracy. When the error is dominated by certain directions we utilize the anisotropic Smolyak algorithm, described in Section 4.4.3 which combines an optimal treatment of the anisotropy of the problem while minimizing function evaluations via the use of sparse grids. Similar to an adaptive interpolation method, we place more points in the directions with the largest contribution to the total error.

The main idea is to link the α_n coefficients with the rate of exponential convergence in the corresponding direction, which for functions that satisfy Assumption 1.3.1 is described by the following Lemma, whose proof can be found in (Babuška et al., 2005c, Lemma 7) and which is an immediate extension of the result given in (DeVore and Lorentz, 1993, Chapter 7, Section 8):

Lemma 5.0.3 *Given a function $v \in C^0(\Gamma^1; W(D))$ which admits an analytic extension in the region of the complex plane $\Sigma(\Gamma^1;\tau) = \{z \in \mathbb{C}, \quad dist(z, \Gamma^1) \leq \tau\}$ for some $\tau > 0$, there holds*

$$E_{m_i} \equiv \min_{w \in V_{m_i}} \|v - w\|_{C^0(\Gamma^1;W(D))} \leq \frac{2}{\varrho - 1} e^{-m_i \log(\varrho)} \max_{z \in \Sigma(\Gamma^1;\tau)} \|v(z)\|_{W(D)}$$

where

$$1 < \varrho = \frac{2\tau}{|\Gamma^1|} + \sqrt{1 + \frac{4\tau^2}{|\Gamma^1|^2}}. \tag{5.0.2}$$

Remark 5.0.4 (Approximation with unbounded random variables) *A related result with weighted norms holds for unbounded random variables whose probability density decays as the Gaussian density at infinity, see Babuška et al. (2005c).*

In the multidimensional case, the value of τ_n will depend, in general, on the direction n, cf. (2.0.7) and (2.0.6). As a consequence of this variation and (5.0.2) the decay coefficients ϱ_n will also depend on the direction, $n = 1, \ldots, N$. We now assume that we know positive numbers $0 < g(n)$, $n = 1, \ldots, N$, such that

$$\varrho_n \geq e^{g(n)}. \tag{5.0.3}$$

Then, we will choose the anisotropic Smolyak weights as

$$\alpha_n = g(n) \quad \text{for all } n = 1, 2, \ldots, N. \tag{5.0.4}$$

On what follows we will use the notations

$$\underline{\alpha} = \underline{g} = \min_{1 \leq n \leq N} \{g(n)\} \quad \text{and} \quad \mathscr{G}(N) = \sum_{n=1}^{N} g(n). \tag{5.0.5}$$

Observe that we have now transformed the problem of choosing $\boldsymbol{\alpha}$ into the one of estimating the decay coefficients $\mathbf{g} = (g(1), \ldots, g(N))$.

Remark 5.0.5 (Optimality of $\boldsymbol{\alpha}$ choice) *As will be seen in Remark 6.2.12 the choice $\boldsymbol{\alpha} = \mathbf{g}$ is optimal with respect to the error bound derived in Theorem 6.2.7.*

In principle, it is possible to consider two kinds of estimation strategies. The first uses *a priori* knowledge, while in the second approach we use *a posteriori* information from computations, i.e. we fit the values of $\mathbf{g}$. The remainder of the Section will explain the choice of $\boldsymbol{\alpha} \in \mathbb{R}_+^N$, for the construction of set of indices $\mathbf{i} \in X_{\boldsymbol{\alpha}}(w, N)$, using these procedures.

5.1 A priori selection

The estimation of the $\boldsymbol{\alpha} = \mathbf{g} \in \mathbb{R}_+^N$ coefficients can be achieved using *a priori* information, i.e. by estimating first suitable values for τ_n, then using (5.0.2), and taking $g(n)$ as in (5.0.3), namely

$$g(n) = \log\left(\frac{2\tau_n}{|\Gamma_n|} + \sqrt{1 + \frac{4{\tau_n}^2}{|\Gamma_n|^2}}\right),$$

for $n = 1, 2, \ldots, N$. For the numerical example 7.0.1 described in Chapter 7 we have used the simpler and more conservative relation

$$g(n) = \log\left(1 + \frac{2\tau_n}{|\Gamma_n|}\right). \tag{5.1.1}$$

On the other hand, a priori estimates for τ_n can be derived for certain classes of problems, see for instance the linear problem (2.0.1) and the estimates (2.0.7) and (2.0.6). In Chapter 7 we implement numerically this *a priori* choice of $\boldsymbol{\alpha} = \mathbf{g} \in \mathbb{R}_+^N$ for the construction of the general simplices $\mathbf{i} \in X_{\boldsymbol{\alpha}}(w, N)$ when solving problem (1.1.2).

In some practical applications it may not be possible to sharply estimate the vector $\mathbf{g}$ using *a priori* information. Therefore, in the next Section we propose a computational alternative.

5.2 A posteriori selection

In order to explain the *a posteriori* estimation of $\boldsymbol{\alpha} = \mathbf{g} \in \mathbb{R}_+^N$ for the construction of the general simplices $\mathbf{i} \in X_{\boldsymbol{\alpha}}(w, N)$ we will describe particular cases with $N = 11$ and focus on both an isotropic and highly anisotropic version of the linear example (7.0.1) described in Chapter 7, corresponding to $L_c = 1/64$ and $L_c = 1/2$ respectively. To compute the weight factor α_n for each stochastic direction $n = 1, 2, \ldots, N$, we inductively utilize Lemma 5.0.3 and investigate the interpolation error, given by (5.0.1).

To analyze this term we employ a one-dimensional argument. We first pass from the norm L^2_ρ to $L^2_{\hat{\rho}}$:

$$\|u_N - \mathcal{I}^N_{\mathbf{p}} u_N\|_{L^2_\rho(\Gamma^N; W(D))} \leq \left\|\frac{\rho}{\hat{\rho}}\right\|^{1/2}_{L^\infty(\Gamma)} \|u_N - \mathcal{I}^N_{\mathbf{p}} u_N\|_{L^2_{\hat{\rho}}(\Gamma^N; W(D))}.$$

Here we adopt the same notation as in Section 1.3, namely we indicate with $\bullet_n$ a quantity relative to the direction y_n and $\bullet_n^*$ the analogous quantity relative to all other directions y_j,

$j \neq n$. We focus on the first direction y_1 and define the Banach space $V = L^2_{\hat{\rho}_1^*}(\Gamma_1^*; W(D))$. Furthermore, we now describe an interpolation operator $\mathcal{I}^1_{p_1} : C^0(\Gamma^1; V) \to L^2_{\hat{\rho}_1}(\Gamma^1; V)$,

$$\mathcal{I}^1_{p_1} v(y_1, y_1^*, x) = \sum_{k=1}^{m_1} v(y_k^1, y_1^*, x)\, l_k^1(y_1),$$

where we recall that $m_n = p_n + 1$, $\forall n = 1, 2, \ldots, N$ and the notation from Section 4.4.1. Then, the global interpolant $\mathcal{I}^N_{\mathbf{p}}$ can be written as the composition of two interpolation operators $\mathcal{I}^N_{\mathbf{p}} = \mathcal{I}^1_{p_1} \circ \mathcal{I}^{(1)}_{\mathbf{p}}$ where $\mathcal{I}^{(1)}_{\mathbf{p}}$ is the interpolation operator in all directions $y_2, y_3, \ldots, y_N$ except y_1, defined by $\mathcal{I}^{(1)}_{\mathbf{p}} : C^0(\Gamma_1^*; W(D)) \to L^2_{\hat{\rho}_1^*}(\Gamma_1^*; W(D))$. We then have

$$\|u_N - \mathcal{I}_p u_N\|_{L^2_{\hat{\rho}}(\Gamma^N; W(D))} \leq \underbrace{\|u_N - \mathcal{I}^1_{p_1} u_N\|_{L^2_{\hat{\rho}}(\Gamma^N; W(D))}}_{\text{(I)}} + \underbrace{\|\mathcal{I}^1_{p_1}(u_N - \mathcal{I}^{(1)}_{\mathbf{p}} u_N)\|_{L^2_{\hat{\rho}}(\Gamma^N; W(D))}}_{\text{(II)}}$$

Let us bound the first term given by (I). We think of u as a function of y_1 with values in the Banach space V, $u \in L^2_{\hat{\rho}_1}(\Gamma^1; V)$. By the construction of the auxiliary function $\hat{\rho}$ given in (4.4.15), the following inclusion holds:

$$C^0(\Gamma^1; V) \subset L^2_{\hat{\rho}_1}(\Gamma^1; V).$$

We also know from (Babuška et al., 2005c, Lemma 5) that the interpolation operator $\mathcal{I}^1_{p_1}$ is continuous as an operator from $C^0(\Gamma^1; V)$ with values in $L^2_{\hat{\rho}_1}(\Gamma^1; V)$. In particular, we can estimate

$$\text{(I)} = \|u_N - \mathcal{I}^1_{p_1} u_N\|_{L^2_{\hat{\rho}_1}(\Gamma^1; V)} \leq C_2 \inf_{w \in V_{m_1}(\Gamma^1; V)} \|u_N - w\|_{C^0(\Gamma^1; V)}.$$

To bound the best approximation error in $C^0(\Gamma^1; V)$, in the case we use Lemma 5.0.3 and the fact that $u_N \in C^0(\Gamma^1; V)$, see (Babuška et al., 2005c, Lemma 3). We also require the analyticity result, for the solution u_N, stated in Assumption 1.3.1. Putting everything together, we can say that

$$\text{(I)} \leq C {\varrho_1}^{-p_1}$$

where the constant C is specified in Lemma 5.0.3. To bound the term (II), we again use (Babuška et al., 2005c, Lemma 5):

$$\text{(II)} \leq C_1 \|u_N - \mathcal{I}^{(1)}_{\mathbf{p}} u_N\|_{C^0(\Gamma^1; V)}.$$

The term on the right hand side is again an interpolation error. So we can bound the interpolation error in all the other $N - 1$ directions, uniformly with respect to y_1 (in the

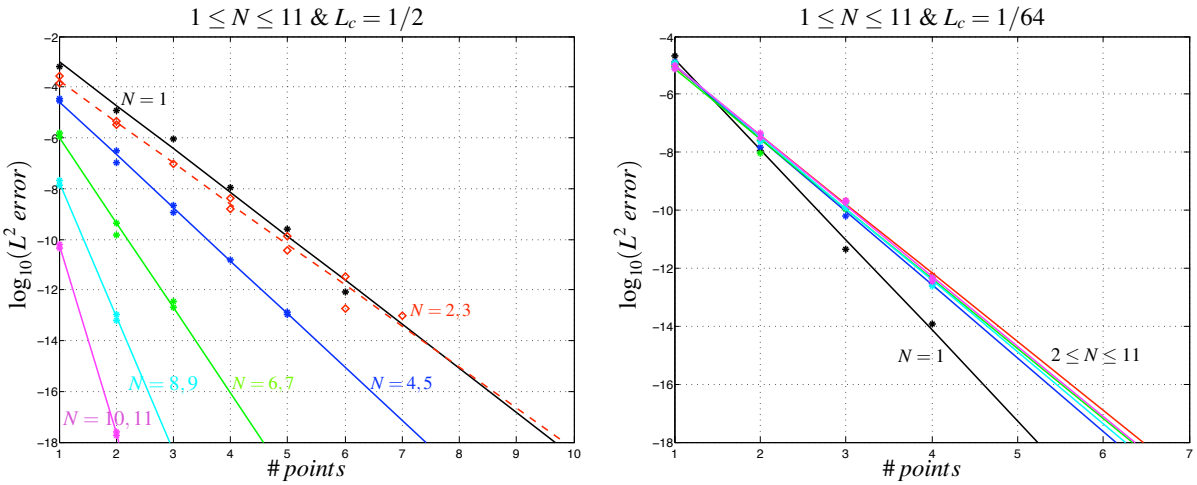

Figure 5.1: A linear least square approximation to fit $\log_{10}(\|E[\varepsilon_n]\|_{L^2(D)})$ versus p_n with ε_n defined by (5.2.2). For $n = 1, 2, \ldots, N = 11$ we plot: on the left, the highly anisotropic case $L_c = 1/2$ and on the right, the isotropic case $L_c = 1/64$.

norm C^0). We can proceed iteratively, defining an interpolation $\mathcal{I}^2_{p_2}$, bounding the resulting error in the direction y_2 and so on. Therefore, we arrive at a bound for the total error, as in (6.1.2). Furthermore, from the above argument we can partition the error in the N directions, i.e.

$$\|u_N - \mathcal{I}^N_{\mathbf{p}} u_N\|_{L^2_\rho(\Gamma^N;W(D))} \le \sum_{n=1}^{N} \varepsilon_n, \tag{5.2.1}$$

where, using Lemma 5.0.3, we expect an error decay in the direction n of the form

$$\varepsilon_n \approx d_n {\varrho_n}^{-p_n}, \quad \text{for all } n = 1, 2, \ldots, N. \tag{5.2.2}$$

Here p_n is the number of collocation points in the direction n. In order to compute the weight vector $\mathbf{g} = \boldsymbol{\alpha} \in \mathbb{R}^N_+$, $\mathbf{g} \approx \log(\boldsymbol{\varrho})$, we first observe from (5.2.2) that

$$\log_{10}(\varepsilon_n) \approx \log_{10}(d_n) - p_n \log_{10}(\varrho_n) \approx \log_{10}(d_n) - p_n \log_{10}(e) g(n).$$

Therefore, if we plot $\log_{10}(\varepsilon_n)$ versus p_n for each stochastic dimension $n = 1, 2, \ldots, N$ and use a linear least squares approximation to fit the data, the slope of each line will give an estimate of $g(n)$. When solving problem (7.0.1), for the cases $L_c = 1/2$ and $1/64$, we employ the same finite element space as in Chapter 7 and we approximate $E[\varepsilon_n]$ in the n-th direction, corresponding to a multi index $\mathbf{p} = (1, 1, \ldots, p_n, 1, \ldots, 1)$ by $E[\varepsilon_n] \approx E[u^N_{h,\mathbf{p}} - u^N_{h,\widetilde{\mathbf{p}}}]$, with

Table 5.1: The $N = 11$ values of the function $g(n) = \alpha_n$ constructed from *a posteriori* information. The values $g(n)$, $n = 1, 2, \ldots, N$ are the slopes of N linear least squares fits to the error of an univariate anisotropic method when solving problem (7.0.1) with correlation lengths $L_c = 1/2$ and $1/64$.

	$g(1)$	$g(2) = g(3)$	$g(4) = g(5)$	$g(6) = g(7)$	$g(8) = g(9)$	$g(10) = g(11)$
$L_c = 1/2$	1.7	1.6	2.1	3.3	5.3	7.4
$L_c = 1/64$	3.1	2.4	2.5	2.4	2.5	2.4

$\widetilde{\mathbf{p}} = (1, 1, \ldots, p_n + 1, 1, \ldots, 1)$. The computational results for the $L^2(D)$ approximation error in the expected value, $E[\varepsilon_n]$, are shown on Figure 5.1.

The results for $g(n)$, $n = 1, 2, \ldots, N = 11$ can be seen in Table 5.1. Table 5.1 reveals that the a posteriori selection of $\boldsymbol{\alpha} \in \mathbb{R}^N_+$ performs well at dictating the behavior of problem (7.0.1) for the cases $L_c = 1/2$ and $1/64$. In the former case the vector $\boldsymbol{\alpha}$ weighs heavily in the higher dimensions as opposed to the latter that approximately weighs equally in all directions. Also, both cases $L_c = 1/2$ and $1/64$, reveal that $\underline{\alpha} = \alpha_2 = \alpha_3$.

Observe that in general the rate ϱ_n also depends on $\mathbf{y}^*_n \in \Gamma^*_n$, i.e. on the values of the other random variables. Our way to estimate the decay coefficient $g(n)$ is not conservative since we only estimate ϱ_n at the single point $E[\mathbf{y}^*_n]$. A more conservative estimate will imply estimating the worst value of $\mathbf{y}^*_n$, i.e. the one that minimizes ϱ_n. This may be critical for nearly singular cases.

Remark 5.2.1 (Applications to piecewise constant random fields) *We also comment that the a posteriori selection of $\boldsymbol{\alpha} \in \mathbb{R}^N_+$, described above, is not only restricted to random fields related to the Karhunen-Loève expansion described in Example 1.2.2 but also can be easily applied in other cases, for instance to the piecewise constant random fields described in Example 1.2.1.*

We finally have a complete description of both the isotropic and anisotropic sparse grid Stochastic Collocation methods for solving problems such as 1.1.2 and 1.1.3. The next Chapter will be devoted to analyzing numerical errors for both methods considered, including cases where the sparse interpolant uses both Clenshaw-Curtis and Gaussian abscissas. In Section 6.1 we concentrate on analyzing the isotropic Smolyak method, described in Section 4.4.2, while Section 6.2 examines the effects of exploiting the possible underlying anisotropic behavior using the $\boldsymbol{\alpha}$-weighted version of the Smolyak, described in Section 4.4.3.

CHAPTER 6

ERROR ANALYSIS

The goal of this chapter is to provide a complete error analysis for both the isotropic and anisotropic sparse grid Stochastic Collocation methods considered, including cases where the sparse interpolant utilizes both Clenshaw-Curtis and Gaussian abscissas. The isotropic case, described in Section 4.4.2, is analyzed in Section 6.1 while the examination of the anisotropic method, described in Section 4.4.3, can be found in Section 6.2. The analysis relies on the regularity of the solution and exploits the behavior of sparse approximations from the anisotropic Smolyak method. We also address the case where the input random variables come from suitably truncated expansions of random fields and discuss how the size of the sparse grid can be algebraically related to the number of random variables retained in the expansion in order to have a discretization error of the same order as that of the error due to the truncation of the input random fields.

Collocation methods can be used to approximate the solution $u_N \in C^0(\Gamma^N; W(D))$ using finitely many function values. By Assumption 1.3.1, u_N admits an analytic extension. Further, each function value will be computed by means of a finite element technique. In general, the semi-dicrete finite element solution $\pi_h u_N$ also satisfies the regularity assumption 1.3.1. In particular, this is obviously true for the linear problem presented in Chapter 2. We now define the numerical approximation $u^N_{h,\mathbf{p}} = \mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N$. Our aim is to give a priori estimates for the total error

$$\epsilon = u - u^N_{h,\mathbf{p}} = u - \mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N$$

where the operator $\mathscr{A}_{\boldsymbol{\alpha}}(w, N)$ is either the isotropic or anisotropic Smolyak sparse interpolant described in Sections 4.4.2 and 4.4.3 respectively. We remind the reader that anisotropic Smolyak algorithm defined (4.4.10) or (4.4.11) is equivalent to the isotropic Smolyak

algorithm given by (4.4.7) or (4.4.8) respectively, when the components of the weight $\boldsymbol{\alpha}$ are equal, i.e. $\alpha_1 = \alpha_2 = \cdots = \alpha_N$. We also define π_h to be the finite element projection operator described by (4.1.2). We will investigate the error

$$\|u - \mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N\| \leq \underbrace{\|u - u_N\|}_{(I)} + \underbrace{\|u_N - \pi_h u_N\|}_{(II)} + \underbrace{\|\pi_h u_N - \mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N\|}_{(III)} \tag{6.0.1}$$

evaluated in the natural norm $L^2_P(\Omega; W(D))$. Since the error functions in (II) and (III) are finite dimensional the natural norm is equivalent to $L^2_\rho(\Gamma^N; W(D))$. By controlling the error in this natural norm we also control the error in the expected value of the solution, for example:

$$\left\|E[u - u^N_{h,\mathbf{p}}]\right\|_{W(D)} \leq E\left[\left\|u - u^N_{h,\mathbf{p}}\right\|_{W(D)}\right] \leq \left\|u - u^N_{h,\mathbf{p}}\right\|_{L^2_P(\Omega; W(D))}.$$

The quantity (I) controls the truncation error for the case where the input data a_N and f_N are suitable truncations of random fields. This contribution to the total error was considered in (Nobile et al., 2006, Section 4.2) and seen in Section 6.1.3. The quantity (I) is otherwise zero if the representation of a_N and f_N is exact, as in Example 1.2.1. The second term (II) controls the convergence with respect to h, i.e. the finite element error, which will be dictated by standard approximability properties of the finite element space $W_h(D)$, given by (4.1.1), and the regularity in space of the solution u, see e.g. Ciarlet (1978); Brenner and Scott (1994). Specifically,

$$\|u_N - \pi_h u_N\|_{L^2_\rho(\Gamma^N; W(D))} \leq C_\pi h^s \left(\int_{\Gamma^N} C(s; u)^2 \rho(y)\, dy\right)^{1/2}$$

by the finite element approximability property (4.1.1).

The full tensor product convergence results are given by (Babuška et al., 2005c, Theorem 1) while the sparse tensor product convergence results for the isotropic and anisotropic Smolyak method can be found in our work, (Nobile et al., 2006, Theorem 4.6 and 4.10) and (Nobile et al., 2007, Theorem 5.7 and 5.13). Therefore, we devote the next two Sections to an extensive explanation of these major results where we will only concern ourselves with the convergence results when implementing the isotropic and anisotropic Smolyak algorithms described in Sections 4.4.2 and 4.4.3 respectively. Namely, our primary concern will be to analyze the interpolation error (III)

$$\|\pi_h u_N - \mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N\|_{L^2_\rho(\Gamma^N; W(D))}, \tag{6.0.2}$$

for both the Clenshaw-Curtis and Gaussian versions of the two sparse grid algorithms.

Under the very reasonable assumption that the semi-discrete finite element solution $\pi_h u_N$ admits an analytic extension as described in Assumption 1.3.1 with the same analyticity region as for u_N, the behavior of the error (6.0.2) will be analogous to $\|u_N - \mathscr{A}_{\boldsymbol{\alpha}}(w, N)u_N\|_{L^2_\rho(\Gamma^N;W(D))}$. For this reason, in the next sections we will analyze the latter.

6.1 Analysis of the sparse interpolation error: The isotropic case

Similar to the approach described in Chapter 5, the technique used to get error bounds for the isotropic Smolyak's algorithm for $N > 1$ utilizes those for the case $N = 1$. Therefore, we first address the case $N = 1$. Let us first recall the best approximation error, described in Lemma 5.0.3, for a function $v : \Gamma^1 \to W(D)$ which admits an analytic extension in the region $\Sigma(\Gamma^1; \tau) = \{z \in \mathbb{C}, \quad \text{dist}(z, \Gamma) < \tau\}$ of the complex plane, for some $\tau > 0$. We will still denote the extension by v; in this case, τ represents the distance between $\Gamma^1 \subset \mathbb{R}$ and the nearest singularity of $v(z)$ in the complex plane. We again recall that $\Gamma^1 = [-1, 1]$ and hence bounded and remind the reader that we can easily extend the analysis of this chapter to include the case of unbounded random variables following Remark 5.0.4.

In the multidimensional case, the size of the analyticity region will depend, in general, on the direction n, for $n = 1, 2, \ldots, N$, and will be denoted by τ_n, as in (2.0.7). The same holds for the decay coefficient ϱ_n. In what follows, we set

$$\varrho \equiv \min_n \varrho_n. \tag{6.1.1}$$

As stated in Section 4.4.2, the Smolyak construction treats all directions equally and is therefore an isotropic algorithm. Moreover, the convergence analysis presented in Sections 6.1.1 and 6.1.2 does not exploit possible anisotropic behaviors of problem (1.1.1). Therefore, we can expect a slower convergence rate for problems that exhibit strong anisotropic effects. This motivated the development of the anisotropic Smolyak method described in Section 4.4.3 and analyzed in Section 6.2. Also see Chapter 7 where we explore numerically the consequences of introducing an anisotropy into the model problems described by Examples 1.1.1 and 1.1.2.

Example 6.1.1 *For the linear problem described in Chapter 2 it was shown in Babuška et al. (2005c) that for a multi-index* $\mathbf{p} = (p_1, \ldots, p_N)$*, a tensor product polynomial interpolation on Gaussian abscissas achieves exponential convergence in each direction* Y_n *and the error can be bounded as*

$$\|u_N - I_{\mathbf{p}}^N u_N\|_{L^2_\rho(\Gamma^N;W(D))} \leq C \sum_{n=1}^{N} \varrho_n{}^{-p_N}. \tag{6.1.2}$$

The constant C *in* (6.1.2) *is independent of* N *and, using* (2.0.7)*, we have*

$$\begin{aligned} \varrho_n &= \frac{2\tau_n}{|\Gamma_n|} + \sqrt{1 + \frac{4\tau_n{}^2}{|\Gamma_n|^2}} \\ &\geq 1 + \frac{2\tau_n}{|\Gamma_n|}. \end{aligned} \tag{6.1.3}$$

where τ_n *can be estimated e.g. as in* (2.0.6) *and* (2.0.7)*.*

For convenience, in the isotropic case only we will let $q = w + N$, where we recall that $w \in \mathbb{N}$. Then, for $q \geq N$ and $\mathbf{i} = (i_1, i_2, \ldots, i_N) \in \mathbb{N}_+^N$ such that $|\mathbf{i}| = i_1 + \cdots + i_N$, the isotropic Smolyak formulas given by (4.4.7) and (4.4.8) can be redefined as:

$$\mathscr{A}(q, N) = \sum_{|\mathbf{i}| \leq q} \left(\Delta^{i_1} \otimes \cdots \otimes \Delta^{i_N}\right)$$

and

$$\mathscr{A}(q, N) = \sum_{q-N+1 \leq |\mathbf{i}| \leq q} (-1)^{q-|\mathbf{i}|} \binom{N-1}{q-|\mathbf{i}|} \cdot \left(\mathscr{U}^{i_1} \otimes \cdots \otimes \mathscr{U}^{i_N}\right).$$

respectively.

6.1.1 Clenshaw-Curtis isotropic estimates

In this section we develop error estimates for interpolating functions $u \in C^0(\Gamma^N; W(D))$ that admit an analytic extension as described by Assumption 1.3.1 using the Smolyak formulations based on the choice (4.4.13) and (4.4.14) described in Section 4.4.4. We remind the reader that in the global estimate (6.0.1) we need to bound the interpolation error (III) in the $L^2_\rho(\Gamma^N; W(D))$ norm. Yet, this norm is always bounded by the $L^\infty(\Gamma^N; W(D))$. Namely, for all $v \in L^\infty(\Gamma^N; W(D))$ we have

$$\|v\|_{L^2_\rho(\Gamma^N;W(D))} \leq \|v\|_{L^\infty(\Gamma^N;W(D))}.$$

In our notation the norm $\|\cdot\|_{\infty,N}$ is shorthand for $\|\cdot\|_{L^\infty(\Gamma^N;W(D))}$ and will be used henceforth. We also define $I_N : \Gamma^N \to \Gamma^N$ as the identity operator on an N-dimensional space.

We begin by letting E_m be the error of the best approximation to functions $u \in C^0(\Gamma^1; W(D))$ by functions $w \in V_m$. Similarly to Barthelmann et al. (2000b), since $\mathscr{U}^i$ is exact on V_{m_i-1} we can apply the general formula

$$\left\|u - \mathscr{U}^i(u)\right\|_{\infty,1} \leq E_{m_i-1}(u) \cdot (1 + \Lambda_{m_i}) \tag{6.1.4}$$

where Λ_m is the Lebesgue constant for our choice (4.4.13). It is known that

$$\Lambda_m \leq \frac{2}{\pi} \log(m-1) + 1 \tag{6.1.5}$$

for $m \geq 2$, see Dzjadyk and Ivanov (1983).

Using Lemma 5.0.3, the best approximation to functions $u \in C^0(\Gamma^1; W(D))$ that admit an analytic extension as described by Assumption 1.3.1 is bounded by:

$$E_{m_i}(u) \leq C\, \varrho^{-m_i} \tag{6.1.6}$$

where C is a constant dependent on τ defined in Lemma 5.0.3. Hence (6.1.4)-(6.1.6) implies

$$\begin{aligned}
\left\|(I_1 - \mathscr{U}^i)(u)\right\|_{\infty,1} &\leq C\, \log(m_i)\varrho^{-m_i} \leq C\, i\varrho^{-2^i}, \\
\left\|(\Delta^i)(u)\right\|_{\infty,1} &= \left\|(\mathscr{U}^i - \mathscr{U}^{i-1})(u)\right\|_{\infty,1} \\
&\leq \left\|(I_1 - \mathscr{U}^i)(u)\right\|_{\infty,1} + \left\|(I_1 - \mathscr{U}^{i-1})(u)\right\|_{\infty,1} \\
&\leq E\, i\varrho^{-2^{i-1}},
\end{aligned}$$

for all $i \in \mathbb{N}_+$ with positive constants C and E depending on u but not on i.

Theorem 6.1.2 *For functions $u \in L^2_\rho(\Gamma^N; W(D))$ that admit an analytic extension as described by Assumption 1.3.1 we obtain*

$$\|(I_N - \mathscr{A}(q,N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq CF^N\, \Psi(q,N)\varrho^{-\frac{p(q,N)}{2}} \tag{6.1.7}$$

where

$$p(q,N) := \begin{cases} N\, 2^{q/N}, & \text{if } q > N\chi, \\ (q-N)\log(2)\cdot 2^\chi, & \text{otherwise,} \end{cases}, \tag{6.1.8}$$

$$\Psi(q,N) := \begin{cases} 1 & \text{if } N = 1 \\ \min\left\{q^{2N-1}, q^3\, e^{q^2}\right\} & \text{otherwise,} \end{cases} \tag{6.1.9}$$

and $\chi = \left(\frac{1+\log(2)}{\log(2)}\right)$.

Proof. First we define, for $j \geq 1$ and $s \geq d$, the two functions

$$f(s,j) := \frac{1}{2}\left(j2^{s/j} + 2^{q-N+j+2-s}\right)$$

and

$$g(s,j) := \sum_{i\in\mathbb{N}_+,|\mathbf{i}|=s} \prod_{n=1}^{j} i_n.$$

We begin by claiming that

$$\|(I_N - \mathscr{A}(q,N))\,(u)\|_{\infty,N} \leq C\sum_{j=1}^{N-1} E^j \sum_{s=j}^{q-N+j} g(s,j)\,(q-N+j+1-s)\,\varrho^{-f(s,j)} + \|(I_1 - \mathscr{A}(q-N+1,1))(u)\|_{\infty,N} \quad (6.1.10)$$

where, for the trivial case, we get

$$\|(I_1 - \mathscr{A}(q-N+1,1))(u)\|_{\infty,N} = \left\|(I_1 - \mathscr{U}^{q-N+1})(u)\right\|_{\infty,N} \leq C(q-N+1)\varrho^{-2^{q-N+1}} \leq Cq\varrho^{-2^{q-N+1}}.$$

This error estimate is computed inductively. For $N > 1$ we use recursively,

$$\begin{aligned} I_{N+1} - \mathscr{A}(q+1,N+1) &= I_{N+1} - \sum_{|\mathbf{i}|\leq q}\left(\bigotimes_{n=1}^{N}\Delta^{i_n}\otimes\mathscr{U}^{q+1-|\mathbf{i}|}\right) \\ &= \sum_{|\mathbf{i}|\leq q}\left(\bigotimes_{n=1}^{N}\Delta^{i_n}\otimes\left(I_1 - \mathscr{U}^{q+1-|\mathbf{i}|}\right)\right) + (I_N - \mathscr{A}(q,N))\otimes I_1. \end{aligned}$$

Furthermore,

$$\begin{aligned} &\left\|\sum_{|\mathbf{i}|\leq q}\left(\bigotimes_{n=1}^{N}(\Delta^{i_n})(u)\otimes\left(I_1 - \mathscr{U}^{q+1-|\mathbf{i}|}\right)(u)\right)\right\|_{\infty,N} \\ &\leq \sum_{|\mathbf{i}|\leq q}\prod_{n=1}^{N}\left\|(\Delta^{i_n})(u)\right\|_{\infty,N}\left\|(I_1 - \mathscr{U}^{q+1-|\mathbf{i}|})(u)\right\|_{\infty,N} \\ &\leq CE^N\sum_{|\mathbf{i}|\leq q}\left(\prod_{n=1}^{N} i_n\right)\varrho^{-\sum_{n=1}^{N}2^{i_n-1}}(q+1-|\mathbf{i}|)\,\varrho^{-2^{q+1-|\mathbf{i}|}} \\ &\leq CE^N\sum_{|\mathbf{i}|\leq q}\left(\prod_{n=1}^{N} i_n\right)(q+1-|\mathbf{i}|)\,\varrho^{-\frac{1}{2}\left(N2^{|\mathbf{i}|/N}+2^{q+2-|\mathbf{i}|}\right)} \\ &\leq CE^N\sum_{s=N}^{q} g(s,N)\,(q+1-s)\,\varrho^{-f(s,N)} \end{aligned}$$

where we have used the convexity estimate

$$\varrho^{-\sum_{n=1}^{N} 2^{in}} \leq \varrho^{-N2^{|\mathbf{i}|/N}}.$$

Then, by the inductive assumption (6.1.10),

$$\begin{aligned}\|((I_N - \mathscr{A}(q,N)) \otimes I_1)(u)\|_{\infty,N} \leq C \sum_{j=1}^{N-1} E^j \sum_{s=j}^{q-N+j} g(s,j)\,(q-N+j+1-s)\,\varrho^{-f(s,j)} \\ + \|(I_1 - \mathscr{A}(q-N+1,1))(u)\|_{\infty,N}.\end{aligned}$$

Therefore,

$$\begin{aligned}\|(I_{N+1} - \mathscr{A}(q+1,N+1))(u)\|_{\infty,N} \leq C \sum_{j=1}^{N} E^j \sum_{s=j}^{q-N+j} g(s,j)\,(q-N+j+1-s)\,\varrho^{-f(s,j)} \\ + \|(I_1 - \mathscr{A}(q-N+1,1))(u)\|_{\infty,N}.\end{aligned}$$

and (6.1.10) is proved. Set $F = \max\{1, E\}$ to obtain

$$\begin{aligned}\|(I_N - \mathscr{A}(q,N))(u)\|_{\infty,N} &\leq C \sum_{j=1}^{N-1} (\max\{1,E\})^N \sum_{s=j}^{q-N+j} g(s,j)\,(q-N+j+1-s)\,\varrho^{-f(s,j)} \\ &\quad + \|(I_1 - \mathscr{A}(q-N+1,1))(u)\|_{\infty,N} \\ &\leq CF^N \sum_{j=1}^{N-1} \sum_{s=j}^{q-N+j} g(s,j)\,(q-N+j+1-s)\,\varrho^{-f(s,j)} \\ &\quad + \|(I_1 - \mathscr{A}(q-N+1,1))(u)\|_{\infty,N}.\end{aligned} \tag{6.1.11}$$

We now turn our attention to finding a maximum for $\varrho^{-f(s,j)}$ on the set $\{(s,j) : j \leq s \leq q-N+j \text{ and } 1 \leq j \leq N-1\}$. Clearly

$$\frac{\partial f}{\partial s} = \left(2^{s/j} - 2^{q-N+j+2-s}\right)\log(2) = 0$$

implies that $s = s(j) = j + \frac{j(q-N+1)}{j+1}$, which satisfies for any $j \in \mathbb{N}_+$

$$j \leq s(j) \leq q-N+1+j.$$

Hence,

$$\begin{aligned}\max_{j \leq s \leq q-N+j} \varrho^{-f(j,s)} &= \varrho^{-f(j,s(j))} \\ &\leq \varrho^{-h(j)}\end{aligned}$$

where $h(j) = (j+1)2^{(q-N+1)/(j+1)}$. Then we get

$$\frac{\mathrm{d}h}{\mathrm{d}j} = 2^{(q-N+1)/(j+1)}\left(1 - \frac{(q-N+1)\log(2)}{j+1}\right) = 0$$

which yields $j = (q-N+1)\log(2) - 1$. For q sufficiently large, the minimum of $h(j)$ falls outside the interval $[1, N-1]$ and the function $h(j)$ is decreasing on this interval. Therefore, there are two cases to consider, the first being the situation when $q > N\left(\frac{1+\log(2)}{\log(2)}\right) = N\chi$ and the second when $N \leq q \leq N\chi$. In either case

$$\max_{1\leq j\leq N-1} \varrho^{-h(j)} = \varrho^{-p(q,N)},$$

hence

$$\max_{\substack{1\leq j\leq N-1\\ j\leq s\leq q-N+j}} \varrho^{-f(j,s)} \leq \varrho^{-p(q,N)}.$$

In conclusion we have, for $q \geq N$

$$\begin{aligned}\|(I_N - \mathscr{A}(q,N))\,(u)\|_{\infty,N} &\leq CF^N\varrho^{-p(q,N)}\sum_{j=1}^{N-1}\sum_{s=j}^{q-N+j} g(s,j)(q-N+j+1-s)\\ &= CF^N\varrho^{-p(q,N)}\,\kappa + \|(I_1 - \mathscr{A}(q-N+1,1))(u)\|_{\infty,N}\end{aligned} \tag{6.1.12}$$

and

$$\begin{aligned}\kappa &= \sum_{j=1}^{N-1}\sum_{s=j}^{q-N+j} g(s,j)(q-N+j+1-s)\\ &= \sum_{j=1}^{N-1}\sum_{s=j}^{q-N+j}(q-N+j+1-s)\sum_{i\in\mathbb{N}_+^j,\,|\mathbf{i}|=s}\left(\prod_{n=1}^{j} i_n\right)\\ &\leq \sum_{j=1}^{N-1}\sum_{s=j}^{q-N+j}(q-N+j+1-s)\sum_{i\in\mathbb{N}_+^j,\,|\mathbf{i}|=s}\left(\frac{|\mathbf{i}|}{j}\right)^j\\ &= \sum_{j=1}^{N-1}\sum_{s=j}^{q-N+j}(q-N+j+1-s)\left(\frac{s}{j}\right)^j\binom{s-1}{j-1}\\ &\leq (q-N+1)\sum_{j=1}^{N-1}\sum_{s=j}^{q-N+j}\left(\frac{q-N+j}{j}\right)^j\frac{(s-1)^{j-1}}{(j-1)!}\\ &\leq (q-N+1)(q-N)\sum_{j=1}^{N-1}(q-1)^j\frac{(q-N+j)^{j-1}}{(j-1)!}\end{aligned}$$

$$
\begin{aligned}
&\leq (q-N+1)(q-N)\sum_{j=1}^{N-1}\frac{(q-1)^{2j-1}}{(j-1)!} \\
&= (q-N+1)(q-N)(q-1)\sum_{j=0}^{N-2}\frac{(q-1)^{2j}}{j!}.
\end{aligned}
\tag{6.1.13}
$$

Since the sum $\sum_{j=0}^{N-2}\frac{(q-1)^{2j}}{j!}$ can be bounded by $e^{(q-1)^2}$ or by $\frac{(q-1)^{2N-2}}{q(q-2)}$, then in either case $\kappa \leq \Psi(q,N)$. Finally, from (6.1.11) and using (6.1.12) and (6.1.13) we conclude that

$$
\|(I_N - \mathscr{A}(q,N))\,(u)\|_{\infty,N} \leq CF^N\,\Psi(q,N)\varrho^{-\frac{p(q,N)}{2}} + \varrho^{-2^{q-N+1}}. \tag{6.1.14}
$$

We also observe that $\Psi(q,N) \geq 1$ and by straightforward calculations

$$
2^{q-N+1} \geq \frac{p(q,N)}{2}, \quad \forall N \geq 1, q \geq N,
$$

to conclude that

$$
\varrho^{-2^{q-N+1}} \leq \Psi(q,N)\varrho^{-\frac{p(q,N)}{2}}
$$

and this completes the proof. □

Now we relate the number of collocation points $\eta = \eta(q,N) = \#\mathscr{H}(q,N)$ to the level q of the Smolyak algorithm. We state the result in the following lemma.

Lemma 6.1.3 *Using the Smolyak interpolant described by (4.4.7) where the abscissas are the Clenshaw-Curtis knots, described in Section 4.4.4, the total number of points required at level q satisfies the following bounds:*

$$
2^{q-N+1} \leq \eta \leq \frac{2^q q^N}{(N-1)!}. \tag{6.1.15}
$$

Proof. The proof follows immediately but will be shown for completeness. By using formula (4.4.7) and exploiting the nested structure of the Clenshaw-Curtis abscissas the number of points $\eta = \eta(q,N) = \#\mathscr{H}(q,N)$ can be counted in the following way:

$$
\eta = \sum_{|\mathbf{i}|\leq q}\prod_{n=1}^{N} r(i_n), \text{ where } r(i) := \begin{cases} 1 & \text{if } i=1 \\ 2 & \text{if } i=2 \\ 2^{i-2} & \text{if } i>2 \end{cases}. \tag{6.1.16}
$$

If we take $i_1 = i_2 = \ldots = i_{N-1} = 1$ then to satisfy the constraint $|\mathbf{i}| \leq q$ we required

$i_N \leq q - N + 1$. Then we get

$$2^{q-N-1} \leq \eta = \sum_{|\mathbf{i}|\leq q} \prod_{n=1}^{N} r(i_n) \leq \sum_{|\mathbf{i}|\leq q} 2^{|\mathbf{i}|} \leq \sum_{j=N}^{q} \sum_{|\mathbf{i}|=j} 2^j = \sum_{j=N}^{q} 2^j \binom{j-1}{N-1}$$
$$\leq \sum_{j=N}^{q} 2^q \frac{(q-1)^{N-1}}{(N-1)!}$$
$$\leq \frac{2^q q^N}{(N-1)!}$$

which completes the proof. □

The next Theorem relates the error bound (6.2.10) to the number of collocation points $\eta = \eta(q, N) = \#\mathscr{H}(q, N)$, described by Lemma 6.1.3.

Theorem 6.1.4 *Assume the conditions of Lemma 6.1.2 and Lemma 6.1.3, and define the function*

$$\gamma(\eta, N) = \log_2(\eta) + N + 1,$$

then for $N \leq q < N\chi$

$$\|(I_N - \mathscr{A}(q, N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq CF^N\,\Psi(\gamma(\eta, N), N)\left(\frac{(2\gamma(\eta, N))^N}{\eta\,(N-1)!}\right)^{\frac{\Theta}{2}\log(\varrho)} \tag{6.1.17}$$

and for $q \geq N\chi$

$$\|(I_N - \mathscr{A}(q, N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq CF^N\,\Psi(\gamma(\eta, N), N)\,\varrho^{-\frac{N}{2}[(N-1)!]^{1/N}\frac{\eta^{1/N}}{\gamma(\eta,N)}}, \tag{6.1.18}$$

where $\Theta = 2^\chi$ and $\eta = \eta(q, N)$ is the number of knots that are used by $\mathscr{A}(q, N)$ and Ψ was defined in (6.1.9).

Proof. Recall that the error bound will be separated into two estimates depending on the domain of definition of (6.1.8). First for $N \leq q < N\chi$ and using (6.1.15) we arrive at

$$\log(\eta) + \log((N-1)!) \leq q\log(2) + N\log(q)$$
$$\leq q\log(2) + N\log\gamma(\eta, N)$$
$$= (q-N)\log(2) + N\log(2\gamma(\eta, N)).$$

Hence,

$$q - N \geq \frac{\log(\eta) + \log((N-1)!) - N\log(2\gamma(\eta, N))}{\log(2)}$$

and using (6.1.8) implies that

$$
\begin{aligned}
p(q,N) &\geq (\log(\eta) + \log((N-1)!) - N\log(2\gamma(\eta,N))) \cdot \Theta \\
&= \Theta \, \log\left(\frac{\eta\,(N-1)!}{(2\gamma(\eta,N))^N}\right).
\end{aligned}
$$

Therefore, using (6.1.15) we deduce that

$$
\begin{aligned}
\|(I_N - \mathscr{A}(q,N))\,(u)\|_{\infty,N} &\leq CF^N\,\Psi(\gamma(\eta,N),N)\,\varrho^{-\Theta\,\log\left(\frac{(N-1)!+\eta}{(2\gamma(\eta,N))^N}\right)} \\
&= CF^N\,\Psi(\gamma(\eta,N),N)\,\left(e^{-\log(\varrho)\,\log\left(\frac{(N-1)!+\eta}{(2\gamma(\eta,N))^N}\right)}\right)^{\Theta} \\
&= CF^N\,\Psi(\gamma(\eta,N),N)\,\left(\frac{(2\gamma(\eta,N))^N}{\eta(N-1)!}\right)^{\Theta\cdot\log(\varrho)}
\end{aligned}
\tag{6.1.19}
$$

and we recover (6.1.17).

On the other hand, for $q \geq N\chi$ and using (6.1.15) we find that

$$
\left(\frac{2^q q^N}{(N-1)!}\right)^{1/N} \geq \eta^{1/N}
$$

which implies that

$$
2^{q/N} \geq \frac{(\eta(N-1)!)^{1/N}}{\gamma(\eta,N)}
$$

and

$$
p(q,N) \geq N\,[(N-1)!]^{1/N}\,\frac{\eta^{1/N}}{\gamma(\eta,N)}.
$$

Therefore, again with (6.1.15) we conclude that

$$
\|(I_N - \mathscr{A}(q,N))\,(u)\|_{\infty,N} \leq CF^N\,\Psi(\gamma(\eta,N),N)\varrho^{-\frac{N}{2}[(N-1)!]^{1/N}\frac{\eta^{1/N}}{\gamma(\eta,N)}}. \tag{6.1.20}
$$

and we recover (6.1.18). □

6.1.2 Gaussian isotropic estimates

Similarly to the previous section we now develop error estimates for interpolating functions $u \in C^0(\Gamma^N; W(D))$ that admit an analytic extension as described by Assumption 1.3.1 using the Smolyak formulations based on Gaussian abscissas described in Section 4.4.5. As before, we remind the reader that in the global estimate (6.0.1) we need to bound the interpolation

error (III) in the norm $L^2_\rho(\Gamma^N; W(D))$. Yet, the Gaussian points defined in Section 4.4.5 are constructed for the more appropriate density $\hat{\rho} = \prod_{n=1}^N \hat{\rho}_n$ and we have

$$\|v\|_{L^2_{\hat{\rho}}(\Gamma^N;W(D))} \leq \left\| \frac{\rho}{\hat{\rho}} \right\|_{L^\infty(\Gamma^N)} \cdot \|v\|_{L^2_{\hat{\rho}}(\Gamma^N;W(D))} \quad \text{for all } v \in C^0(\Gamma^N; W(D)).$$

In what follows we will use the shorthand notation $\|\cdot\|_{\hat{\rho},N}$ for $\|\cdot\|_{L^2_{\hat{\rho}}(\Gamma^N;W(D))}$. Utilizing the work of Erdös and Turan (1937) we present the following lemma:

Lemma 6.1.5 *For every function $u \in C^0(\Gamma^1; W(D))$ the interpolation error satisfies*

$$\|u - \mathscr{U}^i(u)\|_{\hat{\rho},1} \leq 2\sqrt{C_{\hat{\rho}}} \inf_{w \in V_{m_i}} \|u - w\|_{\infty,1},$$

where $C_{\hat{\rho}} = \displaystyle\int_{\Gamma^1} \hat{\rho}(y)\, dy$.

Proof. We have, indeed, for any $v \in V_{m_i}$

$$\begin{aligned} \|u - \mathscr{U}^i(u)\|^2_{\hat{\rho},1} &= \left\|u - v + (v - \mathscr{U}^i(u))\right\|^2_{\hat{\rho},1} \\ &= \left\|u - v + \mathscr{U}^i(v - u)\right\|^2_{\hat{\rho},1} \\ &\leq 2\left(\|u - v\|^2_{\hat{\rho},1} + \left\|\mathscr{U}^i(u - v)\right\|^2_{\hat{\rho},1}\right) \end{aligned} \tag{6.1.21}$$

where we observe that $\forall v \in V_{m_i}$, it holds $\mathscr{U}^i(v) = v$. Then it is easy to see that

$$\begin{aligned} \|u - v\|^2_{\hat{\rho},1} &\leq \int_{\Gamma^1} \hat{\rho}(y)\, |(u - v)(y)|^2\, dy \\ &\leq \|u - v\|^2_{\infty,1} \int_{\Gamma^1} \hat{\rho}(y)\, dy = C_{\hat{\rho}}\, \|u - v\|^2_{\infty,1} \end{aligned}$$

and

$$\begin{aligned} \left\|\mathscr{U}^i(u - v)\right\|^2_{\hat{\rho},1} &= \left\| \sum_{j=1}^{m_i} (u - v)(y^i_j) l^i_j(y) \right\|^2_{\hat{\rho},1} \\ &\leq \sum_{j,j'=1}^{m_i} \left|(u - v)(y^i_j)\right| \left|(u - v)(y^i_{j'})\right| \int_{\Gamma^1} \hat{\rho}(y) l^i_j(y) l^i_{j'}(y)\, dy \\ &\leq \|u - v\|^2_{\infty,1} \sum_{j=1}^{m_i} \int_{\Gamma^1} \hat{\rho}(y)(l^i_j(y))^2\, dy = \left(\int_{\Gamma^1} \hat{\rho}(y) dy\right) \|u - v\|^2_{\infty,1} \end{aligned}$$

where we exploit the orthogonality of the Lagrange polynomial basis. Then from (6.1.21) we conclude that

$$\|u - \mathscr{U}^i(u)\|^2_{\hat{\rho},1} \leq 4 \left(\int_{\Gamma^1} \hat{\rho}(y) dy\right) \|u - v\|^2_{\infty,1}$$

and the result follows directly. $\square$

Similar to Section 6.1.1 we let E_m be the error of the best approximation to functions $u \in C^0(\Gamma^1; W(D))$ that admit an analytic extension as described by Assumption 1.3.1 by functions $w \in V_m$. Then, from Lemma 6.1.5 we begin with

$$\left\| u - \mathscr{U}^i(u) \right\|_{\hat{\rho},1} \leq 2 \sqrt{\left(\int_{\Gamma^1} \hat{\rho}(y) dy \right)} E_{m_i - 1}(u). \tag{6.1.22}$$

Again, from Lemma 5.0.3 the best approximation is bounded by :

$$E_{m_i}(u) \leq C \, \varrho^{-m_i} \tag{6.1.23}$$

where C is a constant dependent on τ defined in Lemma 5.0.3. Hence (6.1.22) and (6.1.23) imply

$$\begin{aligned} \left\| (I_1 - \mathscr{U}^i)(u) \right\|_{\hat{\rho},1} &\leq \widetilde{C} \, \varrho^{-2^i}, \\ \left\| (\Delta^i)(u) \right\|_{\hat{\rho},1} &= \left\| (\mathscr{U}^i - \mathscr{U}^{i-1})(u) \right\|_{\hat{\rho},1} \\ &\leq \left\| (I_1 - \mathscr{U}^i)(u) \right\|_{\hat{\rho},1} + \left\| (I_1 - \mathscr{U}^{i-1})(u) \right\|_{\hat{\rho},1} \\ &\leq \widetilde{E} \, \varrho^{-2^{i-1}} \end{aligned}$$

for all $i \in \mathbb{N}_+$ with positive constants $\widetilde{C}$ and $\widetilde{E}$ depending on u but not on i. We then present the following lemma and theorem whose proofs follow, with minor changes, those given in Lemma 6.1.2 and Theorem 6.1.4 respectively.

Lemma 6.1.6 *For functions $u \in L^2_\rho(\Gamma^N; W(D))$ that admit an analytic extension as described by Assumption 1.3.1 we obtain*

$$\left\| \left(I_N - \mathscr{A}(q, N) \right)(u) \right\|_{L^2_\rho(\Gamma^N; W(D))} \leq \left\| \rho / \hat{\rho} \right\|_{L^\infty(\Gamma^N)} \widetilde{C} \widetilde{F}^N \, \widetilde{\Psi}(q, N) \varrho^{-\frac{p(q,N)}{2}} \tag{6.1.24}$$

where

$$p(q, N) := \begin{cases} N \, 2^{q/N}, & \text{if } q > N\chi, \\ (q - N) \log(2) \cdot 2^\chi, & \text{otherwise} \end{cases}, \tag{6.1.25}$$

$$\widetilde{\Psi}(q, N) := \begin{cases} 1 & \text{if } N = 1 \\ \min\left\{ q^{N-2}, q e^q \right\} & \text{otherwise} \end{cases} \tag{6.1.26}$$

and $\chi = \left(\frac{1 + \log(2)}{\log(2)} \right)$.

Now we relate the number of collocation points $\eta = \eta(q, N) = \#\mathscr{H}(q, N)$ to the level q of the Smolyak algorithm. We state the result in the following lemma:

Lemma 6.1.7 *Using the Smolyak interpolant described by (4.4.8) where the abscissas are the Gaussian knots described in Section 4.4.5, the total number of points required at level q satisfies the following bounds:*

$$2^{q-N} \leq \eta \leq \frac{2^q q^N}{(N-1)!}. \tag{6.1.27}$$

Proof. The proof follows immediately but will be shown for completeness. By using formula (4.4.8), where we collocate using the Gaussian abscissas, the number of points $\eta = \eta(q,N) = \#\mathscr{H}(q,N)$, can be counted in the following way:

$$\eta = \sum_{q-N+1 \leq |\mathbf{i}| \leq q} \prod_{n=1}^{N} \widetilde{r}(i_n), \text{ where } 2^{i-1} \leq \widetilde{r}(i) := \begin{cases} 1 & \text{if } i = 1 \\ 2^{i-1}+1 & \text{if } i \geq 2 \end{cases}. \tag{6.1.28}$$

If we take $i_1 = i_2 = \ldots = i_{N-1} = 1$, then to satisfy the constraint $q - N + 1 \leq |\mathbf{i}| \leq q$ we required $i_N \leq q - N + 1$. Then we get

$$\begin{aligned} 2^{q-N} \leq 2^{q-N} + 1 \leq \eta = \sum_{q-N+1 \leq |\mathbf{i}| \leq q} \prod_{n=1}^{N} \widetilde{r}(i_n) \leq \sum_{|\mathbf{i}| \leq q} 2^{|\mathbf{i}|} \leq \sum_{j=N}^{q} \sum_{|\mathbf{i}|=j} 2^j &= \sum_{j=N}^{q} 2^j \binom{j-1}{N-1} \\ &\leq \sum_{j=N}^{q} 2^q \frac{(q-1)^{N-1}}{(N-1)!} \\ &\leq \frac{2^q q^N}{(N-1)!} \end{aligned}$$

which completes the proof. □

Finally, the next Theorem relates the error bound (6.1.24) to the number of collocation points $\eta = \eta(q,N) = \#\mathscr{H}(q,N)$, described by Lemma 6.1.7.

Theorem 6.1.8 *Assume the conditions of Lemma 6.1.6 and 6.1.7, and define the function*

$$\widetilde{\gamma}(\eta, N) = \log_2(\eta) + N,$$

then for $N \leq q < N\chi$ there holds

$$\|(I_N - \mathscr{A}(q,N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq \widehat{C}\widetilde{F}^N\,\widetilde{\Psi}(\widetilde{\gamma}(\eta,N),N)\left(\frac{(2\widetilde{\gamma}(\eta,N))^N}{\eta(N-1)!}\right)^{\frac{\Theta}{2}\log(\varrho)} \tag{6.1.29}$$

and for $q \geq N\chi$

$$\|(I_N - \mathscr{A}(q,N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq \widehat{C}\widetilde{F}^N\,\widetilde{\Psi}(\widetilde{\gamma}(\eta,N),N)\varrho^{-\frac{N}{2}[(N-1)!]^{1/N}\frac{\eta^{1/N}}{\widetilde{\gamma}(\eta,N)}}, \tag{6.1.30}$$

where $\Theta = 2^\chi$, $\widehat{C} = \widetilde{C}\,\|\rho/\hat{\rho}\|_{L^\infty(\Gamma^N)}$ and $\eta = \eta(q,N)$ is the number of knots that are used by $\mathscr{A}(q,N)$ and $\widetilde{\Psi}$ was defined in (6.1.26).

6.1.3 Influence of truncation errors

In this Section we consider the case where the coefficients a_N and f_N from (1.2.1) are suitably truncated random fields. In this case the truncation error $u - u_N$ is nonzero and contributes to the total error. Such contribution should be considered as well as the relationship between this error and the discretization error.

To this end, if we take the level q to be dimension dependent, i.e. $q = \alpha N$ where $\alpha \geq \chi$ is some constant, then we can estimate the total error $\|u - \mathscr{A}(q,N)(u_N)\|_{L^2_P(\Omega;W(D))}$ in terms of N only. Consider first the case of Gaussian abscissas described in Section 4.4.5. The following theorem holds:

Theorem 6.1.9 *Let $q = \alpha N$ such that $\alpha \geq \chi$ and $\zeta(N)$ is a monotonic decreasing function of N such that $\zeta(N) \to 0$ as $N \to \infty$. Further define $\beta(\alpha) = \alpha + \log(\widetilde{F}) - \log(\varrho)\, 2^{\alpha-1}$, $\widetilde{\alpha}$ the solution to $\beta(\widetilde{\alpha}) = 0$ and $\alpha > \max\{\chi, \widetilde{\alpha}\}$. Under the assumptions of Lemma 6.1.6 and Theorem 6.1.8 and the further assumption that*

$$\|u - u_N\|_{L^2_P(\Omega;W(D))} \leq \zeta(N)$$

where $u \in L^2_P(\Omega;W(D))$ and $u_N \in L^2_\rho(\Gamma^N;W(D))$, we get

$$\|u - \mathscr{A}(q,N)(u_N)\|_{L^2_P(\Omega;W(D))} \leq \zeta(N) + \alpha \widehat{C} N\, e^{\beta(\alpha)N}. \tag{6.1.31}$$

Proof. We begin by writing the total error when approximating $u \in C^0(\Omega;W(D))$ by its N-dimensional interpolant $\mathscr{A}(q,N)(u_N)$. That is, we want to understand

$$\begin{aligned}\|u - \mathscr{A}(q,N)(u_N)\|_{L^2_P(\Omega;W(D))} &\leq \|u - u_N\|_{L^2_P(\Omega;W(D))} + \|(I_N - \mathscr{A}(q,N)(u_N))\|_{L^2_P(\Omega;W(D))} \\ &= \underbrace{\|u - u_N\|_{L^2_P(\Omega;W(D))}}_{\text{(I)}} + \underbrace{\|(I_N - \mathscr{A}(q,N)(u_N))\|_{L^2_\rho(\Gamma^N;W(D))}}_{\text{(II)}}\end{aligned} \tag{6.1.32}$$

By the assumption, the first term (I) is bounded by $\zeta(N)$ for all N and from Lemma 6.1.6, and by the assumption $q = \alpha N \geq \chi N$, the second term (II) can be bounded by

$$\begin{aligned}\|(I_N - \mathscr{A}(q,N))\,(u_N)\|_{L^2_\rho(\Gamma^N;W(D))} &\leq \widehat{C}\widetilde{F}^N q\, e^q \varrho^{-\frac{N}{2}\, 2^{q/N}} \\ &\leq \widehat{C}\widetilde{F}^N \alpha N\, e^{\alpha N} \varrho^{-N\, 2^{\alpha-1}} \\ &\leq \alpha \widehat{C}\widetilde{F}^N N\, e^{\alpha N - \log(\varrho) N\, 2^{\alpha-1}} \\ &= \alpha \widehat{C} N\, e^{N(\alpha + \log(\widetilde{F}) - \log(\varrho)\, 2^{\alpha-1})} \\ &= \alpha \widehat{C} N\, e^{\beta(\alpha)N}\end{aligned} \tag{6.1.33}$$

where $\beta(\alpha) = \alpha + \log(\widetilde{F}) - \log(\varrho)\, 2^{\alpha}$; such that, for sufficiently large α, $\beta(\alpha)$ is negative. With $\alpha > \max\{\chi, \widetilde{\alpha}\}$ equation (6.1.33) becomes

$$\|(I_N - \mathscr{A}(q,N))\,(u_N)\|_{L^2_\rho(\Gamma^N;W(D))} \leq \alpha \widehat{C} N\, e^{\beta(\alpha)N},$$

which substituted into the total error (6.1.32) yields

$$\begin{aligned}\|u - \mathscr{A}(q,N)(u)\|_{L^2_P(\Omega;W(D))} &\leq \|u - u_N\|_{L^2_P(\Omega;W(D))} + \|(I_N - \mathscr{A}(q,N)(u_N))\|_{L^2_\rho(\Gamma^N;W(D))} \\ &\leq \zeta(N) + \alpha \widehat{C} N\, e^{\beta(\alpha)N}\end{aligned}$$

as required by (6.1.31). □

We want to understand the cases where (II) is negligible when compared with (I). In Theorem 6.1.9 we assume that the truncation error $\|u - u_N\|_{L^2_P(\Omega;W(D))}$ is bounded by $\zeta(N)$ for all N. The function $\zeta(N)$ is typically related to the decay of the eigenvalues if one truncates the noise with a Karhunen-Loève expansion, see Frauenfelder et al. (2005b). For example, if

$$\|u - u_N\|_{L^2_P(\Omega;W(D))} \leq \theta N^{-r}, \text{ for } r > 0,$$

for some constant θ, then

$$\|u - \mathscr{A}(q,N)(u)\|_{L^2_P(\Omega;W(D))} \leq \underbrace{\theta N^{-r}}_{\text{(I)}} + \underbrace{\alpha \widehat{C} N\, e^{\beta(\alpha)N}}_{\text{(II)}}.$$

In such a situation the Smolyak error (II) is asymptotically negligible with respect to the truncation (I) as $N \to \infty$. Therefore, the isotropic Smolyak algorithm is an efficient interpolation scheme to choose in computational experiments. On the other hand, if

$$\|u - u_N\|_{L^2_P(\Omega;W(D))} \leq \theta e^{-\gamma N} \text{ where } \gamma > \beta(\alpha)$$

then

$$\|u - \mathscr{A}(q,N)(u)\|_{L^2_P(\Omega;W(D))} \leq \underbrace{\theta e^{-\gamma N}}_{\text{(I)}} + \underbrace{\alpha \widehat{C} N\, e^{\beta(\alpha)N}}_{\text{(II)}},$$

which implies that the truncation error (I) is dominated by the Smolyak error (II). In this case the Smolyak algorithm is an inadequate interpolation scheme and improvements to this algorithm must be investigated. To facilitate faster convergence of such problems we utilize the anisotropic Smolyak algorithm, described in Section 4.4.3 and analyzed in the next section.

Remark 6.1.10 *In the situation in which the Clenshaw-Curtis abscissas are used, the term (II) in* (6.1.32) *can be bounded as*

$$\begin{aligned}\|(I_N - \mathscr{A}(q,N)(u_N))\|_{L^2_\rho(\Gamma^N;W(D))} &\leq C\, F^N q^2 e^{q^2} \varrho^{-\frac{N}{2}2^{q/N}} \\ &= C q^2 e^{N\log(F)+q^2-\frac{N}{2}\log\varrho 2^{q/N}}.\end{aligned}$$

In the presence of the term q^2, global convergence can only be achieved if one takes $q = N^{1+\alpha}$ with $\alpha > 0$.

6.2 Analysis of the sparse interpolation error: The anisotropic case

Similar to the previous section, In order to get error bounds for the anisotropic Smolyaks algorithm in the multidimensional case, we will connect the general case to the case $N = 1$ and then successively apply Lemma 5.0.3 .

As stated in Section 4.4.3, the sparse grid construction treats all directions differently and is therefore an anisotropic algorithm. Moreover, the convergence analysis presented in Section 6.2.1 exploits the possible anisotropic behaviors of problem (1.1.1). Therefore, we can expect a faster convergence rate when compared to our previous isotropic Smolyak algorithm (Nobile et al., 2006), described in Section 6.1, for such problems that exhibit strong anisotropic effects. Since the algorithm exploits this behavior and is a sparse interpolation technique, a similar conclusion can be drawn when making convergence comparisons with the anisotropic full tensor product method, introduced in the work Babuška et al. (2005c). See Chapter 7 where we explore numerically the consequences of introducing an anisotropy into the model problem described by Examples 1.1.1 and 1.1.2 .

6.2.1 Clenshaw-Curtis anisotropic estimates

For $n = 1, 2, \ldots, N$ we begin by recalling that we previously defined E_m to be the error of the best approximation to functions $u \in C^0(\Gamma_n; W(D))$ by functions $w \in V_m$. Similarly to Section 6.1.1, for $n = 1, \ldots, N$, since $\mathscr{U}^{i_n}$ is exact on $V_{m_{i_n}-1}$ we can apply the general formula

$$\left\|u - \mathscr{U}^{i_n}(u)\right\|_{\infty,1} \leq E_{m_{i_n}-1}(u) \cdot \left(1 + \Lambda_{m_{i_n}}\right) \tag{6.2.1}$$

where Λ_m is the Lebesgue constant for our choice (4.4.13), defined previously by (6.1.5).

Using Lemma 5.0.3 and with the assumption (5.0.3) on the decay coefficients, the best approximation to functions $u \in C^0(\Gamma^n; W(D))$ that admit an analytic extension as described by Assumption 1.3.1 is bounded by:

$$E_{m_{i_n}}(u) \leq C\, \varrho_n^{-m_{i_n}} = C\, e^{-g(n)\, m_{i_n}}, \tag{6.2.2}$$

where C is a constant dependent on τ and u but otherwise independent of n, defined in Lemma 5.0.3. For $n = 1, 2, \ldots, N$ we define the one-dimensional indentity operator $I_1^{(n)} : \Gamma_n \to \Gamma_n$, then (6.2.1)-(6.2.2) implies

$$\left\| (I_1^{(n)} - \mathscr{U}^{i_n})(u) \right\|_{\infty,1} \leq C\, \log(m_{i_n}) \varrho_n^{-m_{i_n}} \leq C\, i_n \varrho_n^{-2^{i_n}} = C\, i_n e^{-g(n) 2^{i_n}},$$

$$\begin{aligned} \left\| (\Delta^{i_n})(u) \right\|_{\infty,1} = \left\| (\mathscr{U}^{i_n} - \mathscr{U}^{i_n - 1})(u) \right\|_{\infty,1} &\leq \left\| (I_1^{(n)} - \mathscr{U}^{i_n})(u) \right\|_{\infty,1} + \left\| (I_1^{(n)} - \mathscr{U}^{i_n - 1})(u) \right\|_{\infty,1} \\ &\leq E\, i_n \varrho_n^{-2^{i_n - 1}} = E\, i_n e^{-\frac{g(n)}{2} 2^{i_n}} \end{aligned} \tag{6.2.3}$$

for all $i \in \mathbb{N}_+$ with positive constants C and E depending on u but not on i or n.

The convergence proof will be split in several pieces, the main results being given in Theorems 6.2.3 and 6.2.7 which state the convergence rates in terms of the level w and the total number of collocation points, respectively.

For the purpose of error analysis we first introduce the set

$$\tilde{X}_{\boldsymbol{\alpha}}(w, d) := \left\{ \mathbf{i} \in \mathbb{N}_+^d, \mathbf{i} \geq \mathbf{1} \ : \ \left(\sum_{n=1}^{d} (i_n - 1)\alpha_n - \underline{\alpha} w \right) \in [0, \alpha_d] \right\}.$$

Moreover, we denote by I_d the identity operator applicable to functions which depend on the first d variables $y_1, \ldots, y_d$. Then the following result holds:

Lemma 6.2.1 *For functions $u \in L^2_\rho(\Gamma^N; W(D))$ satisfying the assumption of Lemma 5.0.3 with decay coefficients as in (5.0.3), the anisotropic Smolyak formula (4.4.10) satisfies:*

$$\left\| \left(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w, N) \right)(u) \right\|_{L^2_\rho(\Gamma^N; W(D))} \leq \sum_{d=1}^{N} R(w, d) \tag{6.2.4}$$

with

$$R(w, d) := \sum_{\mathbf{i} \in \tilde{X}_{\boldsymbol{\alpha}}(w,d)} C E^{d-1} \left(\prod_{n=1}^{d} i_n \right) e^{-h(\mathbf{i}, d)} \tag{6.2.5}$$

and

$$h(\mathbf{i}, d) = \sum_{n=1}^{d} g(n) 2^{i_n - 1}, \tag{6.2.6}$$

with constants C and E defined in (6.2.2) and (6.2.3).

Proof. We start providing and equivalent representation of the anisotropic Smolyak formula:

$$\begin{aligned}
\mathscr{A}_{\boldsymbol{\alpha}}(w, N) &= \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} \bigotimes_{n=1}^{N} \Delta^{i_n} \\
&= \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N-1)} \bigotimes_{n=1}^{N-1} \Delta^{i_n} \otimes \sum_{j=1}^{\left\lfloor 1 + w \frac{\underline{\alpha}}{\alpha_N} - \sum_{n=1}^{N-1} (i_n - 1) \frac{\alpha_n}{\alpha_N} \right\rfloor} \Delta^{j} \\
&= \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N-1)} \bigotimes_{n=1}^{N-1} \Delta^{i_n} \otimes \mathscr{U}^{\left\lfloor 1 + w \frac{\underline{\alpha}}{\alpha_N} - \sum_{n=1}^{N-1} (i_n - 1) \frac{\alpha_n}{\alpha_N} \right\rfloor}
\end{aligned}$$

where we have denoted with $\lfloor \cdot \rfloor$ the integer part of a real number.

The error estimate is computed recursively using the previous representation.

$$\begin{aligned}
I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w, N) &= I_N - \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N-1)} \bigotimes_{n=1}^{N-1} \Delta^{i_n} \otimes \left(\mathscr{U}^{\left\lfloor 1 + w \frac{\underline{\alpha}}{\alpha_N} - \sum_{n=1}^{N-1} (i_n - 1) \frac{\alpha_n}{\alpha_N} \right\rfloor} - I_1^{(N)} \right) \\
&\quad - \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N-1)} \bigotimes_{n=1}^{N-1} \Delta^{i_n} \otimes I_1^{(N)} \\
&= \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N-1)} \bigotimes_{n=1}^{N-1} \Delta^{i_n} \otimes \left(I_1^{(N)} - \mathscr{U}^{\left\lfloor 1 + w \frac{\underline{\alpha}}{\alpha_N} - \sum_{n=1}^{N-1} (i_n - 1) \frac{\alpha_n}{\alpha_N} \right\rfloor} \right) \\
&\quad + (I_{N-1} - \mathscr{A}_{\boldsymbol{\alpha}}(w, N-1)) \otimes I_1^{(N)} \\
&= \sum_{d=2}^{N} \left[\tilde{R}(w, d) \bigotimes_{n=d+1}^{N} I_1^{(n)} \right] + \left(I_1^{(1)} - \mathscr{A}_{\boldsymbol{\alpha}}(w, 1) \right) \bigotimes_{n=2}^{N} I_1^{(n)}
\end{aligned}$$

where, for a general dimension d, we define

$$\tilde{R}(w, d) = \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,d-1)} \bigotimes_{n=1}^{d-1} \Delta^{i_n} \otimes \left(I_1^{(d)} - \mathscr{U}^{\hat{i}_d - 1} \right)$$

and, for any $(i_1, \ldots, i_{d-1}) \in X_{\boldsymbol{\alpha}}(w, d-1)$, we have set $\hat{i}_d = \left\lfloor 2 + w \frac{\underline{\alpha}}{\alpha_d} - \sum_{n=1}^{d-1} (i_n - 1) \frac{\alpha_n}{\alpha_d} \right\rfloor$. Observe that with this definition, the d-dimensional vector $\mathbf{j} = (i_1, \ldots, i_{d-1}, \hat{i}_d) \in \tilde{X}_{\boldsymbol{\alpha}}(w, d)$.

The term $\tilde{R}(w,d)$ can now be bounded as

$$\begin{aligned}\left\|\tilde{R}(w,d)(u)\right\|_{\infty,d} &\leq \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,d-1)} \prod_{n=1}^{d-1} \left\|(\Delta^{i_n})(u)\right\|_{\infty,d} \left\|\left(I_1^{(d)} - \mathscr{U}^{\hat{i}_d-1}\right)(u)\right\|_{\infty,d} \\ &\leq \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,d-1)} C\, E^{d-1} \left(\prod_{n=1}^{d-1} i_n\right) (\hat{i}_d - 1) e^{-\sum_{n=1}^{d-1} g(n) 2^{i_n-1} - g(d) 2^{\hat{i}_d-1}} \\ &\leq \sum_{\mathbf{i}\in \tilde{X}_{\boldsymbol{\alpha}}(w,d)} C\, E^{d-1} \left(\prod_{n=1}^{d} i_n\right) e^{-h(\mathbf{i},d)} =: R(w,d).\end{aligned}$$

Hence, the interpolation error with the anisotropic Smolyak construction can be bounded by

$$\left\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))(u)\right\|_{\infty,N} \leq \sum_{d=2}^{N} R(w,d) + \left\|(I_1^{(1)} - \mathscr{A}_{\boldsymbol{\alpha}}(w,1))(u)\right\|_{\infty,1}.$$

Observe that the first term in the recursion (6.2.4) can also be bounded by (6.2.5). Indeed, the set $\tilde{X}_{\boldsymbol{\alpha}}(w,1)$ contains only the point $i_1 = \lfloor 2 + \frac{\underline{\alpha} w}{\alpha_1} \rfloor$ and

$$\begin{aligned}\left\|\left(I_1^{(1)} - \mathscr{A}_{\boldsymbol{\alpha}}(w,1)\right)(u)\right\|_{\infty,1} &= \left\|\left(I_1^{(1)} - \mathscr{U}^{\lfloor 1+\frac{\underline{\alpha} w}{\alpha_1}\rfloor}\right)(u)\right\|_{\infty,1} \\ &\leq C \left\lfloor 1 + \frac{\underline{\alpha} w}{\alpha_1} \right\rfloor e^{-g(1)\, 2^{\lfloor 1+\frac{\underline{\alpha} w}{\alpha_1}\rfloor}} \\ &\leq \sum_{i_1 \in \tilde{X}_{\boldsymbol{\alpha}}(w,1)} C\, i_1\, e^{-g(1)\, 2^{i_1-1}} =: R(w,1)\end{aligned}$$

and this concludes the proof. □

Lemma 6.2.2 *For the choice $\alpha_n = g(n)$ of the weights in the anisotropic Smolyak formula (4.4.10), the following bound holds for the term $R(w,d)$, $d = 1,\ldots,N$:*

$$R(w,d) \leq C_1(\mathbf{g},d) \exp\left\{-\frac{\mathscr{G}(d)}{2} 2^{w\frac{\underline{g}}{\mathscr{G}(d)}} + w\frac{d\underline{g}}{\mathscr{G}(d)}\right\}, \tag{6.2.7}$$

where the function $C_1(\mathbf{g},d)$ does not depend on w.

Proof. First we convert the sum appearing in (6.2.5) into an integral. For that, we define the two subsets of $\mathbb{R}^d$:

$$\begin{aligned}\tilde{Y}_{\boldsymbol{\alpha}}(w,d) &:= \left\{\mathbf{y} \in \mathbb{R}_+^d : \ \lfloor y \rfloor + 1 \in \tilde{X}_{\boldsymbol{\alpha}}(w,d)\right\}, \\ \tilde{Y}_{\boldsymbol{\alpha}}^{+}(w,d) &:= \left\{\mathbf{y} \in \mathbb{R}_+^d : \ \left(\sum_{n=1}^{d} y_n \alpha_n - w\underline{\alpha}\right) \in [0, \alpha_d + \sum_{n=1}^{d} \alpha_n]\right\},\end{aligned}$$

and notice that $\tilde{X}_{\boldsymbol{\alpha}}(w,d) \subset \overline{\tilde{Y}_{\boldsymbol{\alpha}}(w,d)} \subset \tilde{Y}^{+}_{\boldsymbol{\alpha}}(w,d)$. Then the term $R(w,d)$ can be bounded by

$$\begin{aligned} R(w,d) &= \int_{\tilde{Y}_{\boldsymbol{\alpha}}(w,d)} CE^{d-1}\left(\prod_{n=1}^{d}(\lfloor y_n \rfloor + 1)\right) e^{-h(\lfloor \mathbf{y} \rfloor + 1, d)}\, d\mathbf{y} \\ &\leq \int_{\tilde{Y}^{+}_{\boldsymbol{\alpha}}(w,d)} CE^{d-1}\left(\prod_{n=1}^{d}(y_n + 1)\right) e^{-h(\mathbf{y},d)}\, d\mathbf{y}. \end{aligned} \tag{6.2.8}$$

Next, we define $y^* = w\frac{g}{\mathscr{G}(d)}$ and we expand the function $h(\mathbf{y},d)$ up to second order, around the point $\mathbf{y}^* = (y^*, \ldots, y^*)$:

$$h(\mathbf{y},d) = h(\mathbf{y}^*,d) + \underbrace{\nabla h(\mathbf{y}^*,d)\cdot\delta\mathbf{y}}_{\text{(I)}} + \underbrace{\delta\mathbf{y}^T\frac{1}{2}\nabla^2 h(\mathbf{y}^* + s\delta\mathbf{y},d)\delta\mathbf{y}}_{\text{(II)}}, \quad \text{with } \delta\mathbf{y} = \mathbf{y} - \mathbf{y}^* \text{ and } s \in [0,1].$$

The linear term is positive on the set $\tilde{Y}^{+}_{\boldsymbol{\alpha}}(w,d)$; indeed

$$\begin{aligned} \text{(I)} &= \log(2)2^{y^*-1}\sum_{n=1}^{d} g(n)\,(y_n - y^*) \\ &= \log(2)2^{y^*-1}\sum_{n=1}^{d}(y_n\alpha_n - w\underline{\alpha}) \geq 0 \qquad \forall \mathbf{y} \in \tilde{Y}^{+}_{\boldsymbol{\alpha}}(w,d) \end{aligned}$$

Similarly, the second order remainder can be bounded as

$$\begin{aligned} \text{(II)} &= \frac{\log(2)^2}{2}\sum_{n=1}^{d} g(n)2^{[y^* + s(y_n - y^*) - 1]}(y_n - y^*)^2 \\ &\geq \frac{\log(2)^2}{4}\sum_{n=1}^{d} g(n)(y_n - y^*)^2 \\ &= \sum_{n=1}^{d}\frac{(y_n - y^*)^2}{2\sigma_n^2}, \qquad \text{with } \sigma_n^2 = \frac{2}{\log(2)^2 g(n)} \end{aligned}$$

Finally, the bound (6.2.8) becomes

$$R(w,d) \leq CE^{d-1}e^{-h(\mathbf{y}^*,d)}\underbrace{\int_{\tilde{Y}^{+}_{\boldsymbol{\alpha}}(w,d)}\prod_{n=1}^{d}(y_n+1)e^{-\sum_{n=1}^{d}\frac{(y_n - y^*)^2}{2\sigma_n^2}}\, d\mathbf{y}}_{I(d)}.$$

We now turn to estimating the term $I(d)$. For that, let us introduce the function $p_n(y) = e^{-\frac{(y-y^*)^2}{2\sigma_n^2}}/\sqrt{2\pi\sigma_n^2}$ which corresponds to the probability density function of a Normal

random variable with mean y^* and variance σ_n^2. Then, we have

$$
\begin{aligned}
I(d) &= \int_{\tilde{Y}_{\boldsymbol{\alpha}}^+(w,d)} \prod_{n=1}^{d} \sqrt{2\pi\sigma_n^2}(y_n+1)p_n(y_n)\,d\mathbf{y} \qquad \{\text{setting } z_n = y_n+1\} \\
&\le \prod_{n=1}^{d} \left(\sqrt{2\pi\sigma_n^2} \int_0^\infty z_n p_n(z_n-1)\,dz_n \right) \qquad \{\text{by Cauchy-Schwartz ineq.}\} \\
&\le \prod_{n=1}^{d} \left(\sqrt{2\pi\sigma_n^2} \left(\int_0^\infty z_n^2 p_n(z_n-1)\,dz_n \right)^{\frac{1}{2}} \right) \\
&\le \prod_{n=1}^{d} \left(\sqrt{2\pi\sigma_n^2}\sqrt{\sigma_n^2 + (1+y^*)^2} \right) \\
&\le \prod_{n=1}^{d} \left(\sqrt{2\pi\sigma_n^2}\,(1+y^*+\sigma_n) \right) \le \prod_{n=1}^{d} \left(\sqrt{2\pi\sigma_n^2} e^{y^*+\sigma_n} \right) \\
&= C_2(\boldsymbol{\sigma}, d) e^{w\frac{d\underline{g}}{\mathscr{G}(d)}}, \qquad \text{with } C_2(\boldsymbol{\sigma}, d) = \prod_{n=1}^{d} \left(\sqrt{2\pi\sigma_n^2} e^{\sigma_n} \right).
\end{aligned}
$$

From this the final result is easily obtained and inequality (6.2.7) holds with constant

$$
C_1(\mathbf{g}, d) = CE^{d-1} \left(\frac{2\sqrt{\pi}}{\log(2)} \right)^d \left(\prod_{n=1}^{d} \frac{1}{\sqrt{g(n)}} \right) \exp\left\{ \frac{\sqrt{2}}{\log(2)} \sum_{n=1}^{d} \frac{1}{\sqrt{g(n)}} \right\}. \tag{6.2.9}
$$

□

Theorem 6.2.3 *For functions $u \in L^2_\rho(\Gamma^N; W(D))$ satisfying the assumption of Lemma 5.0.3 with decay coefficients as in (5.0.3), the anisotropic Smolyak formula (4.4.10) with the choice $\alpha_n = g(n)$ of the weights satisfies:*

$$
\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \le \hat{C}(\mathbf{g}, N) e^{w-\lambda(w,N)} \tag{6.2.10}
$$

where

$$
\lambda(w,N) := \begin{cases} w\dfrac{\underline{g}\log(2)e}{2}, & \text{if } 0 \le w \le \frac{\mathscr{G}(N)}{\underline{g}\log(2)}, \\[2ex] \dfrac{\mathscr{G}(N)}{2} 2^{w\frac{\underline{g}}{\mathscr{G}(N)}}, & \text{otherwise} \end{cases}, \tag{6.2.11}
$$

and the function $\hat{C}(\mathbf{g}, N)$ does not depend on w.

Proof. From Lemmas 6.2.1 and 6.2.2 we obtain the following bound for the interpolation error

$$
\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \le \sum_{d=1}^{N} C_1(\mathbf{g}, d) \exp\left\{ w - \widehat{\lambda}(w,d) \right\}
$$

with $\widehat{\lambda}(w,d) := \frac{\mathscr{G}(d)}{2} 2^{w\frac{\underline{g}}{\mathscr{G}(d)}}$. We now turn our attention to finding a minimum for $\widehat{\lambda}(w,d)$ for $1 \leq d \leq N$. Let us define, for $s \in \left[1, \frac{\mathscr{G}(N)}{\underline{g}}\right]$, the function

$$p(s) := s\, 2^{\frac{w}{s}}.$$

so that $\widehat{\lambda}(w,d) = \frac{\underline{g}}{2} p\left(\frac{\mathscr{G}(d)}{\underline{g}}\right)$ and

$$\min_{1\leq d\leq N} \widehat{\lambda}(w,d) \geq \min_{s\in[1,\mathscr{G}(N)/\underline{g}]} \frac{\underline{g}}{2} p(s).$$

We have

$$\frac{dp}{ds} = 2^{\frac{w}{s}} \left(1 - \left(\frac{w}{s}\right)\log(2)\right) = 0$$

yielding $s = w\log(2)$. For w sufficiently large, the minimum of $p(s)$ falls outside the interval $\left[1, \frac{\mathscr{G}(N)}{\underline{g}}\right]$ and the function $p(s)$ is decreasing on this interval. Therefore, there are two cases to consider. The first being the situation when $w > \frac{\mathscr{G}(N)}{\underline{g}\log(2)}$ and the second when $0 \leq w \leq \frac{\mathscr{G}(N)}{\underline{g}\log(2)}$. In either case

$$\min_{1\leq d\leq N} \widehat{\lambda}(w,d) \geq \lambda(w,N)$$

and hence,

$$\max_{1\leq d\leq N} e^{-\widehat{\lambda}(w,d)} \leq e^{-\lambda(w,N)}.$$

From this the result (6.2.10) follows, by taking

$$\hat{C}(\mathbf{g},N) = N \max_{d=1,\ldots,N} C_1(\mathbf{g},d). \tag{6.2.12}$$

□

Remark 6.2.4 *From the expressions* (6.2.12) *and* (6.2.9) *we observe that the constant* $\hat{C}(\mathbf{g},N)$ *appearing in the convergence estimates of Theorem 6.2.3 goes to infinity when* $\underline{g}$ *tends to zero. We should note that in such case we loose convergence anyhow.*

On the other hand, for given $\underline{g}$ *and* N*, the more the sequence* $\mathbf{g}$ *is anisotropic (i.e. the larger the ratio* $g_{max}/\underline{g}$ *gets), the smaller becomes the constant* $\hat{C}(\mathbf{g},N)$*. In any case, such constant can be bounded by*

$$\hat{C}(\mathbf{g},N) \leq \frac{C}{E} \exp\left\{\frac{N}{\sqrt{\underline{g}}}\left(\frac{2E\sqrt{\pi}+\sqrt{2}}{\log(2)}\right)\right\},$$

independently of the anisotropy.

Now we relate the number of collocation points $\eta = \eta(w, N) = \#\mathscr{H}_{\boldsymbol{\alpha}}(w, N)$ to the level w of the anisotropic Smolyak algorithm. We state the result in the following lemma:

Lemma 6.2.5 *Using the anisotropic Smolyak interpolant described by* (4.4.7) *where the abscissas are the Clenshaw-Curtis knots, described in Section 4.4.4, the total number of points required at level w satisfies the following bounds:*

$$2^{w-1} \leq \eta \leq 2^w e^{w \sum_{n=1}^{N} \underline{\alpha}/\alpha_n}. \tag{6.2.13}$$

Moreover, as a direct consequence of (6.2.13) *we get that:*

$$\frac{\log(\eta)}{\sum_{n=1}^{N} \frac{\underline{\alpha}}{\alpha_n} + \log(2)} \leq w \leq \log_2(2\eta). \tag{6.2.14}$$

Proof. By using formula (4.4.7) and exploiting the nested structure of the Clenshaw-Curtis abscissas the number of points $\eta = \eta(w, N) = \#\mathscr{H}_{\boldsymbol{\alpha}}(w, N)$ can be counted in the following way:

$$\eta = \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} \prod_{n=1}^{N} r(i_n), \text{ where } r(i) := \begin{cases} 1 & \text{if } i = 1 \\ 2 & \text{if } i = 2 \\ 2^{i-2} & \text{if } i > 2 \end{cases}. \tag{6.2.15}$$

Begin by noticing that for all $n = 1, 2, \ldots, N$ the following bound holds:

$$2^{i_n - 2} \leq r(i_n) \leq 2^{i_n - 1}. \tag{6.2.16}$$

Next, let $m \in [1, N]$ be the index corresponding to the minimum α, i.e. $\alpha_m = \min_{1 \leq n \leq N} \{\alpha_n\} = \underline{\alpha}$. A lower bound on the number η of points can be obtained considering only the contribution from the tensor grid of indices $i_n = 1$, for $n \neq m$ and $i_m = w + 1$. On the other hand, it is easy to see that $|\mathbf{i} - \mathbf{1}| = \sum_{n=1}^{N} (i_n - 1) \leq w$ so the following bounds hold:

$$2^{w-1} \leq \eta = \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} \prod_{n=1}^{N} r(i_n) \leq \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} 2^{|\mathbf{i}-\mathbf{1}|} \leq 2^w \# X_{\boldsymbol{\alpha}}(w, N).$$

We need now a bound for the cardinality of the set $X_{\boldsymbol{\alpha}}(w, N)$. We prove by induction that

$$\# X_{\boldsymbol{\alpha}}(w, N) \leq \prod_{n=1}^{N} \left(w \frac{\underline{\alpha}}{\alpha_n} + 1 \right). \tag{6.2.17}$$

Indeed, the result is obviously true for $N = 1$, and assuming that it holds for $N-1$, we have

$$\begin{aligned}
\#X_{\boldsymbol{\alpha}}(w,N) &= \sum_{j_N=1}^{\lfloor w\underline{\alpha}/\alpha_N\rfloor+1} \#X_{\boldsymbol{\alpha}}\left(w - \frac{(j_N-1)\alpha_N}{\underline{\alpha}}, N-1\right) \\
&= \sum_{j_N=1}^{\lfloor w\underline{\alpha}/\alpha_N\rfloor+1} \prod_{n=1}^{N-1}\left(w\frac{\underline{\alpha}}{\alpha_n} - \frac{(j_N-1)\alpha_N}{\alpha_n} + 1\right) \\
&\leq \sum_{j_N=1}^{\lfloor w\underline{\alpha}/\alpha_N\rfloor+1} \prod_{n=1}^{N-1}\left(w\frac{\underline{\alpha}}{\alpha_n} + 1\right) \\
&\leq \prod_{n=1}^{N}\left(w\frac{\underline{\alpha}}{\alpha_n} + 1\right)
\end{aligned}$$

and this finishes the induction proof. Moreover we have

$$\#X_{\boldsymbol{\alpha}}(w,N) \leq \exp\left\{w\sum_{n=1}^{N}\frac{\underline{\alpha}}{\alpha_n}\right\} \tag{6.2.18}$$

and the inequalities (6.2.13) and (6.2.14) follow. □

Remark 6.2.6 *The bound for the cardinality of the set $\#X_{\boldsymbol{\alpha}}(w,N)$ given in (6.2.17) is not sharp when $w \to \infty$ and actually one has instead the asymptotic behaviour*

$$\#X_{\boldsymbol{\alpha}}(w,N) \lesssim \frac{w^N}{N!}\prod_{n=1}^{N}\frac{\underline{\alpha}}{\alpha_n},$$

which is consistent with the isotropic result given in Nobile et al. (2006), described by Lemma 6.1.3. Moreover, if we use Stirling's approximation for the factorial term, the previous bound becomes

$$\#X_{\boldsymbol{\alpha}}(w,N) \lesssim \frac{1}{\sqrt{N}}\prod_{n=1}^{N}\frac{ew}{N}\frac{\underline{\alpha}}{\alpha_n} \leq \frac{1}{\sqrt{N}}\exp\left\{w\left(\frac{e}{N}\sum_{n=1}^{N}\frac{\underline{\alpha}}{\alpha_n}\right)\right\},$$

which greatly improves (6.2.18).

The next Theorem provides an error bound in terms of the total number η of collocation points. The proof follows directly from the results in Theorem 6.2.3 and Lemma 6.2.5 and is therefore omitted.

Theorem 6.2.7 *For functions $u \in L^2_\rho(\Gamma^N; W(D))$ satisfying the assumption of Lemma 5.0.3 with decay coefficients as in (5.0.3), the anisotropic Smolyak formula (4.4.10) with the choice $\alpha_n = g(n)$ of the weights satisfies:*

- Algebraic convergence $\left(0 \le w \le \frac{\mathscr{G}(N)}{\underline{g}\log(2)}\right)$. *Under the assumption that* $\underline{g} \ge 1/(e\log(2))$,

$$\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \le \hat{C}(\mathbf{g},N)\eta^{-\mu_1},$$
$$\textit{with} \qquad \mu_1 = \frac{\underline{g}\log(2)e - 1}{\log(2) + \sum_{n=1}^N \underline{g}/g(n)}. \tag{6.2.19}$$

- Sub-exponential convergence $\left(w > \frac{\mathscr{G}(N)}{\underline{g}\log(2)}\right)$

$$\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \le \hat{C}(\mathbf{g},N)(2\eta)^{1/\log(2)}e^{-\frac{\mathscr{G}(N)}{2}\eta^{\mu_2}},$$
$$\textit{with} \qquad \mu_2 = \frac{\underline{g}\log(2)}{\mathscr{G}(N)\left(\log(2) + \sum_{n=1}^N \underline{g}/g(n)\right)}. \tag{6.2.20}$$

and constant $\hat{C}(\mathbf{g},N)$ *defined in* (6.2.12) *and independent of* η.

Remark 6.2.8 *The estimates given in* (6.2.20) *may be improved when* $w \to \infty$. *Such asymptotic estimate is obtained using the better counting result described in Remark 6.2.6.*

Remark 6.2.9 *We observe that sub-exponential rate of convergence is always faster than the algebraic one when* $w > \mathscr{G}(N)/(\underline{g}\log(2))$. *Yet, this estimate is of little practical relevance since in practical computations, such a high level* w *is seldom reached.*

Remark 6.2.10 *The condition* $\underline{g} > 1/(log(2)e)$ *in the algebraic regime can be improved following an* L^2 *analysis. Yet it is largely satisfied in all our numerical tests. In the next Section we present results using Gaussian abscissas where this condition is no longer needed.*

Remark 6.2.11 *Suppose now that the stochastic input data are truncated expansions of random fields and that we are able to estimate the values* $\{g(n)\}_{n=1}^\infty$. *Whenever the sum* $\sum_{n=1}^\infty \underline{g}/g(n)$ *is finite then the algebraic exponent in* (6.2.19) *does not deteriorate as the truncation dimension* N *increases. This condition is satisfied for instance by the problem presented in the numerical Section. This is a clear advantage with respect to the isotropic Smolyak method.*

Remark 6.2.12 (Optimal choice of $\boldsymbol{\alpha}$) *Looking at the exponential term* $e^{-h(\mathbf{i},d)}$ *in* (6.2.5), *which is the term determining the rate of convergence, we may try to choose the weight* $\boldsymbol{\alpha}$ *for* $X_{\boldsymbol{\alpha}}(w,N)$ *as the solution to the optimization problem*

$$\max_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^d_+\\ |\boldsymbol{\alpha}|=1}} \min_{\mathbf{i}\in\tilde{X}_{\boldsymbol{\alpha}}(w,d)} h(\mathbf{i},d).$$

This problem has the solution $\boldsymbol{\alpha} = \mathbf{g}$ *and hence, our choice of weights* (5.0.4) *is optimal.*

6.2.2 Gaussian anisotropic estimates

By using a similar approach to our previous work (Nobile et al., 2006, Section 4.1.2) and Section 6.1.2 we can develop error estimates for interpolating functions $u \in C^0(\Gamma^N; W(D))$ that admit an analytic extension as described by Assumption 1.3.1 using the anisotropic Smolyak formulations based on Gaussian abscissas described in Section 4.4.5. We remind the reader that in the global estimate (6.0.1) we need to bound the interpolation error (III) in the norm $L^2_\rho(\Gamma^N; W(D))$. Yet, the Gaussian points defined in Section 4.4.5 are constructed for the more appropriate density $\hat{\rho} = \prod_{n=1}^N \hat{\rho}_n$ and we have the following useful bound:

$$\|v\|_{L^2_{\hat{\rho}}(\Gamma^N;W(D))} \leq \left\| \frac{\rho}{\hat{\rho}} \right\|_{L^\infty(\Gamma^N)} \cdot \|v\|_{L^2_{\hat{\rho}}(\Gamma^N;W(D))} \qquad \text{for all } v \in C^0(\Gamma^N; W(D)).$$

In what follows we will use the shorthand notation $\|\cdot\|_{\hat{\rho},N}$ for $\|\cdot\|_{L^2_{\hat{\rho}}(\Gamma^N;W(D))}$. Following Nobile et al. (2006) and Section 6.1.2, the one-dimensional interpolation error for Gaussian abscissas satisfies

$$\left\|(I_1^{(n)} - \mathscr{U}^{i_n})(u)\right\|_{\hat{\rho},1} \leq \tilde{C}\, e^{-g(n)2^{i_n}} \qquad \text{and} \qquad \left\|(\Delta^{i_n})(u)\right\|_{\hat{\rho},1} \leq \tilde{E}\, e^{-\frac{g(n)}{2}2^{i_n}},$$

for all $i \in \mathbb{N}_+$ with positive constants $\tilde{C}$ and $\tilde{E}$ depending on u but not on i or n.

The following Theorem states the rate of convergence for the anisotropic Smolyak formula based on Gaussian abscissas. Since the proof is mostly similar to the one presented in the previous Section for Clenshaw-Curtis points, it will be just sketched and only the main differences will be highlighted.

Theorem 6.2.13 *For functions $u \in L^2_\rho(\Gamma^N; W(D))$ satisfying the assumption of Lemma 5.0.3 with decay coefficients as in (5.0.3), the anisotropic Smolyak formula (4.4.10), based on Gaussian abscissas and with the choice $\alpha_n = g(n)$ of the weights satisfies:*

- Algebraic convergence $\left(0 \leq w \leq \frac{\mathscr{G}(N)}{\underline{g}\log(2)}\right)$:

$$\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))\,(u)\|_{L^2_{\hat{\rho}}(\Gamma^N;W(D))} \leq \hat{C}^G(\mathbf{g},N)\eta^{-\mu_1},$$
$$\textit{with} \qquad \mu_1 = \frac{\underline{g}\log(2)e}{2\log(2) + \sum_{n=1}^N \underline{g}/g(n)}. \tag{6.2.21}$$

• Sub-exponential convergence $\left(w > \frac{\mathscr{G}(N)}{\underline{g}\log(2)}\right)$:

$$\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq \hat{C}^G(\mathbf{g},N)e^{-\frac{\mathscr{G}(N)}{2}\eta^{\mu_2}},$$
$$\textit{with} \qquad \mu_2 = \frac{\underline{g}\log(2)}{\mathscr{G}(N)\left(2\log(2) + \sum_{n=1}^N \underline{g}/g(n)\right)}. \tag{6.2.22}$$

and constant $\hat{C}^G(\mathbf{g},N)$ *independent of* η.

Proof. The recursion formula (6.2.4) still holds with the term $R(w,d)$ defined now as

$$R(w,d) = \left\|\frac{\rho}{\hat{\rho}}\right\| \sum_{\mathbf{i}\in\tilde{X}_{\boldsymbol{\alpha}}(w,d)} \tilde{C}\tilde{E}^{d-1}e^{-h(\mathbf{i},d)}.$$

Notice that the negative effect of the Lebesgue constant for the Clenshaw-Curtis points, which was responsible for the term $\prod_{n=1}^d i_n$ in the definition of $R(w,d)$ (see (6.2.5)) is not present any more. Then, following the guidelines of the proof of Lemma 6.2.2, this term can be bounded as

$$R(w,d) \leq C_1^G(\mathbf{g},d)\exp\left\{-\frac{\mathscr{G}(d)}{2}2^{w\frac{\underline{g}}{\mathscr{G}(d)}}\right\},$$

with constant

$$C_1^G(\mathbf{g},d) = \left\|\frac{\rho}{\hat{\rho}}\right\| \tilde{C}\tilde{E}^{d-1}\left(\frac{2\sqrt{\pi}}{\log(2)}\right)^d \left(\prod_{n=1}^d \frac{1}{\sqrt{g(n)}}\right).$$

This leads to the estimate in terms of w

$$\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))\,(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq \hat{C}^G(\mathbf{g},N)e^{-\lambda(w,N)} \tag{6.2.23}$$

with the same $\lambda(w,N)$ as in (6.2.11) and $\hat{C}^G(\mathbf{g},N) = N\max_{d=1,\dots,N} C_1^G(\mathbf{g},d)$. Again, notice the great improvement with respect to the bound (6.2.10) holding for Clenshaw-Curtis points. Finally, we observe that for a given level w, the number of Gauss points is larger than the number of Clenshaw-Curtis points (due to the non-nested structure) and, following Lemma 6.1.7, we know that (Nobile et al., 2006):

$$\eta = \sum_{\mathbf{i}\in Y_{\boldsymbol{\alpha}}(w,N)} \prod_{n=1}^N \tilde{r}(i_n), \qquad \text{with} \quad \tilde{r}(i) := \begin{cases} 1 & \text{for } i = 1 \\ 2^{i-1}+1 & \text{for } i > 1 \end{cases} \tag{6.2.24}$$

$$\prod_{n=1}^N \tilde{r}(i_n) \leq \prod_{n=1}^N (2^{i_n-1}+1) \leq 2^{2|i-1|} \leq 2^{2w}$$

and the number of points can be bounded as

$$2^{w-1} \leq \eta \leq 2^{2w} \# Y_{\boldsymbol{\alpha}}(w, N) \leq 2^{2w} \# X_{\boldsymbol{\alpha}}(w, N),$$

which, substituted in (6.2.23) gives the desired result. □

To demonstrate the effectiveness of the proposed methods considered throughout this chapter, in the next chapter we present some numerical results validating our theoretical results. We also include a comparison of our proposed methods with other ensemble-based approaches, including the well-known Monte Carlo.

CHAPTER 7

NUMERICAL EXAMPLES

This Chapter illustrates the convergence of the both the isotropic and anisotropic sparse collocation methods for the stochastic linear and nonlinear elliptic problems in two spatial dimensions, as described in Chapters 2 and Chapters 3 repectively. Most of this chapter will be devoted to a linear problem with applications to a nonlinear problem left until the end. The computational results are in accordance with the convergence rates predicted by the theory. Actually, we observe a faster convergence than stated in Theorems 6.2.7 and 6.2.13, which hints that the current estimates may be improved.

We will also use this chapter to compare the convergence of the anisotropic Smolyak method, described and analyzed in Sections 4.4.3 and 6.2 respectively and the work Nobile et al. (2007), with other ensemble-based methods such as: the isotropic Smolyak method, described and analyzed in Sections 4.4.2 and 6.1 respectively and the work Nobile et al. (2006), the anisotropic adaptive full tensor product method described in the work (Babuška et al., 2005d, Section 9) and finally, the well-known Monte Carlo method.

The linear and nonlinear problems we consider are as follows:

$$\begin{cases} -\nabla \cdot (a(\omega,\cdot)\nabla u(\omega,\cdot)) &= f(\omega,\cdot) \quad \text{in } \Omega \times D, \\ u(\omega,\cdot) &= 0 \quad \text{on } \Omega \times \partial D, \end{cases} \tag{7.0.1}$$

and

$$\begin{cases} -\nabla \cdot (a(\omega,\cdot)\nabla u(\omega,\cdot)) + u^3 &= f(\omega,\cdot) \quad \text{in } \Omega \times D, \\ u(\omega,\cdot) &= 0 \quad \text{on } \Omega \times \partial D, \end{cases} \tag{7.0.2}$$

with $D = [0,d]^2$ and $d = 1$. For these numerical examples we take a deterministic load $f(\omega,x,z) = \cos(x)\sin(z)$ and construct the random diffusion coefficient $a_N(\omega,x)$ with one-dimensional spatial dependence as

$$\log(a_N(\omega,\tilde{x}) - 0.5) = 1 + \mathrm{Y}_1(\omega)\left(\frac{\sqrt{\pi}L}{2}\right)^{1/2} + \sum_{n=2}^{N} \zeta_n\, \varphi_n(x)\, \mathrm{Y}_n(\omega). \tag{7.0.3}$$

where

$$\zeta_n := \left(\sqrt{\pi}L\right)^{1/2} \exp\left(\frac{-\left(\lfloor \frac{n}{2} \rfloor \pi L\right)^2}{8}\right), \quad \text{if } n > 1 \tag{7.0.4}$$

and

$$\varphi_n(x) := \begin{cases} \sin\left(\frac{\lfloor \frac{n}{2} \rfloor \pi x}{L_p}\right), & \text{if } n \text{ even}, \\ \cos\left(\frac{\lfloor \frac{n}{2} \rfloor \pi x}{L_p}\right), & \text{if } n \text{ odd}. \end{cases} \tag{7.0.5}$$

For $x \in [0, d]$ let L_c be a desired physical correlation length for the coefficient a, meaning that the random variables $a(x)$ and $a(y)$ become essentially uncorrelated for $|x - y| >> L_c$. Then, the parameter L_p in (7.0.5) and (7.0.4) is $L_p = \max\{d, 2L_c\}$ and the parameter L in (7.0.3) and (7.0.4) is

$$L = \frac{L_c}{L_p}.$$

In this example, the random variables $\{Y_n(\omega)\}_{n=1}^{\infty}$ are independent, have zero mean and unit variance, i.e. $E[Y_n] = 0$ and $E[Y_n Y_m] = \delta_{nm}$ for $n, m \in \mathbb{N}_+$, and are uniformly distributed in the interval $[-\sqrt{3}, \sqrt{3}]$. Expression (7.0.3) is truncation of a one-dimensional random field with stationary covariance

$$\begin{aligned} \text{cov}[\log(a_N - 0.5)](x_1, x_2) &= E\big[(\log(a)(x_1) - E[\log(a)](x_1))\,((\log(a)(x_2) - E[\log(a)](x_2))\big] \\ &= \exp\left(\frac{-(x_1 - x_2)^2}{L_c^2}\right). \end{aligned}$$

To illustrate the convergence, given in Chapter 6, of isotropic and anisotropic Smolyak algorithms, described in Sections 4.4.2 and 4.4.3 respectively, for solving problem (7.0.1) we need to first, describe the decay coefficient given by (5.0.3) and then second, either *a priori* or *a posteriori* select the N-dimensional weights described in Chapter 5.

To formulate the constant ϱ_n for each $n = 1, 2, \ldots, N$, defined by (5.0.3), for the problem (7.0.1), we investigate a lower bound for ϱ_n. That is (see (6.1.3)) for $n = 1$

$$\begin{aligned} \varrho_1 \geq 1 + \frac{1}{4\sqrt{3}}\left(\frac{2}{\sqrt{\pi}L}\right)^{1/2} &= 1 + \frac{1}{2}\left(\frac{1}{6\sqrt{\pi}L}\right)^{1/2} \\ &\geq 1 + \exp\left(-\frac{1}{2}\log\left(24\sqrt{\pi}L\right)\right) \end{aligned} \tag{7.0.6}$$

and similarly, for $n \geq 2$

$$\begin{aligned} \varrho_n &\geq 1 + \frac{e^{\frac{\lfloor \frac{n}{2} \rfloor^2 \pi^2 L^2}{8}}}{\sqrt{24\sqrt{\pi}L}} \\ &= 1 + \exp\left(\frac{\lfloor \frac{n}{2} \rfloor^2 \pi^2 L^2}{8} - \frac{1}{2}\log\left(24\sqrt{\pi}L\right)\right). \end{aligned} \tag{7.0.7}$$

The rate of convergence of the isotropic Smolyak method is dictated by the decay coefficient ϱ defined by (5.0.3), which from (7.0.6)-(7.0.7) can be defined by

$$\varrho = \min_n \varrho_n \geq 1 + \sqrt{\frac{1}{24\sqrt{\pi}L}}. \tag{7.0.8}$$

We consider now an a priori selection of the weights $g(n)$ based on formula (6.1.3) for solving problem (7.0.1) using the anisotropic Smolyak method. For that, from (2.0.7) we can estimate the parameters τ_n by

$$\tau_n = \frac{1}{4\zeta_n \|\varphi\|_\infty} = \begin{cases} \sqrt{\frac{1}{8\sqrt{\pi}L}}, & \text{for } n = 1 \\ \sqrt{\frac{1}{16\sqrt{\pi}L}} \exp\left(\frac{\lfloor \frac{n}{2} \rfloor^2 \pi^2 L^2}{8}\right), & \text{for } n > 1. \end{cases}$$

Then, using (6.1.3) the weight vector $\mathbf{g}$ becomes

$$g(n) = \begin{cases} \log\left(1 + \sqrt{\frac{1}{24\sqrt{\pi}L}}\right), & \text{for } n = 1 \\ \log\left(1 + \sqrt{\frac{1}{48\sqrt{\pi}L}} \exp\left(\frac{\lfloor \frac{n}{2} \rfloor^2 \pi^2 L^2}{8}\right)\right), & \text{for } n > 1 \end{cases} \tag{7.0.9}$$

and then we have $\varrho_n \geq e^{g(n)}$ for all $n = 1, 2, \ldots, N$.

In Table 7.1 we show the function values $g(n)$ for $n = 1, 2, \ldots, N = 11$ using (7.0.9). With this *a priori* information we can construct the simplices $\mathbf{i} \in X_{\boldsymbol{\alpha}}(w, N)$. These become the indices $\mathbf{i} \in \mathbb{N}_+^N$ used for solving problem (7.0.1) with correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$. This table also yields insight into the anisotropic behavior of each problem. In the case of small correlation lengths, i.e. $L_c = 1/64$, we observe an almost equal weighing of all stochastic directions, except for the first one. The opposite behavior can be seen as we increase the correlation length. For example, when $L_c = 1/2$ the ratio between $g(11)$ and $\underline{g} = g(2) = g(3)$ is approximately $30 : 1$.

Since the random variables Y_n are uniformly distributed, in this case the Gaussian abscissas correspond to the root of the Legendre polynomials. Recall from Section 4.4.4 that the Clenshaw-Curtis abscissas are nested and therefore, in practice, we exploit this fact and construct the isotropic or anisotropic Smolyak interpolant using formula (4.4.7) or (4.4.10) respectively. Hence, the number of points $\eta = \eta(w, N) = \#\mathscr{H}_{\boldsymbol{\alpha}}(w, N)$ can be counted as in formula (6.1.16) or (6.2.15) respectively. On the other hand, the Gaussian abscissas, described in Section 4.4.5, are not nested and to reduce the number of points

Table 7.1: The $N = 11$ values of the function $g(n)$ constructed from *a priori* information given by (7.0.9) for correlation lengths $L_c = 1/2, 1/4, 1/16, 1/64$.

	$g(1)$	$g(2) = g(3)$	$g(4) = g(5)$	$g(6) = g(7)$	$g(8) = g(9)$	$g(10) = g(11)$
$L_c = 1/2$	0.20	0.19	0.42	1.24	3.1	5.8
$L_c = 1/4$	0.27	0.21	0.26	0.36	0.56	0.91
$L_c = 1/16$	0.48	0.36	0.37	0.37	0.38	0.40
$L_c = 1/64$	0.79	0.62	0.62	0.62	0.62	0.62

necessary to build the isotropic or anisotropic Smolyak interpolant one utilizes the variant of (4.4.7) or (4.4.10), given by (4.4.8) or (4.4.11) respectively. Consequently, we can count the number of points η used by the Smolyak interpolant as in (6.1.28) or (6.2.24) respectively.

The finite element space for the spatial discretization is the span of continuous functions that are piecewise polynomials with degree two over a uniform triangulation of D with 4225 unknowns.

Observe, in general, that the collocation method only requires the solution of uncoupled deterministic problems over the set of collocation points, even in the presence of a diffusivity coefficient which depends nonlinearly on the random variables as in (7.0.3). This is a significant advantage that the collocation method offers compared to the classical Stochastic-Galerkin finite element method as considered, for instance, in Babuška et al. (2004b); Frauenfelder et al. (2005b); Matthies and Keese (2005); Xiu and Karniadakis (2002a). To study the convergence of the isotropic and anisotropic Smolyak algorithms we consider a problem with a fixed dimension N and investigate the behavior when the level w of the interpolation in the Smolyak algorithm is increased linearly.

The computational results for the $L^2(D)$ approximation error to the expected value, $E[u]$, using the isotropic Smolyak interpolant, are shown in Figure 7.1. The isotropic Smolyak algorithm for solving problem (7.0.1) was analyzed in Sections 6.1.1 and 6.1.2 and in the work Nobile et al. (2006). To estimate the computational error in the w-th level we approximate $\|E[\epsilon]\| \approx \|E[\mathscr{A}(w, N)\pi_h u_N - \mathscr{A}(w+1, N)\pi_h u_N]\|$ using either Gaussian or Clenshaw-Curtis abscissas. The results reveal, as expected, that for a small non-degenerate correlation length, i.e. $L_c = 1/64$, the error decreases (sub)-exponentially, as the level w increases. We also observe that the convergence rate is dimension dependent and slightly deteriorates as N increases.

To investigate the performance of the isotropic algorithm by varying the correlation

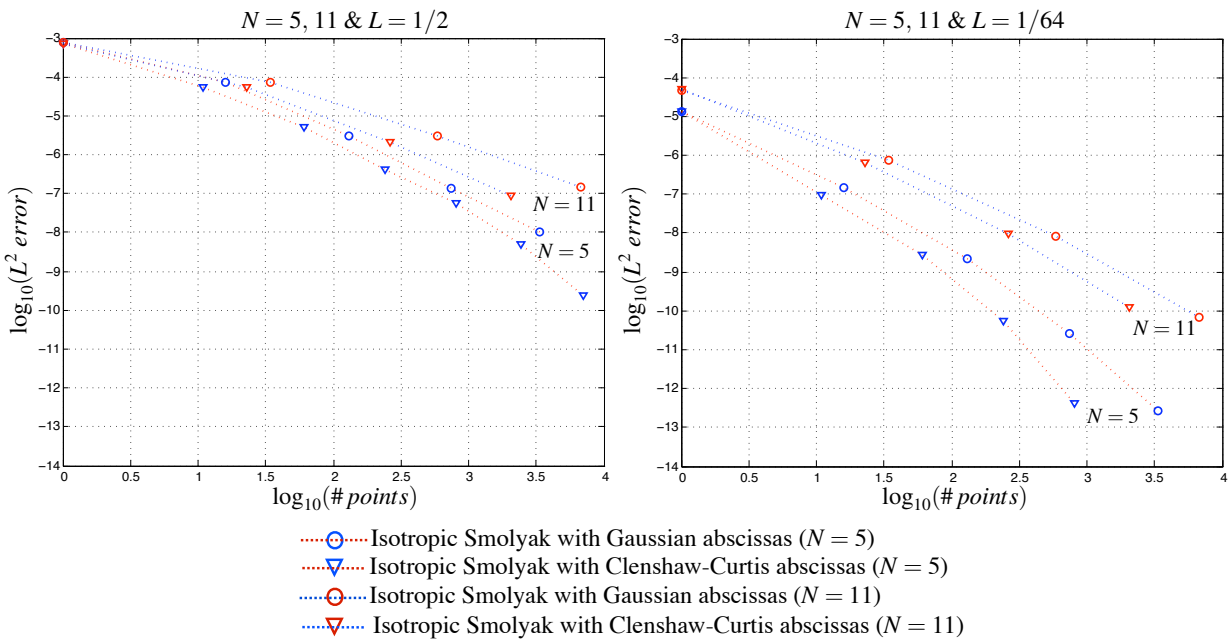

Figure 7.1: The rate of convergence of the isotropic Smolyak algorithm for solving problem (7.0.1) with given correlation lengths $L_c = 1/64$ using both the Gaussian and Clenshaw-Curtis abscissas. For a finite dimensional probability space Γ^N with $N = 5$ and $N = 11$ we plot $\log(\epsilon)$ versus the number of collocation points. The $L^2(D)$ approximation error in the expected value for the anisotropic sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[\mathscr{A}(w, N)\pi_h u_N - \mathscr{A}(w+1, N)\pi_h u_N]\|_{L^2(D)}$ where $w = 0, 1, \ldots, \overline{w}$.

length L_c we also include the cases where $L_c = 1/16$, $L_c = 1/4$ and $L_c = 1/2$ for both $N = 5$ and $N = 11$, seen in Figure 7.2. We notice that the larger correlation lengths have negative effects on the rate of convergence. This can be explained by examining ϱ defined by (7.0.8). From this we see that the coefficient ϱ appearing in the estimates (6.2.19)-(6.2.20) and (6.1.29)-(6.1.30), is approaching 1 as L_c becomes large. Hence, the effect of increasing L_c is a deterioration of the rate of convergence. This motivated developing and analyzing the anisotropic Smolyak method, which will be the primary focus of the remainder of this chapter.

The computational results for the $L^2(D)$ approximation error to the expected value, $E[u]$, using the anisotropic Smolyak interpolant, are shown in Figures 7.3. The anisotropic Smolyak algorithm for solving problem (7.0.1) was analyzed in Sections 6.2.1 and 6.2.2 and in the work Nobile et al. (2007). Here we consider the truncated probability space to have dimensions $N = 5$ and $N = 11$ and we compute approximate solutions up to level $\overline{w}$. To

estimate the computational error in the w-th level, for $0 \leq w \leq \overline{w}$, first we denote the maximum index utilized by $\mathscr{A}_{\boldsymbol{\alpha}}(w, N)$ in each stochastic direction, $\boldsymbol{\kappa}(\boldsymbol{\alpha}, w) \in \mathbb{N}_+^N$, given component-wise as:

$$\kappa_n(\boldsymbol{\alpha}, w) \equiv \max_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} \{i_n\}. \tag{7.0.10}$$

Then we introduce an enriched solution,

$$\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}+1, N)\pi_h u_N, \quad \text{with } \widehat{\alpha}_n = \left(\frac{\kappa_n(\boldsymbol{\alpha}, \overline{w}) - 1}{\kappa_n(\boldsymbol{\alpha}, \overline{w})}\right)\left(\frac{\alpha_n}{\underline{\alpha}}\right), \tag{7.0.11}$$

and approximate the computational error for $w = 0, 1, 2, \ldots, \overline{w}$ as

$$\|E[\epsilon]\| \approx \|E[\mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}+1, N)\pi_h u_N]\|. \tag{7.0.12}$$

By construction, the enriched solution possesses a maximum index $\boldsymbol{\kappa}(\widehat{\boldsymbol{\alpha}}, \overline{w}+1)$ which is larger by one in each direction, i.e. $\kappa_n(\widehat{\boldsymbol{\alpha}}, \overline{w}+1) = \kappa_n(\boldsymbol{\alpha}, \overline{w}) + \mathbf{1}$, $n = 1, \ldots, N$.

Tables 7.2 and 7.3 show the values $\boldsymbol{\kappa}(\boldsymbol{\alpha}, w)$ for computing $\mathscr{A}_{\boldsymbol{\alpha}}(w, N)$, with $w = 0, 1, 2, \ldots, \overline{w}$, as well as the value $\boldsymbol{\kappa}(\widehat{\boldsymbol{\alpha}}, \overline{w})$ for computing the enriched solution, $\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}, N)$, for the cases $L_c = 1/2$ and $L_c = 1/64$, respectively. The convergence plots shown in Figures 7.3 and 7.4 confirm, as expected, that the error decreases sub-exponentially, as the level w increases linearly. For highly anisotropic problems, i.e. $L_c = 1/2$, we observe that the rate of convergence is increased significantly with respect to the isotropic Smolyak method, as we anticipated in Theorem 6.2.7. We also observe that the convergence rate is dimension dependent and slightly deteriorates as N increases.

Table 7.2: The $N = 11$ components of the maximum indices $\boldsymbol{\kappa}(\boldsymbol{\alpha}, w)$ and $\widehat{\boldsymbol{\kappa}} = \boldsymbol{\kappa}(\widehat{\boldsymbol{\alpha}}, \overline{w}+1)$, defined by (7.0.10) and (7.0.11), respectively, used for solving problem (7.0.1) with a correlation length $L_c = 1/2$.

w	$\kappa_1(\boldsymbol{\alpha}, w)$	$\kappa_2{=}\kappa_3(\boldsymbol{\alpha}, w)$	$\kappa_4{=}\kappa_5(\boldsymbol{\alpha}, w)$	$\kappa_6{=}\kappa_7(\boldsymbol{\alpha}, w)$	$\kappa_8{=}\kappa_9(\boldsymbol{\alpha}, w)$	$\kappa_{10}{=}\kappa_{11}(\boldsymbol{\alpha}, w)$
0	1	1	1	1	1	1
1	1	2	1	1	1	1
2	2	3	1	1	1	1
3	3	4	2	1	1	1
4	4	5	2	1	1	1
5	5	6	3	1	1	1
$\widehat{\boldsymbol{\kappa}}$	6	7	4	2	2	2

To investigate the performance of the anisotropic algorithm by varying the correlation length L_c we examine Figure 7.4. We notice that the larger correlation lengths have positive

Table 7.3: The $N = 11$ components of the maximum indices $\boldsymbol{\kappa}(\boldsymbol{\alpha}, w)$ and $\widehat{\boldsymbol{\kappa}} = \boldsymbol{\kappa}(\widehat{\boldsymbol{\alpha}}, \overline{w})$, defined by (7.0.10) and (7.0.11), respectively, used for solving problem (7.0.1) with a correlation length $L_c = 1/64$.

w	$\kappa_1(\boldsymbol{\alpha}, w)$	$\kappa_2{=}\kappa_3(\boldsymbol{\alpha}, w)$	$\kappa_4{=}\kappa_5(\boldsymbol{\alpha}, w)$	$\kappa_6{=}\kappa_7(\boldsymbol{\alpha}, w)$	$\kappa_8{=}\kappa_9(\boldsymbol{\alpha}, w)$	$\kappa_{10}{=}\kappa_{11}(\boldsymbol{\alpha}, w)$
0	1	1	1	1	1	1
1	1	2	2	2	2	2
2	2	3	3	3	3	3
3	3	4	4	4	4	4
4	4	5	5	5	5	5
$\widehat{\boldsymbol{\kappa}}$	5	6	6	6	6	6

effects on the rate of convergence. This can be explained by examining $g(n)$ defined by (7.0.9). From this we see that as L_c becomes large the higher dimensions weigh less which greatly reduces the number of function evaluation required by the anisotropic sparse collocation method. On the other hand, the effect of decreasing L_c is a deterioration of the rate of convergence, due to the equal weighing of all directions. In this case, our anisotropic Smolyak algorithm and the isotropic Smolyak method obtain an equivalent convergence rate. Our final interest is to compare our sparse tensor product methods, both isotropic and anisotropic, with an anisotropic full tensor product method, proposed in Babuška et al. (2005d) and also with the Monte Carlo method.

Recall, the isotropic Smolyak algorithm for solving problem (7.0.1) was analyzed in Section 6.1 and the work Nobile et al. (2006) and can be constructed by equally weighing all stochastic directions, e.g. $\boldsymbol{\alpha} = \mathbf{1} \in \mathbb{R}^N_+$. To estimate the computational error in the w-th level we approximate $\|E[\epsilon]\| \approx \|E[\mathscr{A}(w, N)\pi_h u_N - \mathscr{A}(w+1, N)\pi_h u_N]\|$ using either Gaussian or Clenshaw-Curtis abscissas.

The anisotropic full tensor product algorithm can be described in the following way: given a tolerance *tol* the method computes a multi-index $\mathbf{p} = (p_1, p_2, \ldots, p_N)$, corresponding to the order of the approximating polynomial spaces $\mathcal{P}_{\mathbf{p}}(\Gamma^N)$. This adaptive algorithm increases the tensor polynomial degree with an anisotropic strategy: it increases the order of approximation in one direction as much as possible before considering the next direction. Table 7.4 and Table 7.5 show the values of components of the 11-dimensional multi-index $\mathbf{p}$ for different values of *tol*, corresponding to $L_c = 1/2$ and $L_c = 1/64$ respectively. These tables can also give insight into the anisotropic behavior of each particular problem and

should be compared with Tables 7.4 and Table 7.5, respectively. Observe, in particular, for the case $L_c = 1/64$ the algorithm predicts a multi-index $\mathbf{p}$ which is equal in all directions, i.e. an isotropic tensor product space. A convergence plot for $L_c = 1/2$ and $L_c = 1/64$ can be constructed by examining each row of the Table 7.4 and Table 7.5 respectively, and plotting the number of points in the tensor product grid versus the error in expectation. We estimate the error in expectation by $\|E[\epsilon]\| \approx \|E[u^N_{h,\mathbf{p}} - u^N_{h,\widetilde{\mathbf{p}}}]\|$, with $\widetilde{\mathbf{p}} = (p_1+1, p_2+1, \ldots, p_N+1)$. This entails an additional computational cost, which is bounded by the factor $\exp\left(\sum_{n=1}^N 1/p_n\right)$ times the work to compute $E[u^N_{h,\mathbf{p}}]$.

Table 7.4: The $N = 11$ components of the multi index $\mathbf{p}$ computed by he anisotropic full tensor product algorithm when solving problem (7.0.1) with a correlation length $L_c = 1/2$.

tol	$N = 1$	$N = 2, 3$	$N = 4, 5$	$N = 6, 7$	$N = 8, 9$	$N = 10, 11$
1.0e-04	1	1	1	1	1	1
1.0e-05	2	1	1	1	1	1
1.0e-06	2	2	1	1	1	1
1.0e-07	3	2	2	1	1	1
1.0e-08	4	3	2	1	1	1
1.0e-09	4	4	3	1	1	1
1.0e-10	5	5	3	2	1	1
1.0e-11	5	5	4	2	1	1
1.0e-12	5	6	4	2	1	1

Table 7.5: The $N = 11$ components of the multi index $\mathbf{p}$ computed by the anisotropic full tensor product algorithm when solving problem (7.0.1) with a correlation length $L_c = 1/64$.

tol	$N = 1$	$N = 2, 3$	$N = 4, 5$	$N = 6, 7$	$N = 8, 9$	$N = 10, 11$
1.0e-03	1	1	1	1	1	1
1.0e-06	2	2	2	2	2	2
1.0e-09	3	3	3	3	3	3
1.0e-12	4	4	4	4	4	4

The standard Monte Carlo Finite Element Method is the most common choice for anyone solving SPDEs such as (7.0.1) or (7.0.2) (Burkardt et al., 2007; Babuška et al., 2005d). If the aim is to compute a functional of the solution such as the expected value, one would approximate $E[u]$ numerically by sample averages of iid realizations of the stochastic input data. Given a number of realizations, $M \in \mathbb{N}_+$, we compute the sample average as follows: For each $k = 1, \ldots, M$, sample iid realizations of $a(\omega_k, \cdot)$ and $f(\omega_k, \cdot)$, solve problem (7.0.1) and construct finite element approximations $u^N_h(\omega_k, \cdot)$. We note that once we have

fixed $\omega = \omega_k$, the problem is completely deterministic, and may be solved by standard methods as in the collocation approach. Finally, approximate $E[u]$ by the sample average: $\overline{E}[u_{h,k}^N; M](\cdot) := \frac{1}{M}\sum_{k=1}^{M} u_h^N(\omega_k, \cdot)$.

For the cases $L_c = 1/2, 1/4, 1/16$ and $1/64$ we take $M = 2^i$, $i = 0, 1, 2, \ldots, 11$ realizations and compute the approximation to the error in expectation by $\|E[\epsilon]\| \approx \|\overline{E}[u_{h,k}^N; M] - E[\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}+1, N)\pi_h u_N]\|$, where $\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}+1, N)$ is the enriched anisotropic sparse solution defined previously and $\widehat{\boldsymbol{\alpha}}$ is defined by (7.0.11).

To study the advantages of an anisotropic sparse tensor product space as opposed to the isotropic sparse tensor product space or an anisotropic full tensor product space we show, in Figure 7.5, the convergence of these methods when solving problem (7.0.1), using correlation lengths $L_c = 1/2, 1/4, 1/16$ and $L_c = 1/64$ with $N = 11$. We also include 5 ensembles of the Monte Carlo method described previously. Figure 7.5 reveals that for the isotropic case with $L_c = 1/64$ the anisotropic and isotropic Smolyak method obtain a comparable convergence rate, both faster than the anisotropic full tensor product method.

On the contrary, opposite behavior can be observed for $L_c = 1/2$. Since, in this case, the rate of decay of the expansion is faster, the anisotropic full tensor method weighs heavily the important modes and, therefore, achieves a faster convergence than the isotropic Smolyak method. Similar conclusions can be made for the anisotropic Smolyak: the increased convergence for this method comes from the fact that it combines an optimal treatment of the problem anisotropy while reducing the *curse of dimensionality* via the use of sparse grids.

In all four cases we observe that all the 3 methods out-perform the Monte Carlo method. We know that the amount of work to reach the accuracy ϵ in the Monte Carlo approach can be approximated by $\epsilon \approx O(M^{-1/2})$ times the amount of work per sample, where M is the number of samples. This is only affected by the problem dimension through the increase of the work per sample. Nevertheless, the convergence rate is quite slow and a high level of accuracy is only achieved when an large amount of function evaluations are required. This can been seen from Figure 7.5 where we include reference lines with slopes $-1/2$ and -1, respectively, or in Table 7.6 where, for $N = 11$, we compare the work, proportional to the number of samples, which is the number of collocation points, required by each method to decrease the original error by a factor of 10^4 for all four correlation lengths $L_c = 1/2, 1/4, 1/16$ and $L_c = 1/64$.

The only aspect we have not yet discussed is applications to nonlinear problems such

Table 7.6: For $N = 11$, we compare the number of function evaluations required by the Anisotropic Smolyak (AS) using Clenshaw-Curtis abscissas, Anisotropic Full Tensor product method (AF) using Gaussian abscissas, Isotropic Smolyak (IS) using Clenshaw-Curtis abscissas and the Monte Carlo (MC) method using random abscissas, to reduce the original error of problem (7.0.1), in expectation, by a factor of 10^4.

L_c	AS	AF	IS	MC
1/2	50	2.5×10^2	2.5×10^3	5.0×10^9
1/4	1.6×10^2	1.2×10^3	4.0×10^3	2.0×10^9
1/16	2.0×10^2	2.0×10^3	5.0×10^2	1.6×10^9
1/64	3.1×10^2	2.0×10^5	3.6×10^2	1.3×10^9

as (7.0.2). First, the well-posedness of the nonlinear problem was verified in Chapter 3. We observed in Section 3.4 that the analyticity requirement given by Assumption 1.3.1 is challenging to verify in the case of nonlinear elliptic PDEs. In what follows we adopt the consequences of Assumption 1.3.1 and will show numerically that such regularity requirements are still justified since we obtain similar convergence results as in the linear problem (7.0.1). Moreover, we will also use the *a posteriori* approach, described in Section 5.2 to construct the anisotropic Smolyak method, described in Section 4.4.3. For a finite dimensional probability space with $N = 17$ we will compare this procedure with the Monte Carlo technique for computing the expected value $E[u]$ of the nonlinear SPDE described by (7.0.2).

In Example 6.1.1 we explained that for the linear problem described in Chapter 2, it was shown in the work Babuška et al. (2005c) that for a multi-index $\mathbf{p} = (p_1, \ldots, p_N)$, a tensor product polynomial interpolation on Gaussian abscissas achieves exponential convergence in each direction Y_n. In fact, this result and all the results of this work can be extended to any problem satisfying Assumption 1.3.1. We showed, in Theorem 3.4.1, that for the nonlinear problem, in general, this is not a trivial assumption to extend. However, Figure 7.6 reveals that the convergence of problem 7.0.2 maintains the exponential rate in each direction and that provides us with enough optimism that these results remain valid and can be extended to the general nonlinear case, described in Chapter 3.

In Figure 7.6 we take $L_c = 1/64$, $1/16$, $1/4$ and $1/2$ and use the *a posteriori* method described in Section 5.2 to plot $\log_{10}(\varepsilon_n)$, where the error ε_n is defined by (5.2.2), versus the number of collocation points in each direction $n = 1, 2, \ldots, N = 11$. Similar to Section 5.2 by using a linear least squares approximation to fit the data, the slope of each line will

give an estimate of $g(n)$. The values of $g(n)$ for $n = 1, 2, \ldots, 11$ can be seen in Table 7.7. Table 7.7 reveals that the a posteriori selection of $\boldsymbol{\alpha} \in \mathbb{R}_+^N$ performs well at dictating the behavior of problem (7.0.2) for the cases $L_c = 1/2$, $1/4$, $1/16$ and $1/64$. In comparing with Table 5.1 we see that exponential convergence rate predicted by $g(n)$ for the linear example 7.0.1 is almost indistinguishable to the rate predicted by $g(n)$ for the nonlinear problem 7.0.2. Again, we observe that for the highly anisotropic case, $L = 1/2$ the vector $\boldsymbol{\alpha}$ weighs heavily in the higher dimensions as opposed to the highly isotropic case $L = 1/64$, that approximately weighs equally in all directions.

Table 7.7: The $N = 11$ values of the function $g(n) = \alpha_n$ constructed from *a posteriori* information. The values $g(n)$, $n = 1, 2, \ldots, N$ are the slopes of N linear least squares fits to the error of an univariate anisotropic method when solving problem (7.0.2) with correlation lengths $L_c = 1/2$, $1/4$, $1/16$ and $1/64$.

	$g(1)$	$g(2) = g(3)$	$g(4) = g(5)$	$g(6) = g(7)$	$g(8) = g(9)$	$g(10) = g(11)$
$L_c = 1/2$	1.7	1.6	2.1	3.3	5.3	7.3
$L_c = 1/4$	2.0	1.7	1.6	1.9	2.4	2.8
$L_c = 1/16$	2.5	2.2	1.9	1.9	2.1	2.1
$L_c = 1/64$	3.1	2.4	2.5	2.4	2.5	2.4

To examine the extraordinary capabilities of the anisotropic Smolyak algorithm for computing accurate solutions to problems with a moderately large amount of uncertainty we compute the expected value , $E[u]$, of the nonlinear problem (7.0.2) using a correlation length $L_c = 1/4$. In this particular case, to fully resolve the noise requires $N = 17$ random variables in the expansion given by (7.0.3). We employ the *a posteriori* approach described in Section 5.2 to compute the weight $\boldsymbol{\alpha} \in \mathbb{R}_+^{17}$ used by the anisotropic Smolyak algorithm. For this particular problem the isotropic Smolyak and anisotropic full tensor product methods suffer from the *curse of dimensionality* as the calculation of $E[u]$ becomes computational infeasible. Therefore, in Figure 7.7 we finally compare the anisotropic Smolyak algorithm with the Monte Carlo method when problem (7.0.2) depends on a 17-dimensional probability space. In Figure 7.7 we illustrate the computational results for the $L^2(D)$ approximation error to the expected value, $E[u]$, using the above two approaches. Similar to the linear results described previously we observe that the anisotropic Smolyak method continues to achieve faster convergence rates compared to Monte Carlo, even in the case where $N = 17$ and the underlying SPDE is nonlinear.

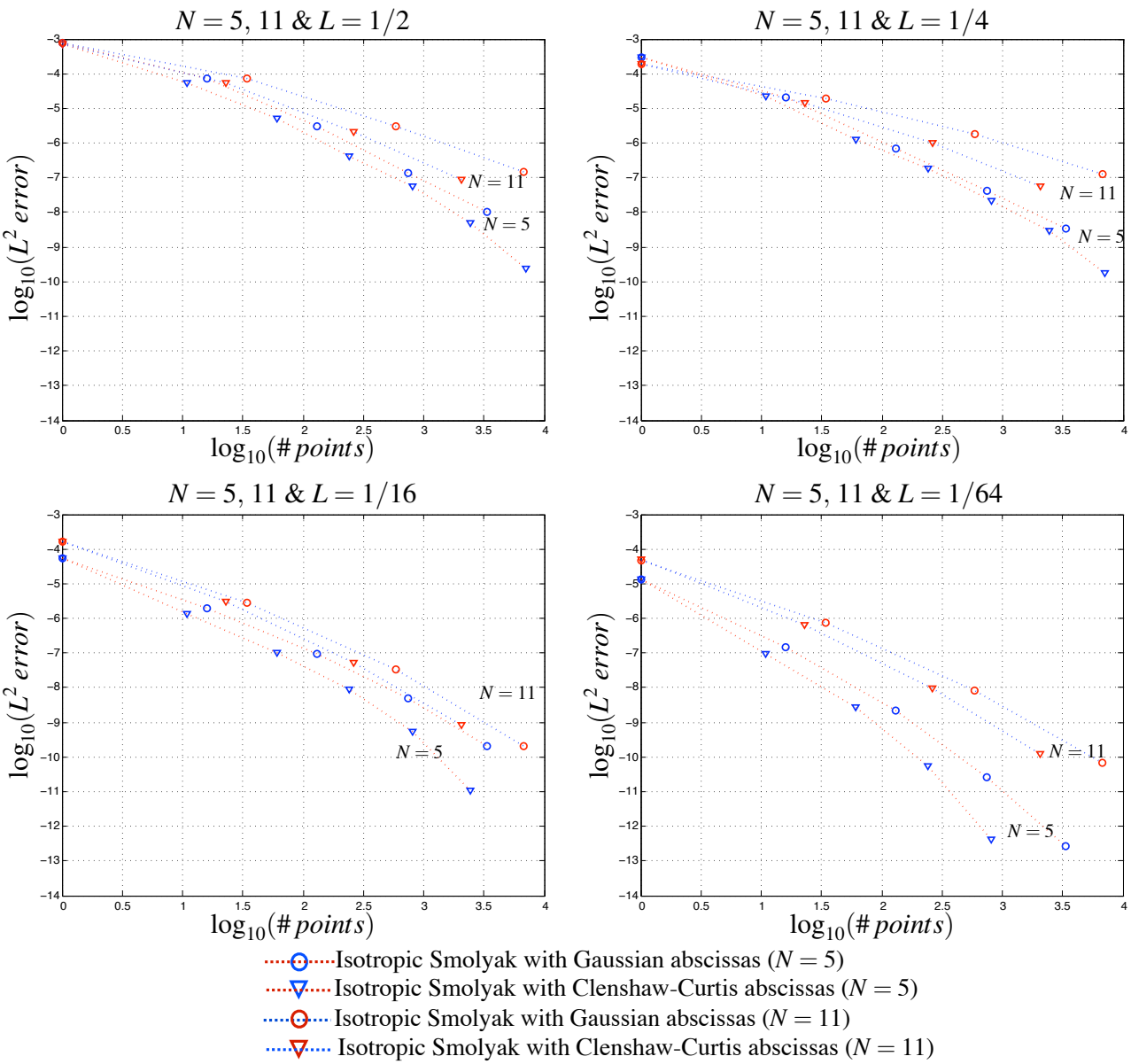

Figure 7.2: The convergence of the isotropic Smolyak algorithm for solving problem (7.0.1) with given correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$ using both the Gaussian and Clenshaw-Curtis abscissas. For a finite dimensional probability space Γ^N with $N = 5$ and $N = 11$ we plot $\log(\epsilon)$ versus the logarithm of the number of collocation points. The $L^2(D)$ approximation error in the expected value for the isotropic sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[\mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N - \mathscr{A}_{\boldsymbol{\alpha}}(w+1, N)\pi_h u_N]\|_{L^2(D)}$ where $w = 0, 1, \ldots, \overline{w}$.

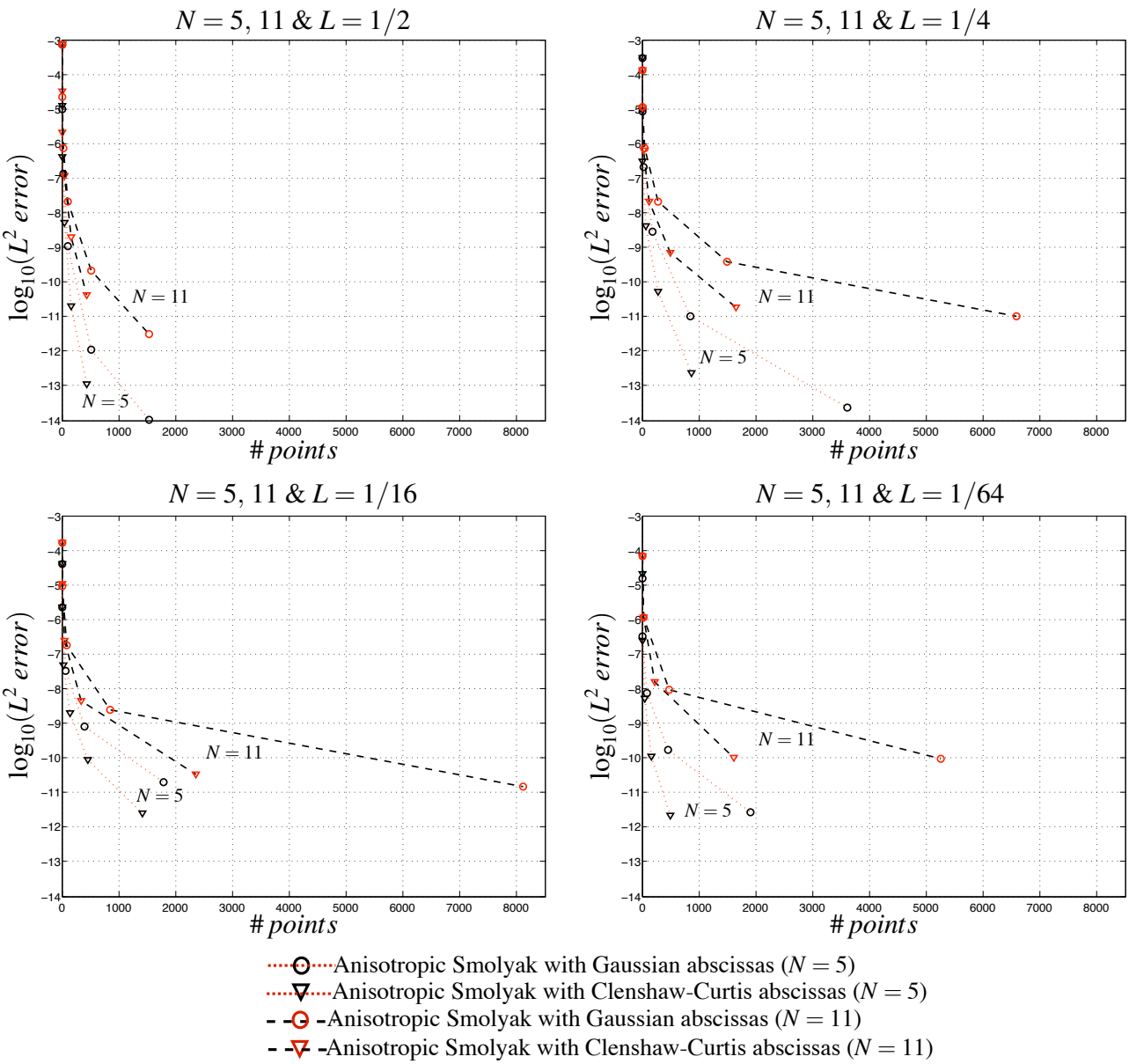

Figure 7.3: The rates of convergence of the anisotropic Smolyak algorithm for solving problem (7.0.1) with given correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$ using both the Gaussian and Clenshaw-Curtis abscissas. For a finite dimensional probability space Γ^N with $N = 5$ and $N = 11$ we plot $\log(\epsilon)$ versus the number of collocation points. The $L^2(D)$ approximation error in the expected value for the anisotropic sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[\mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}, N)\pi_h u_N]\|_{L^2(D)}$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (7.0.11).

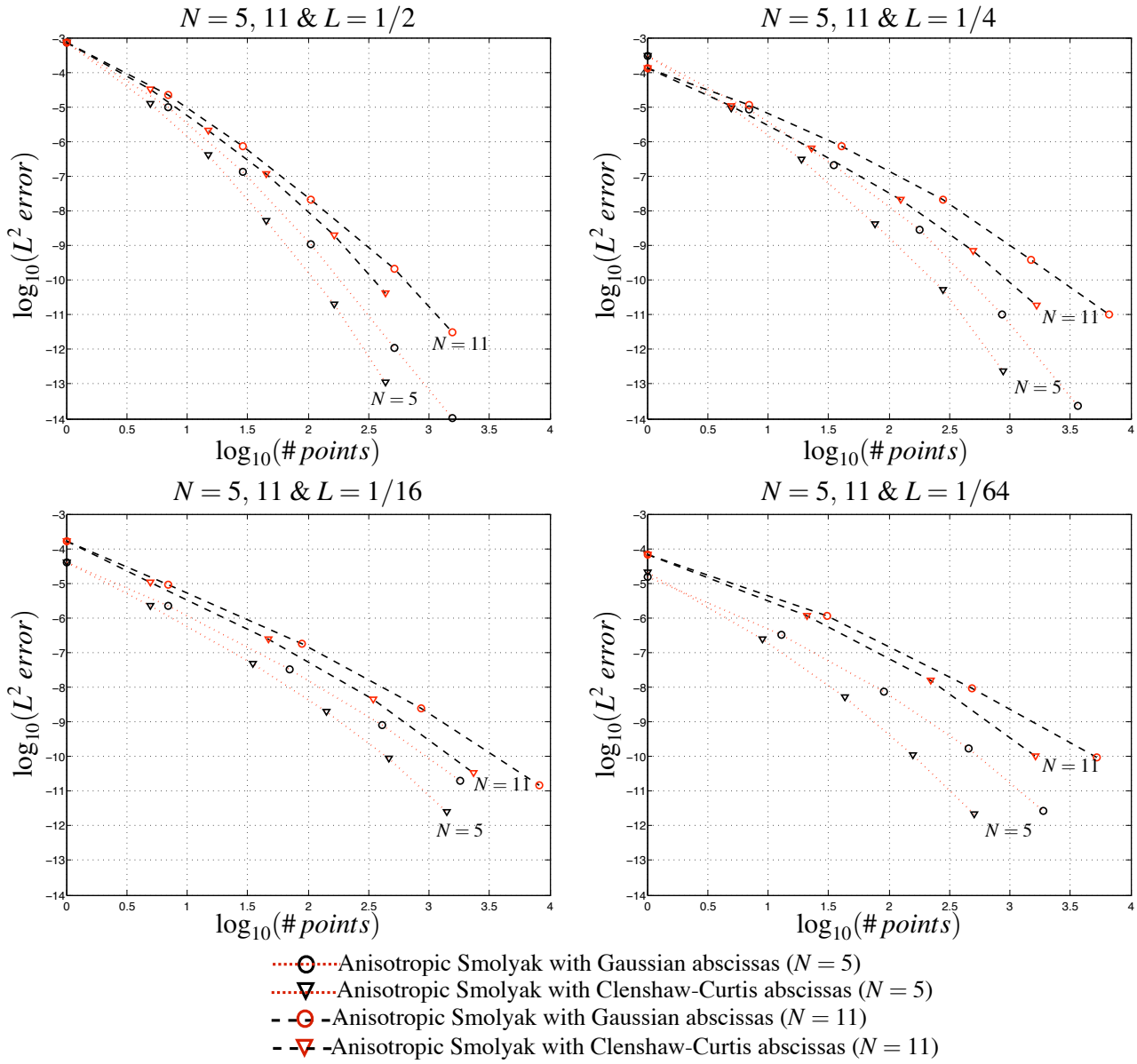

Figure 7.4: The convergence of the anisotropic Smolyak algorithm for solving problem (7.0.1) with given correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$ using both the Gaussian and Clenshaw-Curtis abscissas. For a finite dimensional probability space Γ^N with $N = 5$ and $N = 11$ we plot $\log(\epsilon)$ versus the logarithm of the number of collocation points. The $L^2(D)$ approximation error in the expected value for the anisotropic sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[\mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}, N)\pi_h u_N]\|_{L^2(D)}$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (7.0.11).

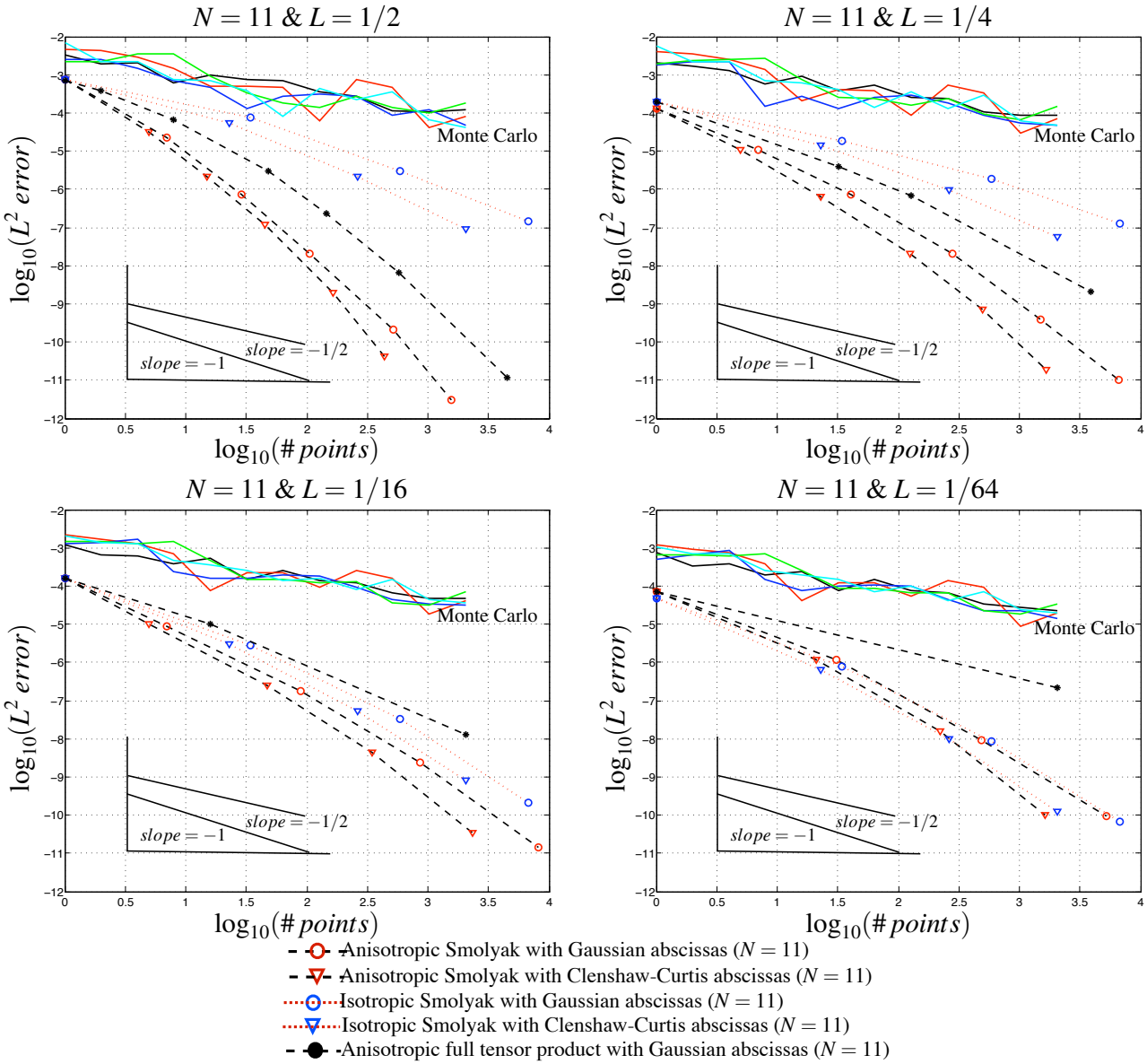

Figure 7.5: A 11-dimensional comparison of the anisotropic Smolyak method, the isotropic Smolyak method, the anisotropic full tensor product algorithm and Monte Carlo approach for solving problem (7.0.1) with correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$. The $L^2(D)$ approximation error in the expected value for the sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx E[\mathscr{A}_{\boldsymbol{\alpha}}(w,N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w},N)\pi_h u_N]$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (7.0.11). The $L^2(D)$ approximation error in the expected value for the anisotropic full tensor product method is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[u^N_{h,\mathbf{p}} - u^N_{h,\widetilde{\mathbf{p}}}]\|_{L^2(D)}$, where $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ and $\widetilde{\mathbf{p}} = (p_1+1, p_2+1, \ldots, p_N+1)$. The $L^2(D)$ approximation error in the expected value for the Monte Carlo method is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|\overline{E}[u^N_{h,k}; M] - E[\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w},N)\pi_h u_N]\|_{L^2(D)}$, where $M = 2^i$, $i = 0, 1, 2, \ldots, 10$.

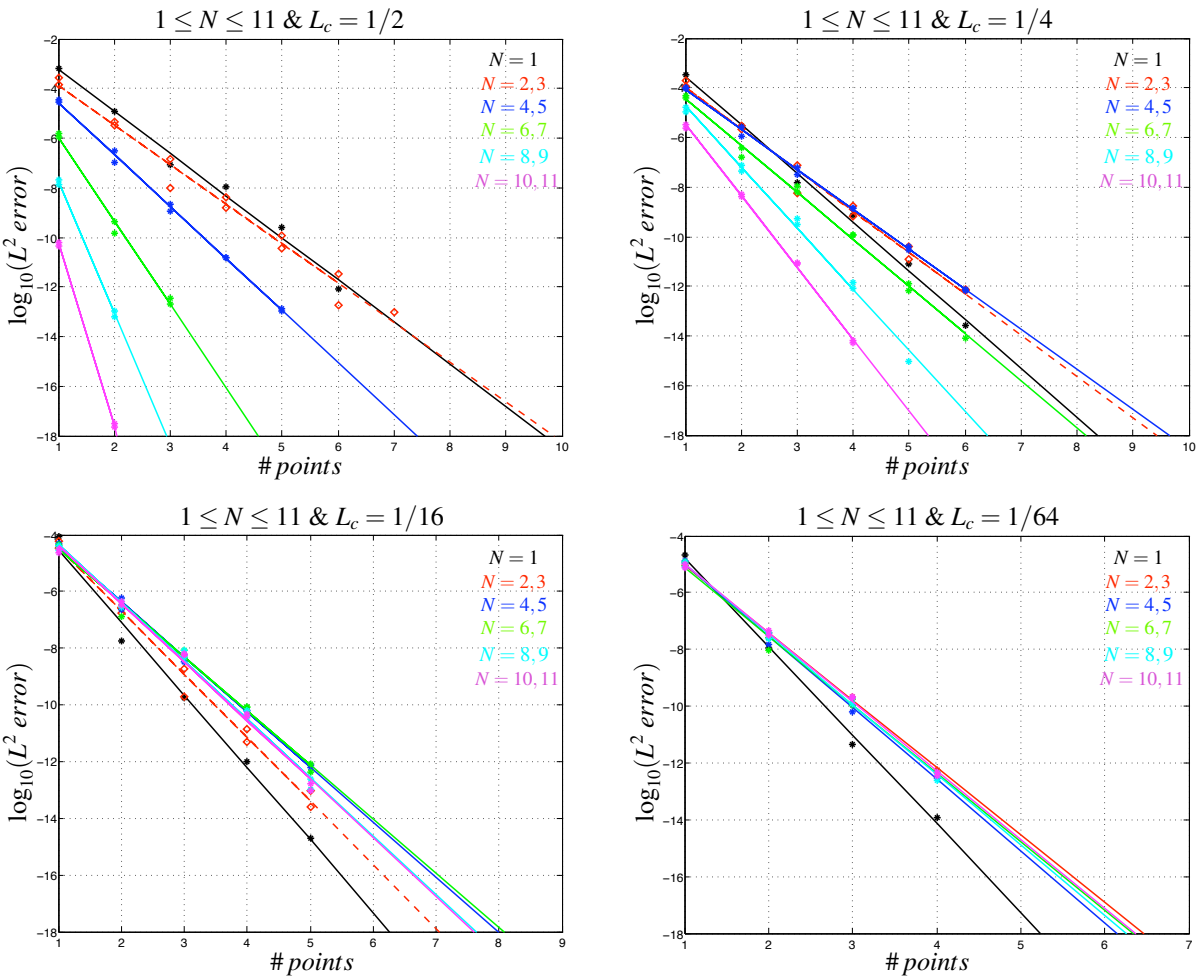

Figure 7.6: A linear least square approximation to fit $\log_{10}(\|E[\varepsilon_n]\|_{L^2(D)})$ versus p_n with ε_n defined by (5.2.2). For $n = 1, 2, \ldots, N = 11$ we plot the cases corresponding to $L_c = 1/2$, $1/4$, $1/16$ and $1/64$ when solving the nonlinear problem given by (7.0.2).

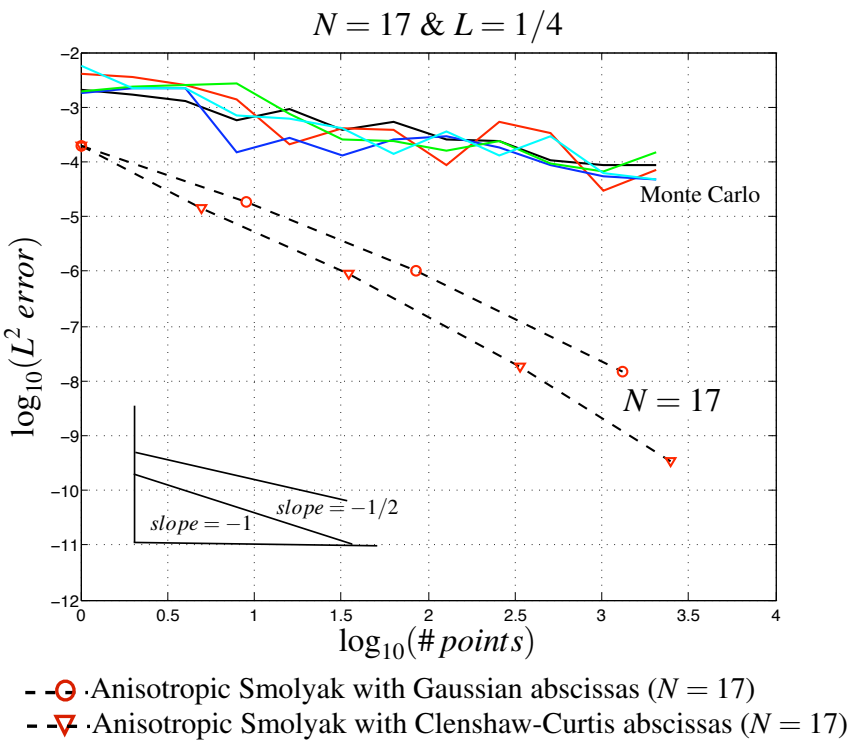

Figure 7.7: A 17-dimensional comparison of the anisotropic Smolyak method with Monte Carlo approach for solving problem (7.0.2) with correlation lengths $L_c = 1/4$ using the *a posterirori* approach described in Section 5.2. The $L^2(D)$ approximation error in the expected value for the anisotropic sparse collocation method is given by: $\|E[\epsilon]\|_{L^2(D)} \approx E[\mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}, N)\pi_h u_N]$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (7.0.11). The $L^2(D)$ approximation error in the expected value for the Monte Carlo method is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|\overline{E}[u^N_{h,k}; M] - E[\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}, N)\pi_h u_N]\|_{L^2(D)}$, where $M = 2^i$, $i = 0, 1, 2, \ldots, 10$.

CHAPTER 8

SUMMARY

In this work we proposed and analyzed several sparse grid stochastic collocation methods for solving both linear and nonlinear elliptic partial differential equations whose coefficients and forcing terms depend on a finite number of random variables. When introducing the novel weighted Smolyak method we also described an optimal choice for the weight parameters. These weights adaptively tune the anisotropy of the method for each given problem. Their systematic choice can be based both on *a priori* or *a posteriori* information and is motivated by the regularity of the solution and the error estimates derived in this work. The sparse grids are constructed from the Smolyak algorithm, utilizing either Clenshaw-Curtis or Gaussian abscissas. Both methods lead to the solution of uncoupled deterministic problems and, as such, are fully parallelizable like a Monte Carlo method.

This method extends the work proposed in Babuška et al. (2005c) where a stochastic collocation method on tensor product grids was proposed to solve problem (1.1.2). Not only do we include nonlinear elliptic problems, as in Example 1.1.2, but the use of sparse grids considered in the present work (as opposed to full tensor grids), reduces considerably the *curse of dimensionality* and allows us to treat effectively problems that depend on a moderately large number of random variables, while keeping a high level of accuracy. The new methods can be found in our latest works Nobile et al. (2006, 2007).

Upon assumption that the solution depends analytically on each random variable (which is a reasonable assumption for a certain class of applications, see Babuška et al. (2004b, 2005c)), we have provided a full convergence analysis and demonstrated (sub)-exponential convergence of the "probability error" in the asymptotic regime and algebraic convergence of the "probability error" in the pre-asymptotic regime, with respect to the total number of collocation points used in the sparse grid.

The main theoretical results for the isotropic and anisotropic methods are given in Theorem 6.1.2, Theorem 6.1.8 and Theorem 6.2.7, Theorem 6.2.13 respectively. The consequences of these results are confirmed numerically by the examples presented in Chapter 7.

The isotropic Smolyak method, introduced in Section 4.4.2 and analyzed in Section 6.1, is very effective for problems whose input data depend on a moderate number of random variables, which "weigh equally" in the solution. For such an isotropic situation the displayed convergence is faster than standard collocation techniques built upon full tensor product spaces.

On the other hand, the convergence rate deteriorates when we attempt to solve highly anisotropic problems, such as those appearing when the input random variables come e.g. from Karhunen-Loève-type truncations of "smooth" random fields. In such cases, a full anisotropic tensor product approximation, as proposed in Babuška et al. (2005c,d), may still be more effective for a small or moderate number of random variables.

To facilitate faster convergence for these types of problems we also introduced an anisotropic version of the Sparse Grid Stochastic Collocation method in Section 4.4.3, which combines an optimal treatment of the anisotropy of the problem while reducing the *curse of dimensionality* via the use of sparse grids. Again, our theoretical results shown in Section 6.2 are confirmed numerically in Chapter 7 and we see that our new method outperforms the previous existing numerical methods, such as Stochastic Galerkin (Section 4.3), Full tensor product interpolation (Section 4.4.1) and Monte Carlo (Section 4.2), to solve problems such as (1.1.2) and (1.1.3) depending on a moderately large number of random variables.

One aspect that we did not consider in this work was the large computation cost that is effected by each realization. Determining accurate statistical information about outputs from ensembles of realizations is very difficult whenever the input-output map involves the (computational) solution of systems of nonlinear partial differential equations, such (1.1.3). Recently, in applications such as control and optimization that also require multiple solutions of PDEs, there has been much interest in reduced-order models (ROMs) that greatly reduce the cost of determining approximate solutions.

Future directions of this research will include the combination of sparse grid techniques in a probability setting with the use of ROMs for determining outputs that depend on solutions of stochastic PDEs. One is then able to cheaply determine much larger ensembles,

but this increase in sample size is countered by the lower fidelity of the ROM used to approximate the state. In the contexts of proper orthogonal decomposition-based ROMs, we will explore these counteracting effects on the accuracy of statistical information about outputs determined from ensembles of solutions. For a particular nonlinear SPDE, we will devote the next chapter to investigating the use of ROMs as an effective technique to further reduce the complexity of such problems. This will serve as the groundwork of much future work and research endeavors (Burkardt et al., 2007).

CHAPTER 9

A REDUCED-ORDER MODEL FOR SOLVING SPDES: A PROMISING RESEARCH DIRECTION

Realistic simulations of complex systems governed by nonlinear partial differential equations must account for the "noisy" features of the modeled phenomena, such as material properties, coefficients, domain geometry, excitations and boundary data. "Noise" can be understood as uncertainties in the specification of the physical model; because of noise, the behavior of a complex system is at least partially unpredictable. A simulation can attempt to capture the noisy aspects of a system by describing the simulation input data as random fields. This turns the problem into a stochastic partial differential equation (SPDE). We will consider such problems, characterized by nonlinear partial differential equations, and for which the input data are not purely deterministic; for example, the coefficients or the right-hand-side of the partial differential equation may be regarded as sums of a deterministic and stochastic function.

For a given system, various stochastic perturbation techniques have been considered (Allen et al., 1998; Alós and Bonaccorsi, 2002; Babuška and Liu, 2001; Babuška et al., 2005b, 2004a; Davie and Gaines, 2000; Deb et al., 2001a; Mytnik, 2002; Yan, 2002, 2003). This chapter will focus on nonlinear SPDE's in which the stochastic inputs are modeled as white noise, i.e., they are not significantly correlated. The aim of our work is to efficiently determine statistical information about the random field $u = u(t, \mathbf{x}; \omega)$ from numerical approximations of the nonlinear SPDE driven by white noise:

$$\frac{\mathrm{d}u}{\mathrm{d}t} = Au - \gamma N(u) + g + \epsilon \frac{\mathrm{d}\mathbf{W}}{\mathrm{d}t}, \quad \mathbf{x} \in D, \quad \omega \in \Omega, \quad t > 0. \tag{9.0.1}$$

Here $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, is a convex, bounded and polygonal spatial domain, $(\Omega, \mathcal{F}, P)$ is a probability space described in section 9.1, and A is a linear second-order elliptic operator

with deterministic coefficients, defined on a space of functions satisfying certain boundary conditions, $N(u)$ is a nonlinear function of the random process u, g represents a deterministic function and $\mathbf{W}$ denotes an infinite dimensional *Brownian motion* or *Wiener process*. The additive noise that appears in (9.0.1) is in the form of space-time Brownian white noise as described in section 9.1.1. The amplitudes of the noise and the nonlinearity are controlled by parameters ϵ and γ, respectively. Once the equation is reformulated into a weak form, the usual Galerkin finite element approach can be used to produce a discretized system suitable for solution on a computer.

Generally, obtaining precise statistics about ensembles of realizations of nonlinear SPDEs such as (9.0.1) entails a high cost in both memory and CPU. This cost is exhibited in many recent attempts on similar problems (Dalang and Mueller, 2003; Gyongy, 1999; Gyongy and Martinez, 2003). Even with the use of reliable nonlinear solvers and carefully chosen solution schemes, these computations involve formidable work. Typical finite element codes may require the use of many thousands of degrees of freedom for the accurate simulation of deterministic PDEs. The situation becomes far worse when the same techniques are extended to SPDEs (Du and Zhang, 2002) for which multiple realizations are usually required.

It is natural to consider a reduced-order model (ROM), such as Burkardt et al. (2003, 2004). A reduced-order model attempts to determine acceptable approximate solutions of a PDE while using very few degrees of freedom. One way to achieve this efficiency is for the models to use basis functions that are in some way intimately connected to the problem being solved. Once a low-dimensional reduced basis has been determined, it may be used in a new Galerkin system to solve related instances of the PDE. In this way, a ROM may be used to efficiently explore the behavior of large ensembles of PDE solutions. This is the kind of efficiency needed when attempting to compute realistic statistics from outputs of the SPDE.

There have been many reduced-order modeling techniques proposed; see Burkardt et al. (2003, 2004); Kunisch and Volkwein (2001); Nagy (1979) and the references cited therein. The most popular reduced-order modeling approach for nonlinear PDEs is based on proper orthogonal decomposition (POD) analysis. POD begins with a set of $\widetilde{m}$ precomputed solutions of the equation, often called *snapshots*; these could be generated by evaluating the computational solution of a transient problem at many instants of time or over a range of values of the problem parameters. These solutions are presumably obtained using costly,

large-scale, high-fidelity codes. The K-dimensional POD basis is then formed from the K eigenvectors corresponding to the dominant eigenvalues of the snapshot correlation matrix. This basis may then be used to construct a new finite element system of much reduced order, suitable for generating approximate solutions, at least within a limited range of the underlying snapshot data. POD-based model reduction has been applied with some success to several problems, most notably in fluid mechanics. For detailed discussions, one may consult Aubry et al. (1993); Berkooz et al. (1993); Berkooz and Titi (1993); Burkardt et al. (2003, 2004); Deane et al. (1991); Graham and Kevrekidis (1996); Holmes et al. (1996, 1997); Kunisch and Volkwein (1999, 2001); Lumley (1970); Park and Cho (1996); Park and Chung (2002); Park (2000); Park and Lee (1998, 2000, 2002); Rathinam and Petzold (2003, 2002); Ravindran (2000a,b); Rodríguez and Sirovich (1990); Sirovich (1987); Smaoui and Armbruster (1997); Volkwein (1999, 2001).

The efficiency of a POD basis comes from its low dimension combined with its good approximating power. However, the ability of a POD-based basis to approximate the state of a system is totally dependent on the information contained in the snapshot set. Certainly, a POD-based basis cannot contain more information than that contained in the snapshot set. Thus, crucial to the success of the POD-based approaches to model reduction is the generation of "good" snapshot sets, which manage to capture a wide range of system behaviors.

In order to present a standard approach, we focus on the case $D = (0,1) \times (0,1) \subset \mathbb{R}^2$, $I := [t^1, T]$, a bounded interval of $\mathbb{R}$ and $Au = \Delta u = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) u$ with homogeneous Dirichlet boundary conditions. However, much of our results and computations can be readily extended to higher spatial dimensions and more general second-order elliptic operators. An effort has been made to present most of this discussion using the familiar terminology from standard finite element methods for the numerical approximation of deterministic PDEs.

This final chapter is organized as follows: we first formulate the idea of a probability space, a Wiener process and Brownian motion in a plane half strip. Next, we discuss discrete Brownian white noise, the model nonlinear stochastic PDE and its finite element approximation. We then apply the Monte Carlo-Galerkin finite element method, introduced in Section 4.2, by sampling input data for the nonlinear SPDE and then approximating the corresponding realization of the solution. Reduced-order models (ROMs) are generated and analyzed with respect to noise driven by the Brownian motion; computational error results

are presented. Finally, some concluding remarks are given.

9.1 Some notation

We begin by recalling the mathematical formulation of a probability space $(\Omega, \mathcal{F}, P)$, where Ω, $\mathcal{F}$ and P are the set of random events, the minimal σ-algebra of subsets of Ω, and the probability measure, respectively. We define $D \subset \mathbb{R}^N$ to be a bounded spatial domain.

If X is a real random variable in $(\Omega, \mathcal{F}, P)$ with $X \in L^1(\Omega)$, we denote its expected value by

$$E[X] = \int_\Omega X(\omega) P(\,\mathrm{d}\omega) = \int_\mathbb{R} x\mu(\,\mathrm{d}x). \tag{9.1.1}$$

Here μ is the distribution probability measure for X, defined on the Borel set $\mathcal{B}$ of $\mathbb{R}$, given by

$$\mu(\mathcal{B}) = P\left(X^{-1}(\mathcal{B})\right). \tag{9.1.2}$$

We will assume that $\mu(\mathcal{B})$ is absolutely continuous with respect to Lebesgue measure; then there exists a density function for X, $\rho : \mathbb{R} \to \mathbb{R}^+$, such that

$$E[X] = \int_\mathbb{R} x\rho(x)\,\mathrm{d}x. \tag{9.1.3}$$

Next, we define what is meant by a measurable *stochastic* and *Wiener* process and formally explain the concept of a Brownian motion on a plane half-strip (otherwise known as a Brownian sheet):

Definition 9.1.1 *An (N, d)-valued stochastic process $\mathbf{U}(\mathbf{z}) = \left\{\mathbf{U}(\mathbf{z}); \mathbf{z} \in \mathbb{R}^N_+\right\} = \{(\mathbf{U}_1(\mathbf{z}), \ldots, \mathbf{U}_d(\mathbf{z}))\}_{\mathbf{z}\in\mathbb{R}^N_+}$, defined on a probability space $(\Omega, \mathcal{F}, P)$, is measurable if*

$$\mathbf{U} : \Omega \times \mathbb{R}^N_+ \to \mathbb{R}^d$$

is $\mathcal{F} \times \mathcal{B}(\mathbb{R}^N_+)$ measurable.

Notice that when $N = 1$, $\mathbf{U}$ is just an $\mathbb{R}^d$-valued stochastic process.

Definition 9.1.2 *An $\mathbb{R}$-valued stochastic process $W(t) = \{W(t)\}_{t\in\mathbb{R}^1_+}$ is called a Brownian motion or Wiener process if*

- $W_0 = 0$ *a.s.,*

- *$W_t - W_s$ is $N(0, t-s)$*[1] *for all $t \geq s \geq 0$*
- *for all times $0 < t_1 < t_2 < \cdots < t_n$, the random variables $W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent ("independent increments").*

Notice that

$$E(W_t) = 0, \quad E(W_t^2) = t \text{ for each } t \geq 0.$$

We complete this section by defining Brownian motion on a plane half strip or Brownian sheet $\mathbf{W}(t, \mathbf{z})$, $t \in \mathbb{R}_+^1$ and $\mathbf{z} \in \mathbb{R}_+^{N-1}$.

Definition 9.1.3 *Let $\mathbf{W}(t, \mathbf{z}) = \{\mathbf{W}(t, \mathbf{z})\}_{t \in \mathbb{R}_+^1, \mathbf{z} \in \mathbb{R}_+^{N-1}}$ denote an (N, d) Brownian sheet. That is, $\mathbf{W}$ is the N-parameter Gaussian random field with values in $\mathbb{R}^d$, its mean-function is zero, and its covariance function is given by the following: for all $t, s \in \mathbb{R}_+^1$ and $\mathbf{z}, \mathbf{w} \in \mathbb{R}_+^{N-1}$ and all $1 \leq i, j \leq d$,*

$$E\left(\mathbf{W}_i(t, \mathbf{z}) \cdot \mathbf{W}_j(s, \mathbf{w})\right) = \begin{cases} (t \wedge s) \times \prod_{k=1}^{N} (z_k \wedge w_k), & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases}$$

We have written $\mathbf{W}(t, \mathbf{z})$ in vector form as $\{(\mathbf{W}_1(t, \mathbf{z}), \ldots, \mathbf{W}_d(t, \mathbf{z}))\}_{t \in \mathbb{R}_+^1, \mathbf{z} \in \mathbb{R}_+^{N-1}}$, as is customary. When $N = 1$, $\mathbf{W}$ is just Brownian motion in $\mathbb{R}^d$.

Now consider the three-dimensional Gaussian white noise

$$\Delta \mathbf{W}(t, \mathbf{x}) = \frac{\partial^3 \mathbf{W}}{\partial t \partial x_1 \partial x_2}(t, \mathbf{x}; \omega), \quad \mathbf{x} = (x_1, x_2) \in D \subset \mathbb{R}^2,$$

where $\mathbf{W}(t, \mathbf{x})$ is Brownian motion on a half plane or a Brownian sheet satisfying, for $t, s \in \mathbb{R}_+^1$ and $\mathbf{x}, \mathbf{y} \in D$:

$$\begin{aligned} E\left(\Delta \mathbf{W}(t, \mathbf{x}) \cdot \Delta \mathbf{W}(s, \mathbf{y})\right) &= \delta(t - s) \times \delta(\mathbf{x} - \mathbf{y}) \\ &= \delta(t - s) \times \delta(x_1 - y_1) \times \delta(x_2 - y_2) \end{aligned} \tag{9.1.4}$$

with δ the usual Dirac δ-function. According to Allen et al. (1998), application of standard finite element techniques requires that $\Delta \mathbf{W}(t, \mathbf{x}) = \Delta \mathbf{W}(t, x_1, x_2)$ be modeled by the piecewise constant random process given in the following section.

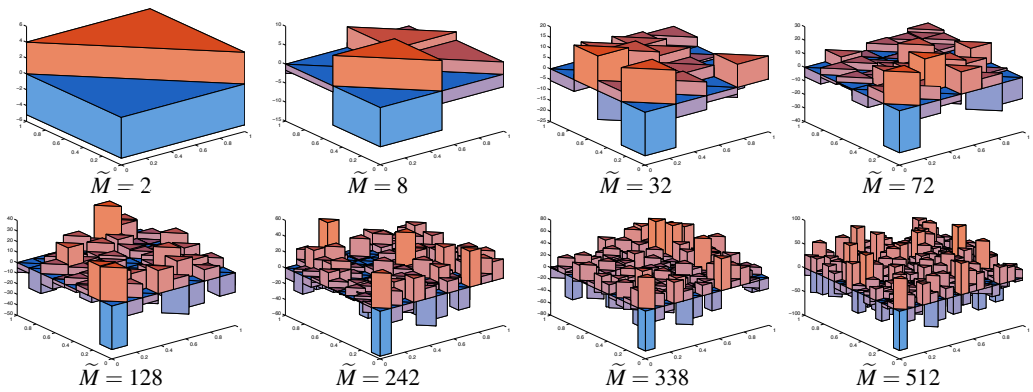

Figure 9.1: Piecewise constant approximation for the noise $\Delta\widehat{\mathbf{W}}(t,\mathbf{x})$, defined by 9.1.5, at the final time $T = 1$ over a uniform finite element triangulation. The number of elements ranges from $\widetilde{M} = 2$ to $\widetilde{M} = 512$.

9.1.1 The approximation of Brownian white noise

Following Allen et al. (1998); Du and Zhang (2002); Yan (2002), we regularize the noise through discretization. For simplicity, we use a uniform discretization $0 = t^1 < t^2 < \cdots < t^n < \cdots < t^{\widetilde{N}} = T$ of the time interval $[t^1, T]$ given by

$$t^{n+1} = t^1 + n\Delta t, \quad t > 1$$

with $\Delta t = (T - t_1)/\widetilde{N}$, for some integer $\widetilde{N}$ large enough so that $\Delta t \in (0,1)$. Likewise, we consider a uniform finite element triangulation $\{\tau_j\}_{j=1}^{\widetilde{M}}$ of the square D characterized by the parameter h, which we take to be the longest side of any triangle. We denote by $\Delta\tau$ the area of any of the finite elements.

Using the finite element triangulation $\{\tau_j\}_{j=1}^{\widetilde{M}}$ of D, and guided by Lemma 9.1.4, a "reasonable" piecewise constant approximation of the noise is given by

$$\Delta\widehat{\mathbf{W}}(t,\mathbf{x}) = \frac{\partial^3\widehat{\mathbf{W}}}{\partial t\partial x_1\partial x_2}(t,\mathbf{x};\omega) = c_{\widetilde{N}\widetilde{M}}\sum_{i=1}^{\widetilde{N}}\sum_{j=1}^{\widetilde{M}}\eta_{ij}(\omega)\chi_i(t)\chi_j(\mathbf{x}), \tag{9.1.5}$$

where

$$c_{\widetilde{N}\widetilde{M}} = \frac{1}{\sqrt{\Delta t}\sqrt{\Delta\tau}}$$

and, for $i = 1, 2, \ldots, \widetilde{N}-1$, $j = 1, 2, \ldots, \widetilde{M}-1$, $\eta_{ij} \in N(0,1)$ is independently and identically distributed (iid),

$$\sqrt{\Delta t}\sqrt{\Delta\tau}\,\eta_{ij} = \int_{t^i}^{t^{i+1}}\int_{\tau_j} \mathrm{d}\mathbf{W}(t,\mathbf{x}),$$

[1] $N(0, t-s)$ is the zero-mean Gaussian (or normal) distribution with variance $t - s$

$$\chi_i(t) = \begin{cases} 1, & \text{if } t^i \le t < t^{i+1}, \\ 0, & \text{otherwise} \end{cases} \qquad \text{and} \qquad \chi_j(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \tau_j, \\ 0, & \text{otherwise} \end{cases}.$$

That is, $\chi_i(t)$ is the characteristic function for the ith time interval and $\chi_j(\mathbf{x})$ is the characteristic function for the jth finite element. For an arbitrary i and j, the discrete analog of (9.1.4) for the piecewise constant approximation to $\Delta\widehat{\mathbf{W}}$ is given by

$$E\left(\Delta\widehat{\mathbf{W}}(t,\mathbf{x}) \cdot \Delta\widehat{\mathbf{W}}(s,\mathbf{y})\right) = \begin{cases} \dfrac{1}{\Delta t} \times \dfrac{1}{\Delta \tau}, & \text{if } t^i \le t, s < t^{i+1} \text{ and } \mathbf{x}, \mathbf{y} \in \tau_j, \\ 0, & \text{otherwise} \end{cases}$$

Therefore,

$$\lim_{\widetilde{N},\widetilde{M}\to\infty} E\left(\Delta\widehat{\mathbf{W}}(t,\mathbf{x}) \cdot \Delta\widehat{\mathbf{W}}(s,\mathbf{y})\right) = \delta(t-s) \times \delta(\mathbf{x}-\mathbf{y})$$

as required. In Figure 9.1.1, some sample realizations of the piecewise constant approximation of the three-parameter white noise are displayed for $\Delta t = 1$ and various values of $\widetilde{M}$ at the final time $T = 1$.

Similarly to Allen et al. (1998), we have the following result for a nonrandom function $f(t,\mathbf{x})$.

Lemma 9.1.4 *Let f be a nonrandom function and Lipschitz continuous on $[0,T] \times [0,1] \times [0,1]$. In particular, assume that there is a positive constant $\gamma \ge 0$ such that $|f(t,\mathbf{x}) - f(s,\mathbf{y})| \le \gamma\left(|t-s| + \|\mathbf{x}-\mathbf{y}\|\right)$ for $(t,\mathbf{x}),(s,\mathbf{y}) \in [0,T] \times [0,1] \times [0,1]$. Then*

$$E\left[\int_0^T \int_D f(t,\mathbf{x})\ d\mathbf{W}(t,\mathbf{x}) - \int_0^T \int_D f(t,\mathbf{x})\ d\widehat{\mathbf{W}}(t,\mathbf{x})\right]^2 \le 2T\gamma^2\left((\Delta t)^2 + h^2\right).$$

Proof. The proof is as follows:

$$\begin{aligned}
&E\left[\int_0^T \int_D f(t,\mathbf{x})\ \mathrm{d}\mathbf{W}(t,\mathbf{x}) - \int_0^T \int_D f(t,\mathbf{x})\ \mathrm{d}\widehat{\mathbf{W}}(t,\mathbf{x})\right]^2 \\
&= E\left[\sum_{i=1}^{\widetilde{N}} \sum_{j=1}^{\widetilde{M}} \int_{t^i}^{t^{i+1}} \int_{\tau_j} f(t,\mathbf{x})\ \mathrm{d}\mathbf{W}(t,\mathbf{x})\right. \\
&\quad \left. - \sum_{i=1}^{\widetilde{N}} \sum_{j=1}^{\widetilde{M}} \int_{t^i}^{t^{i+1}} \int_{\tau_j} f(s,\mathbf{y}) \left(\frac{1}{\Delta t \Delta \tau} \int_{t^i}^{t^{i+1}} \int_{\tau_j} \mathrm{d}\mathbf{W}(t,\mathbf{x})\right)\ \mathrm{d}\mathbf{y}\ \mathrm{d}s\right]^2 \\
&= E\left[\sum_{i=1}^{\widetilde{N}} \sum_{j=1}^{\widetilde{M}} \int_{t^i}^{t^{i+1}} \int_{\tau_j} \left(f(t,\mathbf{x}) - \frac{1}{\Delta t \Delta \tau} \int_{t^i}^{t^{i+1}} \int_{\tau_j} f(s,\mathbf{y})\ \mathrm{d}\mathbf{y}\ \mathrm{d}s\right)\ \mathrm{d}\mathbf{W}(t,\mathbf{x})\right]^2
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{\widetilde{N}} \sum_{j=1}^{\widetilde{M}} \int_{t^i}^{t^{i+1}} \int_{\tau_j} \left(f(t,\mathbf{x}) - \frac{1}{\Delta t \Delta \tau} \int_{t^i}^{t^{i+1}} \int_{\tau_j} f(s,\mathbf{y}) \, \mathrm{d}\mathbf{y} \, \mathrm{d}s \right)^2 \mathrm{d}\mathbf{x} \, \mathrm{d}t \\
&= \sum_{i=1}^{\widetilde{N}} \sum_{j=1}^{\widetilde{M}} \int_{t^i}^{t^{i+1}} \int_{\tau_j} \left(\frac{1}{\Delta t \Delta \tau} \int_{t^i}^{t^{i+1}} \int_{\tau_j} \left(f(t,\mathbf{x}) - f(s,\mathbf{y}) \right) \, \mathrm{d}\mathbf{y} \, \mathrm{d}s \right)^2 \mathrm{d}\mathbf{x} \, \mathrm{d}t \\
&\leq \frac{\gamma^2}{(\Delta t \Delta \tau)^2} \sum_{i=1}^{\widetilde{N}} \sum_{j=1}^{\widetilde{M}} \int_{t^i}^{t^{i+1}} \int_{\tau_j} \left(\int_{t^i}^{t^{i+1}} \int_{\tau_j} \left(|t-s| + \|\mathbf{x}-\mathbf{y}\| \right) \, \mathrm{d}\mathbf{y} \, \mathrm{d}s \right)^2 \mathrm{d}\mathbf{x} \, \mathrm{d}t \\
&\leq \frac{\gamma^2}{(\Delta t \Delta \tau)^2} \sum_{i=1}^{\widetilde{N}} \sum_{j=1}^{\widetilde{M}} \int_{t^i}^{t^{i+1}} \int_{\tau_j} \left(\int_{t^i}^{t^{i+1}} \int_{\tau_j} \left(\Delta t + h \right) \, \mathrm{d}\mathbf{y} \, \mathrm{d}s \right)^2 \mathrm{d}\mathbf{x} \, \mathrm{d}t \\
&= \gamma (\Delta t + h)^2 \sum_{i=1}^{\widetilde{N}} \sum_{j=1}^{\widetilde{M}} \int_{t^i}^{t^{i+1}} \int_{\tau_j} \mathrm{d}\mathbf{x} \, \mathrm{d}t \leq 2T\gamma^2 \left((\Delta t)^2 + h^2 \right).
\end{aligned}
$$

□

9.2 The model nonlinear SPDE and its finite element approximation

In this section we discuss the formulation of the nonlinear stochastic PDE and its finite element approximation. Let $D = D_1 \times D_2$ be a bounded region in $\mathbb{R}^2$ whose boundary is denoted ∂D; let T denote a positive constant. Denote $\mathbf{x} = (x_1, x_2) \in D$ such that $x_1 \in D_1$ and $x_2 \in D_2$. Let ϵ denote the given (constant) perturbation parameter that measures the amplitude of the space-time white noise $\dfrac{\partial^3 \mathbf{W}}{\partial t \partial x_1 \partial x_2}(t, \mathbf{x}; \omega)$. Furthermore, let g be a deterministic function. We consider the model nonlinear stochastic partial differential equation on a bounded interval $I := (0, T) \subset \mathbb{R}$ as follows: find $u : \bar{I} \times D \to \mathbb{R}$ such that

$$
\begin{cases}
\dfrac{\partial u}{\partial t}(t,\mathbf{x}) - \Delta u(t,\mathbf{x}) + \gamma N(u(t,\mathbf{x})) = g(t,\mathbf{x}) + \epsilon \dfrac{\partial^3 \mathbf{W}}{\partial t \partial x_1 \partial x_2}(t,\mathbf{x};\omega), \text{ in } I \times D \\
u(t,\mathbf{x}) = 0, \text{ on } \partial D \\
u(0,\mathbf{x}) = 0, \text{ in } D,
\end{cases}
\tag{9.2.1}
$$

where $N : \bar{I} \times D \to \mathbb{R}$ is a nonlinear function of $u(t,\mathbf{x})$ and γ denotes the strength of the nonlinearity. Setting $\epsilon = 0$ reduces (9.2.1) to a deterministic PDE for which standard reduced-order methods apply. Similarly, setting $\gamma = 0$ reduces (9.2.1) to a linear stochastic PDE.

Generally, when discussing the stochastic process $u(t,\mathbf{x};\omega)$, we will omit the explicit

dependence on the probability space Ω, treating each realization as a deterministic PDE, unless it will be useful for clarity's sake to make this dependence explicit.

9.2.1 The nonlinear Monte Carlo finite element method

For our first attempt at a ROM for nonlinear SPDEs we decided to apply the Monte Carlo Finite Element Method as outlined in Section 4.2. Our aim is to approximate the integral $E\left(u(\cdot\,;\omega)\right)$ numerically by sample averages of iid realizations corresponding to sample white noise functions. In fact, all the statistics we are interested in are really integrals of a function $u(\cdot\,;\omega)$ over Ω.

We begin with a variational formulation, to define a finite element method suitable for producing approximate solutions of (9.2.1); of course, many other methods could also be used. A variational formulation of (9.2.1) is the following: find $u(t,\cdot)$, $t \in I$, such that

$$\begin{cases} \displaystyle\int_D u(t,\mathbf{x})\phi(\mathbf{x})\ \mathrm{d}D + \int_0^T \int_D \{\nabla u \cdot \nabla \phi\}\ \mathrm{d}D\ \mathrm{d}s \\ \qquad\qquad\qquad\qquad + \gamma \displaystyle\int_0^T \int_D N(u(s,\mathbf{x}))\phi(\mathbf{x})\ \mathrm{d}D\ \mathrm{d}s \\ \qquad\qquad\qquad\qquad = \displaystyle\int_0^T \int_D g(s,\mathbf{x})\phi(\mathbf{x})\ \mathrm{d}D\ \mathrm{d}s + \epsilon \int_0^T \int_D \phi(\mathbf{x})\ \mathrm{d}\mathbf{W}(s,\mathbf{x}), \\ \qquad\qquad\qquad u(0,\mathbf{x}) = 0, \text{ in } D. \end{cases} \tag{9.2.2}$$

for all $\phi \in H_0^1(D)$. The last integral in (9.2.2) is understood in the Itô sense.

The semi-discretization in space leads to the following problem: find $u(t,\cdot) \in H_0^1(D)$, $t \in I$, such that

$$(u_t, \phi(\mathbf{x}))_{L^2(D)} + \mathscr{B}(u,\phi) = \mathscr{F}(\phi) \quad \text{ for all } \phi \in H_0^1(D) \tag{9.2.3}$$

where

$$\begin{aligned} \mathscr{B}(u,\phi) &= \int_D \nabla u \cdot \nabla \phi\ \mathrm{d}D + \gamma \int_D N(u(t,\mathbf{x}))\phi(\mathbf{x})\ \mathrm{d}D \\ &= (\nabla u, \nabla \phi)_{L^2(D)} + (N(u),\phi)_{L^2(D)} \end{aligned} \tag{9.2.4}$$

and

$$\begin{aligned} \mathscr{F}(\phi) &= \int_D \left(g(t,\mathbf{x}) + \epsilon \frac{\partial^3 \widehat{\mathbf{W}}}{\partial t \partial x_1 \partial x_2}(t,\mathbf{x};\omega) \right) \phi(\mathbf{x})\ \mathrm{d}D \\ &= (g,\phi)_{L^2(D)} + \left(\Delta \widehat{\mathbf{W}}, \phi\right)_{L^2(D)} \end{aligned} \tag{9.2.5}$$

respectively. We note that $u_t := \frac{\partial u}{\partial t}$.

For each realization of the piecewise constant noise $\Delta\widehat{\mathbf{W}}$, a realization of the approximate solution is computed, using a time-marching scheme such as the standard backward Euler-Galerkin finite element method. Linearization is accomplished by a Newton or Quasi-Newton approach. Here it will be helpful to write the solution as $u(\cdot\,;\omega)$ to emphasize the dependence on the probability space $(\Omega, \mathscr{F}, P)$. The challenge is that we do not know $u(\cdot\,;\omega)$, as this remains an unknown quantity of interest. Hence, we formulate a solution procedure that does not require knowledge of a functional form for $u(\cdot\,;\omega)$ for all $\omega \in \Omega$.

The Monte Carlo method for the problem (9.0.1) is defined as follows:

1. Choose the number of realizations $M \in \mathbb{N}_+$ and a piecewise continuous finite element approximating space of D, S_0^h.

2. For each $k = 1, \ldots, M$, sample iid realizations of the piecewise constant noise $\Delta\widehat{\mathbf{W}}(\cdot\,;\omega_k)$ and produce approximations $u^h(\cdot\,;\omega_k) \in S_0^h$ such that

$$\begin{aligned}&(u_t^h(\cdot\,;\omega_k), \phi^h)_{L^2(D)} + \widetilde{\mathscr{B}}(u^h(\cdot\,;\omega_k), \phi^h)\\ &= \widetilde{\mathscr{F}}(\phi^h) = \Big(g(\cdot) + \epsilon\Delta\widehat{\mathbf{W}}(\cdot\,;\omega_k), \phi^h\Big)_{L^2(D)} \qquad \forall \phi^h \in S_0^h\end{aligned}$$

   where $\widetilde{\mathscr{B}}$ and $\widetilde{\mathscr{F}}$ are the stochastic forms of (9.2.4) and (9.2.5) respectively, defined by

$$\forall \omega \in \Omega \,:\, \widetilde{\mathscr{B}}(\phi_1, \phi_2) = \mathscr{B}(\phi_1, \phi_2; \omega) \quad \forall \phi_1, \phi_2 \in H_0^1(D)$$

   and

$$\widetilde{\mathscr{F}}(\phi_2) = \mathscr{F}(\phi_2; \omega) \quad \forall \phi_2 \in H_0^1(D).$$

   Once we have fixed $\omega = \omega_k$, the problem is completely deterministic, and may be solved by standard methods.

3. Approximate $E[u](\mathbf{x})$ by the sampling average:

$$\overline{E}(u^h; M) = \frac{1}{M}\sum_{k=1}^{M} u^h(\cdot\,;\omega_k) \tag{9.2.6}$$

We only consider the case where S_0^h is fixed for all realizations, in particular, we have fixed the triangulation $\{\tau_j\}$ of D and employ a backward-Euler scheme for the time discretization.

9.3 A Reduced-order model for the nonlinear SPDE

In this section, we briefly describe the reduced-order model (ROM) for the nonlinear SPDE (9.2.1). In Section 9.4, we will use a concrete example to exhibit the relative accuracy and efficiency of the ROM compared to the standard FEM approach. The generation of the reduced-order model requires a set of snapshot solution vectors; see section 9.4.1 for a discussion of snapshots.

9.3.1 POD reduced-order bases

Given a discrete set of snapshot vectors $W = \{\vec{w}_n\}_{n=1}^N$ belonging to $\mathbb{R}^J$, we form the $J \times N$ snapshot matrix $\mathbb{A}$ whose columns are the snapshot vectors $\vec{w}_n$:

$$\mathbb{A} = \big(\vec{w}_1 \ \vec{w}_2 \ \cdots \ \vec{w}_N\big).$$

Let

$$\mathbb{U}^T \mathbb{A} \mathbb{V} = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix},$$

where $\mathbb{U}$ and $\mathbb{V}$ are $J \times J$ and $N \times N$ orthogonal matrices, respectively, and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_{\widetilde{M}})$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\widetilde{M}}$ be the singular value decomposition of $\mathbb{A}$. Here, $\widetilde{M}$ is the rank of $\mathbb{A}$, i.e., the dimension of the snapshot set W, which would be less than N whenever the snapshot set is linearly dependent. It can easily be shown that if

$$\mathbb{U} = \big(\vec{\Phi}_1 \ \vec{\Phi}_2 \ \cdots \ \vec{\Phi}_J\big) \qquad \text{and} \qquad \mathbb{V} = \big(\vec{\psi}_1 \ \vec{\psi}_2 \ \cdots \ \vec{\psi}_N\big),$$

then

$$\mathbb{A}\vec{\psi}_i = \sigma_i \vec{\Phi}_i \qquad \text{and} \qquad \mathbb{A}^T \vec{\Phi}_i = \sigma_i \vec{\psi}_i \qquad \text{for } i = 1, \ldots, \widetilde{M}$$

so that also

$$\mathbb{A}^T \mathbb{A} \vec{\psi}_i = \sigma_i^2 \vec{\psi}_i \qquad \text{and} \qquad \mathbb{A}\mathbb{A}^T \vec{\Phi}_i = \sigma_i^2 \vec{\Phi}_i \qquad \text{for } i = 1, \ldots, \widetilde{M}$$

so that σ_i^2, $i = 1, \ldots, \widetilde{M}$, are the nonzero eigenvalues of $\mathbb{A}^T\mathbb{A}$ (and also of $\mathbb{A}\mathbb{A}^T$) arranged in nondecreasing order. The matrix $\mathbb{C} = \mathbb{A}^T\mathbb{A}$ is simply the correlation matrix of the snapshot vectors $W = \{\vec{w}_n\}_{n=1}^N$, i.e., we have that $\mathbb{C}_{mn} = \vec{w}_m^T \vec{w}_n$.

In the reduced-order modeling context, given a set of snapshots $W = \{\vec{w}_n\}_{n=1}^N$ belonging to $\mathbb{R}^J$, the POD reduced basis of dimension $K \leq N < J$ is the set $\{\vec{\Phi}_k\}_{k=1}^K$ of vectors also

belonging to $\mathbb{R}^J$ consisting of the first K left singular vectors of the snapshot matrix $\mathbb{A}$. Thus, one can determine the POD basis by computing the (partial) singular value decomposition of the $J \times N$ matrix $\mathbb{A}$. Alternately, one can compute the (partial) eigensystem $\{\sigma_k^2, \vec{\psi}_i\}_{i=1}^K$ of the $N \times N$ correlation matrix $\mathbb{C} = \mathbb{A}^T\mathbb{A}$ and then set $\vec{\Phi}_k = \frac{1}{\sigma_k}\mathbb{A}\vec{\psi}_k$, $k = 1, \ldots, K$.

The K-dimensional POD basis has the obvious property of orthonormality. It also has several other important properties which we now mention. Let $\{\vec{s}_k\}_{k=1}^K$ be an arbitrary set of K orthonormal vectors in $\mathbb{R}^J$ and let $\Pi\vec{w}$ denote the projection of a vector $\vec{w} \in \mathbb{R}^J$ onto the subspace spanned by that set. Further, let

$$\mathcal{E}(\vec{s}_1, \ldots, \vec{s}_K) = \sum_{n=1}^{N} |\vec{w}_n - \Pi\vec{w}_n|^2,$$

i.e., $\mathcal{E}$ is the sum of the squares of the error between each snapshot vector $\vec{w}_n$ and its projection $\Pi\vec{w}_n$ onto the span of $\{\vec{s}_k\}_{k=1}^K$. Then, it can be shown that

$$\begin{cases} \textit{the POD basis } \{\vec{\Phi}_k\}_{k=1}^K \textit{ minimizes } \mathcal{E} \textit{ over all possible} \\ K\textit{-dimensional orthonormal sets in } \mathbb{R}^J. \end{cases} \tag{9.3.1}$$

In fact, the POD basis corresponding to a set of snapshots $W = \{\vec{w}_n\}_{n=1}^N$ is often defined by (9.3.1), and then its relation to the singular value decomposition of the matrix $\mathbb{A}$, or to the eigenvalue decomposition of $\mathbb{A}^T\mathbb{A}$, are derived properties. We note that $\mathcal{E}(\vec{\Phi}_1, \ldots, \vec{\Phi}_K)$ is referred to as the "POD energy" or "error in the POD basis." Also, it can be shown that

$$\mathcal{E}(\vec{\Phi}_1, \ldots, \vec{\Phi}_K) = \sum_{k=K+1}^{\widetilde{M}} \sigma_k^2, \tag{9.3.2}$$

i.e., the error in the POD basis is simply the sum of the squares of the singular values corresponding to the neglected POD modes.

Another property of the POD basis is given as follows:

$$\begin{cases} \textit{the POD basis } \{\vec{\Phi}_k\}_{k=1}^K \textit{ solves the sequence of problems:} \\ \textit{for } k = 1, \ldots, K, \qquad \max_{\vec{s}_k \in \mathbb{R}^J} \sum_{n=1}^{N} \left(\vec{w}_n^T \vec{s}_k\right)^2 \\ \textit{subject to } |\vec{s}_k| = 1 \textit{ and } \vec{s}_j^T \vec{s}_k = 0 \textit{ for } j = 1, \ldots, k-1. \end{cases} \tag{9.3.3}$$

Again, (9.3.3) is often used to define the POD basis and then its relation to the singular value decomposition and (9.3.1) are noted as derived properties.

The singular values of the snapshot matrix may be used to determine a practical value for the dimension K of the POD basis. Indeed, it is a simple matter to show that if one requires the error in the POD basis to be less than some prescribed tolerance δ, i.e., that

$$\mathcal{E}(\vec{\Phi}_1, \ldots, \vec{\Phi}_K) \leq \delta,$$

then one need only

$$\begin{array}{ll} \textit{choose } K \textit{ to be the smallest} & \dfrac{\sum_{k=1}^{K} \sigma_k^2}{\sum_{k=1}^{\widetilde{M}} \sigma_k^2} \geq 1 - \delta. \\ \textit{integer such that} & \end{array}$$

The usefulness of POD-based reduced-order modeling is derived from the observation that in many settings one finds that, even if δ is chosen to be relatively small, e.g., 0.01, one can still be able to use a basis of low order K; K is usually much smaller than $\widetilde{M}$ and might be of order 10 or so.

For reduced-order modeling applications, the snapshot vectors are coefficient vectors in the expansion of the finite element approximation of the stochastic process evaluated at different instants in time. Thus, to each snapshot vector $\vec{w}_n$, $n = 1, \ldots, N$, there corresponds a finite element function

$$w_n(\mathbf{x}) = \sum_{j=1}^{J} \mathbf{w}_{j,n} \phi_j(\mathbf{x}) \in S_0^h, \tag{9.3.4}$$

where $\mathbf{w}_{j,n}$ denotes the j-th component of the vector $\vec{w}_n$ and $\phi_j(\mathbf{x}) \in V_0^h$ denotes the j-th finite element basis function. One can define a POD basis with respect to functions instead of vectors, i.e., we could start with a snapshot set $W = \{w_n(\mathbf{x})\}_{n=1}^N$ consisting of finite element functions belonging to S_0^h. Then, instead of (9.3.3), one could define the POD basis $\{\Phi_k(\mathbf{x}) \in V_0^h\}_{k=1}^K$ to be the solution of the sequence of problems: *for* $k = 1, \ldots, K$,

$$\max_{s_k(\mathbf{x}) \in V_0^h} \sum_{n=1}^{N} \langle w_n, s_k \rangle_0^2$$

subject to $\|s_k(\mathbf{x})\|_0 = 1$ *and* $\langle s_j, s_k \rangle_0 = 0$ *for* $j = 1, \ldots, k-1$. Note that $\langle \cdot\,, \cdot \rangle_0 = \langle \cdot\,, \cdot \rangle_{L^2(D)}$. Equivalently, one could define the POD basis to be the solution of the problem: *minimize*

$$\mathcal{E}(s_1, \ldots, s_K) = \sum_{n=1}^{N} \|w_n - \Pi w_n\|_0^2$$

over all possible K*-dimensional* $L^2(D)$*-orthonormal sets* $\{s_k(\mathbf{x})\}_{k=1}^K$ *in* V_0^h, where Πw_n is the $L^2(D)$-projection of w_n onto the span of the functions $\{s_k(\mathbf{x})\}_{k=1}^K$. Moreover, one can instead

determine the POD basis by first solving the $N \times N$ eigenvalue problem: for $k = 1, \ldots, \widetilde{M}$,

$$\mathbb{C}\vec{a}_k = \sigma_k^2 \vec{a}_k, \qquad |\vec{a}_k| = 1, \qquad \vec{a}_\ell^T \vec{a}_k = 0 \text{ if } k \neq \ell, \quad \text{and} \quad \sigma_k \geq \sigma_{k-1} > 0, \tag{9.3.5}$$

then setting

$$\Phi_k(\mathbf{x}) = \sum_{n=1}^{N} \frac{1}{\sigma_k} a_{k,n} w_n(\mathbf{x}) \qquad \text{for } k = 1, \ldots, K.$$

Here, we have that the rank $\widetilde{M} \leq N$ correlation matrix $\mathbb{C}$ is defined by $\mathbb{C}_{mn} = \langle w_m, w_n \rangle_0$, and $a_{k,n}$ is the n-th component of the eigenvector $\vec{a}_k$. Note that in terms of the snapshot matrix $\mathbb{A}$ and the mass matrix $\mathbb{M}$ for the finite element basis, i.e., for $\mathbb{M}_{ij} = (\Phi_i, \Phi_j)_0$, we now have that $\mathbb{C} = \mathbb{A}^T \mathbb{M} \mathbb{A}$. This fact allows us to again use the singular value decomposition to determine the POD basis function. To this end, let $\mathbb{M} = \mathbb{S}^T \mathbb{S}$, where the $J \times J$ matrix $\mathbb{S}$ may be chosen to be a symmetric, positive definite square root of $\mathbb{M}$, i.e., $\mathbb{S} = \mathbb{M}^{1/2}$, or $\mathbb{S}$ could be a Cholesky factor, i.e., $\mathbb{S}^T = \mathbb{L}$. Then, we let $\widetilde{\mathbb{A}} = \mathbb{S}\mathbb{A}$ so that $\mathbb{C} = \mathbb{A}^T \mathbb{M} \mathbb{A} = \widetilde{\mathbb{A}}^T \widetilde{\mathbb{A}}$ and therefore $\vec{a}_k$, $k = 1, \ldots, K$, are the first K right singular vectors of $\widetilde{\mathbb{A}}$.

9.3.2 The POD reduced-order models

We now show how a POD basis is used to define a reduced-order model for the nonlinear stochastic PDE (9.2.1). For the sake of brevity, we only discuss the case for which the snapshot set is viewed as a set of finite element coefficient vectors; the case for which the snapshot set is a set of finite element functions proceeds in similar manner.

Let $\{\vec{\Phi}_k\}_{k=1}^K$ be a K-dimensional POD basis corresponding to the snapshot set $\{\vec{w}_n\}_{n=1}^N$. For each $\vec{\Phi}_k$, $k = 1, \ldots, K$, there is a finite element function

$$\Phi_k(\mathbf{x}) = \sum_{j=1}^{J} \mathbf{\Phi}_{j,k} \phi_j(\mathbf{x}) \in S_0^h, \tag{9.3.6}$$

where $\mathbf{\Phi}_{j,k}$ denotes the jth component of $\vec{\Phi}_k$. Let

$$U_K = \text{span}\{\Phi_k\}_{i=1}^K \subset S_0^h.$$

As will be explained in Section 9.4.1, the reduced basis functions satisfy homogeneous boundary conditions. We then seek a reduced basis approximation of the stochastic process field of the form

$$u^K(t, \cdot) \in U_K.$$

We determine $u^K(t,\cdot)$, $t \in I$, from the discrete problem

$$\begin{cases} \displaystyle\int_D \frac{\partial u^K}{\partial t}\Phi(\mathbf{x})\ \mathrm{d}D + \int_D \{\nabla u^K \cdot \nabla \Phi\}\ \mathrm{d}D \\ \qquad\qquad\qquad\qquad + \gamma \displaystyle\int_D N(u^K(s,\mathbf{x}))\Phi(\mathbf{x})\ \mathrm{d}D \\ \qquad\qquad\qquad\qquad = \displaystyle\int_D \left(g(t,\mathbf{x}) + \epsilon \frac{\partial^3 \widehat{\mathbf{W}}}{\partial t \partial x_1 \partial x_2}(t,\mathbf{x};\omega) \right) \Phi(\mathbf{x})\ \mathrm{d}D \\ \qquad\qquad u^K(0,\mathbf{x}) = 0, \text{ in } \mathcal{D} \quad \forall \Phi(\mathbf{x}) \in U_K. \end{cases} \tag{9.3.7}$$

The reduced basis approximation of the stochastic process takes the form

$$u^K(t,\cdot) = \sum_{k=1}^{K} \alpha_k(t;\omega)\Phi_k$$

and (9.3.7) may be expressed as

$$\begin{cases} \displaystyle\sum_{k=1}^{K} \frac{d\alpha_k}{dt} \int_D \Phi_k(\mathbf{x})\Phi_j(\mathbf{x})\ \mathrm{d}D + \sum_{k=1}^{K} \alpha_k(t) \int_D \nabla\Phi_k \nabla\Phi_j\ \mathrm{d}D \\ \qquad\qquad\qquad\qquad + \gamma \displaystyle\int_D N\left(\sum_{k=1}^{K} \alpha_k(t)\Phi_k(\mathbf{x}) \right) \Phi_j\ \mathrm{d}D \\ \qquad\qquad\qquad\qquad = \displaystyle\int_D \left(g(t,\mathbf{x}) + \epsilon \frac{\partial^3 \widehat{\mathbf{W}}}{\partial t \partial x_1 \partial x_2}(t,\mathbf{x};\omega) \right) \Phi_j(\mathbf{x})\ \mathrm{d}D \\ \qquad \displaystyle\sum_{k=1}^{K} \alpha_k(0;\omega)\Phi_k(\mathbf{x}) = 0, \text{ in } D, \end{cases} \tag{9.3.8}$$

for $j = 1, \ldots, K$. Equivalently, we have the system of nonlinear ordinary differential equations that determine the coefficient functions $\{\alpha_k(t;\omega)\}_{k=1}^K$:

$$\begin{cases} \mathbb{G}\dfrac{d}{dt}\vec{\alpha}(t;\omega) + \widetilde{\mathbb{G}}\vec{\alpha}(t;\omega) + \vec{N}\left(\vec{\alpha}(t;\omega)\right) = \vec{f}(t;\omega) \\ \vec{\alpha}(0;\omega) = \vec{\alpha}_0, \end{cases} \tag{9.3.9}$$

where the Gram matrix $\mathbb{G}$, stiffness matrix $\widetilde{\mathbb{G}}$, nonlinear vector function $\vec{N}\left(\vec{\alpha}(t;\omega)\right)$, and solution vector $\vec{\alpha}(t;\omega)$ are respectively given by

$$\mathbb{G}_{jk} = \int_D \Phi_k(\mathbf{x})\Phi_j(\mathbf{x})\ \mathrm{d}D, \qquad \widetilde{\mathbb{G}}_{jk} = \int_D \nabla\Phi_k\nabla\Phi_j\ \mathrm{d}D,$$

$$\vec{N}\left(\vec{\alpha}(t;\omega)\right) = \gamma \int_D N\left(\sum_{k=1}^{K} \alpha_k(t;\omega)\Phi_k(\mathbf{x}) \right) \Phi_j\ \mathrm{d}D \qquad \text{and} \qquad (\vec{\alpha})_k = \alpha_k(t;\omega)$$

for $j, k = 1, \ldots, K$, while the forcing vector $\vec{f}(t;\omega)$ and initial data vector $\vec{\alpha}_0$ are respectively given by

$$(\vec{f})_j = \int_D \left(g(t,\mathbf{x}) + \epsilon \frac{\partial^3 \widehat{\mathbf{W}}}{\partial t \partial x_1 \partial x_2}(t,\mathbf{x};\omega) \right) \Phi_j(\mathbf{x}) \, \mathrm{d}D$$

and

$$(\vec{\alpha}_0)_j = 0$$

for $j = 1, \ldots, K$. Unlike the matrices in a standard FEM formulation, these matrices are not sparse; but their order K is so much smaller (see Section 9.4) that their density is of no consequence. Another important observation is that matrices $\mathbb{G}$, $\widetilde{\mathbb{G}}$, depend only on the reduced basis functions $\{\Phi_k\}_{k=1}^K$ so that they may all be pre-computed.

9.3.3 The error in a reduced-order solution

At any given time t, we define the "error" $\mathfrak{E}(t)$ in a POD reduced-order solution to be the $L^2(D)$-norm of the difference between the *expected value* of the full finite element solution and the *expected value* of the reduced-order solution, i.e.,

$$\begin{aligned}
\mathfrak{E}(t) &= \left\{ \int_D \left(\int_\Omega u^h(t,\mathbf{x};\omega) P(\,\mathrm{d}\omega) - \int_\Omega u^K(t,\mathbf{x};\omega) P(\,\mathrm{d}\omega) \right)^2 \,\mathrm{d}D \right\}^{1/2} \\
&= \left\{ \int_D \left(E\left[u^h(t,\mathbf{x};\omega)\right] - E\left[u^K(t,\mathbf{x};\omega)\right] \right)^2 \,\mathrm{d}D \right\}^{1/2} \\
&\approx \left\{ \int_D \left(\overline{E}(u^h;M) - \overline{E}(u^K;M) \right)^2 \,\mathrm{d}D \right\}^{1/2}, \qquad (9.3.10)
\end{aligned}$$

where $u^h(t,\mathbf{x};\omega)$ denotes the approximate stochastic process determined using the full finite element simulation code, and $u^K(t,\mathbf{x};\omega)$ denotes the approximate stochastic process determined by a POD reduced-order model. The expected value $E(\cdot)$ is defined in Section 9.1. Using the standard MCFEM described in section 9.2.1, we approximate $E(u^h)$ by sampling averages $\overline{E}(u^h;M)$ where M is the number of sample realizations. We can use a similar Monte Carlo reduced-order modeling technique to approximate $E(u^K)$ by $\overline{E}(u^K;M)$ for the same number of sample realizations. Also of interest is the space-time error

$$\begin{aligned}
\mathfrak{E}_T = \left\{ \int_0^T \mathfrak{E}^2(s) \,\mathrm{d}s \right\}^{1/2} &= \left\{ \int_0^T \int_D \left(E\left[u^h(t,\mathbf{x};\omega)\right] - E\left[u^K(t,\mathbf{x};\omega)\right] \right)^2 \,\mathrm{d}D \,\mathrm{d}s \right\}^{1/2} \\
&\approx \left\{ \int_0^T \int_D \left(\overline{E}(u^h;M) - \overline{E}(u^K;M) \right)^2 \,\mathrm{d}D \,\mathrm{d}s \right\}^{1/2}. \qquad (9.3.11)
\end{aligned}$$

There are two contributions to these "errors." One is due to the fact that the reduced-order model does not exactly reproduce the information contained in the snapshot set; the other to the fact that the snapshot set itself cannot exactly represent the full finite element solution.

Recall that the POD reduced bases are determined from a set of snapshots and that those bases are designed so that *most* of the information in the snapshot set is captured. But because we truncate the singular value decomposition quite early, it is clear that even if the only solutions encountered were linear combinations of snapshot vectors, there would be many such solutions which would lie partially or entirely outside the span of the reduced basis.

Even if a reduced basis could exactly capture all the information in the snapshot set, the errors (9.3.10) and (9.3.11) would not vanish because the snapshot set itself cannot exactly capture all the behaviors allowed in the full finite element space.

A snapshot cannot exactly represent even the finite element solutions used in its construction because that set consists of a time-sampling of those solutions. Also, a snapshot set represents only a discrete set of sampled values of any system parameters or boundary conditions. Strong nonlinearities in those effects can reveal that a snapshot set is inadequate for a given problem.

The analyst must therefore always keep in mind the limitations of the snapshot approach. On the other hand, it is easy to automate, and generally quite practical. It has been demonstrated many times in the literature that POD-based reduced-order models are excellent at exploiting the information contained in a snapshot set; the computational experiments reported in Section 9.4 provide one such example. Thus, a key to designing reduced-order models of the POD type is to ensure that the snapshot set contains sufficient information for the problems to be considered.

9.4 Computational experiments

We compare the use, efficiency, and accuracy of the POD-based reduced-order modeling techniques on an example. We consider a nonlinear stochastic PDE in which a perturbation parameter ϵ controls the intensity of the space-time white noise. The problem has a relatively strong cubic nonlinearity, in which the strength of the nonlinear function $N(u(t,\mathbf{x}) = u(1-u)^2$ is $\gamma = 10$. The nonlinear stochastic PDE is rewritten as follows: find $u : \bar{I} \times D \to \mathbb{R}$

such that

$$\begin{cases} \dfrac{\partial u}{\partial t}(t,\mathbf{x}) - \Delta u(t,\mathbf{x}) + 10\, u(t,\mathbf{x})(1-u(t,\mathbf{x}))^2 \\ \qquad = e^t \sin(x_1)\cos(x_2) + \epsilon \dfrac{\partial^3 \widehat{\mathbf{W}}}{\partial t \partial x_1 \partial x_2}(t,\mathbf{x};\omega), \text{ in } I \times D \\ u(t,\mathbf{x}) = 0, \text{ on } \partial D \\ u(0,\mathbf{x}) = 0, \text{ in } D, \end{cases} \tag{9.4.1}$$

where $\mathbf{x} = (x_1, x_2)$ and the deterministic forcing term $g(t,\mathbf{x}) = e^t \sin(x_1)\cos(x_2)$.

The Galerkin finite element method on a square grid of $J = 1089$ nodes is used in the discrete weak formulation, described in section 9.2.1, to obtain accurate stochastic Galerkin finite element approximations of solutions of (9.4.1). This choice of spatial dimension is arbitrary; however for this initial study, a relatively coarse grid was chosen to allow us to compute multiple realizations of the stochastic example problem (9.4.1). The time derivative is discretized by a backward-Euler method. Newton's method is used to solve the fully discrete nonlinear stochastic PDE. Finite element solutions are used for the generation of snapshots and later for comparison with POD-based reduced-order solutions.

The snapshot solutions are generated by the finite element method applied to problem (9.4.1) for a particular number of realizations; each realization samples the piecewise constant white noise $\Delta\widehat{\mathbf{W}}(\cdot\,;\omega)$ with a specific perturbation parameter ϵ whose assignment is to be discussed next.

9.4.1 Generating snapshots

For all simulations involved in the snapshot generation process, ϵ is chosen to have the value 0.1. The snapshot vectors are determined by the following procedure. First, for $k = 1, \ldots, \widetilde{K} = 10$ realizations and $t \in I = (0,1)$, a finite element approximation is determined $\sum_{j=1}^{J} w(t;\omega_k)\phi(\mathbf{x})$ of the solution of (9.4.1), where J denotes the dimension of the finite element space S_0^h, and $\{\phi_j\}_{j=1}^{J}$ denotes the basis functions for that space. This is the approximate stochastic process used to generate the snapshots. For $S = 200$ and $P = 10$, the $N = S \times P = 2000$ snapshot vectors

$$\vec{w}_n = \vec{w}_{sp} = \begin{pmatrix} w_1(t_s;\omega_p) \\ w_2(t_s;\omega_p) \\ \vdots \\ w_J(t_s;\omega_p) \end{pmatrix}, \quad \text{for } s = 1, \ldots, S, \text{ and } p = 1, \ldots, P \tag{9.4.2}$$

are then determined for P realizations of (9.4.1). For each realization $p = 1, \ldots, P$, the solution is started impulsively, and evaluated at 200 equally spaced time values t_s, $s = 1, \ldots, 200$, ranging from $t = 0$ to $T = 1$, for a total of 2000 snapshots. The time interval used for sampling snapshots is the same used for the time discretization of the nonlinear SPDE (9.4.1). The snapshot vectors $\{\vec{w}_n\}_{n=1}^N$ correspond to the snapshot functions

$$\mathbf{w}_n(\mathbf{x}) = \mathbf{w}_{sp}(\mathbf{x}) = \sum_{j=1}^{J} w_j(t_s; \omega_p)\phi_j(\mathbf{x}) \quad \text{for } n = 1, \ldots, N.$$

It is sometimes convenient to modify the snapshot set $\{\vec{w}_n\}_{n=1}^N$ to satisfy certain boundary conditions. In this case, the snapshot vectors naturally satisfy the Dirichlet homogeneous boundary conditions. It might also seem relevant to compute the sample average of the P realizations and use the resulting S snapshot vectors rather than the complete N snapshot vectors in the reduced-order model. In the next section, it is explained why this approach was not employed.

9.4.2 POD reduced bases

POD reduced bases corresponding to the snapshot set $\{\vec{w}_n\}_{n=1}^N$ are determined as described in Section 9.3.1. As described in the previous section, each POD basis function satisfies zero Dirichlet boundary conditions on the boundary δD. The elements of a K-dimensional POD basis constitute the first K elements of all POD bases of dimension greater that K. For the snapshot set determined as described in Section 9.4.1, the 16-dimensional POD basis functions are displayed in Figure 9.3 and the first 16 singular values of the corresponding

Table 9.1: The first 16 singular values of the snapshot matrix for the example problem (9.4.1).

1	34.0387	5	8.8536	9	6.4498	13	5.5122
2	13.0477	6	8.4762	10	6.1230	14	4.8172
3	11.7138	7	7.3131	11	5.8013	15	4.7471
4	9.7154	8	7.1571	12	5.5966	16	4.6712

snapshot matrix are listed in Table 9.1. In the context of POD-based reduced-order modeling, one expects the singular values to decrease rapidly, indicating that a low-dimension POD basis can capture most of the information in the snapshot set. This behavior cannot, of course, be universal, but it has been observed in many examples. In our particular problem

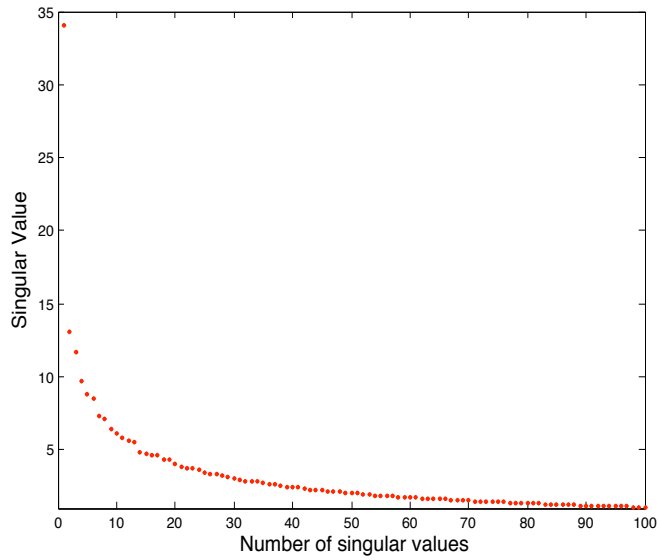

Figure 9.2: The first 100 singular values of the snapshot matrix for the example problem (9.4.1).

we observe from Table 9.1 and Figure 9.2, that the singular values of the corresponding snapshot matrix decay rather slowly to zero. However, we shall show in section 9.4.5 that the POD-based reduced-order models for the nonlinear SPDE still perform adequately for the approximation of ensemble averages of finite element solutions, rather than for the approximate finite element solutions themselves; this is the usual case for POD-basis reduced-order models. That is, a "small-dimensional" POD basis can still be sufficient if our primary interest is approximating the expected value of finite element solutions, and not the approximation of a particular realization.

Note that if one computes the sample average of the P realizations and uses the resulting S snapshot vectors rather than the complete N snapshot vectors, the resulting singular values decay at the same rate. Therefore, the smaller snapshot set does not provide any extra information or improvement over our reduced-order calculation.

9.4.3 Determining POD-based reduced-order approximations

Given a perturbation parameter ϵ and a particular $\omega \in \Omega$, the K-dimensional system of nonlinear ordinary differential equations (9.3.9) is used to determine reduced-order solutions of the example problem (9.4.1). Approximate solutions of the system of ordinary differential equations (9.3.9) can be determined using an ODE solver of appropriate order. The

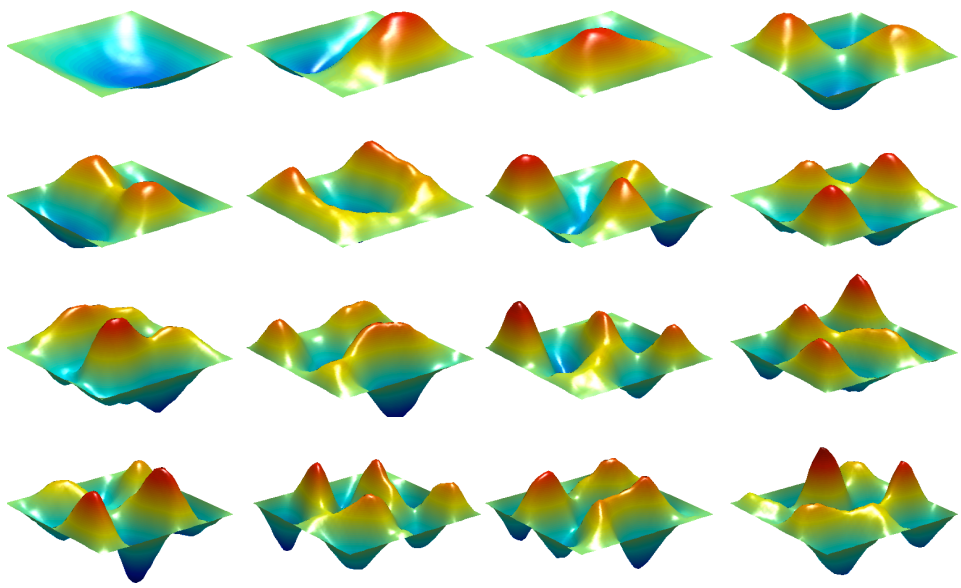

Figure 9.3: The POD reduced basis of cardinality 16 for the nonlinear stochastic problem 9.4.1.

computations described here use the finite element type assembly, described in section 9.2, for the considerably smaller K-dimensional system.

The vectors and matrices appearing in the system (9.3.9) depend only on the choice of reduced basis, so that once a POD basis is determined as described in section 9.4.2 it can be reused for different choices of ϵ and ω. This is a major advantage of POD-based reduced-order models, where the primary interest is the approximation of the expected value $E[u^h(\cdot\,;\omega)](\mathbf{x})$. Similar to the MCFEM described in section 9.2.1, we again use a Monte Carlo sampling method for the reduced-order model. That is, given the number of realizations $M \in \mathbb{N}_+$ we sample iid realizations of the piecewise constant white noise $\Delta\widehat{\mathbf{W}}(\cdot\,;\omega_k)$ for $k = 1, \ldots, M$ and find $u^K(\cdot\,;\omega)$ satisfying (9.3.7). The quantity $E[u^K(\cdot\,;\omega)](\mathbf{x})$ is then approximated by the sampling average:

$$\overline{E}(u^K; M) = \frac{1}{M}\sum_{k=1}^{M} u^K(\cdot\,;\omega_k). \tag{9.4.3}$$

To illustrate the effectiveness of the low-dimensional, POD-based reduced-order models, we employ several perturbation parameters ϵ and several different choices for the dimension of the reduced bases. Note that the set $\{\omega_p\}_{p=1}^{P}$ used to generate the snapshot vectors $\{\vec{w}_n\}_{n=1}^{N}$ is distinct from the set $\{\omega_k\}_{k=1}^{M}$ used to calculate the sample average (9.4.3).

9.4.4 Perturbation parameters for POD-based reduced-order models

The specific choices used for ϵ in the reduced-order simulation are given as follows. For $M = 1000$ realizations, we compute the sampling average $\overline{E}(u^K; M)$, described by (9.4.3), using several different cases: $\epsilon = 0, 0.01, 0.05, 0.1, 0.5, 1$. In each case, the reduced-order solution u^K satisfies (9.3.7).

It is worth noting something about the six perturbation parameters ϵ used in these computational experiments: For all cases except $\epsilon = 0.1$, the perturbation parameter is different from the parameter used to generate the snapshots. This is, of course, how one wants to use a reduced-order model: generate a reduced basis using snapshots determined from finite element simulations for some specific choice of data for the example problem (9.4.1), and then solve the reduced-order model for a variety of different data. For those cases, the reduced-order simulations are carried out over the time interval $[0, 1]$, which is the same time interval used to generate the snapshots.

9.4.5 Computational results

For all six cases, full finite element solutions employing thousands of unknowns are determined so that they may be compared to the POD-based reduced-order solutions for the same data. Specifically, for the comparisons, we use the measures $\mathfrak{E}(t)$ and $\mathfrak{E}_T$ defined in (9.3.10) and (9.3.11), respectively.

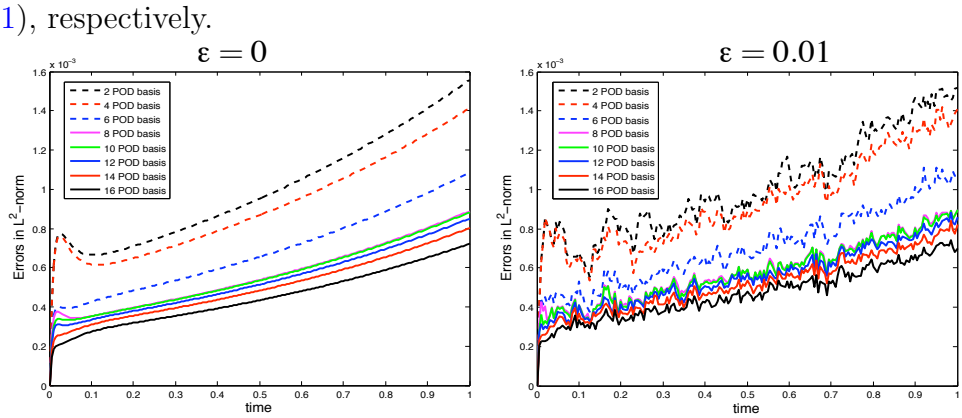

Figure 9.4: $\mathfrak{E}(t)$ for the POD-based reduced-order models with $\epsilon = 0$ and $\epsilon = 0.1$. The plots are on the same scale indicated the effects on the error by introducing the perturbation parameter ϵ into the ROM.

In Figure 9.6, plots of $\mathfrak{E}(t)$ versus the time t are provided for each of the six test problems

described in Section 9.4.4. In each plot, results are displayed for eight values of the POD basis K, ranging from 2 to 16. Each plot has been scaled separately, to exhibit the drop in error as K increases. To consider the effect of ϵ, we also provide, in Figure 9.4, a panel of plots drawn to the same scale, for $\epsilon = 0$ and $\epsilon = 0.01$. A comparison of these plots demonstrates the effect of introducing noise into the system (9.0.1), while also suggesting how well the mean of the reduced-order solution still approximates the mean of the finite element solution of the deterministic PDE. Both plots indicate that the error grows with ϵ but can be reduced by increasing K. Obviously, for a given K, by increasing ϵ, the approximation error $\mathfrak{E}(t)$ will generally be larger.

For the example problem (9.4.1), an examination of Figure 9.6 shows that very low-dimensional POD-based reduced-order models are quite effective at approximating the expected value of the full finite element solutions; even for bases of dimension less than 10, the "error" $\mathfrak{E}(t)$ is small. As the perturbation parameter increases, though, a larger number of basis vectors are required to control the error $\mathfrak{E}(t)$. It should be noted that for small perturbation parameters, $0 < \epsilon \leq 0.01$, the error $\mathfrak{E}(t)$ follows the error for the case $\epsilon = 0$, with the obvious stochastic effects present. The results also suggest that, for this problem, increasing the order of the reduced basis beyond about $K = 10$ produced very little further improvement in accuracy. This conclusion can also be inferred from Table 9.2 where the space-time "error" $\mathfrak{E}_T$ versus K is listed. There is a much slower reduction in the error as K increases from 10 to 16 than the change from 2 to 10.

Table 9.2: $\mathfrak{E}_T$ for the POD-based reduced-order models vs. the dimension K of the reduced basis space for the six test cases.

K	$\epsilon = 0$	$\epsilon = 0.01$	$\epsilon = 0.05$	$\epsilon = 0.1$	$\epsilon = 0.5$	$\epsilon = 1$
2	3.17e-02	3.19e-02	3.90e-02	5.07e-02	1.25e-01	2.08e-01
4	3.03e-02	3.06-02	3.72e-02	4.79e-02	1.15e-01	1.89e-01
6	2.61e-02	2.66e-01	3.44e-02	4.54e-02	1.09e-01	1.76e-01
8	2.37e-02	2.42e-02	3.25e-02	4.34e-02	1.04e-01	1.67e-01
10	2.36e-02	2.41e-02	3.18e-02	4.22e-02	9.99e-02	1.60e-01
12	2.31e-02	2.36e-02	3.10e-02	4.11e-02	9.96e-02	1.53e-01
14	2.24e-02	2.29e-02	3.01e-02	4.00e-02	9.37e-02	1.48e-01
16	2.12e-02	2.17e-02	2.93e-02	3.91e-02	9.13e-02	1.44e-01

Finally, Figure 9.6 and Table 9.2 show that POD-based reduced-order models have the ability to approximate the expected value of full finite element solutions by solving

a very small dimensional system. This is also evident from observing Figure 9.5 which plots the approximate expected value $\overline{E}(u^h)$ using the full FEM versus the approximate expected value $\overline{E}(u^K)$ using the reduced-order model with 2, 8 and 16 POD vectors at times $t = 0, 0.25, 0.5, 0.75$ and 1. From this plot we see that as the dimension K increases from 2 to 16 the quality of the approximation improves dramatically.

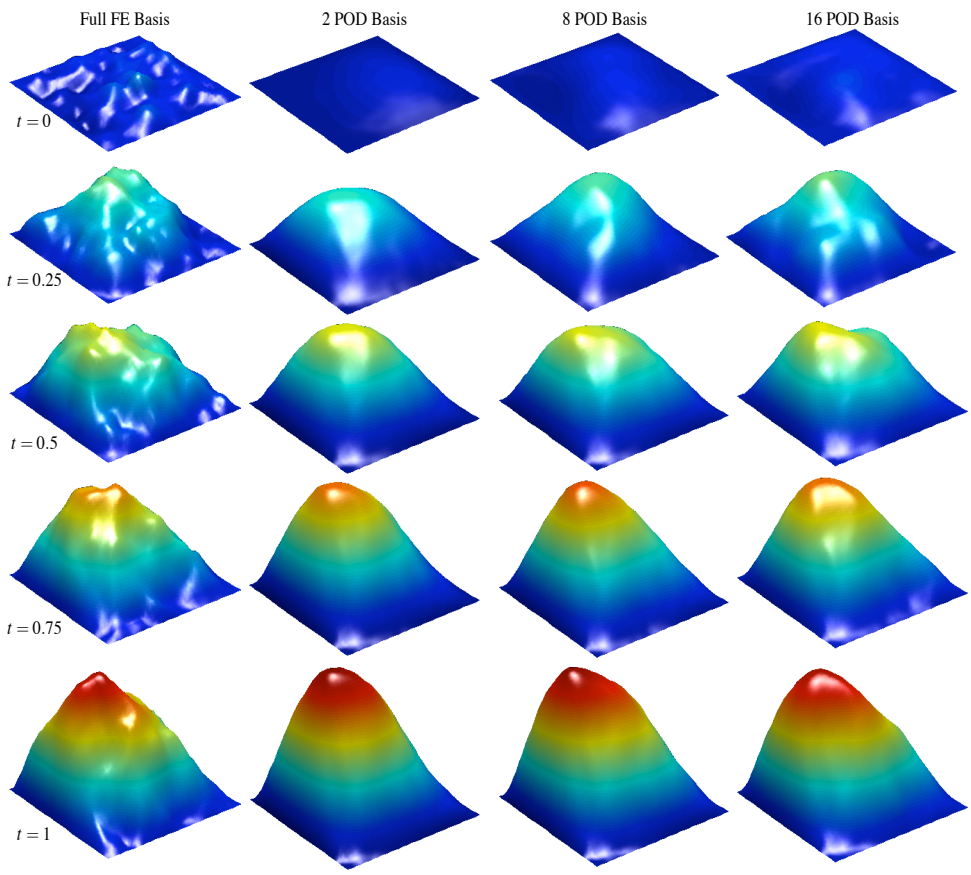

Figure 9.5: The approximate expected value $\overline{E}(u^h)$, using the full FEM versus the approximate expected value $\overline{E}(u^K)$ using the ROM with 2, 8 and 16 POD vectors at times $t = 0, 0.25, 0.5, 0.75$ and 1. All plots are generated for the case $\epsilon = 0.1$.

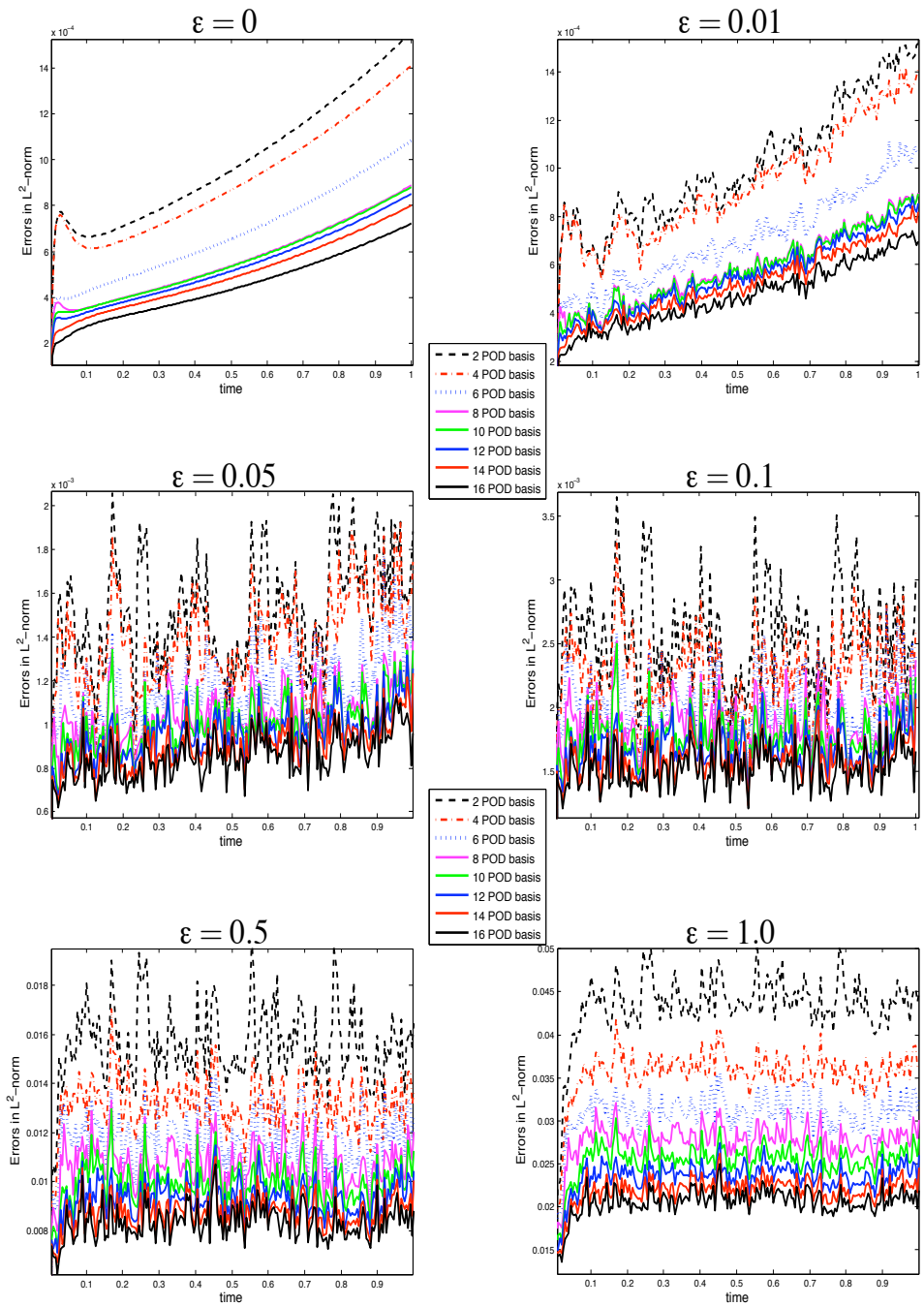

Figure 9.6: $\mathfrak{E}(t)$ for the POD-based reduced-order models for Cases 1-6. As the perturbation parameter ϵ increases from 0 to 1.0, the scale of the plots is adjusted to clearly display the error behavior.

9.5 Summary

We saw in section 9.3.1 that snapshots can be viewed either as finite element coefficient vectors or as finite element functions. The net effect of taking the latter view is the appearance of the finite element mass matrix $\mathbb{M}$ in, e.g., (9.3.5). Although handling the appearance of the mass matrix does not add a significant cost to the determination of a reduced basis, it also does not appreciably improve the performance of the reduced-order model. For this reason, in section 9.4.5, we only provided results based on viewing snapshots as coefficient vectors.

The results given in Section 9.4.5 suggest that POD-based reduced-order modeling can be very effective in approximating the expected value of solutions of a nonlinear stochastic PDE driven by space-time additive white noise. It is therefore appropriate to briefly examine the cost of these POD-based reduced-order models. If one settles on a reduced-order modeling technique that relies on the generation of snapshots, then one only needs to discuss the production of a reduced-order basis when comparing the cost of such a procedure. That is, regardless of the type of reduced-order model employed, one must certainly "pay the price" of computing a snapshot set. The expense of extracting the reduced order basis from the snapshot set is minuscule compared to the cost of generating the snapshot set itself.

POD methods have attracted much research interest in recent years as an inexpensive means of simulating deterministic PDEs. To our knowledge, little work has been done on extending this method to the field of stochastic PDEs, and yet this is a field of study which seems ideal for this approach, given the need to consider a vast ensemble of solutions. We hope that this work invites further inquiry by others into a relatively open area.

There are many worthwhile problems to consider in both linear and nonlinear stochastic PDEs. For instance, using a Karhunen-Loève representation of the input data, described by Example 1.2.2, provides greater correlation and information concerning the random fields. Studies of these cases may lead to a better understanding of the errors that occur when approximating ensemble averages of solutions by a reduced-order model. More importantly, the coupling of this ROM technique with the sparse collocation methods described in Chapters 1 - 7 will be of critical importance since our eventual goal is to solve realistic engineering applications. Such investigations will increase the understanding, interest and application of these methods for the analysis of the stochastic dynamics of linear and

nonlinear problems of practical interest (Deang et al., 2001; Faris and Jona-Lasino, 1982; Sancho et al., 1998; Shardlow, 2005, 2003, 2004).

CHAPTER 10

CONCLUDING REMARKS AND FUTURE RESEARCH

The major contribution of this work was the development and analysis of a novel weighted Smolyak method, describing an optimal choice for the weight parameters. These weights adaptively tune the anisotropy of the method for each given problem. The methodology behind this approach can be based both on *a priori* or *a posteriori* information and is motivated by the regularity of the solution and the error estimates derived in this work.

The method proposed here can be viewed as a natural extension of our previous works; the full tensor Stochastic Collocation method described in Babuška et al. (2005c) and the isotropic Sparse Grid Stochastic Collocation method Nobile et al. (2006).

The new technique consists of a Galerkin approximation in physical space and an anisotropic collocation in probability space, at the zeros of sparse tensor product polynomials utilizing either Clenshaw-Curtis or Gaussian knots. As a consequence of the collocation approach our techniques naturally lead to the solution of uncoupled deterministic problems that are trivially parallelizable, as in the Monte Carlo method.

The error estimates derived in this work predict a rate of convergence that is at least algebraic with respect to the total number of collocation points. Similarly to the Monte Carlo method, the number of collocation points is directly proportional to the computational work required by the algorithm. Furthermore, the derived estimates provide a glimpse into future directions in this area. We observe that for problems possessing coefficients that are truncated expansions with proper decay in each probability direction, there may not be a deterioration in the exponent of the algebraic convergence with respect to the truncation dimension. This is a substantial advantage when comparing to the previous works; see Babuška et al. (2005c) and Nobile et al. (2006).

The numerical examples included in this work gave computational ground to the theoretical results and suggest that the actual convergence rate of the method may be superior to the prediction yielded by the current error estimates. The numerical results include a comparison of the new weighted Smolyak method with the methods from Babuška et al. (2005c), Nobile et al. (2006) and the standard Monte Carlo. In particular, for moderately large dimensional problems, the anisotropic sparse grid approach seems to be very efficient and superior to all examined methods.

Some of the immediate research directions and possible future areas of study we would like to pursue include topics that are extensions of my current efforts, while others are areas we are interested in and look forward to exploring. Each of these areas are discussed below.

- **Regularity of solutions to nonlinear SPDEs:** For certain classes of problems, the solution may have very regular dependence on input random variables. Rigorous analysis of solution regularity with respect to the random variables is fundamental since it dictates the rate of convergence. For instance, it was shown in Babuška et al. (2005c) and Babuška et al. (2004b) that the solution of a linear elliptic PDE with diffusivity coefficient and/or forcing term described as truncated expansions of random fields is analytic in the input random variables. In Chapter 3 we provided a similar proof, under mild assumptions for a nonlinear elliptic problem with a quadratic nonlinearity.

This is an area of extreme interest to us and we intend to pursue several varying approaches in the future, to prove this result in general. In particular, we are optimistic that borrowing techniques from abstract nonlinear parabolic theory will assist in validating this result. For example, the works Lunardi (1987, 2004); Escher and Simonett (2003) prove that the solution to a fully nonlinear parabolic problem admits an analytic extension in some region of the complex plane with respect to the parameter t (time). In this setting we can think of the parameter t as the noise parameter $y_n \in \Gamma_n = [-1, 1]$ and we can attempt to extend these analyticity results to our particular problem, described by 3.4.2.

Broadening these abstract theoretical results will open the door to many new applications. In particular, the work of Takáč et al. (1996) proved the analyticity of essentially bounded solutions to semilinear parabolic systems with applications to the Ginzburg-Landau equations. It is this direction that we predict this area of study to be headed. This thesis is just a beginning to many realistic engineering applications which we intend to devote much

future research.

• **Parameter identification in nonlinear SPDEs:** An important inverse problem in applications (biochemistry, physics, bioengineering) is to fit a system of nonlinear SPDEs to observed data, possibly corrupted by random measuring errors. Let the given data be a set of measurements $\{\eta(t,x,\omega); t \in (0,T), x \in D, \omega \in \Omega\}$ (or a discrete set, even a finite number of snapshots in time, or on a small portion D° of the physical domain D), and suppose that these data are to be modeled by a system of equations of the form

$$\frac{\partial y}{\partial t}(t,x,\omega) = f\Big(t,x,\omega,y(t,x,\omega),\nabla y(t,x,\omega),\Delta y(t,x,\omega);\boldsymbol{\beta}(t,x,\omega)\Big) \tag{10.0.1}$$

with some boundary and initial conditions, where $\boldsymbol{\beta}$ is the parameter to be determined by fitting. In most cases, the system of SPDEs cannot be solved in closed form, so it is necessary to combine a numerical solver with the parameter identification process. We seek $\boldsymbol{\beta}$ so that the corresponding computed solutions $y_{\boldsymbol{\beta}}$ match the measured data η.

The method that we use is a nonlinear least square approach, i.e. we minimize the expected value

$$\mathscr{L}(\boldsymbol{\beta}) = \mathbb{E}\left[\int_{(0,T)\times D}\Big(y_{\boldsymbol{\beta}}(t,x,\omega) - \eta(t,x,\omega)\Big)^2 dxdt\right],$$

with respect to $\boldsymbol{\beta}$, and subject to the restriction $y_{\boldsymbol{\beta}}$ being a solution of (10.0.1). We treat this as a target-tracking optimal control problem, where we will extend our work (Gunzburger et al., 2007) to prove the existence of a global minimizer and use adjoint-based first order necessary conditions for optimality to write a gradient type numerical algorithm to find the parameter $\boldsymbol{\beta}$ that corresponds to the η measurements.

• **Combining deterministic ROMs with Sparse Grid probabilistic techniques :** There are many worthwhile problems to consider in both linear and nonlinear stochastic PDEs. For instance, rather than the white noise approach in Burkardt et al. (2007) described in Chapter 9, using a Karhunen-Loève representation of the random fields discussed in our earlier work (Nobile et al., 2006, 2007; Babuška et al., 2007), provides greater correlation and information concerning the random fields. Studies of these cases may lead to a better understanding of the errors that occur when approximating ensemble averages of solutions by a ROM. More importantly, the coupling of this reduced-order

modeling technique with the sparse collocation methods described in Sections 4.4.2 and 4.4.3 will be of critical importance, since our eventual goal is to solve realistic engineering applications. Such investigations will increase the understanding, interest and application of these methods for the analysis of the stochastic dynamics of linear and nonlinear problems of practical interest.

- **Extending our SPDE techniques to solve complicated applications:** In the future we would like to focus on the development of mathematical theory and efficient numerical algorithms to produce reliable numerical approximations for a wide variety of applied problems. In particular, it has been my goal since first investigating SPDEs to research Multiscale Methods for Fluid-Structure Interaction, where the interaction between distinct regions is random or uncertain. An example application of this type of algorithm is modeling blood-flow through a pulsating heart, and the blood endocardial wall interactions involved.

  The fluid mechanics formulation will be based on multiscale and stabilized finite element methods for the incompressible Navier-Stokes equations with random coefficients and source terms. It is in this setting that we intend to couple my existing stochastic techniques. We already have insight into the regularity of solutions to the stochastic Navier Stokes equations with respect to the noise parameterization. Along with the fact that our stochastic techniques are "embarrassingly parallel" and we can reuse deterministic codes, we are extremely confident that we can expect analytical and computational successes in this area.

- **Verification and Validation (V&V):** The main aim of our stochastic simulations is to derive some predictions, which could be the basis for decision making. A major question is how reliable these predictions are. The areas of Verification and Validation address the question above. In particular, "Validation" relates to the reliability of the mathematical model (a completely formulated mathematical problem) which is solved numerically, while "Verification" relates to the quality of a numerical solution to the given mathematical model. The Verification step is purely a mathematical problem, while the Validation and prediction steps pose much more complex problems.

  Recently, there have been many interesting publications attempting to improve the effectiveness of V&V, all revealing the need to perform nondeterministic computational simulations in comparison with experimental data. It is this parallel that exists to my

own efforts that we believe must be addressed in order for V&V to be more effective in improving confidence in computational predictive capability. We are eager to explore this area further.

In closing, we firmly believe that our research in several areas of applied mathematics, in particular the cutting-edge field of SPDEs and their numerical solutions, will undeniably generate both intellectual interest on the part of our colleagues and fellow researchers in the field, and monetary support as an extension of the current widespread appeal this type of work has invariably gained. Many of the areas in which we have devoted our attention have been recently explored however, there still are vast uncharted areas in these fields where we feel our techniques would be adept at conquering. We are anxious to get started.

REFERENCES

Ainsworth, M. and J.-T. Oden, 2000: *A posteriori Error Estimation in Finite Element Analysis*. Wiley. 1

Allen, E. J., S. J. Novosel, and Z. Zhang, 1998: Finite element and difference approximation of some linear stochastic partial differential equations. *Stochastics Rep.*, **64**, 117–142. 114, 118, 119, 120

Alós, E. and S. Bonaccorsi, 2002: Stochastic partial differential equations with dirichlet white-noise boundary conditions. *Ann. I. H. Poincaré*, **38(2)**, 152–154. 114

Asokan Badri Narayanan, V. and N. Zabaras, 2005: Variational multiscale stabilized FEM formulations for transport equations: stochastic advection-diffusion and incompressible stochastic Navier-Stokes equations. *J. Comput. Phys.*, **202(1)**, 94–133. 3

Aubry, N., W. Lian, and E. Titi, 1993: Preserving symmetries in the proper orthogonal decomposition. *SIAM J. Sci. Comp.*, **14**, 483–505. 116

Babuška, I., 1961: On randomized solutions of Laplace's equation. *Casopis Pest. Mat.*, **86**, 269–276. 44

Babuška, I. and P. Chatzipantelidis, 2002a: On solving elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, **191(37-38)**, 4093–4122. 7

Babuška, I. and J. Chleboun, 2002: Effects of uncertainties in the domain on the solution of Neumann boundary value problems in two spatial dimensions. *Math. Comp.*, **71(240)**, 1339–1370 (electronic). 1

—, 2003: Effects of uncertainties in the domain on the solution of Dirichlet boundary value problems. *Numer. Math.*, **93(4)**, 583–610. 1

Babuška, I. and K.-M. Liu, 2001: On solving stochastic initial-value differential equations. *Mathematical Models and Methods in Applied Sciences*. 114

Babuška, I., F. Nobile, and R. Tempone, 2005a: A stochastic collocation method for elliptic partial differential equations with random input data. Tech. rep., ICES report 05-47, University of Texas at Austin, submitted. 46, 47, 51

Babuška, I., F. Nobile, R. Tempone, and C. Webster, 2007: A stochastic collocation method for nonlinear elliptic partial differential equations with random input data. In progress. 13, 58, 143, 160

Babuška, I. and J.-T. Oden, 2005: The reliability of computer predictions: Can they be trusted? ICES report 05. Submitted to Int. J. of Num. Anal. and Modeling. 3

Babuška, I., R. Tempone, and G. Zouraris, 2002: Galerkin finite element approximations of stochastic elliptic partial differential equations. Tech. rep., ICES report 02, University of Texas at Austin. 47, 49

—, 2005b: Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Comput. Methods Appl. Mech. Engrg.*, **194**, 1251–1294. 114

Babuška, I. M. and P. Chatzipantelidis, 2002b: On solving elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, **191(37-38)**, 4093–4122. 43

Babuška, I. M., K. M. Liu, and R. Tempone, 2003: Solving stochastic partial differential equations based on the experimental data. *Math. Models Methods Appl. Sci.*, **13(3)**, 415–444, dedicated to Jim Douglas, Jr. on the occasion of his 75th birthday. 5

Babuška, I. M., F. Nobile, and R. Tempone, 2005c: A stochastic collocation method for elliptic partial differential equations with random input data. Tech. rep., MOX, Dipartimento di Matematica. 7, 13, 16, 17, 19, 31, 40, 41, 59, 60, 62, 66, 68, 81, 103, 111, 112, 141, 142

Babuška, I. M. and T. Strouboulis, 2001: *The finite element method and its reliability*. Numerical Mathematics and Scientific Computation, The Clarendon Press Oxford University Press, New York, ISBN 0-19-850276-1. 1

Babuška, I. M., R. Tempone, and G. Zouraris, 2004a: Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numeric. Anal.*, **42(2)**, 800–825. 114

Babuška, I. M., R. Tempone, and G. E. Zouraris, 2004b: Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, **42(2)**, 800–825 (electronic). 2, 7, 16, 44, 45, 47, 49, 50, 97, 111, 142

—, 2005d: Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Comput. Methods Appl. Mech. Engrg.*, **194(12-16)**, 1251–1294. 2, 7, 45, 49, 50, 94, 100, 101, 112

Babuška, I. M., F. Nobile, and R. Tempone, 2005: Worst-case scenario analysis for elliptic problems with uncertainty. *Numer. Math.*, **101**, 185–219. 2

Badri Narayanan, V. A. and N. Zabaras, 2004: Stochastic inverse heat conduction using a spectral approach. *Internat. J. Numer. Methods Engrg.*, **60(9)**, 1569–1593. 2

Barthelmann, V., E. Novak, and K. Ritter, 2000a: High dimensional polynomial interpolation on sparse grids. *Adv. Comput. Math.*, **12(4)**, 273–288, multivariate polynomial interpolation. 9, 42, 53

—, 2000b: High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics*, **12**, 273–288. 69

Beck, J. and J. Ching, 2004: Bayesian analysis of the phase ii iascasce structural health monitoring experimental benchmark data. *J. Engrg. Mech.*, **130**, 1233–. 3

Beck, J. L. and S. K. Au, 2002: Bayesian updating of structural models and reliability using markov chain monte carlo simulation. *J. Engrg. Mech.*, **128**, 380–391. 3

Ben-Haim, Y., 1996: *Robust reliability in the mechanical sciences*. Springer. 2

Benth, F. E. and J. Gjerde, 1998a: Convergence rates for finite element approximations of stochastic partial differential equations. *Stochastics Stochastics Rep.*, **63(3-4)**, 313–326. 2, 44, 45

—, 1998b: Numerical solution of the pressure equation for fluid flow in a stochastic medium. In *Stochastic analysis and related topics, VI (Geilo, 1996)*, Birkhäuser Boston, Boston, MA, vol. 42 of *Progr. Probab.*, pp. 175–186. 2

Berkooz, G., P. Holmes, and J. Lumley, 1993: The proper orthogonal decomposition in the analysis of turbulent flows. *Ann. Rev. Fluid. Mech.*, **25**, 539–575. 116

Berkooz, G. and E. Titi, 1993: Galerkin projections and the proper orthogonal decomposition for equivariant equations. *Phys. Let. A*, **174**, 94–102. 116

Bernardini, A., 1999: *What are the random fuzzy sets and how to use them for uncertainty modeling in engineering systems, in "Whys and Hows in Uncertainty Modelling Probability Fuzziness and Anti-Optimization", I. Elishakoff, editor*, Springer–Verlag. No. 338 in CISM Course and Lectures, pp. 63–125. 3

Box, G. E. P., 1973: *Bayesian inference in statistical analysis.*. Wiley and Sons, Reading, Mass. 3

Braack, M. and A. Ern, 2003: A posteriori control of modeling errors and discretization errors. *Multiscale Model. Simul.*, **1(2)**, 221–238 (electronic). 1

Brenner, S. C. and L. R. Scott, 1994: *The Mathematical Theory of Finite Element Methods*. Springer–Verlag, New York. 47, 66

Burkardt, J., M. Gunzburger, and H.-C. Lee, 2003: Centroidal Voronoi tessellation-based reduced-order modeling of complex systems. Tech. rep., Florida State University. 115, 116

—, 2004: Pod and cvt-based reduced-order modeling of Navier-Stokes flows. Tech. rep., Florida State University. 115, 116

Burkardt, J., M. Gunzburger, and C. Webster, 2007: Reduced order modeling of some nonlinear stochastic partial differential equations. *International Journal of Numerical Analysis and Modeling*, **4(3-4)**, 368–391. 101, 113, 143, 160

Caflisch, R. E., 1-49, Cambridge Univ. Press, 1998: Monte Carlo and quasi-Monte Carlo methods. *Acta numerica*. 44

Chang, J. S., P. Looy, and C. Webster, 1998: The effects of dust loadings on the collections of fine particles by an electrostatic precipitator with dc or pulse energized prechargers. *Journal of Aerosol Science*, **29**, S1127–S1128. 160

Ciarlet, P. G., 1978: *The Finite Element Method for Elliptic Problems*. North-Holland, New York. 66

Clenshaw, C. W. and A. R. Curtis, 1960: A method for numerical integration on an automatic computer. *Numer. Math.*, **2**, 197–205. 57

Cullen, A. C. and H. C. Frey, 1999: *Probabilistic Techniques Exposure Assessment*. Plenum Press. 2

Dalang, R. and C. Mueller, 2003: Some nonlinear spde's that are second order in tiime. *Electronic Journal of Probability*, **8(1)**, 1–21. 115

Davie, A. M. and J. G. Gaines, 2000: Convergence of numerical schemes for the solution of parabolic stochastic partial differential equatioins. *Mathematics of Computation*, **70(233)**, 121–134. 114

Deane, A., I. Kevrekidis, G. Karniadakis, and S. Orszag, 1991: Low-dimensional models for complex geometry flows: applications to grooved channels and circular cylinders. *Phys. Fluids A*, **3**, 2337–2354. 116

Deang, J., Q. Du, and M. Gunzburger, 2001: Thermal fluctuations of superconducting vortices. *Phy. Rev. B.*, **64**, 52506—52510. 140

Deb, M.-K., 2000: *Solution of stochastic partial differential equations (SPDEs) using Galerkin method: theory and applications*. Ph.D. thesis, The University of Texas at Austin. 44

Deb, M. K., I. M. Babuška, and J. T. Oden, 2001a: Solution of stochastic partial differential equations using Galerkin finite element. *Comput. Methods Appl. Mech. Engrg. 2*, **190**, 6359–6372. 114

—, 2001b: Solution of stochastic partial differential equations using Galerkin finite element techniques. *Comput. Methods Appl. Mech. Engrg.*, **190(48)**, 6359–6372. 44, 49

DeVore, R. A. and G. G. Lorentz, 1993: *Constructive approximation*, vol. 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, ISBN 3-540-50627-6. 59

di Bruno, C. F. F., 1855: Sullo sviluppo delle funzione. *Ann. di Scienze Matem. et Fisiche di Tortoloni*, **6**, 479–480. 32

Du, Q. and T. Zhang, 2002: Numerical approximation of some linear stochastic partial differential equations driven by special additive noises. *SIAM J. Numer. Anal.*, **40(4)**, 1421–1445. 115, 119

Dubois, D. and H. Prade, eds., 2000: *Fundamentals of fuzzy sets*, vol. 7 of *Handbook of fuzzy sets*. Kluwer. 3

Durrett, R., 1996: *Probability: Theory and Examples*. Duxbury Press, Second ed., ISBN 0-534-24318-5. 44

Dzjadyk, V. K. and V. V. Ivanov, 1983: On asymptotics and estimates for the uniform norms of the Lagrange interpolation polynomials corresponding to the chebyshev nodal points. *Analysis Mathematica*, **9(11)**, 85–97. 69

Elisakoff, I. and Y. Ren, 2003: *Finite element methods for structures with large variations*. Oxford Univ. Press. 2

Elishakoff, I., ed., 1999: *Whys and Hows in Uncertainty Modelling: Probability, Fuzziness and Anti-Optimization*. No. 388 in CISM International Centre for Mechanical Sciences, Springer–Verlag. 2

Elman, H. C., O. G. Ernst, D. P. O'Leary, and M. Stewart, 2005: Efficient iterative algorithms for the stochastic finite element method with application to acoustic scattering. *Comput. Methods Appl. Mech. Engrg.*, **194(9-11)**, 1037–1055. 50

Erdös, P. and P. Turan, 1937: On interpolation. I. Quadrature- and mean-convergence in the Lagrange-interpolation. *Ann. of Math. (2)*, **38(1)**, 142–155. 76

Eriksson, K., D. Estep, P. Hansbo, and C. Johnson, 1995: Introduction to computational methods for differential equations. In *Theory and numerics of ordinary and partial differential equations (Leicester, 1994)*, Oxford Univ. Press, New York, Adv. Numer. Anal., IV, pp. 77–122. 1

Escher, J. and G. Simonett, 2003: Analyticity of solutions to fully nonlinear parabolic evolution equations on symmetric spaces. *J. Evol. Equ.*, **3(4)**, 549–576, dedicated to Philippe Bénilan. 41, 142

Evans, L., 1998: *Partial Differential Equations*, vol. 19 of *Graduate Studies in Mathematics*. AMS. 22

Faris, W. G. and G. Jona-Lasino, 1982: Large fluctuations for a nonlinear heat equation with noise. *J. Phys. A:Math Gen.*, **18**, 3025–3055. 140

Ferson, S., V. Kreinovich, L. Ginzburg, D. Mayers, and K. Sentz, 2003: Constructing probability boxed and Demster-Shafer structures. Sandia Report SAND 2002-4015, Sandia National Laboratories. 3

Fishman, G., 1996a: *Monte Carlo*. Springer Series in Operations Research, Springer-Verlag, New York, ISBN 0-387-94527-X, concepts, algorithms, and applications. 6

—, 1996b: *Monte Carlo: Concepts, Algorithms, Applications*. Springer–Verlag. 43

Frauenfelder, P., C. Schwab, and R. A. Todor, 2005a: Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Engrg.*, **194(2-5)**, 205–228. 2, 44, 45, 49

—, 2005b: Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Engrg.*, **194(2-5)**, 205–228. 5, 7, 9, 15, 20, 42, 80, 97

Ganapathysubramanian, B. and N. Zabaras, 2006: Sparse grid collocation schemes for stochastic natural convection problems. Tech. rep., Cornell. 7, 8

Garcke, J. and M. Griebel, 2002: Classification with anisotropic sparse grids using simplicial basis functions. *Intelligent Data Analysis*, **6**, 483–502. 55

Gaudagnini, A. and S. Neumann, 1999: Nonlocal and localized analysis of conditional mean steady state flow in bounded, randomly nonuniform domains. 1. Theory and computational approach. 2. Computational examples. *Water Resources Research*, **35(10)**, 2999–3039. 7

Gerstner, T. and M. Griebel, 1998a: Numerical integration using sparse grids. *Numer. Algorithms*, **18(3-4)**, 209–232. 9, 42, 57

—, 1998b: Numerical integration using sparse grids. Unpublished. 53

—, 2003: Dimension-adaptive tensor-product quadrature. *Computing*, **71**, 65–87. 55

Ghanem, R., 1999: Ingredients for a general purpose stochastic finite elements implementation. *Comput. Methods Appl. Mech. Engrg.*, **168(1-4)**, 19–34. 7

Ghanem, R. and J. Red-Horse, 1999: Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach. *Phys. D*, **133(1-4)**, 137–144, predictability: quantifying uncertainty in models of complex phenomena (Los Alamos, NM, 1998). 2, 44, 45

Ghanem, R. G. and P. D. Spanos, 1991: *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York, ISBN 0-387-97456-3. 2, 4, 7, 44, 45, 50

Glimm, J., S. Hou, Y.-H. Lee, D. H. Sharp, and K. Ye, 2003: Solution error models for uncertainty quantification. In *Advances in differential equations and mathematical physics (Birmingham, AL, 2002)*, Amer. Math. Soc., Providence, RI, vol. 327 of *Contemp. Math.*, pp. 115–140. 2

Graham, M. and I. Kevrekidis, 1996: Pattern analysis and model reduction: Some alternative approaches to the Karhunen-Lóeve decomposition. *Comp. Chem. Engng.*, **20**, 495–506. 116

Grigoriu, M., 2002: *Stochastic calculus*. Birkhäuser Boston Inc., Boston, MA, ISBN 0-8176-4242-0, applications in science and engineering. 4

Gunzburger, M., C. Trenchea, and C. Webster, 2007: Analysis of an optimal control problem for nonlinear stochastic partial differential equations. In progress. 143, 160

Gyongy, I., 1999: Lattice approximations of stochastic quasi-linear parabolic partial differential equations driven by space-time white noise ii. *Potential Analysis*, **11(1-37)**. 115

Gyongy, I. and T. Martinez, 2003: On the approximation of solutions of stochastic differential equations of elliptic type. Tech. rep., Universidad Autónoma de Madrid. 115

Helton, J. and F. Davis, 2003: Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety*, **81**, 23–69. 6

Helton, J. C., 1997: Analysis in the presence of stochastic and subjective uncertainties. *Journal of Stochastical Computations and Simulations*, **57**, 3–76. 3

Hlaváček, J., I. Chleboun, and I. M. Babuška, 2004: *Uncertain input data problems and the worst scenario method*. Elsevier, Amsterdam. 2

Holmes, P., J. Lumley, G. Berkooz, and P. Holmes, 1996: *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press. 116

Holmes, P., J. Lumley, G. Berkooz, J. Mattingly, and R. Wittenberg, 1997: Low-dimensional models of coherent structures in turbulence. *Phys. Rep.*, **287**, 337–384. 116

Johnson, C., 2000: Adaptive computational methods for differential equations. In *ICIAM 99 (Edinburgh)*, Oxford Univ. Press, Oxford, pp. 96–104. 1

Jouini, E., J. Cvitanić, and M. Musiela, eds., 2001: *Option Pricing, Interest Rates and Risk Management*. Cambridge University Press. 44

Karniadakis, G., C.-H. Su, D. Xiu, D. Lucor, C. Schwab, and R. Todor, 2005: Generalized polynomial chaos solution for differential equations with random inputs. SAM Report 2005-01, ETH. 7

Keese, A. and H. G. Matthies, 2005: Hierarchical parallelisation for the solution of stochastic finite element equations. *Comput. & Structures*, **83(14)**, 1033–1047. 2

Kleiber, M. and T.-D. Hien, 1992: *The Stochastic Finite Element Method*. Wiley. 2, 43

Kramosil, I., 2001: *Probabilistic Analysis of Belief Functions*. Kluwer. 3

Kunisch, K. and S. Volkwein, 1999: Control of burger's equation by a reduced order approach using proper orthogonal decomposition. *JOTA*, **102**, 345–371. 116

—, 2001: Galerkin proper orthogonal decomposition methods for parabolic problems. *Numer. Math.*, **90(1)**, 117–148. 115, 116

Larsen, S., 1986: *Numerical analysis of elliptic partial differential equations with stochastic input data*. Ph.D. thesis, University of Maryland. 44

Le Maître, O. P., O. M. Knio, B. J. Debusschere, H. N. Najm, and R. G. Ghanem, 2003: A multigrid solver for two-dimensional stochastic diffusion equations. *Comput. Methods Appl. Mech. Engrg.*, **192(41-42)**, 4723–4744. 50

Le Maître, O. P., O. M. Knio, H. N. Najm, and R. G. Ghanem, 2004a: Uncertainty propagation using Wiener-Haar expansions. *J. Comput. Phys.*, **197(1)**, 28–57. 2

—, 2004b: Uncertainty propagation using Wiener-Haar expansions. *J. Comput. Phys.*, **197(1)**, 28–57. 7

Le Maître, O. P., H. N. Najm, R. G. Ghanem, and O. M. Knio, 2004c: Multi-resolution analysis of Wiener-type uncertainty propagation schemes. *J. Comput. Phys.*, **197(2)**, 502–531. 2

Lemm, J. C., 2003: *Bayesian field theory*. Johns Hopkins University Press. 3

Loève, M., 1977: *Probability theory*. Springer-Verlag, New York, 4th ed., graduate Texts in Mathematics, Vol. 45 and 46. 4, 14

Lu, Z. and D. Zhang, 2004: A comparative study on uncertainty quantification for flow in randomly heterogeneous media using Monte Carlo simulations and conventional and KL-based moment-equation approaches. *SIAM J. Sci. Comput.*, **26(2)**, 558–577 (electronic). 2, 43

Lucor, D. and G. E. Karniadakis, 2004: Predictability and uncertainty in flow-structure interactions. *Eur. J. Mech. B Fluids*, **23(1)**, 41–49, bluff body wakes and vortex-induced vibrations. 2

Lucor, D., D. Xiu, C.-H. Su, and G. E. Karniadakis, 2003: Predictability and uncertainty in CFD. *Internat. J. Numer. Methods Fluids*, **43(5)**, 483–505, eCCOMAS Computational Fluid Dynamics Conference, Part I (Swansea, 2001). 2

Lumley, J. L., 1970: *Stochastic tools in turbulence*. Academic Press, New York, applied Mathematics and Mechanics, Vol. 12. 116

Lunardi, A., 1987: Time analyticity for solutions of nonlinear abstract parabolic evolution equations. *J. Funct. Anal.*, **71(2)**, 294–308. 41, 142

—, 2004: Nonlinear parabolic equations and systems. In *Evolutionary equations. Vol. I*, North-Holland, Amsterdam, Handb. Differ. Equ., pp. 385–436. 41, 142

Mathelin, L., M. Y. Hussaini, and T. A. Zang, 2005: Stochastic approaches to uncertainty quantification in CFD simulations. *Numer. Algorithms*, **38(1-3)**, 209–236. 3, 7, 51

Matthies, H. G. and A. Keese, 2005: Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, **194(12-16)**, 1295–1331. 2, 4, 7, 45, 50, 97

Melchers, R. E., 1999: *Structural Reliability, Analysis and Prediction*. Wiley. 2

Moon, K.-S., E. von Schwerin, A. Szepessy, and R. Tempone, 2006: An adaptive algorithm for ordinary, stochastic and partial differential equations. In *Recent Advances in Adaptive Computation*, American Mathematical Society, Providence, RI, vol. 381 of *Contemporary Mathematics*, pp. 369–388. 1

Mrczyk, J., ed., 1997: *Computational Mechanics in a Meta Computing Perspective*. Center for Num. Meth. Engrg., Barcelona. 2

Mytnik, L., 2002: Stochastic partial differential equation driven by stable noise. *Probab. Theory Relat. Fields*, **123**, 157–201. 114

Nagy, D., 1979: Modal representation of geometrically nonlinear behavior by the finite element method,. *Comput. &Struct.*, **10(683)**. 115

Niederreiter, H., 1992: *Random number generation and quasi-Monte Carlo methods*, vol. 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, ISBN 0-89871-295-5. 6

Nobile, F., R. Tempone, and C. Webster, 2006: A sparse grid stochastic collocation method for elliptic partial differential equations with random input data. Tech. rep., Technical Report #85, MOX, Dipartimento di Matematica, submitted to SIAM J. Numer. Anal. (May 2006). 5, 7, 8, 15, 51, 55, 58, 59, 66, 81, 89, 91, 92, 94, 97, 100, 111, 141, 142, 143, 160

—, 2007: An anisotropic sparse grid stochastic collocation method for elliptic partial differential equations with random input data. Tech. rep., Technical Report, MOX, Dipartimento di Matematica, submitted to SIAM J. Numer. Anal. (Jan. 2007). 8, 51, 58, 66, 94, 98, 111, 143, 160

Novak, E., 1988: Stochastic properties of quadrature formulas. *Numer. Math.*, **53**, 609–620. 44

Novak, E. and K. Ritter, 1996: High dimensional integration of smooth functions over cubes. *Numer. Math.*, **75**, 79–97. 53

—, 1999: Simple cubature formulas with high polynomial exactness. *Constr. Approx.*, **15**, 499–522. 53

Novak, E., K. Ritter, R. Schmitt, and A. Steinbauer, 1999: On an interpolatory method for high dimensional integration. *Journal of Computational and Applied Mathematics*, **112**, 215–228. 53

Oberkampf, W. L., J. C. Helton, and K. Sentz, 2001: Mathematical representation of uncertainty. Tech. Rep. AIAA 2001-1645, American Institute of Aeronautics and Astronautics. 3

Oden, J. T., I. Babuška, F. Nobile, Y. Feng, and R. Tempone, 2005a: Theory and methodology for estimation and control of errors due to modeling, approximation, and uncertainty. *Comput. Methods Appl. Mech. Engrg.*, **194(2-5)**, 195–204. 1

Oden, J. T., T. Belytschko, B. Ivo, and T. J. R. Hughes, 2003: Research directions in computational mechanics. *Comput. Methods Appl. Mech. Engrg.*, **192(7-8)**, 913–922. 3

Oden, J. T. and S. Prudhomme, 2002: Estimation of modeling error in computational mechanics. *Journal of Computational Physics*, **182**, 496–515. 1

Oden, J. T., S. Prudhomme, and P. Bauman, 2005b: On the extension of goal-oriented error estimation and hierarchical modeling to discrete lattice models. *Comput. Methods Appl. Mech. Engrg.*, **194(34-35)**, 3668–3688. 1

Oden, J. T., S. Prudhomme, D. C. Hammerand, and M. S. Kuczma, 2001: Modeling error and adaptivity in nonlinear continuum mechanics. *Comput. Methods Appl. Mech. Engrg.*, **190(49-50)**, 6663–6684. 1

Oden, J. T. and K. S. Vemaganti, 2000: Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials. I. Error estimates and adaptive algorithms. *J. Comput. Phys.*, **164(1)**, 22–47. 1

Øksendal, B., 2003: *Stochastic differential equations*. Universitext, Springer-Verlag, Berlin, sixth ed., ISBN 3-540-04758-1, an introduction with applications. 15

Park, H. and D. Cho, 1996: Low dimensional modeling of flow reactors. *Int. J. Heat Mass Transf. Int. J. Heat Mass Int. J. Heat Mass Int. J. Heat Mass*, **39**, 3311–3323. 116

Park, H. M. and J. S. Chung, 2002: A sequential method of solving inverse natural convection problems. *Inverse Problems*, **18(3)**, 529–546. 116

Park, H. M. and J. H. Lee, 2000: Solution of an inverse heat transfer problem by means of empirical reduction of modes. *Z. Angew. Math. Phys.*, **51(1)**, 17–38. 116

Park, H. M. and M. W. Lee, 1998: An efficient method of solving the Navier-Stokes equations for flow control. *Internat. J. Numer. Methods Engrg.*, **41(6)**, 1133–1151. 116

Park, H. M. and W. J. Lee, 2002: A new numerical method for the boundary optimal control problems of the heat conduction equation. *Internat. J. Numer. Methods Engrg.*, **53(7)**, 1593–1613. 116

Park, Y. D., H. M.and Jang, 2000: Control of Burgers equation by means of mode reduction. *Internat. J. Engrg. Sci.*, **38(7)**, 785–805. 116

Rathinam, M. and L. R. Petzold, 2002: Dynamic iteration using reduced order models: a method for simulation of large scale modular systems. *SIAM J. Numer. Anal.*, **40(4)**, 1446–1474 (electronic). 116

—, 2003: A new look at proper orthogonal decomposition. *SIAM J. Numer. Anal.*, **41(5)**, 1893–1925 (electronic). 116

Ravindran, S., 2000a: Proper orthogonal decomposition in optimal control of fluids. *Int. J. Numer. Meth. Fluids*, **34**, 425–448. 116

—, 2000b: Reduced-order adaptive controllers for fluid flows using pod. *J. Sci. Comput.*, **15(4)**, 457–478. 116

Reed, M. and B. Simon, 1990: *Methods of Modern Mathematical Physics I: Functional Analysis*. Academic Press. 25

Regan, H. M., S. Ferson, and D. Berleant, 2004: Equivalence of methods for uncertainty propagation of real-valued random variables. *Internat. J. Approx. Reason.*, **36(1)**, 1–30. 2

Riesz, F. and B. Sz-Nagy, 1990: *Functional Analysis*. Dover. 14

Rodríguez, J. and L. Sirovich, 1990: Low-dimensional dynamics for the complex Ginzburg-Landau equations. *Physica D*, **43**, 77–86. 116

Roman, L. and M. Sarkis, 2006: Stochastic Galerkin method for elliptic SPDES: a white noise approach. *DCDS-B Journal*, **6(4)**, 941–955. 3, 7

Romkes, A. and J. T. Oden, 2004: Adaptive modeling of wave propagation in heterogeneous elastic solids. *Comput. Methods Appl. Mech. Engrg.*, **193(6-8)**, 539–559. 1

Rubinstein, R. and M. Choudhari, 2005: Uncertainty quantification for systems with random initial conditions using Wiener-Hermite expansions. *Stud. Appl. Math.*, **114(2)**, 167–188. 2

Sancho, J., J. Garcia-Ojalvo, and H. Guo, 1998: Non-equilibrium Ginzburg-Landau model driven by colored noise,. *Phys. D.*, **113**, 331–337. 140

Schueller, G., ed., 2004: *Special Issue of Comp. Meth. Appl. Mech. Engrg.*, to appear. 2

Schwab, C. and R. Todor, 2007: Stochastic Galerkin fem for the numerical solution of elliptic pdes with stochastic coefficients. In preparation. 40

Schwab, C. and R.-A. Todor, 2002: Finite element method for elliptic problems with stochastic data. Tech. rep., ETH Zurich, Seminar for Applied Mathematics, research Report 2002-05. 44

—, 2003a: Sparse finite elements for elliptic problems with stochastic loading. *Numer. Math.*, **95(4)**, 707–734. 2

Schwab, C. and R. A. Todor, 2003b: Sparse finite elements for stochastic elliptic problems—higher order moments. *Computing*, **71(1)**, 43–63. 2

Shardlow, T., 2003: Weak convergence of a numerical method for a stochastic heat equation. *BIT Numerical Mathematics*, **43**, 179–193. 140

—, 2004: Nucleation of waves in excitable media by noise. *Multiscal Model Simul.*, **3(1)**, 151–167. 140

—, 2005: Numerical simulation of stochastic pdes for excitable media. *Journal of Computational and Applied Mathematics*, **175**, 429–446. 140

Sirovich, L., 1987: Turbulence and the dynamics of coherent structures, i-iii. *Quart. Appl. Math.*, **45**, 561–590. 116

Smaoui, N. and D. Armbruster, 1997: Symmetry and the Karhunen-Loève analysis. *SIAM J. Sci. Comput.*, **18(1526-1532)**. 116

Smolyak, S., 1963: Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR*, **4**, 240–243. 9, 42

Sobol′, I. M., 1994: *A primer for the Monte Carlo method*. CRC Press. 43

—, 1998: On quasi-Monte Carlo integrations. *Math. Comput. Simulation*, **47**, 103–112. 44

Soize, C., 2003: Random matrix theory and non-parametric model of random uncertainties in vibration analysis. *J. Sound Vibration*, **263(4)**, 893–916. 2

—, 2005: Random matrix theory for modeling uncertainties in computational mechanics. *Comput. Methods Appl. Mech. Engrg.*, **194(12-16)**, 1333–1366. 2

Soize, C. and R. Ghanem, 2004: Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM J. Sci. Comput.*, **26(2)**, 395–410 (electronic). 2

Takáč, P., P. Bollerman, A. Doelman, A. van Harten, and E. S. Titi, 1996: Analyticity of essentially bounded solutions to semilinear parabolic systems and validity of the Ginzburg-Landau equation. *SIAM J. Math. Anal.*, **27(2)**, 424–448. 41, 142

Tatang, M., 1995: *Direct incorporation of uncertainty in chemical and environmental engineering systems*. Ph.D. thesis, MIT. 7, 51

Todor, R. A., 2005: *Sparse Perturbation Algorithms for Elliptic PDE's with Stochastic Data*. Ph.D. thesis, Dipl. Math. University of Bucharest. 7, 15

Traub, J. F. and A. G. Werschulz, 1998: *Complexity and Information*. Cambridge University Press, ISBN 0-521-48506-1. 44

Trefethen, L. N., 2006: Is Gauss quadrature better than Clenshaw-Curtis? SIAM Review, to appear. 52

Velamur Asokan, B. and N. Zabaras, 2005: Using stochastic analysis to capture unstable equilibrium in natural convection. *J. Comput. Phys.*, **208(1)**, 134–153. 2

Verfürth, R., 1996: *A review of a posteriori error estimation and adaptive mesh refinement techniques*. Wiley-Teubner. 1

Vick, S. G., 2002: *Degrees of belief : subjective probability and engineering judgment*. Reston, Va. : American Society of Civil Engineers. 3

Volkwein, S., 1999: Proper orthogonal decomposition and singular value decomposition. In *Spezialforschungsbereich F003 Optimierung und Kontrolle, Projektbereich Kontinuierliche Optimierung und Kontrolle, Bericht*. 116

—, 2001: Optimal control of a phase field model using the proper orthogonal decomposition. *ZAMM*, **81**, 83–97. 116

Wang, J. and N. Zabaras, 2005: Hierarchical Bayesian models for inverse problems in heat conduction. *Inverse Problems*, **21(1)**, 183–206. 3

Wasilkowski, G. W. and H. Woźniakowski, 1995: Explicit cost bounds of algorithms for multivariate tensor product problems. *Journal of Complexity*, **11**, 1–56. 53

Webster, C., 2003: *Numerical Modeling of Waveguide Interface*. Master's thesis, McMaster University. 160

Wiener, N., 1938: The homogeneous chaos. *Amer. J. Math.*, **60**, 897–936. 4

Winter, C. L. and D. M. Tartakovsky, 2002: Groundwater flow in heterogeneous composite aquifers. *Water Resour. Res.*, **38(8)**, 23.1 (doi:10.1029/2001WR000450). 4, 7

Winter, C. L., D. M. Tartakovsky, and A. Guadagnini, 2002: Numerical solutions of moment equations for flow in heterogeneous composite aquifers. *Water Resour. Res.*, **38(5)**, 13.1 (doi:10.1029/2001WR000222). 43

Xiu, D. and J. Hesthaven, 2005: High order collocation methods for differential equations with random inputs. Submitted to SIAM J. Sci. Comput. 7, 8, 9, 42, 51

Xiu, D. and G. E. Karniadakis, 2002a: Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Comput. Methods Appl. Mech. Engrg.*, **191(43)**, 4927–4948. 2, 7, 44, 45, 49, 97

—, 2002b: The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, **24(2)**, 619–644 (electronic). 4

—, 2003: Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.*, **187(1)**, 137–167. 2

Xiu, D. and D. M. Tartakovsky, 2004: A two-scale nonperturbative approach to uncertainty analysis of diffusion in random composites. *Multiscale Model. Simul.*, **2(4)**, 662–674 (electronic). 2

Yan, Y., 2002: Semidiscrete Galerkin approximation for a linear stochastic parabolic partial differential equation driven by an additive noise. Tech. rep., The University of Manchester. 114, 119

—, 2003: The finite element method for a linear stochastic parabolic partial differential equation driven by additive noise. Tech. rep., Chalmers University of Technology. 114

Yuen, K. V. and J. L. Beck, 2003: Updating properties of nonlinear dynamical systems with uncertain input. *J. Engrg. Mech.*, **129(1)**, 9–20. 3

Zabaras, N. and D. Samanta, 2004: A stabilized volume-averaging finite element method for flow in porous media and binary alloy solidification processes. *Internat. J. Numer. Methods Engrg.*, **60(6)**, 1103–1138. 2

BIOGRAPHICAL SKETCH

Clayton Garrett Webster

Clayton Garrett Webster, the son of Gerald and Jay-Dee Webster, was born on the 6th of September, 1978, in Brantford, Ontario, Canada. In the spring of 2001 he graduated *summa cum laude*, obtaining a Bachelors degree in Pure Mathematics; Specialist, from the Department of Mathematics and Statistics at McMaster University in Hamilton, Ontario, Canada. Under the advisement of Professors Walter Craig and Dmitry Pelinovsky, he obtained his Masters degree in Applied Mathematics during the summer of 2003, also from the Department of Mathematics and Statistics at McMaster University. He enrolled in the doctoral program in Applied and Computational Mathematics at Florida State University in the fall of 2003, and under the supervision of Max Gunzburger he completed his Ph. D. in the spring, 2007.

Clayton's research interests include both theoretical and numerical solutions of stochastic partial differential equations and the many applications related to this fascinating research area. Thus far in his career Clayton has several publications concentrating in this area (Burkardt et al., 2007; Nobile et al., 2006, 2007; Babuška et al., 2007; Gunzburger et al., 2007). While obtaining his Master's degree, Clayton did some interesting work on numerical modeling of waveguide interfaces.(Webster, 2003). Even before this, his work Chang et al. (1998) in Engineering Physics focussed on *electrostatic precipitators*.

Clayton currently lives in Tallahassee, Florida, and has recently been awarded the John von Neumann fellowship in Mathematics and Algorithms at Sandia National Laboratories in Albuquerque, New Mexico. Upon completion of this work, he will be accepting this position and moving there with his beautiful wife Sarah.