

# A Reduced-Basis Method for Input-Output Uncertainty Propagation in Stochastic PDEs

by

Ferran Vidal-Codina

Llicenciatura de Matemàtiques, UPC (2010)  
Enginyeria de Camins, Canals i Ports, UPC (2011)

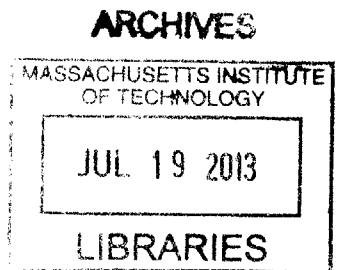
Submitted to the School of Engineering  
in partial fulfillment of the requirements for the degree of  
Master of Science in Computation for Design and Optimization

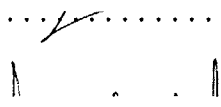
at the

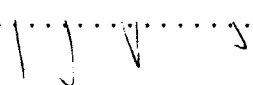
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

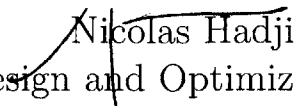
© Massachusetts Institute of Technology 2013. All rights reserved.



Author .....  
 School of Engineering  
May 20, 2013

Certified by .....  
 Jaime Peraire  
Professor  
Thesis Supervisor

Certified by .....  
 Ngoc Cuong Nguyen  
Research Scientist  
Thesis Supervisor

Accepted by .....  
 Nicolas Hadjiconstantinou  
Director, Computation for Design and Optimization (CDO)



# A Reduced-Basis Method for Input-Output Uncertainty Propagation in Stochastic PDEs

by

Ferran Vidal-Codina

Submitted to the School of Engineering  
on May 20, 2013, in partial fulfillment of the  
requirements for the degree of  
Master of Science in Computation for Design and Optimization

## Abstract

Recently there has been a growing interest in quantifying the effects of random inputs in the solution of partial differential equations that arise in a number of areas, including fluid mechanics, elasticity, and wave theory to describe phenomena such as turbulence, random vibrations, flow through porous media, and wave propagation through random media. Monte-Carlo based sampling methods, generalized polynomial chaos and stochastic collocation methods are some of the popular approaches that have been used in the analysis of such problems.

This work proposes a non-intrusive reduced-basis method for the rapid and reliable evaluation of the statistics of linear functionals of stochastic PDEs. Our approach is based on constructing a reduced-basis model for the quantity of interest that enables to solve the full problem very efficiently. In particular, we apply a reduced-basis technique to the Hybridizable Discontinuous Galerkin (HDG) approximation of the underlying PDE, which allows for a rapid and accurate evaluation of the input-output relationship represented by a functional of the solution of the PDE. The method has been devised for problems where an affine parametrization of the PDE in terms of the uncertain input parameters may be obtained. This particular structure enables us to seek an offline-online computational strategy to economize the output evaluation. Indeed, the offline stage (performed once) is computationally intensive since its computational complexity depends on the dimension of the underlying high-order discontinuous finite element space. The online stage (performed many times) provides rapid output evaluation with a computational cost which is several orders of magnitude smaller than the computational cost of the HDG approximation.

In addition, we incorporate two ingredients to the reduced-basis method. First, we employ the greedy algorithm to drive the sampling in the parameter space, by computing inexpensive bounds of the error in the output on the online stage. These error bounds allow us to detect which samples contribute most to the error, thereby enriching the reduced basis with high-quality basis functions. Furthermore, we develop

the reduced basis for not only the primal problem, but also the adjoint problem. This allows us to compute an improved reduced basis output that is crucial in reducing the number of basis functions needed to achieve a prescribed error tolerance. Once the reduced bases have been constructed, we employ Monte-Carlo based sampling methods to perform the uncertainty propagation. The main achievement is that the forward evaluations needed for each Monte-Carlo sample are inexpensive, and therefore statistics of the output can be computed very efficiently. This combined technique renders an uncertainty propagation method that requires a small number of full forward model evaluations and thus greatly reduces the computational burden.

We apply our approach to study the heat conduction of the thermal fin under uncertainty from the diffusivity coefficient and the wave propagation generated by a Gaussian source under uncertainty from the propagation medium. We shall also compare our approach to stochastic collocation methods and Monte-Carlo methods to assess the reliability of the computations.

Thesis Supervisor: Jaime Peraire  
Title: Professor

Thesis Supervisor: Ngoc Cuong Nguyen  
Title: Research Scientist



# Acknowledgments

My first acknowledgement is to 'Obra Social La Caixa' for giving me the once in a lifetime opportunity to come to the United States for a graduate program. When I was awarded the La Caixa fellowship on September 2010 I had little idea of what was awaiting for me. One of the best experiences I have had in my American adventure was the orientation week that La Caixa held for all the fellows before coming to MIT. It was an excellent opportunity to get to know brilliant people from several backgrounds, and some of them have become great friends that have made me feel somehow closer to home.

I would also like to express my gratitude to Professor Jaume Peraire, my advisor during my time here at MIT. He has been very helpful from the first moment, and his profound knowledge and intuition have been very valuable whenever I could not find answers to the problems. The guidance he provided me has definitely made this work possible. Furthermore, it has been his understanding, advice and continuous support what I really value the most. It certainly makes a difference if your supervisor is a person you know he trusts in you. It is a pleasure to work with him and I am confident to continue in this situation for the future.

Another important reason of why I have been able to develop a Master Thesis is Dr. Cuong Nguyen. His expertise and knowledge have been essential in my everyday research life. The majority of the examples shown in this document have been simulated using his code, and he has been crucial in my understanding of the different research topics, always willing to spend time helping me understand any problem that I had along the way. It is a pleasure for me to work and learn with him. Professor Youssef Marzouk has also played a key part in my research project. He has always provided a different scope to the problem, helping me to reach a global level of understanding that would not have been possible without him. I very much appreciate the time he has devoted to me and my project. Finally, I am proud to have had the opportunity to meet and work with Doctor Xevi Roca. I cannot recall how many

engaging discussions we have had and how many times he has helped me fix small (and not so small) problems.

I feel very fortunate for having shared these two years with some wonderful people I could not possibly forget. Joel Saà, my flatmate, friend from college and now labmate, has been very important to share the good and not so good moments. Together with David Moro, Xevi Roca, Carmen García and Hemant Chaurasia they have made this experience very smooth and enjoyable. Apart from them, my friends from 'La Caixa' fellowship have also contributed to making this experience better than I could have expected.

It is of great importance to remember people who have helped you arrive where you are. Back in Barcelona, Professor Antonio Huerta and Professor Rosa Maria Estela encouraged me to take the step I am in, and I am thankful for their thoughtful advice and recommendations that opened the door for this whole adventure.

Finally, last but definitely not least, I have to thank the many friends that, in one way or another, have been of great importance for me. I really appreciate the regular conversations with many of them despite the distance, since they make me feel that I am home, and for the sensation that nothing has ever changed whenever I reencounter them. My family has been of great importance since the beginning, and I could not have possibly make it that far without their constant support and understanding. This work is dedicated to my mother, my father and my brother Adrià, because they are the direct responsables of making me who I am today. And of course I want to dedicate this work to Mariona, the most important person in my life, who is there for my despite my many shortcomings. I am confident that, at the end of the way, everything is going to be alright between us.

# Contents

<b>1</b>	<b>Introduction</b>	<b>17</b>
1.1	Problem definition . . . . .	18
1.1.1	Stochastic PDE . . . . .	18
1.1.2	Input uncertainty . . . . .	19
1.2	Uncertainty Propagation Techniques . . . . .	22
1.2.1	Monte Carlo methods . . . . .	23
1.2.2	Stochastic Galerkin Methods . . . . .	24
1.2.3	Stochastic collocation methods . . . . .	28
1.3	Model Order Reduction . . . . .	33
1.3.1	Proper Orthogonal Decomposition . . . . .	33
1.3.2	Krylov Subspace Methods . . . . .	34
1.4	Reduced-Basis Approach . . . . .	35
1.5	Thesis Outline . . . . .	37
<b>2</b>	<b>Reduced-Basis Methods</b>	<b>39</b>
2.1	Coercive symmetric linear operator: Diffusion problem . . . . .	40
2.1.1	Abstract Formulation . . . . .	40
2.1.2	Reduced-Basis Approach . . . . .	42
2.1.3	<i>A Priori</i> Convergence Results . . . . .	45
2.1.4	Adjoint Problem . . . . .	47
2.1.5	<i>A Posteriori</i> Error Estimation . . . . .	48
2.1.6	Computational Procedure . . . . .	50

2.1.7	Orthogonalization . . . . .	54
2.1.8	Sampling Strategy . . . . .	54
2.1.9	Uncertainty Propagation . . . . .	57
2.2	Noncoercive linear operator: Helmholtz problem . . . . .	58
2.2.1	Abstract Formulation . . . . .	58
2.2.2	Reduced-Basis Approach . . . . .	60
2.2.3	<i>A Priori</i> Convergence Results . . . . .	60
2.2.4	<i>A Posteriori</i> Error Estimation . . . . .	61
2.2.5	Computational Procedure . . . . .	61
2.2.6	Sampling Strategy . . . . .	63
2.2.7	Uncertainty Propagation . . . . .	64
<b>3</b>	<b>The Hybridizable Discontinuous Galerkin Method</b>	<b>65</b>
3.1	Why HDG? . . . . .	66
3.2	The Helmholtz equation . . . . .	67
3.2.1	HDG formulation . . . . .	69
3.2.2	Notation . . . . .	69
3.2.3	Aproximation spaces . . . . .	71
3.2.4	Space discretization . . . . .	72
3.2.5	Weak formulation and matrix system . . . . .	73
3.2.6	Adjoint equation . . . . .	75
3.2.7	Reduced-basis approach . . . . .	78
3.3	The diffusion equation . . . . .	83
3.3.1	HDG formulation . . . . .	83
3.3.2	Weak formulation and matrix system . . . . .	84
3.3.3	Adjoint equation . . . . .	86
3.3.4	Reduced-basis approach . . . . .	86
<b>4</b>	<b>Numerical Results</b>	<b>91</b>
4.1	Thermal Fin . . . . .	92
4.1.1	Definition . . . . .	92

4.1.2	HDG solution . . . . .	93
4.1.3	Reduced-Basis . . . . .	95
4.2	Wave Propagation . . . . .	103
4.2.1	Definition . . . . .	103
4.2.2	HDG solution . . . . .	105
4.2.3	Reduced-Basis . . . . .	106
4.2.4	Uncertainty Propagation . . . . .	111
<b>5</b>	<b>Conclusions and Future Work</b>	<b>117</b>
5.1	Conclusions . . . . .	117
5.2	Future Research . . . . .	118



# List of Figures

2-1	Sketch of the low-dimensional manifold $\mathcal{M}^u$ and the approximation space formed by snapshots . . . . .	43
2-2	Two dimensional sparse grid using Clenshaw-Curtis points. Left: Sparse grid of level 4, total number of points 65. Right: Tensor grid using the same one-dimensional nodes, total number of points 289. . . . .	56
3-1	Degrees of freedom considered by the HDG method for degree $p = 3$ . Classical DG schemes do not consider $\widehat{u}_h$ degrees of freedom . . . . .	68
3-2	Sketch of a high-order triangulation with curved elements. The interior and boundary faces are depicted, together with the triangulation elements $K$ in grey. . . . .	70
3-3	Sketch of a triangulation with curved elements of order $p = 3$ . The degrees of freedom for the corresponding approximation spaces are shown	71
4-1	Geometry of the thermal fin problem. The fin consists of six subfins, and the flux is introduced by the root. . . . .	92
4-2	Discretization of the thermal fin, using a total of 1490 triangles of polynomial degree $p = 3$ . . . . .	93
4-3	Primal solutions to the thermal fin problem for different values of $\boldsymbol{\xi}$ .	94
4-4	Adjoint solutions to the thermal fin problem corresponding to the primal solutions in Figure. 4-3 . . . . .	94
4-5	Output error estimator $\varepsilon_M(\boldsymbol{\xi})/\sqrt{\widehat{\alpha}(\boldsymbol{\xi})}$ and $\varepsilon_{M_d}^d(\boldsymbol{\xi})/\sqrt{\widehat{\alpha}(\boldsymbol{\xi})}$ versus size of the reduced-basis $M, M_d$ . . . . .	96

4-6	Blue dots: $ s(\boldsymbol{\xi}) - s_M(\boldsymbol{\xi}) $ ; red asterisks: $ s(\boldsymbol{\xi}) - \tilde{s}_{M,M_d}(\boldsymbol{\xi}) $ versus realizations of the parameter. Straight line corresponds to desired output tolerance $\varepsilon_{\text{tol}}^2 = 10^{-6}$ . . . . .	97
4-7	Output error estimator $\varepsilon_{\tilde{M}}(\boldsymbol{\xi})/\sqrt{\widehat{\alpha}(\boldsymbol{\xi})}$ versus size of the reduced-basis $\tilde{M}$ . Primal error estimator runs until reaching $\varepsilon_{\text{tol}}^2$ . . . . .	98
4-8	Blue dots: $ s(\boldsymbol{\xi}) - s_{\tilde{M}}(\boldsymbol{\xi}) $ versus realizations of the parameter. Straight line corresponds to desired output tolerance $\varepsilon_{\text{tol}}^2 = 10^{-6}$ . . . . .	98
4-9	Field of sensitivities $\partial s/\partial \kappa_1$ in the $\kappa_1 - \kappa_2$ space . . . . .	101
4-10	Probability density function of the average temperature over the fin, obtained using histogram techniques for the MC simulation . . . . .	102
4-11	Geometry of the wave propagation problem. The source generates a wave that propagates through the medium . . . . .	103
4-12	Decay of eigenvalues of the random density in normalized scale . . . . .	104
4-13	Discretization of the wave propagation problem domain, using a total of 1435 triangles of polynomial degree $p = 4$ . . . . .	105
4-14	Left: Real part of primal solution. Right: Real part of adjoint solution to the wave propagation problem for an arbitrary value of $\boldsymbol{\xi}$ using 10 terms in the expansion of $\rho(\boldsymbol{\xi})$ . . . . .	106
4-15	Left: Density field to generate solutions in Figure 4-14. Right: Source field for the wave propagation problem. . . . .	106
4-16	Maximum relative error for samples in $\tilde{\Theta}$ versus size of reduced-basis, for both primal and primal-adjoint approaches . . . . .	108
4-17	Maximum relative error for samples in $\tilde{\Theta}$ versus size of reduced-basis, for both primal and primal-adjoint approaches. Comparison between sparse grid level 4 (401 points) parameter set and tensor grid (83521 points). Tensor grid is run with a tolerance of $\varepsilon_{\text{tol}} = 10^{-7}$ and sparse grid with a tolerance of $\varepsilon_{\text{tol}} = 10^{-5}$ . . . . .	109
4-18	Maximum relative error for samples in $\tilde{\Theta}$ versus size of reduced-basis, for both primal and primal-adjoint approaches. Sparse grid of level 4 (3937 points) with a tolerance of $\varepsilon_{\text{tol}} = 10^{-5}$ . . . . .	110



4-19	Maximum relative error for samples in $\tilde{\Theta}$ versus size of reduced-basis, for both primal and primal-adjoint approaches. Sparse grid of level 4 (51137 points) with a tolerance of $\varepsilon_{\text{tol}} = 10^{-4}$ . . . . .	112
4-20	Probability density function of the real part of the amplitude at the Neumann boundary, obtained using histogram techniques for the MC simulation . . . . .	114



# List of Tables

- 4.1 Results for the first four raw moments for the reduced-basis uncertainty propagation and stochastic collocation, compared with MC and QMC results . . . . . 100
- 4.2 Values of average and maximum relative error  $|s(\boldsymbol{\xi}) - \tilde{s}_{M,M_d}(\boldsymbol{\xi})| / |s(\boldsymbol{\xi})|$  for the  $10^5$  MC/QMC samples. . . . . 102
- 4.3 First four raw moments of real part of output for the 4-dimensional problem. Results for MC simulation are shown with one standard error. Results for the reduced-basis uncertainty propagation are provided using the same set of MC samples and two different tolerances for the greedy. Results for stochastic collocation are also shown. . . . . 113
- 4.4 Values of average and maximum relative error  $|s(\boldsymbol{\xi}) - \tilde{s}_{M,M_d}(\boldsymbol{\xi})| / |s(\boldsymbol{\xi})|$  for the  $10^5$  MC samples in 4 dimensions. . . . . 113
- 4.5 First four raw moments of real part of output for the 8-dimensional problem. Results for MC simulation are shown with one standard error. Results for the reduced-basis uncertainty propagation are provided using the same set of MC samples and two different tolerances for the greedy. Results for stochastic collocation are also shown. . . . . 115
- 4.6 Values of average and maximum relative error  $|s(\boldsymbol{\xi}) - \tilde{s}_{M,M_d}(\boldsymbol{\xi})| / |s(\boldsymbol{\xi})|$  for the  $10^5$  MC samples in 8 dimensions. . . . . 115

4.7	First four raw moments of real part of output for the 16-dimensional problem. Results for MC simulation are shown with one standard error. Results for the reduced-basis uncertainty propagation are provided using the same set of MC samples and two different tolerances for the greedy. Results for stochastic collocation are also shown. . . . .	115
4.8	Values of average and maximum relative error $ s(\boldsymbol{\xi}) - \tilde{s}_{M, M_d}(\boldsymbol{\xi})  /  s(\boldsymbol{\xi}) $ for the $10^5$ MC samples in 16 dimensions. . . . .	116

# Chapter 1

## Introduction

The analysis of physical and engineering systems is often carried out by mathematical modeling and numerical simulations. For a given system, the corresponding model requires certain input data. Input data may consist of model parameters, forcing terms, boundary conditions, geometry information, etc. The most common approach has been to analyze the mathematical models under the assumption that such input data was deterministic, i.e. precisely known. However, in many situations input information is not known precisely. In these cases, one needs to consider uncertainty in the input data. Mathematical models represented by differential equations which incorporate uncertainty are known as stochastic ordinary/partial differential equations (SODE/SPDE), see [51, 99].

Uncertainty in the input data may come from different sources. It can be that the physical system under study has itself some intrinsic variability, for example uncertainty in the wind and seismic loadings on civil structures or uncertainty in the mechanical properties of materials and fluids. Another possible source of uncertainty arises when data comes from experiments. In this case, observed quantities must be considered in a probabilistic setting. In some cases we may have to infer the values of a certain field (e.g. permeability, porosity) from limited experimental measurements. Furthermore, it is also possible that we are unable to effectively characterize the

physical system with a mathematical model. For instance, we may have errors in geometry, roughness or multiscale behavior that we are unable to capture.

Therefore, there is a growing need to represent the uncertainty in the data and effectively propagate it through the mathematical model (SODE/SPDE). The goal of this probabilistic approach resides in computing statistical moments of the solution, or statistics of some observable output, or quantity of interest. In particular, one is interested in the prediction of moments and probability density function (PDF) of the quantity of interest, which are usually defined as real-valued functionals of the solution of the SODE/SPDE. This quantity of interest may consist of average values of the solution in certain regions, fluxes across boundaries, physical quantities derived from the solution (drag on an airfoil), etc.

## 1.1 Problem definition

### 1.1.1 Stochastic PDE

Following the notation used in [123], we define a complete probability space  $(\Omega, \mathcal{F}, \mathcal{P})$ , where  $\Omega$  is the space of events,  $\mathcal{F} \in 2^\Omega$  is the  $\sigma$ -algebra of subsets (events) in  $\Omega$ , and  $\mathcal{P} : \mathcal{F} \rightarrow [0, 1]$  is the probability measure. Define also a  $d$ -dimensional bounded domain  $D \subset \mathbb{R}^d$  ( $d = 1, 2, 3$ ), with boundary  $\Gamma$ . The goal is to find a stochastic function  $u : \Omega \times \bar{D} \rightarrow \mathbb{R}$ , such that for  $\mathcal{P}$ -almost everywhere (a.e.)  $\omega \in \Omega$ , the following equation holds

$$\mathcal{L}(\omega, \mathbf{x}; u) = f(\omega, \mathbf{x}), \quad \forall \mathbf{x} \in D \tag{1.1a}$$

$$\mathcal{B}(\omega, \mathbf{x}; u) = g(\omega, \mathbf{x}), \quad \forall \mathbf{x} \in \Gamma \tag{1.1b}$$

where  $\mathbf{x} = (x_1, \dots, x_d)$  refers to the spatial coordinates in  $\mathbb{R}^d$ ,  $\mathcal{L}$  is a linear or nonlinear differential operator,  $f$  is the forcing term,  $\mathcal{B}$  is the boundary operator and  $g$  is the boundary data. This boundary operator may be different on different boundary

subsets, for instance  $\mathcal{B} \equiv u$  on Dirichlet subsets,  $\mathcal{B} \equiv \mathbf{n} \cdot \nabla u$  on Neumann subsets and  $\mathcal{B} \equiv \mathbf{n} \cdot \nabla u + u$  on Robin subsets. We shall impose no limitation on the location of the randomness, that is all terms in equations (1.1) may be stochastic. Furthermore, we assume sufficient regularity on the boundary  $\Gamma$  and the terms  $f, g$ , such that (1.1) is well-posed for  $\mathcal{P}$ -a.e.  $\omega \in \Omega$ <sup>1</sup>.

The solution of the SPDE (1.1) is used to define quantities of interest. The general form of a quantity of interest is expressed as

$$s(\omega) = \mathcal{J}(u(\mathbf{x}, \omega), \omega) \tag{1.2}$$

where  $\mathcal{J}$  is a real-valued functional. Note that, besides the implicit dependence of the output on both  $\mathbf{x}, \omega$  via the solution  $u(\mathbf{x}, \omega)$  of the SPDE, we also allow for an explicit dependence on the stochasticity.

### 1.1.2 Input uncertainty

The stochasticity in the problem is represented by a vector of random variables  $\boldsymbol{\xi}(\omega) = (\xi_1(\omega), \dots, \xi_N(\omega))$ , where  $N$  is the dimension of the stochastic space. These random variables represent different forms of uncertainty.

One case that we shall consider is where the mathematical model depends on some unknown parameters, that are taken as random variables with a certain joint probability density function. For instance, we might think of the scattering of a planar wave interacting with a certain object or scatterer, where the incidence angle and the wavenumber of the incident wave are random variables.

---

<sup>1</sup>The incorporation of the time derivative is straightforward by considering an extra dimension  $D \subset \mathbb{R}^{d+1}$  for time-dependent terms.

## Random processes

A different situation may arise if the input data is assumed to vary randomly from one point of the physical domain to another. In this case, uncertainty is usually expressed in terms of a random process. Given a physical space  $D$  and a space of events  $\Omega$ , a real-valued random process  $\mathcal{R}$  is defined as

$$\mathcal{R} : (\mathbf{x}, \omega) \in D \times \Omega \mapsto \mathcal{R}(\mathbf{x}, \omega) \in \mathbb{R} \quad (1.3)$$

where for any  $\mathbf{x} \in D$ ,  $\mathcal{R}(\mathbf{x}, \cdot)$  represents a random variable. Usually only second-order processes are considered, that is processes such that

$$\mathcal{R}(\mathbf{x}, \cdot) \in L^2(\Omega, \mathcal{P}), \quad \forall \mathbf{x} \in D \quad (1.4)$$

where  $L^2(\Omega, \mathcal{P})$  is the space of second-order random variables defined on the complete probability space equipped with the following inner product and norm

$$\langle \xi, \eta \rangle = \int_{\Omega} \xi(\omega)\eta(\omega) d\mathcal{P}(\omega) = \mathbb{E}[\xi\eta], \quad \forall \xi, \eta, \in L^2(\Omega, \mathcal{P}) \quad (1.5)$$

$$\xi \in L^2(\Omega, \mathcal{P}) \rightarrow \langle \xi, \xi \rangle = \|\xi\|_{\Omega}^2 < \infty \quad (1.6)$$

where  $\mathbb{E}$  is the expectation operator. Random processes are defined together with a mean  $\overline{\mathcal{R}}(\mathbf{x})$  and a spatial correlation structure  $\mathcal{C} : \bar{D} \times \bar{D} \rightarrow \mathbb{R}$  that is real, symmetric and positive definite.

The most common representation for stochastic processes is the Karhunen-Loève (KL) expansion [57, 58]. The KL expansion, which exists provided that the random field has bounded second moments [2], represents the stochastic process as an infinite linear combination of functions and coefficients, similarly to a Fourier series. The KL expansion chooses a basis such that it minimizes the total mean squared error. The coefficients for the KL expansion are uncorrelated random variables, whereas functions are deterministic continuous and real-valued, and form a complete orthogonal set with



respect to the  $L^2$  norm. In the special case where  $\mathcal{R}$  is a Gaussian process, then the random variables are independent Gaussian random variables.

The functions in the KL expansion are closely related to the spectrum of  $\mathcal{C}$ . The general form of a KL expansion is

$$\mathcal{R}(\mathbf{x}, \omega) = \overline{\mathcal{R}}(\mathbf{x}) + \sum_{i \geq 1} \sqrt{\lambda_i} \phi_i(\mathbf{x}) \xi_i \quad (1.7)$$

where  $\lambda_i, \phi_i$  are the eigenpairs of  $\mathcal{C}$ , and can be retrieved by solving the homogeneous Fredholm integral equation of the second kind

$$\int_D \mathcal{C}(\mathbf{x}, \bar{\mathbf{x}}) \phi_i(\bar{\mathbf{x}}) d\bar{\mathbf{x}} = \lambda_i \phi_i(\mathbf{x}) \quad (1.8)$$

In order to reduce the complexity of dealing with an infinite number of variables, the usual approach is to resort to a reduced-order representation, that is truncate the spectral expansion of the stochastic process, which is commonly known as *finite dimensional noise assumption* [2, 123], that is

$$\mathcal{R}(\mathbf{x}, \omega) = \overline{\mathcal{R}}(\mathbf{x}) + \sum_{i=1}^N \sqrt{\lambda_i} \phi_i(\mathbf{x}) \xi_i \quad (1.9)$$

The number of terms retained in the expansion depends on the decay of the eigenvalues (see [26]) of the correlation function  $\mathcal{C}$  of the random process  $\mathcal{R}$ , which is associated to the spatial variation of the input data. Rapid decay of eigenvalues leads to only a few terms needed to reproduce the behavior of the random field. Elliptic problems where the diffusivity field on a certain medium is expressed as a truncated KL expansion have been widely studied in [1, 3, 4, 2, 61, 92, 93, 123, 124].

Either considering parametric uncertainty or a truncated spectral expansion of a random field, we assume that the random inputs are described by a set of  $N$  random variables  $\boldsymbol{\xi} = (\xi_1(\omega), \dots, \xi_N(\omega))$ . Let  $\Xi$  be the support of  $\boldsymbol{\xi}$ , which can in principle be unbounded (e.g. gaussian or exponential random variables). We shall define the joint probability density of the input random variables as  $\pi(\boldsymbol{\xi})$  with support  $\Xi$ . The special

case where all the random variables are uncorrelated reduces the latter expression to  $\pi(\boldsymbol{\xi}) = \prod_{i=1}^N \pi_i(\xi_i)$ ,  $\forall \boldsymbol{\xi} \in \Xi = \prod_{i=1}^N \Xi_i$ , i.e. the density and the support factorize.

Therefore the problem defined in (1.1) may be recast as

$$\mathcal{L}(\boldsymbol{\xi}, \mathbf{x}; u) = f(\boldsymbol{\xi}, \mathbf{x}), \quad \forall (\mathbf{x}, \boldsymbol{\xi}) \in D \times \Xi \quad (1.10a)$$

$$\mathcal{B}(\boldsymbol{\xi}, \mathbf{x}; u) = g(\boldsymbol{\xi}, \mathbf{x}), \quad \forall (\mathbf{x}, \boldsymbol{\xi}) \in D \times \Xi \quad (1.10b)$$

with quantity of interest

$$s(\boldsymbol{\xi}) = \mathcal{J}(u(\mathbf{x}, \boldsymbol{\xi}), \boldsymbol{\xi}) \quad (1.11)$$

which is a reduction of the initially infinite-dimensional stochastic problem (1.1) to a finite-dimensional one. To solve the latter problem, a number of methodologies have been developed, which are briefly reviewed below. These methodologies fall into two groups: uncertainty propagation techniques, which employ the full forward model in various ways, and reduced order modeling and reduced-basis techniques, which aim to construct a surrogate for the input-output model to economize the forward evaluations.

## 1.2 Uncertainty Propagation Techniques

In this section, we will provide an overview of the most common numerical methods available to solve problem (1.10). These methods can be classified in two different ways: statistical versus non-statistical and intrusive versus non-intrusive, depending on the information that they provide or the nature of the implementation itself. The three main families of methods that are described here are Monte Carlo methods, stochastic Galerkin methods and stochastic collocation methods.

Apart from these methods, it is worth mentioning the second moment analysis [56] and the perturbation method [36, 49]. These methods, which are non-statistical and intrusive, are usually only valid for small ranges of the uncertainty. Henceforth, they

are not adequate for complex or non-linear problems, where small values of the input uncertainty could translate in large output uncertainty.

### 1.2.1 Monte Carlo methods

The Monte Carlo (MC) simulation [24] is perhaps the most popular method for solving partial differential equations. It is a statistical non-intrusive method, in the sense that it gives access to the complete statistics and that it only requires repetitive evaluations of a deterministic code.

#### Algorithm

The procedure of applying the MC simulation to a stochastic problem like (1.10) where randomness is represented by  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_N)$  is straightforward

- Draw  $M$  independent identically distributed (i.i.d.) samples from the input probability density function  $\pi(\boldsymbol{\xi})$ , such that we have a collection of  $M$  realizations of a random variable of dimension  $N$ ,  $\Theta_M = \{\boldsymbol{\xi}^j\}_{j=1}^M$ ;
- For each realization  $\boldsymbol{\xi}^j$  perform a deterministic solve of (1.10), obtaining  $u_j = u(\boldsymbol{\xi}^j, \boldsymbol{x})$  or more directly  $s_j = s(\boldsymbol{\xi}^j; \theta)$ ;
- Once all realizations have been computed, perform a trivial postprocessing to obtain the desired statistics of either the solution or the quantity of interest. For example, the k-th moment of the output is easily computed as

$$\mu_k \equiv \mathbb{E}[s^k] = \frac{1}{M} \sum_{j=1}^M s_j^k \quad (1.12)$$

#### Pros, cons and improvements

MC methods are simple to implement and are embarrassingly parallel. It is also important to stress that the simplicity of MC methods ensures robustness, i.e. that

if the deterministic solver works, so will MC methods.

Furthermore, another interesting feature of MC methods is that the convergence rate is not influenced by the dimension of the stochastic space, therefore they are especially appealing for problems with a large number of random variables. Conversely, since the convergence rate is  $\mathcal{O}\left(1/\sqrt{M}\right)$ , where  $M$  is the number of samples considered, it takes a large amount of realizations to acquire good results. Moreover, for cases where the deterministic problem is complex and expensive to solve, the computational work required to obtain accurate solutions quickly becomes intractable. Therefore, MC methods are appropriate whenever the desired accuracy in the computations is not very high, but the dimension of the problem is large.

This slow convergence has in part been mitigated by the introduction of several techniques. Quasi-Monte Carlo methods (using low-discrepancy quasi-random sequences) [25, 90, 42, 14] produce convergence rates of order  $\mathcal{O}\left((\log M)^r/M\right)$ , where  $r$  is a linear function of the stochastic dimension. In addition to that, stratified sampling techniques such as Latin Hypercube Sampling [29, 45, 59, 109] retain the  $\mathcal{O}\left(1/\sqrt{M}\right)$  rate of convergence, but significantly improve the constant for certain response functions. Other techniques include importance sampling [14, 37], Markov chain Monte Carlo [27, 68] or the sensitivity derivative enhanced Monte Carlo [15, 16, 29].

## 1.2.2 Stochastic Galerkin Methods

The stochastic Galerkin methods (SGM) generalize the theory of Wiener-Hermite polynomial chaos expansion, developed by Wiener in [118], and combine it with a finite element method to model uncertainty in a stochastic PDE. The SGM were first introduced by Ghanem *et al.* in [36], and are a type of non-statistical intrusive methods.

## Polynomial space

This method uses a set of orthogonal multi-variate polynomials that span the stochastic space  $\Xi$ , of dimension  $N$ . Since the set of orthogonal multi-variate polynomials is typically infinite dimensional, a finite dimensional subspace is used in computations. The two most common approaches include the complete polynomial space [36, 34, 72, 104, 124, 126, 127] and the tensor product space [1, 3, 20, 26, 104].

The complete polynomial space  $W_N^p$  considers only  $N$ -variate orthogonal polynomials of degree up to  $p$

$$\Phi_{\mathbf{p}}(\boldsymbol{\xi}) = \prod_{|\mathbf{p}| \leq p} \phi_{p_i}(\xi_i) \quad \mathbf{p} = (p_1, \dots, p_N) \quad (1.13)$$

Note that this expression includes all possible combinations of the multiindex  $\mathbf{p}$  satisfying

$$|\mathbf{p}| = \sum_{i=1}^N p_i \leq p \quad (1.14)$$

where  $p_i$  is the degree of the univariate polynomial  $\phi_{p_i}$  in the  $i$ th dimension. The dimension of  $W_N^p$  is given by  $N_p + 1 = \binom{N+p}{p}$ .

On the other hand, the full tensor product space  $Z_N^p$  considers all possible combinations of univariate orthogonal polynomials of degree  $p$  in each dimension, that is

$$\Phi_{\mathbf{p}}(\boldsymbol{\xi}) = \prod_{\max p_i = p} \phi_{p_i}(\xi_i) \quad \mathbf{p} = (p_1, \dots, p_N) \quad (1.15)$$

with total dimension  $\dim Z_N^p = (p+1)^N$ . Note that  $W_N^p \subset Z_N^p$ .

## Orthogonal polynomials

In the early work by Wiener only Gaussian random variables were considered, leading to a basis of Hermite polynomials. The first applications of the SGM with Hermite polynomials were in solid mechanics [36, 33, 32, 35]. In order to consider random

distributions other than Gaussian, the SGM has been extended into the generalized polynomial chaos expansion (gPCE) [125, 117].

For a general one-dimensional random variable  $\xi_i$  with density function  $\pi_i$  and support  $\Xi_i$ , the set of orthogonal polynomials used in the gPCE satisfy

$$\mathbb{E}[\phi_m \phi_n] = \int_{\Xi_i} \pi_i(\xi_i) \phi_n(\xi_i) \phi_m(\xi_i) \quad (1.16)$$

where  $\delta_{mn}$  is the Kronecker delta. The orthogonal polynomials are defined by the integration weight  $\pi_i$  (or density function). Therefore, uniform random variables are best represented by Legendre polynomials, Gaussian random variables by Hermite polynomials, Gamma random variables by Laguerre polynomials, etc.

### Stochastic Galerkin projection

By choosing an approximation space and a family of polynomials we may approximate the solution of (1.10) as

$$u(\mathbf{x}, \boldsymbol{\xi}) \approx \sum_{i=0}^{N_p} u_i(\mathbf{x}) \Phi_i(\boldsymbol{\xi}) \quad (1.17)$$

where  $u_i$  are the (deterministic) spectral coefficients or PC coefficients, and each  $i$  is an admissible multiindex according to (1.14). Plugging expression (1.17) into equation (1.10) we arrive at

$$\mathcal{L} \left( \sum_{i=0}^{N_p} \alpha_i \Phi_i, \mathbf{x}; \sum_{j=0}^{N_p} u_j \Phi_j \right) = f \left( \sum_{k=0}^{N_p} \beta_k \Phi_k, \mathbf{x} \right), \quad \forall (\mathbf{x}, \boldsymbol{\xi}) \in D \times \Xi \quad (1.18a)$$

$$\mathcal{B} \left( \sum_{i=0}^{N_p} \gamma_i \Phi_i, \mathbf{x}; \sum_{j=0}^{N_p} u_j \Phi_j \right) = g \left( \sum_{k=0}^{N_p} \lambda_k \Phi_k, \mathbf{x} \right), \quad \forall (\mathbf{x}, \boldsymbol{\xi}) \in D \times \Xi \quad (1.18b)$$

where the coefficients  $\alpha_i, \beta_i, \gamma_i, \lambda_i$  are zero if there is no explicit dependence on  $\boldsymbol{\xi}$ , and are otherwise easy to compute. Furthermore, by performing a stochastic Galerkin projection of the system above using all the elements in the basis  $\Phi_l$  and the inner

product defined by (1.16), we arrive to

$$\mathbb{E} \left[ \mathcal{L} \left( \sum_{i=0}^{N_p} \alpha_i \Phi_i, \mathbf{x}; \sum_{j=0}^{N_p} u_j \Phi_j \right) \Phi_l \right] = \mathbb{E} \left[ f \left( \sum_{k=0}^{N_p} \beta_k \Phi_k, \mathbf{x} \right) \Phi_l \right], \quad \forall (\mathbf{x}, \boldsymbol{\xi}) \in D \times \Xi \quad (1.19a)$$

$$\mathbb{E} \left[ \mathcal{B} \left( \sum_{i=0}^{N_p} \gamma_i \Phi_i, \mathbf{x}; \sum_{j=0}^{N_p} u_j \Phi_j \right) \Phi_l \right] = \mathbb{E} \left[ g \left( \sum_{k=0}^{N_p} \lambda_k \Phi_k, \mathbf{x} \right) \Phi_l \right], \quad \forall (\mathbf{x}, \boldsymbol{\xi}) \in D \times \Xi \quad (1.19b)$$

which can be reduced to a system of  $N_p + 1$  deterministic equations for the coefficients  $u_i$ . Using a similar expansion for the objective function (1.11) an approximation may be obtained

$$\hat{s}(\boldsymbol{\xi}) = \sum_{i=0}^{N_p} s_i \Phi_i(\boldsymbol{\xi}) \quad (1.20)$$

where  $s_i$  can be computed using  $u_i$ . The mean and the variance of the quantity of interest can be analytically computed from this expression, whereas high order statistics may be obtained by sampling (1.20) using a Monte Carlo approach, which is an inexpensive operation.

### Pros, cons and improvements

This methodology has proven to be very effective when solving PDEs in a broad range of applications, such as diffusion problems and heat conduction [33, 124, 127], structural dynamics [35], transport in random media [32] and fluid dynamics [126, 17, 70]. These methods converge exponentially fast with increasing order of the expansions, whenever the solution is sufficiently smooth in the stochastic space, as numerous studies [3, 4, 20, 128] have shown. Therefore, they provide much more accurate solutions for simple and smooth problems than Monte Carlo.

On the contrary, if the solution or the QoI exhibits discontinuities on the random space, gPCE loses the exponential convergence, and may even fail to converge. To overcome this situation, approaches based on local polynomials have been devised.

Some examples include the hat functions in finite elements [3, 4, 20, 71], wavelet basis expansions [55, 54] or a multi-element generalized polynomial chaos expansion [116].

There are, however, some drawbacks in the use of gPCE. Firstly, note that the number of expansion terms grows combinatorially for the polynomial order  $p$  and the stochastic dimension  $N$ . This fact, combined with the coupled nature of the equations (1.19), makes the solution of the problem very expensive for large dimensions. Furthermore, the intrusive nature of the polynomial chaos expansion implies that for each application considered, a robust stochastic solver needs to be coded. Apart from the extra coding work involved, for complex nonlinear PDEs the coupled equations resulting from Galerkin projections in (1.19) may have very complicated forms (see [17, 70]), and they may also present numerical instabilities (see [21]). Therefore, gPCE is not suitable for complex problems with a large number of stochastic dimensions.

### 1.2.3 Stochastic collocation methods

The coupled nature of the final equations is definitely a big challenge for gPCE approaches. To overcome this limitation, collocation methods have been introduced. The goal of these methods is to combine the strength of Stochastic Galerkin methods of using a polynomial approximation in the stochastic space with the simplicity of implementation of Monte Carlo methods by sampling in the random space.

The stochastic collocation method, first introduced in [69] as a deterministic sampling method and further developed in [123], computes independently deterministic solutions of the stochastic PDE at certain points in the stochastic space and then builds an interpolation function to approximate the desired solution. Therefore, they may be classified as non-statistical and non-intrusive.



## Lagrange interpolation

The idea of the stochastic collocation method is to construct an interpolating function of the stochastic dependent variables in problem (1.10) employing the value of these variables at certain points in the random space. Consider a set of points in the  $N$ -dimensional random space,  $\Theta_M = \{\boldsymbol{\xi}^j\}_{j=1}^M$  and the set of  $N$ -variate polynomials of degree at most  $p$ ,  $\Pi_N^p$ . The Lagrange interpolation problem of a smooth function  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  consists in finding the polynomial  $\mathcal{I}(f) \in \Pi_N^p$  such that  $\mathcal{I}(f)(\boldsymbol{\xi}_j) = f(\boldsymbol{\xi}_j)$ ,  $j = 1, \dots, M$ . Using the Lagrange interpolation polynomials  $\mathcal{I}(f)$  is expressed as

$$\mathcal{I}(f)(\boldsymbol{\xi}) = \sum_{i=1}^M f(\boldsymbol{\xi}_i) \mathcal{L}_i(\boldsymbol{\xi}), \quad \mathcal{L}_i(\boldsymbol{\xi}_j) = \delta_{ij}, \quad \forall i, j = 1, \dots, M \quad (1.21)$$

The convergence of the interpolation is not guaranteed for any distribution of the interpolation nodes. The quality of the approximation, and therefore the interpolation error, in 1D ( $N = 1$ ), are described by the Lebesgue theorem, namely

$$\|f(\xi) - p^*(\xi)\|_\infty \leq \|f(\xi) - \mathcal{I}(f)(\xi)\|_\infty \leq (1 + \Lambda(\xi)) \|f(\xi) - p^*(\xi)\|_\infty \quad (1.22)$$

where  $p^*(\xi)$  is the best approximating polynomial and  $\Lambda(\xi) = \max_{\xi \in \Xi} \sum_{i=1}^M |\mathcal{L}_i(\xi)|$  is known as the Lebesgue constant. This constant depends solely on the location of the interpolation points, and its determination is nontrivial even for  $N = 1$ . Distributions of points with small  $\Lambda$  include the Clenshaw-Curtis points (located at the extrema of Chebyshev polynomials) and the Gauss quadrature points (see [12, 50, 110] for more details).

## Formulation

Interpolating the solution of (1.10) using Lagrange polynomials we write

$$u(\mathbf{x}, \boldsymbol{\xi}) \approx \sum_{i=1}^M u_i(\mathbf{x}) \mathcal{L}_i(\boldsymbol{\xi}) \quad (1.23)$$

which can be inserted in equation (1.10). By the interpolative nature of the solution, we readily obtain a set of  $M$  decoupled deterministic problems

$$\mathcal{L}(\boldsymbol{\xi}_i, \mathbf{x}; u) = f(\boldsymbol{\xi}_i, \mathbf{x}), \quad \forall \mathbf{x} \in D \quad (1.24a)$$

$$\mathcal{B}(\boldsymbol{\xi}_i, \mathbf{x}; u) = g(\boldsymbol{\xi}_i, \mathbf{x}), \quad \forall \mathbf{x} \in D \quad (1.24b)$$

for each node  $\boldsymbol{\xi}_i$ . For quantities of interest the formulation is similar

$$\hat{s}(\boldsymbol{\xi}) = \sum_{i=1}^M s(\boldsymbol{\xi}_i) \mathcal{L}_i(\boldsymbol{\xi}), \quad s(\boldsymbol{\xi}_i) = \mathcal{J}(u(\mathbf{x}, \boldsymbol{\xi}_i)) \quad (1.25)$$

Once the deterministic problems have been solved, the  $k$ -th moment of the output may be computed as

$$\hat{\mu}_k \equiv \mathbb{E}[\hat{s}^k] = \sum_{i=1}^M s^k(\boldsymbol{\xi}_i) \int_{\Xi} \mathcal{L}_i(\boldsymbol{\xi}) \boldsymbol{\pi}(\boldsymbol{\xi}) d\boldsymbol{\xi} \quad (1.26)$$

To evaluate these integrals we need explicit knowledge of the Lagrange polynomials. A possible alternative is to choose the nodal set  $\Theta_M$  to be a cubature set. Therefore the integrals in (1.26) are reduced to

$$\mathbb{E}[\hat{s}^k] = \sum_{i=1}^M s^k(\boldsymbol{\xi}_i) w_i \quad (1.27)$$

where  $\{w_i\}$  are the integration weights. Note that if the cubature is of order  $q$  it will introduce errors whenever the integrals cannot be evaluated exactly, i.e. if they are not polynomials of order at most  $q$ .

Note that the complexity of the SCM relies on the computation of  $M$  deterministic problems, corresponding to solving equation (1.24) for each point in the nodal set  $\Theta_M$ . Therefore the objective is to minimize the number of model evaluations, provided that there is sufficient accuracy in the interpolation.

## Tensor product grids

A straightforward way to extend the Lagrange interpolation polynomials to the multidimensional case  $N > 1$  is to use the tensor product of the unidimensional nodal set. Let  $M_1, \dots, M_N$  be the number of collocation points in each dimension. The multidimensional interpolation formula for the quantity of interest can be written as

$$\mathcal{I}(s)(\boldsymbol{\xi}) = (\mathcal{I}_1 \otimes \dots \otimes \mathcal{I}_N)(s)(\boldsymbol{\xi}) = \sum_{i_1=1}^{M_1} \dots \sum_{i_N=1}^{M_N} s(\xi_1^{i_1}, \dots, \xi_N^{i_N}) \cdot (\mathcal{L}_1^{i_1} \otimes \dots \otimes \mathcal{L}_N^{i_N}) \quad (1.28)$$

where  $\mathcal{I}_k$  is the interpolation formula in the  $k$  direction and  $\xi_l^k$  is the  $k$ th point in the  $l$ th coordinate. This method is used in [2], providing also the first rigorous error estimate for elliptic SPDEs. Clearly, if we are using the same number of points in each dimension, this methodology requires  $M^N$  problem evaluations. Therefore, it should only be used for a small number of dimensions, e.g.  $N \leq 5$ .

## Sparse grids

For a moderately large number of random variables, instead of tensor product grids one should resort to sparse grids, first introduced by Smolyak [108], and further analyzed by [7, 30, 96, 97, 98, 50].

The Smolyak algorithm provides an efficient construction of multidimensional interpolative functions based on a linear combination of product formulas, rendering a nodal set with a substantial reduction on the number of nodes compared with the tensor product grid (see [123]). For the explicit construction of the sparse grid, the reader should refer to [28, 123, 93].

The implementation of the SCM using sparse grids has been widely used for many applications, including elliptic problems [93, 123], stochastic ODEs [122], parabolic problems [91], natural convection problems [28] and the wave equation [74, 75].

## Pros, cons and improvements

The notorious advantages of the stochastic collocation methods, either utilizing full tensor product or sparse grids, are the exponential convergence, proved in [2], which makes them attractive for elliptic and parabolic problems with smooth data on the stochastic space. Furthermore, the non-intrusive nature of the method is also very advantageous, as it opens the possibility of reusing existing deterministic codes.

Nevertheless, the SCM also suffers from the curse of dimensionality as the gPCE. In fact, as shown in [123], the SCM has more degrees of freedom than the intrusive polynomial chaos, even for the sparse grids case. The influence of dimensionality on the cost is equivalently bad for both approaches (but with a clearly differentiated coding effort).

Some techniques have been devised to alleviate the influence of the dimension. The most common is the inclusion of anisotropy in the sparse grid. Adaptivity on sparse grids has been discussed by [31, 41, 50], and recent research has been devoted to applying adaptivity techniques on sparse grids for uncertainty propagation. The usual approach is to detect which dimensions are relevant to the problem and weigh them unequally, since the classic Smolyak algorithm prescribes an isotropic weighting for all dimensions. The detection is either done on-the-fly [28] or found by a combination of *a priori* and *a posteriori* information [92]. The results presented in [28, 92] above cover the case of truncated KL expansions of random fields, and for situations with rapid decay of eigenvalues the adaptive approach allows to consider a great number of stochastic dimensions with just a mild dependence on the dimensionality. An alternative is the procedure presented in [61]. This approach uses an adaptive piecewise linear hierarchical basis that allows to detect singular local behavior -otherwise impractical to capture when using global polynomials such as Lagrange- and scales linearly with dimensions. Although for elliptic problems the extension to multiple dimensions is straightforward with these anisotropic approaches, a general extension for more complicated problems remains unresolved.

## 1.3 Model Order Reduction

Model order reduction (MOR) is a necessary tool for simulating large-scale dynamical systems where full forward evaluations are expensive. The idea is to develop a surrogate that is efficient to evaluate and yet produces accurate solutions. This becomes necessary in the event that repeated simulations are required, since multiple evaluations of the full model may be prohibitive. The issue of how to create a model that is faithful, especially when the dimension of the input parameter space is large, is a challenge. The basic idea behind these methods relies on projecting the high-dimensional state space onto a very-low dimensional state space that renders the reduced-order model.

### 1.3.1 Proper Orthogonal Decomposition

The proper orthogonal decomposition (POD) has been widely used to obtain a low-dimensional representation of dynamical systems. The basic idea is to form a basis by a set of state solutions, usually called snapshots, obtained by numerically solving the full forward model for certain values of the input parameters. The idea behind the POD method is to capture the dominant dynamics of the system, which will then produce a more accurate reduced model [107]. The issue of choosing the snapshots is critical to the quality of the model. The model is then obtained using a Galerkin projection onto the subspace spanned by these snapshots. The POD method has been used for several large-scale dynamical systems, e.g. in CFD and aerodynamic applications [8, 44, 119], optimal control of fluids [60, 103], turbomachinery flows [22, 120]. Furthermore, research has been devoted into extending POD for nonlinear systems and nonlinear structural dynamics [52, 73].

However, the issue of sampling the parameter space is still a challenge, especially if high-dimensional inputs are considered. It is obvious that uniform sampling (tensor product grids) quickly become computationally intractable due to the exponential dependence on the dimensionality. Another choice may be random sampling,

amongst which Latin Hypercube Sampling and quasi Monte Carlo sequences are good candidates. Recently, the greedy sampling method introduced by Patera et. al. [40, 38, 114, 111] has proved to be an efficient alternative to adaptively select the snapshots based on inexpensive estimates of the error. The greedy sampling method has been applied to incompressible Navier-Stokes equations [111], parabolic and time-dependent PDEs [114, 111], noncoercive and nonlinear operators [76, 114]. Recently, the greedy approach has also been reformulated as a sequence of adaptive model-constrained optimization problems [13], being advantageous from the point of view that the sample space is treated as a continuous and not discrete.

### 1.3.2 Krylov Subspace Methods

Krylov subspace-based methods is an alternative approach for efficient modeling and simulation of dynamical systems. The idea underlying these methods resides in approximating the transfer function of the original system by a subspace spanned by orthogonal basis functions and projecting the original system onto this subspace. Krylov-subspace methods are robust and have a low computational cost, therefore have been widely used in various engineering applications [5, 53, 121].

Furthermore, Krylov-subspace methods have also been extended to deal with nonlinear problems [18], by linearizing the original nonlinear system using Taylor expansions [100] that represent the nonlinear model as a combination of linear models, generated at different linearization points in the state space. These approaches allow for a treatment of nonlinear systems using standard linear MOR methodologies.

Although MOR methods have been widely used in the study of large-scale systems, they do present some weaknesses. Firstly, it is difficult to efficiently characterize and develop a reduced-order model for highly nonlinear problems without incurring in excessive computational complexity. Moreover, only *a priori* bounds have been derived for the linear case, therefore there is no guarantee in the quality of the model via *a posteriori* bounds even for the linear case. Finally, most MOR approaches

focus on large-scale dynamical systems, i.e. considering time-variation, and not many techniques have been devised for parametric applications.

## 1.4 Reduced-Basis Approach

The reduced-basis method is a technique to obtain rapid yet accurate approximations of functional outputs of parametrized PDEs. In fact, a stochastic PDE (1.10) can also be seen as a parametrized PDE, where the parameters are endowed with some probability distribution. The basic idea is to detect some underlying patterns in the solution of the PDE as a consequence of the parametric dependence. That is, the solution will not in general be an arbitrary member of the true space  $X$ , but rather live in a much lower dimensional manifold. If such structure can be detected, a reduced basis of solutions may be constructed, where each solution corresponds to one realization of the input parameter. Provided that the basis is sufficiently rich, approximations of both the solution field and the output can be sought on this low-dimensional space.

The reduced-basis method was first introduced in the early 1980s by Noor [94, 95] for single and multiple parameter problems in nonlinear analysis of structures. Further work was developed to include *a priori* error analysis [23, 101], although at the time no rigorous *a posteriori* analysis of the error had been introduced, thereby incapacitating the certification of the computations. Recently, a lot of work has been devoted to reduced-basis methods by Patera et. al. [6, 76, 40, 38, 39, 63, 65, 66, 102, 105, 115, 114, 113, 111], introducing several new concepts that have greatly developed these techniques, such as the use of global approximation spaces based on snapshots of the solution for the full governing equations; rigorous *a posteriori* error estimators to certify the quality of the approximation; and the exploitation of an offline/online strategy to improve computational efficiency. The first theoretical *a priori* convergence results by Maday et.al. [66] demonstrated exponential convergence of the reduced-basis. The method was developed for linear elliptic problems with

affine parametrization [62] and for eigenvalue problems [63]. Extensions to include nonlinear and noncoercive elliptic and parabolic problems were developed by Rovas, Veroy et al. [105, 115, 114, 112], together with developing rigorous and sharp error estimators. New error estimation methods for linear and nonlinear time dependent problems have been developed by Grepl [40, 38].

The reduced-basis method has been applied to a broad range of areas, from nonlinear analysis of structures [23, 94, 101], fluid flow problems [48, 47], thermal fin problems [65, 67] and steady incompressible Navier-Stokes [111].

The inclusion of *a posteriori* error estimation is of vital importance to guarantee the quality of the approximation performed by the reduced-basis method. The work by Patera et. al. [63, 102, 76, 115, 114, 113] introduces rigorous error estimators for a wide variety of partial differential equations, i.e. linear elliptic, noncoercive linear, nonaffine noncoercive and even highly nonlinear monotonic elliptic equations. The combination of rigorous *a posteriori* error estimates, which are inexpensive to compute, with the greedy algorithm introduced in [40, 38, 114, 111] constitutes a powerful tool for constructing a surrogate for the input-output relationship in parametrized PDEs.

The approach proposed in this work is to construct a reduced-basis for stochastic PDEs with affine dependence on the parameters to predict quantities of interest defined as linear functionals of the solution of (1.10). The reduced-basis is computed in a fully automated manner using a greedy approach sampling strategy driven by *a posteriori* error estimates of the quantity of interest, which are computed inexpensively. Furthermore, the reduced-basis is also built for the adjoint problem, rendering the opportunity to compute an enhanced output approximation that is of vital importance in reducing the number of basis functions needed to attain a prescribed error tolerance, thus minimizing the number of full model evaluations. The basic ingredients and the algorithm for the reduced-basis approach are presented in great detail in Chapter 2.

In this thesis a reduced-basis approach for uncertainty propagation (RBUP) is



devised. The goal of this work is to develop a non-intrusive method for uncertainty propagation that relies on a reduced-basis to economize the input-output evaluation. One of the main disadvantages of non-intrusive methods is the number of full model evaluations that are required to attain a prescribed error in the moments of the output. If the model is very expensive to evaluate, or if the dimensionality of the input parameter space is large, the uncertainty propagation problem can become computationally inefficient. The objective of the method here proposed is to mitigate the cost of uncertainty propagation by constructing a surrogate of the input-output map using a reduced-basis that leads to significant savings in computational cost. Reduced-basis techniques have already been applied to uncertainty propagation by Boyaval et.al. [10, 11, 43], and variance reduction strategies have been proposed to accelerate the convergence of Monte Carlo [9]. The novelty introduced here is the application of high-fidelity finite element solvers to the reduced-basis, and the incorporation of the adjoint problem to achieve superior convergence of the reduced-basis. Finally, the reduced-basis uncertainty propagation is also applied to wave propagation problems to demonstrate its performance for noncoercive cases.

## 1.5 Thesis Outline

The two central themes in this thesis are the application of reduced-basis methods on Hybridizable Discontinuous Galerkin (HDG) methods for solving parametrized differential equations with an affine dependence of input parameters, and the application of the reduced-basis HDG approach for input-output uncertainty propagation employing Monte Carlo techniques. In Chapter 2 the basic concepts for the reduced-basis method are presented, together with *a posteriori* error bounds, sampling strategy and implementation. Furthermore, the procedure to apply the reduced-basis method to uncertainty propagation is also described. In Chapter 3 the HDG method is reviewed for the Helmholtz and the diffusion equation, the examples that will be analyzed in depth. The application of reduced-basis approaches to the HDG method is also

introduced here, together with the derivation of the adjoint equation that will be of great interest in the computations. In Chapter 4 numerical results are presented for both the performance of the reduced-basis applied to HDG and for the Reduced-Basis Uncertainty Propagation compared to the Stochastic Collocation method. For both comparisons the Helmholtz and the diffusion equation are used as model problems. Finally, we conclude in Chapter 5 with the summary of the work presented and provide some guidelines for future work.

# Chapter 2

## Reduced-Basis Methods

In this chapter we address the issue of computing the solution of the forward problem in an efficient way. The main focus is on stochastic partial differential equations where an affine parametrization in terms of the uncertain input parameters may be found. In general, the goal of the computation is to obtain an estimate for a certain output or quantity of interest. Throughout this thesis we will assume that the quantity of interest may be computed as linear functional of the solution of the stochastic PDE. Extension to nonaffine parametrized PDEs has already been studied in [6, 76, 39, 64] with the introduction of the empirical interpolation method (EIM). Furthermore, Nguyen et.al. [89] presented a "best points" interpolation method for an optimal approximation of parametrized functions.

The reduced-basis methods allow a rapid yet reliable evaluation of a certain input-output relationship induced by a parametrized partial differential equation. Following the work by Patera et. al. [76, 40, 38, 39, 63, 65, 66, 102, 105, 115, 114, 113] the method presented in this Chapter has three important aspects that differentiate it from the initial work in reduced-basis introduced in the 1980s. Firstly, the use of global approximation spaces, i.e. members of the reduced-basis constitute solutions for the full governing PDE; second, the use of rigorous *a posteriori* error estimators to not only certify the quality of the approximation, but also to adaptively enrich the

approximation space; and third, we exploit the structure of the problem to devise an offline/online computational strategy to economize the output evaluation. Furthermore, we also include adjoint techniques that greatly accelerate the construction of the reduced-basis. The method is reviewed for the simplest case of a coercive elliptic linear operator with compliant output, although it is extended for a coercive elliptic operator with noncompliant output and for noncoercive and nonsymmetric linear elliptic equations.

Finally, we propose a strategy for propagating uncertainty in stochastic PDEs that relies on a reduced-basis surrogate of the input-output map constructed applying the greedy algorithm on sparse grids combined with Monte Carlo techniques. This approach allows us to inexpensively evaluate a large amount of samples using the surrogate to obtain the statistics of the output.

## 2.1 Coercive symmetric linear operator: Diffusion problem

### 2.1.1 Abstract Formulation

Let us consider a parametrized PDE in its weak form

$$a(u^e(\boldsymbol{\xi}), v; \boldsymbol{\xi}) = f(v; \boldsymbol{\xi}), \quad \forall v \in X^e \quad (2.1)$$

where  $u^e(\boldsymbol{\xi})$  is the exact solution (the superscript  $e$  represents exact) of the PDE and  $\boldsymbol{\xi} \in \Xi \subset \mathbb{R}^N$  represents the parametric dependence of the PDE. Note that for many problems the stochasticity can be reduced to parametric dependency, e.g. using the finite-dimensional noise assumption of KL expansions of random processes, or just simple parametric uncertainty. The problem that we aim to solve reads: find

$$s^e(\boldsymbol{\xi}) = l(u^e(\boldsymbol{\xi})) \quad (2.2)$$

where  $s^e(\boldsymbol{\xi})$  is the (exact) quantity of interest;  $u^e(\boldsymbol{\xi})$  is the (exact) field variable;  $X^e$  is an associated Hilbert space defined over a suitably regular physical domain  $D \in \mathbb{R}^d$ ,  $d = 1, 2, 3$  (independent of the parameter space  $\Xi$ ), where we can define an associated inner product  $(w, v)_{X^e}$  and norm  $\|w\|_{X^e} = \sqrt{(w, w)_{X^e}}$ ; and finally  $a(\cdot, \cdot)$  and  $f(\cdot), l(\cdot)$  are  $X$ -continuous bilinear and linear forms, respectively. We also require the linear functionals to be bounded.

Throughout this work we will focus only on second-order PDEs, therefore the function space required must satisfy  $X^e \subset (H^1(D))^{\nu 1}$ , where  $\nu$  is the dimension of the field variable  $u^e(\boldsymbol{\xi})$ .

For approximation purposes, the infinite-dimensional space  $X^e$  is replaced by a high-order discontinuous finite element approximation space  $X$ ,  $X \subset X^{e2}$ , of dimension  $\mathcal{N}$ . Problem (2.1)-(2.2) is restated as evaluating

$$s(\boldsymbol{\xi}) = l(u(\boldsymbol{\xi})) \quad (2.3)$$

where the field variable  $u(\boldsymbol{\xi}) \in X$  is the solution of the discretized weak form

$$a(u(\boldsymbol{\xi}), v; \boldsymbol{\xi}) = f(v; \boldsymbol{\xi}), \quad \forall v \in X \quad (2.4)$$

for a given  $\boldsymbol{\xi} \in \Xi$ . We shall denote  $(\cdot, \cdot)_X$  and  $\|\cdot\|_X$  the inner product and the norm associated with the finite element space  $X$ . Furthermore, we shall also assume that the high-order discontinuous finite element space  $X$  used to approximate the true solution  $u^e$  by  $u$  is sufficiently rich, therefore  $s \rightarrow s^e$ ,  $u \rightarrow u^e$  as  $\mathcal{N} \rightarrow \infty$ . The dual functional space  $X'$  is given by

$$\|f\|_{X'} \equiv \sup_{v \in X} \frac{f(v)}{\|v\|_X}, \quad \forall f \in X' \quad (2.5)$$

Furthermore, we shall assume that the bilinear form is symmetric,  $a(w, v; \boldsymbol{\xi}) =$

---

<sup>1</sup>The space  $H^1(D)$  is the space of functions  $v$  that are square-integrable and its gradient is also square integrable, i.e.  $v \in L^2(D)$ ,  $\nabla v \in L^2(D)^d$

<sup>2</sup>The usual notation would be to consider  $X$  the real space and  $X_h$  the finite element space, where  $h$  is the size of the discretization, but in here we shall drop this subscript for a simplified notation

$a(w, v; \boldsymbol{\xi}), \forall w, v \in X, \forall \boldsymbol{\xi} \in \Xi$ , continuous

$$a(w, v; \boldsymbol{\xi}) \leq \gamma(\boldsymbol{\xi}) \|w\|_X \|v\|_X \leq \gamma_0 \|w\|_X \|v\|_X, \quad \forall \boldsymbol{\xi} \in \Xi \quad (2.6)$$

and coercive

$$0 < \alpha_0 \leq \alpha(\boldsymbol{\xi}) = \inf_{v \in X} \frac{a(v, v; \boldsymbol{\xi})}{\|v\|_X^2}, \quad \forall \boldsymbol{\xi} \in \Xi \quad (2.7)$$

The constant  $\alpha(\boldsymbol{\xi})$  is the minimum singular value associated with the differential operator (coercivity constant) and  $\gamma(\boldsymbol{\xi})$  is the continuity constant. Even though they are referred to as constants, they do depend on the parameter  $\boldsymbol{\xi}$ . Proof of existence and uniqueness follows from the application of the Lax-Milgram theorem, assuming enough regularity of the domain and the source  $f$ .

Finally, the key assumption is that the parametric dependence of  $a, f$  may be expressed, for finite (small) integers  $Q, \tilde{Q}$ , as

$$a(w, v; \boldsymbol{\xi}) = \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) a_q(w, v), \quad \forall w, v \in X, \forall \boldsymbol{\xi} \in \Xi \quad (2.8a)$$

$$f(v, \boldsymbol{\xi}) = \sum_{q=1}^{\tilde{Q}} \tilde{\sigma}_q(\boldsymbol{\xi}) f_q(v) \quad \forall v \in X, \forall \boldsymbol{\xi} \in \Xi \quad (2.8b)$$

for some real-valued functions  $\sigma_q, \tilde{\sigma}_q : \Xi \rightarrow \mathbb{R}$  continuous, differentiable and parameter-dependent, whereas the forms  $a_q : X \times X \rightarrow \mathbb{R}, f_q : X \rightarrow \mathbb{R}$  are parameter-independent. This affine parametrization is crucial in the computational savings that are described below. Nonetheless, reduced-basis methods allow for non-affine dependence of the parameters, but this discussion is beyond the scope of this work.

### 2.1.2 Reduced-Basis Approach

The reduced-basis method relies on the fact that the field variable  $u(\boldsymbol{\xi})$  is not an arbitrary member of the infinite-dimensional space  $X$  associated to the underlying partial differential equation; but instead resides on a low dimensional manifold

$\mathcal{M}^u \equiv \{u(\boldsymbol{\xi}) | \boldsymbol{\xi} \in \Xi\}$  that is induced by the parametric dependence. In fact, the approximation space  $X$  may in general contain solutions to the PDE that are not relevant to our interest, since they do not lie on this manifold. The idea is to further reduce the dimension approximation space  $X$  by focusing solely on the manifold of solutions  $\mathcal{M}^u$ , therefore economizing the computations.

The idea behind the reduced-basis method resides in constructing an approximation space to the manifold of solutions of the parametrized PDE. This approximation space is constructed using global solutions to the PDE at selected points  $\boldsymbol{\xi}$  in the parameter space  $\Xi$ , usually known as snapshots. Thereby, for any parameter in the parameter space the field variable  $u(\boldsymbol{\xi})$  and the output  $s(\boldsymbol{\xi})$  may be computed by a suitable projection onto the approximation space, see Figure 2-1. The Lagrangian reduced-basis approximation space is constructed as  $W_M \equiv \text{span}\{\zeta_m \equiv u(\boldsymbol{\xi}_m), m = 1, \dots, M\}$ , for a certain collection of samples  $S_M = \{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_M \in \Xi\}$ . Due to the nature of the reduced-basis described above, it is required to orthogonalize the basis with respect to the inner product  $(\cdot, \cdot)_X$  to avoid conditioning problems of the basis.

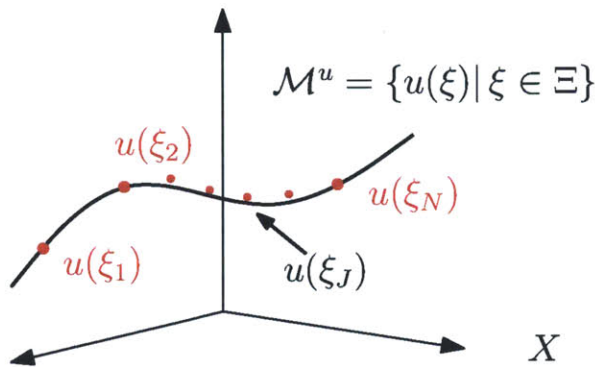


Figure 2-1: Sketch of the low-dimensional manifold  $\mathcal{M}^u$  and the approximation space formed by snapshots

The reduced-basis space  $W_M$  seeks to approximate the manifold of solutions  $\mathcal{M}^u$  of the parametric PDE, which has very low dimensionality. The reduced-basis benefits from this low dimensionality, allowing to approximate the field variable  $u(\boldsymbol{\xi})$  by a linear combination of elements of the basis  $\zeta_m$ . The reduced-basis approximation to the field variable is denoted as  $u_M(\boldsymbol{\xi})$ . Furthermore, we also expect rapid conver-

gence (in some cases, exponential [63, 66, 102]), i.e. using a total number of basis functions  $M$  significantly smaller than the dimension of the finite element space  $\mathcal{N}$ . The reduced-basis solution  $u_M(\boldsymbol{\xi})$  for an arbitrary parameter  $\boldsymbol{\xi}_M$  is obtained using a Galerkin projection onto the reduced-basis space  $W_M$

$$a(u_M(\boldsymbol{\xi}), v; \boldsymbol{\xi}) = f(v; \boldsymbol{\xi}), \quad \forall v \in W_M \quad (2.9)$$

Indeed, let us denote by  $\Phi_M$  the matrix for the reduced-basis space  $W_M$

$$\Phi = \begin{bmatrix} | & & | \\ \zeta_1 & \cdots & \zeta_M \\ | & & | \end{bmatrix} \quad (2.10)$$

where  $(\Phi^T, \Phi)_X = \mathbb{I}_M$ , that is the basis' members are properly orthonormalized with respect to the inner product associated with the space  $X$ . Choosing as trial functions the basis of  $W_M$ , we may express  $u_M = \sum_{m=1}^M \lambda_m \zeta_m = \Phi \lambda$ . If the test space is the same as the trial space, we recover a Galerkin projection

$$\Phi^T \mathbb{A} \Phi \lambda = \Phi^T \mathbb{F} \quad (2.11)$$

where  $\mathbb{A}, \mathbb{F}$  are the matrices corresponding to the forms  $a(\cdot, \cdot), f(\cdot)$ , using as test and trial functions the ones corresponding to the high-order discontinuous finite element space  $X$ . Note that the dimensionality of the system has been reduced to  $M$ , which is typically  $\ll \mathcal{N}$ , and although it is no longer sparse, the orthogonalization of the snapshots ensures good conditioning properties. Once system (2.11) is solved, the approximate output for  $\boldsymbol{\xi}_M$  may be easily evaluated as

$$s_M(\boldsymbol{\xi}) = l(u_M) = \mathbb{L}^T \Phi \lambda \quad (2.12)$$

where  $\mathbb{L}$  is the matrix corresponding to the linear functional  $l(\cdot)$ .

Naturally, a crucial point in the reduced-basis approach is to decide which snap-



shots should form the basis. Ideally, one would like to choose the field variables containing the maximum amount of information about the manifold  $\mathcal{M}^u$  in order to minimize the number of elements in the basis, thus reducing the computational effort in solving (2.11)-(2.12). Therefore the issue of selecting the values of the parameters  $\boldsymbol{\xi}$  that render a better reduced-basis is of great importance, and care must be taken in order to effectively explore  $\mathcal{M}^u$  to extract the best candidates. The strategy pursued here differs from the one used in POD procedures, in the sense that we use inexpensive error bounds that serve as indicators to choose potential new candidates to optimally enrich the basis. The procedure for enriching the basis is described below.

### 2.1.3 *A Priori* Convergence Results

We consider here the convergence rates of the solution  $u_M(\boldsymbol{\xi}) \rightarrow u(\boldsymbol{\xi})$  and the output  $s_M(\boldsymbol{\xi}) \rightarrow s(\boldsymbol{\xi})$ . It is simple to prove optimality of the reduced-basis approximate solution  $u_M(\boldsymbol{\xi})$  in the  $X$ -norm, that is

$$\|u(\boldsymbol{\xi}) - u_M(\boldsymbol{\xi})\|_X \leq \sqrt{\frac{\gamma(\boldsymbol{\xi})}{\alpha(\boldsymbol{\xi})}} \inf_{w_M \in W_M} \|u(\boldsymbol{\xi}) - w_M(\boldsymbol{\xi})\|_X \quad (2.13)$$

which follows from Galerkin orthogonality  $a(u(\boldsymbol{\xi}) - u_M(\boldsymbol{\xi}), v; \boldsymbol{\xi}) = 0$ ,  $\forall v \in W_M$ , symmetry, continuity and coercivity. From now on, let us denote the error in the approximation as  $e(\boldsymbol{\xi}) = u(\boldsymbol{\xi}) - u_M(\boldsymbol{\xi})$ . The convergence of the approximated output highly depends on whether the output is or not compliant.

## Compliant output

For the special compliance case  $l = f$  one can prove optimal convergence of the approximated output  $s_M(\boldsymbol{\xi})$  to  $s(\boldsymbol{\xi})$  in the  $X$ -norm

$$\begin{aligned}
s(\boldsymbol{\xi}) - s_M(\boldsymbol{\xi}) &= l(u(\boldsymbol{\xi}) - u_M(\boldsymbol{\xi})) \\
&= a(u(\boldsymbol{\xi}), e(\boldsymbol{\xi}); \boldsymbol{\xi}) \quad (l = f) \\
&= a(e(\boldsymbol{\xi}), e(\boldsymbol{\xi}); \boldsymbol{\xi}) \quad (\text{symmetry and Galerkin orthogonality}) \quad (2.14) \\
&\leq \gamma(\boldsymbol{\xi}) \|e(\boldsymbol{\xi})\|_X^2 \quad (\text{continuity}) \\
&\leq \frac{\gamma^2(\boldsymbol{\xi})}{\alpha(\boldsymbol{\xi})} \inf_{w_M \in W_M} \|u(\boldsymbol{\xi}) - w_M(\boldsymbol{\xi})\|_X^2 \quad (\text{result (2.13)})
\end{aligned}$$

one can readily see the convenience of result (2.14), since the quantity of interest converges as the square of the error in the reduced-basis approximation. Unfortunately, this is a rather special case, because in general the quantity of interest may have nothing to do with the forcing term.

## Noncompliant output

In the more general noncompliance case, the approximated output  $s_M(\boldsymbol{\xi})$  converges to  $s(\boldsymbol{\xi})$  in the  $X$ -norm as

$$|s(\boldsymbol{\xi}) - s_M(\boldsymbol{\xi})| = |l(e(\boldsymbol{\xi}))| \leq \|l\|_{X'} \|e(\boldsymbol{\xi})\|_X \leq \|l\|_{X'} \sqrt{\frac{\gamma(\boldsymbol{\xi})}{\alpha(\boldsymbol{\xi})}} \inf_{w_M \in W_M} \|u(\boldsymbol{\xi}) - w_M(\boldsymbol{\xi})\|_X \quad (2.15)$$

which follows from the boundedness of the functional output and result (2.13). This error bound is definitely worse than (2.14), since the square effect is lost. For certain problems the error bound may suffice to satisfy a prescribed error tolerance, provided rapid convergence of the reduced basis. However, for some applications this slow convergence may imply a large amount of functions in the basis, therefore losing the property of representing the manifold  $\mathcal{M}^u$  with a very low-dimensional basis. To overcome this limitation, adjoint techniques are applied.

### 2.1.4 Adjoint Problem

For noncompliant symmetric problems (and for nonsymmetric problems in general), the optimal convergence of the output, and thus the "square" effect, may be regained by the introduction of adjoints. The adjoint problem is defined as the dual problem of (2.4): for a given  $\boldsymbol{\xi} \in \Xi$ , find the dual field variable  $\psi(\boldsymbol{\xi})$  satisfying

$$a(v, \psi(\boldsymbol{\xi}); \boldsymbol{\xi}) = -l(v), \quad \forall v \in X \quad (2.16)$$

The purpose is to build a reduced-basis approach for the dual problem as well. Therefore we introduce an adjoint reduced-basis approximation space<sup>3</sup>  $W_{M_d}^d \equiv \text{span}\{\zeta_{m_d}^d \equiv \psi(\boldsymbol{\xi}_{m_d}^d)\}$ ,  $m_d = 1, \dots, M_d$ . Similarly, the basis is formed by snapshots of the dual solution for some certain values of the parameters. Even though the parameter space  $\Xi$  is the same, the set of snapshots may very well differ between the primal and the adjoint. The error of the adjoint problem is defined as  $e^d(\boldsymbol{\xi}) = \psi(\boldsymbol{\xi}) - \psi_{M_d}(\boldsymbol{\xi})$ .

To prove the *a priori* convergence result for the output, we apply a Galerkin projection for both the primal and the adjoint problems onto both the primal  $W_M$  and the adjoint  $W_{M_d}^d$  approximation space

$$a(u_M(\boldsymbol{\xi}), v; \boldsymbol{\xi}) = f(v; \boldsymbol{\xi}), \quad \forall v \in W_M \quad (2.17a)$$

$$a(v, \psi_{M_d}(\boldsymbol{\xi}); \boldsymbol{\xi}) = -l(v), \quad \forall v \in W_{M_d}^d \quad (2.17b)$$

and define the optimal or enhanced output that will be used in the computations

$$\tilde{s}_{M, M_d}(\boldsymbol{\xi}) = l(u_M(\boldsymbol{\xi})) - r(\psi_{M_d}(\boldsymbol{\xi})) \quad (2.18)$$

where  $r(\cdot; \boldsymbol{\xi}) = f(\cdot) - a(u_M(\boldsymbol{\xi}), \cdot; \boldsymbol{\xi})$  is the residual of the primal problem (2.17a). Recalling the *a priori* convergence result of the reduced-basis approximate solution

---

<sup>3</sup>From now on the subscript/superscript <sup>d</sup> stands for dual

in (2.13), a similar statement may be derived for the adjoint problem

$$\|\psi(\boldsymbol{\xi}) - \psi_{M_d}(\boldsymbol{\xi})\|_X \leq \sqrt{\frac{\gamma(\boldsymbol{\xi})}{\alpha(\boldsymbol{\xi})}} \inf_{w_{M_d} \in W_{M_d}^d} \|\psi(\boldsymbol{\xi}) - w_{M_d}(\boldsymbol{\xi})\|_X. \quad (2.19)$$

Then convergence of the enhanced output can be proven optimal

$$\begin{aligned} |s(\boldsymbol{\xi}) - \tilde{s}_{M, M_d}(\boldsymbol{\xi})| &= |l(e(\boldsymbol{\xi})) + f(\psi_{M_d}(\boldsymbol{\xi})) - a(u_M(\boldsymbol{\xi}), \psi_{M_d}(\boldsymbol{\xi}); \boldsymbol{\xi})| \\ &= |-a(e(\boldsymbol{\xi}), \psi(\boldsymbol{\xi}); \boldsymbol{\xi}) + a(e(\boldsymbol{\xi}), \psi_{M_d}(\boldsymbol{\xi}); \boldsymbol{\xi})| \\ &= |-a(e(\boldsymbol{\xi}), e^d(\boldsymbol{\xi}); \boldsymbol{\xi})| \\ &\leq \gamma(\boldsymbol{\xi}) \|e(\boldsymbol{\xi})\|_X \|e^d(\boldsymbol{\xi})\|_X \\ &\leq \frac{\gamma^2(\boldsymbol{\xi})}{\alpha(\boldsymbol{\xi})} \inf_{w_M \in W_M} \|u(\boldsymbol{\xi}) - w_M(\boldsymbol{\xi})\|_X \inf_{w_{M_d} \in W_{M_d}^d} \|\psi(\boldsymbol{\xi}) - w_{M_d}(\boldsymbol{\xi})\|_X \end{aligned} \quad (2.20)$$

using the definition for both the primal and the dual problems, Galerkin orthogonality and continuity. Thanks to the incorporation of the adjoint, we recover the 'square' effect because the output converges as the product of the primal and the dual errors.

### 2.1.5 *A Posteriori* Error Estimation

Although the *a priori* convergence theory introduced above is useful for estimating the convergence of the reduced-basis approach, in practice *a posteriori* error estimation is more useful. Indeed, we need a tool to know how many snapshots need to be incorporated in the reduced-basis to achieve a prescribed tolerance on the output for all parametric inputs. It is in principle expensive to compute  $|s(\boldsymbol{\xi}) - \tilde{s}_{M, M_d}(\boldsymbol{\xi})|$  for all possible parametric points to verify the quality of our reduced-basis, therefore rigor in the bounds is necessary. On the other hand, having an excessively large reduced-basis may result in an intense computational complexity, hence sharpness of the *a posteriori* error bounds is also required.

We shall assume that we are given a lower bound of the coercivity constant  $\alpha(\boldsymbol{\xi})$ ,

denoted by  $\widehat{\alpha}(\boldsymbol{\xi})$ , which depends on the sample in the parameter space, such that  $\alpha(\boldsymbol{\xi}) \geq \widehat{\alpha}(\boldsymbol{\xi}) \geq \alpha_0 > 0$ . The general computation of  $\widehat{\alpha}(\boldsymbol{\xi})$ , different for the symmetric and for the nonsymmetric case, will not be addressed here. Efficient techniques for computing it for the both the symmetric and nonsymmetric case have been developed in [76, 46, 65].

The residual of the adjoint problem is defined as  $r^d(\cdot; \boldsymbol{\xi}) = -l(\cdot) - a(\cdot, \psi_{M_d}(\boldsymbol{\xi}); \boldsymbol{\xi})$ . Using the primal and the adjoint residuals, we shall define the dual norm of these residuals

$$\varepsilon_M(\boldsymbol{\xi}) = \|r(v; \boldsymbol{\xi})\|_{X'} = \sup_{v \in X} \frac{r(v; \boldsymbol{\xi})}{\|v\|_X}, \quad (2.21a)$$

$$\varepsilon_{M_d}^d(\boldsymbol{\xi}) = \|r^d(v; \boldsymbol{\xi})\|_{X'} = \sup_{v \in X} \frac{r^d(v; \boldsymbol{\xi})}{\|v\|_X}. \quad (2.21b)$$

The error bounds associated with the field variable, or energy bounds, are defined by

$$\Delta_M(\boldsymbol{\xi}) = \frac{\varepsilon_M(\boldsymbol{\xi})}{\widehat{\alpha}(\boldsymbol{\xi})}, \quad (2.22a)$$

$$\Delta_{M_d}(\boldsymbol{\xi}) = \frac{\varepsilon_{M_d}^d(\boldsymbol{\xi})}{\widehat{\alpha}(\boldsymbol{\xi})}. \quad (2.22b)$$

And finally, recalling expression (2.20), the enhanced output error estimator is defined as

$$\Delta_{M, M_d}^s(\boldsymbol{\xi}) = \frac{\varepsilon_M(\boldsymbol{\xi}) \cdot \varepsilon_{M_d}^d(\boldsymbol{\xi})}{\widehat{\alpha}(\boldsymbol{\xi})}. \quad (2.23)$$

It can be proven [76] that the above error bounds are both sharp and rigorous. It is important that the bounds are rigorous, in order to certify the quality of the approximations using the reduced-basis. Sharpness is crucial to ensure that the number of snapshots needed is not unnecessarily large, thereby improves efficiency of the method. Finally, in order to compute the dual norm of the residuals needed for the

output bounds, we resort to duality arguments

$$\|r(v; \boldsymbol{\xi})\|_{X'} = \|y(\boldsymbol{\xi})\|_X, \quad \text{where } (v, y(\boldsymbol{\xi}))_X = r(v; \boldsymbol{\xi}), \quad (2.24a)$$

$$\|r^d(v; \boldsymbol{\xi})\|_{X'} = \|z(\boldsymbol{\xi})\|_X, \quad \text{where } (v, z(\boldsymbol{\xi}))_X = r^d(v; \boldsymbol{\xi}). \quad (2.24b)$$

Once these classical bounds have been reviewed, the computational procedure must be addressed. The key idea used in the reduced-basis context is to pursue an *offline-online* stage decoupling that allows us to greatly economize the output evaluation and associated error bounds.

## 2.1.6 Computational Procedure

The assumption of affine parametric dependency is of key importance for achieving significant computational savings. In general, even though for some problems  $M, M_d$  may be small, we would like that the surrogate model constructed for the input-output relation is independent of the dimension of the underlying finite element approximation  $\mathcal{N}$ . To this end, efficient offline-online computational procedures are developed, allowing for a full exploitation of the dimension reduction. We first express  $u_M(\boldsymbol{\xi}), \psi_{M_d}(\boldsymbol{\xi})$  as a linear combination of the reduced-basis functions as

$$u_M(\boldsymbol{\xi}) = \sum_{m=1}^M \lambda_m \zeta_m, \quad (2.25a)$$

$$\psi_{M_d}(\boldsymbol{\xi}) = \sum_{m_d=1}^{M_d} \lambda_{m_d}^d \zeta_{m_d}^d. \quad (2.25b)$$

For simplicity, we adopt the matrix notation already introduced in (2.10)-(2.11). Galerkin projections (2.17) are used for both the primal and the adjoint problem,

thus the latter coefficients  $\lambda, \lambda_d$  satisfy the following  $M \times M, M_d \times M_d$  systems

$$\Phi^T \left[ \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{A}_q \right] \Phi \lambda = \Phi^T \left[ \sum_{q=1}^{\tilde{Q}} \tilde{\sigma}_q(\boldsymbol{\xi}) \mathbb{F}_q \right] \quad (2.26a)$$

$$\Phi_d^T \left[ \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{A}_q \right] \Phi_d \lambda_d = -\Phi_d^T \mathbb{L} \quad (2.26b)$$

recalling the affine dependency of the parameters for  $a(\cdot, \cdot), f(\cdot)$ . The reduced-basis enhanced output is evaluated as

$$\tilde{s}_{M, M_d}(\boldsymbol{\xi}) = \mathbb{L}^T \Phi \lambda - (\Phi_d \lambda_d)^T \left( \left[ \sum_{q=1}^{\tilde{Q}} \tilde{\sigma}_q(\boldsymbol{\xi}) \mathbb{F}_q \right] - \left[ \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{A}_q \right] \Phi \lambda \right) \quad (2.27)$$

finally, we define the dual norm of the residuals. It follows from (2.24) that

$$\mathbb{Y} = \mathbb{X}^{-1} \left[ \sum_{q=1}^{\tilde{Q}} \tilde{\sigma}_q(\boldsymbol{\xi}) \mathbb{F}_q - \left( \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{A}_q \right) \Phi \lambda \right] \quad (2.28a)$$

$$\mathbb{Z} = \mathbb{X}^{-1} \left[ -\mathbb{L} - \left( \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{A}_q \right) \Phi_d \lambda_d \right] \quad (2.28b)$$

where  $\mathbb{X}$  is the matrix of the inner product associated with the finite element space  $X$ , and  $\mathbb{Y}, \mathbb{Z}$  represent vectors  $y(\boldsymbol{\xi}), z(\boldsymbol{\xi})$ . Inserting these values into (2.24) we obtain

$$\begin{aligned} \varepsilon_M(\boldsymbol{\xi}) &= \sum_{q=1}^{\tilde{Q}} \sum_{q'=1}^{\tilde{Q}} \tilde{\sigma}_q(\boldsymbol{\xi}) \tilde{\sigma}_{q'}(\boldsymbol{\xi}) \mathbb{F}_q^T \mathbb{X}^{-1} \mathbb{F}_{q'} - 2 \sum_{q=1}^{\tilde{Q}} \sum_{q'=1}^Q \tilde{\sigma}_q(\boldsymbol{\xi}) \sigma_{q'}(\boldsymbol{\xi}) \mathbb{F}_q^T \mathbb{X}^{-1} \mathbb{A}_{q'} \Phi \lambda + \\ &+ \sum_{q=1}^Q \sum_{q'=1}^Q \sigma_q(\boldsymbol{\xi}) \sigma_{q'}(\boldsymbol{\xi}) (\Phi \lambda)^T \mathbb{A}_{q'} \mathbb{X}^{-1} \mathbb{A}_q \Phi \lambda \end{aligned} \quad (2.29)$$

for the primal and

$$\begin{aligned} \varepsilon_{M_d}^d(\boldsymbol{\xi}) &= \mathbb{L}^T \mathbb{X}^{-1} \mathbb{L} + 2 \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{L}^T \mathbb{X}^{-1} \mathbb{A}_q \Phi_d \lambda_d + \\ &+ \sum_{q=1}^Q \sum_{q'=1}^Q \sigma_q(\boldsymbol{\xi}) \sigma_{q'}(\boldsymbol{\xi}) (\Phi_d \lambda_d)^T \mathbb{A}_{q'} \mathbb{X}^{-1} \mathbb{A}_q \Phi_d \lambda_d \end{aligned} \quad (2.30)$$

for the dual. Expressions (2.26)-(2.30) clearly suggest an offline-online strategy to effectively minimize the computational cost of evaluating the output and the error bounds for any parameter  $\boldsymbol{\xi}$ . Note that there should also be a strategy for evaluating the lower bound of coercivity constant  $\widehat{\alpha}(\boldsymbol{\xi})$ . We shall only consider the simplest form of "bound conditioner", introduced by Veroy et. al. [115, 112], suited exclusively for symmetric coercive operators where  $\sigma_q(\boldsymbol{\xi}) > 0, \forall \boldsymbol{\xi} \in \Xi$ . Given this rather restrictive conditions, the lower bound for the coercivity parameter may be expressed as

$$\widehat{\alpha}(\boldsymbol{\xi}) = \min_{q=\{1,\dots,Q\}} \frac{\sigma_q(\boldsymbol{\xi})}{\sigma_q(\bar{\boldsymbol{\xi}})} \alpha(\bar{\boldsymbol{\xi}}) \quad (2.31)$$

for a certain value of  $\bar{\boldsymbol{\xi}}$ . For problems where the positivity condition of  $\sigma_q(\boldsymbol{\xi})$  is not satisfied, or for more general nonsymmetric it becomes more expensive to compute these bounds. This issue is later discussed.

## Offline Stage

The offline stage, performed only *once*, is computationally very expensive, since it depends on the dimension  $\mathcal{N}$  of the underlying high-order discontinuous finite element space  $\mathbb{X}$ . The key part is that the offline stage computes all quantities independent of the parameter  $\boldsymbol{\xi}$ . Firstly, the matrices and vectors corresponding to the bilinear and linear forms are precomputed, that is  $\mathbb{A}_q, 1 \leq q \leq Q, \mathbb{F}_q, 1 \leq q \leq \widetilde{Q}, \mathbb{L}$  and the matrix of the inner product  $\mathbb{X}$ . In addition, we also calculate the matrices involved in the computation of the residuals (2.29)-(2.30),  $\mathbb{F}_q^T \mathbb{X}^{-1} \mathbb{F}_{q'}, 1 \leq q, q' \leq \widetilde{Q}, \mathbb{F}_q^T \mathbb{X}^{-1} \mathbb{A}_{q'}, 1 \leq q \leq \widetilde{Q}, 1 \leq q' \leq Q, \mathbb{L}^T \mathbb{X}^{-1} \mathbb{L}, \mathbb{L}^T \mathbb{X}^{-1} \mathbb{A}_q, 1 \leq q \leq Q$ .

Secondly, we start solving for  $\zeta_m, \zeta_{m_d}^d$  for appropriate values of the parameter set<sup>4</sup>. Furthermore, on the offline stage we also compute  $\Phi^T \mathbb{A}_q \Phi, \Phi_d^T \mathbb{A}_q \Phi_d, 1 \leq q \leq Q, \mathbb{L}^T \Phi, \Phi^T \mathbb{F}_q, \Phi_d^T \mathbb{F}_q, 1 \leq q \leq \widetilde{Q}$  and finally  $\Phi^T \mathbb{A}_{q'} \mathbb{X}^{-1} \mathbb{A}_q \Phi, \Phi_d^T \mathbb{A}_{q'} \mathbb{X}^{-1} \mathbb{A}_q \Phi_d, 1 \leq q, q' \leq Q$ . Note that these structures are updated on-the-fly as more elements in the basis are being incorporated. The computational intensity is clear for this second

---

<sup>4</sup>The choice of such values is later addressed



set of matrices, since it needs  $M + M_d$  expensive high-order finite element solutions,  $\mathcal{O}(QM^2 + QM_d^2)$  high-order discontinuous finite element vector inner products and  $\mathcal{O}(Q^2M^2 + Q^2M_d^2)$  high-order discontinuous finite element vector inner products with the inverse of the inner product matrix  $\mathbb{X}$ .

## Online stage

The online stage will in general be performed multiple times -for each new value of the parameter  $\boldsymbol{\xi} \in \Xi$ - in the context of uncertainty quantification, where a great number of independent deterministic solves of the problem are needed. Once the offline quantities have been precomputed and stored, each online solve is very simple. For each  $\boldsymbol{\xi}$  values of  $\sigma_q(\boldsymbol{\xi}), \tilde{\sigma}_q(\boldsymbol{\xi})$  are known, therefore systems (2.26) are readily assembled and solved for  $\lambda, \lambda_d$ . These values are used to compute the enhanced output for the new sample using (2.27), just by assembling the precomputed units. Moreover, the *a posteriori* estimates of the error may be easily obtained assembling (2.29)-(2.30) and evaluating for  $\lambda, \lambda_d$ , and the lower bound of the coercivity constant is obtained using (2.31).

For each new parameter, the operation count is  $\mathcal{O}(QM^2 + \tilde{Q}M + QM_d^2)$  to assemble the systems,  $\mathcal{O}(M^3 + M_d^3)$  to invert them and  $\mathcal{O}(M + \tilde{Q}M_d + QMM_d)$  to evaluate the enhanced output. It should be pointed out that the systems to invert are full, since whichever sparsity structure arises from the PDE is no longer preserved after the Galerkin projection. Even though the systems are full, they are very well-conditioned if the orthogonalization procedure described below is carried out for the members of the basis. The online cost for the *a posteriori* error bounds is  $\mathcal{O}(Q^2M^2 + Q\tilde{Q}M + Q^2M_d^2 + QM_d)$ . The power of these bounds is that for any new parameter we can inexpensively evaluate a sharp error bound of the real error in the output, therefore assessing the quality of our reduced-basis.

The essential point of this offline-online strategy is to eliminate the  $\mathcal{N}$  contamination in the online stage, which leads to large computational savings if the problem

needs to be solved for a number of parameters. For cases where  $M, M_d \ll \mathcal{N}$ , the reduced-basis approach leads to computational savings of several orders of magnitude relative to classical standard finite-element (or high-order) approaches.

### 2.1.7 Orthogonalization

The reduced-basis is computed by solving the underlying PDE (and its adjoint) on a certain set of parameters. Each function in the basis is a solution of the very same PDE for a different parametric value, therefore it is reasonable to think that they have a strong linear dependency. This dependency translates into the matrix  $\Phi$  (2.10) (and the one formed by the adjoint solutions  $\Phi_d$ ) being very ill-conditioned. This ill-conditioning, which usually grows exponentially with the size of the basis, introduces an additional penalty in solving systems (2.26), which inherit the conditioning properties of the reduced-basis, see [76].

To overcome this drawback, Gram-Schmidt orthonormalization is applied to the original reduced-basis, thus conserving the approximation properties and improving the conditioning of the subsequent systems. The basis is orthonormalized with respect to the inner product defined by our high-order discontinuous finite element space  $X$ , i.e.

$$(\zeta_i, \zeta_j)_X = \delta_{ij}, \quad 1 \leq i, j, \leq M \quad (2.32a)$$

$$(\zeta_i^d, \zeta_j^d)_X = \delta_{ij}, \quad 1 \leq i, j, \leq M_d \quad (2.32b)$$

The orthonormalization procedure is applied in the offline stage, since systems (2.26) must be solved for all possible parameters at each iteration of the greedy algorithm.

### 2.1.8 Sampling Strategy

In order to construct the reduced-basis using snapshots that contain the more information about the parametrized PDE, we review here the greedy algorithm [40, 38,

112, 114] to find an optimal set of samples. We shall begin by discretizing the parameter space  $\Xi$  by selecting a collection of points  $\Theta$ , to which the greedy algorithm will be applied. The algorithm relies on the *a posteriori* error bounds introduced above, and is applied simultaneously to both the primal and adjoint problem separately, therefore we shall start with two equal set of points,  $\Theta^p, \Theta^d$ .

To start the algorithm, any point on the discretized sets is picked for both problems (usually the same for simplicity). For these value, the discretized PDEs (2.4),(2.16) are solved and the solutions are normalized with respect to  $(\cdot, \cdot)_X$ . The greedy procedure is now applied, by visiting every point in both discretizations and computing the dual norm of the residual (2.21) and a lower bound for the coercivity constant for both problems. The points (one in the primal and one in the adjoint parameter set) with greatest *a posteriori* error bound are incorporated to the reduced-basis, i.e. the exact solution is found and orthonormalized. The reduced-basis structures corresponding to the offline stage are updated with the newly computed values. The algorithm iterates until both the primal and the adjoint *a posteriori* error bounds for all the points are below a prescribed tolerance  $\varepsilon_{\text{tol}}$ .

The key point of the greedy algorithm is how the parameter set is discretized. The obvious, but naive choice is to do uniform gridding in all dimensions. Nonetheless, the number of points grows exponentially with the number of dimensions, thus if we seek a reduced-basis approach for a problem with multiple parameters (typically  $N \geq 8$ ), the dimensional explosion of the tensor product grid will hinder the computation on the offline stage, even if the error bounds are cheap to evaluate.

The approach we propose is to use sparse grids for the greedy algorithm. The sparse grids are used for high-dimensional integration and interpolation. They were introduced by Smolyak [108], although they have been further developed, analysed and implemented in an efficient way by Gerstner, Griebel, Novak et.al. [30, 31, 41, 96, 97, 98]. The Smolyak algorithm is a linear combination of product formulas in a way that the resulting number of points in the set is significantly smaller than that of the tensor product grids.

Let  $\mathbf{i} = (i_1, \dots, i_N) \in \mathbb{N}^N$  be a multiindex. The general formula for computing a sparse grid in  $N$  dimensions of level  $p$  reads

$$\Theta_N \equiv \bigcup_{p+1 \leq |\mathbf{i}| \leq N+p} (\Theta_1^{i_1} \times \dots \times \Theta_1^{i_N}) \quad (2.33)$$

where  $\Theta_1^i$  are one dimensional nodal sets. Details on how to construct sparse grids may be found in [108]. These one dimensional nodal sets are usually defined as the Clenshaw-Curtis or Gaussian quadrature points, especially for interpolation and integration purposes. Furthermore, it is known that the sparse grid of level  $l$  integrates exactly  $N$ -variate polynomials of degree at most  $p$  (represented as  $\Pi_N^p$ ). Furthermore, the size of  $\Theta_N$  for  $N \gg 1$  is of order  $2^p N^p / p!$ , hence it has roughly  $2^p$  more points than the dimension of  $\Pi_N^p$ .

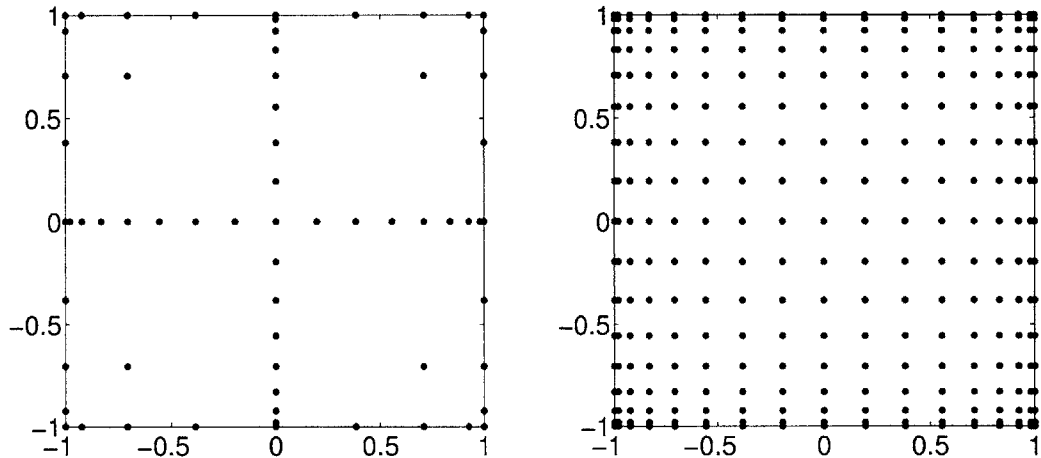


Figure 2-2: Two dimensional sparse grid using Clenshaw-Curtis points. Left: Sparse grid of level 4, total number of points 65. Right: Tensor grid using the same one-dimensional nodes, total number of points 289.

The goal of using sparse grids for the greedy algorithm is to perform a cheaper yet accurate exploration of the parameter space, enabling us to deal with moderate-dimensional problems. Furthermore, in the context of stochastic PDE, depending on the probability distribution of the random variables, different choices of one-dimensional nodes can be made: Clenshaw-Curtis/Gauss-Legendre for uniform, Gauss-

Hermite for Gaussian, Gauss-Laguerre for exponential, etc.

Another possible choice could have been a set of points obtained with low-discrepancy sequences or stratified sampling. Both approaches have been tested for the numerical examples presented in Chapter 4, and no significant improvement with respect to sparse grids has been observed. Furthermore, sparse grids offer not only an efficient exploration of the space (in the absence of localized singularities), but also the possibility to compute the reduced-basis on the same set of points used by stochastic collocation methods, enabling to perform an easy and yet meaningful comparison.

### 2.1.9 Uncertainty Propagation

The procedure described above represents the offline stage. The reduced-basis resulting is essentially a surrogate of the real input-output map, since for every  $\boldsymbol{\xi}$  an approximation  $\tilde{s}_{M, M_d}(\boldsymbol{\xi})$  to  $s(\boldsymbol{\xi})$  may be computed inexpensively. The purpose is to use this reduced-basis surrogate to propagate the uncertainty in a stochastic PDE. The approach pursued here is applying Monte Carlo techniques to exhaustively sample the parameter space  $\Xi$  and get the output response by performing an inexpensive surrogate solve. The main bottleneck of Monte Carlo methods is their slow convergence rate -partially alleviated by Quasi-Monte Carlo and Latin Hypercube Sampling- thus requiring an enormous number of full model solutions to acquire good statistics of the quantity of interest. However, if a cheap surrogate model is employed, many more evaluations can be performed within a reasonable time. This approach was first used by Boyaval et.al. for a heat conduction problem [10], and by Haasdonk for parametric uncertainty in elliptic PDEs arising from KL expansions [43]. Furthermore, a variance reduction technique using control variates has been developed and applied to heat conduction problems [9], which is interesting when multiple online stages are needed for different control parameters. However, we shall consider only Monte Carlo methods.

The procedure is fairly simple. For every new sample  $\boldsymbol{\xi}_j$ , systems (2.26) are solved

for  $\lambda, \lambda_d$  and the enhanced output (2.27) is evaluated. Note that this procedure, which is nothing but a Monte Carlo method applied on a surrogate input-output map, is embarrassingly parallel. Once we have performed the surrogate evaluation for each sample, statistics of the output can be readily obtained. Furthermore, the output's probability density function (PDF) may be approximately computed using either histograms or kernel density estimation techniques.

In general, we shall expect sparse grids of a certain level to provide accurate surrogate models for smooth input-output functions. Nonetheless, localized discontinuities in the parameter space, or phenomena that are not aligned with the cartesian axis may be missed by a coarse sparse grid. The introduction of adaptivity and anisotropy to the sparse grid is as an attractive choice for problems where the latter may occur. Much research has been devoted to adaptive sparse grids for integration and interpolation [31, 41] and for uncertainty propagation [28, 61, 92], however adaptive sparse grids are beyond the scope of this work.

## 2.2 Noncoercive linear operator: Helmholtz problem

### 2.2.1 Abstract Formulation

Let us consider a parametrized PDE in its weak form

$$a(u^e(\boldsymbol{\xi}), v; \boldsymbol{\xi}) = f(v; \boldsymbol{\xi}), \quad \forall v \in X^e \quad (2.34)$$

where  $u^e(\boldsymbol{\xi})$  is the exact solution (the superscript  $e$  represents exact) of the PDE and  $\boldsymbol{\xi} \in \Xi \subset \mathbb{R}^N$  represents the parametric dependence of the PDE, that can also be seen as the stochastic parameters defined on a certain closed space  $\Xi$ . The problem that we aim to solve reads: find

$$s^e(\boldsymbol{\xi}) = l(u^e(\boldsymbol{\xi})) \quad (2.35)$$

where  $s^e(\boldsymbol{\xi})$  is the (exact) quantity of interest;  $u^e(\boldsymbol{\xi})$  is the (exact) field variable;  $X^e$  is an associated Hilbert space defined over a suitably regular physical domain  $D \in \mathbb{R}^d$ ,  $d = 1, 2, 3$  (independent of the parameter space  $\Xi$ ), where we can define an associated inner product  $(w, v)_{X^e}$  and norm  $\|w\|_{X^e} = \sqrt{(w, w)_{X^e}}$ ; and finally  $a(\cdot, \cdot)$  and  $f(\cdot), l(\cdot)$  are  $X$ -continuous bilinear and linear forms, respectively. We also require the linear functionals to be bounded. Note that for the Helmholtz problem we may usually consider complex-valued fields. The function space required must satisfy  $X^e \subset (H^1(D))^\nu$ , depending on whether the field variable  $u^e(\boldsymbol{\xi})$  is scalar  $\nu = 1$  or vector  $\nu = d$ .

For approximation purposes, the infinite-dimensional space  $X^e$  is replaced by a high-order discontinuous finite element approximation space  $X$ ,  $X \subset X^e$ , of dimension  $\mathcal{N}$ . We shall denote  $(\cdot, \cdot)_X$  and  $\|\cdot\|_X$  the inner product and the norm associated with the finite element space  $X$ . The dual functional space  $X'$  is given by

$$\|f\|_{X'} \equiv \sup_{v \in X} \frac{f(v)}{\|v\|_X}, \quad \forall f \in X' \quad (2.36)$$

Furthermore, we shall assume that the bilinear form satisfies a continuity and an inf-sup condition for all parametric values.

$$\beta(\boldsymbol{\xi}) \equiv \inf_{w \in X} \sup_{v \in X} \frac{a(w, v; \boldsymbol{\xi})}{\|w\|_X \|v\|_X}, \quad (2.37a)$$

$$\gamma(\boldsymbol{\xi}) \equiv \sup_{w \in X} \sup_{v \in X} \frac{a(w, v; \boldsymbol{\xi})}{\|w\|_X \|v\|_X}. \quad (2.37b)$$

The constant  $\beta(\boldsymbol{\xi})$  is the minimum singular value associated with the differential operator, known also as the inf-sup stability constant, and  $\gamma(\boldsymbol{\xi})$  is the continuity constant. We shall assume that these constants are bounded for any  $\boldsymbol{\xi} \in \Xi$ , that is  $\beta(\boldsymbol{\xi}) \geq \beta_0 > 0$  and  $\gamma(\boldsymbol{\xi}) \leq \gamma_0 < \infty$ .

For computational efficiency we also assume that parametric dependence of  $a, f$  may be expressed, for some small integers  $Q, \tilde{Q}$ , as (2.8) for some complex-valued functions  $\sigma_q, \tilde{\sigma}_q : \Xi \rightarrow \mathbb{C}$  continuous, differentiable and parameter-dependent, whereas

the forms  $a_q : X \times X \rightarrow \mathbb{C}$ ,  $f_q : X \rightarrow \mathbb{C}$  are parameter-independent.

## 2.2.2 Reduced-Basis Approach

The reduced-basis approach pursued here is the same as below, hence the details will not be discussed. For the reduced-basis spaces  $W_M$ ,  $W_{M_d}^d$ , Galerkin projections are applied to obtain the reduced-basis solutions for both the primal and the adjoint problem  $u_M(\boldsymbol{\xi}) \in W_M$ ,  $\psi_{M_d}(\boldsymbol{\xi}) \in W_{M_d}^d$  using expressions (2.17). The enhanced output is then evaluated using (2.18). However, for the noncoercive case the discrete inf-sup parameter associated with the operator may not guarantee stability. There are more sophisticated techniques, such as minimum-residual [105, 65], and in particular Petrov-Galerkin approaches, which restore stability at the expense of additional computational complexity. For the numerical examples presented in this work only Galerkin projections will be considered.

## 2.2.3 *A Priori* Convergence Results

Optimal convergence rates of the primal and dual solutions  $u_M(\boldsymbol{\xi}) \rightarrow u(\boldsymbol{\xi})$ ,  $\psi_{M_d}(\boldsymbol{\xi}) \rightarrow \psi(\boldsymbol{\xi})$  and the enhanced output  $\tilde{s}_{M,M_d}(\boldsymbol{\xi}) \rightarrow s(\boldsymbol{\xi})$  have the following expressions

$$\|u(\boldsymbol{\xi}) - u_M(\boldsymbol{\xi})\|_X \leq \left(1 + \frac{\gamma_0}{\beta_0}\right) \inf_{w_M \in W_M} \|u(\boldsymbol{\xi}) - w_M(\boldsymbol{\xi})\|_X, \quad (2.38a)$$

$$\|\psi(\boldsymbol{\xi}) - \psi_{M_d}(\boldsymbol{\xi})\|_X \leq \left(1 + \frac{\gamma_0}{\beta_0}\right) \inf_{w_{M_d} \in W_{M_d}^d} \|\psi(\boldsymbol{\xi}) - w_{M_d}(\boldsymbol{\xi})\|_X. \quad (2.38b)$$

Using a similar argument than that of (2.20), optimal convergence of the enhanced output reads

$$|s(\boldsymbol{\xi}) - \tilde{s}_{M,M_d}(\boldsymbol{\xi})| \leq \gamma_0 \left(1 + \frac{\gamma_0}{\beta_0}\right)^2 \inf_{w_M \in W_M} \|u(\boldsymbol{\xi}) - w_M(\boldsymbol{\xi})\|_X \inf_{w_{M_d} \in W_{M_d}^d} \|\psi(\boldsymbol{\xi}) - w_{M_d}(\boldsymbol{\xi})\|_X. \quad (2.39)$$

Note that the enhanced output converges to the true output as the product of the errors in the primal and in the dual, thus we maintain the square effect observed for



the coercive case.

### 2.2.4 *A Posteriori* Error Estimation

The error bounds are presented assuming that a parameter-dependent lower bound  $\widehat{\beta}(\boldsymbol{\xi})$  of the inf-sup stability constant  $\beta(\boldsymbol{\xi})$ , such that  $\beta(\boldsymbol{\xi}) \geq \widehat{\beta}(\boldsymbol{\xi}) \geq \beta_0 > 0$ ,  $\forall \boldsymbol{\xi} \in \Xi$  may be computed.

The dual norm of the residual for both the primal and the adjoint problem is given by (2.21), and the enhanced output error estimator may be computed as

$$\Delta_{M, M_d}^s(\boldsymbol{\xi}) = \frac{\varepsilon_M(\boldsymbol{\xi}) \cdot \varepsilon_{M_d}^d(\boldsymbol{\xi})}{\widehat{\beta}(\boldsymbol{\xi})}. \quad (2.40)$$

These bounds are sharp and rigorous.

### 2.2.5 Computational Procedure

The computational strategy is the same as for the coercive case. The main difference is that conjugate transpose operator needs to be used for the adjoint, since the forms can in general be complex-valued. The reduced-basis solutions  $u_M(\boldsymbol{\xi})$ ,  $\psi_{M_d}(\boldsymbol{\xi})$  may be expressed as a linear combination of the reduced-basis functions as (2.25). Again, Galerkin projections (2.17) are used for both the primal and the adjoint problem, thus the latter coefficients  $\lambda, \lambda_d$  satisfy the following  $M \times M, M_d \times M_d$  systems

$$\Phi^* \left[ \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{A}_q \right] \Phi \lambda = \Phi^* \left[ \sum_{q=1}^{\tilde{Q}} \tilde{\sigma}_q(\boldsymbol{\xi}) \mathbb{F}_q \right] \quad (2.41a)$$

$$\Phi_d^* \left[ \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{A}_q^* \right] \Phi_d \lambda_d = -\Phi_d^* \mathbb{L} \quad (2.41b)$$

recalling the affine dependency of the parameter for  $a(\cdot, \cdot)$ ,  $f(\cdot)$ . The reduced-basis enhanced output is evaluated as

$$\tilde{s}_{M, M_d}(\boldsymbol{\xi}) = \mathbb{L}^* \Phi \lambda - (\Phi_d \lambda_d)^* \left( \left[ \sum_{q=1}^{\tilde{Q}} \tilde{\sigma}_q(\boldsymbol{\xi}) \mathbb{F}_q \right] - \left[ \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{A}_q \right] \Phi \lambda \right). \quad (2.42)$$

Furthermore, due to the loss of symmetry of the operator the dual norm of the residuals change, and may be computed as

$$\begin{aligned} \varepsilon_M(\boldsymbol{\xi}) &= \sum_{q=1}^{\tilde{Q}} \sum_{q'=1}^{\tilde{Q}} \tilde{\sigma}_q(\boldsymbol{\xi}) \tilde{\sigma}_{q'}^*(\boldsymbol{\xi}) \mathbb{F}_{q'}^* \mathbb{X}^{-1} \mathbb{F}_q - \sum_{q=1}^{\tilde{Q}} \sum_{q'=1}^Q \tilde{\sigma}_q^*(\boldsymbol{\xi}) \sigma_{q'}(\boldsymbol{\xi}) \mathbb{F}_q^* \mathbb{X}^{-1} \mathbb{A}_{q'} \Phi \lambda - \\ &\quad - \sum_{q=1}^{\tilde{Q}} \sum_{q'=1}^Q \tilde{\sigma}_q(\boldsymbol{\xi}) \sigma_{q'}^*(\boldsymbol{\xi}) (\Phi \lambda)^* \mathbb{A}_{q'}^* \mathbb{X}^{-1} \mathbb{F}_q + \sum_{q=1}^Q \sum_{q'=1}^Q \sigma_q(\boldsymbol{\xi}) \sigma_{q'}^*(\boldsymbol{\xi}) (\Phi \lambda)^* \mathbb{A}_{q'}^* \mathbb{X}^{-1} \mathbb{A}_q \Phi \lambda, \end{aligned} \quad (2.43)$$

for the primal and

$$\begin{aligned} \varepsilon_{M_d}^d(\boldsymbol{\xi}) &= \mathbb{L}^* \mathbb{X}^{-1} \mathbb{L} - \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{L}^* \mathbb{X}^{-1} \mathbb{A}_q \Phi_d \lambda_d - \sum_{q=1}^Q \sigma_{q'}^*(\boldsymbol{\xi}) (\Phi_d \lambda_d)^* \mathbb{A}_{q'}^* \mathbb{X}^{-1} \mathbb{L} \\ &\quad + \sum_{q=1}^Q \sum_{q'=1}^Q \sigma_q(\boldsymbol{\xi}) \sigma_{q'}^*(\boldsymbol{\xi}) (\Phi_d \lambda_d)^* \mathbb{A}_{q'}^* \mathbb{X}^{-1} \mathbb{A}_q \Phi_d \lambda_d, \end{aligned} \quad (2.44)$$

for the adjoint. Furthermore, the same offline-online computational strategy presented above is pursued here, taking into account the minor modification aforementioned due to the nonsymmetry of the operator.

Finally, it only remains to compute the lower bound  $\hat{\beta}(\boldsymbol{\xi})$  of the inf-sup stability constant  $\beta(\boldsymbol{\xi})$ . Extensive description of methods for computing  $\hat{\beta}(\boldsymbol{\xi})$  may be found in [76, 114]. In general, it is not a trivial task, since it involves developing a reduced-scheme for a generalized eigenvalue problem that allows to compute a lower bound of the inf-sup constant. Another technique is the Successive Constraint Method, introduced by Huynh et.al. [46]. Although both approaches allow for an offline-online stage decomposition, there is an increase in computational cost. The computation

of the inf-sup constant is especially important in the Helmholtz equation, since its value approaches zero near resonances. The implementation of efficient computation of the inf-sup constant is left as future work, since the example presented in Chapter 4 avoids the problem of resonances. Thus, the bounds obtained will be neither rigorous nor sharp, hence they will be merely used as error indicators to drive the sampling.

### 2.2.6 Sampling Strategy

The greedy algorithm will be employed as before on sparse grids  $\Theta^p$ ,  $\Theta^d$  (one for the primal and the other for the adjoint), using only the dual norm of the residuals as error indicators for the sampling. Indeed, since we lack the estimation of a lower bound for the inf-sup constant, the termination criterion will not be related to the error bound. Instead, a more inefficient but simpler technique is used.

The solution for an arbitrary set, or test set, of parameters  $\tilde{\Theta}$ , randomly chosen, i.e. drawn from the underlying probability density function using a pseudo-random generator, are precomputed and stored. During the execution of the greedy, we (1) choose the point that contributes most to the error (for the primal and the adjoint separately) by evaluating *a posteriori* error estimates for all points in  $\Theta^p$ ,  $\Theta^d$ , (2) solve the full primal and adjoint PDE for the primal and adjoint parameters, (3) update the reduced-basis orthogonalizing the new snapshots, (4) compute the absolute error between the true output  $s(\boldsymbol{\xi})$  and the approximate enhanced output  $\tilde{s}_{M,M_d}(\boldsymbol{\xi})$  for all the points  $\boldsymbol{\xi}$  in  $\tilde{\Theta}$ . The algorithm terminates whenever the  $\|s(\boldsymbol{\xi}) - \tilde{s}_{M,M_d}(\boldsymbol{\xi})\|_\infty < \varepsilon_{\text{tol}}$ ,  $\forall \boldsymbol{\xi} \in \tilde{\Theta}$ .

This criterion is obviously not optimal, since it involves the solution of the full problem for a number of parameters, which is precisely what the reduced-basis approach seeks to avoid. However, we shall use it as a heuristic approach until the computation of the inf-sup stability constant is implemented.

### 2.2.7 Uncertainty Propagation

Once the surrogate has been constructed (offline stage), we proceed to the online stage for the uncertainty propagation. The procedure relies on applying Monte Carlo techniques to obtain independent samples, and for each sample solve equations (2.41) for the coefficients  $\lambda$ ,  $\lambda_d$  and then recover the approximate quantity of interest (2.42).

In the current status, since the *a posteriori* error bounds computed are neither rigorous not sharp, the reliability of the surrogate for any new sample cannot be evaluated. For the numerical examples presented in Chapter 4, we will rely solely on the capacity of the sparse grid to explore the most important regions of the parameter space. Comparison with other methods are provided to test the performance of the reduced-basis surrogate.

## Chapter 3

# The Hybridizable Discontinuous Galerkin Method

The hybridizable Discontinuous Galerkin (HDG) method is a new emerging DG approach firstly introduced by Cockburn, Gopalakrishnan and Lazarov in 2009 [19] and further analyzed and developed by Nguyen, Peraire and Cockburn in [79, 80, 81, 83, 84, 82, 85, 86, 87, 88, 78, 77]. This method generalizes the classic Discontinuous Galerkin methods by introducing hybrid variables at the faces of the elements, decoupling the interaction between neighboring elements. Once the problem is decoupled, a reduced global problem is solved to find the global variables, which are then used for recovering the local variables element-wise.

In this chapter the HDG method will be presented as the method to which the reduced-basis approach will be applied for solving the PDE. The specifications of the method will be provided together with a detailed formulation for the numerical examples that are introduced in the following chapter. Furthermore, since the reduced-basis method employs the adjoint equation, we will also address the issue of how to compute it efficiently using the HDG method.

### 3.1 Why HDG?

The basic and key idea of the HDG method resides in introducing new variables on the faces called the numerical traces, which become the globally coupled variables of the problem. The numerical fluxes for the elemental problems are defined in terms of these traces, involving an additional stabilization parameter  $\tau$ . In the end, these new variables are such that decouple the interaction among neighboring elements and thus, the problem can be locally solved very efficiently. At the end, a global problem is solved for the numerical traces. It must be said that this new set of variables is smaller (especially when going to high order polynomial approximation spaces) than the original set of DG variables, since they are only defined on the edges. So we end up solving a number of inexpensive local problems and just one global problem for variables which are defined only on the edges, and is therefore cheaper than a globally coupled DG.

The Finite Element Method (FEM) has been a popular choice to obtain the solution for PDEs given its ability to handle complex geometries and interface/boundary conditions. The Finite Differences (FD) method would not offer the desired flexibility for complicated geometries, and also present notorious difficulties for the treatment of boundary conditions and the extension to high-order. This latter drawback is also a limitation for Finite Volumes (FV) methods. Nonetheless, it should be remarked that the FV and HDG methods share the property of being formulated as conservative schemes, and are therefore adequate for systems of conservation laws. Boundary Integral methods have been discarded due to the fact that the resulting matrices are dense and its inability to handle nonlinear problems.

Once the FE method has been chosen there are still several spatial discretization strategies that can be considered. They include continuous Galerkin/Petrov-Galerkin methods, spectral element methods, mixed finite element methods, extended finite element methods and finally discontinuous Galerkin/Petrov-Galerkin methods. They could all have been used and they all have their strengths and weaknesses. However,

due to its ability to combine complex geometry and high order solutions the Discontinuous Galerkin Finite Element methods, seem to be most suitable. They also offer stability and low dispersion for discretizations of hyperbolic systems, allow for a simple imposition of boundary conditions and are very flexible to future parallelization and adaptivity.

One of the main drawbacks of DG methods is the duplication of nodal degrees of freedom at the element boundary interfaces. This limitation is overcome in the HDG method thanks to the introduction of the numerical traces at the element interfaces, leading to block diagonal systems of equations can be very efficiently solved in a local sense (only need to invert local matrices which are small).

For time dependent problems, time integration can be carried out either using the well-known class of Newmark-finite element methods or just transforming all higher order time dependent semidiscretized PDEs into first order systems of ODEs that can be efficiently time integrated using either a backwards difference scheme or a Runge-Kutta. However, in this document we will consider only equations with no time dependency.

## 3.2 The Helmholtz equation

Firstly, we present the HDG method for the Helmholtz equation to introduce the notation and the machinery necessary for this class of methods. Once the method has been properly described, we will show its adaptation to the reduced-basis method and derive the adjoint equation. The method is presented for this equation because it will be of importance later on, and the fact that it will be adapted to the reduced-basis approach requires insight on the nature and features of HDG.

The HDG method has already been introduced for the linear and nonlinear convection diffusion equations by Nguyen, Peraire and Cockburn in [79, 80] respectively, and by Saà-Seoane for the Helmholtz equation in [106], and here we follow closely its

discussion. As it is typical for the HDG methods, in order to carry out the discretization in space we basically proceed in two main steps. First of all, we formulate and solve the local problems where the approximate scalar variable and flux are expressed in an element-by-element fashion in terms of an approximate trace of the scalar variable along the element boundary. Then, we formulate and solve the global problem, which is just obtaining a unique value for the trace at the interfaces by enforcing flux continuity. Figure 3-1 shows the extra degrees of freedom considered in the HDG ( $\hat{u}_h$ ) versus the classical DG variables. Note that they are all on the boundaries and they are used to decouple all local problems and are then obtained by solving the global problem.

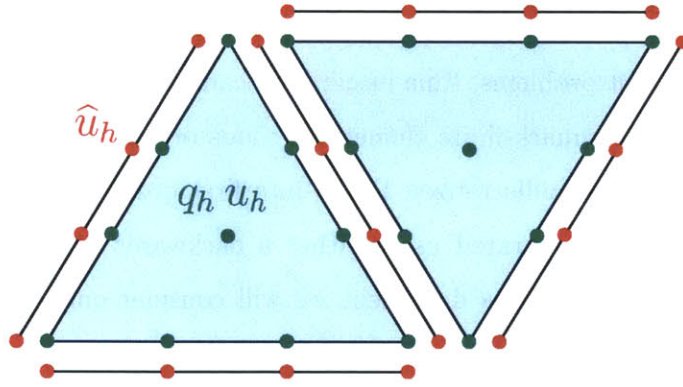


Figure 3-1: Degrees of freedom considered by the HDG method for degree  $p = 3$ . Classical DG schemes do not consider  $\hat{u}_h$  degrees of freedom

Consider now the Helmholtz equation, where  $D \in \mathbb{R}^d$  represents the physical domain with boundary  $\Gamma$ , and  $\mathbf{x}$  the spatial variable. Then the strong formulation of the Helmholtz equation is the one in (3.1).

$$-\nabla(\rho\nabla u) - k^2 u = h \quad \forall \mathbf{x} \in D, \quad (3.1a)$$

$$\rho\nabla u \cdot \mathbf{n} = u g_R \quad \forall \mathbf{x} \in \Gamma_R, \quad (3.1b)$$

$$\rho\nabla u \cdot \mathbf{n} = g_N \quad \forall \mathbf{x} \in \Gamma_N, \quad (3.1c)$$

where  $k = \omega/c$  is the wavenumber, obtained as a quotient of the frequency of the wave and the sound speed of the medium  $c$ ,  $\rho(\mathbf{x})$  represents the heterogeneity in the



propagation medium and  $h$  the source term generating the perturbation. The details of the equation are given in Chapter 4. The boundaries are represented by  $\Gamma_R$  and  $\Gamma_N$  where Robin and Neumann boundary conditions are applied, respectively represented by  $g_R, g_N$ . They are such that  $\bar{\Gamma}_R \cup \bar{\Gamma}_N = \Gamma$  and  $\Gamma_R \cap \Gamma_N = \emptyset$ .

### 3.2.1 HDG formulation

The first step is to rewrite our problem (3.1) as a first order system of equations. To that end, we shall define an auxiliary variable  $\mathbf{q} = \nabla u$  corresponding to the gradient of the state field. The system with the new variable shapes

$$\mathbf{q} - \nabla u = 0 \quad \forall \mathbf{x} \in D, \quad (3.2a)$$

$$-\nabla \cdot (\rho \mathbf{q}) - k^2 u = h \quad \forall \mathbf{x} \in D, \quad (3.2b)$$

$$\rho \mathbf{q} \cdot \mathbf{n} = u g_R \quad \forall \mathbf{x} \in \Gamma_R, \quad (3.2c)$$

$$\rho \mathbf{q} \cdot \mathbf{n} = g_N \quad \forall \mathbf{x} \in \Gamma_N. \quad (3.2d)$$

### 3.2.2 Notation

Let  $\mathcal{T}_h$  be such that  $\mathcal{T}_h = \bigcup_{i=1}^n K_i^1$  and  $\mu_d(K_i \cap K_j) = \delta_{ij}^2$  where  $\dim(D) = d$ , that is, the intersection of two different elements of  $\mathcal{T}_h$  (also called discretization space) can only contain a face. If  $\dim(D) = d$ , such intersection can only be, at most, of dimension  $d - 1$ . In fact, if  $\mu_{d-1}(K_i \cap K_j) \neq 0$  it means that  $K_i$  and  $K_j$  are neighbors and therefore we define their common interface as  $F_{ij}^o = K_i \cap K_j = \partial K_i \cap \partial K_j$  and it will be an interior or a boundary face. The set of interior faces is denoted as  $\mathcal{E}_h^o$ . Similarly, the set of elements such that  $\mu_{d-1}(K_i \cap \partial D) \neq 0$  are elements with an edge on the boundary and hence we define such faces through  $F_i^\partial = K_i \cap \partial D = \partial K_i \cap \partial D$ . The set of all boundary faces is denoted as  $\mathcal{E}_h^\partial$ . Finally we consider the set of all faces in the discretization as  $\mathcal{E}_h = \mathcal{E}_h^o \cup \mathcal{E}_h^\partial$ . An example of triangulation is depicted

<sup>1</sup>It can be the case where  $D$  is curved and  $\mathcal{T}_h$  only considers linear elements, therefore  $D \neq \mathcal{T}_h$

<sup>2</sup>Note that  $\mu_d(\cdot)$  indicates the  $d$ -dimension Lebesgue measure of the set.

in Figure 3-2. As usual in DG methods, we need to define the averages  $\{\{\cdot\}\}$  and

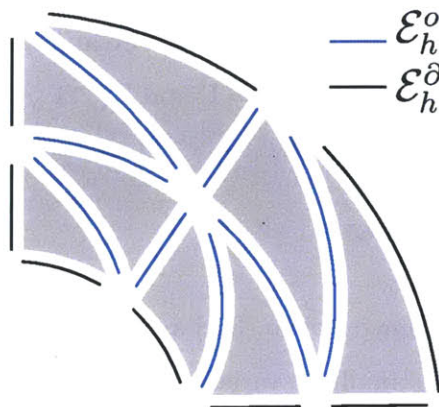


Figure 3-2: Sketch of a high-order triangulation with curved elements. The interior and boundary faces are depicted, together with the triangulation elements  $K$  in grey.

the jumps  $\llbracket \cdot \rrbracket$  on the interior and boundary faces. Consider two neighboring elements  $K^+, K^-$  and their common edge  $\partial K^+ \cap \partial K^-$ . Let then  $(\mathbf{q}^+, u^+)$  be the values of the gradient (vector) and state (scalar) on the face considered as a face of  $\partial K^+$ , and  $(\mathbf{q}^-, u^-)$  be the values of the gradient and displacement on the same edge considered as a face of  $\partial K^-$ . Then we introduce for the interior edges  $F \in \mathcal{E}_h^o$ :

$$\begin{aligned} \{\{\mathbf{q}\}\} &= (\mathbf{q}^+ + \mathbf{q}^-)/2 & \{\{u\}\} &= (u^+ + u^-)/2 \\ \llbracket \mathbf{q} \cdot \mathbf{n} \rrbracket &= \mathbf{q}^+ \cdot \mathbf{n}^+ + \mathbf{q}^- \cdot \mathbf{n}^- & \llbracket u \rrbracket &= u^+ \mathbf{n}^+ + u^- \mathbf{n}^- \end{aligned} \quad (3.3)$$

Note how the average of a vector is a vector and the average of a scalar is a scalar but the jumps are defined for the magnitudes times the normal so for the gradient it becomes a scalar and for the displacement it is a vector. We still need to extend this definition to the boundary edges and we do that as follows. For  $F \in \mathcal{E}_h^\partial$ :

$$\begin{aligned} \{\{\mathbf{q}\}\} &= \mathbf{q} & \{\{u\}\} &= u \\ \llbracket \mathbf{q} \cdot \mathbf{n} \rrbracket &= \mathbf{q} \cdot \mathbf{n} & \llbracket u \rrbracket &= u \cdot \mathbf{n} \end{aligned} \quad (3.4)$$

Finally let's introduce the notation used for the  $L^2$  inner products of functions over elements and faces. For a given  $\mathbf{u}, \mathbf{v} \in [L^2(D)]^d$  and  $u, v \in L^2(D)$ , then we denote

the scalar products in the interior and over the edges as in (3.5)

$$\begin{aligned}
 (\mathbf{u}, \mathbf{v})|_D &= \int_D \mathbf{u} \cdot \mathbf{v} \, dV, & \langle \mathbf{u}, \mathbf{v} \rangle|_{\partial D} &= \int_{\partial D} \mathbf{u} \cdot \mathbf{v} \, dS, \\
 (u, v)|_D &= \int_D u \cdot v \, dV, & \langle u, v \rangle|_{\partial D} &= \int_{\partial D} u \cdot v \, dS.
 \end{aligned} \tag{3.5}$$

### 3.2.3 Approximation spaces

First let  $\mathcal{P}^p(D)$  be the set of polynomials of degree at most  $p$  over a domain  $D$ . The discontinuous finite element spaces are defined in (3.6). Note that we need three spaces: one for the scalar displacement, another for the gradient which is a vector, and yet another scalar space need be defined on the edges of the discretization.

$$W_h^p = \{w \in L^2(D) : w|_K \in \mathcal{P}^p(K), \forall K \in \mathcal{T}_h\}, \tag{3.6a}$$

$$\mathbf{V}_h^p = \{\mathbf{v} \in [L^2(D)]^d : \mathbf{v}|_K \in [\mathcal{P}^p(K)]^d, \forall K \in \mathcal{T}_h\}, \tag{3.6b}$$

$$M_h^p = \{\mu \in L^2(\mathcal{E}_h) : \mu|_F \in \mathcal{P}^p(F), \forall F \in \mathcal{E}_h\}. \tag{3.6c}$$

Recall that if  $u \in L^2(D)$ , then  $\int_D u^2 < +\infty$ . In Figure 3-3 we show the same mesh as before, considering a polynomial order equal to 3, together with the degrees of freedom of the approximation functions living in the spaces (3.6).

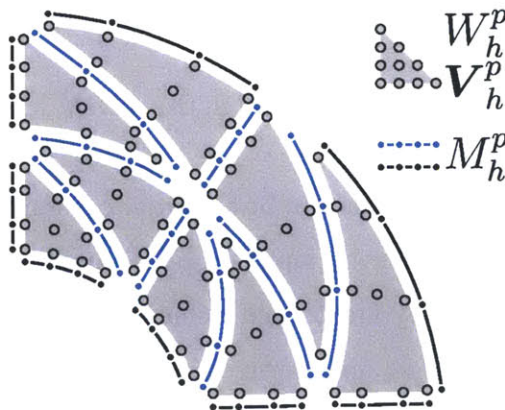


Figure 3-3: Sketch of a triangulation with curved elements of order  $p = 3$ . The degrees of freedom for the corresponding approximation spaces are shown

### 3.2.4 Space discretization

Once the notation and the approximation spaces are introduced, we need to recast the strong formulation in (3.2) into a weak formulation. The weak formulation is obtained by multiplying the strong form of the equations for the FE discretized solution  $(u_h, \mathbf{q}_h)$  by test functions in the corresponding spaces. Using the notation defined above we have

$$(\mathbf{q}_h, \mathbf{v})_K - (\nabla u_h, \mathbf{v})_K = 0 \quad \forall \mathbf{v} \in [\mathcal{P}^p(K)]^d, \quad (3.7a)$$

$$-(\nabla \cdot (\rho \mathbf{q}_h), w)_K - (k^2 u, w)_K = (h, w)_K \quad \forall w \in \mathcal{P}^p(K). \quad (3.7b)$$

Integrating by parts the volume terms containing the nabla operator we obtain (3.8), that is the system satisfied at each element  $K$  of the triangulation  $\mathcal{T}_h$  by our approximation  $(\mathbf{q}_h, u_h) \in \mathbf{V}_h^p \times W_h^p$

$$(\mathbf{q}_h, \mathbf{v})_K + (u_h, \nabla \cdot \mathbf{v})_K - \langle \hat{u}_h, \mathbf{v} \cdot \mathbf{n} \rangle_{\partial K} = 0 \quad \forall \mathbf{v} \in [\mathcal{P}^p(K)]^d, \quad (3.8a)$$

$$(\rho \mathbf{q}_h, \nabla w)_K - \langle \rho \hat{\mathbf{q}}_h \cdot \mathbf{n}, w \rangle_{\partial K} - (k^2 u, w)_K = (h, w)_K \quad \forall w \in \mathcal{P}^p(K). \quad (3.8b)$$

Note how, in the integration by parts, the degrees of freedom at the boundaries have been replaced by the numerical traces or fluxes since the variables themselves are not well defined there. The HDG requires us to define  $\hat{\mathbf{q}}_h$  in terms of  $\mathbf{q}_h, u_h, \hat{u}_h$ .

$$\hat{\mathbf{q}}_h = \mathbf{q}_h - \tau(u_h - \hat{u}_h)\mathbf{n}, \text{ on } \partial K. \quad (3.9)$$

The stabilization parameter  $\tau$  is taken to be a positive constant of order unity. Further analysis on how to choose  $\tau$  can be found in [79]. Equations defined in (3.8) refer to the local problem, since they are defined element-wise. The local problem may be solved if the numerical traces  $\hat{u}_h$  are known. Therefore, we define the global problem for the traces  $\hat{u}_h \in M_h^p$  as

$$\langle \rho \hat{\mathbf{q}}_h \cdot \mathbf{n}, \mu \rangle_{\partial \mathcal{T}_h \setminus \Gamma} + \langle \rho \hat{\mathbf{q}}_h \cdot \mathbf{n} - g_N, \mu \rangle_{\Gamma_N} + \langle \rho \hat{\mathbf{q}}_h \cdot \mathbf{n} - \hat{u}_h g_R, \mu \rangle_{\Gamma_R} = 0 \quad \forall \mu \in \mathcal{P}^p(F). \quad (3.10)$$

Finally by adding up all contributions of (3.8) over the elements on the triangulation  $\mathcal{T}_h$  and using the expression in (3.9), we obtain the global system of equations (3.11): Find  $(\mathbf{q}_h, u_h, \hat{u}_h) \in \mathbf{V}_h^p \times W_h^p \times M_h^p$  such that

$$\langle \mathbf{q}_h, \mathbf{v} \rangle_{\mathcal{T}_h} + (u_h, \nabla \cdot \mathbf{v})_{\mathcal{T}_h} - \langle \hat{u}_h, \mathbf{v} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} = 0, \quad (3.11a)$$

$$\langle \rho \mathbf{q}_h, \nabla w \rangle_{\mathcal{T}_h} - \langle \rho \mathbf{q}_h \cdot \mathbf{n}, w \rangle_{\partial \mathcal{T}_h} - (k^2 u_h, w)_{\mathcal{T}_h} + \langle \rho \tau (u_h - \hat{u}_h), w \rangle_{\partial \mathcal{T}_h} = (h, w)_{\mathcal{T}_h}, \quad (3.11b)$$

$$\langle \rho \mathbf{q}_h \cdot \mathbf{n}, \mu \rangle_{\partial \mathcal{T}_h} - \langle \rho \tau (u_h - \hat{u}_h), w \rangle_{\partial \mathcal{T}_h} - \langle \hat{u}_h g_R, \mu \rangle_{\Gamma_R} = \langle g_N, \mu \rangle_{\Gamma_N}, \quad (3.11c)$$

holds for all  $(\mathbf{v}, w, \mu) \in \mathbf{V}_h^p \times W_h^p \times M_h^p$ .

### 3.2.5 Weak formulation and matrix system

The system of equations in (3.11) is rewritten for convenience in terms of several bilinear forms. The weak formulation reads: find  $(\mathbf{q}_h, u_h, \hat{u}_h) \in \mathbf{V}_h^p \times W_h^p \times M_h^p$  such that

$$p(\mathbf{q}_h, \mathbf{v}) + b(u_h, \mathbf{v}) - c(\hat{u}_h, \mathbf{v}) = 0, \quad (3.12a)$$

$$s(\mathbf{q}_h, w) + d(u_h, w) - e(\hat{u}_h, w) = h(w), \quad (3.12b)$$

$$n(\mathbf{q}_h, \mu) - e(u_h, w) + m(\hat{u}_h, \mu) = g(\mu), \quad (3.12c)$$

holds for all  $(\mathbf{v}, w, \mu) \in \mathbf{V}_h^p \times W_h^p \times M_h^p$ . The bilinear functions are given by

$$\begin{aligned} p(\mathbf{q}, \mathbf{v}) &= (\mathbf{q}, \mathbf{v})_{\mathcal{T}_h}, & b(\eta, \mathbf{v}) &= (\eta, \nabla \cdot \mathbf{v})_{\mathcal{T}_h}, \\ c(\lambda, \mathbf{v}) &= \langle \lambda, \mathbf{v} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h}, & d(u, w) &= -(k^2 u, w)_{\mathcal{T}_h} + \langle \rho \tau u, w \rangle_{\partial \mathcal{T}_h}, \\ e(\lambda, w) &= \langle \rho \tau \lambda, w \rangle_{\partial \mathcal{T}_h}, & m(\eta, \mu) &= \langle \rho \tau \eta, \mu \rangle_{\partial \mathcal{T}_h} - \langle \eta g_R, \mu \rangle_{\Gamma_R}, \\ g(\mu) &= \langle g_N, \mu \rangle_{\Gamma_N}, & s(\mathbf{v}, \eta) &= (\rho \mathbf{v}, \nabla \eta)_{\mathcal{T}_h} - \langle \rho \mathbf{v} \cdot \mathbf{n}, \eta \rangle_{\partial \mathcal{T}_h}, \\ h(w) &= (h, w)_{\mathcal{T}_h}, & n(\mathbf{v}, \mu) &= \langle \rho \mathbf{v} \cdot \mathbf{n}, \mu \rangle_{\partial \mathcal{T}_h}. \end{aligned} \quad (3.13)$$

The discretization of the system of equations given in (3.12) is expressed by the matrix equation denoted in (3.14).

$$\begin{bmatrix} \mathbf{P} & \mathbf{B} & -\mathbf{C} \\ \mathbf{S} & \mathbf{D} & -\mathbf{E} \\ \mathbf{N} & -\mathbf{E}^* & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{Q} \\ \mathbf{U} \\ \widehat{\mathbf{U}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{H} \\ \mathbf{G} \end{bmatrix} \quad (3.14)$$

Note that  $\mathbf{Q}$  represents the variables associated with the  $\mathbf{q}_h$  degrees of freedom,  $\mathbf{U}$  those related to  $u_h$  and similarly  $\widehat{\mathbf{U}}$  those of  $\widehat{u}_h$ . Also note that the subblock  $\mathbf{E}^*$  is the adjoint of  $\mathbf{E}$ , that is the conjugate transpose, for reasons that will become clear later. As it was already pointed out in equations (3.8) and (3.10), the degrees of freedom regarding  $\mathbf{q}_h, u_h$  have only a local dependence, and once the numerical traces  $\widehat{u}_h$  are found in the global problem, they may be recovered separately for each element. This circumstance constitutes one of the great advantages of HDG over other DG methods.

Thanks to this local dependence, the submatrix

$$\begin{bmatrix} \mathbf{P} & \mathbf{B} \\ \mathbf{S} & \mathbf{D} \end{bmatrix} \quad (3.15)$$

is block diagonal and invertible if grouped by elements, provided  $\tau > 0$ . Using this feature and manipulating (3.14), the degrees of freedom for  $\mathbf{Q}, \mathbf{U}$  may be eliminated, rendering the following global problem

$$\mathbf{T}\widehat{\mathbf{U}} = \mathbf{R} \quad (3.16)$$

where the matrices  $\mathbf{T}, \mathbf{R}$  are given by

$$\mathbf{T} = \mathbf{M} + [\mathbf{N} \quad -\mathbf{E}^*] \begin{bmatrix} \mathbf{P} & \mathbf{B} \\ \mathbf{S} & \mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C} \\ \mathbf{E} \end{bmatrix} \quad (3.17a)$$

$$\mathbf{R} = \mathbf{G} - [\mathbf{N} \quad -\mathbf{E}^*] \begin{bmatrix} \mathbf{P} & \mathbf{B} \\ \mathbf{S} & \mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{H} \end{bmatrix} \quad (3.17b)$$

once we have solved for the global degrees of freedom, the local ones may be recovered using expression (3.18).

$$\begin{bmatrix} \mathbf{Q} \\ \mathbf{U} \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{B} \\ \mathbf{S} & \mathbf{D} \end{bmatrix}^{-1} \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{H} \end{bmatrix} + \begin{bmatrix} \mathbf{C} \\ \mathbf{E} \end{bmatrix} \hat{\mathbf{U}} \right). \quad (3.18)$$

### 3.2.6 Adjoint equation

In order to attain optimality in the reduced-basis algorithm, the adjoint equation needs to be solved as well. To define the adjoint equation we first need to explicit the boundary conditions of the problem

$$g_R = ik \quad (3.19a)$$

$$g_N = 0 \quad (3.19b)$$

where  $i$  is the imaginary unit and  $k$  is the wavenumber. The Robin boundary condition for this case is also known as absorbing boundary condition. Further details on its definition may be found in Chapter 4. Note that the presence of an absorbing boundary condition implies that the solution of the problem will no longer be real-valued, therefore the need to use the adjoint  $x^*$  operator instead of the transpose  $x^T$ .

The final ingredient needed is the output or quantity of interest  $s$ . For the Helmholtz problem, we will consider

$$s^e = l(u) = \int_{\Gamma_N} u dS \quad (3.20)$$

This output represents the total amplitude of the wave on the Neumann boundary.

The adjoint equation is very simple for the Helmholtz problem, since it is self-adjoint. For the sake of clarity, a thorough derivation for the adjoint is provided

below. To obtain the adjoint equation, we first linearize the governing equations (3.1) by introducing perturbations to the solution field  $u + \delta u$  and the density field  $\rho + \delta \rho$  and cancelling second order terms, rendering

$$-\nabla \cdot (\rho \nabla \delta u + \delta \rho \nabla u) - k^2 \delta u = 0 \quad \forall \mathbf{x} \in D, \quad (3.21a)$$

$$(\rho \nabla \delta u + \delta \rho \nabla u) \cdot \mathbf{n} = ik \delta u \quad \forall \mathbf{x} \in \Gamma_R, \quad (3.21b)$$

$$(\rho \nabla \delta u + \delta \rho \nabla u) \cdot \mathbf{n} = 0 \quad \forall \mathbf{x} \in \Gamma_N. \quad (3.21c)$$

The objective function (3.20) is also linearized

$$\delta s^e = \int_{\Gamma_N} \delta u dS. \quad (3.22)$$

Now we shall introduce an additional variable  $\psi^*$ , referred to as the adjoint variable, that multiplies (3.21a)

$$-\psi^* [\nabla \cdot (\rho \nabla \delta u + \delta \rho \nabla u) + k^2 \delta u] = 0 \quad (3.23)$$

equality that is satisfied for any  $\psi$ . Therefore, we have

$$\delta s^e = \int_{\Gamma_N} \delta u dS - \int_D \psi^* [\nabla \cdot (\rho \nabla \delta u + \delta \rho \nabla u) + k^2 \delta u] dV. \quad (3.24)$$

Using the product rule for vector calculus the latter is rearranged as,

$$\begin{aligned} \delta s^e = & \int_{\Gamma_N} \delta u dS - \int_D \nabla \cdot (\psi^* (\rho \nabla \delta u + \delta \rho \nabla u)) dV + \\ & + \int_D \nabla \psi^* \cdot (\rho \nabla \delta u + \delta \rho \nabla u) dV - \int_D k^2 \psi^* \delta u dV. \end{aligned} \quad (3.25)$$

Applying the divergence theorem, we get

$$\begin{aligned} \delta s^e = & \int_{\Gamma_N} \delta u dS - \int_{\Gamma} \psi^* (\rho \nabla \delta u + \delta \rho \nabla u) \cdot \mathbf{n} dS + \int_D \delta \rho \nabla \psi^* \cdot \nabla u dV + \\ & + \int_D \rho \nabla \psi^* \cdot \nabla \delta u dV - \int_D k^2 \psi^* \delta u dV. \end{aligned} \quad (3.26)$$



Repeating this two-step process with the last term we obtain,

$$\begin{aligned} \delta s^e &= \int_{\Gamma_N} \delta u \, dS - \int_{\Gamma} \psi^* (\rho \nabla \delta u + \delta \rho \nabla u) \cdot \mathbf{n} \, dS + \int_D \delta \rho \nabla \psi^* \cdot \nabla u \, dV + \\ &+ \int_{\Gamma} \rho \delta u \nabla \psi^* \cdot \mathbf{n} \, dS - \int_D \delta u \nabla \cdot (\rho \nabla \psi^*) \, dV - \int_D k^2 \psi^* \delta u \, dV . \end{aligned} \quad (3.27)$$

Separating volume and surface integrals, we have

$$\begin{aligned} \delta s^e &= \int_{\Gamma_N} \delta u \, dS - \int_D \delta u \nabla \cdot (\rho \nabla \psi^*) \, dV + \int_D \delta \rho \nabla \psi^* \cdot \nabla u \, dV - \\ &- \int_{\Gamma} (\psi^* (\rho \nabla \delta u + \delta \rho \nabla u) - \rho \delta u \nabla \psi^*) \cdot \mathbf{n} \, dS - \int_D k^2 \psi^* \delta u \, dV . \end{aligned} \quad (3.28)$$

Splitting the boundary terms and applying the linearized boundary conditions

$$\begin{aligned} \delta s^e &= \int_{\Gamma_N} \delta u \, dS - \int_D \delta u \nabla \cdot (\rho \nabla \psi^*) \, dV + \int_D \delta \rho \nabla \psi^* \cdot \nabla u \, dV - \int_D k^2 \psi^* \delta u \, dV - \\ &- \int_{\Gamma_R} (ik\psi^* \delta u - \rho \delta u \nabla \psi^* \cdot \mathbf{n}) \, dS \\ &+ \int_{\Gamma_N} \rho \delta u (\nabla \psi^* \cdot \mathbf{n}) \, dS . \end{aligned} \quad (3.29)$$

The equation and boundary conditions for the adjoint problem yield automatically by enforcing the terms in the latter equation to vanish. Therefore, the adjoint variable  $\psi^*$  is the one solving the strong equation

$$-\nabla \cdot (\rho \nabla \psi^*) - k^2 \psi^* = 0, \quad \forall \mathbf{x} \in D \quad (3.30a)$$

$$\rho \nabla \psi^* \cdot \mathbf{n} = ik\psi^*, \quad \forall \mathbf{x} \in \Gamma_R \quad (3.30b)$$

$$\rho \nabla \psi^* \cdot \mathbf{n} = -1, \quad \forall \mathbf{x} \in \Gamma_N \quad (3.30c)$$

In order to solve it, we apply the same procedure as above. We introduce an additional unknown for the gradient of the adjoint field  $v = \nabla \psi$ . We seek for the approximate solution  $v_h, \psi_h, \widehat{\psi}_h \in \mathbf{V}_h^p \times W_h^p \times M_h^p$ , where the spaces are the ones in (3.6). The

matrix equation for the adjoint problem reads

$$\begin{bmatrix} \Upsilon^* & \Psi^* & \widehat{\Psi}^* \end{bmatrix} \begin{bmatrix} \mathbb{P} & \mathbb{B} & -\mathbb{C} \\ \mathbb{S} & \mathbb{D} & -\mathbb{E} \\ \mathbb{N} & -\mathbb{E}^* & \mathbb{M} \end{bmatrix} = - \begin{bmatrix} 0 & 0 & \mathbb{L}^* \end{bmatrix} \quad (3.31)$$

where  $\Upsilon$  refers to the degrees of freedom of  $\nabla\psi_h$  and  $\Psi, \widehat{\Psi}$  to the degrees of freedom of  $\psi_h, \widehat{\psi}_h$ , and the vector  $\mathbb{L}$  arises from the linear form corresponding to the Neumann boundary condition (3.30c),  $l(\mu) = (\widetilde{g}_N, \mu)_{\Gamma_N}$ , where  $\widetilde{g}_N = 1$ . To solve (3.31), a similar procedure to the one described in (3.16)-(3.18) is used. Just note that since in this particular problem there are complex numbers, special care needs to be taken when dealing with the adjoint  $*$  operator.

### 3.2.7 Reduced-basis approach

The standard way of solving an equation using HDG once the weak form is known relates back to equations (3.16)-(3.18). However, if a reduced-basis approach is pursued, the latter procedure can no longer be applied, since we need an affine parametrization of the matrix equation. Indeed, even if such parametrization can be attained in (3.14), this structure will most likely not be conserved in (3.16).

For the Helmholtz equation we will be interested in treating as a parameter the medium density  $\rho$ . In general, instead of the rather simplistic assumption that  $\rho$  is merely a parameter, we shall consider a spatial dependency  $\rho(\mathbf{x})$ , which can be expressed in terms of a finite expansion such as

$$\rho(\mathbf{x}) = \sum_{i=1}^N \xi_i \rho_i(\mathbf{x}) \quad (3.32)$$

where  $\xi_i$  are the unknown parameters and  $\rho_i(\mathbf{x})$  are modes describing the complexity of the field. Such an expansion may be obtained, for instance, as the truncated KL expansion of a certain random process  $\mathcal{R}$ . This truncation, already presented as the

finite dimensional noise assumption, requires more modes as the field becomes more complicated to represent.

### Affine reduction

The Helmholtz problem (3.2) is written in a way such that the equations corresponding to  $\mathbf{q}$  (3.2a) are independent of the medium density  $\rho$ . This allows us to express the degrees of freedom for  $\mathbf{U}$ ,  $\widehat{\mathbf{U}}$  in terms of degrees of freedom for  $\mathbf{Q}$ , using the first equation of (3.14)

$$\mathbf{Q} = \mathbf{P}^{-1} (\mathbf{C}\widehat{\mathbf{U}} - \mathbf{B}\mathbf{U}) \quad (3.33)$$

This elimination, which can be performed elementwise, renders the following system of equations

$$\begin{bmatrix} \mathbf{D} - \mathbf{S}\mathbf{P}^{-1}\mathbf{B} & \mathbf{S}\mathbf{P}^{-1}\mathbf{C} - \mathbf{E} \\ -\mathbf{E}^* - \mathbf{N}\mathbf{P}^{-1}\mathbf{B} & \mathbf{M} + \mathbf{N}\mathbf{P}^{-1}\mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \widehat{\mathbf{U}} \end{bmatrix} = \begin{bmatrix} \mathbf{H} \\ \mathbf{G} \end{bmatrix}, \quad (3.34)$$

which can also be locally computed and assembled. Solving the system (3.34) renders a solution  $\zeta$  in terms of the state field  $u_h$  and its trace  $\widehat{u}_h$ , which will be a potential candidate for the primal reduced-basis. Furthermore, since the Helmholtz equation is self-adjoint, performing a similar elimination for (3.31) we arrive to a reduced system for the adjoint

$$\begin{bmatrix} \mathbf{D} - \mathbf{S}\mathbf{P}^{-1}\mathbf{B} & \mathbf{S}\mathbf{P}^{-1}\mathbf{C} - \mathbf{E} \\ -\mathbf{E}^* - \mathbf{N}\mathbf{P}^{-1}\mathbf{B} & \mathbf{M} + \mathbf{N}\mathbf{P}^{-1}\mathbf{C} \end{bmatrix}^* \begin{bmatrix} \Psi \\ \widehat{\Psi} \end{bmatrix} = - \begin{bmatrix} \mathbf{0} \\ \mathbf{L} \end{bmatrix} \quad (3.35)$$

The solution  $\zeta^d$  of this system, expressed in terms of the adjoint state field  $\psi_h$  and its trace  $\widehat{\psi}_h$ , is in turn a potential candidate for the adjoint reduced-basis. To be consistent with the notation introduced in the previous chapter, systems (3.34)-(3.35)

are rewritten as

$$\mathbb{A}\zeta = \mathbb{F} \tag{3.36a}$$

$$\mathbb{A}^*\zeta^d = -\mathbb{L} \tag{3.36b}$$

The latter formulation is convenient because the degrees of freedom of  $\mathbb{Q}$  have been eliminated, thus significantly reducing the size of the matrices that need to be stored, yet conserving the affine parametric dependency. Indeed, the abstract parametric formulation of the bilinear and linear forms introduced in (2.8) is now expressed in terms of the HDG formulation as

$$\left[ \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{A}^q \right] \zeta = \mathbb{F} , \tag{3.37a}$$

$$\left[ \sum_{q=1}^Q \sigma_q(\boldsymbol{\xi}) \mathbb{A}^q \right]^* \zeta^d = -\mathbb{L} . \tag{3.37b}$$

Note that for the Helmholtz problem  $\tilde{Q} = 1, \tilde{\sigma}_q(\boldsymbol{\xi}) = 1$ , since the source term does not have parametric dependency. Therefore, in the offline stage matrices  $\mathbb{A}^q$  are computed by alternatively setting  $\sigma_q = 1, \sigma_{q'} = 0, \forall q' \in \{1, \dots, Q\} \neq q$ , together with vectors  $\mathbb{F}, \mathbb{L}$ .

## Inner Product Matrix

Furthermore, we need to address the computation of the matrix corresponding to the inner product associated with the high-order finite element space, denoted in the last chapter as  $\mathbb{X}$ . Since the operator is noncoercive, to define an inner product matrix,

(symmetric and positive definite) we resort to the following equation

$$\mathbf{q} - \nabla u = 0, \quad \forall \mathbf{x} \in D \quad (3.38a)$$

$$-\nabla^2 \mathbf{q} + u = h, \quad \forall \mathbf{x} \in D \quad (3.38b)$$

$$\mathbf{q} \cdot \mathbf{n} = 0, \quad \forall \mathbf{x} \in \Gamma_R \quad (3.38c)$$

$$\mathbf{q} \cdot \mathbf{n} = 0, \quad \forall \mathbf{x} \in \Gamma_N \quad (3.38d)$$

Applying the same methodology as for the Helmholtz equation, we arrive to a matrix system

$$\begin{bmatrix} \mathbf{P} & \mathbf{B} & -\mathbf{C} \\ -\mathbf{B}^* & \tilde{\mathbf{D}} & -\mathbf{E} \\ \mathbf{C}^* & -\mathbf{E}^* & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{Q} \\ \mathbf{U} \\ \hat{\mathbf{U}} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{H} \\ 0 \end{bmatrix}, \quad (3.39)$$

where the matrices are the same as before except  $\tilde{\mathbf{D}}$ , which arises from the modified bilinear form  $\tilde{d}(u, w) = (u, w)_{\mathcal{T}_h} + \langle \rho \tau u, w \rangle_{\partial \mathcal{T}_h}$ . Eliminating the degrees of freedom corresponding to  $\mathbf{Q}$ , the matrix for the inner product is defined as

$$\mathbf{X} = \begin{bmatrix} \tilde{\mathbf{D}} + \mathbf{B}^* \mathbf{P}^{-1} \mathbf{B} & -\mathbf{B}^* \mathbf{P}^{-1} \mathbf{C} - \mathbf{E} \\ -\mathbf{E}^* - \mathbf{C}^* \mathbf{P}^{-1} \mathbf{B} & \mathbf{M} + \mathbf{C}^* \mathbf{P}^{-1} \mathbf{C} \end{bmatrix}. \quad (3.40)$$

It is symmetric and positive definite, since it arises from a stiffness matrix plus a mass matrix in the high-order discontinuous finite element space.

## Solving HDG System

While the affine parametric dependency is well suited for the offline-online computational strategy, the duplicity of degrees of freedom typical from the DG methods leads to systems (3.36) that may be very large, thus compromising computational efficiency and losing the characteristic HDG property of solving only for the global degrees of freedom. Nonetheless, this property may be recovered for computing the snapshot solutions  $\zeta, \zeta^d$ .

For a chosen value of the parameter  $\xi_J$ , the systems (3.36) are assembled for the corresponding values of  $\sigma_q(\xi)$ . Now a Schur complement procedure is applied, since the first subblock  $\mathbb{D} - \mathbb{S}\mathbb{P}^{-1}\mathbb{B} = \mathbb{A}_1$  of  $\mathbb{A}$  is *block-diagonal* due to its local definition, thus trivial to invert. To simplify the notation, let us rename the subblocks of  $\mathbb{A}$  row-wise as  $\mathbb{A}_i$ ,  $i = 1, 2, 3, 4$ . The global problem is recovered by substituting the first equation of (3.34) into the second one, that is

$$\mathbb{U} = \mathbb{A}_1^{-1} \left( \mathbb{H} - \mathbb{A}_2 \widehat{\mathbb{U}} \right) \quad (3.41a)$$

$$[\mathbb{A}_4 - \mathbb{A}_3 \mathbb{A}_1^{-1} \mathbb{A}_2] \widehat{\mathbb{U}} = \mathbb{G} - \mathbb{A}_3 \mathbb{A}_1^{-1} \mathbb{H} \quad (3.41b)$$

for the primal and

$$\Psi = -\mathbb{A}_1^{-*} \mathbb{A}_3^* \widehat{\Psi} \quad (3.42a)$$

$$[\mathbb{A}_4^* - \mathbb{A}_2^* \mathbb{A}_1^{-*} \mathbb{A}_3^*] \widehat{\Psi} = -\mathbb{L} \quad (3.42b)$$

for the adjoint, hence we only solve for the global degrees of freedom in (3.41b),(3.42b).

To summarize, we (1) eliminate the degrees of freedom of  $\mathbb{Q}$ ,  $\Upsilon$  to reduce the storage and the dimensionality of the finite-element space. (2) Precompute the matrices and vectors independent of parameters  $\mathbb{A}^q$ ,  $\mathbb{X}$ ,  $\mathbb{F}$ ,  $\mathbb{L}$ . (3) Assemble the matrices for the chosen parameter  $\xi_J$ . (4) Perform a Schur complement decomposition to solve smaller problems (3.41b),(3.42b) involving only the global degrees of freedom  $\widehat{\mathbb{U}}$ ,  $\widehat{\Psi}$ . (5) Recover the local degrees of freedom  $\mathbb{U}$ ,  $\Psi$  to define a snapshot, or solution for the reduced-basis, in terms of both  $\mathbb{U}$ ,  $\widehat{\mathbb{U}}$  and  $\Psi$ ,  $\widehat{\Psi}$  using equations (3.41a),(3.42a). Note that the aforementioned procedure allows for an effective decouple of the reduced-basis code from the HDG code.

### 3.3 The diffusion equation

The HDG method is also presented here for the diffusion equation, which will be of interest in the next chapter. Most of the steps are straightforward after the development presented above, with only minor changes. Again, we shall consider the diffusion equation defined on a certain physical domain  $D \in \mathbb{R}^d$  with boundary  $\partial D = \Gamma$ . The strong formulation of the diffusion equation reads

$$-\nabla \cdot (\kappa \nabla u) = h \quad \forall \mathbf{x} \in D, \quad (3.43a)$$

$$u = 0 \quad \forall \mathbf{x} \in \Gamma_D, \quad (3.43b)$$

$$\kappa \nabla u \cdot \mathbf{n} = g \quad \forall \mathbf{x} \in \Gamma_{N_1}, \quad (3.43c)$$

$$\kappa \nabla u \cdot \mathbf{n} = 0 \quad \forall \mathbf{x} \in \Gamma_{N_2}. \quad (3.43d)$$

#### 3.3.1 HDG formulation

The HDG formulation for the diffusion equation resembles the one obtained for the Helmholtz equation. As done above, we define an additional variable equal to the gradient of the state field times the diffusivity field  $\mathbf{q} = \kappa \nabla u$ , which enables to rewrite equation (3.43) in terms of the system

$$\kappa^{-1} \mathbf{q} - \nabla u = 0 \quad \forall \mathbf{x} \in D, \quad (3.44a)$$

$$-\nabla \cdot \mathbf{q} = h \quad \forall \mathbf{x} \in D, \quad (3.44b)$$

$$u = 0 \quad \forall \mathbf{x} \in \Gamma_D, \quad (3.44c)$$

$$\mathbf{q} \cdot \mathbf{n} = g \quad \forall \mathbf{x} \in \Gamma_{N_1}, \quad (3.44d)$$

$$\mathbf{q} \cdot \mathbf{n} = 0 \quad \forall \mathbf{x} \in \Gamma_{N_2}. \quad (3.44e)$$

Using the notation introduced above, we choose the approximation spaces defined in (3.6a)-(3.6b) for the scalar and vector test functions on the elements. However, since we now have a Dirichlet boundary condition, the space of test functions on the faces

will be described by  $M_h^p(0) = \{\mu \in M_h^p : \mu|_{\Gamma_D} = 0\}$ , where  $M_h^p$  is given by (3.6c). Finally, multiplying the governing equations (3.44) for the FE discretized solution by test functions and integrating by parts we obtain the local problem at each element  $K$  for the local variables  $(\mathbf{q}_h, u_h) \in \mathbf{V}_h^p \times W_h^p$

$$(\kappa^{-1}\mathbf{q}_h, \mathbf{v})_K + (u_h, \nabla \cdot \mathbf{v})_K - \langle \widehat{u}_h, \mathbf{v} \cdot \mathbf{n} \rangle_{\partial K} = 0 \quad \forall \mathbf{v} \in [\mathcal{P}^p(K)]^d, \quad (3.45a)$$

$$(\mathbf{q}_h, \nabla w)_K - \langle \widehat{\mathbf{q}}_h \cdot \mathbf{n}, w \rangle_{\partial K} = (h, w)_K \quad \forall w \in \mathcal{P}^p(K), \quad (3.45b)$$

and the global problem for the traces  $\widehat{u}_h \in M_h^p(0)$

$$\langle \widehat{\mathbf{q}}_h \cdot \mathbf{n}, \mu \rangle_{\partial \mathcal{T}_h \setminus (\Gamma_D \cup \Gamma_{N_2})} + \langle \widehat{\mathbf{q}}_h \cdot \mathbf{n} - g, \mu \rangle_{\Gamma_{N_1}} = 0, \quad \forall \mu \in \mathcal{P}^p(F). \quad (3.46)$$

Summing the contributions (3.45) over all the elements in the triangulation, using again (3.9) to define the vector fluxes and redoing some integration by parts we obtain the global system of equations (3.47). Find  $(\mathbf{q}_h, u_h, \widehat{u}_h) \in \mathbf{V}_h^p \times W_h^p \times M_h^p$  such that

$$(\kappa^{-1}\mathbf{q}_h, \mathbf{v})_{\mathcal{T}_h} + (u_h, \nabla \cdot \mathbf{v})_{\mathcal{T}_h} - \langle \widehat{u}_h, \mathbf{v} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} = 0, \quad (3.47a)$$

$$- (\nabla \cdot \mathbf{q}_h, w)_{\mathcal{T}_h} - \langle \tau(u_h - \widehat{u}_h), w \rangle_{\partial \mathcal{T}_h} = (h, w)_{\mathcal{T}_h}, \quad (3.47b)$$

$$\langle \mathbf{q}_h \cdot \mathbf{n}, \mu \rangle_{\partial \mathcal{T}_h \setminus \Gamma_D} - \langle \tau u_h, \mu \rangle_{\partial \mathcal{T}_h \setminus \Gamma_D} + \langle \tau \widehat{u}_h, \mu \rangle_{\partial \mathcal{T}_h \setminus \Gamma_D} = \langle g, \mu \rangle_{\Gamma_{N_1}}. \quad (3.47c)$$

holds for all  $(\mathbf{v}, w, \mu) \in \mathbf{V}_h^p \times W_h^p \times M_h^p$ .

### 3.3.2 Weak formulation and matrix system

Similarly as done above for the Helmholtz equation, the system of equations in (3.47) is rewritten for convenience in terms of several bilinear forms. The weak formulation



reads: find  $(\mathbf{q}_h, u_h, \hat{u}_h) \in \mathbf{V}_h^p \times W_h^p \times M_h^p(0)$  such that

$$p(\mathbf{q}_h, \mathbf{v}) + b(u_h, \mathbf{v}) - c(\hat{u}_h, \mathbf{v}) = 0, \quad (3.48a)$$

$$-b(\mathbf{q}_h, w) + d(u_h, w) - e(\hat{u}_h, w) = h(w) \quad (3.48b)$$

$$c(\mathbf{q}_h, \mu) - e(u_h, w) + m(\hat{u}_h, \mu) = g(\mu) \quad (3.48c)$$

holds for all  $(\mathbf{v}, w, \mu) \in \mathbf{V}_h^p \times W_h^p \times M_h^p(0)$ . The bilinear functions are given by

$$\begin{aligned} p(\mathbf{q}, \mathbf{v}) &= (\kappa^{-1}\mathbf{q}, \mathbf{v})_{\mathcal{T}_h}, & b(\eta, \mathbf{r}) &= (\eta, \nabla \cdot \mathbf{r})_{\mathcal{T}_h}, \\ c(\lambda, \mathbf{r}) &= \langle \lambda, \mathbf{r} \cdot \mathbf{n} \rangle_{\partial\mathcal{T}_h}, & d(u, w) &= \langle \tau u, w \rangle_{\partial\mathcal{T}_h}, \\ e(\lambda, w) &= \langle \tau \lambda, w \rangle_{\partial\mathcal{T}_h}, & h(w) &= (h, w)_{\mathcal{T}_h}, \\ m(\eta, \mu) &= \langle \tau \eta, \mu \rangle_{\partial\mathcal{T}_h}, & g(\mu) &= \langle g, \mu \rangle_{\Gamma_{N_1}}. \end{aligned} \quad (3.49)$$

The discretization of the system of equations given in (3.48) is expressed by the matrix equation denoted in (3.50).

$$\begin{bmatrix} \mathbf{P} & \mathbf{B} & -\mathbf{C} \\ -\mathbf{B}^T & \mathbf{D} & -\mathbf{E} \\ \mathbf{C}^T & -\mathbf{E}^T & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{Q} \\ \mathbf{U} \\ \hat{\mathbf{U}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{H} \\ \mathbf{G} \end{bmatrix} \quad (3.50)$$

Proceeding in the same way as above, the degrees of freedom for  $\mathbf{Q}, \mathbf{U}$  are eliminated to form the global problem  $\mathbf{T}\hat{\mathbf{U}} = \mathbf{R}$ , where the matrices  $\mathbf{T}, \mathbf{R}$  are defined by

$$\mathbf{T} = \mathbf{M} + [\mathbf{C}^T \quad -\mathbf{E}^T] \begin{bmatrix} \mathbf{P} & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C} \\ \mathbf{E} \end{bmatrix}, \quad (3.51a)$$

$$\mathbf{R} = \mathbf{G} - [\mathbf{C}^T \quad -\mathbf{E}^T] \begin{bmatrix} \mathbf{P} & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{H} \end{bmatrix}. \quad (3.51b)$$

Solving the global problem renders the numerical traces, which are used to retrieve the local degrees of freedom using (3.52)

$$\begin{bmatrix} \mathbf{Q} \\ \mathbf{U} \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{D} \end{bmatrix}^{-1} \left( \begin{bmatrix} 0 \\ \mathbf{H} \end{bmatrix} + \begin{bmatrix} \mathbf{C} \\ \mathbf{E} \end{bmatrix} \hat{\mathbf{U}} \right). \quad (3.52)$$

### 3.3.3 Adjoint equation

Again, the adjoint equation is needed to assure optimal converge of the reduced-basis output. The output of interest  $s^e$  for this case is defined as the average of the field variable over some subdomain  $D^* \subset D$ , which can be generally expressed as

$$s^e = l(u) = \int_D u \cdot \theta \, dV \quad (3.53)$$

where  $\theta$  is a function defined over  $D$  which can be, for instance, an indicator function for the subset  $D^*$  (non-smooth), or a Gaussian pulse defined around the region of interest (smooth), or even the average over the whole domain (smooth). Smoothness of the output is crucial in the convergence rate of the output. For this particular problem, we will consider parametric dependance of both the diffusivity field  $\kappa$  and the Neumann boundary condition  $g$ . Since the derivation procedure is very similar to the one developed above, we simply provide the adjoint equation

$$-\nabla \cdot (\kappa \nabla \psi^T) = -\theta \quad \forall \mathbf{x} \in D, \quad (3.54a)$$

$$\psi^T = 0 \quad \forall \mathbf{x} \in \Gamma_D, \quad (3.54b)$$

$$\kappa \nabla \psi^T \cdot \mathbf{n} = 0, \quad \forall \mathbf{x} \in \Gamma_{N_1} \cup \Gamma_{N_2}. \quad (3.54c)$$

### 3.3.4 Reduced-basis approach

Using the same procedure as before, we will present the several steps to achieve an affine decomposition of (3.50) together with an efficient solution of the system. For the diffusion equation, the parameters considered are the diffusivity  $\kappa$  and the flux  $g$ .

Whereas the flux  $g$  is a single parameter, the diffusivity will have different values for different spatial subdomains, hence we still need an affine expansion of the system's matrix. However, for this case we shall assume no explicit spatial dependency of the diffusivity field, e.g. for each region  $\kappa$  will be 'constant'. The problem is precisely defined in Chapter 4.

### Affine reduction

Eliminating (element-wise) the degrees of freedom of  $\mathbb{Q}$  using the first equation of (3.50) as before, the following reduced system of equations is obtained

$$\begin{bmatrix} \mathbb{D} + \mathbb{B}^T \mathbb{P}^{-1} \mathbb{B} & -\mathbb{B}^T \mathbb{P}^{-1} \mathbb{C} - \mathbb{E} \\ -\mathbb{E}^T - \mathbb{C}^T \mathbb{P}^{-1} \mathbb{B} & \mathbb{M} + \mathbb{C}^T \mathbb{P}^{-1} \mathbb{C} \end{bmatrix} \begin{bmatrix} \mathbb{U} \\ \widehat{\mathbb{U}} \end{bmatrix} = \begin{bmatrix} \mathbb{H} \\ \mathbb{G} \end{bmatrix}, \quad (3.55)$$

which can also be locally computed and assembled. Solving the system (3.34) renders a solution  $\zeta$  in terms of the state field  $u_h$  and its trace  $\widehat{u}_h$ , which will be a potential candidate for the primal reduced-basis. The adjoint matrix system is even simpler, since (3.55) is already symmetric

$$\begin{bmatrix} \mathbb{D} + \mathbb{B}^T \mathbb{P}^{-1} \mathbb{B} & -\mathbb{B}^T \mathbb{P}^{-1} \mathbb{C} - \mathbb{E} \\ -\mathbb{E}^T - \mathbb{C}^T \mathbb{P}^{-1} \mathbb{B} & \mathbb{M} + \mathbb{C}^T \mathbb{P}^{-1} \mathbb{C} \end{bmatrix} \begin{bmatrix} \Psi \\ \widehat{\Psi} \end{bmatrix} = - \begin{bmatrix} \Theta \\ 0 \end{bmatrix}, \quad (3.56)$$

where the vector  $\Theta$  arises from the output functional-dual source, defined as  $\theta(w) = (\theta, w)_{\mathcal{T}_h}$ . The solution  $\zeta^d$  of this system, expressed in terms of the adjoint state field  $\psi_h$  and its trace  $\widehat{\psi}_h$ , is in turn a potential candidate for the dual reduced-basis. To be consistent with the notation introduced in the previous chapter, systems (3.55)-(3.56) are rewritten as

$$\mathbb{A}\zeta = \mathbb{F}, \quad (3.57a)$$

$$\mathbb{A}\zeta^d = -\mathbb{L}, \quad (3.57b)$$

and the affine parametric expansion reads

$$\left[ \sum_{q=1}^{Q-1} \sigma_q(\boldsymbol{\xi}) \mathbb{A}^q \right] \zeta = \sigma_Q(\boldsymbol{\xi}) \mathbb{F}^Q, \quad (3.58a)$$

$$\left[ \sum_{q=1}^{Q-1} \sigma_q(\boldsymbol{\xi}) \mathbb{A}^q \right] \zeta^d = -\mathbb{L}, \quad (3.58b)$$

since the Neumann boundary condition has parametric dependency. Therefore, in the offline stage matrices  $\mathbb{A}^q$  are computed by alternatively setting  $\sigma_q = 1, \sigma_{q'} = 0, \forall q' \in \{1, \dots, Q-1\} \neq q$ , together with vectors  $\mathbb{F}^Q, \mathbb{L}$ .

### Inner Product Matrix

For this case the matrix  $\mathbb{X}$  corresponding to the inner product associated with the high-order finite element space is very simple, since the operator is itself symmetric and coercive. The matrix is the same as (3.50) setting  $\kappa = 1$  everywhere, that is

$$\mathbb{X} = \begin{bmatrix} \mathbb{D} + \mathbb{B}^T \tilde{\mathbb{P}}^{-1} \mathbb{B} & -\mathbb{B}^T \tilde{\mathbb{P}}^{-1} \mathbb{C} - \mathbb{E} \\ -\mathbb{E}^T - \mathbb{C}^T \tilde{\mathbb{P}}^{-1} \mathbb{B} & \mathbb{M} + \mathbb{C}^T \tilde{\mathbb{P}}^{-1} \mathbb{C} \end{bmatrix}, \quad (3.59)$$

where matrix  $\tilde{\mathbb{P}}$  arises from the bilinear form  $\tilde{p}(\mathbf{q}, \mathbf{v}) = (\mathbf{q}, \mathbf{v})_{\mathcal{T}_h}$ .

### Solving HDG System

To solve the HDG problem for a given parameter  $\boldsymbol{\xi}_J$ , the same Schur complement procedure as before is applied. Renaming the subblocks of  $\mathbb{A}$  as  $\mathbb{A}_i, i = 1, 2, 3$ , where  $\mathbb{A}_2$  is the off-diagonal subblock (note that  $\mathbb{A}$  is symmetric), the global problem reads

$$\mathbb{U} = \mathbb{A}_1^{-1} \left( \mathbb{H} - \mathbb{A}_2 \hat{\mathbb{U}} \right), \quad (3.60a)$$

$$\left[ \mathbb{A}_3 - \mathbb{A}_2^T \mathbb{A}_1^{-1} \mathbb{A}_2 \right] \hat{\mathbb{U}} = \mathbb{G} - \mathbb{A}_2^T \mathbb{A}_1^{-1} \mathbb{H}, \quad (3.60b)$$

for the primal and

$$\Psi = -\mathbb{A}_1^{-1} \left( -\Theta - \mathbb{A}_2 \widehat{\Psi} \right), \quad (3.61a)$$

$$[\mathbb{A}_3 - \mathbb{A}_2^T \mathbb{A}_1^{-1} \mathbb{A}_2] \widehat{\Psi} = \mathbb{A}_2^T \mathbb{A}_1^{-1} \Theta, \quad (3.61b)$$

for the adjoint. Observe that an effective decoupling of the reduced-basis and the HDG method is accomplished.



# Chapter 4

## Numerical Results

In the previous chapters the reduced-basis approach for the HDG method has been introduced. Furthermore, the application of the reduced-basis method for uncertainty propagation has also been described. The objective in this chapter is to apply the HDG reduced-basis uncertainty propagation method to realistic problems and to compare the performance to well-known uncertainty propagation techniques, such as stochastic collocation.

The structure of this chapter is as follows. First, we will study diffusion on a thermal fin, which corresponds to a coercive linear symmetric operator. Convergence of the reduced-basis is analysed together with convergence of the uncertainty propagation, using the sharp and rigorous *a posteriori* error bounds available for this operator. Finally, the Helmholtz equation for wave propagation on a random medium will be studied, for a low wave frequency. Convergence studies of the reduced-basis algorithm are performed to be continued

## 4.1 Thermal Fin

### 4.1.1 Definition

The thermal fin is used as a test problem for the case of coercive linear symmetric operators. We study the heat conduction in a thermal fin. This problem has been widely used as a test problem in the context of reduced-basis, see [65, 67]. The problem under study is depicted in Figure 4-1. Note that the boundaries correspond to the ones defined in the strong formulation (3.43). A heat flux  $g$  is prescribed at the root. This flux, that may be the heat generated by an external device, will be considered as an uncertain parameter. The Dirichlet boundary remains at a fixed temperature  $u = 0$ , and if we assume no heat loss through the other boundary  $\Gamma_{N_2}$ , we model the heating of the fin. The fin under study has three different stages. Since each

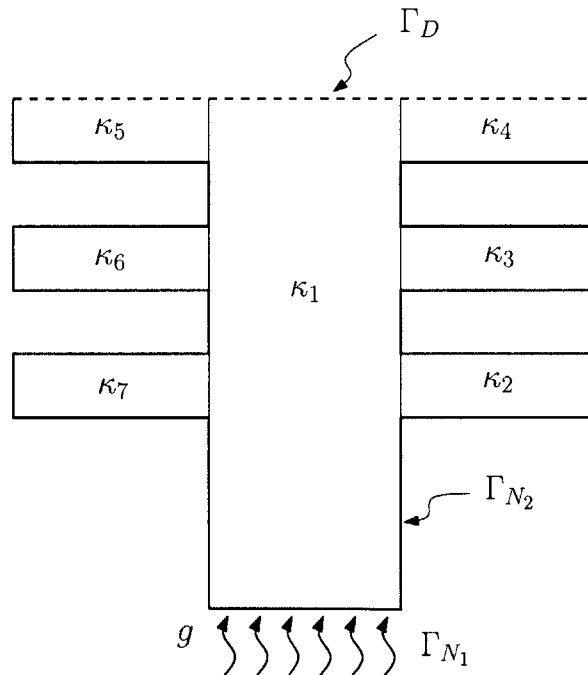


Figure 4-1: Geometry of the thermal fin problem. The fin consists of six subfins, and the flux is introduced by the root.

part of the fin may be assembled using different materials, we shall assume that the thermal diffusivities  $\kappa_i$  are different for each subfin  $D_i$ ,  $i = 1, \dots, 7$ . The diffusivity on the central part of the fin is denoted as  $\kappa_1$ . Note that we allow diffusivities of subfins



at the same stage to be different. Hence, we will not have symmetric solutions. All diffusivities are assumed positive  $\kappa_i > 0$ . We also assume no heat generation within the thermal fin, hence the heating will come from the flux  $g$  entering at the root  $\Gamma_{N_1}$  in its entirety. We will also assume the flux to be positive.

The quantity of interest of the problem is defined as the average heat over the entire domain  $D$ , hence  $\theta = \frac{1}{\int_D 1 dV}$  in equation (3.53). This output is smooth, thus we should expect the adjoint problem to converge at similar rate to the primal.

### 4.1.2 HDG solution

For all the subsequent computations, we will assume that the physical domain is discretized using a triangular mesh  $\mathcal{T}_h$  of  $\mathbf{ne} = 1490$  straight-edged elements of order  $p = 3$ . The mesh is shown in Figure 4-2, it is conforming and generated by ensuring that the subdomains  $D_i$  are separated. The dimension of the high-order discontin-

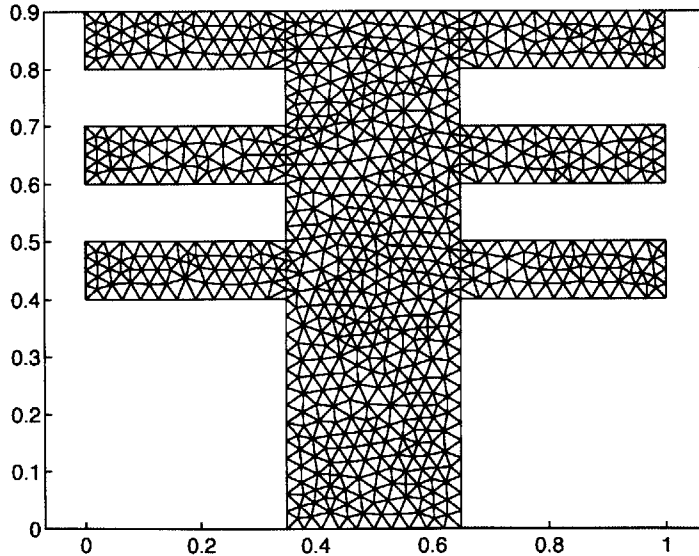


Figure 4-2: Discretization of the thermal fin, using a total of 1490 triangles of polynomial degree  $p = 3$

uous finite element space is  $\mathcal{N} = 24148$ , divided into 14900 degrees of freedom for

both  $u_h, \psi_h$  and 9428 for the numerical traces  $\widehat{u}_h, \widehat{\psi}_h$  (excluding the Dirichlet nodes). Hence, each HDG problem involves solving a sparse system of size 9428. Through-

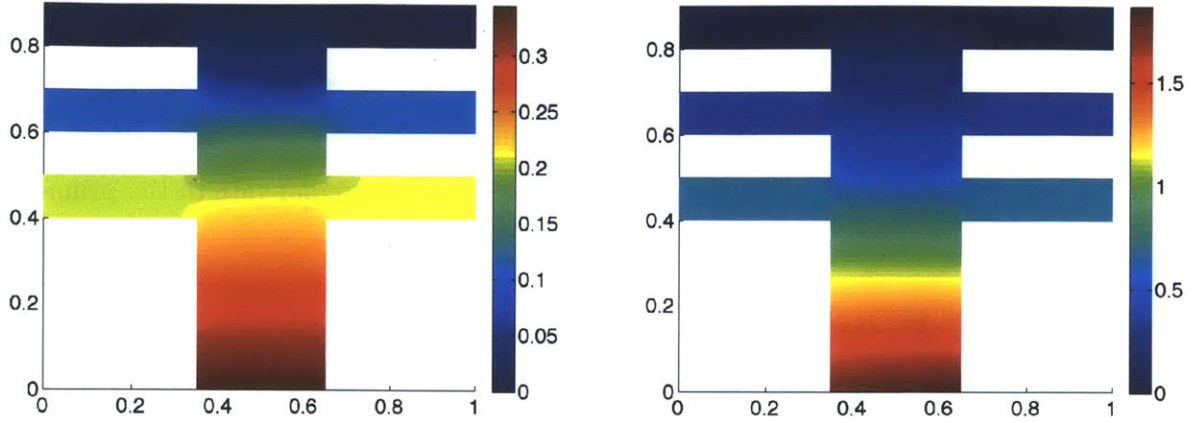


Figure 4-3: Primal solutions to the thermal fin problem for different values of  $\xi$

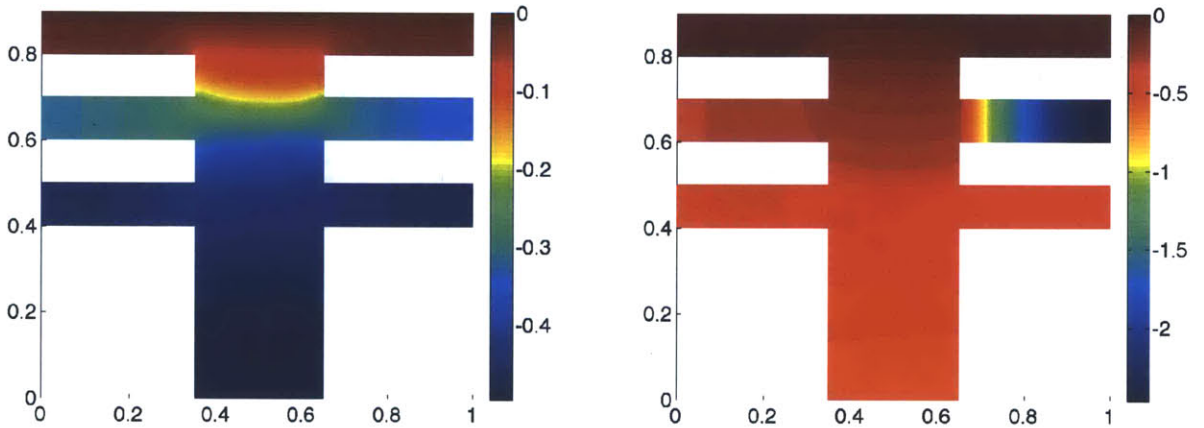


Figure 4-4: Adjoint solutions to the thermal fin problem corresponding to the primal solutions in Figure 4-3

out the problem, we shall assume that the parameters are uniformly distributed as  $(\kappa_1, \dots, \kappa_7, g) = \xi \in \Xi = [0.01, 4]^8$ . In Figures 4-3, 4-4 we show the primal and the adjoint solution to the thermal fin problem using two different realizations of the uncertain parameters  $\xi$ .

### 4.1.3 Reduced-Basis

#### Primal-Adjoint Approach

First we shall evaluate the performance of the reduced-basis method for this problem. To initialize the greedy algorithm, we first need a discretization  $\Theta$  of the continuous parameter space  $\Xi$ . As proposed in the previous chapter a sparse grid of level 4 (3937 points) is used as nodal set  $\Theta$ . The one-dimensional points are obtained with Clenshaw-Curtis abscissae, since the parameters are assumed uniformly distributed. The 8-dimensional problem is run separately for the primal and the adjoint, and the stopping criterion is set when the output error estimator for both the primal and the adjoint are below  $\varepsilon_{\text{tol}} = 10^{-3}$ , i.e.

$$\frac{\|r(v; \boldsymbol{\xi})\|_{X'}}{\sqrt{\hat{\alpha}(\boldsymbol{\xi})}} < 10^{-3} \quad , \quad \frac{\|r^{\text{d}}(v; \boldsymbol{\xi})\|_{X'}}{\sqrt{\hat{\alpha}(\boldsymbol{\xi})}} < 10^{-3} . \quad (4.1)$$

The computation of these error estimators is trivial, since all the parameters are strictly positive, hence a lower bound for the coercivity constant is obtained using equation (2.31) for  $\boldsymbol{\xi} = (1, \dots, 1)$ , that is recovering the Laplacian operator. Furthermore, since the primal and the adjoint need not converge at the same rate, the algorithm is designed to run until both error estimators are below the prescribed tolerance.

In Figure 4-5 the convergence results for both error estimators are shown. For this particular example, note that the adjoint converges slower than the primal, and in particular takes 4 more steps in the greedy algorithm to achieve the desired tolerance.

To assess the power of the reduced basis approach we have performed the following numerical experiment. On the offline stage, we have solved the exact PDE for a total of 500 parameters randomly chosen from the underlying probability density function (in this case multivariate uniform), thus the exact value of the output has been obtained for this set of parameters. After the reduced-basis has been computed, we perform the online stage evaluation for the same set of 500 parameters and compare

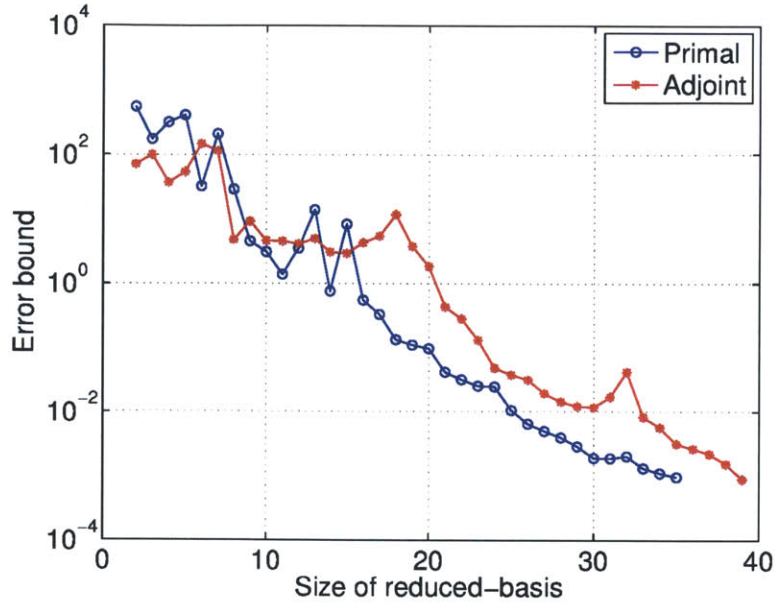


Figure 4-5: Output error estimator  $\varepsilon_M(\boldsymbol{\xi})/\sqrt{\widehat{\alpha}(\boldsymbol{\xi})}$  and  $\varepsilon_{M_d}^d(\boldsymbol{\xi})/\sqrt{\widehat{\alpha}(\boldsymbol{\xi})}$  versus size of the reduced-basis  $M$ ,  $M_d$

the absolute error using both the reduced-basis output  $s_M(\boldsymbol{\xi})$  and the reduced-basis enhanced output  $\tilde{s}_{M,M_d}(\boldsymbol{\xi})$ . Results are depicted in Figure 4-6. Note that the primal-adjoint approach performs as expected, since roughly the 32% of the sample points have an output error greater than  $10^{-6}$ , and the rest lie below. However, convergence of the reduced-basis output  $s_M(\boldsymbol{\xi})$  is much better than expected. In fact, the vast majority of samples have an output error below  $10^{-4}$ , one order of magnitude less than the one imposed by the greedy.

### Primal Approach

In light of the results shown in Figure 4-6 we can consider, instead of performing  $M$  steps of the greedy algorithm for the primal and  $M_d$  for the adjoint, just performing  $\widetilde{M} \approx M + M_d$  steps for the primal only. The problem is solved as before, but now using only the primal for the greedy algorithm. The greedy terminates whenever the output error estimator is below the prescribed tolerance  $\varepsilon_{\text{tol}}^2$ . The convergence of the output error estimator versus the number of reduced basis, here denoted  $\widetilde{M}$

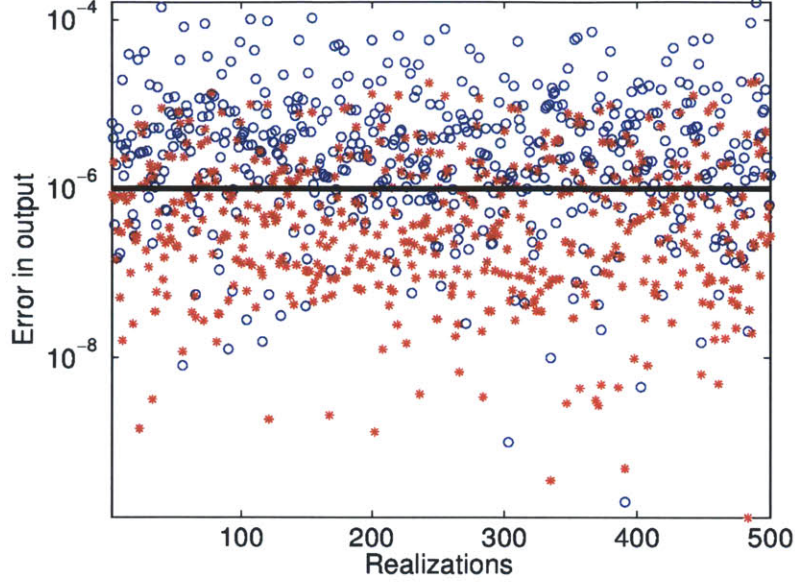


Figure 4-6: Blue dots:  $|s(\boldsymbol{\xi}) - s_M(\boldsymbol{\xi})|$ ; red asterisks:  $|s(\boldsymbol{\xi}) - \tilde{s}_{M,M_d}(\boldsymbol{\xi})|$  versus realizations of the parameter. Straight line corresponds to desired output tolerance  $\varepsilon_{\text{tol}}^2 = 10^{-6}$

to differentiate from the latter approach, is presented in Figure 4-7, superimposed to the results already displayed in Figure 4-5. Similarly as before, the value of the true output error using the same 500 realizations as before is shown in Figure 4-8. Note that in this case the 97% of realizations have an output error below the desired tolerance. It seems that for this particular problem it may be also interesting to consider only the primal equation, since it gives results much better than expected. The primal approach needs  $\tilde{M} = 61$  full problem evaluations, whereas the primal-adjoint approach needs  $M + M_d = 35 + 39 = 74$ . In addition to that at iteration  $n$  of the greedy algorithm we need to solve an  $n \times n$  system for all the remaining candidates in our discrete parameter set  $\Theta$ . It is obvious that the primal approach requires a larger number of system solves (of increasing size), since it needs to run for more iterations, whereas the primal-adjoint terminates before. This involves that although we need to solve for systems for both the primal and the adjoint, these systems are smaller. The difference becomes more obvious as the discrete parameter set contains more points.



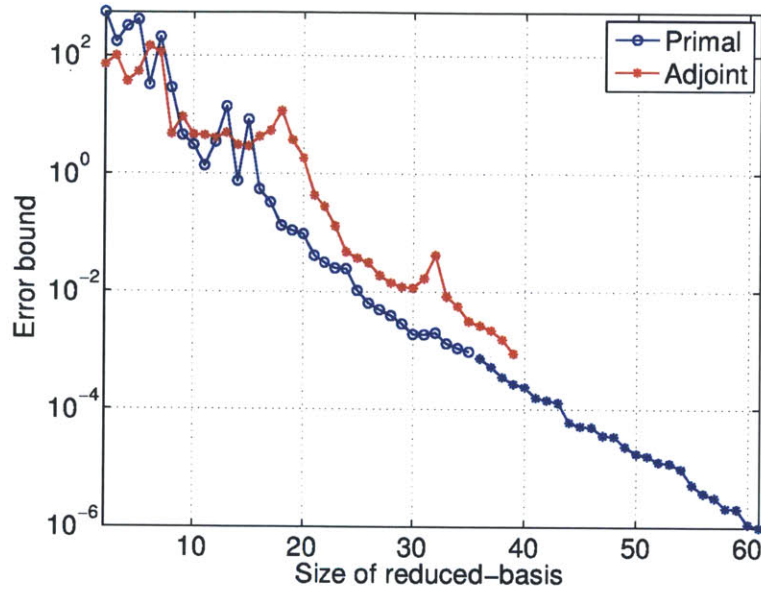


Figure 4-7: Output error estimator  $\varepsilon_{\widetilde{M}}(\boldsymbol{\xi})/\sqrt{\widehat{\alpha}(\boldsymbol{\xi})}$  versus size of the reduced-basis  $\widetilde{M}$ . Primal error estimator runs until reaching  $\varepsilon_{\text{tol}}^2$

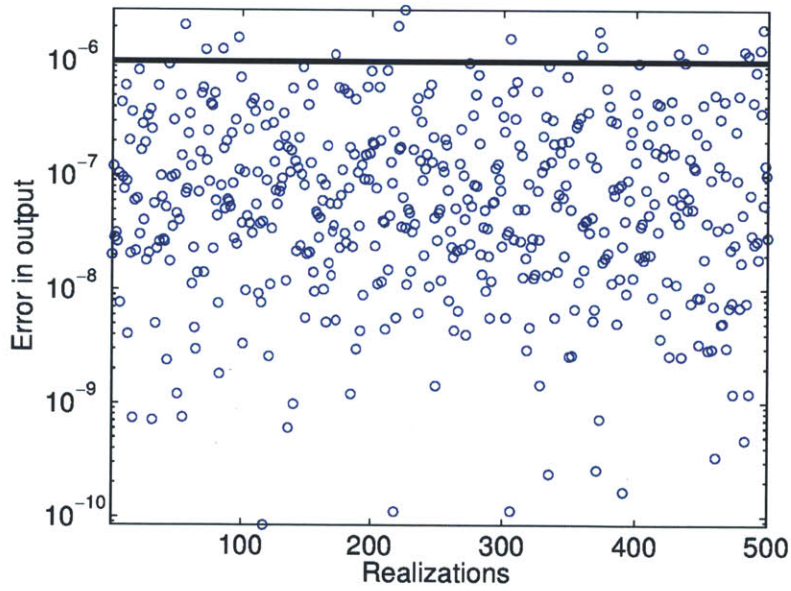


Figure 4-8: Blue dots:  $|s(\boldsymbol{\xi}) - s_{\widetilde{M}}(\boldsymbol{\xi})|$  versus realizations of the parameter. Straight line corresponds to desired output tolerance  $\varepsilon_{\text{tol}}^2 = 10^{-6}$

Furthermore, the size of the final reduced-basis plays an important role in the online stage, since for every new parameter systems (2.26) must be solved. Indeed, the primal approach involves solving a full system of size  $61 \times 61$  for each new sample, while the primal-adjoint approach requires solving a system  $35 \times 35$  and a system  $39 \times 39$ , also full. However, as Figures 4-6, 4-8 point out, convergence of the reduced-basis output is superior to convergence of the reduced-basis enhanced output for a prescribed tolerance on the greedy. For this particular problem, both approaches will be used for the uncertainty propagation.

### Uncertainty Propagation

The final step is to assess the uncertainty in the quantity of interest, that is the average temperature over the fin, given uncertainty in the thermal diffusivity of the subfins and in the flux. Assuming that the random variables are uniformly distributed in  $(\kappa_1, \dots, \kappa_7, g) = \boldsymbol{\xi} \in \Xi = [0.01, 4]^8$ , we shall use the surrogate constructed with reduced-basis to evaluate the moments and the probability density function of the output. For this particular example, we use the approaches introduced before, that is the primal-adjoint reduced-basis or just the primal one. Both surrogates are used to propagate uncertainty with two different Monte Carlo techniques: a set of pseudo-random numbers and a set of quasi-random numbers.

The results are compared to those obtained using a stochastic collocation method on the same 8-dimensional sparse grid of several levels. The sparse grid has been constructed as the sparse product of 1D Clenshaw-Curtis points, hence  $\Theta$  also constitutes a cubature set. Therefore, the  $k$ -th moment of the output may be computed using numerical integration as  $\hat{\mu}_k \equiv \mathbb{E}[\hat{s}^k] = \sum_{j=1}^J s^k(\boldsymbol{\xi}_j) w_j$ , where  $s(\boldsymbol{\xi}_j)$  is the value of the output for  $\boldsymbol{\xi} = \boldsymbol{\xi}_j$  solving the corresponding HDG problem and  $w_j$  is the integration weight.

The results are presented for the first four raw moments  $\hat{\mu}_k$ ,  $k = 1, 2, 3, 4$  of the output in Table 4.1. Furthermore, we also include comparisons with respect to stan-

standard Monte Carlo and Quasi-Monte Carlo sampling. Robustness of the Monte Carlo methods is used to assess the validity of our model. The reference value for each moment is the one obtained using standard Monte Carlo techniques (pseudorandom numbers), since we can also compute errors in the moments according to equation (4.2), where  $J$  is the total number of samples

$$\hat{\mu}_k = \sum_{j=1}^J \frac{s^k(\xi_j)}{J} \quad (4.2a)$$

$$\epsilon_k = \sqrt{\frac{\sum_{j=1}^J \frac{(s^k(\xi_j) - \hat{\mu}_k)^2}{J-1}}{J}} \quad (4.2b)$$

where  $\epsilon_k$  is an unbiased estimator of the standard deviation of  $\hat{\mu}_k$  derived using the Central Limit Theorem, known as the Monte Carlo standard error. Results from a Quasi-Monte Carlo simulation are also shown, although no standard error can be computed, since there exists no Central Limit Theorem for low-discrepancy sequences. For all QMC computations a Sobol sequence generated by **MATLAB** has been used.

<b>Method</b>	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$
MC	0.732 ± 0.004	2.47 ± 0.08	31.5 ± 0.8	680 ± 71
QMC	0.734	2.56	34.3	773
RB Primal-Adjoint MC	0.732	2.47	31.5	680
RB Primal-Adjoint QMC	0.734	2.56	34.3	773
RB Primal MC	0.732	2.47	31.5	680
RB Primal QMC	0.734	2.56	34.3	773
SC level 4	0.330	-0.013	-11.6	-856
SC level 5	0.331	0.076	-7.1	-944

Table 4.1: Results for the first four raw moments for the reduced-basis uncertainty propagation and stochastic collocation, compared with MC and QMC results

Computations for both the full model and the reduced-basis model are set  $10^5$  model evaluations, even though the running time is significantly different. The reduced-basis employed are the ones previously obtained, that is with size  $\widetilde{M} = 61$  for the primal only and  $M + M_d = 35 + 39 = 74$  for the primal-adjoint approach, both



computed with the greedy algorithm on a discrete candidate set consisting of a sparse grid of level 4 in 8 dimensions. To further assess the quality of the reduced-basis, the same set of MC/QMC  $10^5$  samples will be used for propagating uncertainty with the reduced-basis.

Note that the stochastic collocation method fails to converge for this example. Stochastic collocation methods usually offer algebraic convergence, and even exponential if the output function is very smooth. However, for the problem presented above, the response surface is not smooth. To verify it, a simple test has been performed considering only two uncertain parameters,  $\kappa_1$  in the main domain and  $\kappa_2$  for the rest of subfins, with prescribed flux  $g = 1$ . In Figure 4-9 a contour plot of the field of sensitivities of the output with respect to  $\kappa_1$  is depicted. The kink that can be observed compromises the smoothness of the function, hence the performance of the stochastic collocation method. The reduced-basis results are much better, since

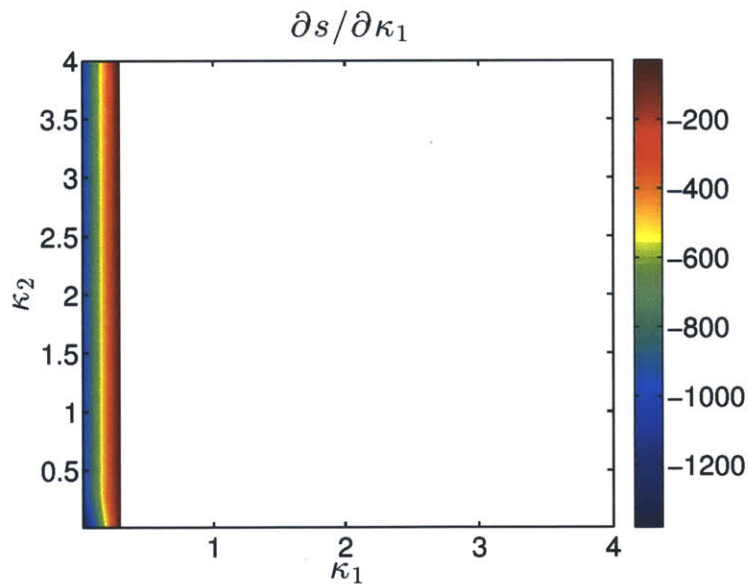


Figure 4-9: Field of sensitivities  $\partial s / \partial \kappa_1$  in the  $\kappa_1 - \kappa_2$  space

the results obtained using the surrogate are indistinguishable to those computed with MC/QMC simulations of the full model. The reliability is assessed by computing the maximum and average relative error for all the MC samples. Results are presented in Table 4.2 for both the primal/primal-adjoint and MC/QMC simulations.

Surrogate	Average relative error	Maximum relative error
RB Primal-Adjoint MC	$2.35 \cdot 10^{-6}$	$1.67 \cdot 10^{-5}$
RB Primal-Adjoint QMC	$2.35 \cdot 10^{-6}$	$1.73 \cdot 10^{-5}$
RB Primal MC	$2.58 \cdot 10^{-7}$	$3.54 \cdot 10^{-6}$
RB Primal QMC	$2.58 \cdot 10^{-7}$	$3.29 \cdot 10^{-6}$

Table 4.2: Values of average and maximum relative error  $|s(\boldsymbol{\xi}) - \tilde{s}_{M,M_d}(\boldsymbol{\xi})| / |s(\boldsymbol{\xi})|$  for the  $10^5$  MC/QMC samples.

The reduced-basis approach is more robust in the sense that once the dominant modes of the solution have been captured, the uncertainty propagation is carried out using MC techniques. Furthermore, it is also much more computationally efficient, since the number of full model evaluations to obtain an accurate prediction of the output moments is several orders of magnitude smaller than the full model evaluations needed for MC methods. For this particular case, it is either  $35 + 39$  or  $61$  vs  $10^5$ , which implies a dramatic saving. Finally, the probability density function (PDF) of the output is shown in Figure 4-10.

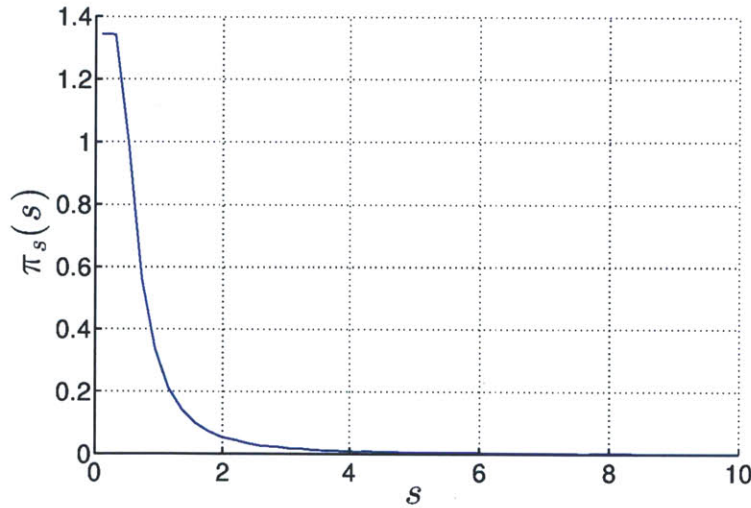


Figure 4-10: Probability density function of the average temperature over the fin, obtained using histogram techniques for the MC simulation

## 4.2 Wave Propagation

### 4.2.1 Definition

To test the reduced-basis uncertainty propagation for noncoercive and nonsymmetric linear problems we utilize a wave propagation problem modeled with the Helmholtz equation, i.e. in the frequency domain. The problem setting is depicted in Figure 4-11. The boundaries correspond to the ones defined in the strong formulation (3.1). A wave is generated from the point source (modeled as a Gaussian source to avoid regularity problems). The wave propagates through an heterogeneous medium of variable density  $\rho(\xi)$  and constant sound speed  $c$ . The wavenumber  $k = \omega/c$  of the

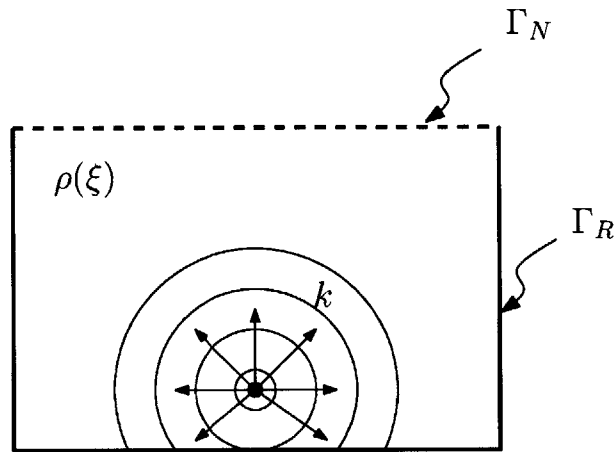


Figure 4-11: Geometry of the wave propagation problem. The source generates a wave that propagates through the medium

propagating wave is assumed constant. The unknown of this problem is the acoustic field  $u$ , and satisfies a Neumann boundary condition at the surface  $\Gamma_N$  and a Sommerfeld radiation (also known as absorbing) condition on the remaining boundaries, defined as  $\lim_{r \rightarrow \infty} r^{1/2} \left( \frac{\partial}{\partial r} - ik \right) u = 0$ . The radiation condition guarantees uniqueness of the solution to the Helmholtz equation. It characterizes radiating solutions, that is the energy radiated from the sources must scatter to infinity.

The uncertainty in the problem arises from the heterogeneity in the medium. We

will assume that the random density  $\rho(\boldsymbol{\xi})$  has the form

$$\rho(\boldsymbol{\xi}) = \rho_0 + \sigma_0 \sum_{i \geq 1} \frac{L}{2i\pi} \cos\left(\frac{2i\pi x}{L}\right) \xi_i \quad (4.3)$$

this form is similar to those obtained from a KL expansion of a random process with eigenvalues decaying like  $\frac{L^2}{(2i\pi)^2}$ . The decay of the eigenvalues is depicted in Figure 4-12. Note that the decay is quite slow, since it takes approximately 57 modes to retain 99% of the energy.

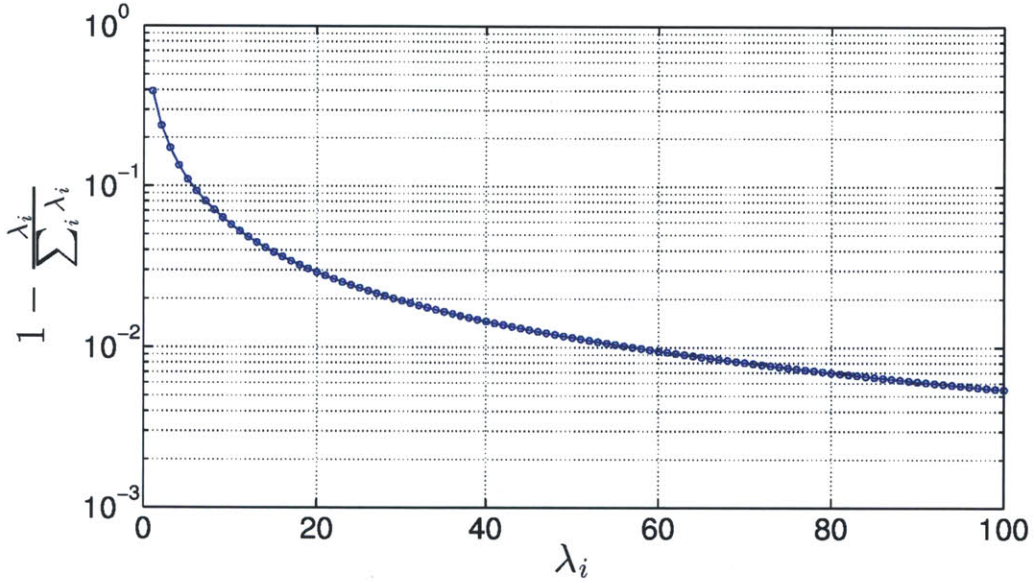


Figure 4-12: Decay of eigenvalues of the random density in normalized scale

The random variables are assumed independent and identically distributed uniform between  $[-1, 1]$ . For the computations presented here, we have used the values  $\sigma = 0.3$ ,  $L = 10$  and  $\rho_0 = 2$ . We employ expression (4.3) to eliminate the errors arising from a KL expansion and to have an analytic formula to simplify the computation of the integrals required in the HDG method.

The quantity of interest for the wave propagation problem is defined as the average amplitude on the wave along the surface  $\Gamma_N$ . The output is smooth, thus a similar rate of convergence for both the primal and the adjoint should be expected. Throughout

this example we will assume the wavenumber is constant and equal to  $k = \sqrt{2}$ , hence we focus on the low-frequency case.

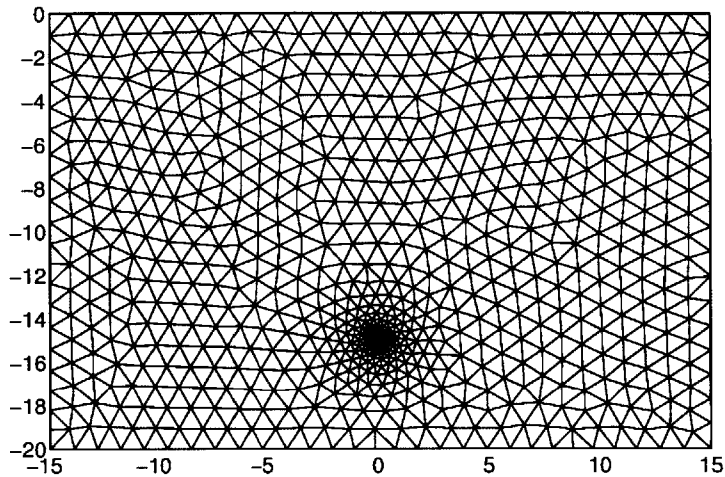


Figure 4-13: Discretization of the wave propagation problem domain, using a total of 1435 triangles of polynomial degree  $p = 4$

## 4.2.2 HDG solution

The subsequent computations are performed with the mesh shown in Figure 4-13. Note that the mesh is refined around the location of the source. The mesh is also triangular with a total of  $\mathbf{ne} = 1435$  straight elements of order  $p = 4$ . The dimension of the high-order discontinuous finite element space is  $\mathcal{N} = 32525$ , divided into 21525 degrees of freedom for both  $u_h, \psi_h$  and 11000 degrees of freedom for the numerical traces  $\widehat{u}_h, \widehat{\psi}_h$ . Therefore the size of the HDG global problems to solve is 11000.

In Figure 4-14 there is a primal and an adjoint solution (only the real part) to the Helmholtz problem for an arbitrary value of the parameter, using a total of 10 terms in the density  $\rho(\boldsymbol{\xi})$ . Solutions in Figure 4-14 are generated using the density field shown in Figure 4-15a, for the source depicted in Figure 4-15b.



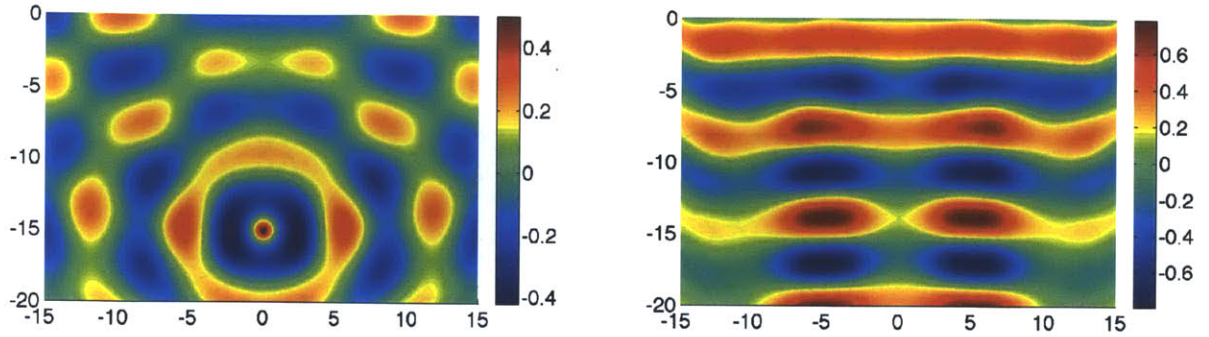


Figure 4-14: Left: Real part of primal solution. Right: Real part of adjoint solution to the wave propagation problem for an arbitrary value of  $\xi$  using 10 terms in the expansion of  $\rho(\xi)$ .

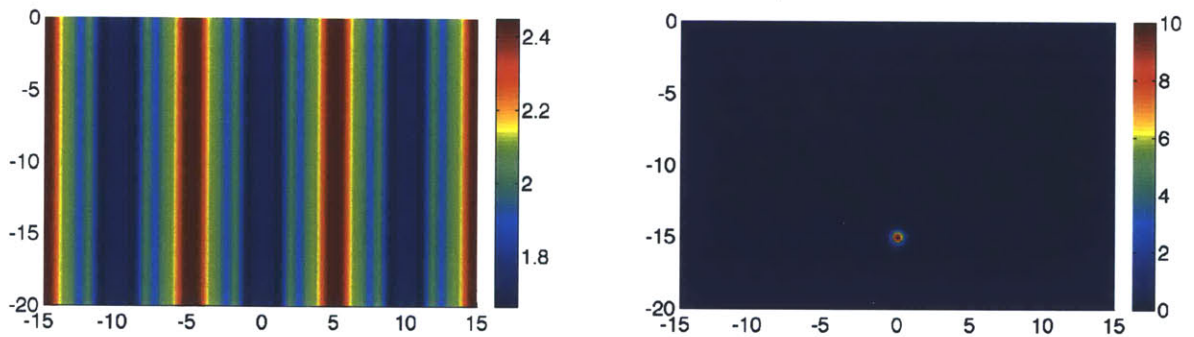


Figure 4-15: Left: Density field to generate solutions in Figure 4-14. Right: Source field for the wave propagation problem.

### 4.2.3 Reduced-Basis

The reduced-basis approach for the Helmholtz problem is first assessed. The starting point of the greedy algorithm is a sparse grid, denoted as  $\Theta$ , using again sparse products of Clenshaw-Curtis points. The greedy is run separately for both the primal and the dual, but since do not include the estimation of a lower bound inf-sup stability

constant  $\beta(\boldsymbol{\xi})$ , the output error estimates are neither sharp nor rigorous for this case. Instead, we will use as a termination criterion one similar to the verification criterion used for the thermal fin problem. We will precompute the exact solution to the problem for a total of 500 arbitrary samples that constitute the test set  $\tilde{\Theta}$ . At each step of the greedy, we will compute  $|s(\boldsymbol{\xi}) - \tilde{s}_{M, M_d}(\boldsymbol{\xi})|$  for all the arbitrary samples, and the greedy will terminate whenever

$$\frac{|s(\boldsymbol{\xi}) - \tilde{s}_{M, M_d}(\boldsymbol{\xi})|}{|s(\boldsymbol{\xi})|} < \varepsilon_{\text{tol}}, \quad \forall \boldsymbol{\xi} \in \tilde{\Theta}.$$

Note that although the choice of the points differs between the primal and the adjoint, since we do not have a separate termination criterion, the final reduced-basis will have the same size, i.e.  $M = M_d$ . For this test case we can consider several number of stochastic dimensions, since the expansion of the density field can be truncated for the desired number of uncertain parameters.

### Primal vs Primal-Adjoint Approach

Firstly, we perform the same comparison as that of the thermal fin problem, comparing the number of greedy iterations to reach the desired tolerance using either just the primal problem or both the primal and the adjoint. For this test case, we resort to a low-dimensional parametric space  $N = 3$ , being the discrete set a sparse grid of level 6 (total of 1073 points), to make sure is reach enough to guarantee convergence of both approaches. The results are depicted in Figure 4-16. For this particular problem is obvious that the computation of a reduced-basis for the adjoint greatly benefits the convergence results. Indeed, the primal only approach needs almost four times (132 vs 34) the number of greedy iterations compared to the primal-adjoint approach. Therefore, not only the full model solves are lower for the primal-adjoint approach and the greedy iterations are more expensive, but also the size of the systems to solve on the online stage is significantly smaller. The results presented below are computed with the primal-adjoint approach, since for this particular problem it

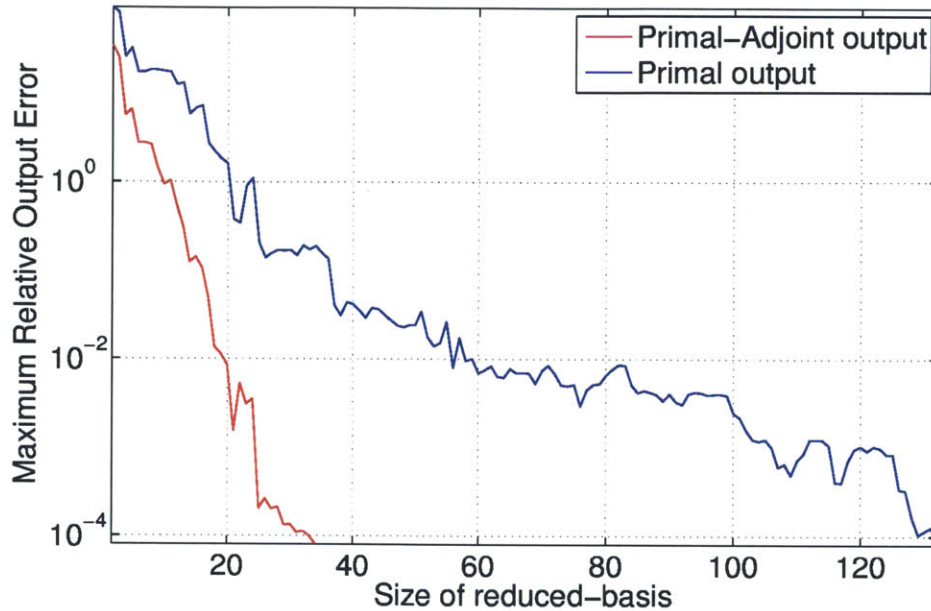


Figure 4-16: Maximum relative error for samples in  $\tilde{\Theta}$  versus size of reduced-basis, for both primal and primal-adjoint approaches

has proven much more effective.

#### $N = 4$ Dimensions

We first consider a low-dimensional case, where the density expansion is truncated using only 4 parameters. Retaining 4 expansion modes of the density is equivalent to considering 87% of the total energy. For this case, we shall assess the effectivity of the sparse grid for this particular problem as a discretization of the parameter set for the greedy approach. Low-dimensionality allows us to compare the sparse grid with a tensor grid that contains the sparse grid. This comparison is interesting because the tensor grid covers the parameter space, whereas the sparse grid focuses on the cartesian directions and on the boundaries of the domain. In Figure 4-17 the comparison between both grids and outputs is shown, using a base 1D Clenshaw-Curtis grid of level 4, consisting of 17 points. Observe that despite the tensor grid has more candidates to choose for the greedy, convergence of the reduced-basis has



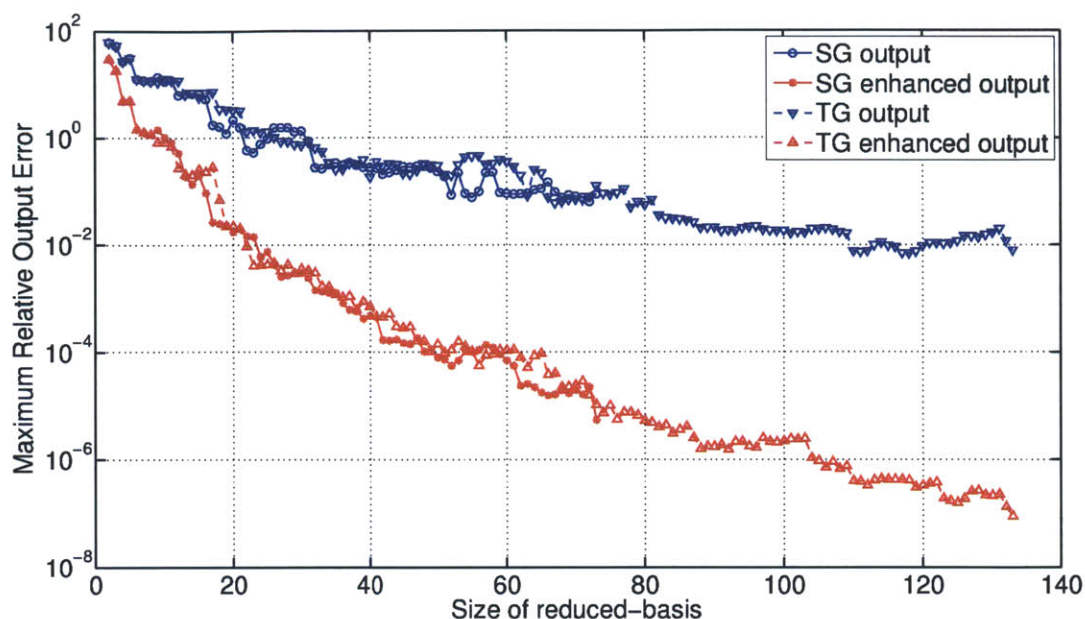


Figure 4-17: Maximum relative error for samples in  $\tilde{\Theta}$  versus size of reduced-basis, for both primal and primal-adjoint approaches. Comparison between sparse grid level 4 (401 points) parameter set and tensor grid (83521 points). Tensor grid is run with a tolerance of  $\varepsilon_{\text{tol}} = 10^{-7}$  and sparse grid with a tolerance of  $\varepsilon_{\text{tol}} = 10^{-5}$ .

approximately the same rate for both grids. Having observed that the convergence does not deteriorate if we substitute uniform gridding by sparse products of grids, the idea of training the greedy algorithm on a sparse set of points seems reasonable. Indeed, the computational savings are enormous, since at each step of the greedy the number of error bounds to compute differ by two orders of magnitude, thus greatly economizing the construction of the reduced-basis.

Furthermore, a straight comparison between the reduced-basis method and the stochastic collocation method can be performed, since the discrete set of points used for both approaches is the same. The main difference is that for stochastic collocation methods the solution of the PDE has to be found at each point in the discrete set  $\Theta$ , whereas for the reduced-basis approach we will solve only for the points chosen by the greedy.

## $N = 8$ Dimensions

Convergence results for the reduced-basis using a truncation of 8 parameters (93% of energy) are presented. For this case no comparison with the tensor grid is done, since the exponential explosion of points constitutes a major obstacle for uniform gridding. The greedy algorithm is trained here on a sparse grid of level 4 (3937 points), and convergence results are shown in Figure 4-18.

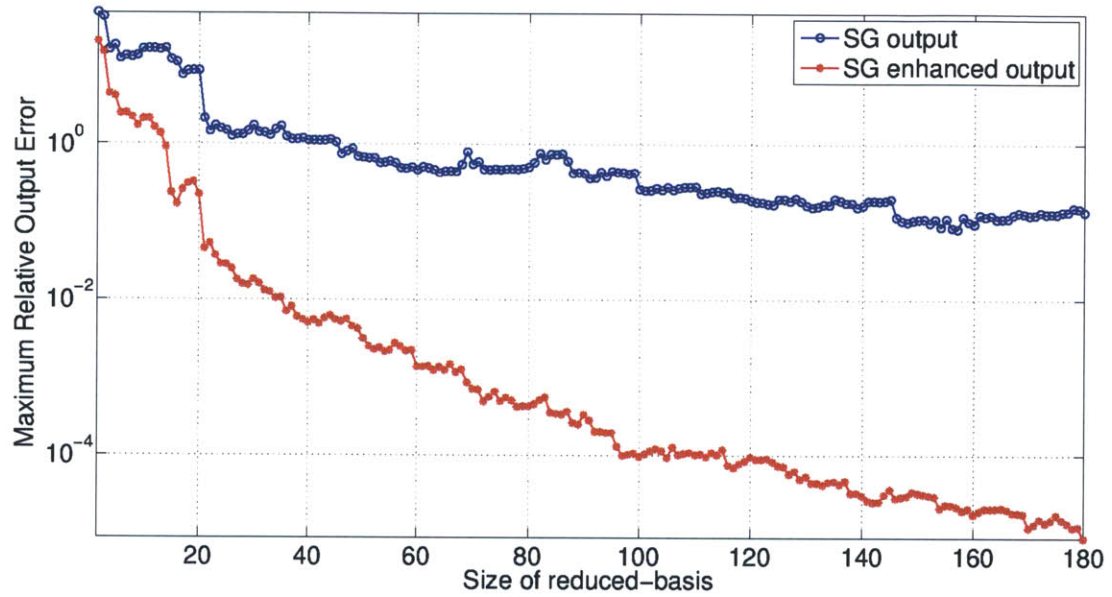


Figure 4-18: Maximum relative error for samples in  $\tilde{\Theta}$  versus size of reduced-basis, for both primal and primal-adjoint approaches. Sparse grid of level 4 (3937 points) with a tolerance of  $\varepsilon_{\text{tol}} = 10^{-5}$ .

The slow decay of the eigenvalues becomes obvious as we increase the dimensionality. In order to reach the same level of accuracy as before, the size of the primal-adjoint reduced-basis is 180 (hence 360 full model evaluations), whereas for the low-dimensional case only 73 basis functions (total of 146 full model evaluations) are needed.

Note the power of combining a sparse grid and the adjoint equation. Applying the greedy algorithm on a sparse grid allows us to explore a high-dimensional space

avoiding the computational complexity of a tensor grid. Furthermore, developing the reduced-basis for the adjoint equation has proven very advantageous for this particular problem. Besides reducing the number of full model evaluations, see Figure 4-16, it also decreases the computational effort to evaluate the surrogate, since the systems (2.41) that we need to solve in the online stage are significantly smaller.

### **$N = 16$ Dimensions**

For the last test case we consider a density field truncated at 16 parameters (96% of energy). The result is presented only for the sparse grid, again with level 4 (51137 points) in Figure 4-19. The greedy algorithm is run up to a tolerance of  $\varepsilon_{\text{tol}} = 10^{-4}$ , since each iteration of the greedy is much more expensive for this case. The analysis to be performed here is the same as for 8 dimensions, and the advantages of solving the adjoint problem become even more clear. Indeed, the maximum relative error in the normal output when the greedy terminates is larger than  $10^{-1}$ . To reach the prescribed tolerance a total of 181 basis functions (362 full model evaluations) are needed.

#### **4.2.4 Uncertainty Propagation**

Once the reduced-basis have been computed for several number of dimensions, the final step is to assess the uncertainty in the quantity of interest by performing online evaluations of the surrogate reduced-basis model using MC techniques. The study is based on the moments of the real part of the total amplitude, that is  $s = \Re(s)$ . Results for the imaginary part present the same features, and are therefore omitted.

For all cases, we will assume that the random variables are independent and uniformly distributed in  $[-1, 1]$ , and three sets of results are presented. Firstly, we compute the moments employing standard MC simulation with  $10^5$  samples. Secondly, we evaluate the reduced-basis on the same set of  $10^5$  MC samples, hence we can also assess the reliability of our surrogate. Finally, the moments of the output

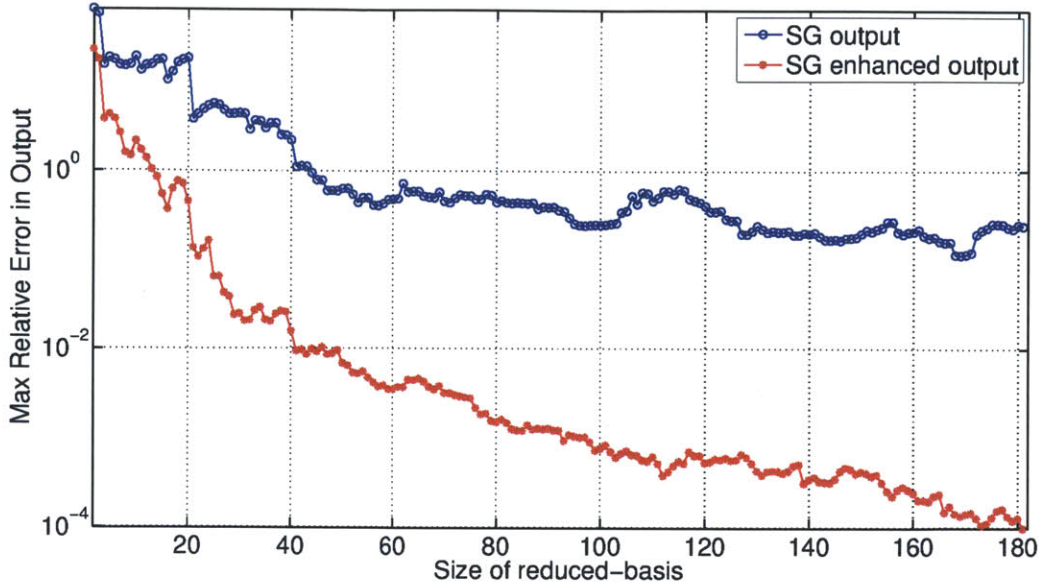


Figure 4-19: Maximum relative error for samples in  $\tilde{\Theta}$  versus size of reduced-basis, for both primal and primal-adjoint approaches. Sparse grid of level 4 (51137 points) with a tolerance of  $\varepsilon_{\text{tol}} = 10^{-4}$ .

(4.2a) are also computed using the stochastic collocation method on the same sparse grid used for the greedy algorithm.

#### $N = 4$ Dimensions

Results for both the reduced-basis uncertainty propagation (RBUP) and the stochastic collocation method for  $N = 4$  are presented in Table 4.3. All computations are performed using a sparse grid of level 4 (401 points), either as the discrete set for the greedy or as the cubature set for the stochastic collocation method. The results obtained using the reduced-basis surrogate are very accurate for both tolerances, which suggests that our reduced-basis effectively captures the dominant information. Obviously, the computational complexity is smaller the lower the tolerance because not only we need more full model solves, but also the greedy iterations are more expensive as the basis grows. Indeed, for the case  $\varepsilon_{\text{tol}} = 10^{-3}$  a total of 36 basis functions are needed, whereas for  $\varepsilon_{\text{tol}} = 10^{-4}$  we need 50. Furthermore, to assess how accurate is

Method	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$
MC	$-1.252 \pm 0.003$	$2.486 \pm 0.007$	$-5.22 \pm 0.02$	$11.52 \pm 0.04$
RB ( $\varepsilon_{\text{tol}} = 10^{-3}$ ) + MC	-1.252	2.486	-5.22	11.52
RB ( $\varepsilon_{\text{tol}} = 10^{-4}$ ) + MC	-1.252	2.486	-5.22	11.52
SC	-1.245	2.484	-5.22	11.51

Table 4.3: First four raw moments of real part of output for the 4-dimensional problem. Results for MC simulation are shown with one standard error. Results for the reduced-basis uncertainty propagation are provided using the same set of MC samples and two different tolerances for the greedy. Results for stochastic collocation are also shown.

the surrogate that we have constructed, we evaluate the relative error between the exact output and the reduced-basis outputs (primal only and primal-adjoint) at each MC sample. Results are shown in Table 4.4.

Size of RB	Average relative error	Maximum relative error
$M = M_d = 36$	$1.17 \cdot 10^{-4}$	3.53
$M = M_d = 50$	$1.04 \cdot 10^{-6}$	$1.76 \cdot 10^{-2}$

Table 4.4: Values of average and maximum relative error  $|s(\boldsymbol{\xi}) - \tilde{s}_{M, M_d}(\boldsymbol{\xi})| / |s(\boldsymbol{\xi})|$  for the  $10^5$  MC samples in 4 dimensions.

Based on these results we can conclude that the surrogate constructed is very accurate. The fact that the termination criterion for the greedy is based on a maximum relative error is very strong, and the results in Table 4.4 suggest that even with less basis functions a very descriptive surrogate may be constructed. The introduction of sharp and rigorous *a posteriori* error estimates will help us overcome this issue.

The stochastic collocation method has a very good convergence in this case, since the response surface is smooth. The evaluation of the output moments is straightforward having chosen the interpolation nodes to be cubature nodes, and the complexity relies solely on performing 401 full model evaluations. However, if we want to compute the PDF, the interpolant needs to be constructed and evaluated at a set of random points. Efficient ways to do so may be found in [50]. The PDF for the MC simulation is shown in Figure 4-20. The PDF is obtained using histogram and kernel density estimation techniques. Histograms will usually render non-smooth PDFs,



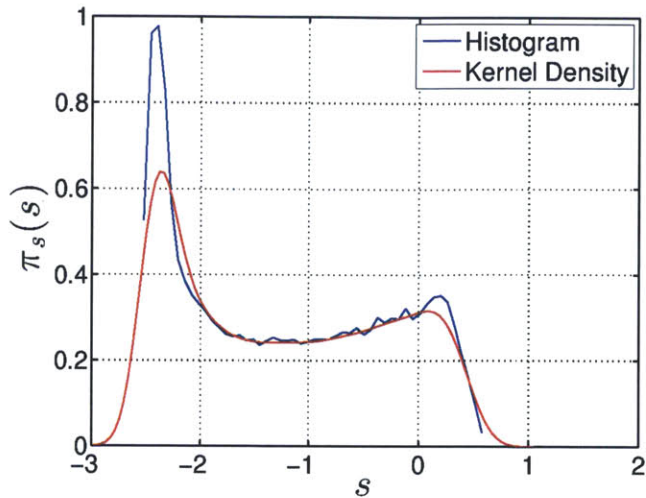


Figure 4-20: Probability density function of the real part of the amplitude at the Neumann boundary, obtained using histogram techniques for the MC simulation

but the representation of the overall behavior is accurate. Kernel density estimation smooths the PDF even when it is not, hence we may lose locality. Both phenomena can be appreciated in 4-20.

### $N = 8$ Dimensions

Results for both the reduced-basis uncertainty propagation and the stochastic collocation method are presented in Table 4.5. All computations are performed using a sparse grid of level 4 (3937 points), either as the discrete set for the greedy or as the cubature set for the stochastic collocation method. The setting is the same as before. Note for this case the size of the reduced-basis is 69 with tolerance  $\varepsilon_{\text{tol}} = 10^{-3}$  and 97 with tolerance  $\varepsilon_{\text{tol}} = 10^{-4}$ . Results in Table 4.6 reflect the accuracy of the reduced-basis surrogate, which is very high. The stochastic collocation method produces moments that lie again within one standard deviation of the MC result.

Method	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$
MC	$-1.246 \pm 0.003$	$2.484 \pm 0.007$	$-5.22 \pm 0.02$	$11.51 \pm 0.04$
RB ( $\varepsilon_{\text{tol}} = 10^{-3}$ ) + MC	-1.246	2.484	-5.22	11.51
RB ( $\varepsilon_{\text{tol}} = 10^{-4}$ ) + MC	-1.246	2.484	-5.22	11.51
SC	-1.248	2.490	-5.23	11.55

Table 4.5: First four raw moments of real part of output for the 8-dimensional problem. Results for MC simulation are shown with one standard error. Results for the reduced-basis uncertainty propagation are provided using the same set of MC samples and two different tolerances for the greedy. Results for stochastic collocation are also shown.

Size of RB	Average relative error	Maximum relative error
$M = M_d = 69$	$8.36 \cdot 10^{-5}$	3.35
$M = M_d = 97$	$8.06 \cdot 10^{-7}$	$3.72 \cdot 10^{-3}$

Table 4.6: Values of average and maximum relative error  $|s(\boldsymbol{\xi}) - \tilde{s}_{M, M_d}(\boldsymbol{\xi})| / |s(\boldsymbol{\xi})|$  for the  $10^5$  MC samples in 8 dimensions.

### $N = 16$ Dimensions

Results for both the reduced-basis uncertainty propagation and the stochastic collocation method are presented in Table 4.7. All computations are performed using a sparse grid of level 4 (51137 points), either as the discrete set for the greedy or as the cubature set for the stochastic collocation method. The size for the reduced-basis

Method	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$
MC	$-1.250 \pm 0.003$	$2.496 \pm 0.007$	$-5.25 \pm 0.02$	$11.59 \pm 0.04$
RB ( $\varepsilon_{\text{tol}} = 10^{-3}$ ) + MC	-1.250	2.496	-5.25	11.59
RB ( $\varepsilon_{\text{tol}} = 10^{-4}$ ) + MC	-1.250	2.496	-5.25	11.59
SC	-1.249	2.494	-5.24	11.58

Table 4.7: First four raw moments of real part of output for the 16-dimensional problem. Results for MC simulation are shown with one standard error. Results for the reduced-basis uncertainty propagation are provided using the same set of MC samples and two different tolerances for the greedy. Results for stochastic collocation are also shown.

is 93 with tolerance  $\varepsilon_{\text{tol}} = 10^{-3}$  and 181 with tolerance  $\varepsilon_{\text{tol}} = 10^{-4}$ . The accuracy of the reduced-basis is assessed with the relative errors of the output for the MC samples, provided in Table 4.8.

Size of RB	Average relative error	Maximum relative error
$M = M_d = 93$	$8.33 \cdot 10^{-5}$	$5.36 \cdot 10^{-1}$
$M = M_d = 181$	$1.04 \cdot 10^{-5}$	$2.91 \cdot 10^{-1}$

Table 4.8: Values of average and maximum relative error  $|s(\boldsymbol{\xi}) - \tilde{s}_{M,M_d}(\boldsymbol{\xi})| / |s(\boldsymbol{\xi})|$  for the  $10^5$  MC samples in 16 dimensions.

Moments computed with stochastic collocation method converge to the moments predicted by MC. However, the fact that results are very similar for all three cases suggest that for this particular problem dimensionality does not play a big part. Even though the expansion for the density decays very slowly, additional dimensions are not really relevant to the quantity of interest. Therefore, a more fair comparison should be made with anisotropic sparse grids, since it would weigh dimensions differently. Comparisons with anisotropic sparse grids, as well as adaptivity for the greedy algorithm is left as future work.



# Chapter 5

## Conclusions and Future Work

In this final chapter we draw conclusions from the research presented above, situating this work in the reduced-basis and uncertainty propagation framework. Furthermore, future lines of research are also introduced.

### 5.1 Conclusions

In this thesis a reduced-basis method for uncertainty propagation using the Hybridizable Discontinuous Galerkin method has been devised. One of the main contributions of this thesis is to apply reduced-basis methods to the HDG method, and it has been demonstrated for linear elliptic equations. Reduced-basis methods have been widely used along with Continuous Galerkin Finite Element discretizations, but with the recent developments in powerful simulation tools for partial differential equations, it is natural to try to adapt existing algorithms to new high fidelity methods, such as HDG. The resulting formulation uses the HDG natural norm, and retains the properties of only solving for the global degrees of freedom.

Furthermore, we have applied reduced-basis techniques for uncertainty propagation purposes. The main idea of reduced-basis methods is to detect the underlying structure of a parametrized PDE (creating a reduced order model) using error es-

timates in a way that the full model is only solved for the values of the parameter that we really need. The offline-online computational strategy allows to decouple the computational complexity of the full model for the online stage, hence the method is suited for multiple queries. Therefore, if the stochasticity in the PDE can be reformulated into parametric uncertainty, e.g. finite dimensional noise assumption, stochastic PDEs can be treated as parametric PDEs, and MonteCarlo techniques may be used for uncertainty propagation. This approach, first pursued by Boyaval et.al. [10] is the one that has driven this work.

The main contribution of this work is to extend the reduced-basis method for stochastic PDEs to the more general noncoercive non-compliance case, taking advantage of adjoint techniques to accelerate the convergence of the reduced-basis, combined with the use of high-order simulation tools (HDG). This method has proven successful for the numerical examples presented above, enabling a reliable assessment of the uncertainty in the quantity of interest. Furthermore, provided an accurate discretization of the parameter space is achieved, it has proven to outperform stochastic collocation methods when the response surface lacks smoothness. The reduced-basis method for uncertainty propagation retains the robustness of classical MonteCarlo methods, with a significant reduction in computational cost.

For the noncoercive case the results are also satisfying, although the stochastic collocation method provides accurate solutions due to regularity of the response surface. The main gain is in function evaluations as we increase dimensionality (although the greedy iterations are more expensive). For instance, the test case with 16 dimensions involves a number of full model solves for the reduced-basis which is only 0.37% of the total solves needed for stochastic collocation, using the same isotropic sparse grid.

## 5.2 Future Research

The work presented here is just the first step towards a more generalized and rigorous reduced-basis uncertainty propagation algorithm using HDG simulations. In

the immediate future, there is a need to incorporate rigorous and sharp *a posteriori* error estimators for the wave propagation problem. The lack of estimation of a lower bound of the inf-sup stability constant has rendered a method that is not self-stopping. Instead, it needs a set of precomputed solutions to assess the reliability of the reduced-basis, hence losing competitiveness. Furthermore, for examples where we have resonances in the domain, estimation of  $\widehat{\beta}(\boldsymbol{\xi})$  is of vital importance, since it becomes close to zero near resonances. For this kind of problems the *a posteriori* error estimators need to be accurate, otherwise the greedy algorithm might fail to recognize the most predominant information in the problem. Techniques such as the Successive Constraint Method (SCM) and the discrete eigenvalue problem are the most common approaches in the reduced-basis setting to estimate the lower bound of the stability constant. Finally, we want to study the wave propagation problem on a high-frequency setting, which becomes more difficult to treat with reduced-basis. High-frequency wave phenomena are challenging because the system itself has more information and a more complicated structure, hence a larger basis is needed to capture this information.

Furthermore, the work developed in this document assumes an affine parametric dependency, which is quite a restrictive assumption. It would therefore be interesting to incorporate nonaffine parametric dependency to the reduced-basis uncertainty propagation, and to assess the competitiveness with respect to stochastic collocation methods. Moreover, extension to nonlinear PDEs is much more challenging, especially for reduced-basis, and would also be an interesting future research line.

On another level, there is the crucial issue of how to select the points for the greedy algorithm. Although the offline-online strategy simplifies the evaluation of the error estimators, as the number of candidates grows the greedy search can become infeasible (although it can be parallelized). The use of sparse grids helps alleviate the explosion of points when dealing with high dimensions, but it is not an optimal choice. The fixed structure of sparse grids and its alignment along the axes may be a major difficulty whenever the response surface presents coupling between dimensions or

very localized information. For the low-moderate dimensional cases presented here, the computational effort is bearable, but if we need a not-that-coarse sparse grid on a very high dimensional space the curse of dimensionality makes its appearance. Stochastic collocation methods are usually applied on anisotropic or adaptive sparse grids to partially overcome the curse of dimensionality whenever a certain structure can be detected in the space. Much research has been devoted to develop efficient algorithms to adapt the sparse grids, a procedure that is usually done on-the-fly, since it is very difficult to estimate the relevant dimensions *a priori*. To test the competitiveness of the reduced-basis uncertainty propagation comparisons should be made with stochastic collocation methods on adaptive sparse grids. For very high-dimensional cases where many dimensions are not relevant, anisotropic sparse grids will most likely beat the reduced-basis if no adaptation is performed for the greedy.

Nonetheless, one of the most important differences between reduced-basis for uncertainty propagation and stochastic collocation is whether the points serve interpolatory purposes or not. The fact that for reduced-basis we just need a set of 'meaningful' points to feed the greedy gives us much more freedom. For a high-dimensional space where sparse grids may fail, we can resort to random or stratified sampling on the high-dimensional space, by prescribing the allowed number of candidates for the greedy subject to a certain computational budget. However, this does not solve the problem of how to pick the points in high dimensions. Recent work by Narayan et.al. based on Leja sequences and least orthogonal interpolation sheds some light onto the choice of points in high dimensions, and it may be an interesting path to explore.

A different approach that could potentially be useful would be to perform a greedy search on a continuous space. This setting is very interesting, since the limitation to a prescribed set of points vanishes. However, the response surface is in general nasty, with a lot of local minima and therefore difficult to optimize. Instead, we might resort to less ambitious but hopefully yet effective approaches, for example

$$\xi_{M+1} = \arg \max_{\xi \in \Xi} |s_{M+P} - s_M| \quad (5.1)$$

where  $s_{M+P}$  is the output computed by a collateral reduced-basis of dimension  $M+P$  that contains the original reduced-basis of dimension  $M$ . Note that derivatives are trivial to compute in a reduced-basis setting, hence problem (5.1) would be inexpensive to solve.



# Bibliography

- [1] I. Babuška and P. Chatzipantelidis. On solving elliptic stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 191:4093–4122, 2002.
- [2] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45(3):1005–1034, 2007.
- [3] I. Babuška, R. Tempone, and G. E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM Journal on Numerical Analysis*, 42(2):800–825, 2004.
- [4] I. Babuška, R. Tempone, and G. E. Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Computer methods in applied mechanics and engineering*, 194(12):1251–1294, 2005.
- [5] Zhaojun Bai. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics*, 43(1):9–44, 2002.
- [6] Maxime Barrault, Yvon Maday, Ngoc Cuong Nguyen, and Anthony T Patera. An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, 2004.
- [7] V. Barthelmann, E. Novak, and K. Ritter. High dimensional polynomial interpolation on sparse grids. *Adv. Compu. Math.*, 12:273–288, 2000.
- [8] Gal Berkooz, Philip Holmes, and John L Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics*, 25(1):539–575, 1993.
- [9] Sébastien Boyaval. A fast monte-carlo method with a reduced basis of control variates applied to uncertainty propagation and bayesian estimation. *Computer Methods in Applied Mechanics and Engineering*, 2012.
- [10] Sébastien Boyaval, Claude Le Bris, Yvon Maday, Ngoc Cuong Nguyen, and Anthony T Patera. A reduced basis approach for variational problems with stochastic parameters: Application to heat conduction with variable robin coefficient.

- Computer Methods in Applied Mechanics and Engineering*, 198(41):3187–3206, 2009.
- [11] Sébastien Boyaval, Claude Le Bris, Tony Lelièvre, Yvon Maday, Ngoc Cuong Nguyen, and Anthony T Patera. Reduced basis techniques for stochastic problems. *Archives of Computational methods in Engineering*, 17(4):435–454, 2010.
- [12] L Brutman. Lebesgue functions for polynomial interpolation—a survey.(english summary). *Ann. Numer. Math*, 4(1-4):111–127, 1997.
- [13] Tan Bui-Thanh, Karen Willcox, and Omar Ghattas. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM Journal on Scientific Computing*, 30(6):3270–3288, 2008.
- [14] R. Caflisch. Monte carlo and quasi-monte carlo methods. *Acta numerica*, 1998:1–49, 1998.
- [15] Y. Cao, M. Y. Hussaini, and T. Zang. Exploitation of sensitivity derivatives for improving sampling methods. *AIAA journal*, 42(4):815–822, 2004.
- [16] Y. Cao, M. Y. Hussaini, T. Zang, and A. Zatezalo. A variance reduction method based on sensitivity derivatives. *Applied numerical mathematics*, 56(6):800–813, 2006.
- [17] Q.-Y. Chen, D. Gottlieb, and J. Hesthaven. Uncertainty analysis for the steady-state flows in a dual throat nozzle. *J. Comp. Physics*, 204:387–398, 2005.
- [18] Y Chen, J White, et al. A quadratic method for nonlinear model order reduction. 2000.
- [19] B. Cockburn, J. Gopalakrishnan, and R. Lazarov. Unified hybridization of discontinuous galerkin, mixed, and continuous galerkin methods for second order elliptic problems. *SIAM Journal on Numerical Analysis*, 47(2):1319–1365, 2009.
- [20] M. K. Deb, I. M. Babuška, and J. T. Oden. Solution of stochastic partial differential equations using galerkin finite element techniques. *Computer Methods in Applied Mechanics and Engineering*, 190(48):6359–6372, 2001.
- [21] B. Debusschere, H.N. Najm, P. Pébay, O. Knio, R. Ghanem, and O. P. Le Maître. Numerical challenges in the use of polynomial chaos representations for stochastic processes. *SIAM Journal on Scientific Computing*, 26(2):698–719, 2004.
- [22] BI Epureanu, EH Dowell, and KC Hall. A parametric analysis of reduced order models of potential flows in turbomachinery using proper orthogonal decomposition. In *Proceedings of ASME turbo expo*, volume 2001, 2001.
- [23] JP Fink and WC Rheinboldt. On the error behavior of the reduced basis technique for nonlinear finite element approximations. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 63(1):21–28, 1983.



- [24] G. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer-Verlag, New York, 1996.
- [25] B. Fox. *Strategies for Quasi-Monte Carlo*, volume 22. Springer, 1999.
- [26] P. Frauenfelder, C. Schwab, and R. A. Todor. Finite elements for elliptic problems with stochastic coefficients. *Computer Methods in Applied Mechanics and Engineering*, 194:205–228, 2005.
- [27] D. Gamerman. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman Hall, London, 1997.
- [28] B. Ganapathysubramanian and N. Zabaras. Sparse grid collocation schemes for stochastic natural convection problems. *J. Comp. Physics*, 225:652–685, 2007.
- [29] V. Garg. *Coupled Flow Systems, Adjoint Techniques and Uncertainty Quantification*. PhD thesis, University of Texas at Austin, August 2012.
- [30] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numerical Algorithms*, 18:209–232, 1998.
- [31] T. Gerstner and M. Griebel. Dimension adaptive tensor product quadrature. *Computing*, 71:65–87, 2003.
- [32] R. Ghanem. Probabilistic characterization of transport in heterogeneous media. *Computer Methods in Applied Mechanics and Engineering*, 158(3):199–220, 1998.
- [33] R. Ghanem. Higher-order sensitivity of heat conduction problems to random data using the spectral stochastic finite element method. *Am. J. Heat Transfer*, 121:290–298, 1999.
- [34] R. Ghanem. Ingredients for a general purpose stochastic finite elements implementation. *Computer Methods in Applied Mechanics and Engineering*, 168(1):19–34, 1999.
- [35] R. Ghanem and A. Sarkar. Mid-frequency structural dynamics with parameter uncertainty. *Computer Methods in Applied Mechanics and Engineering*, 191(47):5499–5513, 2002.
- [36] R. Ghanem and P.D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, 1991.
- [37] P. Glynn and D. Iglehart. Importance sampling for stochastic simulations. *Management Science*, 35(11):1367–1392, 1989.
- [38] M Grepl. *Reduced-basis approximations for time-dependent partial differential equations: application to optimal control*. PhD thesis, Ph. D. thesis, Massachusetts Institute of Technology, 2005.
- [39] Martin A Grepl, Yvon Maday, Ngoc C Nguyen, and Anthony T Patera. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equa-

- tions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 41(03):575–605, 2007.
- [40] Martin A Grepl and Anthony T Patera. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 39(01):157–181, 2005.
- [41] M. Griebel. Adaptive sparse grid multilevel methods for elliptic pdes based on finite differences. *Computing*, 61:151–179, 1998.
- [42] G. Larcher H. Niederreiter, P. Hellekalek and P. Zinterhof. *Monte Carlo and Quasi-Monte Carlo methods*. Springer-Verlag, New York, 1996.
- [43] Bernard Haasdonk, Karsten Urban, and Bernhard Wieland. Reduced basis methods for parametrized partial differential equations with stochastic influences using the karhunen-loeve expansion. *Preprint, Ulm University*, 2012.
- [44] Kenneth C Hall, Jeffrey P Thomas, and Earl H Dowell. Proper orthogonal decomposition technique for transonic unsteady aerodynamic flows. *AIAA journal*, 38(10):1853–1862, 2000.
- [45] J. C. Helton and F. J. Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab. Engrg. System Safety*, 81:23–69, 2003.
- [46] Dinh Bao Phuong Huynh, Gianluigi Rozza, Sugata Sen, and Anthony T Patera. A successive constraint linear optimization method for lower bounds of parametric coercivity and inf–sup stability constants. *Comptes Rendus Mathematique*, 345(8):473–478, 2007.
- [47] K Ito and SS Ravindran. Reduced basis method for optimal control of unsteady viscous flows. *International Journal of Computational Fluid Dynamics*, 15(2):97–113, 2001.
- [48] Kazufumi Ito and SS Ravindran. A reduced-order method for simulation and control of fluid flows. *Journal of computational physics*, 143(2):403–425, 1998.
- [49] M. Kleiber and T. Hien. *The stochastic finite element method*. John Wiley, New York, 1992.
- [50] A. Klimke. *Uncertainty Modeling using Fuzzy Arithmetic and Sparse Grids*. PhD thesis, Universitat Stuttgart, Aachen, 2006.
- [51] P. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer-Verlag, New York, 1999.
- [52] ME Kowalski and J-M Jin. Karhunen-loève based model order reduction of nonlinear systems. In *Antennas and Propagation Society International Symposium, 2002. IEEE*, volume 2, pages 552–555. IEEE, 2002.

- [53] Guillaume Lassaux and K Willcox. Model reduction for active control design using multiple-point arnoldi methods. *AIAA Paper*, 616:2003, 2003.
- [54] OP Le Maître, OM Knio, HN Najm, and RG Ghanem. Uncertainty propagation using wiener-haar expansions. *Journal of Computational Physics*, 197(1):28–57, 2004.
- [55] OP Le Maître, HN Najm, RG Ghanem, and OM Knio. Multi-resolution analysis of wiener-type uncertainty propagation schemes. *Journal of Computational Physics*, 197(2):502–531, 2004.
- [56] W. Liu, T. Belytschko, and A. Mani. Probabilistic finite elements for nonlinear structural dynamics. *Computer Methods in Applied Mechanics and Engineering*, 23:1831–1845, 1986.
- [57] M. Loève. *Probability Theory I*, volume 45. Springer-Verlag, New York, 4th edition, 1977.
- [58] M. Loève. *Probability Theory II*, volume 46. Springer-Verlag, New York, 4th edition, 1978.
- [59] W. Loh. On latin hypercube sampling. *Ann. Statist.*, 24:2058–2080, 1996.
- [60] Hung V Ly and Hien T Tran. Modeling and control of physical processes using proper orthogonal decomposition. *Mathematical and computer modelling*, 33(1):223–236, 2001.
- [61] X. Ma and N. Zabararas. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *Journal of Computational Physics*, 228(8):3084–3113, 2009.
- [62] L Machiels, Y Maday, and AT Patera. Output bounds for reduced-order approximations of elliptic partial differential equations. *Computer methods in applied mechanics and engineering*, 190(26):3413–3426, 2001.
- [63] Luc Machiels, Yvon Maday, Ivan B Oliveira, Anthony T Patera, and Dimitrios V Rovas. Output bounds for reduced-basis approximations of symmetric positive definite eigenvalue problems. *Comptes Rendus de l’Academie des Sciences Series I Mathematics*, 331(2):153–158, 2000.
- [64] Yvon Maday, Ngoc Cuong Nguyen, Anthony T Patera, and George SH Pau. A general, multipurpose interpolation procedure: the magic points. 2007.
- [65] Yvon Maday, Anthony T Patera, and Dimitrios V Rovas. A blackbox reduced-basis output bound method for noncoercive linear problems. *Studies in Mathematics and its applications*, 31:533–569, 2002.
- [66] Yvon Maday, Anthony T Patera, and Gabriel Turinici. Global a priori convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic partial differential equations. *Comptes Rendus Mathematique*, 335(3):289–294, 2002.

- [67] Yvon Maday and Einar M Ronquist. The reduced basis element method: application to a thermal fin problem. *SIAM Journal on Scientific Computing*, 26(1):240–258, 2004.
- [68] N. Madras. *Lectures on Monte Carlo methods*, 2002.
- [69] L. Mathelin and M. Y. Hussaini. A stochastic collocation algorithm for uncertainty analysis. Technical report, NASA/CR-2003-0212153, NASA Langley Research Center, 2003.
- [70] L. Mathelin, M. Y. Hussaini, and T. A. Zang. Stochastic approaches to uncertainty quantification in cfd simulations. *Numerical Algorithms*, 38:209–236, 2005.
- [71] L. Mathelin and O. Le Maître. Dual-based a posteriori error estimation for stochastic finite element methods. *Commun. Appl. Math. Comput. Sci.*, 2(1):83–115, 2007.
- [72] H. G. Matthies and A. Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 194:1295–1331, 2005.
- [73] Marcus Meyer and Hermann G Matthies. Efficient model reduction in nonlinear dynamics using the karhunen-loeve expansion and dual-weighted-residual methods. *Computational Mechanics*, 31(1-2):179–191, 2003.
- [74] M. Motamed, F. Nobile, and R. Tempone. A stochastic collocation method for the second order wave equation with a discontinuous random speed. *Numerische Mathematik*, pages 1–44, 2011.
- [75] M. Motamed, F. Nobile, and R. Tempone. Analysis and computation of the elastic wave equation with random coefficients. 2012.
- [76] N. C. Nguyen. *Reduced-Basis Approximations and A Posteriori Error Bounds for Nonaffine and Nonlinear Partial Differential Equations: Application to Inverse Reduced-Basis Approximations and A Posteriori Error Bounds for Nonaffine and Nonlinear Partial Differential Equations: Application to Inverse Analysis*. PhD thesis, Singapore-MIT Alliance, June 2005.
- [77] N. C. Nguyen and J. Peraire. Hybridizable discontinuous galerkin methods for partial differential equations in continuum mechanics. *J. Comput. Phys.*, 231(18):5955–5988, July 2012.
- [78] N. C. Nguyen and J. Peraire. A shock-capturing hdg method for cfd applications. *J. Comp. Physics*, Working paper.
- [79] N. C. Nguyen, J. Peraire, and B. Cockburn. An implicit high-order hybridizable discontinuous galerkin method for linear convection-diffusion equations. *J. Comput. Phys.*, 228(9):3232–3254, 2009.

- [80] N. C. Nguyen, J. Peraire, and B. Cockburn. An implicit high-order hybridizable discontinuous galerkin method for nonlinear convection-diffusion equations. *J. Comput. Phys.*, 228(23):8841–8855, 2009.
- [81] N. C. Nguyen, J. Peraire, and B. Cockburn. A comparison of hdg methods for stokes flow. *J. Sci. Comput.*, 45(1-3):215–237, 2010.
- [82] N. C. Nguyen, J. Peraire, and B. Cockburn. A hybridizable discontinuous galerkin method for stokes flow. *Computer Methods in Applied Mechanics and Engineering*, 199(9-12):582–597, 2010.
- [83] N. C. Nguyen, J. Peraire, and B. Cockburn. A hybridizable discontinuous galerkin method for the compressible euler and navier-stokes equations. In *AIAA Paper*, volume 363, 2010.
- [84] N. C. Nguyen, J. Peraire, and B. Cockburn. A hybridizable discontinuous galerkin method for the incompressible navier-stokes equations. In *AIAA Paper*, editor, *Proceedings of the 48th AIAA Aerospace Sciences Meeting and Exhibit, Orlando, Florida, Paper*, volume 2010-362, 2010.
- [85] N. C. Nguyen, J. Peraire, and B. Cockburn. High-order implicit hybridizable discontinuous galerkin methods for acoustics and elastodynamics. *J. Comp. Physics*, 230(10):3695–3718, 2011.
- [86] N. C. Nguyen, J. Peraire, and B. Cockburn. Hybridizable discontinuous galerkin methods. *Spectral and High Order Methods for Partial Differential Equations*, pages 63–84, 2011.
- [87] N. C. Nguyen, J. Peraire, and B. Cockburn. Hybridizable discontinuous galerkin methods for the time-harmonic maxwell’s equations. *J. Comp. Physics*, 230(19):7151–7175, 2011.
- [88] N. C. Nguyen, J. Peraire, and B. Cockburn. An implicit high-order hybridizable discontinuous galerkin method for the incompressible navier-stokes equations. *J. Comp. Physics*, 230(4):1147–1170, 2011.
- [89] NC Nguyen, AT Patera, and J Peraire. A ‘best points’ interpolation method for efficient approximation of parametrized functions. *International journal for numerical methods in engineering*, 73(4):521–543, 2008.
- [90] H. Niederreiter. *Random number generation and Quasi-Monte Carlo methods*. SIAM, Philadelphia, 1992.
- [91] F. Nobile and R. Tempone. Analysis and implementation issues for the numerical approximation of parabolic equations with random coefficients. *International journal for numerical methods in engineering*, 80(6-7):979–1006, 2009.
- [92] F. Nobile, R. Tempone, and C. Webster. An anisotropic sparse grid collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2411–2442, 2008.

- [93] F. Nobile, R. Tempone, and C. Webster. A sparse grid collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.
- [94] Ahmed K Noor and Jeanne M Peters. Reduced basis technique for nonlinear analysis of structures. *AIAA Journal*, 18(4):455–462, 1980.
- [95] Ahmed K Noor and Jeanne M Peters. Multiple-parameter reduced basis technique for bifurcation and post-buckling analyses of composite plates. *International journal for numerical methods in engineering*, 19(12):1783–1803, 1983.
- [96] E. Novak and K. Ritter. High dimensional integration of smooth functions over cubes. *Numer. Math.*, 75:79–97, 1996.
- [97] E. Novak and K. Ritter. The curse of dimension and a universal method for numerical integration. *International Series of Numerical Mathematics*, pages 177–188, 1997.
- [98] E. Novak, K. Ritter, R. Schmitt, and A. Steinbauer. On an interpolatory method for high dimensional integration. *J. Comp. Appl. Math*, 112:215–228, 1999.
- [99] B. Oksendal. *Stochastic Differential Equations. An Introduction with Applications*. Springer-Verlag, New York, 5 edition, 1998.
- [100] Joel R Phillips. Projection-based approaches for model reduction of weakly nonlinear, time-varying systems. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 22(2):171–187, 2003.
- [101] TA Porsching. Estimation of the error in the reduced basis method solution of nonlinear equations. *Mathematics of Computation*, 45(172):487–496, 1985.
- [102] C Prud’homme, D Rovas, Y Maday, AT Patera, and G Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. 2002.
- [103] SS Ravindran. A reduced-order approach for optimal control of fluids using proper orthogonal decomposition. *International journal for numerical methods in fluids*, 34(5):425–448, 2000.
- [104] L. J. Roman and M. Sarkis. Stochastic galerkin method for elliptic spdes: A white noise approach. *Discrete Contin. Dyn. Syst. Ser. B*, 6:941–955, 2006.
- [105] Dimitrios Vasileios Rovas. *Reduced-basis output bound methods for parametrized partial differential equations*. PhD thesis, Massachusetts Institute of Technology, 2003.
- [106] J. Saà. Simulation and design optimization for linear wave phenomena on metamaterials. Master’s thesis, MIT, Cambridge, June 2011.
- [107] Lawrence Sirovich. Turbulence and the dynamics of coherent structures. i-coherent structures. ii-symmetries and transformations. iii-dynamics and scaling. *Quarterly of applied mathematics*, 45:561–571, 1987.

- [108] S. Smolyak. Quadrature and interpolation formulas for tensor product of certain classes of functions. *Soviet Math. Dokl.*, 4:240–243, 1963.
- [109] M. Stein. Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29:143–151, 1987.
- [110] L. N. Trefethen. Is gauss quadrature better than clenshaw-curtis? *SIAM Rev.*, 50:67–87, 2008.
- [111] K Veroy and AT Patera. Certified real-time solution of the parametrized steady incompressible navier–stokes equations: rigorous reduced-basis a posteriori error bounds. *International Journal for Numerical Methods in Fluids*, 47(8-9):773–788, 2005.
- [112] Karen Veroy. *Reduced-basis methods applied to problems in elasticity: Analysis and applications*. PhD thesis, Massachusetts Institute of Technology, 2003.
- [113] Karen Veroy, Christophe Prud’homme, and Anthony T Patera. Reduced-basis approximation of the viscous burgers equation: rigorous a posteriori error bounds. *Comptes Rendus Mathematique*, 337(9):619–624, 2003.
- [114] Karen Veroy, Christophe Prud’homme, DV Rovas, and AT Patera. A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations. In *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference*, volume 3847, 2003.
- [115] Karen Veroy, Dimitrios V Rovas, and Anthony T Patera. A posteriori error estimation for reduced-basis approximation of parametrized elliptic coercive partial differential equations: “convex inverse” bound conditioners. *ESAIM: Control, Optimisation and Calculus of Variations*, 8:1007–1028, 2002.
- [116] X. Wan and G. E. Karniadakis. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *J. Comp. Physics*, 209:617–642, 2005.
- [117] X. Wan and G. E. Karniadakis. Beyond wiener-asker expansions: Handling arbitrary pdfs. *J. Sci. Comput.*, 27:455–464, 2006.
- [118] N. Wiener. The homogeneous chaos. *Am. J. Math.*, 60:897–936, 1938.
- [119] Karen Willcox and Jaime Peraire. Balanced model reduction via the proper orthogonal decomposition. *AIAA journal*, 40(11):2323–2330, 2002.
- [120] Karen Willcox, Jaime Peraire, and James D Paduano. Application of model order reduction to compressor aeroelastic models. *Journal of Engineering for Gas Turbines and Power(Transactions of the ASME)*, 124(2):332–339, 2002.
- [121] Karen Willcox, Jaime Peraire, and Jacob White. An arnoldi approach for generation of reduced-order models for turbomachinery. *Computers & fluids*, 31(3):369–389, 2002.

- [122] D. Xiu. Efficient collocational approach for parametric uncertainty analysis. *Communications in computational physics*, 2(2):293–309, 2007.
- [123] D. Xiu and J. S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27:1118–1139, 2005.
- [124] D. Xiu and G. E. Karniadakis. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Computer Methods in Applied Mechanics and Engineering*, 191(43):4927–4948, 2002.
- [125] D. Xiu and G. E. Karniadakis. The wiener-asky polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.
- [126] D. Xiu and G. E. Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics*, 187(1):137–167, 2003.
- [127] D. Xiu and G. E. Karniadakis. A new stochastic approach to transient heat conduction modeling with uncertainty. *International Journal of Heat and Mass Transfer*, 46(24):4681–4693, 2003.
- [128] D. Xiu, D. Lucor, C-H. Su, and G. E. Karniadakis. Performance evaluation of generalized polynomial chaos. *Computational Science—ICCS 2003*, pages 723–723, 2003.